Considerations for the interdisciplinary development of environmental system models

by

Takuya Iwanaga

Submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy
of the Australian National University
June 2021

Candidate's Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university. To the best of the author's knowledge, it contains no material previously published or written by another person, except where due reference is made in the text.

Takuya Iwanaga Date: 07/01/2021

Included Publications

Details of each publication included in this thesis, as well additional citable works, are provided below. Statement of Contributions detailing the extent of contribution and endorsement for inclusion in this thesis are included in Appendix 1.

Chapter 2:

Douglas-Smith, D., **Iwanaga, T.***, Croke, B.F.W., Jakeman, A.J., 2020. Certain trends in uncertainty and sensitivity analysis: An overview of software tools and techniques. Environmental Modelling & Software 124, 104588. https://doi.org/10.1016/j.envsoft.2019.104588

Software:

Iwanaga, T., Douglas-Smith, D., 2019. Wosis: beta-release (v0.1.3). https://github.com/ConnectedSystems/Wosis. Zenodo. doi: 10.5281/zenodo.3406947

Published data and analysis:

Douglas-Smith, D., Iwanaga, T., 2019. UASA Trends. https://github.com/frog7/uasa-trends. doi: 10.5281/zenodo.3406946

Chapter 3:

Iwanaga, T.*, Zare, F., Croke, B., Fu, B., Merritt, W., Partington, D., Ticehurst, J., Jakeman, A., 2018. Development of an integrated model for the Campaspe catchment: a tool to help improve understanding of the interaction between society, policy, farming decision, ecology, hydrology and climate, in: Proceedings of the International Association of Hydrological Sciences. Presented at the Innovative water resources management - understanding and balancing interactions between humankind and nature - 8th International Water Resources Management Conference of ICWRS, Beijing, China, 13-15 June 2018, Copernicus GmbH, pp. 1–12. https://doi.org/10.5194/piahs-379-1-2018

Chapter 4:

Iwanaga, T.*, Partington, D., Ticehurst, J., Croke, B.F.W., Jakeman, A.J., 2020. A socio-environmental model for exploring sustainable water management futures: Participatory and collaborative modelling in the Lower Campaspe catchment. Journal of Hydrology: Regional Studies 28, 100669. https://doi.org/10.1016/j.ejrh.2020.100669

Chapter 5:

Iwanaga, T.*, Sun, X., Wang, Q., Guillaume, J.H.A., Croke, B.F.W., Rahman, J., Jakeman, A.J., 2021. Property-based Sensitivity Analysis: An approach to identify model implementation and integration errors. Environmental Modelling and Software 139, 105013. https://doi.org/10.1016/j.envsoft.2021.105013

Published data and analysis:

Iwanaga, T., 2020. diagnostic-sa. https://doi.org/10.17605/OSF.IO/NUZAF

Iwanaga, T., Sun, X., Wang, Q., 2020. oat-diagnostic. OSF. https://doi.org/10.17605/OSF.IO/85EDC

Chapter 6:

Iwanaga, T.*, Wang, H.-H., Hamilton, S.H., Grimm, V., Koralewski, T.E., Salado, A., Elsawah, S., Razavi, S., Yang, J., Glynn, P., Badham, J., Voinov, A., Chen, M., Grant, W.E., Peterson, T.R., Frank, K., Shenk, G., Barton, C.M., Jakeman, A.J., Little, J.C., 2021. Socio-technical scales in socio-environmental modeling: managing a system-of-systems modeling approach. Environmental Modelling & Software 104885. https://doi.org/10.1016/j.envsoft.2020.104885

Chapter 7:

Iwanaga, T.*, Wang, H-H., Koralewski, T.E., Jakeman A.J., Little, J.C., 2021. Towards a complete interdisciplinary treatment of scale: reflexive lessons from socio-environmental systems modeling. Elementa: Science of the Anthropocene 9(1). https://doi.org/10.1525/elementa.2020.00182

Addendum:

Koo, H., **Iwanaga, T.**, Croke, B.F.W., Jakeman, A.J., Yang, J., Wang, H-H., Sun, X., Lü, G., Li, X., Yue, T., Yuan, W., Liu, X., Chen, M.*, 2020. Sensitivity analysis of spatially distributed environmental models - a pragmatic framework for the exploration of uncertainty, Environmental Modelling and Software 134, 104857. https://doi.org/10.1016/j.envsoft.2020.104857

Acknowledgements

Those who have undertaken a PhD will know of the sardonic saying that goes “A good dissertation is a completed dissertation. A great dissertation is a published dissertation. A perfect one is neither”. I would thank my supervisors, Tony Jakeman, Barry Croke, and Joel Rahman for sharing their wisdom and support whenever needed which helped to ensure a published dissertation. Their experience and leadership enabled my involvement with several publications, not all included as a part of this thesis, and opportunities to engage and collaborate with colleagues from around Australia and the world. My thanks also to Prof. Shayne Flint for his early input and advice. I would be remiss if I did not take this opportunity to state my appreciation and thanks to Drs. Daniel Partington, Hsiao-Hsuan “Rose” Wang, Tomasz Koralewski, Joseph Guillaume, William E. Grant, Volker Grimm, Jin Teng, Alexey Voinov, Saman Razavi, and John Little for their support, encouragement, and the interesting discussions had. Their advice and feedback were invaluable and a cornerstone to many a publication.

It is common knowledge that a PhD is not for everyone. Tenacity and determination are common attributes said to be necessary for those who take this path, but equally true is the importance of friendship that sustains and motivates one to push forward. My heartfelt thanks to my friends at the Fenner School who made lunch time discussions that much more enjoyable. Special thanks to Qian Wang and Meena Sritharan for their company and friendship through this journey, as well as their involvement in organizing the various social events at the Fenner School, which I know many appreciate and enjoy. Ciao also to friends made along the way including Mattia Amadio, Anna Pupi, Marko Kallio, Ignacio Fuentes, and Antonie Botes.

Last but not least, I would like to thank my family, most especially my ever-patient wife, Tamara, who supported me in my quest to obtain two letters before my name despite being not entirely sure what this whole thing is about. Special acknowledgements go to the cats – Chomjun, Kiki, and Adora – whose collective judgemental faces served to remind me that I have other responsibilities beyond the PhD, especially around feeding time.

Abstract

Effective decision making and policy development requires holistic consideration of the modelling context. This thesis explores how consideration of multiple disciplinary perspectives and concerns lead to an integrative and holistic model development process for the purpose of socio-environmental systems (SES) management. The research is presented through two frames: (1) Integrated Environmental Model (IEM) development through a System-of-Systems (SoS) approach, and (2) the socio-technical considerations within an interdisciplinary modelling process. This is accomplished by incorporating the perspectives of the modelling, systems engineering, and software development paradigms.

IEMs are developed for the purpose of integrating knowledge across the various disciplines involved, whereas traditional approaches focus on single systems within the SES, such as hydrology, economics, social dynamics, or climatic drivers. Use of IEMs allows for the consideration of the flow-on effects due to system changes and interaction, and how these may affect long-term SES behaviour. Pathways that are robust – i.e., lead to beneficial or desirable outcomes – under a range of plausible but uncertain conditions can then be identified and assessed.

An SoS approach to IEM development leverages the separate specialized disciplinary knowledge to enable parallel development of the multiple models that make up an IEM. An interconnected network of such models thus makes up an SoS model allowing consideration of higher-order effects. In practice, however, the method and approach used in the development of constituent models may influence integrated system behaviour once coupled.

The socio-technical modelling concerns within the SoS/SES modelling context, including the methods to assess and manage model validity, complexity, and uncertainty, with respect to model purpose and intended outcomes are explored through a series of publications. This thesis contributes to the growing body of knowledge through:

1. An expansive overview of the currently available software for model uncertainty and sensitivity analysis, and the techniques they encompass.
2. The development of an integrated environmental model for the Lower Campaspe catchment in North-Central Victoria, Australia. The model explores long-term implications of water management decisions and potential policy changes (primarily through an agricultural lens), including conjunctive use of surface and groundwater under a range of uncertain futures.
3. Demonstration of a property-based sensitivity analysis approach to model diagnostics that combines software testing and sensitivity analysis to validate model behaviour. The approach is useful as a first-pass screening tool. Failure to reproduce expected model behaviour indicates issues with the model to be corrected and avoids the necessity of more computationally demanding diagnostics.

4. A pragmatic step-by-step framework for the sensitivity analysis of spatially distributed environmental models.
5. Exploration and discussion of the modelling practices, issues and challenges that arise when dealing with the various influences and effects of scale within the interdisciplinary SoS context. The discussion adopts a socio-technical lens incorporating knowledge and experiences of 20 co-authors and calls for a grander vision for SoS-IEM modelling (and commensurate funding) to better enable interdisciplinary, and integrative, socio-environmental research to occur.
6. A shared reflexive account of two case studies that draws out the considerations and decisions regarding scale to arrive at five shared lessons learnt to foster an effective and interdisciplinary modelling process.

The key conclusion is the need for researchers involved in SoS modelling of SESs to actively consider and address cross-disciplinary concerns through improved interdisciplinary communication, documentation practices, and explicit consideration of the interplay between defined scales and resulting influence on uncertainty. Integrative consideration of these would then lower or avoid barriers that hamper the development and application of integrated environmental system models.

Table of Contents

List of acronyms and abbreviations

Glossary and Terms

Term	Definition
ANU	Australian National University
IEM	Integrated Environmental Model
IAM	Integrated Assessment Model
SoS	System of Systems
SES	Socio-Environmental System
ESM	Exploratory Scenario Modelling
SA	Sensitivity Analysis
UA	Uncertainty Analysis
OAT (SA)	One-at-a-Time sensitivity analysis
PbSA	Property-based sensitivity analysis
SDEM	Spatially Distributed Environmental Model

Chapter 1: Introduction

Modelling of socio-environmental systems is a complex endeavour given the multitude of social, technical, and natural systems involved. Models constructed to represent socio-environmental systems (SESs) are often used to inform long-term policy and management decisions which require careful, and continual, consideration and planning due to a high risk of unintended consequences. The act of addressing concerns may in turn exacerbate or be the cause of further issues, a concept known as "wicked problems" (Peters, 2017). Discovery and consideration of the flow-on effects due to system drivers as early as possible requires holistic consideration of the interlinked relationships between the systems at play; known as an integrated assessment (Risbey et al., 1996). These include considerations of the social, economic, ecological, agricultural, (geo)hydrological and climate systems as well as the systemic uncertainties and the engineering concerns that underlie the modelling (Laniak et al., 2013). Here the social system refers to the networks of people and their relationships involved in both the modelling processes (e.g. the modellers themselves) as well as those being modelled (e.g. the people "on the ground").

1.1 Environmental Software and Model Development

Model development is commonly described as undergoing a "cycle", in which multiple and often concurrent activities are conducted (Hamilton et al., 2015; Jakeman et al., 2006). To those with a software engineering background, the modelling cycle has parallels with the iterative spiral model of development (Boehm, 1986) in which, broadly speaking: (1) requirements towards meeting objectives are gathered; (2) the model designed; (3) prototype constructed; and (4) the process and/or results evaluated by stakeholders and risks analysed, where risk relates to sources of uncertainty that threaten successful project completion. Project cost is additionally recognized to cumulatively increase with each iteration. The process allows for earlier phases to be revisited, and that subsequent iterations build on previous efforts where feasible such that the lessons learnt and new knowledge and information are incorporated (Boehm, 1986; Jakeman et al., 2006).

Although there are now many alternative conceptualisations of iterative software development cycles (overviews may be found in Ambler, 2002; Ruparelia, 2010) in most, if not all, modern conceptualisations, it is recognised that the development process is not only iterative but can also be concurrently applied, and that input of stakeholders and experts are more often than not crucial to informing and validating the process. Once a minimum set of requirements are known, the subsequent phases of design, construction, and evaluation for each model component can begin in parallel, each with their own iterative processes, with stakeholder input considered throughout.

In environmental modelling, the practice of iterative and concurrent model development is coupled with the explicit consideration of the complexity and uncertainty inherent in environmental systems management (Jakeman et al., 2006; Kragt et al., 2013; Robson et al., 2008). Stakeholder knowledge is leveraged to verify the validity, appropriateness, and plausibility of the modelling particularly around the scenarios and parameter ranges considered – in a sense reducing model and scenario uncertainty – thereby improving the chances of achieving beneficial outcomes (Jakeman and Letcher, 2003; McIntosh et al., 2011; Voinov et al., 2016). The process introduced for environmental modelling in Jakeman et al. (2006) bridges the engineering (represented by the above described iterative and concurrent model development) with the social and scientific concerns to allow a holistic integrated environmental assessment and modelling to be conducted.

Development of environmental models often adopt two broad pathways: an integral or an integrated approach (Voinov and Shugart, 2013). An *integral* approach is characterised by the application of a single modelling paradigm (e.g., agent-based, systems dynamics, etc), and is often purpose-developed for a specific context with no further re-use envisioned. An *integrated* approach to model development may leverage existing models and computational frameworks, mix multiple modelling paradigms, and are commonly intended as general-purpose applications with the intentions of further re-use and development. The approach originates from Component-Based Software Engineering (cf. Vale et al., 2016) and treats individual scientific models as composable, and reusable, blocks allowing for quick development and application relative to starting development from scratch (Holzworth et al., 2015; Midingoyi et al., 2020; Whelan et al., 2014).

A further perspective from the field of systems engineering is the System-of-Systems (SoS) approach which has stemmed from the need to model the interactions between "sectors", for example the transport, electrical production, and agricultural systems (Nielsen et al., 2015). From the perspective of the system engineer, each constituent model is a separate and independently functioning entity that represents a specific system. Thus, a point of distinction between integrated models and SoS models is that the former does not imply that its components represent separate systems.

Connecting these individual models together, referred to as "coupling" or "integrating", for the purpose of integrated assessment results in an Integrated Environmental Model (IEM) that can represent the interactions between systems and their flow-on effects. Such IEMs can be regarded as an SoS model as each individual model is separable, independently functional, and represents individual systems within the SES. As these representations are computational in nature, environmental SoS models can therefore be characterised as a collection of interoperating software which represents the salient properties of the SES under investigation.

1.2 Considerations within an SoS approach

A shared concern amongst SoS practitioners is ensuring the validity of system representations under integrated contexts (Belete et al., 2017; Nielsen et al., 2015). As each constituent model operates independently the overall emergent behaviour of the SoS model is not prescribed (Kinder et al., 2012). Validity and plausibility of the emergent behaviour may be compromised if the conceptual linkages between systems are not appropriate, even if technical integration (e.g., data interoperation between models) is achieved (Nielsen et al., 2015; Wirtz and Nowak, 2017; Voinov and Shugart, 2013).

The situation is more complex compared to the single-system context given the independent and concurrent development cycle that each model, and its components, undergo. Decisions made in the development of one constituent model may affect the operation or validity of another due to the interactions between models. Each model is likely developed by a team of specialists with preferred (often disciplinary-specific) terminologies and language such that inter-team communication is made difficult. Moreover, each model developer may have different concerns to address and differing levels of resources allocated to accomplish their goal(s). Opportunities to ensure validity of both technical and conceptual integration may be scarce and may only be guaranteed for a model version several iterations behind. Resolving such issues requires that those involved in the modelling process must work in tandem to construct plausible (1) system representations, (2) representations of their interactions, as well as (3) behaviour of the overall SES being represented.

Once constructed, the behaviour of the SoS model can then be explored through an Exploratory Scenario Modelling (ESM) approach. The ESM approach involves the simulation of multiple scenarios, with each scenario representing a plausible "future". The overarching "theme" for scenarios should be co-developed with stakeholders (Mahmoud et al., 2009). Exploring the range of possible futures allows system vulnerabilities under the given range of scenarios to be assessed and the robustness and sensitivity of decision pathways to a range of (uncertain) conditions to be considered and described (Horne et al., 2019; Maier et al., 2016). Application of ESM with SoS models requires the consideration of three intertwined aspects:

- that model complexity is manageable for its purpose
- the represented relationship and interactions between systems are conceptually and technically valid, and
- that the model, system, and scenario uncertainties are adequately assessed or addressed, and communicated

The first, unwarranted complexity, can exacerbate the latter two, as excessive complexity interferes with the validation and verification of model behaviour (Nielsen et al., 2015), assessment of uncertainty (Tscheikner-Gratl et al., 2019), and consequently can compromise the interpretability and effectivity of the modelling (John et al., 2020; Voinov and Shugart, 2013). For exploratory purposes, technical implications of an unnecessarily complex (computational) model include increased runtimes, and a codebase that requires (significantly) more testing and maintenance. Conceptual implications include difficulties in interpreting and communicating results and ensuring validity of the interactions that the SoS model represents. These have the effect of impeding the overall development cycle.

Given SoS models are constructed through a mix of constituent models, both pre-existing and purpose-built, each constituent model may be complex in and of themselves. Valid and appropriate model behaviour in the *un*integrated context does not necessarily translate to the same when models are integrated. It is then challenging to determine if modelled representations of a system, when coupled, are valid and adequate for the given purpose. Determining correct functionality of constituent models, even in the decoupled context, may not be possible at all, or at least compromised, if the original model developers are not involved, or otherwise available. It is often the case that models are treated as black or, at best, grey boxes, and issues of appropriate model complexity for integration may not be addressed given the resources available.

Usual recommendations for managing model complexity include screening for unimportant model parameters through uncertainty and sensitivity analysis (Norton, 2009; Pianosi et al., 2016; Razavi et al. 2021) or the development of, and subsequent replacement with, surrogate models so as to have a simplified, but behaviourally similar, constituent model (Lam et al., 2020). Note that these two activities are not mutually exclusive. The nature of SoS modelling implies that its constituent models must be integrated before such activities are conducted as parameters that appear to be insensitive – and thus have little impact on model output – may in fact be highly sensitive depending on the scenarios being considered.

Complexity in SoS models are not completely addressable nor avoidable, resulting in uncertainty. High uncertainty, however, does not necessarily preclude effective decision making or policy development (Reichert and Borsuk, 2005). A necessity is that the treatment of uncertainty, and communication of its influence on model results, be outlined clearly lest the lessons learnt be lost or otherwise render the modelling ineffectual in resolving the issues of concern. In the context and general purpose of IEMs – that is, to provide information towards the development of policy and aid in decision making – insufficient attention to any one of the above will likely compromise the effectivity of the modelling for its purpose and in achieving intended outcomes. From a “big picture” perspective these serve as barriers that slow down development effective policies and actions (Arnold et al., 2020; Arnold et al., 2016). How these concerns of model complexity, validity, and uncertainty are managed with available resources then shapes the path forward.

1.3 Scope of thesis

Integrated modelling requires interdisciplinary involvement and collaboration to provide a holistic assessment of the socio-environmental system under study. This thesis explores how cohesive consideration of multiple disciplinary perspectives leads to an integrative IEM development process. These perspectives include not just the modelling concerns but additionally the systems engineering and software development that underpins modern computational science such as for environmental modelling. Consideration across these multiple perspectives can lower or avoid barriers that preclude successful management of model complexity, validity, and uncertainty. The bulk of the thesis revolves around a case study conducted in the Lower Campaspe catchment in North-Central Victoria, Australia. The case study serves as a springboard from which ideas and their implications are explored to clear a path towards holistic socio-environmental systems modelling.

The pathway taken represented as a sequence of publications is depicted in Figure 1. Not all publications depicted in this figure are included in this thesis but it serves to showcase the development of ideas and their inter-relation. Publications related to this thesis are marked with their respective chapter number. The primary modelling themes of each chapter are also outlined in Figure 2.

1.4 Thesis Outline

The structure of the thesis is as follows. In Chapter 2 the current state of uncertainty and sensitivity analysis (UA/SA) as applied to environmental research is explored through a hybrid bibliometric approach. The primary theme of the chapter are the approaches and methods for uncertainty and sensitivity analysis within the environmental sciences, with a focus on the software available to conduct such analyses.

A key contribution of Chapter 2 are descriptions of the current state of available UA/SA tooling for common programming environments used in the sciences, and an overview of needed improvements in terms of usability and accessibility to further increase uptake of such tooling. Common terms used in UA/SA research are described and defined therein. The chapter was published in the journal Environmental Modelling and Software as a review paper.

Chapter 3 describes the participatory development process undertaken for an IEM for the investigation of possible water management futures in the Lower Campaspe basin of North-Central Victoria, Australia. The model, then undergoing development was designed to examine the influence of hypothetical conjunctive water use policies and a changing climate to future water security, farm productivity, and the consequent impact on recreational and ecological outcomes. The chapter was published as a refereed conference paper in the Proceedings of the International Association of Hydrological Sciences (PIAHS) and serves to provide context for the following chapter.

Chapter 4 expands on the previous chapter with the application of the IEM through an exploratory scenario modelling approach. The paper details the choices made in the technical implementation and subsequent application of the model as well as the findings from the modelling and its management implications. The paper further details the complexities encountered in the development of an expansive IEM and associated model, system, and scenario uncertainties. The paper was published as a research article in the Journal of Hydrology: Regional

Studies. A supplement is included as Chapter 4b which addresses concerns raised regarding the calibration of the hydrological model during the thesis examination process.

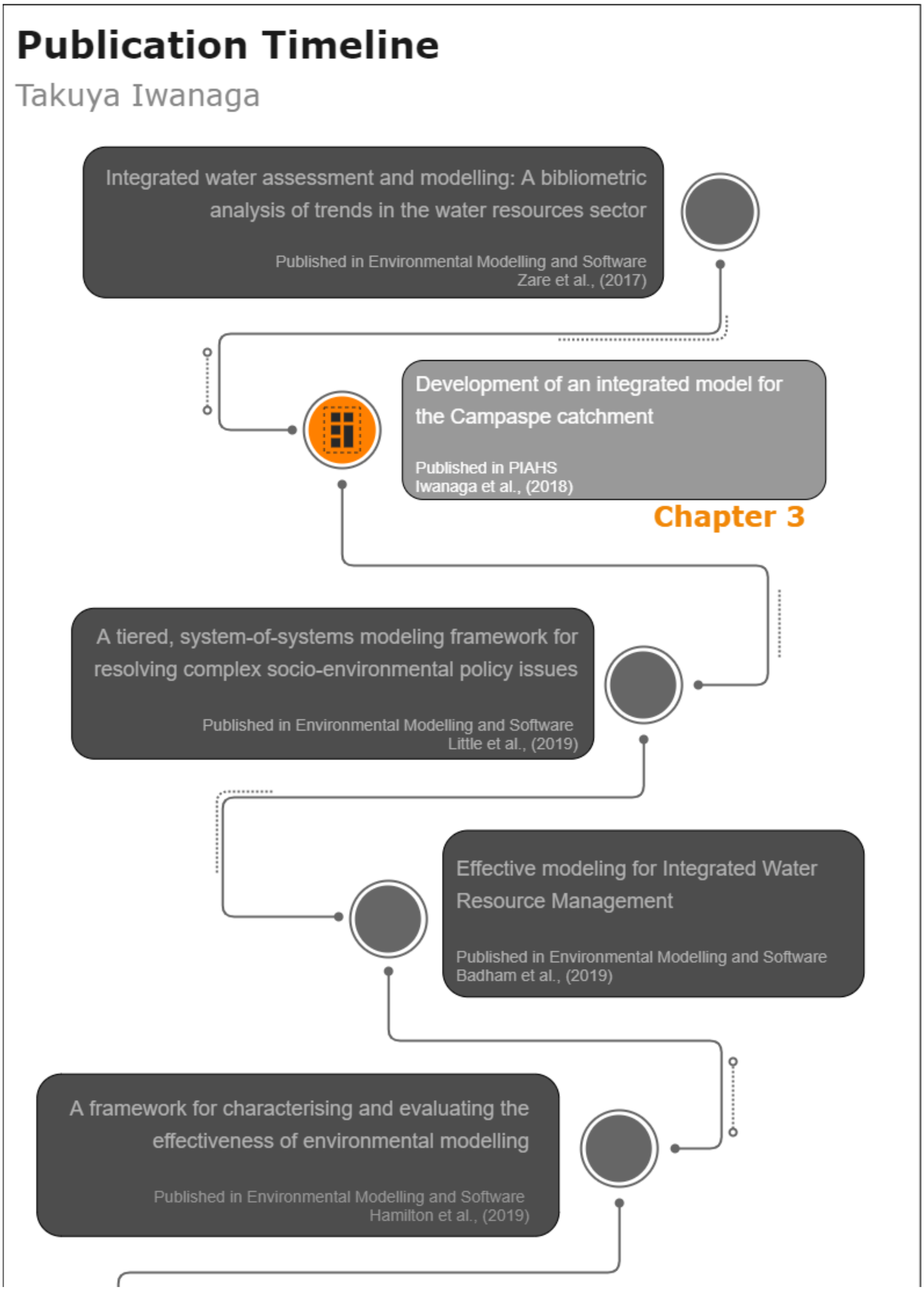
Chapter 5 illustrates a practical and effective diagnostic testing approach for complex IEMs that utilises property-based sensitivity analysis. The approach takes a software testing perspective of sensitivity analysis to aid in reducing the overall computational effort expended in the model development and validation process. The paper illustrates the use of a form of diagnostic screening referred to as “extremity testing” and, separately, uses an “activity threshold” to qualitatively confirm correctness of model behaviour. Such tests ensure the existence of known conceptual relationships between parameters and quantities of interest can be checked. It is intended to complement a more expansive global analyses by first confirming expected model behaviour within a restricted area of parameter space. The paper has been published in Environmental Modelling and Software.

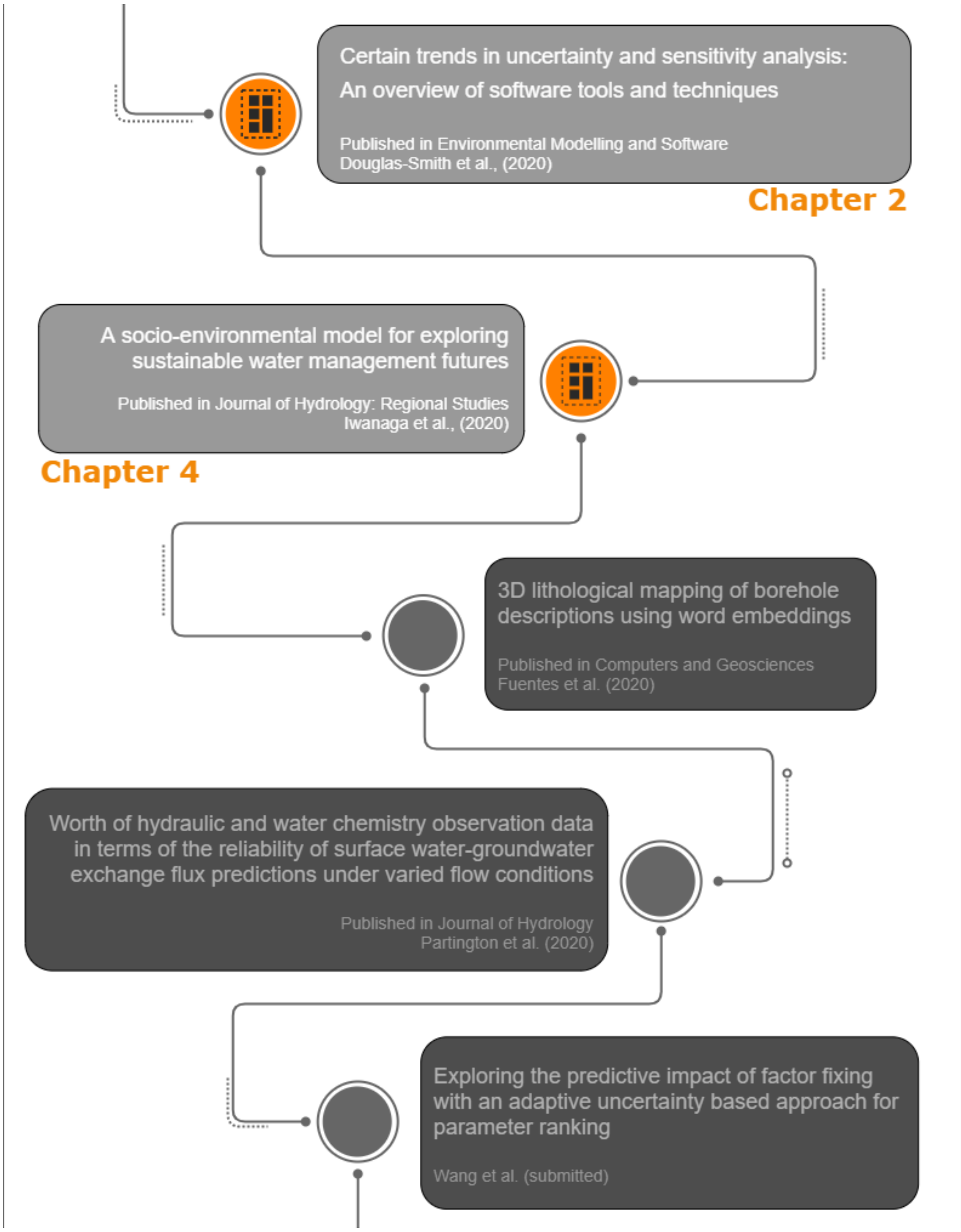
Chapter 6 takes an overarching view of the socio-technical scales involved in the interdisciplinary socio-environmental SoS modelling process. The practices and challenges that arise in the consideration of scale are explored. Here, “scale” is defined as the scope of work to be conducted in the treatment and representation of the system under investigation. The variety and disparity of models, modelers and their disciplinary perspectives then leads to issues of conceptual and technical mismatches and constrains the level of knowledge integration achieved. A key contribution here is the view that socio-technical scales encompass the interactions between the people involved and their choices in scale, not simply the technical decisions made. Such scale choices must be actively considered when developing a holistic system representation. Key paths forward are identified to resolve socio-technical scale issues. The paper is published in Environmental Modelling and Software.

Chapter 7 furthers the viewpoint introduced in Chapter 6 by outlining the process of considering scale in two separate integrated modelling case studies. The implications for future SoS modelling studies in practical terms are additionally explored and discussed. It is unusual in the sense that a reflexive approach is used to explore the experiences had and to elicit the lessons learnt. Such in-depth reflexive analyses are a rarity in the environmental sciences. The paper has been published in the open-access journal Elementa: Science of the Anthropocene.

Chapter 8 then concludes with a summary and overview of the contributions presented in this thesis.

An addendum to the thesis details a pragmatic step-by-step sensitivity analysis framework. The framework guides modellers in the assessment of the strength of potential sources of uncertainty with respect to spatially distributed environmental models (SDEMs), with a focus on spatial datasets. The paper was published as a Position Paper in Environmental Modelling and Software. The addendum was originally included as Chapter 6 but has since been appended instead to conform with Section 5 of ANU procedure 003405 (https://policies.anu.edu.au/ppl/document/ANUP_003405).




Certain trends in uncertainty and sensitivity analysis:
An overview of software tools and techniques
Published in Environmental Modelling and Software
Douglas-Smith et al., (2020)
Chapter 2
A socio-environmental model for exploring sustainable water management futures
Published in Journal of Hydrology: Regional Studies
Iwanaga et al., (2020)
Chapter 4
3D lithological mapping of borehole descriptions using word embeddings
Published in Computers and Geosciences
Fuentes et al. (2020)
Worth of hydraulic and water chemistry observation data in terms of the reliability of surface water-groundwater exchange flux predictions under varied flow conditions
Published in Journal of Hydrology
Partington et al. (2020)
Exploring the predictive impact of factor fixing with an adaptive uncertainty based approach for parameter ranking
Wang et al. (submitted)

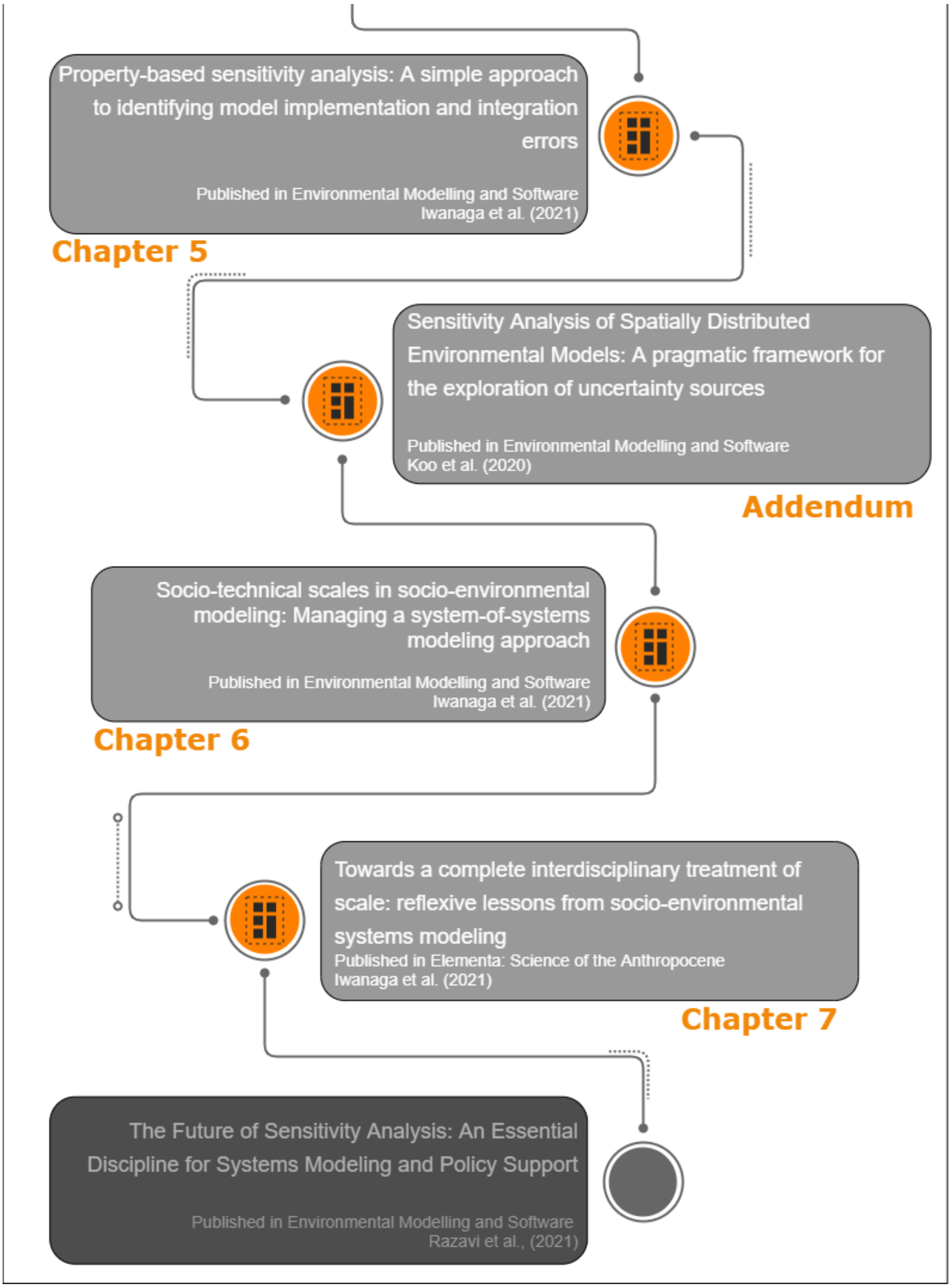


Figure 1. Publication timeline with thesis chapters indicated. Not all publications are included as part of this thesis. Those that are included are marked with their respective chapter numbers. Note that the chapters do not match publication order.

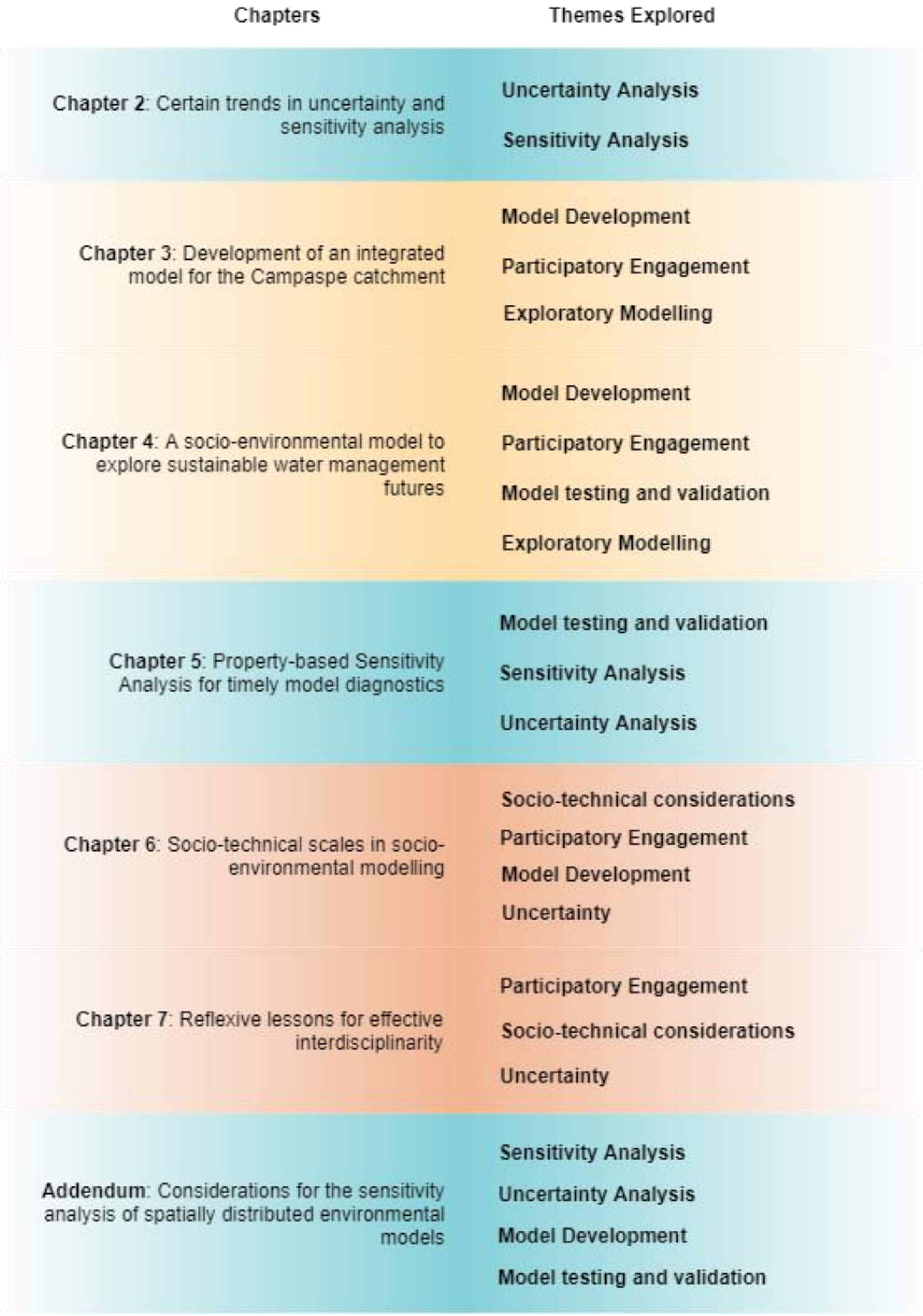


Figure 2. Each chapter can be associated with three core themes of sensitivity and uncertainty analysis (aqua blue; Chapters 2, 5, and the addendum), model development processes (light orange; Chapters 3 and 4), and the social and socio-technical factors in model development (red orange; Chapters 6 and 7).

1.5 References for Chapter 1

Ambler, S., 2002. Agile Modeling: Effective Practices for eXtreme Programming and the Unified Process. John Wiley & Sons, Inc., New York.

Arnold, T., 2016. Modelling for reproducibility, transparency, and continuous learning: Revisiting Jakeman's 10 steps.

Arnold, T., Guillaume, J.H.A., Lahtinen, T.J., Vervoort, R.W., 2020. From ad-hoc modelling to strategic infrastructure: A manifesto for model management. Environmental Modelling & Software 123, 104563. https://doi.org/10.1016/j.envsoft.2019.104563

Belete, G.F., Voinov, A., Laniak, G.F., 2017. An overview of the model integration process: From pre-integration assessment to testing. Environmental Modelling & Software 87, 49–63. https://doi.org/10.1016/j.envsoft.2016.10.013

Boehm, B., 1986. A spiral model of software development and enhancement. SIGSOFT Softw. Eng. Notes 11, 14–24. https://doi.org/10.1145/12944.12948

Dey, R., Lewis, S.C., Arblaster, J.M., Abram, N.J., 2019. A review of past and projected changes in Australia's rainfall. WIREs Climate Change 10, e577. https://doi.org/10.1002/wcc.577

Hamilton, S.H., ElSawah, S., Guillaume, J.H.A., Jakeman, A.J., Pierce, S.A., 2015. Integrated assessment and modelling: Overview and synthesis of salient dimensions. Environmental Modelling & Software 64, 215–229. https://doi.org/10.1016/j.envsoft.2014.12.005

Holzworth, D.P., Snow, V., Janssen, S., Athanasiadis, I.N., Donatelli, M., Hoogenboom, G., White, J.W., Thorburn, P., 2015. Agricultural production systems modelling and software: Current status and future prospects. Environmental Modelling & Software 72, 276–286. https://doi.org/10.1016/j.envsoft.2014.12.013

Horne, A.C., Nathan, R., Poff, N.L., Bond, N.R., Webb, J.A., Wang, J., John, A., 2019. Modeling Flow-Ecology Responses in the Anthropocene: Challenges for Sustainable Riverine Management. BioScience 69, 789–799. https://doi.org/10.1093/biosci/biz087

Jakeman, A.J., Letcher, R.A., 2003. Integrated assessment and modelling: features, principles and examples for catchment management. Environmental Modelling & Software, Applying Computer Research to Environmental Problems 18, 491–501. https://doi.org/10.1016/S1364-8152(03)00024-0

Jakeman, A.J., Letcher, R.A., Norton, J.P., 2006. Ten iterative steps in development and evaluation of environmental models. Environmental Modelling and Software 21, 602–614. https://doi.org/10.1016/j.envsoft.2006.01.004

John, A., Horne, A., Nathan, R., Stewardson, M., Webb, J.A., Wang, J., Poff, N.L., 2020. Climate change and freshwater ecology: Hydrological and ecological methods of comparable complexity are needed to predict risk. WIREs Climate Change n/a, e692. https://doi.org/10.1002/wcc.692

Kinder, A., Barot, V., Henshaw, M., Siemieniuch, C., 2012. System of Systems: "Defining the system of interest," in: 2012 7th International Conference on System of Systems Engineering (SoSE). Presented at the 2012 7th International Conference on System of Systems Engineering (SoSE), pp. 463–468. https://doi.org/10.1109/SYSoSE.2012.6384211

Lam, R.R., Zahm, O., Marzouk, Y.M., Willcox, K.E., 2020. Multifidelity Dimension Reduction via Active Subspaces. SIAM J. Sci. Comput. 42, A929–A956. https://doi.org/10.1137/18M1214123

Laniak, G.F., Olchin, G., Goodall, J., Voinov, A., Hill, M., Glynn, P., Whelan, G., Geller, G., Quinn, N., Blind, M., Peckham, S., Reaney, S., Gaber, N., Kennedy, R., Hughes, A., 2013. Integrated environmental modeling: A vision and roadmap for the future. Environmental Modelling and Software 39, 3–23. https://doi.org/10.1016/j.envsoft.2012.09.006

Mahmoud, M., Liu, Y., Hartmann, H., Stewart, S., Wagener, T., Semmens, D., Stewart, R., Gupta, H., Dominguez, D., Dominguez, F., Hulse, D., Letcher, R., Rashleigh, B., Smith, C., Street, R., Ticehurst, J., Twery, M., van Delden, H., Waldick, R., White, D., Winter, L., 2009. A formal framework for scenario development in support of environmental decision-making. Environmental Modelling & Software 24, 798–808. https://doi.org/10.1016/j.envsoft.2008.11.010

Maier, H.R., Guillaume, J.H.A., van Delden, H., Riddell, G.A., Haasnoot, M., Kwakkel, J.H., 2016. An uncertain future, deep uncertainty, scenarios, robustness and adaptation: How do they fit together? Environmental Modelling & Software 81, 154–164. https://doi.org/10.1016/j.envsoft.2016.03.014

McIntosh, B.S., Ascough, J.C., Twery, M., Chew, J., Elmahdi, A., Haase, D., Harou, J.J., Hepting, D., Cuddy, S., Jakeman, A.J., Chen, S., Kassahun, A., Lautenbach, S., Matthews, K., Merritt, W., Quinn, N.W.T., Rodriguez-Roda, I., Sieber, S., Stavenga, M., Sulis, A., Ticehurst, J., Volk, M., Wrobel, M., van Delden, H., El-Sawah, S., Rizzoli, A., Voinov, A., 2011. Environmental decision support systems (EDSS) development – Challenges and best practices. Environmental Modelling & Software 26, 1389–1402. https://doi.org/10.1016/j.envsoft.2011.09.009

Midingoyi, C.A., Pradal, C., Athanasiadis, I.N., Donatelli, M., Enders, A., Fumagalli, D., Garcia, F., Holzworth, D., Hoogenboom, G., Porter, C., Raynal, H., Thorburn, P., Martre, P., 2020. Reuse of process-based models: automatic transformation into many programming languages and simulation platforms. in silico Plants 2. https://doi.org/10.1093/insilicoplants/diaa007

Nielsen, C.B., Larsen, P.G., Fitzgerald, J., Woodcock, J., Peleska, J., 2015. Systems of Systems Engineering: Basic Concepts, Model-Based Techniques, and Research Directions. ACM Comput. Surv. 48, 18:1-18:41. https://doi.org/10.1145/2794381

Norton, J.P., 2009. Selection of Morris Trajectories for Initial Sensitivity Analysis. IFAC Proceedings Volumes, 15th IFAC Symposium on System Identification 42, 670–674. https://doi.org/10.3182/20090706-3-FR-2004.00111

Peters, B.G., 2017. What is so wicked about wicked problems? A conceptual analysis and a research program. Policy and Society 36, 385–396. https://doi.org/10.1080/14494035.2017.1361633

Pianosi, F., Beven, K., Freer, J., Hall, J.W., Rougier, J., Stephenson, D.B., Wagener, T., 2016. Sensitivity analysis of environmental models: A systematic review with practical workflow. Environmental Modelling & Software 79, 214–232. https://doi.org/10.1016/j.envsoft.2016.02.008

Reichert, P., Borsuk, M.E., 2005. Does high forecast uncertainty preclude effective decision support? Environmental Modelling & Software, Methods of Uncertainty Treatment in Environmental Models 20, 991–1001. https://doi.org/10.1016/j.envsoft.2004.10.005

Risbey, J., Kandlikar, M., Patwardhan, A., 1996. Assessing integrated assessments. Climatic Change 34, 369–395. https://doi.org/10.1007/BF00139298

Robson, B., Hamilton, D., Webster, I., Chan, T., 2008. Ten steps applied to development and evaluation of process-based biogeochemical models of estuaries. Environmental Modelling & Software 23, 369–384. https://doi.org/10.1016/j.envsoft.2007.05.019

Ruparelia, N.B., 2010. Software development lifecycle models. SIGSOFT Softw. Eng. Notes 35, 8–13. https://doi.org/10.1145/1764810.1764814

Tscheikner-Gratl, F., Bellos, V., Schellart, A., Moreno-Rodenas, A., Muthusamy, M., Langeveld, J., Clemens, F., Benedetti, L., Rico-Ramirez, M.A., de Carvalho, R.F., Breuer, L., Shucksmith, J., Heuvelink, G.B.M., Tait, S., 2019. Recent insights on uncertainties present in integrated catchment water quality modelling. Water Research 150, 368–379. https://doi.org/10.1016/j.watres.2018.11.079

Vale, T., Crnkovic, I., de Almeida, E.S., Silveira Neto, P.A. da M., Cavalcanti, Y.C., Meira, S.R. de L., 2016. Twenty-eight years of component-based software engineering. Journal of Systems and Software 111, 128–148. https://doi.org/10.1016/j.jss.2015.09.019

Voinov, A., Kolagani, N., McCall, M.K., Glynn, P.D., Kragt, M.E., Ostermann, F.O., Pierce, S.A., Ramu, P., 2016. Modelling with stakeholders - Next generation. Environ. Model. Softw. 77, 196–220. https://doi.org/10.1016/j.envsoft.2015.11.016

Voinov, A., Shugart, H.H., 2013. “Integronsters”, integral and integrated modeling. Environmental Modelling and Software 39, 149–158. https://doi.org/10.1016/j.envsoft.2012.05.014

Whelan, G., Kim, K., Pelton, M.A., Castleton, K.J., Laniak, G.F., Wolfe, K., Parmar, R., Babendreier, J., Galvin, M., 2014. Design of a component-based integrated environmental modeling framework. Environmental Modelling and Software 55, 1–24. https://doi.org/10.1016/j.envsoft.2014.01.016

Wirtz, D., Nowak, W., 2017. The rocky road to extended simulation frameworks covering uncertainty, inversion, optimization and control. Environmental Modelling and Software 93, 180–192. https://doi.org/10.1016/j.envsoft.2016.10.003

Chapter 2: Certain trends in uncertainty and sensitivity analysis

Assessment of uncertainty and sensitivity are regarded as crucial practices that aid in determining the level of (model) complexity that is warranted for a given purpose, and the level of trust one can have in model predictions. Yet, typical analyses have repeatedly been found not to be sufficiently expansive to fulfil this role. A higher degree of mathematical and statistical knowledge is required to conduct a sufficiently rigorous analysis. At the same time, software tools that ease the cognitive burden of more robust approaches are now available such that a lack of awareness of such tools may be one possible contributing factor to the current situation.

In this chapter the research trends regarding uncertainty analysis techniques and the software underpinning such work is examined through a hybrid bibliometric literature review. In this process we investigated the general research trends through the lens of environmental science and synthesised an overview of the available software tools, techniques, and their apparent uptake.

This chapter was peer-reviewed by three anonymous reviewers and published in Environmental Modelling and Software. The publication additionally introduced a purpose-built software package called "Wosis", for **W**eb **o**f **S**cience Analys**is,** to ease interaction and analysis of data obtained through the Web of Science platform (operated by Clarivate Analytics). The work is openly accessible, with the software used hosted in a publicly accessible code repository. Representative datasets and the code used for analysis are also made available in a separate code repository. The publication itself is open-access under the creative commons CC-BY-4.0 licence. The authors acknowledge the use of funding from the Fenner School Publication Fund to make this possible.

Douglas-Smith, D., **Iwanaga, T.***, Croke, B.F.W., Jakeman, A.J., 2020. Certain trends in uncertainty and sensitivity analysis: An overview of software tools and techniques. Environmental Modelling & Software 124, 104588. https://doi.org/10.1016/j.envsoft.2019.104588

Software:

Iwanaga, T.*, Douglas-Smith, D., 2019. Wosis: beta-release (v0.1.3). https://github.com/ConnectedSystems/Wosis. Zenodo. doi: 10.5281/zenodo.3406947

Data and Analysis:

Douglas-Smith, D., **Iwanaga, T.**, 2019. UASA Trends. https://github.com/frog7/uasa-trends. doi: 10.5281/zenodo.3406946

Environmental Modelling and Software 124 (2020) 104588

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: http://www.elsevier.com/locate/envsoft

Certain trends in uncertainty and sensitivity analysis: An overview of software tools and techniques

Dominique Douglas-Smith [a], Takuya Iwanaga [a,*], Barry F.W. Croke [a,b], Anthony J. Jakeman [a]

[a] The Integrated Catchment Assessment and Management Centre, Fenner School of Environment and Society, The Australian National University, Canberra, Australia
[b] Mathematical Sciences Institute, The Australian National University, Canberra, Australia

ARTICLE INFO

Keywords:
Uncertainty analysis
Sensitivity analysis
Analysis software packages
Bibliometric trends
Environmental modeling

ABSTRACT

Uncertainty and sensitivity analysis (UA/SA) aid in assessing whether model complexity is warranted and under what conditions. To support these analyses a variety of software tools have been developed to provide UA/SA methods and approaches in a more accessible manner. This paper applies a hybrid bibliometric approach using 11 625 publications sourced from the Web of Science database to identify software packages for UA/SA used within the environmental sciences and to synthesize evidence of general research trends and directions. Use of local sensitivity approaches was determined to be prevalent, although adoption of global sensitivity analysis approaches is increasing. We find that interest in uncertainty management is also increasing, particularly in improving the reliability and effectiveness of UA/SA. Although available software is typically open-source and freely available, uptake of software tools is apparently slow or their use is otherwise under-reported. Longevity is also an issue, with many of the identified software appearing to be unmaintained. Improving the general usability and accessibility of UA/SA tools may help to increase software longevity and the awareness and adoption of purpose-appropriate methods. Usability should be improved so as to lower the "cost of adoption" of incorporating the software in the modelling workflow. An overview of available software is provided to aid modelers in choosing an appropriate software tool for their purposes. Code and representative data used for this analysis can be found at https://github.com/frog7/uasa-trends (10.5281/zenodo.3406946).

1. Introduction

Computational modeling has become a key activity in many areas of research. In the environmental sciences the amount of available computational power and speed has led to the development of environmental models with ever-increasing level of detail and complexity. In this context complexity is reflected by the number of parameters a model incorporates as inputs. These parameters may also be referred to as ‘parameter factors’, ‘factors’ or simply ‘inputs’ in the literature (Norton, 2015). Increasing the number of parameters allows for a more detailed representation of the investigated system while also increasing computational cost and model complexity at an exponential rate. Increased detail (and thus complexity) may reduce the identifiability of parameters - the ability to apportion model results to specific parameter values – but is not always justified or necessary with respect to the aims of the modeling exercise.

Increased complexity has led modelers to better appreciate the issue of model identifiability (Guillaume et al., 2019) and to recognize the importance of understanding the contribution of model inputs with respect to model performance and purpose. Uncertainty and Sensitivity Analysis (UA/SA) refer to the methods and approaches used to help researchers better understand the relative importance of each parameter factor within a given problem context. Put simply, “[S]ensitivity analysis assesses how variations in input parameters, model parameters or boundary conditions affect the model output” (Bennett et al., 2013). With these approaches, it is possible to better understand how *sensitive* model results are to parameter factors and how *uncertain* the model results are (Saltelli et al., 2019; Saltelli and Annoni, 2010). Individual parameter factors may influence one or more outputs and could (conditionally) affect the importance of other factors; referred to as parameter interaction. The practice of analyzing uncertainty and sensitivity is now considered standard modeling practice. The interested reader is directed to (Bennett et al., 2013; Norton, 2015; Pianosi et al., 2016; Razavi and Gupta, 2015) for introductory overviews and further information.

Understanding the relative ‘sensitivity’ of parameters can aid in the

* Corresponding author.
E-mail address: iwanaga.takuya@anu.edu.au (T. Iwanaga).

https://doi.org/10.1016/j.envsoft.2019.104588
Received 13 June 2019; Received in revised form 13 September 2019; Accepted 13 November 2019
Available online 18 November 2019

Table 1
Descriptions of common UA/SA sampling and analysis techniques and key texts. Qualitative assessment and indications of sampling requirements are provided. Where indicated, p refers to the number of parameters and N the number of parameter sets.

Name	Abbreviation	Description	More information
One-at-a-time SA	OAT	Each parameter is perturbed from its baseline point. May also be known as variations of the name such as one-factor-at-a-time, one-variable-at-a-time, etc. Required number of model evaluations range from $p+1$ to $(N \times p)+1$ where p is the number of parameters and N is the number of desired perturbations.	Czitrom (1999)
Derivative-based SA	–	Family of methods that take partial derivatives of each input parameter with respect to the output.	(Helton, 1993; Norton, 2015)
Variance-based SA	–	Family of methods that attempt to map statistical properties of the output distribution to the inputs used – how variance in the inputs explains variance in the outputs. Variance-based approaches may require an exponentially increasing number of N samples with increasing p to obtain reliable results (e.g. Razavi and Gupta, 2016).	(Norton, 2015; Pianosi et al., 2015)
Monte Carlo sampling	MC	Random sampling: statistically independent samples of the parameter space. Used for variance-based SA.	(Fedra, 1983; Metropolis and Ulam, 1949)
Latin Hypercube sampling	LHS	The range of each parameter in the parameter space is partitioned into N equal-probability divisions. One sample is taken from each of the N partitions, generating N samples per parameter, and a sample from each set of N samples is chosen for each parameter. The process is repeated to obtain the desired number of samples (Norton, 2015).	McKay et al. (1979)
Importance/stratified sampling	–	Estimate the probability density of the parameters, usually uniform or Gaussian, and determine the importance of resulting outcomes in order to define regions in the parameter space. Each region is given an equal quota of randomly distributed samples. Used for variance-based GSA.	Castaings et al. (2012)
Morris method	Morris	An elementary effects method, derivative-based GSA. Ranks parameters by influence on output and non-linearity. Each parameter is stepped along trajectories. The starting points are random and uniformly distributed. The parameters are perturbed once in succession along the trajectory, in random order. The resulting sample consists of the changes in model output caused by each parameter's perturbation (Norton, 2015). Several variations have been proposed (Pianosi et al., 2016) and convergence of sensitivity indices are said to occur with a relatively small number of model evaluations and so is commonly used for factor screening (Gan et al., 2014; Sun et al., 2012). It requires $N(p+1)$ model evaluations, where typically $N \approx p$ or less (Norton, 2015)	(Campolongo et al., 2011; Morris, 1991)
Derivative-based Global Sensitivity Measure	DGSM	Derivative-based GSA. Sensitivity indices are computed by taking the integral over the function domain of the square of the partial derivatives of each factor with respect to the function (Sobol and Kucherenko, 2009). The method is described as being effective at screening parameters for high-dimensional models with low sample sizes (Becker et al., 2018).	Sobol and Kucherenko (2009)
Sobol' method	Sobol'	Variance-based GSA. An MC-based method: analyzes how the variability of a parameter or combination of parameters influences the variability of the output.	Sobol (1993)
Fourier Amplitude Sensitivity Test	FAST	Variance-based GSA. By estimating the output by a sum of sinusoids, each parameter becomes a function of a chosen variable ranging from $-\pi$ to π. Rather than computing the variance and mean of the output as multiple integrals, these now become a single integral with respect to the chosen variable. Variations include eFAST and Random Balance Design (RBD).	FAST: (Cukier et al., 1973) eFAST: (Saltelli et al., 1999; Wang et al., 2013) RBD: (Tarantola et al., 2006)
Distributed Evaluation of Local Sensitivity	DELSA	Derivative- and variance-based GSA. An elementary effects method. DELSA is a multiple starts perturbation method in which squared finite differences are the metric of sensitivity. DELSA is said to be able to obtain the full distribution of sensitivities at a lower cost compared to the Sobol' method (i.e. a comparatively lower N).	Rakovec et al. (2014)
Regression- and correlation-based SA	–	Statistical-based GSA. The sensitivity metric is the regression/correlation coefficient between the input parameters and output after Monte Carlo sampling.	(Iman and Helton, 1988), (Saltelli and Marivoet, 1990)
Regional SA	RSA	Statistical-based GSA. A binary split of the input parameters from a Monte Carlo sample is determined by whether the resulting output respective to an input sample exhibit required behaviors. A cumulative distribution function is applied to the non-behavioral input samples as a metric of sensitivity. Said to have low computational requirements (Sun et al., 2012).	(Spear and Hornberger, 1980; Young et al., 1978)
Generalized Likelihood Uncertainty Estimate	GLUE	Model results are given as probability distributions of possible outcomes. Assesses how accurate these model results are as a representation of uncertainty.	Beven and Binley (1992)
Emulators	–	A simplified model is fit to a sample in order to give a general indication of parameter sensitivity.	(Crestaux et al., 2009; Oakley and O'Hagan, 2004; Oladyshkin and Nowak, 2012; Ratto and Pagano, 2010; Storlie and Helton, 2008)

development of better monitoring strategies and experiment design, for example indicating the priority and amount of data to be collected (Saltelli and Tarantola, 2002). The practice of SA can also help to constrain the parameter space by identifying parameters that may be 'insensitive' or 'inactive', having little to no effect on model results, at least for the purpose of the modeling. Identifying such parameters can help constrain model complexity which in turn eases the computational cost of model evaluations, for example to facilitate uncertainty analysis and the development of surrogate models.

In recent years a wide variety of software tools to support UA/SA processes have become available that make such analyses more accessible to modelers. To gain an overview of the available methods and tools, we applied a hybrid bibliometric approach using publications from the Web of Science database. While reviews of sensitivity analysis practice have been published (see for example Ferretti et al., 2016; Saltelli et al., 2019) and comparisons between UA/SA methods conducted (for example Gan et al., 2014; Sun et al., 2012), to our knowledge there does not appear to be an overview of the available UA/SA tools currently in use across different platforms and programming languages. This paper follows on from and is distinguished from existing reviews (such as Matott et al., 2009; Refsgaard et al., 2007) as it surveys UA/SA in environmental modeling, with a specific focus on SA. We then provide information on the available tools, as revealed through the bibliometric analysis and expert knowledge, including implemented UA/SA methods, programming language, and software features. The aim here is then to provide 1) a brief introduction to the field of UA/SA and its relevance to environmental modeling for those new to the field, 2) an overview of UA/SA research trends, and 3) a guide to the development trends of UA/SA tools, their availability, and relevance.

2. Key UA/SA terminologies and methods

Often the first hurdle for those new to a research area is to grasp the multitude of acronyms and terms used. In this section we briefly outline some common terminology, UA/SA methods, and relevant publications for further reference. These are provided here to contextualize the analysis and discussion later in this paper. The information provided in this section is not exhaustive. Interested readers are directed to Norton (2015) for a more thorough introduction to UA/SA, the descriptions of sensitivity analysis methods in Pianosi et al. (2016), the citations in Table 1, and the citations in (Bennett et al., 2013, p. 3).

Pianosi et al. (2015) identify three stages in a sensitivity analysis: selecting a sample of input values from the variability space, running a model evaluation against these input values, and applying a sensitivity analysis method to the input/output samples to compute *sensitivity indices*, i.e. values which indicate each parameter's sensitivity. For more information about the calculations for various sensitivity indices, see Norton (2015). Here, the *variability space* refers to all possible combinations of values that can be assigned to a model's input parameter set. By running the model with the values sampled from the variability space and taking note of the resultant outputs, analyses can be conducted to calculate the influence that a specific input, or set of inputs, may have, i.e. their sensitivities. The focus of this paper is on providing an overview of tools that aid in conducting these analyses.

Methods to select the sample of input values are often characterized as being either 'local' or 'global'. Global methods (GSA) consider all dimensions of a model "in one grand exercise" (Leamer, 1985), achieved by varying all parameter values at the same time. GSA methods are themselves commonly categorized as being statistical, derivative, or variance based. Statistical methods use statistical analysis of the parameter space as a measure of sensitivity (Pianosi et al., 2016). Derivative-based methods provide indices which characterize the distributional properties of partial derivatives (Razavi et al., 2019). Variance-based approaches determine how different factors contribute to model variance by analyzing and decomposing the variance in model outputs (Razavi et al., 2019). For brevity, a full exploration of these methods is not provided here, but a brief overview, with references to relevant papers, is given in Table 1.

The strength of GSA methods is that they provide a more robust depiction of model uncertainty by comprehensively accounting for parameter interactions (Saltelli and Annoni, 2010). Such approaches assume a random distribution of output values in the parameter space and that such a distribution is plausible. GSA methods can also be computationally expensive to perform as the parameter space being explored can be very large. Sampling methods ('schemes') are used to aid in limiting the number of model runs involved whilst adequately representing the parameter space. The computational cost of applying GSA methods may explain, at least in part, why their use is relatively uncommon compared to their local counterparts.

Local SA methods (LSA) are anchored around a particular point in the parameter space with analysis involving comparisons against a known 'baseline' output (Razavi and Gupta, 2015). The simplest, most naïve, and most common, method of SA is one-at-a-time (OAT). As the name suggests, this approach involves changing the value of a single parameter factor at a time (referred to as 'perturbing') whilst keeping all other parameters constant at their nominal values. This approach could be described as taking samples along a single dimension with the changes to the output then attributed to the factor that was modified. There are different approaches to how much the parameter value is perturbed but often a proportional increment is used – e.g. increase or decrease a parameter by 10% of the nominal value up to and including a given bound (Razavi and Gupta, 2015).

Other LSA methods examine the partial derivatives of output with respect to each input parameter. These are computed at one point in the sample space to determine sensitivity indices. The simplicity of the procedure is advantageous, as well as being computationally inexpensive for first order derivatives as they often do not require a formal sampling approach. Monte Carlo (MC) – a simple random sampling – is commonly used, although it offers a limited representation of the total parameter space (Gan et al., 2014). The downside is that LSA only provides a robust indication of sensitivity for linear or additive models (Saltelli and Annoni, 2010): they do not account for parameter interactions and become computationally expensive when higher order and non-linear effects are considered. To resolve this issue several other sampling approaches have been developed and applied. Given each method and approach have their pros and cons, multiple methods could be applied to obtain complementary results à la ensemble analysis (Sagi and Rokach, 2018) and should be considered where appropriate (Sun et al., 2012). Brief descriptions of commonly employed methods are given in Table 1. Methods are taken to be "common" where they are indicated to be so in recent review papers (specifically, Gan et al., 2014; Pianosi et al., 2016), the references found within these, and those found within the identified corpora (detailed in the next section).

3. Method: The hybrid bibliometric approach

To conduct this bibliometric review a collection of publications (the 'corpora') was gathered from Clarivate Analytics' Web of Science (WoS) database using the available web-based Application Programming Interface (API). Use of the API enabled programmatic access to the publication data and metadata including titles, abstract text, author-supplied keywords, and DOIs. Data was retrieved with the use of Wosis (Web of Science Analysis), a Python package developed to simplify the process of querying the WoS database and aid in data analysis and visualization (Iwanaga and Douglas-Smith, 2019). Publications in the resulting corpora were taken to represent the field of uncertainty and sensitivity analysis in the overarching field of environmental modeling.

To ensure as much transparency as possible, much of the data collection and subsequent analysis was conducted programmatically in the Python programming language. The complete dataset cannot be made available as it is subject to Clarivate Analytics' license terms. Representative datasets are provided instead, along with the code developed for the analysis; these can be viewed as a collection of Jupyter Notebooks and associated files at https://github.com/frog7/uasa-trends (Douglas-Smith and Iwanaga, 2019). Names of specific notebooks will be referred to throughout this text where further detail can be found.

The corpora was iteratively and incrementally refined through a semi-autonomous process of topic identification, keyword search, and subsequent manual analysis of the publications with the aid of key phrase extraction. Topic modeling (briefly described in Section 3.2) was used to aid in identifying a collection of papers relevant to uncertainty and sensitivity analysis and their overarching focus, be it an application of or guiding frameworks for UA/SA. The publication and citation trends within these topic areas were then analyzed. Additional topic modeling, complemented by a keyword search process, was used to identify papers related to the use of UA/SA software. These were manually combed through with the aid of an automated key phrase identifier that helped to reduce the amount of text to be examined. A subset of these papers were investigated for mention of software tools and packages. The general search and analysis approach is depicted in Fig. 1, with further detail on topic modeling and key phrase identification provided within this section.

3.1. Initial search

The initial corpora for the analysis was identified by specifying the search phrase (with search fields bolded): **TS**=("sensitivity analysis" OR "uncertainty analysis" OR "uncertainty quantification" OR "uncertainty propagation" OR "local sensitivity analysis" OR "LSA" OR "one-at-a-time" OR "exploratory modeling" OR "OAT" OR "global sensitivity analysis" OR "GSA" OR "all-at-a-time" OR "AAT") AND **WC**=("ENVIRONMENTAL SCIENCES" OR "WATER RESOURCES" OR "ENGINEERING ENVIRONMENTAL" OR "INTERDISCIPLINARY APPLICATIONS"). This returns publications that use at least one of the specified terms (those listed for the "**TS**" field) within the title, abstract, or author supplied keywords for publications in the the WoS defined subject areas; specified for the "**WC**" field. The raw search string is supplied for transparency and can be used to obtain the corpora from WoS.

Only English language publications between 2000 and 2017 were considered for this study, with the ending year selected as the data request occurred in December of 2018. The approach taken at the time was to include full year datasets only. The final search phrase applied with the specified time frame reduced the number of matches from over 500 000 to 11 718 publications. The number of results obtained through the unrestricted search were far too many to comprehensively review, at least in a timely manner. The initial corpora for this study (of 11 718 publications) were then further constrained through the process depicted in Fig. 1 and is described in more detail below.

3.2. Topic identification

A key focus in this study is the software tools and packages available to support UA/SA processes, the methods they implement, and the trends of these. To this end, topic modeling was applied to constrain the corpora to relevant publications for further consideration. Topic models attempt to cluster texts into similar or related topics based on commonly occurring words and can aid in identifying new and emerging fields whilst also reducing the likelihood of bias and the required hours for a systematic review (Achakulvisut et al., 2016; Westgate et al., 2018). Topic modeling has been applied before to reduce the time and difficulties encountered when conducting systematic reviews (Westgate and Lindenmayer, 2017), however their use is still relatively limited and perhaps underutilized. Although software is available to aid in these bibliometric approaches, currently no single software package provides all necessary functionality to conduct end-to-end systematic mapping – the classification of articles based on their contents – of research literature from data collection through to summarization and visualization. Arguably the conjunctive application of systematic mapping and bibliometric analysis is still in its infancy (as evidenced by Nakagawa et al., 2018).

Topics are identified by the common co-occurrence of semantics within a discipline. For example, "sensitivity" in the context of SA would conceptually be expected to appear in texts containing words such as "analysis", "uncertainty", and "modeling". The term "sensitivity" may also appear in relation to physical/psychological response to stimuli, in which case the term will appear alongside terms associated with the medical and therapy fields. Topics can be identified and represented through their common semantics. The topic modeling approach provided within Wosis – Non-negative Matrix Factorization (NMF) – is implemented with the scikit-learn Python package (Pedregosa et al., 2011). The approach allows publications to be assigned to one or more topics (Arora et al., 2012) and has been shown to be appropriate for collections of short texts (Chen et al., 2019). This process was complemented with a traditional keyword search to help identify publications related to specific subjects.

Tokens, meaning specific words or terms, for topic modeling consisted of the text found in the document titles, abstracts, and keywords. The top 1000 tokens found within the corpora based on Term Frequency-Inverse Document Frequency (TF-IDF) rankings were selected for topic modeling. TF-IDF is a common ranking method used in text mining (Beel et al., 2016). A high TF-IDF score indicates that the word token has a high frequency within specific document(s), but a low number of occurrences within the entire corpora. Weighting the score in such a manner has the effect of filtering out commonly used tokens which may not have high semantic importance.

3.3. Key phrase identification

Once a topic area is identified, the resulting sub-corpora can be further constrained through automated key phrase identification. The approach summarizes text, aiding reviewers to identify irrelevant publications by reducing the amount of text for manual review. The implemented approach attempts to identify these phrases of interest by scoring sentences based on their similarity with other sentences throughout the abstract text.

To elaborate, each sentence (s_i) is compared with other sentences in the abstract (s_y), which are initially filtered based on the presence of a root token which is taken to be the token that appears in the middle of s_i. This root token selection approach is used in Rabby et al. (2018) for its simplicity and computational efficiency. The similarity between s_i and s_y is then scored based on the ratio of the intersection of the two sentences. Sentences with three or less tokens (i.e. words, numbers, or other counted by splitting the text on individual spaces) are ignored.

The approach assumes that important features of the publication, such as its key findings, will be repeated throughout the considered fields (title, abstract, and author-supplied keywords). These may, for example, be introduced or alluded to, framed and the implications discussed. The implemented approach is therefore dependent on the abstract length, with longer texts preferred. Poor performance can be expected for very short abstracts (e.g. 3 sentences or less) and these were ignored for the purpose of this study. Comparisons with an established key phrase identification approach, RAKE: Rapid Automatic Keyword Extraction (Rose et al., 2010), implemented through the rake-nltk Python package – indicate that the above approach produces, subjectively, key phrases that were more useful for the purpose of this study (see Table 2).

3.4. Citation and trend analysis

Citation analysis indicates the papers being referred to by other papers within the corpora as well as the overall number of citations the given publication has received, the assumption here being that impactful papers are more likely to be cited. The number of citations is then used to indicate papers that are of high importance to the subject at hand. Both the total number of citations and the average citations since publication were used in the analysis. Publication trends within topic areas aided in identifying the general focus and direction taken by the research community. Plotted publication trends were used for this

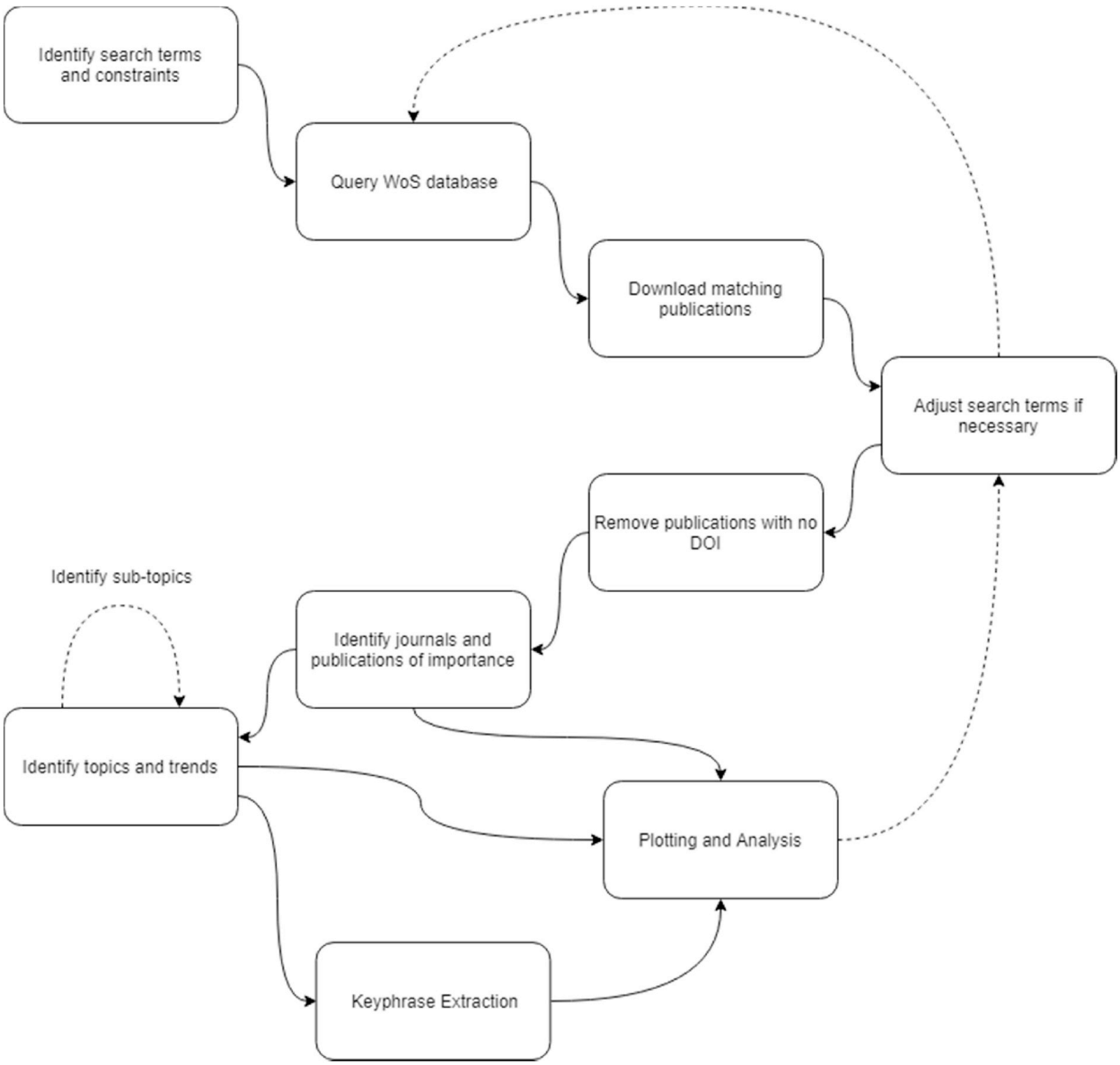


Fig. 1. The hybrid bibliometric analysis process. Identification and subsequent analysis of the final corpora followed an iterative process through which publications were progressively filtered to arrive at a relevant subset.

purpose.

4. Results: UA/SA packages

Of particular interest to this paper were the trends of software packages implementing UA/SA methods and these are discussed here. The final corpora was broadly categorized into two topics – "Applications" and "Frameworks" using the topic model described in the "Method" section. Publications focused on UA/SA frameworks and guidelines were placed into the "Frameworks" sub-corpora, while "Applications" included those taken to be focused on the application of UA/SA methods. From each of these a keyword search was applied to identify publications related to model sensitivity, optimization, uncertainty quantification, or toolboxes, in order to build a sub-corpora related to the software.

Manually sorting the identified publications with the aid of the automated key phrase extraction tool reduced the corpora to 193 papers (referred to as the "Software corpora", see Notebook 5c "Software packages analysis"). Papers were regarded as relevant if they: included direct reference to UA/SA or optimization software packages; were theory, review, or framework papers that recommended software implementation to a given field; or referred to other methods and packages of interest to expert opinion. Further detail and a general bibliometric overview are provided in a later subsection.

There does not appear to be a strong correlation between the Applications and Software corpora (Fig. 2). The Software corpora has a stable publication trend relative to those focusing on applications over the surveyed timeframe. A spike in publications in 2007 proportional to the full final corpora can be seen (Fig. 3). While publications on the software for UA/SA have been increasing (see Fig. 4) the trend relative to the Applications corpora and the full corpora could be indicative of 1) a general ambivalence towards reporting use, or development of, general UA/SA software, 2) a common set of UA/SA software, 3) a reliance on self-coded analysis software, or 4) increased tendency to release software in a directly citable manner, e.g. with an attached DOI which the WoS database does not include but this is considered unlikely however in the authors' opinion.

The slow uptake of software packages, relative to the Applications corpora, could also be due to 1) a lack of documentation for beginner users and 2) a lack of awareness of available software packages. In the first case, beginner users may not use software that requires significant learning time for effective use, especially when no clear user guide, examples to draw from, or community to engage with exists. In the latter case, modelers should be made aware of the available software that can reduce the time required to conduct UA/SA and promote better practices in UA/SA.

The software evident in the literature range from those specific to a field, general-purpose packages, to custom-made code. Fields such as hydrology, climate, chemistry, and more general environmental modeling and engineering used field-specific packages. A complete list of reviewed software publications and their related software packages can be found in Notebook 5a "Finding software packages by keyphrase extraction". The most common analysis method provided by UA/SA software was found to be Sobol' with the R sensitivity package providing the widest mix of methods (see Table 3). Surveyed software tools typically did not provide OAT analysis, perhaps due to its simplicity or a sign of its decline. Publication of software related papers is relatively stable, with a proportional spike in 2007 (Fig. 3).

Software for the development of emulators did not feature heavily within the Software corpora although they are present, the HDMR method being one example (described later on). Software to develop emulators include ChaosPy (Feinberg and Langtangen, 2015), the PRISM Uncertainty Quantification framework (Hunt et al., 2015), GTApprox (Belyaev et al., 2016) and UQ-PyL (Wang et al., 2016). A collection of functions presented as a Matlab toolbox is also introduced in Vu-Bac et al. (2016). All of these with the exception of Vu-Bac et al. (2016) were developed in the Python programming language. Application of Artificial Neural Networks and similar approaches did appear in the corpora but is not a topic of focus here.

As aforementioned, current trends have shown an increased interest in best practices. Three SA packages, released within the past five years, reflect these changing attitudes: PSUADE (Gan et al., 2014), SAFE (Pianosi et al., 2015), and VARS-TOOL (Razavi et al., 2019). PSUADE (a Problem Solving Environment for Uncertainty Analysis and Design Exploration) provides users with implementations of UQ methods, including sampling techniques and SA methods (both local and global). The package has had general application to various modeling scenarios. SAFE (Sensitivity Analysis For Everybody) provides users with implementations of global SA methods, with the ability to perform multiple SAs, robustness assessment, and convergence analysis without further model runs. As reflected in its name, this package was designed to allow global SA to be accessible to a more general audience.

The most recently released package in the survey, VARS-TOOL, provides implementations of sampling techniques and global SA methods, including derivative-, variance-, and variogram-based, which can all be performed from a single sample. The variogram approach to SA reportedly links both local and global approaches.

4.1. Survey of packages in common programming languages

Brief descriptions of software found in the corpora are provided here, categorized by their implementation language. Some packages may be listed more than once as various implementations may exist, or interoperability between languages is supported. We decided to categorize the packages based on the implementation languages as most packages are not standalone tools with user interfaces ready to be used and are often provided as a library to be incorporated programmatically. Indicating the implementation language also allows readers to identify packages in a familiar language for potential adoption. Very few packages were found to provide a Graphical User Interface (GUI) so some amount of programming ability and experience is the baseline expectation. In the vast majority of cases, users are expected to have a passing familiarity with the UA/SA methods being applied as very little protection against improper use is provided (a further brief discussion is in the Recent developments section). Table 3 and 4 provide summary overviews of the software and packages.

4.1.1. Fortran

Fortran was one of the earliest programming languages available and arguably still dominates the scientific programming landscape. Fortran modules from the surveyed literature are JUPITER API and UCODE. There is also a Fortran repository of UA/SA functions supported by the Joint Research Centre (Pianosi et al., 2015). The JUPITER API (Joint Universal Parameter IdenTification and Evaluation of Reliability Application Programming Interface) attempts to provide a standard set of programmatic functions for developing UA/SA software and serves as the underlying "engine" for other UA/SA packages (UCODE 2005/2014 were developed on top of this API). The provided modules are developed in Fortran-90 and support parallelization and local (derivative) sensitivity analysis. JUPITER API was first released in 2006 and its latest release was 2013. Its affiliated webpage was last updated in 2016, suggesting an active community. It is provided freely and under an open-source license with a user manual and examples of applications.

First released in 1998, UCODE (Universal inverse CODE) was developed in Fortran90, Fortran95, and Perl. It originally implemented inverse modeling methods, and by its first revision (2005) consisted of post-processing modules for (and not limited to) SA, calibration, and UA. The second revision (2014) included MCMC in the UA module and made the platform more compatible with models developed in Matlab or using a GUI. This can be viewed as a response to changing trends in model development, particularly the proliferation of Matlab-based models. User documentation is available for download. Although the software is still available for download, its development has ceased.

Table 2

Example of key phrases identified and extracted by Wosis compared with the RAKE method provided in the 'rake-nltk' package. The original abstract was taken from Roos et al. (2015). Both RAKE and Wosis approaches were limited to a minimum of 3 words per phrase. Identified phrases are ordered by score and do not follow the original paragraph structure.

Wosis	RAKE
We propose a novel formal approach to prior sensitivity analysis, which is fast and accurate.	hoc modified base prior parameter values
	parameters within deeper layers
Other formal approaches to prior sensitivity analysis suffer from a lack of popularity in practice, mainly due to their high computational cost and absence of software implementation.	identifiability issues may imply
	prior sensitivity examination plays
	detect high prior sensitivities
Despite its importance, informal approaches to prior sensitivity analysis are currently used.	prior sensitivity analysis suffer
	parametrized Bayesian hierarchical models
This is especially true for Bayesian hierarchical models, where interpretability of the parameters within deeper layers in the hierarchy becomes challenging.	prior sensitivity analysis
	bayesian hierarchical models
	quantifies sensitivity without high computational cost
They require repetitive re-fits of the model with ad-hoc modified base prior parameter values.	applied bayesian analyses
	novel formal approach
	hierarchy becomes challenging

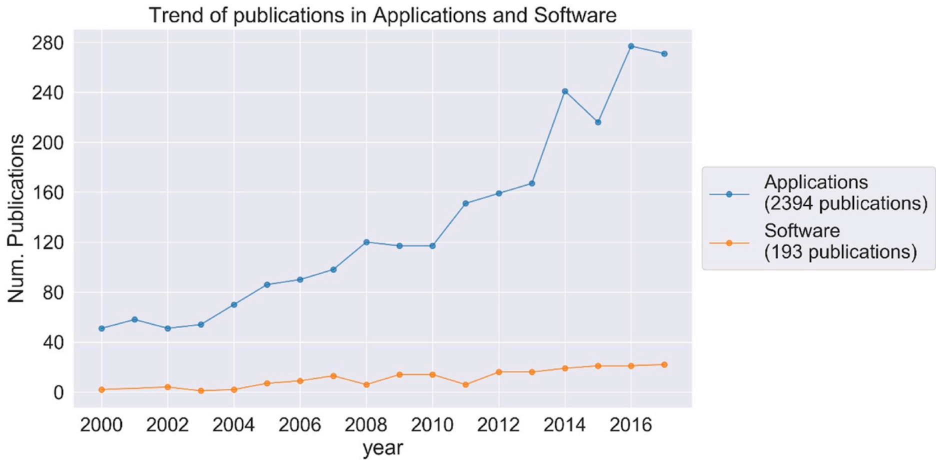


Fig. 2. A comparison of the publication trend of Applications and Software. Both increase during the timeframe, although it is unclear whether the trends are related.

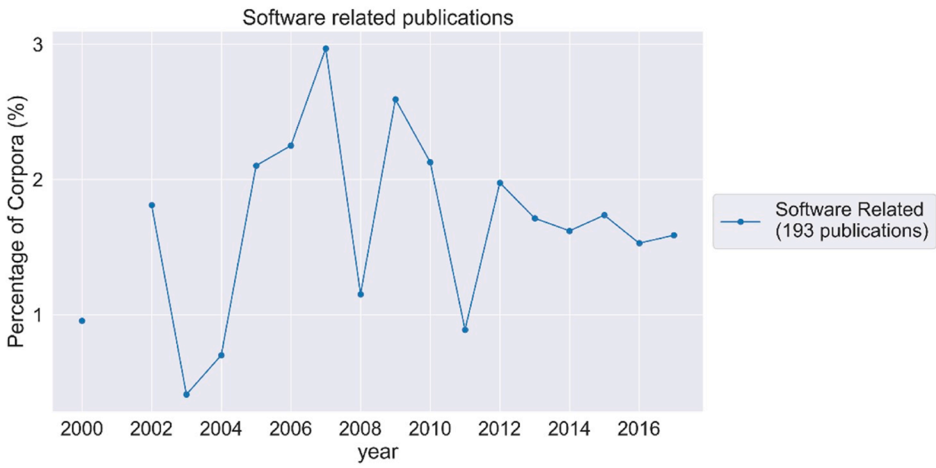


Fig. 3. Relative trend of software publications. The trend is relatively flat but note the initial spike in publications in 2007.

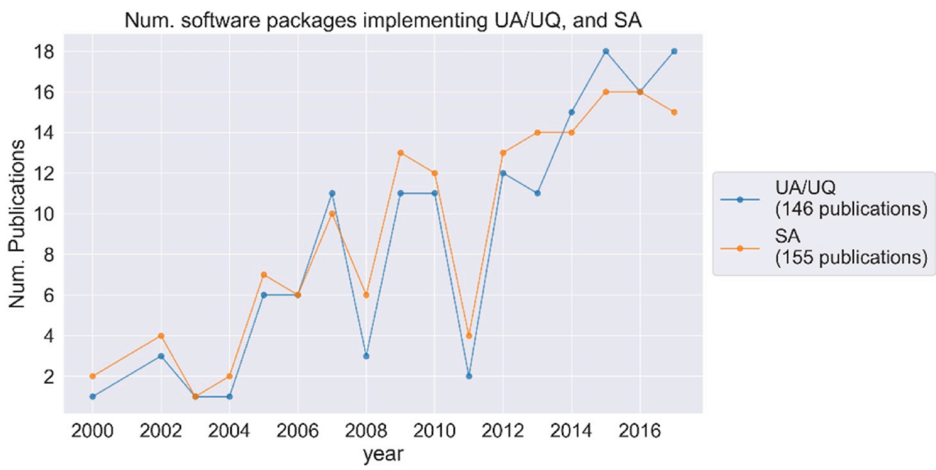


Fig. 4. Absolute publication trends for software packages implementing UA/UQ, and SA methods.

4.1.2. C/C++

Surveyed software available in C/C++ include Dakota, PSUADE, PEST, and VARS-TOOL.

The Dakota toolkit had its initial release in 1994 to provide optimization tools for engineers. With further development, it now includes sampling methods, global SA methods, parameter estimation, and UQ. The software can be tightly-, semi-, or loosely-coupled to the target model, requiring the user in the first two cases to modify their code or use a direct interface. The package is presented as being accessible to beginners and involves advanced features for more competent users. It operates on Linux, Windows, and Unix. Parallelization is possible and there is a GUI option. It is freely available for academic use and open

Table 3
Comparison table of available UA/SA methods in the surveyed packages.

Name[a] (Language)	MC	LHC	Morris	DGSM	Sobol'	FAST	RSA	Regression/ Correlation SA	Other
SimLab (Matlab)		✓	✓		✓	✓		✓	
MCAT (Matlab)					✓			✓	GLUE and many others (UA, parameter estimation)
GUI-HDMR (Matlab)					✓			✓	HDMR Emulation
UQ Lab (Matlab with R plugin)			✓		✓			✓	Bayesian Inversion, Kriging, Support Vector Machines, and more
SAFE (Matlab/R)			✓		✓	✓	✓		Dynamic Identifiability Analysis, PAWN
VARS-TOOL (Matlab, C++, Python, and built into OSTRICH)		✓	✓		✓		✓		STAR-VARS, Generalized Global Sensitivity Matrix
Dakota (C/C++, Fortran77/Fortran90)	✓	✓	✓		✓				Supports emulation and many other UA/SA methods
PSUADE (C++)	✓	✓	✓			✓			Fractional Factorial, Central Composite, Probabilistic methods, and others
PEST/PEST++ (C/C++, Fortran)	✓		✓		✓				
R Sensitivity (R)	✓	✓	✓	✓	✓	✓			DELSA, Kriging, many other variations available
SALib (Python)		✓	✓	✓	✓	✓			Delta Moment Independent Measure, Fractional Factorial, Finite Difference
MADS (Julia, C/C++)	✓	✓			✓	✓			Kriging, Bayesian Information Gap Decision Theory, Support Vector Regression
GANetXL (Excel)									Single- and multi-objective genetic algorithm
UCODE (Fortran90/95, Perl)								✓	
MOUSE (Java)			✓			✓	✓		GLUE
GLUE (R, Matlab)									GLUE
OSTRICH (standalone)								✓	GLUE, user defined evaluations also possible

[a] As documentation can lag behind releases, software may include implementations of methods not listed in the table (at time of writing).

source. A user community exists, including mailing lists and interaction with developers. Documentation includes user manuals, examples, and release notes. Dakota is well maintained, its most recent release and webpage update being in 2018. It is an example of software that has kept up to date with the latest trends in UA/SA and software implementation.

PSUADE (Problem Solving environment for Uncertainty Analysis and Design Exploration) can link to simulation code in any language. It provides users with 14 sampling methods and 12 SA methods, both local and global SA. It was developed for large complex systems models and has been applied to various fields. The software has a free public license and is open source. A collaborative user community exists. The software and documentation (a user manual) are available for web download. The package is well maintained, with its latest release and update in 2018.

PEST (Parameter EStimation Toolkit) is designed primarily for model calibration. Originally released in 2003, and with its most recent release in 2019 it has remained up to date with the latest research in environmental modeling. The current package provides parameter estimation and uncertainty analysis, including Monte Carlo analysis, and has parallelization capabilities. The software is designed for complex environmental models, and other models. Models written in C, C++, Fortran, and Python have interoperable interfaces available. It is free, although the license does not appear to be specified, and well-documented for ease of use. Developer-user interaction is encouraged, and training courses are offered.

4.1.3. Matlab

Identified packages of interest written in Matlab are Simlab, MCAT, GUI-HDMR, UQLab, SAFE, and VARS-TOOL.

SimLab is a package for Monte Carlo-based SA, written in Matlab and supplied by the Joint Research Centre. Initially released in 1985, its latest release was 2008 and its associated webpage was last updated in 2016. It provides Monte Carlo and other random sampling methods, test functions for educational purposes, and GSA (correlation-, regression-, and variance-based). The SA follows a loosely-coupled approach requiring only the model output to be fed in. It is freely available for academic use and open source. No user community appears to exist. The documentation consists of a reference manual and the software is available for web download.

MCAT (Monte Carlo Analysis Toolbox) implements Monte Carlo SA. Its first release was 2001 and a companion paper, highlighting the importance of best practices in SA, was released in 2007. However, no further research appears to have been conducted since this time and links to software download provided in the companion paper have expired. This package is of interest as an example of software tooling designed to promote modeling SA best practices. The software provides implementations of UA/SA methods, including regional SA, Monte Carlo analysis, and GLUE. A GUI was developed for it in 2007. The package is free and open source, and documentation includes a manual and examples. No user community appears to exist, however, there is an unofficial GitHub page (see Table 4).

GUI-HDMR (Graphical User Interface-High Dimensional Model Representation) provides HDMR, a variance-based SA method, which the developers advertise as an alternative to other contemporary SA methods. The user must supply an appropriate sample of the model output (there is a complementary package, RS-HDMR [Random Sampling-HDMR] for this purpose). Users have the choice of using a GUI or a script-based interface. The software is reportedly user-friendly and has been applied to various fields. It is freely available for academic use, but not open source. The software and user documentation are available for web download. Although the related publication is highly cited, this software appears to be abandoned, having its first and last release in 2008. A lack of user community and implementation of a single SA method could be a cause for this.

UQLab (Uncertainty Quantification Laboratory) provides, among

Table 4
Summary details (specifications, usability) of some available software of interest. Dash (–) indicates the information could not be found.

Name and Language	First and Last Release (latest update[a])	License	Community	Docs	Indicated required expertise	Related publication	Link to source/ software	Comments
SimLab (Matlab)	1985, 2008 (2016)	Freely available for academic use, End User free license.	–	Manual and examples	Professional tool for model developers, scientists, and professionals	JRC (2015)	https://ec.europa.eu/jrc/en/samo/simlab	Development and simulation tool for UA/SA No GUI.
MCAT (Matlab)	2001 (2007)	Free and open source	–	Manual and examples		Wagener and Kollat (2007)	Unofficial GitHub page https://github.com/ICHydro/MCAT	Monte Carlo Analysis Toolbox GUI is available
GUI-HDMR (Matlab)	2008	Freely available for academic use, but not open source.	–	Manual		Ziehn and Tomlin (2009)	http://www.gui-hdmr.de/	Software does not appear to be actively developed but still available and in use GUI is available
UQ Lab (Matlab with R plugin)	2014, 2018 (2018)	Free for academic use. Content management system is licensed, scientific modules are open source.	User collaboration encouraged, users can contribute to code with revision by developers	Manuals, examples, release notes.	Beginner to advanced functionality	Marelli and Sudret (2014)	https://www.uqlab.com/download	Tagline states "make uncertainty quantification available for anybody, in the field of applied science and engineering" Plugin for R Sensitivity package available Supports parallelized analysis. Has a GUI
SAFE (Matlab/ R)	2015, 2015 (2018)	Freely available for academic use, open source.	No user community, easily adapted to personal use	Pianosi et al., workflow scripts	Beginner to advanced functionality	Pianosi et al. (2015)	https://www.safetoolbox.info/register-for-download/	Designed for users with limited global SA/ Matlab experience Has various GUIs available
VARS-TOOL (Matlab, C++, Python, and built into OSTRICH)	2016, 2018 (2018)	Free for non-commercial use.	–	Manual	Beginner to advanced functionality	Razavi et al. (2019)	http://vars-tool.com	Supports parallelization
Dakota (C/ C++, Fortran77/ Fortran90)	1994, 2018 (2018)	Freely available for academic use, open source with various levels of user interaction. GNU LGPL from version 5.0.	User mailing list and user-developer interaction.	Manual and examples	For users experienced with UA/SA	Adams et al. (2010)	https://dakota.sandia.gov/content/getting-dakota-source-code	Toolkit for optimization, experimental design, and UA/SA Supports parallelization Has a GUI Linkages with Python, Matlab, and Scilab available
PSUADE (C++)	2013, 2018 (2018)	Free public license, open source, LGPL.	User community	Manual	Said to be beginner friendly	Gan et al. (2014)	https://github.com/LLNL/psuade https://computation.llnl.gov/projects/psuade/software	Supports parallelization
PEST (C/C++, Fortran)	2003, 2019 (2019)	Free	User-developer interaction, training courses	Manual, tutorial		Doherty (2018)	http://www.pesthomepage.org/Downloads.php	Supports parallelization Linkages with Python available
R – Sensitivity package	2006, 2018 (2018)	Free public licence (GPL-2), open source.	Developer community.	Manual	Assumes knowledge of R	Iooss et al. (2018)	https://CRAN.R-project.org/package=sensitivity	
SALib (Python)			User community	Manual, examples,		Herman and Usher (2017)		

(continued on next page)

Table 4 (*continued*)

Name and Language	First and Last Release (latest update[a])	License	Community	Docs	Indicated required expertise	Related publication	Link to source/ software	Comments
	2013, 2018 (2018)	Free public licence (MIT), open source.		release notes	Assumed knowledge of Python		https://github.com/SALib/SALib	Has some visualization methods chiefly for the Morris method
MADS (Julia, C/C++)	2016, 2018 (2018)	Free public licence (GPL), open source.	User community	Manual and examples	Beginner to advanced functionality	Various publications listed under https://mads.lanl.gov/#research	https://github.com/madsjulia/Mads.jl	Supports parallelization
JUPITER API (Fortran90)	2006, 2013 (2016)	Free and open source	User-developer interaction	Manual, examples		http://water.usgs.gov/software/JupiterApi		Application Programming Interface to improve model analysis software development Supports parallelization Used to develop other tools, including UCODE (below)
UCODE (Fortran90/ 95, Perl)	1998, 2015 (2016)	Free and open source	–	Manual		Poeter and Hill (1999)	https://igwmc.mines.edu/ucode/	Software appears to be abandoned but download still available
MOUSE (Java)	2014, 2016	Free and open source	–			Ascough II et al. (2015)		Affiliated webpage unavailable Software not under active development but is being maintained Reportedly has a GUI
GLUE (R, Matlab)	1992, 2013 (2016)	Free for academic use, open source	–	Manual and examples		Beven and Binley (1992)	http://www.uncertain-future.org.uk/?page_id=131	Method implemented in software developed by creators (R) and users (Matlab) with implementations found in other packages
SWAT (Fortran)	2000, 2018 (2019)	Free public licence, open source	User community, user-developer interaction, workshops/ conferences	Manual	May be difficult for beginners	https://swat.tamu.edu/software/	https://swat.tamu.edu/software/plus/	Externally developed tools/interfaces developed to implement e.g. SA, GUI
OSTRICH (standalone)	2017	Free and open source	User community for hydrologists	Manual, examples		Matott (2017)	http://www.eng.buffalo.edu/~lsmatott/Ostrich/OstrichMain.html	Supports parallelization

[a] Latest update refers to last identified date in which documentation or code was released. Documentation refers to user/technical manuals, publications specifically on the software or other.

other tools for UQ, tools for statistical analysis, such as sampling and global SA. Global SA methods are supplied through a linkage with the R Sensitivity Package. Parallelization is supported. The package is user friendly and adaptable to various levels of computational experience. Collaboration amongst users is encouraged and users can contribute to code, with revision by the major developers. It is portable between operating systems and freely available for academic use, however documentation is not freely available. The software is well-maintained, with its latest release and update in 2018.

SAFE (Sensitivity Analysis For Everyone) is compatible with the GNU Octave environment and a version implemented in R exists, making it the most openly accessible of all the surveyed Matlab packages. It runs on any operating system. The toolbox was designed to make global SA accessible to users with limited knowledge of global SA or Matlab, whilst also allowing more advanced users to explore, research, and better understand SA. Users are provided with various sampling methods, local and global SA methods, and a GUI (see Table 3). Although there appears to be no collaborative user community, user-developer interaction is possible via email. The software is freely available for academic use and is open source. Documentation includes the companion paper (Pianosi et al., 2015) and additional information provided in workflow scripts. There have been no recent releases, however the website is maintained (last update 2018).

VARS-TOOL is also available in C++ and OSTRICH (a user-independent interface). It features off-line and on-line mode options for running models in any language or operating system. Numerous sampling and SA methods are supplied, including VARS. It is said to be user-friendly and accessible to various levels. It appears to operate as a command-line interface, without a GUI. Although recently developed, there is no collaborative community. The software is freely available for non-commercial use and is open-source. There are capacities for parallelization and reporting and visualization tools; its documentation consists of a manual.

4.1.4. R statistical language

The main SA package for the R language is the R 'sensitivity' package. Like Python, the R language is widely used in the sciences and so many of the tooling support interoperability with R (see the section on Python below, and Table 4). The R 'sensitivity' package supplies various SA and sampling methods. It offers loose coupling with models implemented in other languages as well as in R. Test cases are supplied for research and comparison purposes. The package requires knowledge of R, which itself is portable between operating systems and freely available. A developer community exists and the available documentation consists of a reference manual. Since its initial release in 2006 more recently developed methods have been implemented and included in its latest release (in 2018).

4.1.5. Python

As with R, users of Python have a large assortment of options generally due to Python being a general-purpose language often used for interoperability across languages (see Table 4). The principal SA package developed in Python appears to be SALib (Sensitivity Analysis Library) which provides global sampling and analysis methods and is distributed under a free public license. Model runs can be invoked directly or separately ("offline"). SALib is most applicable to systems modeling and knowledge of Python is assumed. It is a freely available, open-source package, with a collaborative user community. SALib is well documented and well maintained: documentation includes an installation guide, basic usage guide, a complete module reference, and release notes; its latest release was 2018. SALib supports visualization of Morris results only, although this feature appears to be under-documented. A separate visualization tool is available for analysis of Sobol results called "savvy" (Hough et al., 2016), however this package was not examined in-depth.

4.1.6. Java

There appears to be limited SA packages implemented in Java, at least in the reviewed corpora. A response to this limitation is the MOUSE (Model Optimization, Uncertainty and SEnsitivity Analysis) package. This is an implementation of MCAT and OPTAS model calibration software for modelers using Java. It is indicative of the continued influence of the packages MCAT and OPTAS. Its first release was in 2014 and was last updated in 2016. Although claiming to be free and open-source, we could not find relevant information to access the package.

4.1.7. Julia

MADS (Model Analysis & Decision Support) is an SA package available for the Julia programming language. The analyses it supports can be tightly- or loosely-coupled with an existing model. In the module documentation, extensive information is provided for all functions included in the main module ("Mads.jl"). The documentation details modules and examples and, although extensive, was found not to be user-friendly, with functions and methods often lacking meaningful descriptions. MADS is said to support use in High-Performance Computing (HPC) environments. It is a freely available open-source package with a collaborative user community.

An inherent advantage of MADS is the relative youth of the Julia language, with v1.0 released in 2018. Due to its relative youth, it leverages lessons learnt in older programming languages and was developed with modern computational architecture in mind. This means that concurrent and parallel programs are relatively easy to develop in Julia (Bezanson et al., 2017) and it has had demonstrable success on HPC platforms (see for example Regier et al., 2019). The disadvantage of this youth, however, is that the user community – while growing quickly – is still relatively small compared to that of established languages. As such, the language ecosystem is undergoing continual development and may still be immature.

4.2. Active use and development

To gauge the level of support and active development occurring for each software tool, we attempted to identify websites, evidence of userbases, public code repositories, journal publications which specifically mention the software tool, and other indications of activity. Through this process we found that many of the packages present in the literature are no longer under active development, although the code and software may still be available for use.

A key issue in developing software for UA/SA is longevity. We find that those packages that are currently used and under active development and maintenance have the advantages of being open source, well documented for transparency and ease of use, have an active user-community, and offer implementations of a range of UA/SA methods for general-purpose application as opposed to providing a specific method for a specific model. Packages that have fallen into disuse may still be useful with the caveat that there is no supportive community to rely on (for bug-fixes, troubleshooting, user-support, and so on). Table 3 provides an overview of the available UA/SA methods in the surveyed packages, while details of the software can be found in Table 4.

4.4. Bibliometric overview

The initial corpora from WoS consisted of 11 718 publications from which journals deemed to be unrelated to the topic areas of interest (as specified by the search terms used), journals with less than three identified publications, and those without a valid DOI were removed. The final corpora consisted of 11 625 publications. Knowing that researchers build on prior work and given the exponential growth of published material (Bornmann and Mutz, 2015; Haddaway and Westgate, 2018), we assume in this analysis that the identified corpora is representative of the UA/SA field. Full details of this process can be found in Notebook 2 "Create filtered corpora". The number of publications in the environmental UA/SA field have been increasing at an exponential rate (depicted in Fig. 5) with Journal of Hydrology having the most publications overall and experiencing the largest year-on-year gain within the analyzed time frame (Fig. 6).

To facilitate analysis, the final corpora was broadly categorized into two topic sub-corpora – "Applications" and "Frameworks" – using the topic model. As a reminder, the final corpora represents a collection of UA/SA research. Publications focused on UA/SA frameworks and guidelines were placed into the "Frameworks" sub-corpora, while "Applications" included those taken to be focused on the application of UA/SA methods. The topic model was iteratively applied and key phrases from top-cited papers were qualitatively examined to determine the focus of the publications. The specifics of the undertaken process can be seen in Notebook 4 "UASA topic modeling".

A keyword search was applied within these topic corpora to sort publications further into those relevant to uncertainty quantification (UQ), UA, and SA. The resulting collections contained 1 940, 2 751, and 1 360 publications, respectively. To distinguish between LSA and GSA methods, specific keywords were searched for in the combined corpora, including, for example, "local sensitivity", "OAT", "one-at-a-time" for local methods and "global sensitivity" and "GSA" to indicate global methods. In addition to these, newer SA methods identified through manual inspection of the corpora were also searched for, such as "active subspaces" and "variograms".

4.4.1. Trends and directions

As suggested by the general publication trends (in Fig. 7), all topics (UA, SA, Frameworks, and Applications) saw large increases in the absolute number of publications over the 2000–2017 timeframe. Within the same time period the proportional share of the filtered corpora has declined for SA (by 4.5%), while UA has increased (by 5%), which may indicate a gradual shift towards being more inclusive of uncertainty related matters in analyses as well as a general need for uncertainty guidelines in environmental modeling (see Notebook 4 "UASA topic modelling").

The five most active journals in the Frameworks sub-corpora were Structural and Multidisciplinary Optimization, Journal of Computational Physics, Environmental Modeling & Software, and Journal of Hydrology (see Fig. 8). The 10 most cited papers from across these top five journals came from Environmental Modeling & Software (2), Structural and Multidisciplinary Optimization (3), Journal of Hydrology (2), Journal of Computational Physics (1), and Computer Methods in Applied Mechanics and Engineering (2), and are detailed in Table 5 under Supplementary Material.

The top-cited "framework" related papers from these journals (Table 7) showcase a range of issues but particularly address the lack of uniformity in the UA/SA approaches used in their respective fields. These fields include:

- environmental modeling – evaluating performance (Bennett et al., 2013), improving confidence in model outcomes, and handling uncertainty (Bennett et al., 2013; Kuczera et al., 2006; Refsgaard et al., 2007), UA for hydrological (SWAT) models (Yang et al., 2008),
- optimization – topology optimization (Sigmund and Maute, 2013), Finite Element Methods (Blatman and Sudret, 2011; Moens and Vandepitte, 2005), level set methods for structural topology optimization (van Dijk et al., 2013), high-dimensional computationally expensive black-box problems (Shan and Wang, 2010), and
- scientific computing - handling uncertainty (Roy and Oberkampf, 2011).

Outlines of procedures, guidelines, comparisons of methods, and suggestions for future research resolve the issues raised in these papers. These papers are highly-cited, indicating that they have had an impact on the research community, at least within their respective fields. It should be noted here that existence of highly cited papers itself does not indicate widespread application of suggested good or best practice and should not be taken as evidence. The review conducted by Saltelli et al. (2019) concludes that there is a "worrying lack of standards and good practices", although it is acknowledged that the review focuses on older papers and may not capture recent trends. Certainly awareness appears to have increased, if not adoption of practices.

Similarly, a keyword search for "best practices" identified 132 papers across the surveyed period. By "best practices" we refer to practices in modeling and uncertainty management that promote transparency and reliability of results. The high citation counts of papers relating to frameworks (Table 7) and the growth in best practices publications in absolute terms (Fig. 9) suggest increasing interest in uncertainty management, particularly improving the reliability and effectiveness of UA/SA. Whether the modelers take up the suggestions in these papers is yet to be seen. Modelers can be encouraged to follow guidelines for reliable and effective treatment of UA/SA if the available software implementing UA/SA is designed in accordance with these guidelines (and if modelers

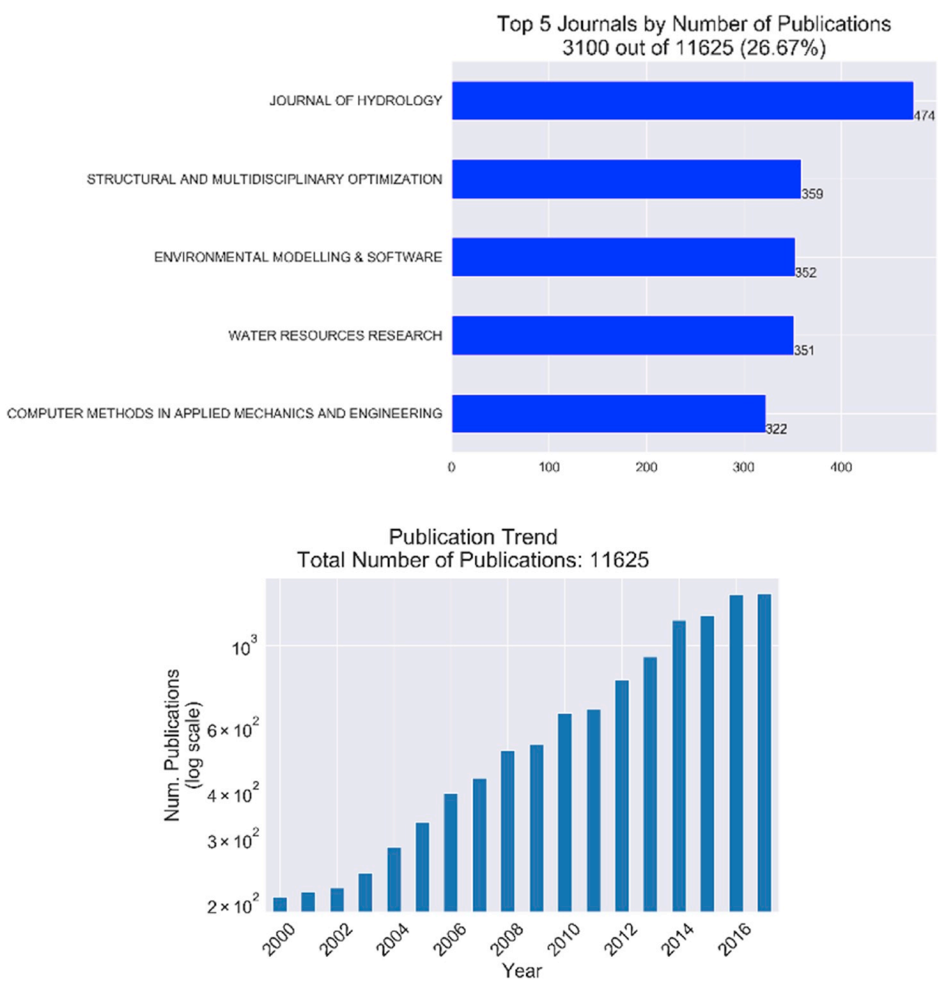


Fig. 5. Publication trends over 2000 to 2017. Journal of Hydrology contributed the most publications in the time frame (474, see top panel). Publications within the field have been occurring at an exponential rate (bottom panel).

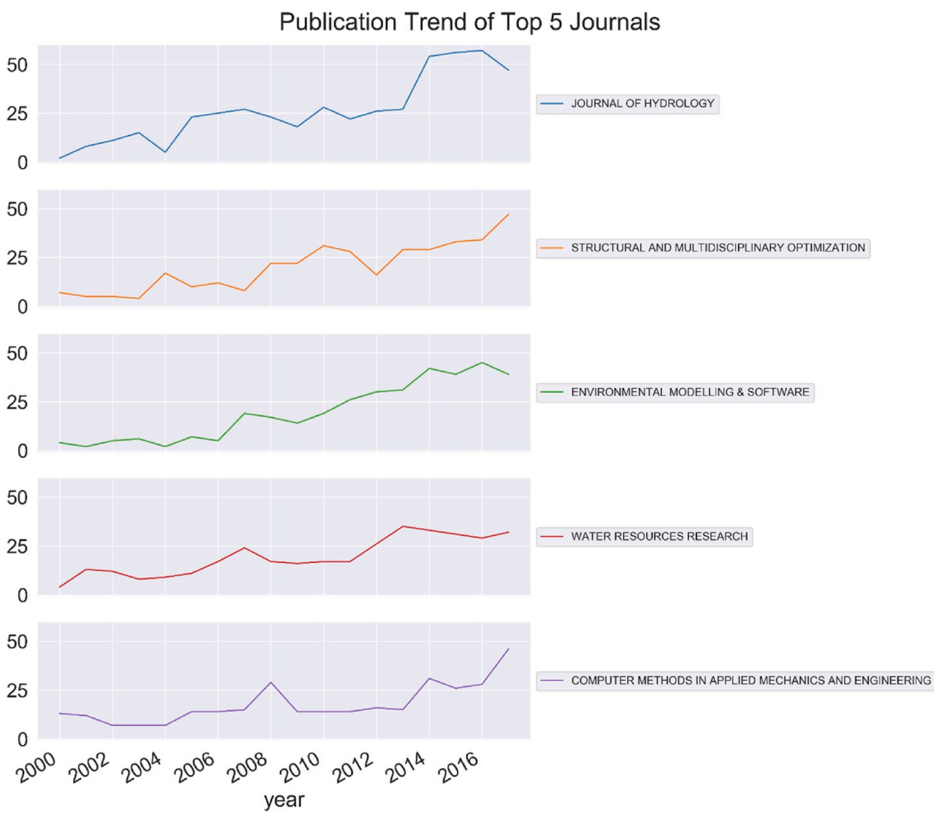


Fig. 6. Publication trend by journal across the timeframe. All journals in the top 5 (by number of publications) saw an increase in publications related to the keywords used.

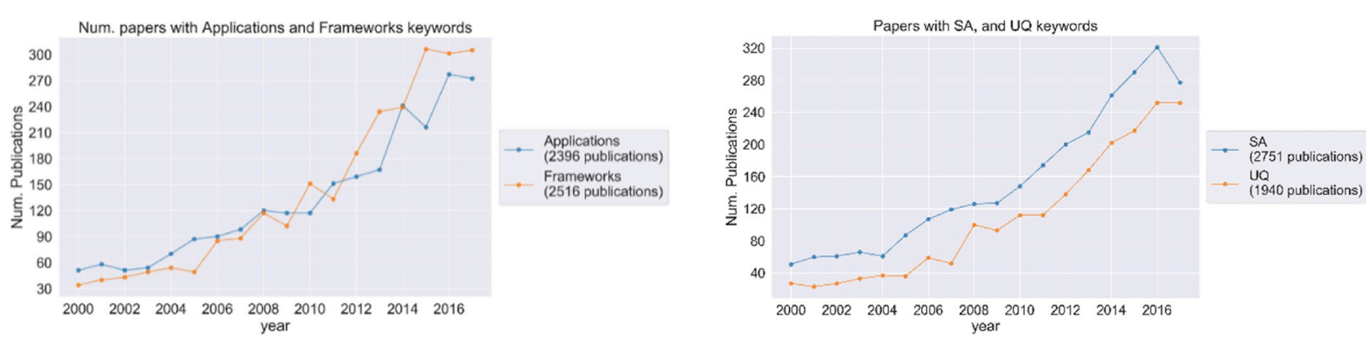


Fig. 7. Publication trends of papers relating to application of UA/SA and frameworks (left) as well as those related to sensitivity analysis, and uncertainty analysis and quantification (right). While publications are increasing in absolute terms, relative to their respective corpora, works on uncertainty frameworks are increasing while SA related papers have decreased, indicating a shift in focus.

make use of such software).

4.4.2. *Recent developments*

Recent impactful publications in sensitivity analysis suggest a shift away from local sensitivity methods. Prior to 2010, 'one-factor-at-a-time' (OAT) local SA was the most prevalent practice in the literature (Saltelli and Annoni, 2010) with a later revisit indicating that while this was still the case for papers published in Science and Nature, GSA methods were gaining traction (Ferretti et al., 2016). A more recent bibliometric review conducted by Saltelli et al. (2019) comes to a similar conclusion across 19 subject areas in which modeling features heavily, although the growth of OAT-related publications is shown to significantly out-pace GSA related publications. Within the presented corpora publications with OAT related keywords do decrease slightly over the past two decades (down roughly 1% compared to the entire corpora), with an uptick in the absolute number of publications post-2010 (see Fig. 10).

Although OAT is said to be a common method (see for example Shin et al., 2013) it may not have featured heavily prior to 2010 due to 1) researchers not reporting OAT use, 2) modelers using custom implementations of OAT, and 3) the software surveyed in our analysis did not support OAT, which discourages modelers from using this method (i.e. they select from available methods). Analysis conducted here indicates an increase in reported GSA keywords post-2010 (Fig. 11) – after the publication of "How to avoid a perfunctory sensitivity analysis" (Saltelli and Annoni, 2010). This paper was identified as a highly cited publication in the initial corpora (Table 6 in the supplementary material), a key contribution being the demonstrated inefficacy of OAT analyses using a geometric proof. The uptick in publications with the OAT-related keywords appears to correlate with the number of papers citing the paper by Saltelli and Annoni (2010), shown in Fig. 12. This may contribute to the rise in publications with OAT related keywords in the

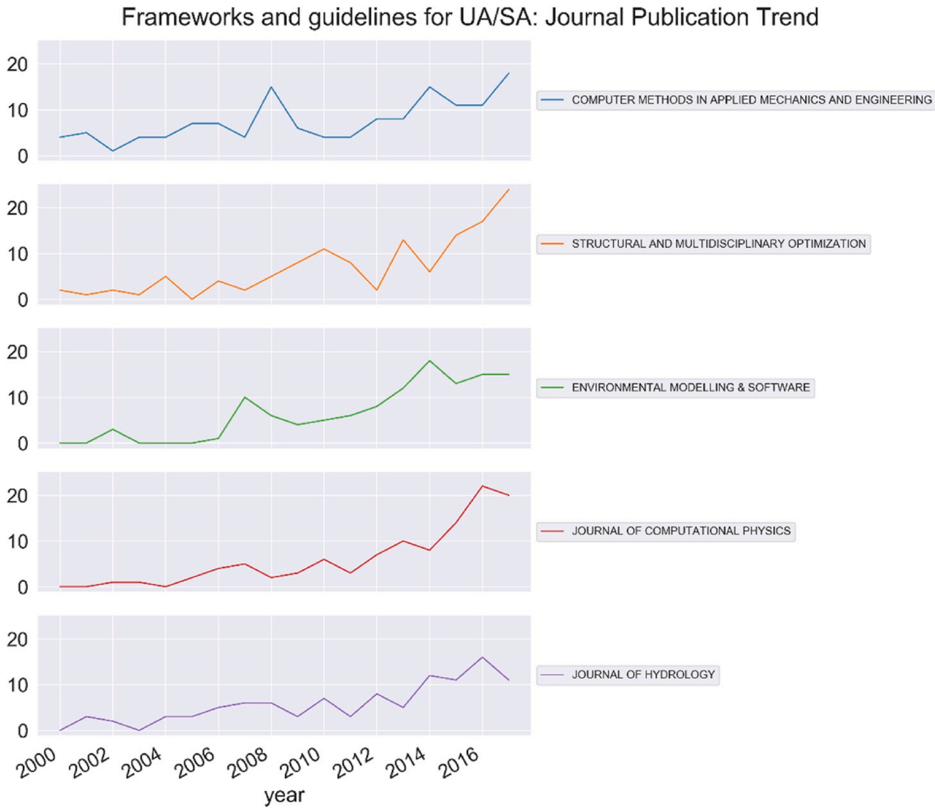


Fig. 8. The top five active journals publishing papers related to UA/SA frameworks. All journals have an increasing publication trend over the given timeframe.

corpora and as identified by Saltelli et al. (2019). The detected increase in GSA papers may reflect the start of changing attitudes towards SA in recognition of the importance of global sensitivity analyses. Increased awareness in the past decade has led to the use and development of more efficient and comprehensive UA/SA techniques and approaches.

Improved approaches put forth in the past decade attempt to enhance the computational efficiency of generating a global sensitivity measure (or range of measures as the case may be) from a single sample set, itself said to be more representative of the possible parameter space (e.g. Razavi et al., 2019). In particular there has been a renewed interest in GSA based on (statistical) design of experiment approaches, as these methods are capable of producing global sensitivity measures at an

Fig. 9. Publication trend of papers with keywords relating to best practices. Notice the larger volume of publications in the years 2014–2017.

acceptable computational cost (Gan et al., 2014; Saltelli, 2017). Such approaches refer to methods that utilize a deterministic sample set, for example the aforementioned Sobol', Latin Hypercube, and Morris methods (Saltelli, 2017).

Despite the increased interest in GSA evidenced by the bibliometric analysis, local SA and OAT methods are still in widespread use, if any SA is conducted at all. Shin et al. (2013) for example, found that only 7% (11 of 164) of papers surveyed conducted any SA, of which five applied OAT. It is difficult to ascertain the full extent of OAT analysis through keyword analysis, as researchers applying this technique may not make explicit reference to this form of analysis. Possible reasons for the relatively slow uptake of GSA methods are listed in Ferretti et al. (2016), including perceived complexity in the application of GSA. Modelers were characterized as being hesitant due to a lack of experience with GSA methods. We also find in the literature a prevalence of self-implemented UA/SA; that is, modelers using their own code in place of existing and often open-source software tools. Not using, or otherwise contributing to, readily available, widely used, and well-tested software represents a duplication of work. This can be somewhat alleviated by greater awareness of and access to the available software tools that simplify the application and use of such analyses. Those developing tools and methods, for their part, could strive to improve ease of use and lower the technical and conceptual barriers to uptake of their software.

Pianosi et al. (2016) outline three principles of good practice for a sensitivity analysis package: 1) the ability to apply multiple sensitivity analyses to one sample, 2) provision of tools to assess and revise user choices, and 3) inclusion of visualization tools. Regarding point 1, early software releases tended to be platform, method, or model specific (see Table 4 for specific examples). In recent years the available software has been made for more general-purpose use, offering a more comprehensive approach to UA/SA with multiple methods supported. The lack of collaborative development is also reportedly an issue, with researchers preferring to develop their own toolset and as a consequence siloing advances, at least in the short to medium term. Usability, especially for

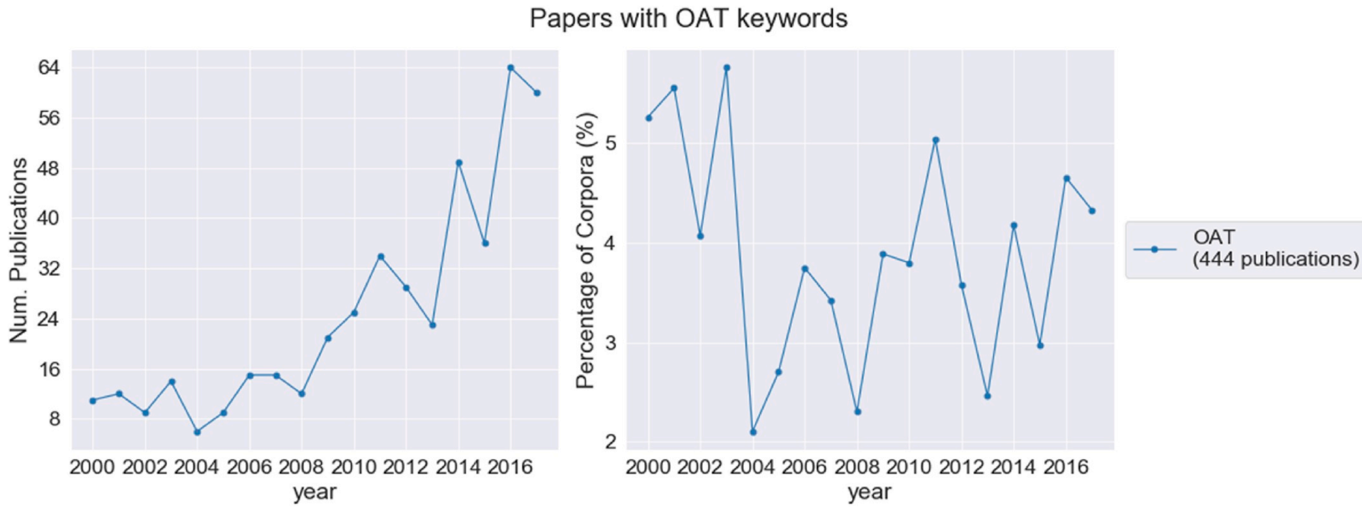


Fig. 10. Absolute and relative trends of publications with OAT keywords. Although publications increase in absolute terms, relative to the corpora yearly publications with OAT keywords decrease over the timeframe.

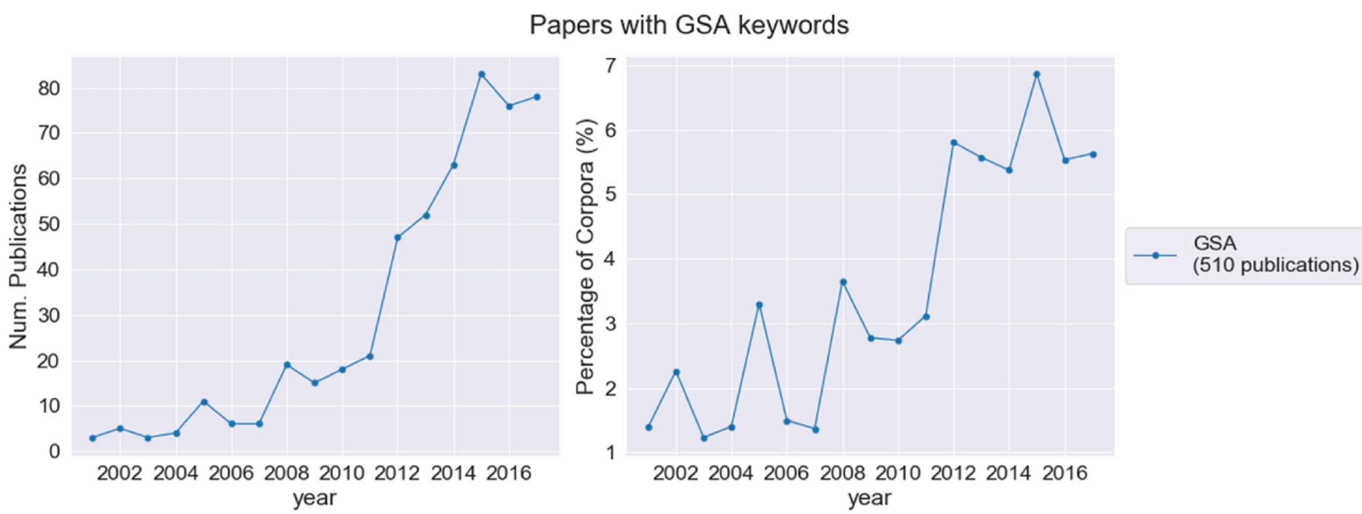


Fig. 11. Absolute and relative trends of publications with GSA keywords. Publications with GSA keywords increase over the timeframe both in absolute and relative terms.

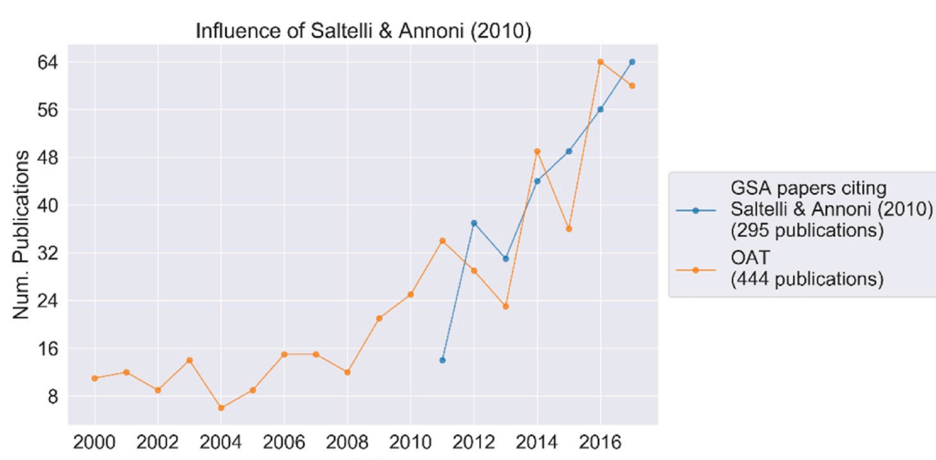


Fig. 12. Papers citing Saltelli and Annoni (2010) appear to be driving the uptick in publications with OAT-related keywords post-2010, possibly as authors give their reason(s) for not relying on OAT.

novices, is an ongoing concern. While many sampling and analysis approaches are amenable to cross-use (e.g. a mix-and-match approach) there is often no limitation in the application of methods (within the packages) which safeguards a user against inappropriate and incompatible mixes, e.g. Sobol' analysis on a Latin Hypercube sample.

Efforts to address these issues and criticisms are evident in the various communities, however, with later packages often offering detailed documentation including usage examples and tutorials (see previous section). Well-known test functions, such as the Ishigami function (Ishigami and Homma, 1990), Sobol' G-Function (Saltelli and Sobol, 1995), the example Lake Problem (Hadka et al., 2015) as well as case studies for research and educational purposes, are often included.

PEST (Doherty, 2018), for example, provides a UA tutorial including two worked examples of hydrological models. Another example is SAFE (Pianosi et al., 2015), which, by providing commented code in workflow scripts, allows beginner users to implement UA/SA more easily and advanced users to improve their methodology. The packages in our survey did not appear to provide guidance through UA/SA theory outside of extensive reading lists, although it is acknowledged that this may be out of scope for those maintaining the tool. A lack of guidance as such may hinder uptake by practitioners of both the software and GSA in general. Generally, the available packages still operate on an understanding that users have appropriate background knowledge of or experience with UA/SA. Many of the packages identified in this review appear to have been abandoned: this perhaps indicates the importance of an active user community to share knowledge and update code.

The corpora give evidence to newer methods that are in development and reflects a continued interest in improving UA/SA. Examples of more recently developed techniques are VARS and active subspaces (developed in 2016 and 2011 respectively). VARS (Variogram Analysis of Response Surfaces) uses variograms as a measure of sensitivity. A variogram is a function describing the spatial dependence of, in the case of SA, the parameter space, that is, how much the variance in parameter values is dependent on the distance between parameters in parameter space. Variogram-related publications began in 2001,however, those specifically relating to the application of variograms to SA only appear from 2016. There are five publications relevant to variogram-based SA in the corpora, totaling 74 citations and the highest citation average is 13 (Razavi and Gupta, 2016).

The top-cited variogram paper (Razavi and Gupta, 2016) presents the method as a linkage between existing derivative- and variance-based GSA methods and demonstrates that the approach reduces computational cost. Over approximately 20 000 and 100 000 model runs, the VARS sensitivity estimates had less uncertainty than Sobol and Morris indices. These relatively new methods, though currently lacking citations, do appear to be methods with development potential due to, for example, current user interest and improvements to the efficiency and comprehensibility of UA/SA methods.

Active subspaces is a dimension reduction technique that identifies directions in parameter space that have a greater influence on the model output. These directions are described as being "active" and their identification aids in reducing the dimensionality of a model by avoiding perturbations across inactive areas of parameter space, thereby reducing computational cost (Constantine et al., 2015). Through this method parameters of importance and their rankings can be obtained (Jefferson et al., 2015). Papers relating to active subspaces first appear in 2015, there are eight in total in the corpora. Citation analysis does not indicate particularly that this new method is being taken up quickly, total citations for all publications was 83, and the publication with the highest citation average had an average of 7.67 (Constantine et al., 2015). The top-cited active subspaces paper (Constantine et al., 2015) details an application of the method to numerical simulation, and an implementation may be found in the 'Effective Quadratures' package for Python (Seshadri and Parks, 2017)

Another technique of interest is HDMR (High Dimensional Model Reduction): the companion paper (Ziehn and Tomlin, 2009) for the method and supporting software came through as a highly cited publication in this analysis (see Table 7). HDMR is an emulation method that improves variance-based SA methods, such as the Sobol' method. Citing articles for Ziehn and Tomlin (2009) continue up to 2019 (identified through manual processes). In fact, 7 of the 32 returned publications in the corpora were published in 2017, indicating a continued interest in the method. In the corpora, the publication with the most citations has 158 (Alış and Rabitz, 2001) and the highest citation average is 14 (Ziehn and Tomlin, 2009).

Furthermore, alternative methods for handling uncertainty have been developed, especially to handle scenarios in which there is large uncertainty, but in which accurate predictions are necessary for future policy making. Software for these alternate methods is deemed out-of-scope for this study but for completeness sake, one such proposed approach is Exploratory Modeling and Analysis. Rather than simply minimizing uncertainty in an attempt to produce an accurate or precise prediction, uncertainty is treated as inevitable. Decision making processes are guided through the exploration of possible outcomes generated through computational experiments and responses planned (Eker et al., 2018; Kwakkel and Pruyt, 2013).

5. Limitations

The bibliometric analysis presented here is limited by the scope of the WoS database, the specific search terms used, the initial time frame and the included fields of study (with the analysis focused on applications in environmental modeling). Search query results may also differ over time due to indexing artefacts with implications for the resulting trend and citation analysis. A bias towards open-source software literature may be perceived as these were the easiest to analyze. That said, it is not claimed that the analysis conducted herein uncovered all software packages currently in use or the full extent to which they are being used.

A known issue is the lack of attributions, citations, and reporting of software used for research, making it difficult to find their mention, especially when the analysis relied on abstract text. Other software may not be referenced simply because their use is taken to be a fundamental part of the (programming) language ecosystem, for example, the R 'sensitivity' package or 'sci-kit learn' (for Python). It was also difficult to search within the corpora for packages with names common to other applications (taking as a particularly difficult example, the R 'sensitivity' package).

In our own process of sorting the generated database, decisions whilst manually sorting and choosing the software collection papers were subject to inherent bias – although this process was kept as transparent and objective as possible (see Notebook 5a "Finding software packages by keyphrase extraction"). Another process which limited the generality of our findings was that of refining the search terms and results. Limiting the scope of the results was necessary to facilitate analysis of the most relevant publications. Iterative use of the topic model achieved this, however, it is entirely possible that relevant

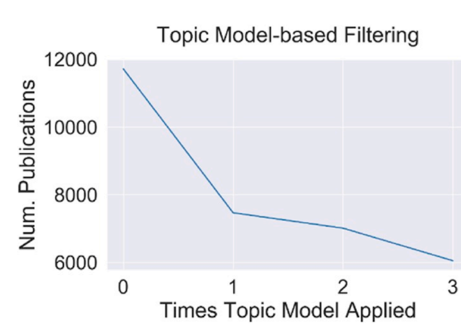


Fig. 13. Plot of the topic model filtering. Note the decrease in publications with each application, which aided in limiting the scope of results but also risked removing relevant publications.

publications will have been removed (Fig. 13). Of particular note is the possible under-representation of articles on emulators and surrogate modeling within the Software corpora. Omitted publications were assumed to be irrelevant or that relevant issues were captured by the papers that remained in the desired corpora. For more information, see Notebook 4 "UASA Topic modeling".

6. Conclusion and future directions

The analysis presented here indicates that UA considerations are increasingly included in the published literature with a slight decrease in the reported use of OAT methods. The identified literature reflects greater attention paid to guidelines for the use of UA/SA over the past decade, itself perhaps indicating advances in the application of UA/SA. Greater interest in the use of UA/SA for rigorous model testing is apparent, although whether modelers embrace and adopt the suggested guidelines towards the treatment, assessment and analysis of UA/SA (e.g. as discussed in Eker et al., 2018; Saltelli et al., 2019) remains to be seen.

The literature also suggests that a wide variety of software has become available in the past two decades, aimed at both non-programmatic audiences and for specific programming languages. The majority of these identified software packages does not support local OAT analyses, which may indicate a general move away from depending on local SA. More recently developed software packages that implement multiple methods with open source code and documentation, with little restriction (in terms of software licencing) to the end-user, are becoming the prevalent distribution format.

While there is a variety of software tools available, the trend of publications on UA/SA tooling has remained largely flat. This trend may be due to the relative infancy of the available tools, or a perceived complexity in their application. For one, while many of the surveyed software provide usage examples and documentation, their use typically assumes 1) experience with the underlying programming language, or 2) intimate familiarity with the methods provided, their pros and cons and contextual suitability. Little guidance is available, aside from extensive reading lists.

The indicated lack of uptake in this analysis may also be because software-specific publications have been largely filtered out from the corpora. These relevant publications may be concentrated within conference proceedings (which were removed from the corpora) or other topic areas not included in the initial publication search. Publications that are application/method focused may not explicitly mention the software used in the abstract. For these reasons it is difficult to concretely conclude whether those involved in environmental modeling are embracing the available UA/SA software tools or if custom "home-grown" solutions are preferred, itself indicating perhaps a lack of awareness of the available software packages.

That said, usability and user-friendliness were found to be a general issue. Users are expected to be adept and experienced enough to produce and interpret results themselves. Even in cases where visualization processes are provided, users may require a different approach for their analyses. In the case of novices, interpreting provided analyses requires first understanding a body of work usually provided in the form of a (often large) reading list of relevant papers. This may explain, in part, a preference for custom "home-grown" solutions where the developers write tools specific to their needs to avoid "adoption cost"; time needed to learn how to use an existing tool effectively. The complexity of existing tools, real or perceived, may contribute to the issue of (lack of) uptake. In cases where the perceived cost of adoption is high, the prospective user may find it easier to apply OAT or otherwise implement their own custom solution to perform common UA/SA methods, which amounts to duplication of effort across the scientific community.

This then raises the question of what constitutes a 'thorough' UA/SA package. In this survey, the most comprehensive software (R sensitivity, Simlab, and SALib) provide users with the widest assortment of UA/SA methods with (limited) visualization capability and test functions. These target languages prevalent in the sciences (R, Matlab, and Python respectively) that are supported by an active community, which may explain their longevity and/or popularity. The prevalence of open-source, community-led efforts evident in more recent software tools suggests that an open development culture is a prerequisite to widespread adoption – perhaps an unremarkable observation due to the scientific context and focus of UA/SA research.

Developers and maintainers of UA/SA tools could support and encourage wider application of GSA processes by moving towards 1) an open development process, 2) placing further attention on expanding documentation, preferably in an easily digestible form, and 3) improving usage guidelines and promoting user-centric interfaces and workflows.

Point 1 is to encourage the sharing of knowledge and experience across the disciplines that rely on modeling, to leverage expertise and experience globally rather than siloing advances. On point 2, UA/SA software developers could further leverage the open-collaboration model and (re-)use explanations and examples from one another. Examples of both simple and complex workflows could be given (e.g. in a "cookbook" or "recipe" documentation style). Point 3 should not be taken to mean that all packages should provide a GUI. Rather, general-purpose UA/SA tools should have processes in place to prevent or limit unintentional or ill-informed analyses from occurring. A particular pain point is the ability to mix-and-match sampling and analysis methods regardless of whether it makes sense to do so.

While UA/SA tools have largely addressed the three steps defined by Pianosi et al. (2015) (sample parameter space, run model, analyze results), the workflow - that is implicit or explicit steps in the use and application of the software - could be improved so that modelers are able to move from each step without issue. Recently developed packages indicate that such improvements to the workflow are being made, with attention to usability, open-source code, and tools for analyzing results. Researchers and modelers, particularly those new to UA/SA, need software designed with usability in mind. It is expected that such software will support UA/SA in more areas and encourage rigorous and reliable UA/SA, which will in turn allow for more informed decision-making.

Software availability

Code and representative data used for this analysis can be found at https://github.com/frog7/uasa-trends (10.5281/zenodo.3406946).

Software used to support analysis can be found at https://github.com/ConnectedSystems/wosis (10.5281/zenodo.3406947).

Declaration of competing interest

The second author has contributed usability and performance improvements to the SALib Python library. All other authors declare no potential sources of conflict.

Acknowledgements

The corresponding author (second in the author list) is supported through an Australian Government Research Training Program (AGRTP) Scholarship, and a top-up scholarship from the Hilda-John Endowment Fund.

The authors would like to thank and acknowledge Joseph Guillaume, Barbara Robson, and the anonymous reviewers for their highly valued comments and suggestions. The use of work by Titipat Achakulvisut (author of the 'wos_parser' Python package, https://github.com/titipata/wos_parser) and Enrico Bacis (author of the 'WoS Client' package, https://github.com/enricobacis/wos) is acknowledged. We would also thank and acknowledge Clarivate Analytics for providing access and use of the Web of Science database, and to their staff for providing clarifications to technical details of the available API without which this work would not have been possible.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envsoft.2019.104588.

References

Achakulvisut, T., Acuna, D.E., Ruangrong, T., Kording, K., 2016. Science concierge: a fast content-based recommendation system for scientific publications. PLoS One 11, e0158423. https://doi.org/10.1371/journal.pone.0158423.

Adams, B.M., Ebeida, M.S., Eldred, M.S., Gianluca, G., Jakeman, J.D., Maupin, K.A., Monschke, J.A., Adam Stephens, J., Swiler, L.P., Vigil, D.M., Wildley, T.M., Bohnhoff, W.J., Dalbey, K.R., Eddy, J.P., Frye, J.R., Hooper, R.W., Hu, K.T., Hough, P.D., Khalil, M., Ridgway, E.M., Winokur, J.G., Rushdi, A., 2010. Dakota : a Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis. Version 5.0, User's Manual. (No. SAND2010-2183, 991842). https://doi.org/10.2172/991842.

Alış, Ö.F., Rabitz, H., 2001. Efficient implementation of high dimensional model representations. J. Math. Chem. 29, 127–142. https://doi.org/10.1023/A:1010979129659.

Arora, S., Ge, R., Moitra, A., 2012. Learning Topic Models - Going beyond SVD.

Ascough II, J.C., Fischer, C., Lighthart, N.P., David, O., Green, T.R., Kralisch, S., 2015. The model optimization, uncertainty, and sensitivity analysis (MOUSE) Toolbox: overview and application. In: Proc. Hydrology Days. Presented at the Proc. Hydrology Days 2015. Colorado State University, Fort Collins, Colorado, pp. 17–28.

Becker, W.E., Tarantola, S., Deman, G., 2018. Sensitivity analysis approaches to high-dimensional screening problems at low sample size. J. Stat. Comput. Simul. 88, 2089–2110. https://doi.org/10.1080/00949655.2018.1450876.

Beel, J., Gipp, B., Langer, S., Breitinger, C., 2016. Research-paper recommender systems: a literature survey. Int. J. Digit. Libr. 17, 305–338. https://doi.org/10.1007/s00799-015-0156-0.

Belyaev, M., Burnaev, E., Kapushev, E., Panov, M., Prikhodko, P., Vetrov, D., Yarotsky, D., 2016. GTApprox: surrogate modeling for industrial design. Adv. Eng. Software 102, 29–39. https://doi.org/10.1016/j.advengsoft.2016.09.001.

Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S. A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising performance of environmental models. Environ. Model. Softw 40, 1–20.

Beven, K., Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrol. Process. 6, 279–298. https://doi.org/10.1002/hyp.3360060305.

Bezanson, J., Edelman, A., Karpinski, S., Shah, V., 2017. Julia: a fresh approach to numerical computing. SIAM Rev. 59, 65–98. https://doi.org/10.1137/141000671.

Blatman, G., Sudret, B., 2011. Adaptive sparse polynomial chaos expansion based on least angle regression. J. Comput. Phys. 230, 2345–2367. https://doi.org/10.1016/j.jcp.2010.12.021.

Bornmann, L., Mutz, R., 2015. Growth rates of modern science: a bibliometric analysis based on the number of publications and cited references. J. Assoc. Inf. Sci. Technol. 66, 2215–2222. https://doi.org/10.1002/asi.23329.

Campolongo, F., Saltelli, A., Cariboni, J., 2011. From screening to quantitative sensitivity analysis. A unified approach. Comput. Phys. Commun. 182, 978–988. https://doi.org/10.1016/j.cpc.2010.12.039.

Castaings, W., Borgonovo, E., Morris, M.D., Tarantola, S., 2012. Sampling strategies in density-based sensitivity analysis. Environ. Model. Softw 38, 13–26. https://doi.org/10.1016/j.envsoft.2012.04.017.

Chen, Y., Zhang, H., Liu, R., Ye, Z., Lin, J., 2019. Experimental explorations on short text topic mining between LDA and NMF based Schemes. Knowl. Based Syst. 163, 1–13. https://doi.org/10.1016/j.knosys.2018.08.011.

Constantine, P.G., Emory, M., Larsson, J., Iaccarino, G., 2015. Exploiting active subspaces to quantify uncertainty in the numerical simulation of the HyShot II scramjet. J. Comput. Phys. 302, 1–20. https://doi.org/10.1016/j.jcp.2015.09.001.

Crestaux, T., Le Maitre, O., Martinez, J.-M., 2009. Polynomial chaos expansion for sensitivity analysis. Reliab. Eng. Syst. Saf. 94, 1161–1172. https://doi.org/10.1016/j.ress.2008.10.008.

Cukier, R.I., Fortuin, C.M., Shuler, K.E., Petschek, A.G., Schaibly, J.H., 1973. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I theory. J. Chem. Phys. 59, 3873–3878. https://doi.org/10.1063/1.1680571.

Czitrom, V., 1999. One-Factor-at-a-Time versus designed experiments. Am. Stat. 53, 126–131. https://doi.org/10.1080/00031305.1999.10474445.

Doherty, J., 2018. Model-Independent Parameter Estimation User Manual Part I: PEST, SENSAN and Global Optimisers.

Douglas-Smith, D., Iwanaga, T., 2019. UASA Trends.

Eker, S., Rovenskaya, E., Obersteiner, M., Langan, S., 2018. Practice and perspectives in the validation of resource management models. Nat. Commun. 9, 1–10. https://doi.org/10.1038/s41467-018-07811-9.

Fedra, K., 1983. Environmental Modeling under Uncertainty: Monte Carlo Simulation [WWW Document]. URL. http://pure.iiasa.ac.at/id/eprint/2152/ (accessed 4.10.19).

Feinberg, J., Langtangen, H.P., 2015. Chaospy: an open source tool for designing methods of uncertainty quantification. J. Comput. Sci. 11, 46–57. https://doi.org/10.1016/j.jocs.2015.08.008.

Ferretti, F., Saltelli, A., Tarantola, S., 2016. Trends in sensitivity analysis practice in the last decade. Sci. Total Environ. 568, 666–670. https://doi.org/10.1016/j.scitotenv.2016.02.133.

Gan, Y., Duan, Q., Gong, W., Tong, C., Sun, Y., Chu, W., Ye, A., Miao, C., Di, Z., 2014. A comprehensive evaluation of various sensitivity analysis methods: a case study with a hydrological model. Environ. Model. Softw 51, 269–285. https://doi.org/10.1016/j.envsoft.2013.09.031.

Guillaume, J.H.A., Jakeman, J.D., Marsili-Libelli, S., Asher, M., Brunner, P., Croke, B., Hill, M.C., Jakeman, A.J., Keesman, K.J., Razavi, S., Stigter, J.D., 2019. Introductory overview of identifiability analysis: A guide to evaluating whether you have the right type of data for your modeling purpose. Environmental Modelling & Software 119, 418–432. https://doi.org/10.1016/j.envsoft.2019.07.007.

Haddaway, N.R., Westgate, M.J., 2018. Predicting the time needed for environmental systematic reviews and systematic maps. Conserv. Biol. 0 https://doi.org/10.1111/cobi.13231.

Hadka, D., Herman, J., Reed, P., Keller, K., 2015. An open source framework for many-objective robust decision making. Environ. Model. Softw 74, 114–129. https://doi.org/10.1016/j.envsoft.2015.07.014.

Helton, J.C., 1993. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. Reliab. Eng. Syst. Saf. 42, 327–367. https://doi.org/10.1016/0951-8320(93)90097-I.

Herman, J., Usher, W., 2017. SALib: an open-source Python library for Sensitivity Analysis [WWW Document] J. Open Source Softw. 2 (9), 97. https://doi.org/10.21105/joss.00097.

Hough, B., Fu, C., Paliwal, S., 2016. savvy: visualize high dimensionality sensitivity analysis data. Updated with full sensitivity analysis from ligpy model. Zenodo. https://doi.org/10.5281/zenodo.53099.

Hunt, M., Haley, B., McLennan, M., Koslowski, M., Murthy, J., Strachan, A., 2015. PUQ: a code for non-intrusive uncertainty propagation in computer simulations. Comput. Phys. Commun. 194, 97–107. https://doi.org/10.1016/j.cpc.2015.04.011.

Iman, R.L., Helton, J.C., 1988. An investigation of uncertainty and sensitivity analysis techniques for computer models. Risk Anal. 71–90. https://doi.org/10.1111/j.1539-6924.1988.tb01155.x.

Iooss, B., Pujol, A.J., Boumhaout, G., with contributions from K, Veiga, S.D., Delage, T., Fruth, J., Gilquin, L., Guillaume, J., Gratiet, L.L., Lemaitre, P., Nelson, B.L., Monari, F., Oomen, R., Rakovec, O., Ramos, B., Roustant, O., Song, E., Staum, J., Sueur, R., Touati, T., Weber, F., 2018. Sensitivity: Global Sensitivity Analysis of Model Outputs.

Ishigami, T., Homma, T., 1990. An importance quantification technique in uncertainty analysis for computer models. In: [1990] Proceedings. First International Symposium on Uncertainty Modeling and Analysis. Presented at the [1990] Proceedings. First International Symposium on Uncertainty Modeling and Analysis, pp. 398–403. https://doi.org/10.1109/ISUMA.1990.151285.

Iwanaga, T., Douglas-Smith, D., 2019. Wosis: Beta-Release. https://doi.org/10.5281/zenodo.3406947.

Jefferson, J.L., Gilbert, J.M., Constantine, P.G., Maxwell, R.M., 2015. Active subspaces for sensitivity analysis and dimension reduction of an integrated hydrologic model. Comput. Geosci. 83, 127–138. https://doi.org/10.1016/j.cageo.2015.07.001.

JRC, 2015. SIMLAB and Other Software - EU Science Hub - European Commission [WWW Document]. EU Science Hub. URL. https://ec.europa.eu/jrc/en/samo/simlab (accessed 12.5.18).

Kuczera, G., Kavetski, D., Franks, S., Thyer, M., 2006. Towards a Bayesian total error analysis of conceptual rainfall-runoff models: characterising model error using storm-dependent parameters. J. Hydrol. 331, 161–177. https://doi.org/10.1016/j.jhydrol.2006.05.010.

Kwakkel, J.H., Pruyt, E., 2013. Exploratory Modeling and Analysis, an approach for model-based foresight under deep uncertainty. Technol. Forecast. Soc. Chang. 80, 419–431. https://doi.org/10.1016/j.techfore.2012.10.005.

Leamer, E.E., 1985. Sensitivity analyses would help. Am. Econ. Rev. 75, 308–313.

Marelli, S., Sudret, B., 2014. UQLab: a framework for uncertainty quantification in Matlab. In: Vulnerability, Uncertainty, and Risk. Presented at the Second International Conference on Vulnerability and Risk Analysis and Management (ICVRAM) and the Sixth International Symposium on Uncertainty, Modeling, and Analysis (ISUMA). American Society of Civil Engineers, Liverpool, UK, pp. 2554–2563. https://doi.org/10.1061/9780784413609.257.

Matott, L.S., 2017. OSTRICH: an Optimization Software Tool, Documentation and User's Guide. Version 17.12.19 [WWW Document]. URL. http://www.eng.buffalo.edu/~lsmatott/Ostrich/OstrichMain.html (accessed 3.24.19).

Matott, L.S., Babendreier, J.E., Purucker, S.T., 2009. Evaluating uncertainty in integrated environmental models: a review of concepts and tools. Water Resour. Res. 45 https://doi.org/10.1029/2008WR007301.

McKay, M.D., Beckman, R.J., Conover, W.J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21, 239–245. https://doi.org/10.2307/1268522.

Metropolis, N., Ulam, S., 1949. The Monte Carlo method. J. Am. Stat. Assoc. 44, 335–341. https://doi.org/10.2307/2280232.

Moens, D., Vandepitte, D., 2005. A survey of non-probabilistic uncertainty treatment in finite element analysis. Comput. Methods Appl. Mech. Eng. 194, 1527–1555. https://doi.org/10.1016/j.cma.2004.03.019.

Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments. Technometrics 33, 161–174. https://doi.org/10.2307/1269043.

Nakagawa, S., Samarasinghe, G., Haddaway, N.R., Westgate, M.J., O'Dea, R.E., Noble, D. W.A., Lagisz, M., 2018. Research weaving: visualizing the future of research

synthesis. Trends Ecol. Evol. 34 (3), 224–238. https://doi.org/10.1016/j.tree.2018.11.007.
Norton, J., 2015. An introduction to sensitivity assessment of simulation models. Environ. Model. Softw 69, 166–174. https://doi.org/10.1016/j.envsoft.2015.03.020.
Oakley, J., O'Hagan, A., 2004. Probabilistic sensitivity analysis of complex models: a Bayesian approach. J. R. Stat. Soc. 66, 751–769. https://doi.org/10.1111/j.1467-9868.2004.05304.x.
Oladyshkin, S., Nowak, W., 2012. Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion. Reliab. Eng. Syst. Saf. 106, 179–190. https://doi.org/10.1016/j.ress.2012.05.002.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.
Pianosi, F., Beven, K., Freer, J., Hall, J.W., Rougier, J., Stephenson, D.B., Wagener, T., 2016. Sensitivity analysis of environmental models: a systematic review with practical workflow. Environ. Model. Softw 79, 214–232. https://doi.org/10.1016/j.envsoft.2016.02.008.
Pianosi, F., Sarrazin, F., Wagener, T., 2015. A Matlab toolbox for global sensitivity analysis. Environ. Model. Softw 70, 80–85. https://doi.org/10.1016/j.envsoft.2015.04.009.
Poeter, E.P., Hill, M.C., 1999. UCODE, a computer code for universal inverse modeling1Code. Comput. Geosci. 25, 457–462. https://doi.org/10.1016/S0098-3004(98)00149-6 available at: http://water.usgs.gov/software/ground_ water.html1.
Rabby, G., Azad, S., Mahmud, Mufti, Zamli, K.Z., Mostafizur Rahman, M., 2018. A flexible keyphrase extraction technique for academic literature. In: Procedia Computer Science, the 3rd International Conference on Computer Science and Computational Intelligence (ICCSCI 2018) : Empowering Smart Technology in Digital Era for a Better Life, vol. 135, pp. 553–563. https://doi.org/10.1016/j.procs.2018.08.208.
Rakovec, O., Hill, M.C., Clark, M.P., Weerts, A.H., Teuling, A.J., Uijlenhoet, R., 2014. Distributed evaluation of local sensitivity analysis (DELSA), with application to hydrologic models. Water Resour. Res. 50, 409–426. https://doi.org/10.1002/2013WR014063.
Ratto, M., Pagano, A., 2010. Using recursive algorithms for the efficient identification of smoothing spline ANOVA models. ASA Adv. Stat. Anal. 94, 367–388. https://doi.org/10.1007/s10182-010-0148-8.
Razavi, S., Gupta, H.V., 2016. A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory. Water Resour. Res. 52, 423–439. https://doi.org/10.1002/2015WR017559.
Razavi, S., Gupta, H.V., 2015. What do we mean by sensitivity analysis? The need for comprehensive characterization of "global" sensitivity in Earth and Environmental systems models: a Critical Look at Sensitivity Analysis. Water Resour. Res. 51, 3070–3092. https://doi.org/10.1002/2014WR016527.
Razavi, S., Sheikholeslami, R., Gupta, H.V., Haghnegahdar, A., 2019. VARS-TOOL: a toolbox for comprehensive, efficient, and robust sensitivity and uncertainty analysis. Environ. Model. Softw 112, 95–107. https://doi.org/10.1016/j.envsoft.2018.10.005.
Refsgaard, J.C., van der Sluijs, J.P., Hojberg, A.L., Vanrolleghem, P.A., 2007. Uncertainty in the environmental modelling process - a framework and guidance. Environ. Model. Softw 22, 1543–1556. https://doi.org/10.1016/j.envsoft.2007.02.004.
Regier, J., Fischer, K., Pamnany, K., Noack, A., Revels, J., Lam, M., Howard, S., Giordano, R., Schlegel, D., McAuliffe, J., Thomas, R., Prabhat, 2019. Cataloging the visible universe through Bayesian inference in Julia at petascale. J. Parallel Distrib. Comput. 127, 89–104. https://doi.org/10.1016/j.jpdc.2018.12.008.
Roos, M., Martins, T.G., Held, L., Rue, H., 2015. Sensitivity analysis for bayesian hierarchical models. Bayesian Anal. 10, 321–349. https://doi.org/10.1214/14-BA909.
Rose, S., Engel, D., Kogan, J., 2010. Automatic keyword extraction from individual documents. In: Text Mining: Applications and Theory. John Wiley & Sons, Ltd, pp. 1–20.
Roy, C.J., Oberkampf, W.L., 2011. A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing. Comput. Methods Appl. Mech. Eng. 200, 2131–2144. https://doi.org/10.1016/j.cma.2011.03.016.
Sagi, O., Rokach, L., 2018. Ensemble learning: a survey. Wiley Interdiscip. Rev.: Data Min. Knowl. Discov. 8, e1249. https://doi.org/10.1002/widm.1249.
Saltelli, A., 2017. Discussion Paper: Should Statistics Rescue Mathematical Modelling? arXiv:1712.06457 [stat].
Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., Wu, Q., 2019. Why so many published sensitivity analyses are false: a systematic review of sensitivity analysis practices. Environ. Model. Softw 114, 29–39.
Saltelli, A., Annoni, P., 2010. How to avoid a perfunctory sensitivity analysis. Environ. Model. Softw 25, 1508–1517. https://doi.org/10.1016/j.envsoft.2010.04.012.
Saltelli, A., Marivoet, J., 1990. Non-parametric statistics in sensitivity analysis for model output: a comparison of selected techniques. Reliab. Eng. Syst. Saf. 28, 229–253. https://doi.org/10.1016/0951-8320(90)90065-U.
Saltelli, A., Sobol, I.M., 1995. About the use of rank transformation in sensitivity analysis of model output. Reliab. Eng. Syst. Saf. 50, 225–239. https://doi.org/10.1016/0951-8320(95)00099-2.
Saltelli, A., Tarantola, S., 2002. On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal. J. Am. Stat. Assoc. 97, 702–709.
Saltelli, A., Tarantola, S., Chan, K.P.S., 1999. A quantitative model-independent method for global sensitivity analysis of model output. Technometrics 41, 39–56.
Seshadri, P., Parks, G., 2017. Effective-Quadratures (EQ): Polynomials for Computational Engineering Studies. Journal of Open Source Software 2 (11), 166. https://doi.org/10.21105/joss.00166.
Shan, S., Wang, G.G., 2010. Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions. Struct. Multidiscip. Optim. 41, 219–241. https://doi.org/10.1007/s00158-009-0420-2.
Shin, M.-J., Guillaume, J.H.A., Croke, B.F.W., Jakeman, A.J., 2013. Addressing ten questions about conceptual rainfall–runoff models with global sensitivity analyses in R. Journal of Hydrology 503, 135–152. https://doi.org/10.1016/j.jhydrol.2013.08.047.
Sigmund, O., Maute, K., 2013. Topology optimization approaches: a comparative review. Struct. Multidiscip. Optim. 48, 1031–1055. https://doi.org/10.1007/s00158-013-0978-6.
Sobol, I.M., 1993. Sensitivity analysis for non-linear mathematical models. Math. Model. Civ. Eng. 1, 407–414.
Sobol, I.M., Kucherenko, S., 2009. Derivative based global sensitivity measures and their link with global sensitivity indices. Math. Comput. Simulat. 79, 3009–3017. https://doi.org/10.1016/j.matcom.2009.01.023.
Spear, R.C., Hornberger, G.M., 1980. Eutrophication in peel inlet—II. Identification of critical uncertainties via generalized sensitivity analysis. Water Res. 14, 43–49.
Storlie, C., Helton, J., 2008. Multiple predictor smoothing methods for sensitivity analysis: description of techniques. Reliab. Eng. Syst. Saf. 93, 28–54. https://doi.org/10.1016/j.ress.2006.10.012.
Sun, X.Y., Newham, L.T.H., Croke, B.F.W., Norton, J.P., 2012. Three complementary methods for sensitivity analysis of a water quality model. Environ. Model. Softw 37, 19–29. https://doi.org/10.1016/j.envsoft.2012.04.010.
Tarantola, S., Gatelli, D., Mara, T.A., 2006. Random balance designs for the estimation of first order global sensitivity indices. Reliab. Eng. Syst. Saf. 91, 717–727. https://doi.org/10.1016/j.ress.2005.06.003.
van Dijk, N.P., Maute, K., Langelaar, M., van Keulen, F., 2013. Level-set methods for structural topology optimization: a review. Struct. Multidiscip. Optim. 48, 437–472. https://doi.org/10.1007/s00158-013-0912-y.
Vu-Bac, N., Lahmer, T., Zhuang, X., Nguyen-Thoi, T., Rabczuk, T., 2016. A software framework for probabilistic sensitivity analysis for computationally expensive models. Adv. Eng. Software 100, 19–31. https://doi.org/10.1016/j.advengsoft.2016.06.005.
Wagener, T., Kollat, J., 2007. Numerical and Visual Evaluation of Hydrological and Environmental Models Using the Monte Carlo Analysis Toolbox.
Wang, C., Duan, Q., Tong, C.H., Di, Z., Gong, W., 2016. A GUI platform for uncertainty quantification of complex dynamical models. Environ. Model. Softw 76, 1–12. https://doi.org/10.1016/j.envsoft.2015.11.004.
Wang, J., Li, X., Lu, L., Fang, F., 2013. Parameter sensitivity analysis of crop growth models based on the extended Fourier Amplitude Sensitivity Test method. Environ. Model. Softw 48, 171–182. https://doi.org/10.1016/j.envsoft.2013.06.007.
Westgate, M.J., Haddaway, N.R., Cheng, S.H., McIntosh, E.J., Marshall, C., Lindenmayer, D.B., 2018. Software support for environmental evidence synthesis. Nat. Ecol. Evol. (in press).
Westgate, M.J., Lindenmayer, D.B., 2017. The difficulties of systematic reviews. Conserv. Biol. 31, 1002–1007. https://doi.org/10.1111/cobi.12890.
Yang, J., Reichert, P., Abbaspour, K.C., Xia, J., Yang, H., 2008. Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. J. Hydrol. 358, 1–23. https://doi.org/10.1016/j.jhydrol.2008.05.012.
Young, P.C., Spear, R.C., Hornberger, G.M., 1978. Modeling badly defined systems: soem further thoughts. In: Proceedings SIMSIG Conference. Presented at the SIMSIG, Canberra, pp. 24–32.
Ziehn, T., Tomlin, A.S., 2009. GUI–HDMR – a software tool for global sensitivity analysis of complex models. Environ. Model. Softw 24, 775–785. https://doi.org/10.1016/j.envsoft.2008.12.002.

Chapter 3: Development of an integrated model for the Campaspe catchment

In this chapter the approach undertaken to develop an integrated socio-environmental model and the surrounding context is outlined. The model described was applied to investigate management of water resources within the Lower Campaspe catchment, a predominantly agricultural area in North-Central Victoria, Australia. Climatic and policy factors are considered for their importance, along with other human drivers which also influence economic and ecological outcomes. Development of the model necessitated cooperation and collaboration with several disciplinary and subject matter experts spanning the natural and human systems. The model was thus developed as a component-based integrated model wherein each system represented as a component within the socio-environmental system-of-systems.

This chapter provides much of the background and context for Chapter 4 and is published as a conference paper published after peer review by two anonymous reviewers in the Proceedings of the International Association of Hydrological Sciences (PIAHS), 8th International Water Resources Management Conference of ICWRS, Beijing, China. The publication is open access under Creative Commons licence CC-BY 4.0.

Iwanaga, T.*, Zare, F., Croke, B., Fu, B., Merritt, W., Partington, D., Ticehurst, J., Jakeman, A., 2018. Development of an integrated model for the Campaspe catchment: a tool to help improve understanding of the interaction between society, policy, farming decision, ecology, hydrology and climate, in: Proceedings of the International Association of Hydrological Sciences. Presented at the Innovative water resources management - understanding and balancing interactions between humankind and nature - 8th International Water Resources Management Conference of ICWRS, Beijing, China, 13-15 June 2018, Copernicus GmbH, pp. 1–12. https://doi.org/10.5194/piahs-379-1-2018

Proc. IAHS, 379, 1–12, 2018
https://doi.org/10.5194/piahs-379-1-2018

Development of an integrated model for the Campaspe catchment: a tool to help improve understanding of the interaction between society, policy, farming decision, ecology, hydrology and climate

Takuya Iwanaga[1], Fateme Zare[1], Barry Croke[1,2], Baihua Fu[1], Wendy Merritt[1], Daniel Partington[3], Jenifer Ticehurst[1], and Anthony Jakeman[1]

[1]Fenner School of Environment and Society, Australian National University, Canberra, Australia
[2]Mathematical Sciences Institute, Australian National University, Canberra, Australia
[3]National Centre for Groundwater Research and Training, Flinders University, Adelaide, Australia

Correspondence: Barry Croke (barry.croke@anu.edu.au)

Received: 20 December 2017 – Revised: 23 January 2018 – Accepted: 3 February 2018 – Published: 5 June 2018

Abstract. Management of water resources requires understanding of the hydrology and hydrogeology, as well as the policy and human drivers and their impacts. This understanding requires relevant inputs from a wide range of disciplines, which will vary depending on the specific case study. One approach to gain understanding of the impact of climate and society on water resources is through the use of an integrated modelling process that engages stakeholders and experts in specifics of problem framing, co-design of the underpinning conceptual model, and discussion of the ensuing results. In this study, we have developed such an integrated modelling process for the Campaspe basin in northern Victoria, Australia. The numerical model built has a number of components:

- Node/link based surface water hydrology module based on the IHACRES rainfall-streamflow model
- Distributed groundwater model for the lower catchment (MODFLOW)
- Farm decision optimisation module (to determine irrigation requirements)
- Policy module (setting conditions on availability of water based on existing rules)
- Ecology module (determining the impacts of available streamflow on platypus, fish and river red gum trees)

The integrated model is component based and has been developed in Python, with the MODFLOW and surface water hydrology model run in external programs, controlled by the master program (in Python). The integrated model has been calibrated using historical data, with the intention of exploring the impact of various scenarios (future climate scenarios, different policy options, water management options) on the water resources. The scenarios were selected based on workshops with, and a social survey of, stakeholders in the basin regarding what would be socially acceptable and physically plausible options for changes in management. An example of such a change is the introduction of a managed aquifer recharge system to capture dam overflows, and store at least a portion of this in the aquifer, thereby increasing the groundwater resource as well as reducing the impact of existing pumping levels.

1 Introduction

Effective, and holistic, water management is contingent on understanding the stressors that affect water resources. Such stressors may come from a variety of physical and social (anthropogenic) sources. In the field of water resource management physical influences typically include the hydrology, geology, ecology/biology, and climatic processes. Social systems that influence and affect water management include agricultural enterprises, water policies, and the social factors that influence the acceptability of water use and management practices. Due to the complexity and interconnected nature of this system of systems, water resource managers and researchers often turn to integrated models to assess potential management actions within the specific context.

Managing such system of systems requires the consideration of a wide range of factors across the interconnected physical and social domains. Integrated Assessment (IA) should not be conducted by individual disciplines in isolation as the issues faced do not fall neatly into traditional academic disciplines. Each represented system may have differing problem frames which influence and affect each other due to the interconnected nature of socio-enviro systems. Therefore, the development of Integrated Assessment Models (IAM) should not be conducted by model developers alone but in conjunction with experts and stakeholders so that the specifics of the problem frame(s) are accounted for. Collectively these problem frames make up a management context, a term used here to refer to the physical properties, the social elements and activities that have influence, and the management scope and objectives of concern.

The lower Campaspe basin in the north-central region of Victoria, Australia is an example of a highly interconnected socio-enviro system. We describe herein an iterative integrative process used to develop an Integrated Assessment Model (IAM) suited for the specific management context. Model development was continuously informed by stakeholder and expert knowledge throughout the process from initial conceptualization through to completion. As new information and knowledge became available and challenges encountered, stakeholders and experts were re-engaged to update the problem frames and model design; a beneficial co-design process. Incorporation of feedback at each iterative stage then helps to ensure that the model remains relevant for its given purpose and to the stakeholders themselves.

The principal aim of the model is to inform stakeholders of the impacts of a range of possible combinatory policy and on-farm water management decisions under a variety of climate conditions. These collectively represent a set of possible "futures". The model will be used in an exploratory manner through which a multitude of such possible "futures" are generated. The combination of factors that led to positive (or at least effective compromises) and negative future conditions can then be identified and communicated to stakeholders through this exploratory process. Because the geophysical, geographical, and social elements are found in a range of contexts, this iterative process is a generally applicable integrative water management approach.

2 Management context

Defining the management context through systems analysis with the aid of stakeholder knowledge is a crucial first step in an integrated assessment process, and a key aspect of Integrated Assessment Modelling (IAM, as in Jakeman and Letcher, 2003). Kraft et al. (2010) argues the importance of stakeholder involvement as incorporation of local domain knowledge ensures that key features of the management context are captured and subsequently represented in the model. Stakeholders further represent an important source of local knowledge which may in turn drive both information need and data accessibility, as well as playing an important role in validating model outputs (Krueger et al., 2012).

The involvement of stakeholders increases the transparency of the development process as it is exposed for critique and review by stakeholders. Through this stakeholder engagement process the scope and objectives of the model can be iteratively developed and refined so that the final model is suitable and relevant (and therefore useful) for the end purpose and users (Jakeman and Letcher, 2003). The process for gathering information and knowledge of the management context and the subsequent influences and implications on the model design and approach is described in later sections.

One motivation for this study was the adoption of the Murray-Darling Basin Plan developed under the Australian Government Water Act 2007. The Basin Plan defines environmental objectives which includes increasing water availability for the environment. To this end the Basin Plan sets Sustainable Diversion Limits, which will be applicable from 1 July 2019 for both ground and surface water (NCCMA, 2014a).

2.1 The Lower Campaspe

The Lower Campaspe catchment covers the northern portion of the Campaspe catchment in North-Central Victoria, an area that is approximately 150 km long and 25 km wide (NCCMA, 2014b), and is itself a part of the Murray-Darling Basin. The Campaspe River starts from the Great Dividing Range in the south, flowing in a northerly direction into Lake Eppalock from which the Lower Campaspe River begins. The Lower Campaspe River then continues northwards (downstream) into the Murray River which flows in a westerly direction. Population centres along the Lower Campaspe River include (from south to north) Axedale, Barnadown, Elmore, Rochester, and Echuca. The Lower Campaspe River itself is highly regulated by the operation of a dam at Lake Eppalock, the Campaspe Weir and Siphon located north of

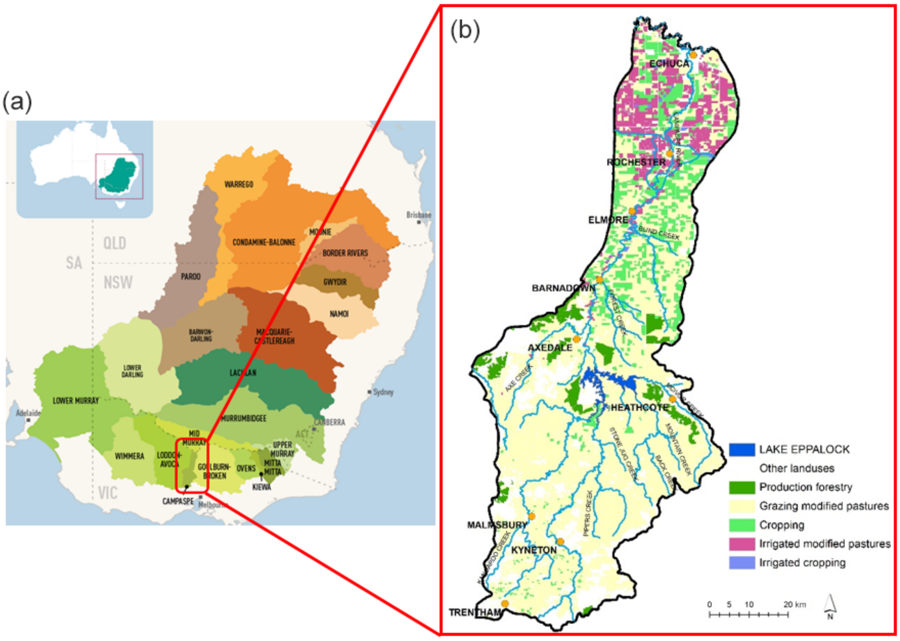


Figure 1. Panel **(a)** depicts the Campaspe catchment in the Murray-Darling Basin (adapted from MDBA, 2017) while **(b)** shows the Campaspe catchment proper with land use in the region.

Rochester (NCCMA, 2014b). A map indicating the catchment location is shown in Fig. 1.

Climate

The lower Campaspe is reported to be a dry semi-arid area which is evident in the historic rainfall records, with the median yearly rainfall being 434 mm (see Fig. 2). Two notable dry periods are identifiable in the historic rainfall records for the past 30 years which have influenced irrigators and water management. The first is a severe drought that occurred during 1982/1983 during which almost no rainfall occurred during the growing season resulting in severe (wheat) crop loss across eastern Australia (ABS, 1988; Arad and Evans, 1987; BoM, 2009). The second was the millennium drought, described as the worst drought on record for southeast Australia, defined as starting in 2001 with the drought eventually broken in 2009 (Van Dijk et al., 2013).

Climate projections for Northern Victoria, of which the Campaspe catchment is a part of, describe drier conditions with rainfall expected to decrease compared to the historic 20-year average. Decreases in mean rainfall of 12–13 % across south and east Australia compared to the 100-year average (1900–2000) have already been experienced within the first decade of the new millennium (2001 to 2009, Van Dijk et al., 2013).

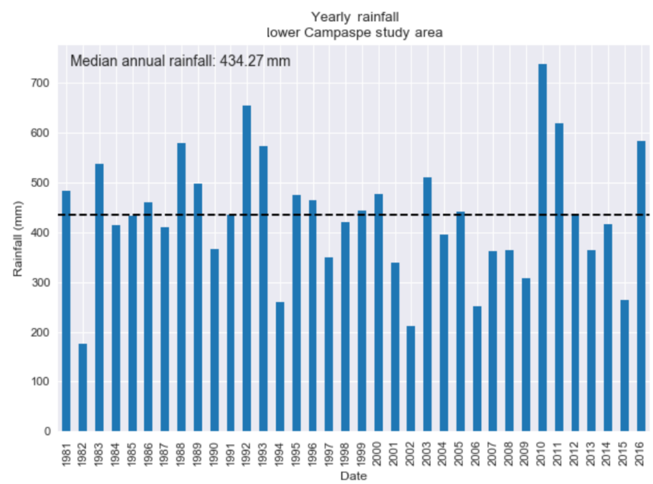


Figure 2. Yearly rainfall in the Lower Campaspe study area. The median amount was found to be 434 mm yr^{-1} (indicated by the dashed black line).

2.2 Hydrology

The area of the Campaspe basin is 4179 km^2, with a river length of 220 km, and a mean annual streamflow volume of 352 GL. The elevation in the southern part of the basin is around 600 m AHD (Australian Height Datum), with mean annual rainfall up to 1000 mm, and estimated mean annual pan evaporation of approximately 1300 mm. Near the catchment outlet (elevation 98 m AHD), the mean annual rainfall is approximately 430 mm, while the estimated mean annual pan evaporation is approximately 1700 mm. The main storage in the basin is Lake Eppalock, which has a catchment area of 2124 km^2, a storage capacity of 304 GL, and is located about 135 km from the catchment outlet at an elevation of about 160 m AHD. A further 3 large storages are located on the Coliban River (upstream of Lake Eppalock), with a total storage of 70 GL (BoM, 2017; GM-W, 2017a, b; MDBA, 2017).

2.3 Hydrogeology

The Campaspe region comprises the recent Coonambidgal Formation incised by the Campaspe River through the Shepparton formation, Parilla/Loxton sands, Newer Volcanic Basalts, with the primary productive aquifers of the region in the Calivil Formation and Renmark Group (collectively known as the Deep Lead)which overlay the Palaeozoic bedrock. The majority of the lower Campaspe consists of the Shepparton formation and the Deep Lead. The Deep Lead aquifers are the primary source of groundwater in the lower Campaspe irrigation areas, in which the Shepparton Formation has low permeability and is not very transmissive. Further details of the local hydrogeology may be found in Chiew et al. (1995).

2.4 Stakeholders

The local water corporation, Goulburn-Murray Water (GM Water), manages both surface and groundwater resources in the Lower Campaspe. Management includes the operation of the dam, water delivery infrastructure maintenance and investment, and the water accounts and licences in the region. GM Water is additionally responsible for determining the amount of water allocations – a percentage of water that an irrigator is entitled to – during each irrigation season.

The Department of Economic Development, Jobs, Transport and Resources (EcoDev) is a state level Government department that is interested in the water resource management and policy aspects, as well as providing advice and assistance to farmers regarding on-farm activities. The North Central Catchment Authority (NCCMA) and an expert from the Australian Platypus Conservancy were engaged for their input and feedback on the ecological system. Farmers themselves are an important stakeholder group to include as they will be impacted by any policy and climatic changes as well as being an important influencer of ecological and recreational water availability. Recreational users of the reservoir at Lake Eppalock were involved due to concerns that over-allocation of water for agricultural purposes have, and will exacerbate, negative impacts on recreational activities.

To gain further insights into the socio-agricultural system Ticehurst and Curtis (2016, 2017) conducted a survey of irrigators during 2016. The survey gathered responses from 254 participants (of 754 surveys sent out) that were later determined to be representative of irrigators in the region. The findings relevant to the model development process are repeated here, however readers are directed to Ticehurst and Curtis (2016, 2017) for further detail on the survey process.

These stakeholder groups were all engaged with through a series of workshops and discussions from late 2015 onwards. The latest workshop was run in October 2017, and another scheduled for March 2018. These stakeholder engagement activities aided in the selection of scenarios which describe plausible, and socially acceptable, options for changes in water management. Examples of such scenarios include the introduction of a managed aquifer recharge system to capture dam overflows for storage in the aquifer for use in times of water scarcity. Recharging the groundwater resource in this manner increases the availability of groundwater as well as reducing the impact of existing pumping levels.

3 Modelling process

The model development process followed a participatory process in which multi-disciplinary practitioners engaged with stakeholders. Through this process the model and its purpose was collaboratively defined and developed. Stakeholders play an additional important role in the development of scenarios of interest and validating model scope and behaviour (Krueger et al., 2012). Participatory engagement elicited a key set of management objectives including holistic management of water resources to improve crop yields, reliability of water availability, and beneficial improvements to environmental and socio-economic outcomes. Inclusion of stakeholders in the design and development process additionally fosters trust between stakeholders and modellers, and as a consequence model results (Franzén et al., 2011).

Another perspective is that of a software developer, as model implementations will largely be expressed in computer code. It is perhaps of interest to note that both software and model development best practices suggest an iterative process and arrived at these processes seemingly independently of each other. Sletholt et al. (2012) for example details and identifies software development practices that can be found in the model development process that have direct counterparts to software development practices.

A key point of interest is that iterative development is regarded as best practice in both model and software development paradigms. From a model developers' perspective continuous engagement with stakeholders has a hand in early detection and correction of faulty assumptions (Jakeman et al., 2006). Continuous exposure to the development process and incorporating feedback can drive stakeholder acceptance of the model by ensuring that the modelling process is transparent and relevant (Chan et al., 2010; Voinov and Gaddis, 2008). Jakeman et al. (2006) suggest an iterative approach to model development, where progress is reviewed at various steps, and part of the process repeated if issues are found. For software developers, iterative processes enable continu-

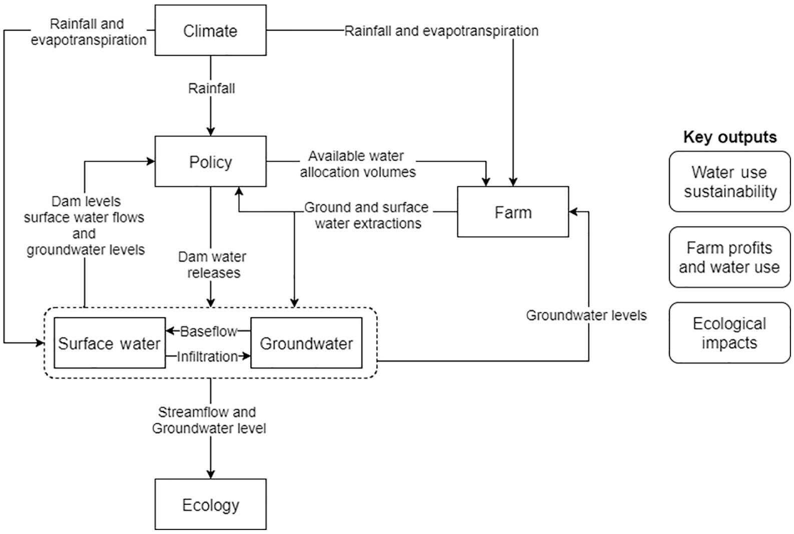


Figure 3. The interactions between component models and key model outputs. The dashed box around surface and groundwater models was inserted to simplify the model diagram and is not intended to indicate a separate coupled model.

ous validation of model implementation and the adjustments necessary to incorporate stakeholder feedback (Sletholt et al., 2012). To this end a component-based approach was applied for the implementation of the model, and so the development process could be described as applying a component-based participatory approach.

Component-based approaches compose a collection of compartmentalized models coupled loosely through a common framework. Loose coupling is achieved using specifically defined interfaces which handle data exchange and refers to the fact that the connections, and thus the feedbacks, between models are no longer "hardwired" to specific models. Malard et al. (2017) refers to the use of interfaces as a "wrapper approach" wherein the individual component models are "wrapped" and interactions channelled through the interfaces. Benefits of such an approach include the ability to reuse or "swap" a given component model for another (either new or pre-existing) as the need arises (de Kok et al., 2015). Changes within a model that do not affect the interface (i.e. the inputs and outputs) are safely abstracted and as such do not propagate and affect other component models. Incorporation of stakeholder feedback then becomes less problematic due to this model compartmentalization allowing model developers to focus on the modelling process instead of issues that may arise from direct coupling. Expected behaviour can then be verified through testing and comparisons against previous model outputs. As a consequence modellers are then able to progress through the iterative loops at a faster pace.

4 Model framing

To support these water reforms Federal and State Governments invested heavily in a modernization program in 2007 (State of Victoria, 2011), what is now known as the Connections Project and managed by GM Water. This infrastructure investment was described as the largest investment in irrigation infrastructure by the Australian Government (a total of AUD 1.1 Billion as reported in Bowler, 2015). A primary aim of the Connections Project was to improve the efficiency of water delivery and on-farm water use to meet sustainable water use goals as defined in the national Murray-Darling Basin Plan introduced in 2012 (Bowler, 2015).

Conjunctive use of water resources were identified by Ticehurst and Curtis (2017) as one method of improving water availability in the catchment. Here conjunctive water use was broadly defined as the multi-use of water sourced from both surface and groundwater for agricultural, recreational, and environmental purposes.

5 Model components

Each component represents a system of interest which collectively describes a system of systems. The developed integrated model represents a sociohydrological-environmental system including a farm model, surface water representing the lower Campaspe River and tributaries, groundwater hydrology, and a water management policy model. A climate component is also included which serves to provide the necessary rainfall and evapotranspiration data at the requisite spatial and temporal scales. The component models are coupled through a common framework developed in the Python programming language. Component models are not required to be developed in the same language as the framework as Python has robust language interoperability capabilities. The interactions between component models through their interfaces are depicted in Fig. 3.

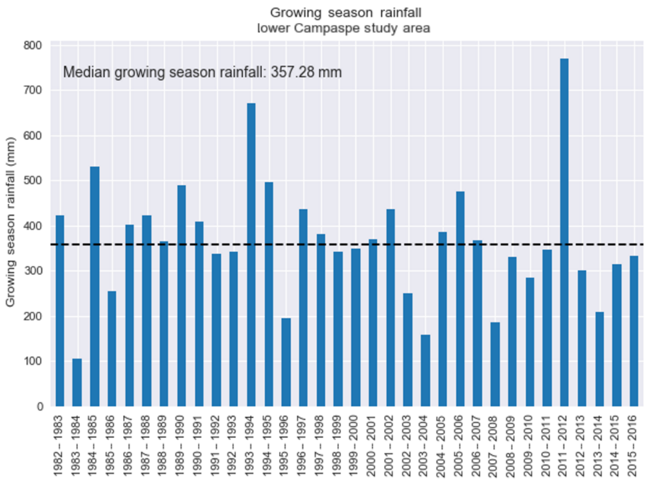


Figure 4. Growing season rainfall over the 1982–2015 growing seasons. Median growing season (dashed black line) was found to be 357.28 mm, below the usual growing season rainfall of 420 mm (see Fig. 5 below).

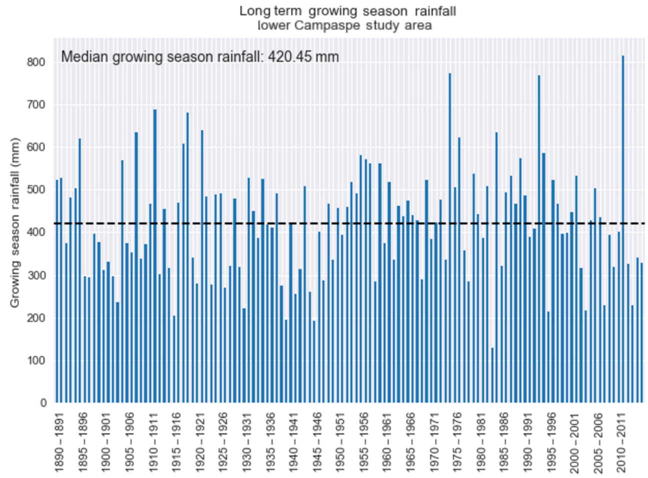


Figure 5. Long term growing season rainfall. Median in-season rainfall was found to be 420.45 mm (dashed black line).

In this section examples of the implications and influences from stakeholder feedback on each of the component models are given. The model continues to be developed in light of findings described herein, and as such is not made publicly available, although public release is intended. Additionally the data comes from various sources and so possible issues regarding intellectual property and data ownership will have to be cleared before public availability is possible. Model development utilizes version control which allows for the release of the model (in its current and future state) and requisite data at a later date.

5.1 Climate

Ticehurst and Curtis (2016) found that over 80 % of farmers surveyed believed that the impact of drought and changing rainfall patterns were important or very important. This finding in conjunction with the observed decrease in rainfall (see Sect. 2.1.1) shows that it is necessary to consider the impact of further climate variability. To this end (30 year) historic and future climate data were sourced via Climate Change in Australia (https://www.climatechangeinaustralia.gov.au, CSIRO, 2017). These datasets are described as being application ready. Long term ($\sim$ 100 years) climate records were developed through the use of interpolated historic rainfall and pan evapotranspiration data (see Vaze et al., 2011). The recent decrease in rainfall is evident within a typical growing season (defined as May to February). The median in-season rainfall during 1982 to 2016 was found to be 357 mm (see Fig. 4), compared to the reported usual growing season rainfall of 400 to 500 mm (EcoDev, 2015) and as indicated in the long term growing season rainfall records (see Fig. 5).

5.2 Surface water

The surface water module estimates the flows and water levels at selected nodes in the Campaspe Catchment. The nodes have been selected based on the location of gauges with suitable data, taking into consideration the needs of the integrated model (Fig. 6). As the focus of the integrated model is the lower Campaspe Catchment (below Lake Eppalock), the majority of the nodes are located in that region. To model the surface water flows, this means having information on releases and spills from Lake Eppalock, thereby requiring an estimate of the inflows to the reservoir. The resulting nodes are shown in Fig. 7. The surface water flows also depend on interaction with the groundwater, requiring a comparison of surface water levels with groundwater levels. This means that the surface water module needs to estimate the surface water levels at the nodes, and that this information is passed to the groundwater model in order to estimate the infiltration loss/baseflow contribution to surface water flow.

The surface water module has three components: a rainfall-streamflow model, a routing module and a rating curve module. Inputs required by the model are climate data (rainfall and potential evaporation), as well as estimates of the groundwater/surface water interactions (from the groundwater module), releases from Lake Eppalock reservoir (from the policy and the farm modules) and extractions from the surface water flows (from the farm model).

The rainfall-streamflow model used here is a variant of the IHACRES model, incorporating a non-linear loss module which converts rainfall into effective rainfall (rainfall that contributes to streamflow), and a unit hydrograph module that represents the dynamics of the water moving through the catchment (river network and landscape). The non-linear loss module used is based on the CMD version of the non-linear module (Croke and Jakeman, 2004), modified to produce two inputs to the unit hydrograph module (Croke et al., 2015): a

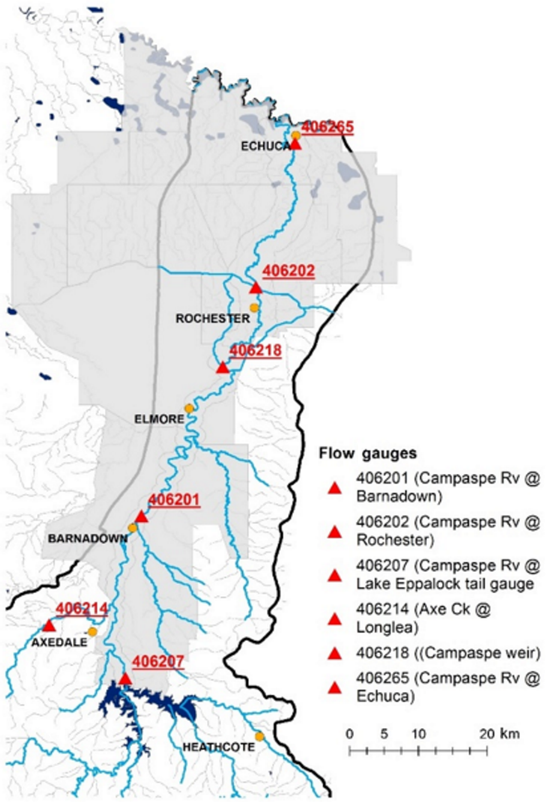


Figure 6. Flow gauges in the lower Campaspe.

contribution to the quick flow component (u_k) and a contribution to the slow flow component (r_k). This permits the model to partition effective rainfall between the two components based on the modelled catchment moisture status. The unit hydrograph module comprises two exponentially decaying stores arranged in parallel (a quick and a slow flow component), modified from the original to take the inputs to each store directly from the CMD module outputs.

An exponentially decaying store is also used to route the flows between nodes (a lag-route approach, Croke et al., 2006). In both the routing and the rainfall-streamflow models, the impact of losses from the river network are taken into consideration using the approach of Ivkovic et al. (2014). The rating module makes use of the rating curve data available at most gauge sites (the exception is gauge 406218, where only water level data is available).

5.3 Groundwater

The groundwater flow module is used to estimate the surface water-groundwater exchanges along the Campaspe River between flow gauges, and to provide information on the groundwater levels at specific locations as well as groundwater levels averaged over larger areas. The groundwater flow module interacts with the hydrology, farm, ecology and policy modules as detailed in Fig. 7.

The first stakeholder engagement workshop and subsequent communication with stakeholders (in particular, a local hydrogeologist) led to the definition of the groundwater model boundary, delineation of the hydrogeological units and provision of input data for groundwater pumping as well as observational head and chemistry data. Defining the model boundaries through this engagement ensured that the area covered as well as representation of hydrogeological units was consistent with their interpretation of the system and met the requirements for areas of interest. Furthermore the boundary conditions used came out of this initial consultation, in particular the consideration of the Campaspe River and smaller inflowing tributaries, the latter of which are not represented in the groundwater model due to their ephemeral nature and low flows.

The groundwater flow model of the Lower Campaspe Valley region is a finite difference representation. The model was constructed with Python scripts utilising Flopy (Bakker et al., 2016), and uses MODFLOW NWT (Niswonger et al., 2011). Representation of the hydrogeologic units (HGUs) is based on rasters (100 m resolution) from the Victorian Aquifer Framework (DSE, 2012). The model is made up of 7 layers, with a horizontal resolution of 1 km, and vertical spacing of the model grid informed by the HGU rasters. The 5 km resolution was chosen for computational speed to avoid the groundwater model becoming a computational bottleneck for the integrated model. Some HGUs span multiple layers where they are not overlain by other HGUs. There are 41 209 active cells within the model.

The groundwater model is driven by rainfall, river stage, groundwater extraction via pumping wells and groundwater head data via a series of boundary conditions shown in Fig. 7. Recharge in the model is implemented in the top layer of the model with the RCH package, and is calculated as a reduction of rainfall using a rainfall reduction parameter, and as such evapotranspiration is not directly modelled. River boundary conditions are implemented using the RIV package for the Campaspe River and Murray River. To allow outflow below the Murray River, through the subsurface in the north of the catchment, a general head boundary condition is implemented with the GHB package.

The model was calibrated using PEST (Doherty, 2016) to groundwater head data by modifying the HGU properties (i.e. hydraulic conductivity, specific yield and specific storage) and also a rainfall reduction parameter, applied statically from the period 1966–2015 and based on monthly stress periods. Initial conditions for the model were established by running the model in steady-state using long-term average rainfall and river stages.

As depicted in Fig. 7, the groundwater model is forced by:

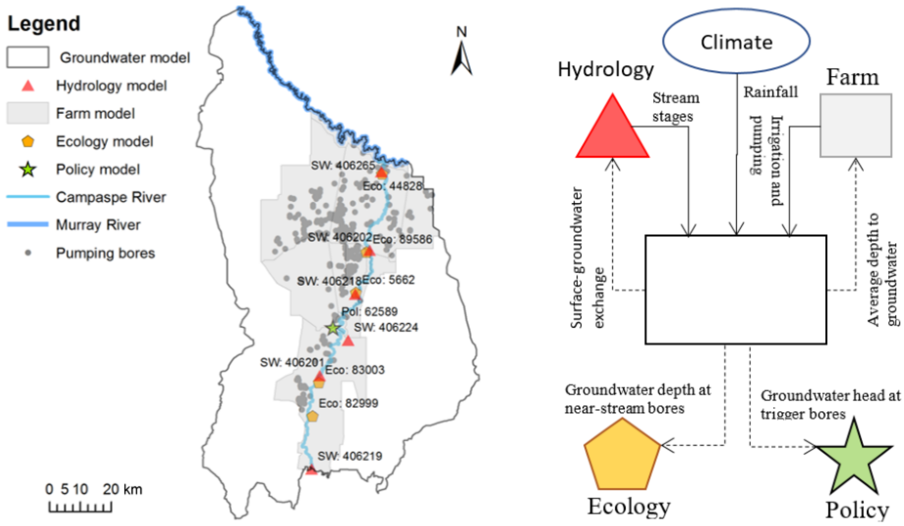


Figure 7. Groundwater model components and model area as well as points and interactions with other component models from the perspective of the groundwater model.

- Distributed rainfall (to be reduced through the rainfall reduction parameter) and irrigation water from the Farm model;
- Pumping volume from the Farm model (uniformly applied across pumps in the area);
- River stages from the Hydrology model
- After running each daily time-step, the model returns the:
- Surface water-groundwater exchange along reaches of the river consistent with the Hydrology model
- the average depth to groundwater for the Farm model
- depth to groundwater at key sites dictated by the Ecology model
- groundwater head at trigger level bores as dictated by the Policy model

Increases and decreases to pumping driven by the Farm model were applied to relevant wells within each farm zone. Surface water-groundwater responses lag behind the surface water forcing from the Hydrology model due to the use of a sequential coupling; it was assumed that a daily lag would not create significant differences in model behaviour. Outputs from the groundwater model, while not precise at the scale of local wells due to model resolution, were fit for purpose for indicative average groundwater levels at points of interest. In the case of the Ecology model, this is subject to the most variability as the levels are near-stream where the depth to groundwater table can change rapidly as it converges to the river. For the Policy model the trigger bores are chosen to be indicative of larger scale behaviour and hence the use of the average head in cells that correspond with the trigger bores is deemed adequate.

5.4 Policy

The current policy setting in the Campaspe is quite sophisticated reflecting extensive water reforms which introduced water trading, carryover, and environmental water provisions (Alston and Whittenbury, 2011; McKay, 2005; Wheeler and Cheeseman, 2013). The policy component of the integrated model provides a representation of policies determining the water allocation and carryover for entitlement holders (farm and environment). Use of these policies as a scenario supports further investigation of the implications and viability, as well as the opportunities, of the given policy condition(s) in the context of climate variability. The design of the policy component was such that it would allow scenarios that fit with current policies (e.g. increased groundwater use) but also the capability to explore alternate policy futures (e.g. conjunctive management of surface and groundwater, Managed Aquifer Recharge, and inter-catchment transfers). The latter include some of the conjunctive use opportunities explicitly identified with Campaspe stakeholders. For example, groundwater and surface water are managed separately in the current policy space. One option identified with Campaspe stakeholders was the temporary relaxation of groundwater restriction trigger levels during dry times when surface water allocations are low, with compensatory actions to increase recharge when climate conditions improve.

Within the current policy setting, groundwater use can be increased as most irrigators surveyed in the region are primarily reliant on surface water resources. Reported figures include 91 % of irrigators holding surface water licences compared to 22 % that additionally hold groundwater licences. Groundwater use historically reach a maximum of 60 % of allocated volumes, although this has increased to 80 % in recent times (2016 water usage, reported in Ticehurst and Curtis, 2017).

5.5 Farming

The farm component was developed with the aim of enabling investigation into the effect of water policy and water availability on farm financial performance under variable climate scenarios. Key attributes that are represented include the crop – currently one of wheat, barley or canola and tomato – irrigation system, pumping systems and soil types. The farming system is represented in a lumped manner with the study area divided into 12 zones configured to represent a mix of applicable surface and groundwater policy, water entitlements and usual cropping practices.

The principal agricultural enterprise in the Lower Campaspe is dairy farming with 55 % of land use devoted to annual and perennial pastures, 70 % (i.e. 38.5 % of reported farming area) of which is irrigated. Cereal cropping amounted to 35.8 % of land use, although the majority of this (68 %, i.e. ∼ 24 % of reported farming area) is dryland. Dairy farming is to be represented in the model through the use of an indicator crop to represent annual and perennial pasture crops and discussions with local experts are ongoing to determine how best to implement this.

Historically, the Campaspe region was an irrigation intensive area however most irrigators (90 %, concentrated in the middle of the study area) stopped irrigation practices in 2010 (NVIRP, 2010). This exit occurred during the millennium drought period (2001–2009) during which irrigators' water allocations were significantly reduced (NCCMA, 2014a). Irrigation is currently concentrated in the lower portion of the catchment the northern area surrounding Echuca, with dryland cropping in the mid and upper areas. A return to irrigation practices in the future remains a possibility due to the network of accessible irrigation infrastructure modernized under the Connections Project.

In the North Central region flood irrigation is the most common irrigation system in use accounting for 99 % of irrigated area (Ash, 2006). Flood irrigation is said to be 50–80 % water use efficient, meaning that 50–20 % of water applied to the field is lost (Clemmens, 2000; Finger and Morris, 2005; Tennakoon et al., 2013). Of those surveyed 77 % of respondents reported having undertaken additional improvements to flood irrigation such as laser grading and tail-water reuse, increasing the water use efficiency. Flood irrigation was then modelled as being 70 % water use efficiency based on this information. Other improvements can be achieved through the adoption of a piped system or investing in spray irrigation which is said to be 80 % water use efficient (Clemmens, 2000; Finger and Morris, 2005).

Outside of the survey, further information was gained through stakeholder engagement. It was highlighted, for example, that the choice to invest in a more water efficient irrigation system depends on the soil type. As such, generating a suitable representation of the soil textures in the modelled farming zones becomes a necessity and acts as a constraint to the choice of irrigation system adopted at each zone. It was also initially assumed that the vast majority of pumping systems in the area were diesel based rather than electric due to the substantial capital costs involved in developing the necessary infrastructure to operate electric pumps. This assumption was also corrected with local knowledge – electric pumping is in reality quite common and is used over the weekends due to off-peak electricity prices. The model will be modified to incorporate this elicited information.

5.6 Ecology

Management decisions that affect the lower Campaspe ecosystems flow on to the Murray River as the Campaspe flows into the Murray. The local ecology has historically been neglected due to over allocation of water resources for agricultural purposes. Decline in riverine health have been reported over the years including substantial decreases in biodiversity (MDBA, 2012; NCCMA, 2014b). The Murray-Darling Basin Plan includes provisions for increased environmental flows to support ecosystem maintenance and recovery (Bowler, 2015; GM-W, 2013; Hughes et al., 2015), and to meet these obligations up to 75 GL of water savings through infrastructure improvements through the GM Water Connections Project (formerly NVIRP) were intended for environmental purposes (NVIRP, 2011).

The local ecology along the Lower Campaspe River include communities of River Red Gum, a eucalyptus tree that is considered iconic (NCCMA, 2014b), platypus colonies, and two native fish populations: the Murray Cod and Golden Perch. Conceptualization and design of the model incorporated feedback from ecology experts from the NCCMA and the Australian Platypus Conservancy. Stakeholder feedback in combination with prior ecological studies and data availability resulted in the development of methods to generate indices that indicate the suitability of water flow for these flora and fauna. These consist of three indices for the River Red Gum which represent the suitability of groundwater availability and surface water flows for the maintenance and regeneration of the iconic tree. The fish indices capture key flow requirements as recommended in the Campaspe River Environmental Water Management Plan (NCCMA, 2014a). Indices developed for platypi indicate flow conditions that sustain food supply and movement, breeding cycles, and the avoidance of burrow flooding during the mating season.

6 Conclusion

Both model and software development best practices recommend working within an iterative cycle that moves the project towards a continually (re)defined goal, informed by stakeholders. Including stakeholders in the iterative development of integrated models was found to be useful in ensuring model validity, relevance, transparency and acceptability. Participatory engagement acts as a peer review process within each iteration of the model development cycle whilst also fostering trust between all participants, modellers and stakeholders alike.

Component-based development processes were found to be complementary to the participatory modelling approach. Throughout each iteration the implementation of the described component models were influenced by stakeholder knowledge and information. The compartmentalization of models that collectively represent a system of systems disentangles their implementation allowing specific and targeted modifications based on stakeholder feedback. Such changes then do not propagate throughout the model as a whole, allowing modellers to progress through each iteration quicker whilst simultaneously ensuring that the model development process is transparent.

Data availability. Climate time series data presented here is available from the Australian Bureau of Meteorology http://www.bom.gov.au/.

Competing interests. The authors declare that they have no conflict of interest.

Special issue statement. This article is part of the special issue "Innovative water resources management – understanding and balancing interactions between humankind and nature". It is a result of the the 8th International Water Resources Management Conference of ICWRS, Beijing, China, 13–15 June 2018.

Acknowledgements. The primary author is supported through an Australian Government Research Training Program (RTP) Scholarship. The work was carried out as part of the Murray-Darling Basin Authority's partnership with the National Centre for Groundwater Research and Training (NCGRT). The authors would like to thank representatives from Goulburn-Murray Water and the Victorian Department of Economic Development, Jobs, Transport and Resources for their valuable insights and assistance over the course of this work. The authors would like to additionally thank the reviewers for their valuable comments and suggestions to improve the quality of this paper.

Edited by: Dingzhi Peng
Reviewed by: Qinglan Li and one anonymous referee

References

ABS: Drought in Australia, available at: http://www.abs.gov.au/ausstats/abs@.nsf/0/068F13BCCD03F27BCA2569DE001F1072?OpenDocument (last access: 26 February 2018), 1988.

Alston, M. and Whittenbury, K.: Climate change and water policy in Australia's irrigation areas: a lost opportunity for a partnership model of governance, Environ. Polit., 20, 899–917, https://doi.org/10.1080/09644016.2011.617175, 2011.

Arad, A. and Evans, R.: The hydrogeology, hydrochemistry and environmental isotopes of the Campaspe River aquifer system, North-Central Victoria, Australia, J. Hydrol., 95, 63–86, https://doi.org/10.1016/0022-1694(87)90116-8, 1987.

Ash, L.: Irrigation Farm Survey 2004/2005 Summary Extract from the final document 17/10/06, NSW Department of Primary Industries, Sydney, Australia, 2006.

Bakker, M., Post, V., Langevin, C. D., Hughes, J. D., White, J. T., Starn, J. J., and Fienen, M. N.: Scripting MODFLOW Model Development Using Python and FloPy, Groundwater, 54, 733–739, 2016.

BoM: Western Australia to the Rescue, avaiable at: http://pandora.nla.gov.au/pan/96122/20090317-1643/www.bom.gov.au/lam/climate/levelthree/c20thc/flip4.html (last access: 11 May 2016), 2009.

BoM: Climate and past weather information, available at: http://www.bom.gov.au/climate/ (last access: 18 December 2017), 2017.

Bowler, M.: Report for the Department of Agriculture and Water Resources – Goulburn-Murray Water Connections Project Stage 2 – Mid Term Review, GHD, Melbourne, Victoria, Australia, 2015.

Chan, T., Ross, H., Hoverman, S., and Powell, B.: Participatory development of a Bayesian network model for catchment-based water resource management, Water Resour. Res., 46, W07544, https://doi.org/10.1029/2009WR008848, 2010.

Chiew, F. H. S., McMahon, T. A., Dudding, M., and Brinkley, A. J.: Technical and economic evaluation of the conjunctive use of surface and groundwater in the Campaspe Valley, north-central Victoria, Australia, Water Resour. Manage., 9, 251–275, https://doi.org/10.1007/BF00872487, 1995.

Clemmens, A. J.: Measuring and improving irrigation system performance at the field level, in: Connellan, G. J., Irrigation Association of Australia 2000 National Conference and Exhibition: Water – Essential for Life. Proceedings. Melbourne Exhibition and Convention Centre, 23–25 May 2000, Irrigation Association of Australia, Melbourne, 190–199, 2000.

Croke, B. F. W. and Jakeman, A. J.: A Catchment Moisture Deficit module for the IHACRES rainfall-runoff model, Environ. Modell. Softw., 19, 1–5, 2004.

Croke, B. F. W., Letcher, R. A., and Jakeman, A. J.: Development of a flow network for the Namoi River Basin, Australia, J. Hydrol., 319, 51–71, https://doi.org/10.1016/j.jhydrol.2005.07.001, 2006.

Croke, B., Cornish, P., and Islam, A.: Modeling the Impact of Watershed Development on Water Resources in India, in: Integrated Assessment of Scale Impacts on Watershed Intervention: Assessing Hydrogeological and Bio-physical Influences on Livelihoods, edited by: Reddy, V. R. and Syme, G. J., Elsevier, 100–148, 2015.

CSIRO: Climate Change in Australia, available at: https://www.climatechangeinaustralia.gov.au/en/ (last access: 13 December 2017), 2017.

de Kok, J. L., Engelen, G., and Maes, J.: Reusability of model components for environmental simulation – Case studies for integrated coastal zone management, Environmental Modelling and Software, 68, 42–54, https://doi.org/10.1016/j.envsoft.2015.02.001, 2015.

Doherty, J. E.: PEST, SENSAN and Global Optimisers, in: PEST, Model-Independent Parameter Estimation User Manual Part I, Watermark Numerical Computing, Brisbane, Australia, available at: http://www.pesthomepage.org/getfiles.php?file=newpestman2.pdf&usg=AOvVaw0XsrKp9kwYUaX_DkwULET9 (last access: 26 February 2018), 2016.

DSE: Victorian Aquifer Framework: Updates for Seamless Mapping of Aquifer Surfaces, available at: http://www.vvg.org.au/cb_pages/files/207086.pdf (last access: 26 February 2018), 2012.

EcoDev: Victorian Winter Crop Summary 2015, Department of Economic Development, Jobs, Transport and Resources, Horsham, Victoria, Australia, available at: https://grdc.com.au/resources-and-publications/all-publications/publications/2015/03/nvt-victorian-winter-crop-summary-2015 (last access: 26 February 2018), 2015.

Finger, L. and Morris, M.: Water Management Options: Assisting irrigators with Stream Flow Management Plan implementation, Department of Primary Industries (DPI), State of Victoria, Tatura, Victoria, available at: http://vro.agriculture.vic.gov.au/dpi/vro/vrosite.nsf/0d08cd6930912d1e4a2567d2002579cb/adbda16966d0e530ca2574c8002ccbfa/$FILE/WaterManagementOptionsReport_2005.pdf (last access: 22 March 2018), 2005.

Franzén, F., Kinell, G., Walve, J., Elmgren, R., and Söderqvis, T.: Participatory social-ecological modeling in eutrophication management: The case of Himmerfjärden, Sweden, Ecol. Soc., 16, 27, https://doi.org/10.5751/ES-04394-160427, 2011.

GM-W: Lake Eppalock Land and On-Water Management Plan, Tatura, Victoria, 2013.

GM-W: Campaspe Basin – Goulburn Murray Water, available at: https://www.g-mwater.com.au/water-resources/catchments/gmwcatchments/campaspebasin (last access: 18 December 2017), 2017a.

GM-W: Lake Eppalock – Goulburn Murray Water, available at: https://www.g-mwater.com.au/water-resources/catchments/storages/campaspe/lakeeppalock (last access: 18 December 2017), 2017b.

Hughes, L., Steffen, W., Rice, M., and Pearce, A.: Feeding a hungry nation: Climate change, food and farming in Australia, Climate Council of Australia Limited, Sydney, Australia, available at: https://www.climatecouncil.org.au/uploads/7579c324216d1e76e8a50095aac45d66.pdf (last access: 26 February 2018), 2015.

Ivkovic, K. M., Croke, B. F. W., Kelly, R. A.: Overcoming the challenges of using a rainfall-runoff model to estimate the impacts of groundwater extraction on low flows in an ephemeral stream, Hydrol. Res., 45, 58–72, https://doi.org/10.2166/nh.2013.204, 2014.

Jakeman, A. J. and Letcher, R. A.: Integrated assessment and modelling: Features, principles and examples for catchment management, Environ. Modell. Softw., 18, 491–501, https://doi.org/10.1016/S1364-8152(03)00024-0, 2003.

Jakeman, A. J., Letcher, R. A., and Norton, J. P.: Ten iterative steps in development and evaluation of environmental models, Environ. Modell. Softw., 21, 602–614, https://doi.org/10.1016/j.envsoft.2006.01.004, 2006.

Kraft, P., Multsch, S., Vaché, K. B., Frede, H.-G., and Breuer, L.: Using Python as a coupling platform for integrated catchment models, Adv. Geosci., 27, 51–56, https://doi.org/10.5194/adgeo-27-51-2010, 2010.

Krueger, T., Page, T., Hubacek, K., Smith, L., and Hiscock, K.: The role of expert opinion in environmental modelling, Environ. Modell. Softw., 36, 4–18, https://doi.org/10.1016/j.envsoft.2012.01.011, 2012.

Malard, J. J., Inam, A., Hassanzadeh, E., Adamowski, J., Tuy, H. A., and Melgar-Quiñonez, H.: Development of a software tool for rapid, reproducible, and stakeholder-friendly dynamic coupling of system dynamics and physically-based models, Environ. Modell. Softw., 96, 410–420, https://doi.org/10.1016/J.ENVSOFT.2017.06.053, 2017.

McKay, J.: Water insititional reforms in Australia, Water Policy, 7, 35–52, https://doi.org/10.1080/02508069908692147, 2005.

MDBA: Sustainable Rivers Audit 2: The ecological health of rivers in the Murray–Darling Basin at the end of the Millennium Drought (2008–2010), Summary, available at: https://www.mdba.gov.au/sites/default/files/pubs/SRA2-SUMMARY-FINAL_0.pdf (last access: 26 February 2018), 2012.

MDBA: Campaspe/Murray-Darling Basin Authority, available at: https://www.mdba.gov.au/discover-basin/catchments/campaspe (last access: 18 December 2017), 2017.

NCCMA: The North Central CMA Region Environmental Water Management Plan for the Campaspe River System, available at: http://www.vewh.vic.gov.au/__data/assets/pdf_file/0004/368698/Campaspe-River-EWMP-FINAL.pdf (last access: 26 February 2018), 2014a.

NCCMA: Campaspe River resource Guide, available at: http://www.nccma.vic.gov.au/sites/default/files/publications/campaspe_river_resource_guide.pdf (last access: 26 February 2018), 2014b.

Niswonger, R. G., Panday, S., and Ibaraki, M.: MODFLOW-NWT, A Newton Formulation for MODFLOW-2005, in: Section A, Groundwater; Book 6 Modeling Techniques (p. 44), available at: https://pubs.usgs.gov/tm/tm6a37/pdf/tm6a37.pdf (last access: 26 February 2018), 2011.

NVIRP: Majority of Campaspe irrigators to exit permanently, available at: http://www.g-mwater.com.au/downloads/gmw/connections/Communication/Media/MediaReleases2010/Majority_of_Campaspe_irrigators_to_exit_permanently.pdf (last access: 26 February 2018), 2010.

NVIRP: NVIRP Annual Report 2010–2011, Northern Victoria Irrigation Renewal Project (NVIRP), Victoria, Australia, 2011.

Sletholt, M. T., Hannay, J. E., Pfahl, D., and Langtangen, H. P.: What do we know about scientific software development's agile practices?, 24–36, https://doi.org/10.4324/9780203894644, London, Routledge, 2012.

State of Victoria: NVIRP Stage 2 On Farm Project.pdf, State of Victoria, Northern Victoria Irrigation Renewal Project: On Farm Priority Project, available at: http://www.agriculture.gov.au/SiteCollectionDocuments/water/vic-farm-project.pdf (last access: 1 March 2018), 2011.

Tennakoon, S., Richards, D., Milroy, S., and Harris, G.: Water use efficiency in the Australian cotton industry, in: Waterpak: A Guide for Irrigation Management in Cotton and Grain Farming Systems, edited by: Wiggington, D., 22–27, Cotton Research and Development Corporation, 2013.

Ticehurst, J. and Curtis, A.: Improving on-farm water use efficiency amongst irrigators in the Campaspe River system and the Rochester Irrigation Area, Wagga Wagga, NSW, 2650, Australia, 2016.

Ticehurst, J. and Curtis, A.: Preliminary assessment of conjunctive use opportunities for in the Murray Darling Basin: a case study in the lower Campaspe catchment, Wagga Wagga, NSW, 2650, Charles Sturt University, 2017.

Van Dijk, A. I. J. M., Beck, H. E., Crosbie, R. S., De Jeu, R. A. M., Liu, Y. Y., Podger, G. M., Timbal, B., and Viney, N. R.: The Millennium Drought in southeast Australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society, Water Resour. Res., 49, 1040–1057, https://doi.org/10.1002/wrcr.20123, 2013.

Vaze, J., Teng, J., and Chiew, F. H. S.: Assessment of GCM simulations of annual and seasonal rainfall and daily rainfall distribution across south-east Australia, Hydrol. Process., 25, 1486–1497, https://doi.org/10.1002/hyp.7916, 2011.

Voinov, A. and Gaddis, E. J. B.: Lessons for successful participatory watershed modeling: A perspective from modeling practitioners, Ecol. Model., 216, 197–207, https://doi.org/10.1016/j.ecolmodel.2008.03.010, 2008.

Wheeler, S. A. and Cheeseman, J.: Key Findings from a Survey of Sellers to the Restoring the Balance Programme, Economic Papers, 32, 340–352, https://doi.org/10.1111/1759-3441.12038, 2013.

Chapter 4: A socio-environmental model to explore sustainable water management futures

In this chapter the application of a component-based integrated environmental model for the purpose of exploring possible water management futures is presented. The integrated model is applied to identify potential pathways to improved farm profitability, recreational, groundwater and ecological outcomes relative to modelled baselines. In particular, the influence of conjunctive use and management of water resources is investigated. Results indicate that improved farm level knowledge and management regarding crop water requirements, soil water capacity, and irrigation are the most significant factors towards achieving outcomes that are robust to a range of future conditions. Conjunctive use is shown to further improve the likelihood of achieving robust outcomes.

Exploratory scenario modelling (ESM) was leveraged for the model purpose, results from which additionally serve to communicate the level of uncertainty. The model can be considered a system-of-systems model as it is made up of multiple independent and interacting constituent models to form a unique whole with its own (emergent) behaviour. The model and the results are intended to raise awareness and facilitate discussion with and amongst stakeholders.

This chapter was published in the Journal of Hydrology: Regional Studies after reviews conducted by two anonymous reviewers. The publication is open-access under the Creative Commons CC-BY-4.0 licence. The authors acknowledge the use of funding from the Fenner School Publication Fund to make this possible.

Iwanaga, T.*, Partington, D., Ticehurst, J., Croke, B.F.W., Jakeman, A.J., 2020. A socio-environmental model for exploring sustainable water management futures: Participatory and collaborative modelling in the Lower Campaspe catchment. Journal of Hydrology: Regional Studies 28, 100669. https://doi.org/10.1016/j.ejrh.2020.100669

Journal of Hydrology: Regional Studies 28 (2020) 100669

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

A socio-environmental model for exploring sustainable water management futures: Participatory and collaborative modelling in the Lower Campaspe catchment

Takuya Iwanaga[a,*], Daniel Partington[b], Jenifer Ticehurst[a], Barry F.W. Croke[a,c], Anthony J. Jakeman[a]

[a] Fenner School of Environment and Society, Institute for Water Futures, The Australian National University, Canberra, Australia
[b] National Centre for Groundwater Research and Training, Flinders University, Adelaide, Australia
[c] Mathematical Sciences Institute, The Australian National University, Canberra, Australia

ARTICLE INFO

Keywords:
Integrated modelling
Participatory modelling
Exploratory modelling
Scenario discovery
Conjunctive water use

ABSTRACT

Study region: Lower Campaspe, North Central Victoria, Australia
Study focus: This paper presents a component-based integrated environmental model developed through participatory processes to explore sustainable water management options. Possible futures with improved farm profitability and ecological outcomes relative to modelled baselines were identified through exploratory modelling. The integrated model and the results produced are intended to raise awareness and facilitate discussion with and amongst stakeholders.
New hydrological insights: The modelling illustrates that improved farm level knowledge and management with regard to crop water requirements, soil water capacity, and irrigations are the most significant factors towards achieving outcomes that are robust to a range of climate and water policy futures. Assuming farmer management with regard to these factors are at their most optimal, increasing irrigation efficiency alone did not lead to improved farm profitability and ecological outcomes under drier climate conditions. Likelihood of achieving robust outcomes were further improved through the conjunctive use of surface and groundwater, with increased consideration of groundwater use a key factor. Further discussion on the viability and impact of increased groundwater use and conjunctive use policies should be further considered.

1. Introduction: Sustainable water management and aims of the integrated modelling

Management of water resources takes place within the context of a complex socio-environmental system. Sustainable management of water resources requires the needs of several agricultural, environmental, and social domains to be balanced with explicit consideration of a multitude of interacting factors. Here, “sustainable” refers to water usage that is both beneficial and robust - featuring improved farm profitability and environmental outcomes and maintaining these under changing, possibly adverse, climatic conditions within hypothetical policy contexts (revisited in Section 4).

Holistic modelling of this socio-environmental system then requires an integrated approach due to the number of system domains under consideration, the level of interactions that can occur at different (spatial and temporal) scales, and the uncertainty that comes with it (Letcher et al., 2007; Schlüter et al., 2019). A well-considered integrated approach can reduce the risk of unintentionally

* Corresponding author.
E-mail address: iwanaga.takuya@anu.edu.au (T. Iwanaga).

https://doi.org/10.1016/j.ejrh.2020.100669
Received 4 December 2019; Received in revised form 29 January 2020; Accepted 1 February 2020

disregarding crucial aspects of the management context and the flow-on effects without which the conclusions reached may be compromised (Kelly et al., 2013). To this end Integrated Environmental Models (IEMs) are often constructed to aid in informing management and policy decisions (Elsawah et al., 2020a; Janssen et al., 2010; Voinov and Shugart, 2013).

Stakeholder engagement is an important step in integrated modelling, particularly in developing socio-economic scenarios that are acceptable and relevant to stakeholders. Such participatory approaches within water resource modelling processes are now considered best practice in facilitating stakeholder buy-in, credibility of the modelling and integration of local knowledge and information into the model (Kelly et al., 2013; Megdal et al., 2017; Refsgaard et al., 2007). To holistically develop the model, an approach such as the one described in Badham et al. (2019) can be used to elicit stakeholder knowledge to aid in defining the problem frame and key issues.

The Murray-Darling Basin Plan (introduced in 2012) increases the amount of water allocated for environmental purposes by decreasing the volume for consumptive use (Bowler, 2015; North Central CMA, 2014). The intention of the Plan is to rectify the observed long-term degradation of environmental health of river systems within the Murray-Darling Basin generally. To this end, beneficial future scenarios would improve, or at least maintain, current levels of water availability for agricultural, environmental, and recreational purposes in the face of uncertain future climate conditions.

The primary aims of the study presented within this paper were to identify these future pathways ("scenarios") to improved environmental and socio-economic outcomes under a variety of climate conditions for the Lower Campaspe catchment in North Central Victoria (Australia). On-farm practices and water allocation policies were modelled and an exploratory modelling approach (Haasnoot et al., 2013) adopted to identify these possible robust outcomes. Such scenario discovery approaches have been utilised previously to identify viable adaptation strategies with indication of trade-offs between scenarios (Kwakkel et al., 2016).

The variety of data and knowledge sources, number of systems involved and the interactions and feedbacks between them required to represent such a system made the use of an integrated model a natural fit. An IEM was developed for the study, which we refer to as the CIM (Campaspe Integrated Model). The CIM includes representations of relevant systems across the socio-environmental spectrum and their interactions. These include policy, farm, surface and groundwater hydrology, ecology, and recreational values. Climate factors are represented by rainfall and evapotranspiration data, which drive the modelled system. The CIM is used to explore the mix of considered farm and policy level options that are robust in the long-term, in terms of successfully achieving desirable improvements from a baseline across climate scenarios. In this paper we detail the model components developed, discuss the integration process, and finally present the model results and their implications. The specifics of the management context and modelling process are reported in Iwanaga et al. (2018), however relevant elements will be repeated herein for context.

2. Lower Campaspe study area

The Lower Campaspe study area is a semi-arid region situated in the North-Central region of Victoria, Australia and is named for its primary river (the Campaspe), which flows northwards, joining the Murray River. The primary water source for the Lower Campaspe is the dam at Lake Eppalock, which divides the Campaspe into its Lower and Upper sub-catchments. The dam is operated by Goulburn-Murray Water (GM Water) subject to local and federally mandated policies such as the aforementioned Murray-Darling Basin Plan. Under current policies the environment is regarded as a water user with its own water entitlements (North Central CMA, 2014). Aside from managing dam operations and other responsibilities GM Water is the local irrigation authority, determining water allocations (for both agricultural and environmental users) and managing licencing for water use and access.

The Campaspe catchment is a mixed-farming area with a focus on dairy farming, with 55 % of its land use devoted to annual and perennial pastures. Cereal cropping amounts to 36 % of reported agricultural land area. Fig. 1 displays a map of the Lower Campaspe in context of the North Central region and the Murray-Darling Basin. Historically the Campaspe region was an irrigation intensive area, but a decade-long drought – the Millennium Drought – reduced water availability such that 90 % of irrigators elected to cease irrigation practices in 2010 (North Central CMA, 2014; NVIRP, 2010). Approximately 38 % of the Lower Campaspe is under dryland farming (Ticehurst and Curtis, 2016) with the north of the catchment focused on cropping activities (depicted in Fig. 1).

Water resources have been described as historically over-allocated for agricultural purposes, and the possibility of a drier climate in the future (van Dijk et al., 2013v) implies balancing available water resources between competing needs and interests is expected to become increasingly difficult. Ecological health of the Campaspe river system has been in decline over the past decades as water was historically prioritised for agricultural purposes. This has had the effect of substantially decreasing local biodiversity (MDBA, 2012; North Central CMA, 2014). Communities of the iconic River Red Gum eucalypts, platypus colonies, and populations of native fish (such as the Murray Cod and Golden Perch) exist along the Lower Campaspe system.

Recent water reforms have included provisions for increased environmental flows to support recovery and maintenance of ecosystem health (GHD, 2015; GM-W, 2013; Hughes et al., 2015). Water to support environmental flows include 75 G L reallocated from agriculture as well as estimated water savings due to infrastructure improvements conducted through the Goulburn Murray Connections project (NVIRP, 2011). Recreational use of the dam is an additional area of concern, with viability of recreational activities (e.g. boating and yachting) suffering as the water levels at Lake Eppalock fall. There has been public outcry in this regard, as evidenced by local media reports (ABC News, 2015; Wines, 2015).

Decisions made in managing the lower Campaspe River affect river systems downstream, the Campaspe being a tributary of the Murray. Therefore, beneficial ecological outcomes within the Campaspe are likely to support ecological recovery elsewhere downstream. The CIM was designed and developed to inform management and decision-making processes within this context through an exploratory process.

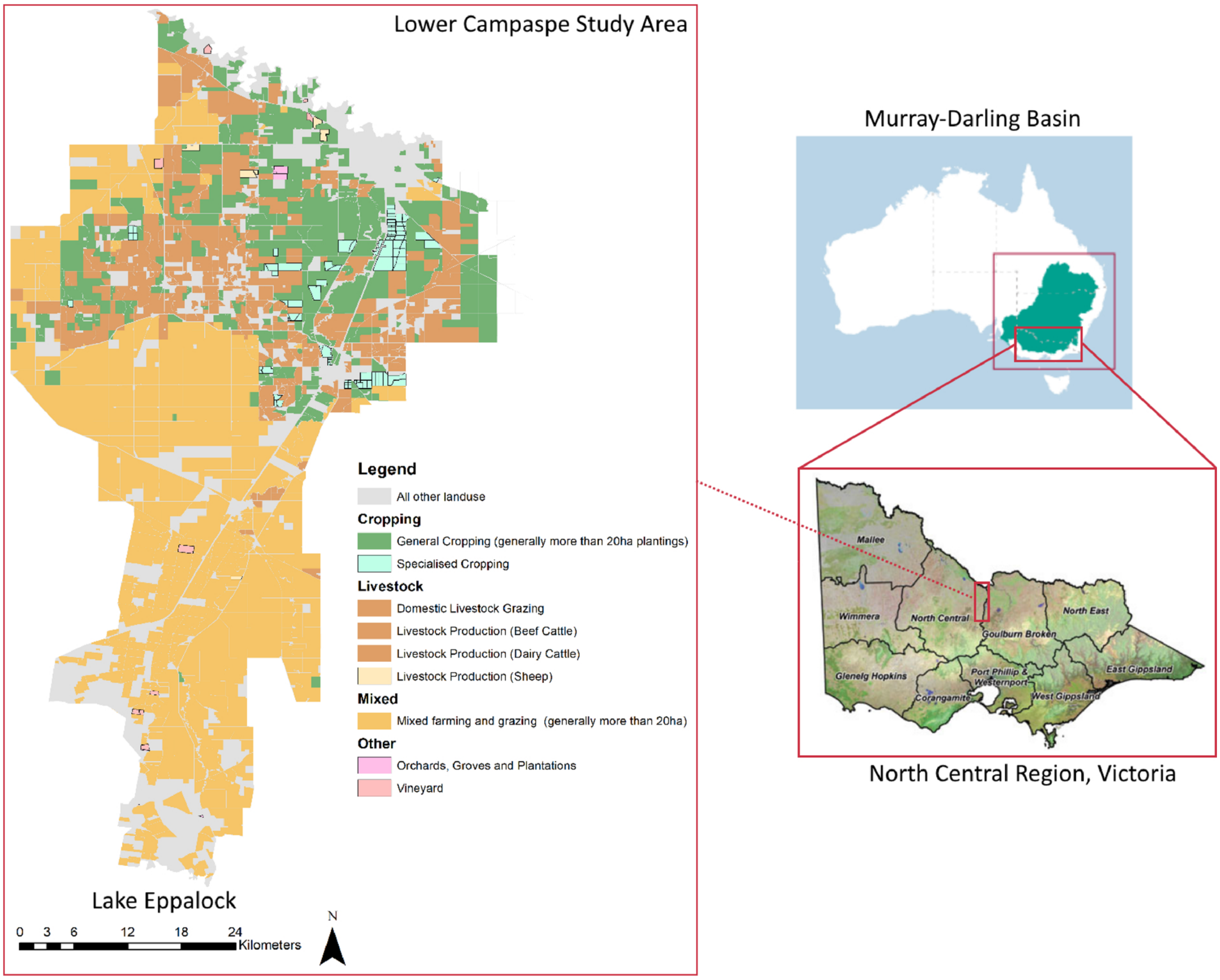


Fig. 1. Map of the Lower Campaspe catchment (left panel) in relation to the Murray-Darling Basin (top right) and the North Central Region of Victoria (bottom right). The Lower Campaspe constitutes the area north of Lake Eppalock, which is the primary reservoir for the study area. The Campaspe River flows south to north, into the Murray River.

3. Integrated model development

To explore possible futures in the Campaspe, changes to on-farm practices and water allocation policies were modelled and the subsequent effects on farmer income, streamflow conditions for platypus colonies, native fish and river red gums (trees), and recreational use of the dam were analysed. Specifically, these scenarios represent the conjunctive management of both surface and groundwater (Pulido-Velazquez et al., 2011), encouragement of further use of groundwater resources in general, and further improvements to irrigation (water application) efficiency (Ticehurst and Curtis, 2017, 2016). The CIM comprises models to represent climate sequences, policy rules, agricultural activities, surface and groundwater hydrology, and indicators of ecological and recreational suitability. Individual model domains deal with their own unique issues and conceptualise the system, and their interactions with other models, in separate ways. This is most obvious in the represented spatial and temporal scales. In building the CIM, compromises were necessary in order to suit the purpose of the modelling and in the face of inter-linked requirements and available resources. For this reason, further detailed framing of each model domain and, where relevant, model inputs and indicators of interest are described in the sub-sections below.

3.1. Stakeholders and engagement process

As noted in the introduction, engagement with local stakeholders was particularly important to both the development and validation of the CIM. Relevant stakeholders were identified through sectoral interests and relevance (e.g. farmers will be interested in policies that affect farming), as well as a snowball sampling approach where known experts where recruited to suggest other experts of interest relevant to the study. Stakeholders involved both prior to and during the modelling process included local farmers, GM Water, and representatives from government departments and non-profit organisations. These represent actors within the system that use the water, managers of the water, and those with specialised knowledge of the system, including farm management and irrigation specialists, ecologists, geo-hydrologists and catchment managers to name a few.

A stakeholder group of representatives from GM Water, the local Catchment Management Authority (North Central CMA) and a relevant state government department (at the time, the Department of Economic Development, Jobs, Transport and Resources) attended workshops to identify potential opportunities for conjunctive use in the region. Farmers were engaged through interviews and surveys prior to model development, details of which can be found in Ticehurst and Curtis (2017, 2016). This engagement identified the current and future intention to adopt various management options including the use of groundwater, various irrigation practices, and the technical feasibility and social acceptability of the range of conjunctive use opportunities identified in previous workshops. GM Water also provided irrigator data and aided in defining the spatial scope and boundaries of the study with respect to the represented groundwater catchment and management zones. The stakeholder group also took part in later workshops and served to provide information and knowledge which corrected earlier assumptions and provided further feasibility assessment of on-farm scenarios. Issues and concerns surrounding ecological aspects were elicited through engagement with ecologists from the Australian Platypus Conservancy and the North Central CMA. Further details may be found in Iwanaga et al. (2018).

3.2. Technical implementation

A requirement of the integrated model development was to be flexible in the face of changing and evolving understanding of the system due to the amount and spread of knowledge being engaged with through the participatory engagement process that was described in the previous section. To facilitate this an iterative component-based approach was adopted in the development of the CIM. Construction of the CIM involved the use of a mix of programming languages including Fortran and Python (and its compiled cousin Cython), incorporation of pre-existing models, and the development of a purpose-built (software) framework through which each component model was coupled.

Interfaces (commonly referred to as wrappers) were developed for the purpose of invoking component model runs and thus provide the necessary linkage between the framework and the component models. Fig. 2 depicts the inter-relationship between the component models. Inter-model communication (i.e. data exchange) occurs with the framework acting as an intermediary. Conversion of data, such as between types or units of measurement, occurs where necessary and is specifically coded. While not adhering to all aspects, the structure of the interfaces is similar to those specified by the Basic Modeling Interface (Peckham et al., 2013) in that each interface provides a method to invoke a run of the component model for a given time step. This design pattern was selected for its flexibility, simplicity and ease of implementation. Component models are run as a serial process with one model run after the other, with feedback occurring across (daily) time steps. Further details on the choice of modelled spatial and temporal scale is given in later sub-sections for each system component.

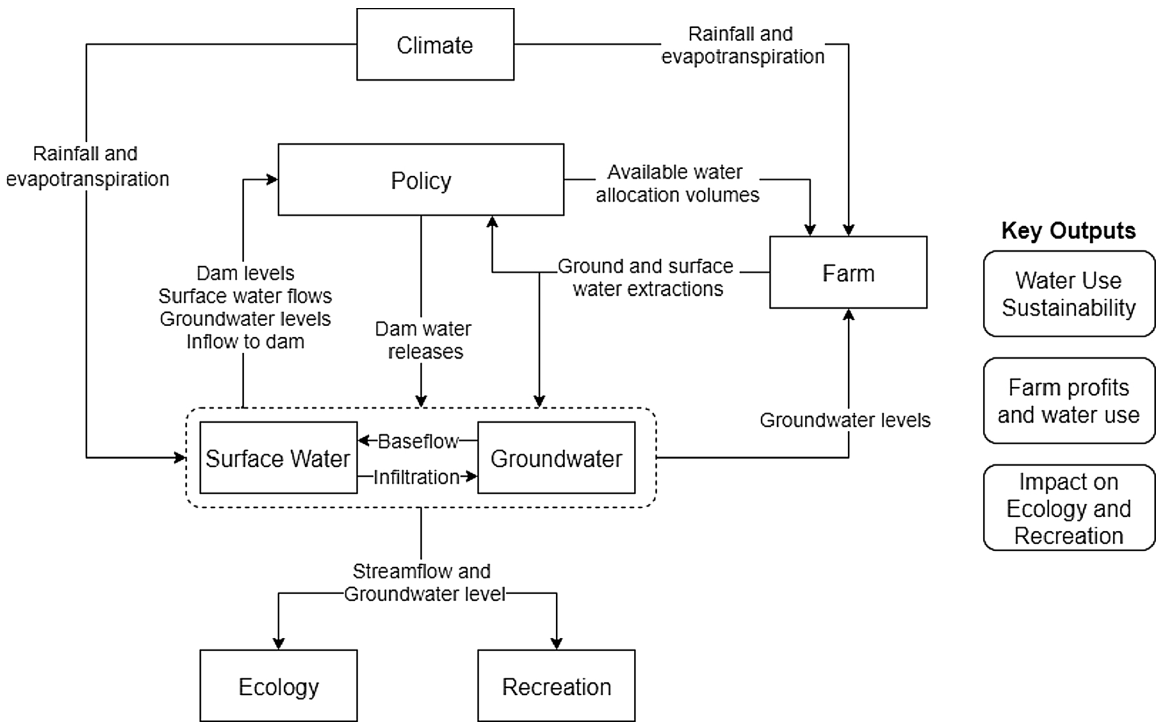


Fig. 2. Component diagram indicating the interactions between model components in the Campaspe Integrated Model (CIM) and the key outputs relevant to the modelled Lower Campaspe system.

Currently, the model takes approximately 30 min. to run for a single scenario on a desktop computer with a Core i5-7500 CPU. The groundwater model based on MODFLOW-NWT is run as an external program and is currently the primary bottleneck limiting further increases to runtime efficiency. Although MODFLOW itself is written in Fortran – considered to be a "fast" language – numerical solution of groundwater models is time consuming and the MODFLOW software itself uses several input files which must be written out for each time step and the results read back in. Overheads due to the constant file operations takes a large proportion of the runtime. As much as 43 % of the CIM runtime can be attributed to the groundwater model. MODFLOW's position as an industry standard was the primary reason for its use.

The exploratory approach conducted for the study involved many runs and so the CIM was designed to run multiple scenarios in parallel in order to overcome computational runtime issues. Model runs are invoked via the command-line and is compatible with both Linux and Windows systems. No graphical user interface was developed for the study as use by non-technical "end-users" was not planned and is not expected.

3.3. Climate and scenarios

Changes in rainfall and temperature influence water availability and their trends can impact the volume of irrigation water necessary to achieve optimum crop growth and yield throughout a growing season. The importance of considering climatic influences in managing farm processes is reflected in the survey conducted by Ticehurst and Curtis (2016). Over 80 % of respondents ranked change in rainfall patterns and the impact of drought as "important" or "very important".

Comparing long term (100 year) average growing season rainfall against more recent trends highlights the impact of the Millennium Drought (1996–2010). A growing season refers to the time span in which crops are usually sown and harvested, defined here as May to February. Average long-term growing season rainfall between 1892–2013 amounted to ~420 mm in line with the reported usual growing season rainfall of 400–500 mm (EcoDev, 2015). Growing season rainfall from 1982 to 2016 however shows a decrease of 67 mm to ~353 mm (see Fig. 3). The trend of decreased rainfall during crop growth may continue with agricultural water management becoming increasingly challenging.

To investigate the impacts of a changing climate, historic and future climate data were sourced from the Climate Change in Australia data service (https://www.climatechangeinaustralia.gov.au; CSIRO, 2017). The provided datasets consist of daily rainfall and evapotranspiration in 5 km grid format for a 30-year period. The future climate data provided were developed through a process of scaling historic observations (described in Mitchell, 2003) and thus exhibit similar rainfall trends to historic observations. Pearson correlations between climate scenarios are depicted in Fig. 4 with a minimum correlation value of 0.88 and 0.99 for rainfall and evaporation data respectively. Future climate datasets are based on the historic timeframe from 1981 to 2011, and thus cover the Millennium Drought period. Therefore, each climate scenario includes a representation of a multi-year drought at differing levels of severity. For the purpose of calibration and analysis, the historic climate dataset was extended to 2016 in order to capture the post-drought recovery. The set of climate scenarios covers RCP4.5 and 8.5 for 2016–2045, 2036–2065 obtained from multiple climate models, "best case", "worst case" are wettest and driest across models, whilst "maximum consensus" represents conditions somewhat comparable to historic experience.

To determine the range of conditions that the climate sequences represent, the aridity index (AI) developed by the UN

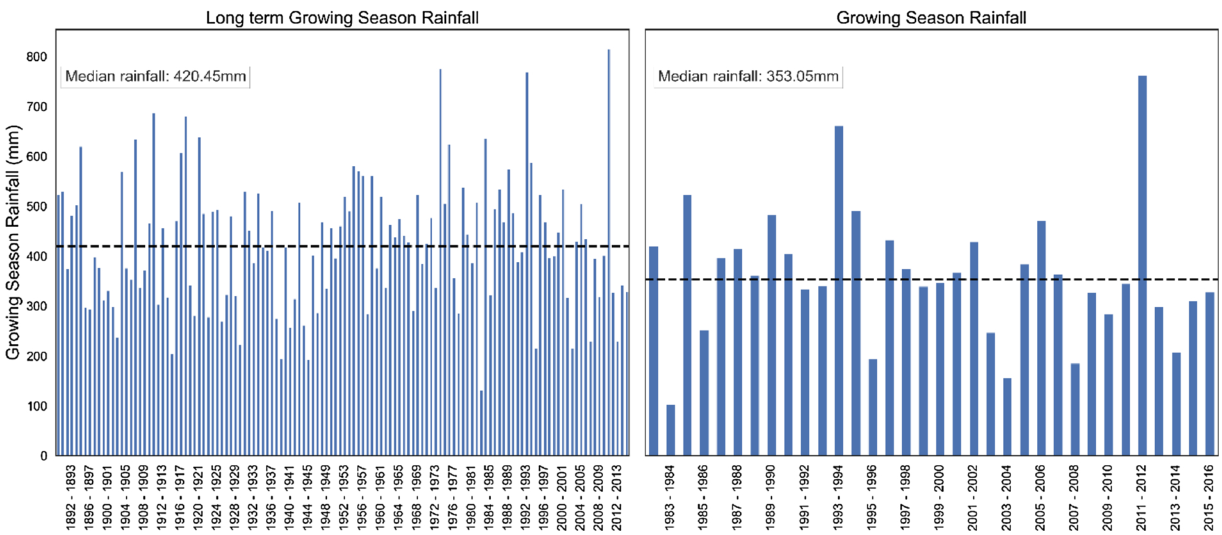


Fig. 3. Comparisons of growing season rainfall over the long-term (100+ years, left panel) and a shorter (30+ year) period (right panel). Dashed black line indicates median values which were found to be 420 mm over the long term which reduced to 353 mm in the recent past (1982–2015 growing seasons) indicating the effect of the Millennium Drought. Long term data were developed through interpolated historic rainfall data (see Vaze et al., 2011).

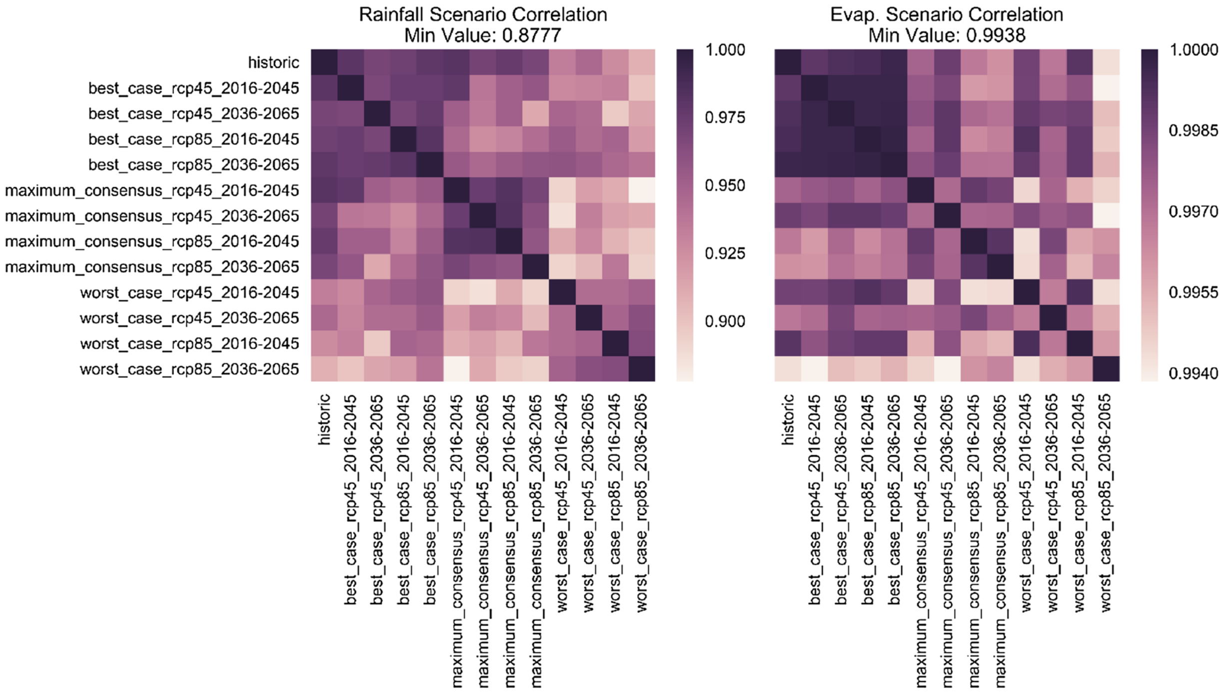


Fig. 4. Pearson correlation matrix between future climate scenarios and historic observations.

Environment Programme (UNEP) was used to compare the climate scenarios. The index is calculated as P/PET, where P is the annual average rainfall and PET is the annual average potential evapotranspiration (Gamo et al., 2013). An AI value of 0.2 to 0.5 indicates a semi-arid climate condition (Sahin, 2012), and the AI value for the historic climate dataset falls within these bounds with a value of 0.34, as is expected for this semi-arid location.

Of the procured climate datasets, the driest conditions are represented by the notation "worst_case_rcp45_2016-2045″ ($AI = 0.26$). The "best_case_rcp45_2016-2045″ scenario was most comparable to historic aridity ($AI = 0.34$), while "best_case_rcp45_2036–2065" was the least arid ($AI = 0.35$). The indicated climate scenarios are taken to represent the extremes of future climate variability (i.e. the most and least arid conditions) and one that is most similar to historically observed conditions. For this reason, modelling with the other climate scenarios were not conducted in-depth. Interpolation between the extremes would likely provide indicative intermediary results. These climate scenarios are referred to as "wet", "usual", and "dry" for brevity from here on and were used to drive the CIM in order to investigate the impact of such changing climate conditions.

3.4. Policy model

Significant reforms to the water policy in the Campaspe introduced water trading, user carryover (specified below), and environmental water provisions (Alston and Whittenbury, 2011; McKay, 2005; Wheeler and Cheeseman, 2013). As part of these reforms the environment is regarded as a water user with water entitlements and yearly water allocations similar to any other entity with a water licence. The policy model provides representations of policies that determine water allocation and carryover.

Access and use of water resources are governed by a licencing system in which a licence held by a water user specifies a given volume of water called an *entitlement*. Separate licences are issued for surface and groundwater. Access to the full entitlement is not guaranteed and is subject to water availability. An *allocation* is announced by GM Water each irrigation season indicating the percentage of the entitlement an irrigator can use – i.e. 100 % allocation is equal to the full entitlement volume. Carryover here refers to the amount of unused water allocation which licence holders (irrigators and the environment) are able to add to the following year's allocation.

Current carryover rules indicate that irrigators are able to carry over 95 % of unused surface water, with 5 % deducted to account for evaporation loss (DSE, 2012a, 2012b), and 25 % of unused groundwater (Goulburn-Murray Water, 2015). While irrigators can decide not to carryover (DSE, 2012b), it is assumed in the model that carryover always occurs. As indicated in Fig. 2, the policy model dictates availability of water for agricultural and environmental purposes and is itself influenced by climate conditions, dam and groundwater levels.

3.4.1. Surface water policy

Surface water allocations in the Lower Campaspe are calculated based on the available volume of water in the dam as well as projected inflows. GM Water holds rights to 82 % of the dam volume as well as 82 % of projected inflows and the sum of these is used

Table 1
Campaspe entitlements in ML for each farm zone used in the modelling (see Fig. 6). Entitlements for a given water system were proportionally distributed based on zonal area.

Zone ID	High Reliability Entitlement (ML)	Low Reliability Entitlement (ML)	Goulburn High Reliability Entitlement (ML)	Water System Name
1	0	0	0	–
2	12050.87	4008.06	0	Campaspe River (Eppalock to Weir)
3	0	0	0	–
4	5703.12	1896.83	0	Campaspe River (Eppalock to Weir)
5	0	0	0	–
6	1086.30	353.10	0	Campaspe River (Weir to WWC)
7	674.90	383.80	215.10	Campaspe Irrigation Area
8	46.71	7.68	39249.62	Rochester Irrigation Area
9	707.05	0	0	Campaspe River (WWC to Murray)
10	96.68	15.91	81223.37	Rochester Irrigation Area
11	224.44	0	0	Campaspe River (WWC to Murray)
12	0	0	0	–

to calculate the total allocation for the Lower Campaspe catchment. The allocation is announced on 1 July and is re-calculated throughout the irrigation season. As such, a water user's allocation may increase over the course of the growing season if larger-than-expected rainfall and inflows occur, and so the volume of allocations will not decrease retrospectively. Dam operation requirements as dictated by local policy and legislative rules, including environmental watering and minimum flow rules, are also factored into the allocation calculations. The model represents this process on a two-week basis (rather than 6-weekly as happens in reality) to allow for better interoperation with the farm model.

A complication in the surface water policy model is the fact that irrigators in the Lower Campaspe hold additional licences for water from the neighbouring Goulburn catchment (east of the Campaspe). Rather than model another catchment in its entirety, linear relationships at three different levels were developed between the Campaspe and Goulburn allocations. Details on this relationship may be found in Appendix C. These are referred to as "high", "median" or "low" and represent the level of allocation volume available from the Goulburn catchment for a modelled future scenario. The separate allocation scenarios allow the model to account for conditions which influence availability of water from the Goulburn including variability between the two catchments. These different conditions may be due to climate variability, water demand, and policies that affect Goulburn allocations which in turn may affect the indicators of interest for the Lower Campaspe. Associated costs were assumed to be identical to that of GM Water. Water entitlements for individual farm zones (detailed in Section 3.5 below) are shown in Table 1 and were calculated by proportionally distributing the entitlement based on the Water System area under which the given zone falls under.

3.4.2. Groundwater policy

Groundwater allocations are restricted by policy rules surrounding what are referred to as "trigger levels". These indicate the groundwater level at which allocations are reduced to a specified percentage or, at its lowest level, no allocation at all. Groundwater use is progressively restricted in this manner as the water table decreases in order to prevent salinity intrusion. Allocations are determined based on the recorded groundwater level at two reference bores with the IDs 62589 in the south and 79324 in the north Goulburn-Murray Water, 2015) and are detailed in Table 2 (under "Current Modelled Allocation").

Table 2
Groundwater allocation rule sets (current and hypothetical). Reduced allowable allocations under surface water allocations > 60 % or 80 % depending on scenario then saves water for use in dry (drought) periods. Groundwater head at reference bores influence allocations for separate Zones.

					Proposed Trigger Rules	
Farm Zone(s)	Reference bore	Depth from natural surface (m)	Water level (mAHD)	Current Modelled Allocation	Non-drought Allocation*	Drought Allocation
3 – 12	79324	0 – 16	82.1 – 97.1	100 %	80 % or 100 %	100 %
		16 – 19	79.1 – 82.1	75 %	65 %	100 %
		19 – 22	76.1 – 79.1	50 %	35 %	100 %
		22 – 25	73.1 – 76.1	40 %	25 %	40 %
		< 22	< 73.1	0 %	0 %	0 %
1 and 2	62589	0 – 16	115.8 – 131.8	100 %	80 % or 100 %	100 %
		16 – 18	113.8 – 115.8	75 %	65 %	100 %
		18 – 20	111.8 – 113.8	50 %	35 %	100 %
		< 18	< 111.8	0 %	0 %	0 %

* non-drought is defined as > 60 % or > 80 % surface water allocation as a scenario option.

In the current policy context surface and groundwater are managed separately, however stakeholders have explicitly suggested a move towards a conjunctive management approach. A possible change in the policy space includes the relaxation of trigger levels during dry periods (thereby enabling irrigators to use more groundwater) followed by compensatory actions to increase recharge during wet periods. Hypothetical trigger rules for conjunctive use were developed for the model. These rules define two sets of trigger levels (referred to as the "drought" and "non-drought" rulesets) and management is switched between these depending on the surface water allocation. "Drought" trigger rules come into effect after consecutive years of surface water allocations below 60 % or 80 %. Depending on scenario, this may be after 1 or 3 years after which irrigators are able to use their full groundwater entitlement for all but the lowest trigger level. These are detailed in Table 2 and referred to collectively as the "proposed trigger rules".

Farmers surveyed in the area have indicated that such increased use of groundwater is technically feasible. Most irrigators are reliant on surface water resources, with 91 % of those surveyed holding surface water licences compared to 22 % who additionally hold groundwater licences (Ticehurst and Curtis, 2017). Groundwater use historically has reached a maximum of 60 % of allocated volumes, although this has increased to 80 % in recent times (2016 water usage as reported in Ticehurst and Curtis, 2017). Recalling that 25 % of unused water can be carried over, the current view is that irrigators are reserving water for drier times accepting a trade-off between enhanced yields in wet years in return for water security in dry conditions. Hence the farming community could be described as managing water resources in a conjunctive manner, albeit informally.

The CIM includes a scaling factor (60–100%) to adjust the amount of allocated groundwater considered for use by the farm model to reflect this informal management approach. Note that setting the scaling factor to 100 % only makes all groundwater available, it does not enforce its use. Limiting allocation to 60 % then reflects a return to historic behaviour in which farmers intentionally reserve water to enhance future water security.

Water usage and its impact is indicated with percentage increase or decrease in water use (both ground and surface water) compared to the baseline scenario, and the long-term decline (or improvement as the case may be) in terms of depth to groundwater across the farm zones. In addition, two quantities of interest relate to groundwater: average groundwater allocation ("GW Allocation Index") and the relative depth to groundwater (head), normalised by the lowest trigger level at each reference bore and averaged over the time series ("GW Level Change"). Average allocations indicate reductions in access to groundwater, whereas the change in relative depth provides an indication of how sustainable groundwater use is under the given condition. Negative changes to the GW Level Change imply increased reliance on groundwater and/or reduced groundwater recharge. Notable model inputs are given in Table 3.

3.5. Farm model

The farm model represents the agricultural system by dividing the catchment into 12 Zones based on their water entitlements, policy rules governing water allocations, and land use (depicted in Fig. 6), corresponding to the Management Zones in use by GM Water (see Goulburn-Murray Water, 2015). Key attributes of a modelled farm include the costs and capital returns of crop sown, nature and cost of irrigation and pumping systems, and the soil water holding properties of the soils found within the Zone. The model focusses on cropping enterprises; water use for dairying is not considered here. Cropping is modelled as a three year rotation of wheat, barley and canola. Further details on the model formulation are provided in Appendix B.

Irrigation systems considered in the model include gravity, pipe and riser, and spray. These represent the cheapest and least water efficient to the most expensive and efficient option available to farmers. Gravity is regarded as the dominant form of irrigation in the Lower Campaspe, although it is known through stakeholder feedback that there is increasing adoption of spray irrigation in the area. Parameters representing soil water capacity were raised by stakeholders as particularly important as they determine what irrigation system is suitable and influence frequency of irrigation events. Farms with sandier soils see a greater benefit from a switch to more efficient irrigation systems as such soils have a lower water retention capacity. While farmers cannot change the soil type, it is possible to improve soil health such that water retention is enhanced through best management practices (Bruyn, 2019). Table 4 lists the considered irrigation options for each farm zone based on the weighted zonal average soil water capacity. For the presented study, only Zones 4 and 9 were modelled as being suitable for spray irrigation.

Farmer decisions are modelled through linear programming at the start of and during a growing season. The available area is allocated to dryland cropping and irrigation with surface and groundwater by optimising profit calculated with assumed per hectare revenues and costs, constrained by water availability. The model then optimises for profitable water usage for each subsequent (two-

Table 3
Policy model inputs of interest and their description.

Factor Name	Description
restriction	Groundwater allocation ruleset ("current" or "proposed") used for the scenario run.
drought_trigger	Surface water allocation threshold for switching to the proposed allocation ruleset under conjunctive use scenarios (see *restriction* factor above), defined as a percentage of entitlements. Either 60% or 80%.
max_drought_years	The number of years surface water allocations must be below *drought_trigger* before switching allocation ruleset. Either 1 or 3.
goulburn_allocation_scenario	Assumed allocation relationship between Campaspe and Goulburn ("high", "median", or "low"). Informs allocation volume from the Goulburn catchment.
gw_cap	Scaling factor limiting the volume of groundwater allocation the farm model considers. Set to 100 %, but can vary between 60–100%.

Table 4
Irrigation options considered for each farm zone. All zones can also consider switching to dryland farming. Zones 1, 3, 5 and 12 are modelled with no water entitlements and are assumed to be dryland only. Assumed efficiency ratings for each are given in brackets (see details in Appendix B). See Zone definitions in Fig. 6.

	Irrigation Option		
Zone	Gravity (50–90 %)	Pipe and Riser (70–90 %)	Spray (80–90 %)
1	–		
2	✓	✓	
3	–		
4	✓	✓	✓
5	–		
6	✓	✓	
7	✓	✓	
8	✓	✓	
9	✓	✓	✓
10	✓	✓	
11	✓	✓	
12	–		

week) time step for each farm zone until the pre-determined harvest date. Again, further details of the model formulation are given in Appendix B.

The assumption of this approach is that a series of short-term decisions that are financially sound will result in a profitable outcome. The two-week time step was also necessary to facilitate representation of cross-domain interactions as agricultural irrigation/pumping activities influence stream and groundwater systems. All events (irrigation, planting/sowing, harvest) are assumed to take place within two week periods.

In terms of interactions with other component models, the farm model is influenced by the water allocation determined by the policy model, and subsequently influences the surface and groundwater models through irrigation water extraction (pumping and irrigation activities). Further interaction with the policy model is included due to local policies allowing carryover in which a portion of unused water is made available to irrigators in the following year.

From the perspective of the farm model, three categories of factors can be identified: uncontrollable, limited control, and direct control. Uncontrollable factors are those that cannot be influenced in any direct manner such as climate condition, water allocations, or crop water requirements. Limited control factors are those which a farmer may influence to some degree, such as how much water a soil type can hold, by applying soil management practices. Direct control factors are those that a farmer can influence directly, such as the irrigation system in use, their efficiencies (through irrigation design and monitoring), and so on.

Groundwater cap (`gw_cap`) is a model input of note, and acts to force the model to limit available water for consideration by the farm model. As noted above in the groundwater policy section (Section 3.4.2), only 60 % of allocated groundwater has been used by irrigators in the past. The considered parameter range was 60–100%, where 60 % represents past behaviour and 100 % reflects consideration of all available groundwater for use without regard to future water security. The cap can be applied in combination with the conjunctive management scenario detailed in Section 3.4 above.

Three key indicators were used to represent farm performance: 1) the average seasonal farm profit for the scenario, 2) the average seasonal water use in ML, and 3) the coefficient of variation (σ/μ) for farm income taken as a measure of fluctuations in yearly income (i.e. a volatility index). Larger values for the volatility index indicate scenarios in which farm income undergoes large variations from year to year. Irrigation area (relative to catchment area) and water use are also included.

3.6. Surface water model

The surface water system is represented by a purpose-built variant of the IHACRES rainfall-runoff model (Croke and Jakeman, 2004) written in Fortran and interfaced with a Python wrapper. The component model comprises a rainfall-streamflow model, a routing module and a rating curve module. The component is driven by rainfall and potential evapotranspiration (climate data), estimated groundwater and surface water interactions (from the groundwater component), dam releases (as given by the policy and farm components) and water extractions from the stream (farm component). As water orders are lumped due to the farm model operating on a two-week time step, the volume extracted is disaggregated uniformly across the 14 days as a daily average.

Effective rainfall i.e. rainfall that contributes to streamflow is calculated from rainfall via a non-linear loss module from Croke and Jakeman (2004), modified to partition effective rainfall between quick and slow flows based on modelled catchment moisture status (Croke et al., 2015). Movement of water through the river network is represented by two parallel exponentially decaying stores representing quick and slow flow contributions. A further exponentially decaying store is used to route the flows between nodes (Croke et al. 2006). Modelled flows are converted to stage heights by the rating curve module using data available at most gauge sites. The single exception is gauge 406218 where available data were limited to water level. Losses from the river network to groundwater are considered via interactions with the groundwater component, similar to the approach in Ivkovic et al. (2014).

Dam level, streamflow and level were important quantities to represent as these influence water allocations, ecological health, and recreational use of the dam (described in later sections). Stream flow and level are estimated at specific nodes (show in Fig. 7) which were selected based on the location of gauges with suitable data, taking into consideration the needs of the integrated model (see Fig. 2). These values are passed to the groundwater model in order to estimate infiltration loss and baseflow contribution to surface flow.

Calibrating the model against historic observations proved difficult due to variation of parameter values estimated depending on climate sequence, and so a decision was made to divide the climate sequence into six time periods and calibrate parameter values for each separately using the Differential Evolution algorithm (Storn and Price, 1997). A good fit with historic dam levels was achieved with this approach (with a NSE value of 0.96, depicted in Fig. 8). Calibrated values were then used for the corresponding time periods for all climate scenarios. Each set of parameters have a limited scope of applicability, but their use for the segmented time periods is arguably justified for the modelling purpose as similar patterns can be associated with the provided climate scenarios (as previously described in Section 3.3). Time periods defined for calibration purposes are detailed in Appendix D.

3.7. Groundwater model

The groundwater component serves two roles in the CIM. The first is to provide estimations of the exchanges between surface water and groundwater along the Lower Campaspe River at specific gauge locations. The second is to provide estimations of the groundwater levels for each farm zone and at specific bore locations, which influence allocations of groundwater and pumping costs and is a factor in the ecological indicator metric (see Section 3.8). Interactions at a daily time step with the surface water, farm, ecology, and policy components are included in this manner, and are forced by rainfall, irrigation, pumping, and river stages (depicted in Fig. 9).

Stakeholders provided knowledge and data with respect to pumping, observational groundwater head, and chemistry, which informed the groundwater model boundaries ensuring that the necessary spatial area is captured in the model. The groundwater model represents the Lower Campaspe at a 5 km grid resolution with 7 layers comprising 1386 active cells, spacing of which was informed by hydrogeologic unit rasters from the Victorian Aquifer Framework DSE, 2012c). The relatively large grid resolution was selected as 1) it aligns with the provided climate data, and 2) for computational reasons, as higher resolutions required prohibitively longer run times.

MODFLOW-NWT (Niswonger et al., 2011) was used to construct the groundwater component and is interfaced with via the Flopy package (Bakker et al., 2016). MODFLOW was selected due to its open-source and well-documented nature, which complements our iterative development, and because the modelling platform is regarded as an industry standard. Modelled groundwater extraction is distributed across known wells within each farm zone. Response to input from the streamflow lags the surface water model as component models are run sequentially. It is assumed that the daily lag will not have a significant impact on model behaviour. Recharge is derived by multiplying the sum of rainfall and irrigation by a rainfall reduction parameter between 0 (no recharge) and 1 (no diversion) with values being in the range of 0.001 – 0.43. These rainfall reduction values were derived from a map of recharge ranges for recharge zones defined by rainfall, land use and soil type (detailed in Xie et al., 2019).

Outputs from the groundwater model, while not precise at the scale of local wells due to model resolution, were fit for purpose for indicative average groundwater levels in areas of interest. In the case of the ecology model, the requisite groundwater heads obtained from the groundwater model would be subject to the largest variance due to the proximity to the Campaspe River and variability of heads within a groundwater model cell. For the policy model the trigger bores are chosen to be indicative of larger scale behaviour, and hence the use of the average head in cells that correspond with the trigger bores is deemed appropriate.

Effectiveness of policies in maintaining groundwater levels above the lowest trigger level is indicated by two metrics. The first indicator is a weighted score based on the percent irrigation season allocation across the scenario period (GW Allocation Index) where a score of 1.0 indicates that water users were given 100 % of their entitlements every year, and values close to zero indicate consistently low water table such that pumping is disallowed. The second indicator (GW Level Change) is the averaged normalised change in groundwater level between the start and the end of the model period relative to the lowest trigger level. Negative values then indicate that the lowest trigger level was breached. These two indicators are referred to in the results as the “GW Allocation Index” and “GW Level Change” (see Section 4 for further detail).

Poor performance of the groundwater model in the integrated context compared to historic observations can be expected due to the integrated model structure and design. For one, both the policy and farm models were developed to represent a more recent socio-hydrological context (i.e. post 2010) as required by the purpose of the modelling. This temporal mismatch has the effect of reducing the overall volume of groundwater extraction and recharge compared to historic occurrences. As noted in Section 2, 90 % of the irrigators had stopped irrigation practices leading to the closure of the Campaspe Irrigation District and so the farm model represents this current level of irrigation activity. The policy model includes reforms introduced to allow carryover of unused water to the next irrigation season as well as environmental water provisions which has an influence on the level of groundwater extractions. Contemporary accounts estimate 35,000 ML of water was used for irrigation within the Campaspe Irrigation District, including approximately 3000 ML extracted from the aquifers (Chiew et al., 1992).

We also note reported difficulties in representing the groundwater system of the Campaspe region in earlier studies. It has been previously noted that the Campaspe region is a difficult region to model (Beverly and Hocking, 2014; Goode and Barnett, 2008). One confounding factor is the long history of system regulation and incomplete data with respect to groundwater extraction and usage at a fine(r) scale. Another reason is the local topography and landscape. Towards the south, large elevation changes can be seen which are not captured by the 5 km grid resolution. Due to the coarse resolution in use for the purpose of the model, there are significant differences in elevation between model cells ($>$ 20 m). These model conditions preclude the use of the groundwater model for accurate forecasting. The purpose here, however, is in representing *futures* (expanded on further in Section 4). Thus, the focus in model development was on representing hypothetical groundwater level response to farmer behaviour and the water policies in place. These are discussed in Sections 4 and 5.

3.8. Ecology and recreation indicator models

The models developed to represent ecological and recreational aspects of the Lower Campaspe system produce indicator values reflecting the suitability of stream and groundwater flow for key ecological purposes and dam water levels for recreational purposes. Ecological indicators represent the suitability of flow conditions for breeding, feeding, nesting, and dispersal of platypus (4 indices) and native fish (2 indices), as well as maintenance and regeneration of the iconic River Red Gum trees (2 indices). Suitability thresholds were based on the North Central CMA Environmental Water Management Plan (North Central CMA, 2014) in conjunction with feedback from ecologists from the Australian Platypus Conservancy and the North Central CMA (as detailed in Iwanaga et al., 2018). It was suggested by stakeholders (local representatives from the government department, EcoDev) to lump ecological indicators into a single metric to ease interpretability of results.

The recreation indicator is based on a linear relationship between dam water level and perceived recreational suitability elicited through interviews with stakeholders from GM Water, caravan park managers and other recreational users. All indicator models produce index values which range between 0 (being unsuitable) and 1 (most preferable). Although ecology and recreation are separate and distinct systems they are lumped together here as the recreation index model constitutes a small aspect of the modelling presented here. Details specific to each ecological and recreational component are provided in Appendix E, with further detail available in Fu (2017).

3.9. Conceptual integrated testing

An issue in the development of CIM was the lack of data to validate coupled model behaviour – known as integration testing in software engineering parlance. One example is the lack of farm-level crop yield data (partly due to privacy concerns). Data that were available describe average yield from both irrigated and dryland cropping on a “per farm average” basis for the entire North-Central region of the state of Victoria, an area of approximately 47,000 km^2 compared to 2600 km^2 for the Lower Campaspe. Additionally, the temporal scale of available data is relatively short, with records starting from 1990 to the early 2000s.

Under such circumstances model testing and validation chiefly involved high-level conceptual checks selected to indicate unacceptable model behaviour. Data to develop a strict quantitative metric was not generally available across the range of component model contexts, and so the adopted approach was to contextualise performance against relevant criteria to provide an indication of model acceptability.

For example, while financial data at the individual farm level were not available, regional economic production data are available for the Campaspe (via Australian Department of Agriculture, 2019). Performance of the farm modelling was considered unacceptable if calculated crop profits exceed or comprised an unjustifiably high proportion of the agricultural output of the Campaspe catchment. Modelled crop profits under historic climate conditions amounted to 20.54 % of the Campaspe agricultural activity ($13.81 M of $67.26 M), which stands to reason given that the Campaspe region is principally a dairy farming community. Long-term average values were also used as a qualitative measure of acceptability, wherein long-term behaviour of the model was in line with historic observations (example given in Fig. 10).

It must be noted here that such an approach is subject to available expert knowledge. Tests at this conceptual level may be the only kind of integration tests applicable given the available resources, data and domain knowledge, and the purpose of the model. For this reason, the scope of testing must be carefully targeted and in line with available resources.

4. Methods for the scenario modelling

Exploratory scenario modelling was used to identify conditions which led to beneficial outcomes. The term “scenario” here refers to the specific set of model inputs used for a model run, which in turn leads to a specific spatial and temporal evolution of system variables and final outcomes. From the perspective of farmers and policy managers, these scenarios represent a possible farm and policy context. The scenarios are run through the model and those with beneficial outcomes are subsequently identified. Thus, the model is used in an exploratory manner rather than to obtain precise prediction of events.

Only factors pertaining to the farm and policy models are considered in the scenario modelling. Calibration was not conducted for the integrated model in its entirety. Instead component models were calibrated individually against relevant historic observations of crop yields, dam level, and observed groundwater head and feedback obtained from stakeholders in cases where quantitative data were not available (e.g. for the recreational indicator). The approach is justified as the data used for calibration encompass a wide range of conditions and the trends (rather than absolute values) of the climate scenarios are similar to observed historical data, as explained earlier in Section 3.3. The ecology model takes no inputs other than the modelled streamflow and groundwater level. The total number of model inputs that are directly considered in the CIM comes to 266 of which 212 factors are regarded as constants. This still leaves 54 parameters that can vary between scenarios and are largely inputs for the farm models for the different zones.

Covariance analysis was carried out to determine the number of scenarios to run. It is desirable for the input values not to be correlated across the sampled scenarios as this could induce artificial correlation within the results for those scenarios, i.e. the covariance value should be as close to zero as possible. A threshold of 0.02 for off-diagonal values was selected as a compromise between parameter independence and model evaluation time (see Appendix F). Latin hypercube sampling – a stratified Monte Carlo sampling approach – was used to generate each scenario. In total, 5625 model evaluations were performed, consisting of 1875 scenario samples for each selected climate scenario.

Apart from the 5625 scenarios above, the model was initially run using "best guess" (default) values for each model parameter for all climate scenarios. These are referred to as the "baseline" results. Indicator scores from these then establish a baseline against which further comparisons can be made and represent the "business as usual" case in which the status quo is maintained in terms of farm management and policy under changing climate conditions. Further scenarios are then run allowing parameter values to vary to explore the possible outcomes under a range of conditions. Baseline results are compared against modelled historic outcomes while results from exploratory modelling are presented as comparisons to their respective baseline scenario.

Comparing results to a baseline allows like-for-like comparisons. It is assumed that change in a beneficial direction (see Table 6) is always desired and a scenario which exhibits improvements to system state relative to the baseline scenario is considered a *desirable* outcome. For the purpose of this analysis, a desirable outcome is one in which all indicators perform comparatively better when compared against their respective baselines. To be explicit, Avg. Annual Profit, Ecology Index, and Recreation Index were all required to have scores above 1, whilst Income Volatility should be below 1.

Choice of farm and policy inputs that lead to these desirable outcomes of improved farm, water, and environmental conditions across the possible climate futures and policy conditions are then regarded as *robust*; a system state in which desirable outcomes are achieved under a range of plausible conditions (McPhail et al., 2018). The indicators of interest are described under relevant subsections in Section 3 above, and are summarised in Table 6 with their "beneficial" direction indicated. Once desirable scenarios are identified, the contributing model inputs were ranked using a random forest feature scoring approach. Feature scoring identifies factors of relative import towards a given result. Dimensional stacking (Suzuki et al., 2015) is then used to identify the inputs that led to the outcome of interest. Appendix A gives a description of terms used in referring to outcomes.

The above approach allows communication of the model inputs that lead to the outcomes under different scenario conditions. The range of results then represent the uncertainty in achieving desirable outcomes and increases confidence that a given condition leads to a desirable change, at least under the model assumptions (Reichert and Borsuk, 2005). The adopted method also avoids prescribing specific conditions to be undertaken, and instead considers a range of alternate decisions that fulfil stakeholder requirements (Herman et al., 2015).

5. Results and discussion

Overall, the results indicate that further pressures will be placed on farmers and the environment without changes to adapt to uncertain climate conditions, as have previous studies and reports in the Murray-Darling Basin (Austin et al., 2018; MDBA, 2019; Steffen et al., 2018). Comparisons of results against the modelled historic climate scenario are depicted in Fig. 11. Surface water allocations are lower in the majority of scenarios, which is expected here as they are typically more arid compared to the historic condition (as shown in Fig. 5). In order to cope with this situation the area under irrigation is reduced so as to maintain the necessary ML/ha volume to attain crop growth and maximum yields. Doing so however comes with a commensurate reduction to (average

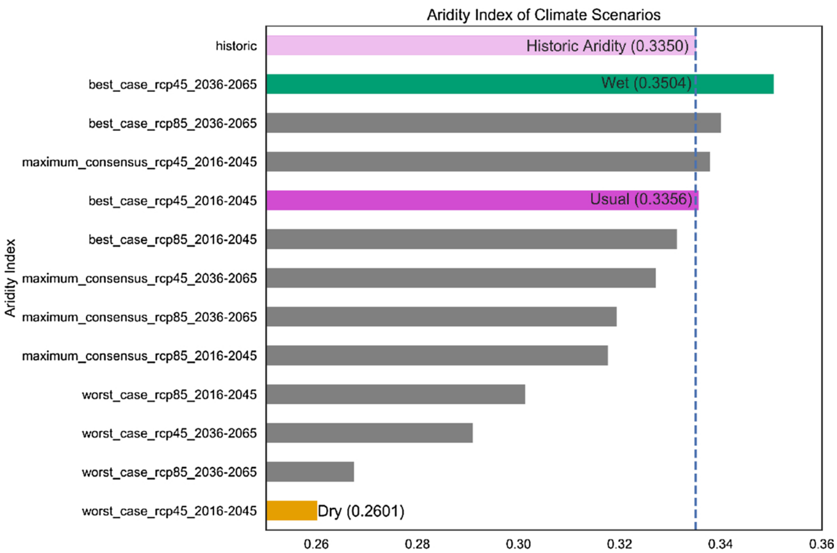


Fig. 5. Future climate scenarios sorted by calculated aridity index value in descending order. The dashed line indicates historic aridity for comparison. The semi-transparent magenta bar (top) represents the historic aridity index for the Lower Campaspe catchment (0.335). Coloured bars indicate the wettest (green), comparable to historic conditions (opaque magenta), and driest conditions (orange). Climate scenario data that were not used extensively in the modelling are coloured grey (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

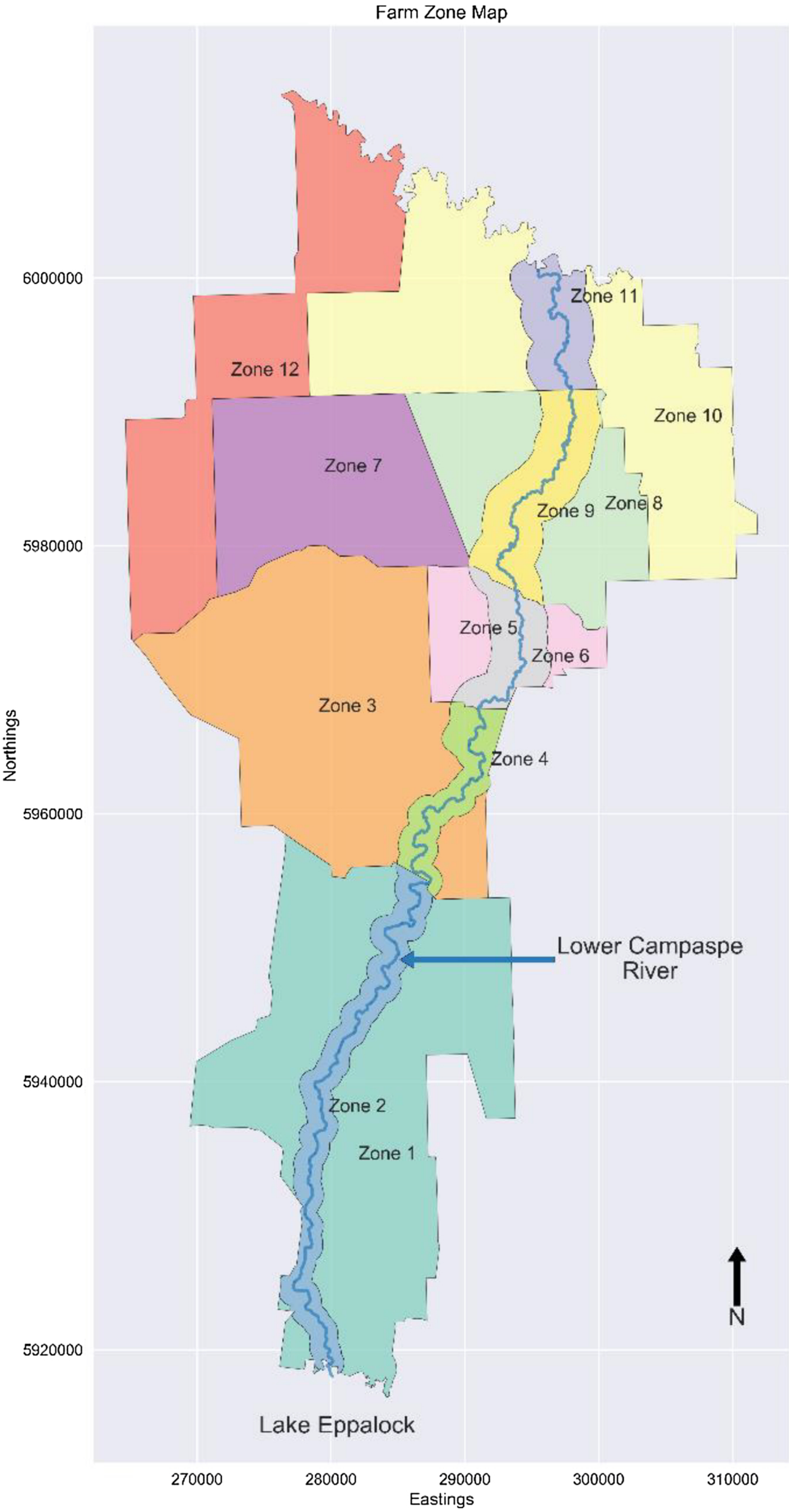


Fig. 6. Lower Campaspe study area divided into farm zones for the farm model.

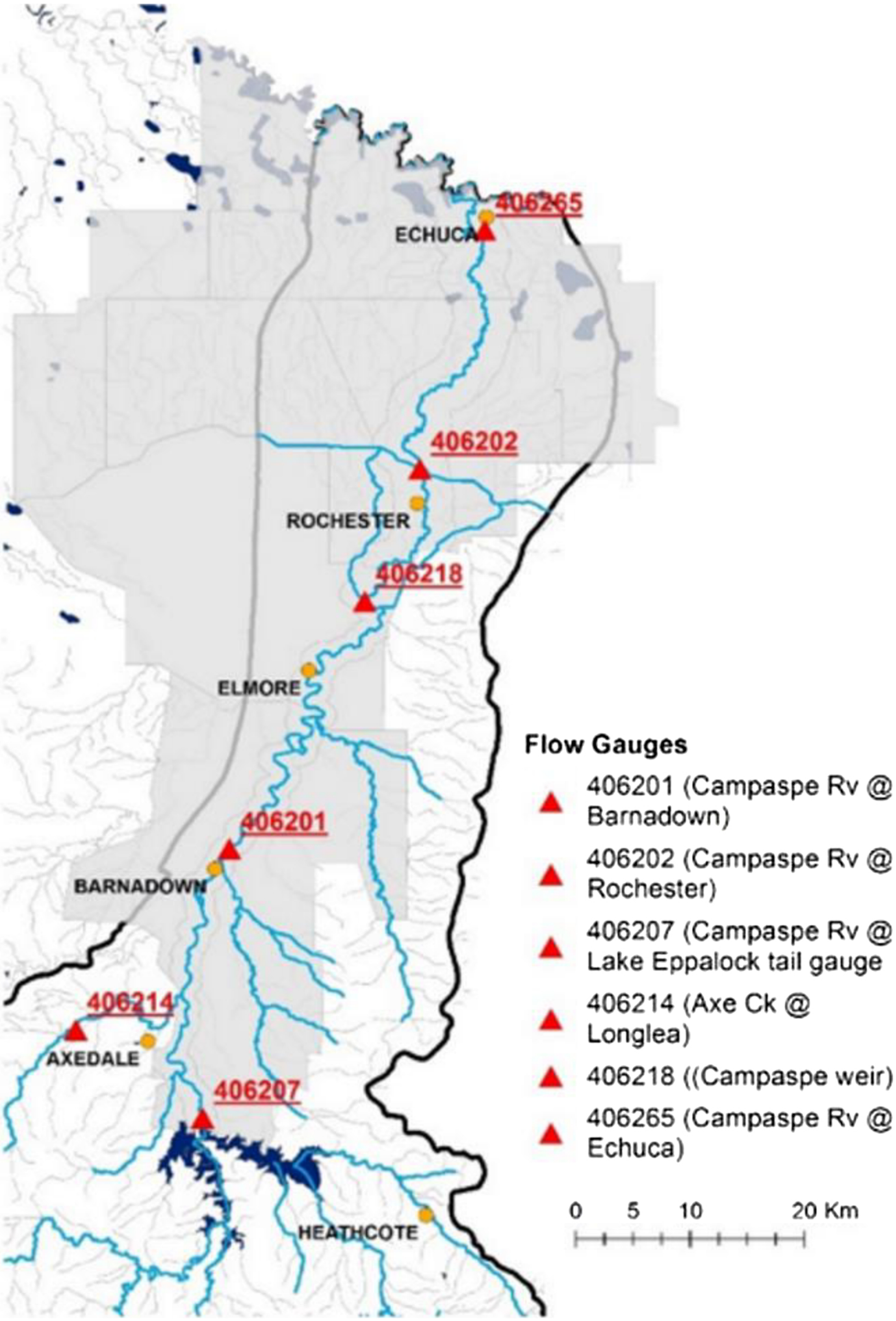


Fig. 7. The six stream gauge locations used to indicate streamflow and level for the Lower Campaspe model. Figure from Iwanaga et al. (2018).

annual) income along with larger fluctuations from year to year (depicted by the Income Volatility indicator in Fig. 11).

A reduction in groundwater use can also be seen (compared to the modelled historic scenario), attributable to the reduction in irrigated area. Groundwater use under drier conditions is higher than their wetter counterparts but does not offset the lack of available surface water. As would be expected, ecological and recreational indicators diverge from the historic baseline results based on the climatic condition. It should be noted that any decrease to the ecological indicator is undesirable given that riverine health is already considered to be highly stressed.

Comparisons of the 5625 scenario results against the historic baseline revealed 796 with desirable outcomes, however none of these were robust, i.e. beneficial under all climate conditions, with no desirable outcomes identified under dry climate conditions.

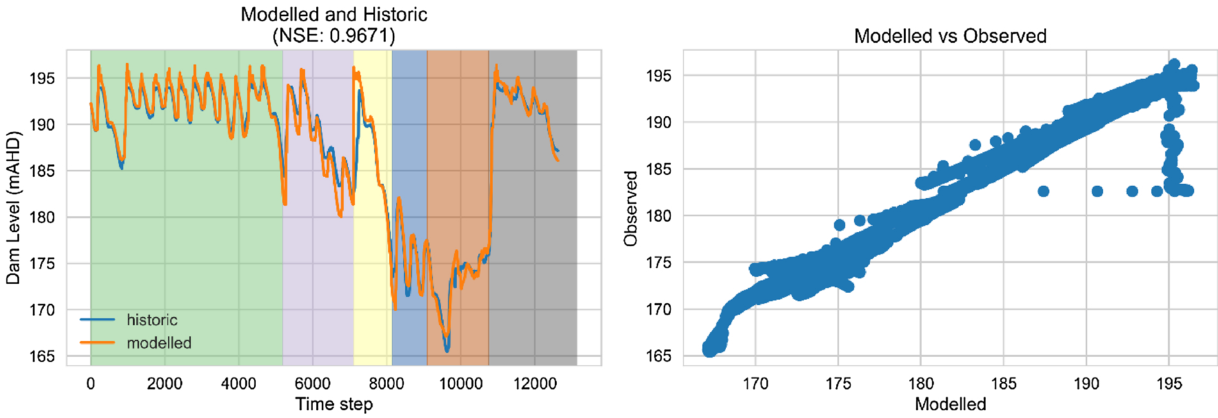


Fig. 8. Calibrated model compared to observed historic dam level (in mAHD) with an NSE of 0.967. Model parameters were estimated separately for each of the six segmented time periods indicated with the background colours in the left-hand plot.

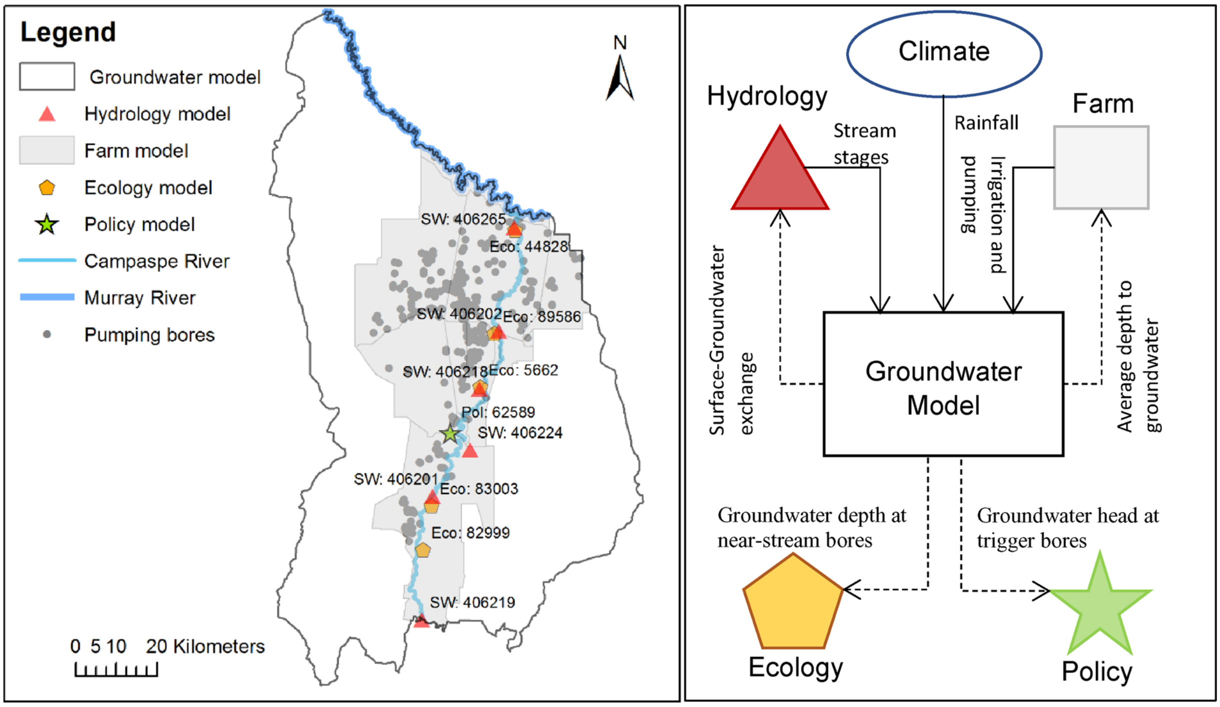


Fig. 9. Groundwater model components and model area as well as points and interactions with other component models from the perspective of the groundwater model. Figure adapted from Iwanaga et al. (2018).

The lack of desirable outcomes for the dry climate scenario suggests that improved outcomes compared to historic conditions become increasingly improbable as the climate becomes drier, regardless of what changes are made (Fig. 12). One suggested scenario was the adoption of irrigation systems with higher water application efficiencies to increase water savings across the catchment. Although soil factors may make such a scenario unlikely (see Section 3.5), no beneficial outcomes were identified in such cases even in the case where all non-dryland zones (i.e. all zones except for 1, 3, 5, and 12) adopted improved irrigations.

Given no robust scenarios were identified against historic results, we then conduct comparisons against baseline results for specific climate conditions. In other words, scenarios under “dry” climate conditions were compared against the “dry” climate baseline in order to identify conditions that led to relative improvements. A small number of scenarios were identified as being robust under all climate conditions (93 scenarios). Input parameters of most import leading to these are related to crop, field, and irrigation factors, followed by policy factors and the maximum amount of groundwater considered. These farm level factors are listed in Table 5 (under Section 3.5). The high feature score attributed to crop, field, and irrigation factors is perhaps unsurprising, but it does highlight the importance of a well-informed farming community. Over-estimating crop water requirements and costs reduces the irrigation area considered, and poorly set up or poorly maintained irrigation infrastructure (including pumps) and soil health

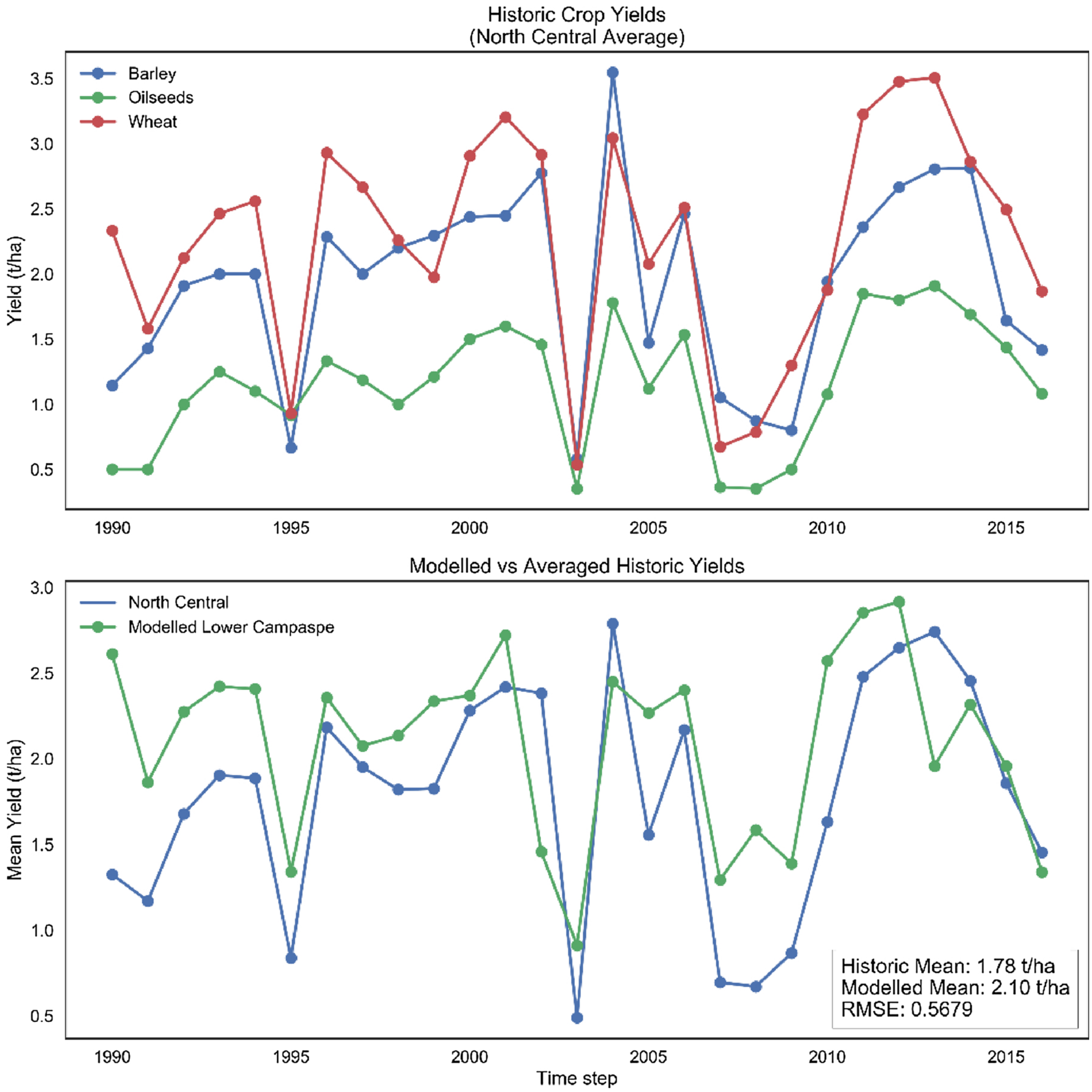


Fig. 10. Modelled catchment average yields produced by CIM under historic climate conditions against available historic data (per farm average for the North Central region) for the crops considered. The Campaspe region reportedly produces higher yields compared to the North Central average.

increases costs and water usage thereby impacting farm profitability (Fig. 13).

At the field level, improving the availability of accurate information on soil water holding capacity has the greatest contribution to farm profitability. Specifically, soil water retention within Zones 5, 10, and 9 are indicated to influence outcomes more than at other locations in the Lower Campaspe (Fig. 14). At the same time, targeted adoption of spray irrigation within Zones 4 and 9 (the zones indicated to be most amenable to spray irrigation systems in our analysis) did not necessarily lead to robust outcomes, further highlighting the point that simply improving irrigation systems across the catchment is not a viable adaptation strategy. Possible reasons for this include operational costs incurred with spray irrigation and the lower water efficiency of pipe and riser systems contributing to increased recharge or streamflow under certain conditions, having the effect of improving ecological outcomes (i.e. due to higher return flows).

One particular aspect of interest was the percentage of groundwater allocation which irrigators consider for use (`gw_cap`, previously explained in Section 3.4.2). While irrigators have historically used 60 % of groundwater allocations in a bid to enhance future water security, increased consideration of groundwater use improves the likelihood of desirable outcomes to be experienced under all climate conditions (Fig. 15).

Increasing the volume of groundwater to be considered for use compensates for the decreased surface water availability under more arid conditions. Consideration of all available groundwater resources (100 %) may be necessary under proposed conjunctive use rulesets, likely due to the reduction in accessible groundwater in wet periods enforced by the proposed rules. This suggests that

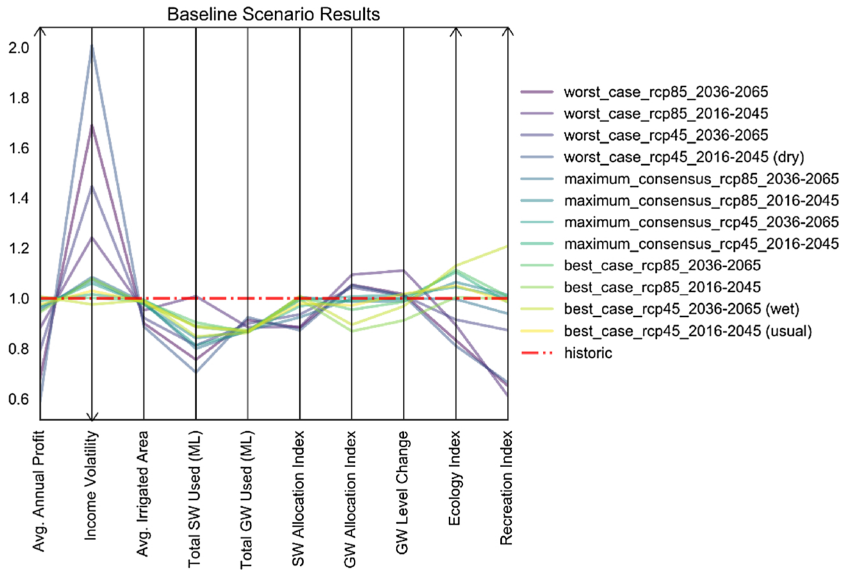


Fig. 11. Baseline future scenario results compared against modelled historic outcomes (relative values) with beneficial directions indicated as directional arrows. Farm profitability is reduced in all but the least arid scenario, while volatility of income generally increases. Ecological indicators generally only improve under "wet" conditions.

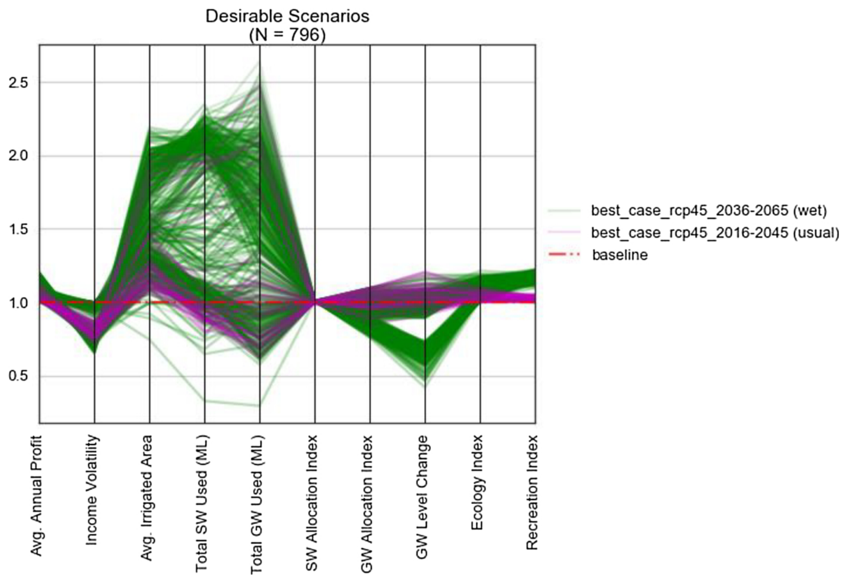


Fig. 12. Desirable scenarios identified when compared to historic baseline (796 of 5625). No desirable results were identified for the dry climate scenario.

while the informal and disparate approach to conjunctive use may be adequate, farmers who are overly cautious and restrict groundwater use may not experience the maximum possible yields and capital returns in the long term.

The results shown in Fig. 13 emphasize the importance of farm level factors towards achieving robust outcomes. Stakeholders have specified various programmes designed to engage with and promote farm management best practices (e.g. DEDJTR, 2015a; Department of Economic Development, 2017). Assuming then that farms have little room for further improvement in terms of soil water holding capacity, knowledge and consideration of crop properties, and on-farm infrastructure such as irrigation and pumping systems, we can then identify which factors contribute towards robust outcomes. Focusing the analysis on factors external to the farming system itself, the amount of allocations from the Goulburn system is identified as playing a large role, perhaps unsurprising given the volume of entitlements which Campaspe irrigators rely on, particularly in zones 8 and 10 (see Table 1). Aside from the Goulburn catchment allocations, the conjunctive use policy rules in place for a given scenario (`gw_restriction`) have a notable effect (Fig. 16).

In scenarios where Goulburn allocations have lowered we see a concentration of robust scenarios wherein conjunctive use of

Table 5
Description of notable farm model inputs.

Factor Name	Category	Description
gw_cap	Direct Control	Groundwater cap – described in the policy section above. Reflects maximum volume of groundwater allocation considered by a farmer. The farmer may choose to favour future water security under dry conditions by carrying over (25 % of) unused water potentially sacrificing the ability to achieve maximum crop yields in the current season.
irrigation_efficiency	Direct Control	Irrigation efficiency rating of a given irrigation system. More efficient irrigation systems require less water to irrigate the same area, but cost more to operate. Efficiency can also be improved by adopting best management practices, which irrigators have been doing.
pumping_costs	Direct Control	Perceived cost of pumping on a $/kilowatt basis. Farmers can adopt more efficient pumps, irrigation designs, or other practices which reduce this cost.
Irrigation	Direct Control	Indicates adopted irrigation system for a Zone
crop_water_use	Limited Control	Perceived crop water requirements. Under-estimating crop water requirements leads to too large an irrigated area relative to available water. Farmers can become better informed of crop attributes but cannot directly control these.
crop_root_depth	Limited Control	Perceived average crop root depth of a crop at each growth stage. Affects irrigation scheduling, as crops with deeper roots typically require less irrigation events.
TAW_mm	Limited Control	Total Available Water – represents the soil water holding properties of a given soil type. Farmers may improve soils through best management practices or invest in monitoring but cannot change the soil type at the landscape/field level.

Table 6
Summary of indicator metrics and their beneficial directions. All values should be taken as relative to a baseline, either modelled historic outcomes or a relevant scenario baseline. Indicators provided without a beneficial direction are included to contextualise outcomes, with respect to area irrigated and water used.

Indicator	"Beneficial" direction	Purpose/Description
Avg. Annual Profit	Up	General farm profitability
Income Volatility	Down	Severity of fluctuation in profit from year to year
Avg. Irrigated Area	–	Average total irrigated area over time
Total SW Used	–	Comparative surface water use
Total GW Used	–	Comparative groundwater use
SW Allocation Index	–	Surface water allocations throughout scenario period.
GW Allocation Index	–	Groundwater allocations throughout scenario period.
GW Level Change	–	Averaged normalised change in groundwater level between the start and end of a scenario period.
Ecology Index	Up	Assessment of ecological outcomes of streamflow
Recreation Index	Up	Assessment of impacts of dam levels on recreation

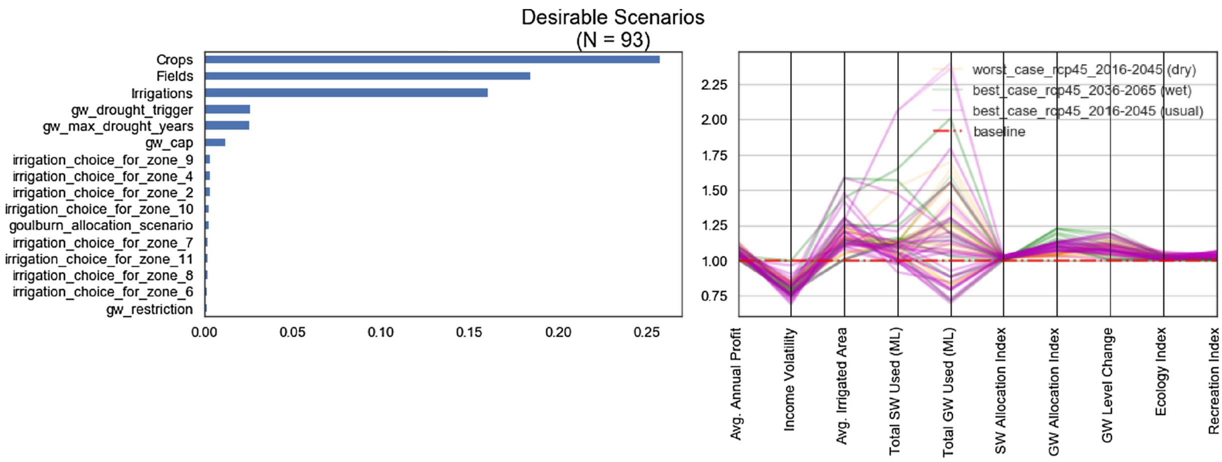


Fig. 13. Robust scenario results (93 of 5625) under all climate conditions. Salient factors leading to robust outcomes include crop, field, and irrigation factors followed by policy factors and the amount of groundwater considered for use.

water is allowed. Under median allocation situations, conjunctive use allows robust outcomes to be achieved while maintaining groundwater use in line with historic behaviour. Without conjunctive use enabled however, considered use of groundwater has to increase to 90–100% in order for the changes to be robust (Fig. 17). As Goulburn water availability further decreases, groundwater use becomes especially important towards achieving robust outcomes. Robust outcomes are more likely if conjunctive use is enabled along with high levels of groundwater use (Fig. 18). The modelling suggests that groundwater levels can be maintained above the lowest trigger level, however careful consideration is required especially with regard to the effect on salinity and water quality issues

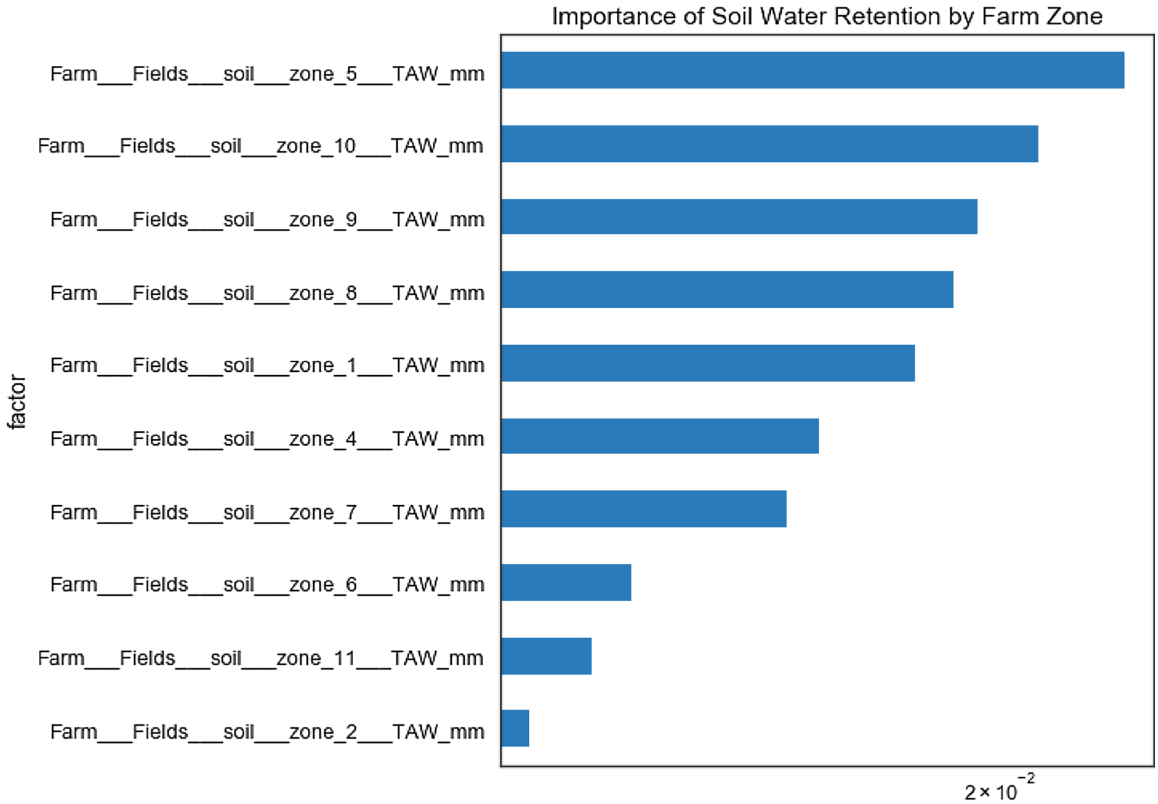


Fig. 14. Influence of soil total available water (TAW) towards achieving robust outcomes (shown in log scale).

– aspects which the modelling presented here did not cover.

The results suggest that improvements to farm soils and infrastructure will be beneficial within the Campaspe, and additional communication, training, and (financial) incentive programmes beyond what has already been occurring may increase benefits. Any such programme should consider possible issues surrounding social acceptability and be cognisant of issues with previous approaches to appraising the cost-benefits (Grafton and Wheeler, 2018). While increasing groundwater use is generally beneficial, possible issues surrounding social acceptability of increased use and water quality, particularly salinity, should be fully considered. The results raise the possibility of increasing groundwater allocations in the Lower Campaspe, especially if Managed Aquifer Recharge is adopted in the region (Chiew et al., 1995; Ticehurst and Curtis, 2017).

5.1. Model and scenario uncertainty

Integrated models are constructed through the interfacing of models that collectively cross disciplinary lines and their respective system boundaries. Intuitively, uncertainty will not decrease if more models are added, simply due to compound uncertainty. This is the uncertainty that arises as outputs from one model are used as inputs to another, with each interaction propagating some amount of error (Dunford et al., 2015; Refsgaard et al., 2007).

In the context of the CIM the sources of uncertainty and their contributions to total model uncertainty are too great to list out individually within the confines of this paper. Formal analyses of individual model components and total model uncertainty including structural uncertainty with regards to model selection, is the subject of another paper. Qualitatively, however, the farm model represents the largest source of (compound) uncertainty as all components, except for climate, are influenced by mechanisms internal to the farm model. In other words, the farm model behaves as a nexus point between models and thus the errors in the interoperated data may be cancelled out or compounded and subsequently propagated through. An over estimation of streamflow may be "corrected" in a sense by over estimation of required crop water and under estimation of irrigation efficiency. Similarly, the opposite may also be true. Such influence may occur directly (e.g. streamflow reduced due to farm water extraction) or indirectly (e.g. ecological suitability influenced by streamflow). The reader is once again referred to Fig. 2 which indicates the component interactions.

Uncertainty within the models was addressed through participatory engagement processes (Section 3.1) and the conceptual testing process (Section 3.9), both of which ensured model behaviour is qualitatively plausible (as judged by stakeholders), and which involved changes in response to their feedback. On top of this, exploratory modelling was applied to hypothetical policy contexts identified by stakeholders and the range of on-farm activities considered, using results from the farmer survey (Ticehurst and Curtis, 2017, 2016).

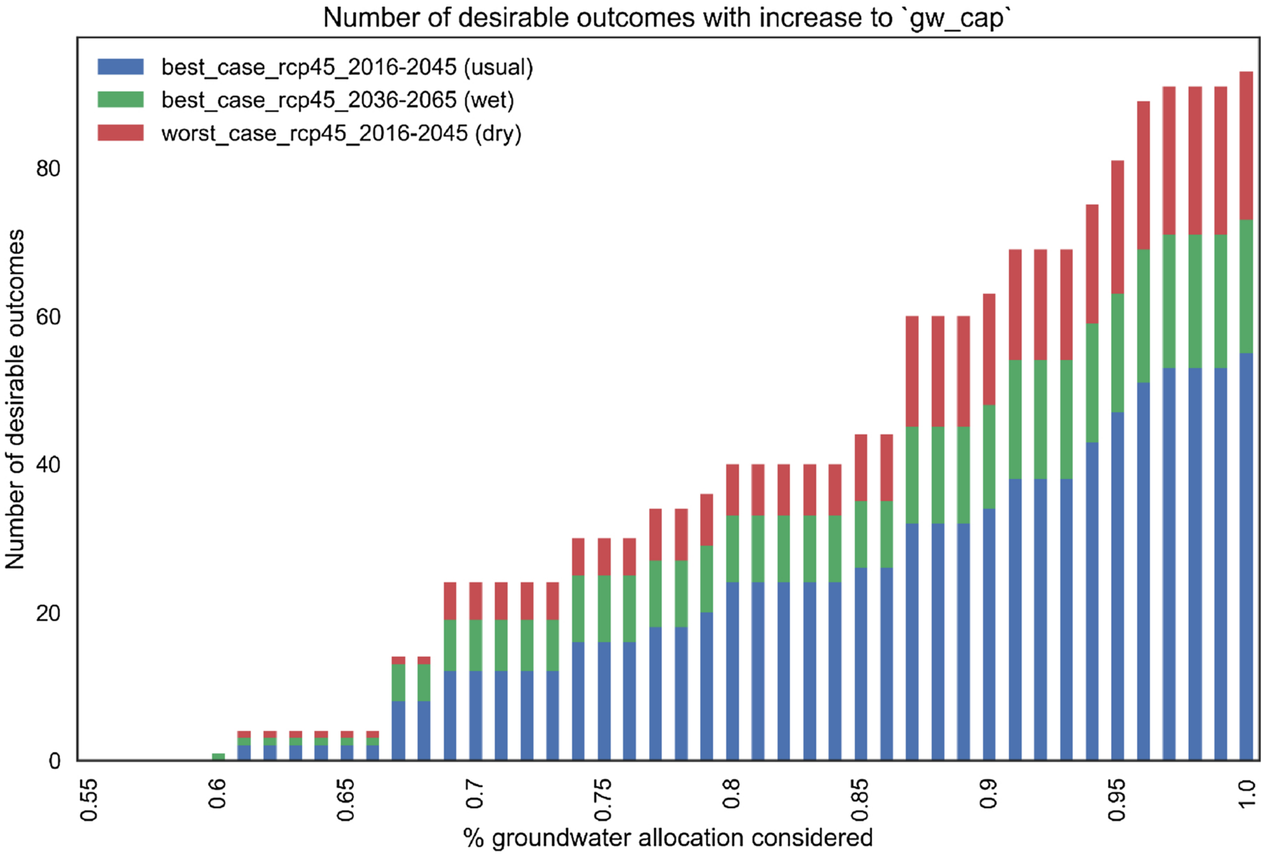


Fig. 15. Number of desirable outcomes experienced under each modelled climate condition. The number of scenarios leading to desirable outcomes increase as the amount of groundwater considered increases (compared to the historic 60 %).

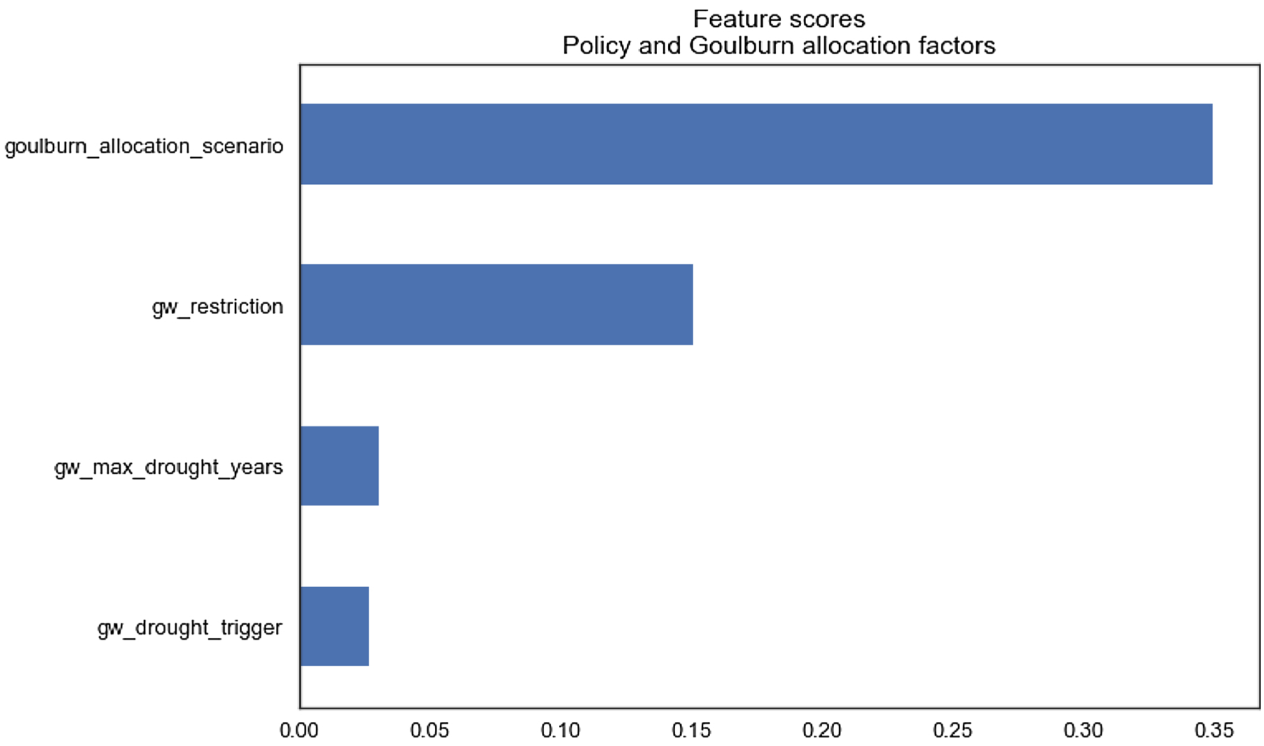


Fig. 16. Features scores when considering policy factors alone.

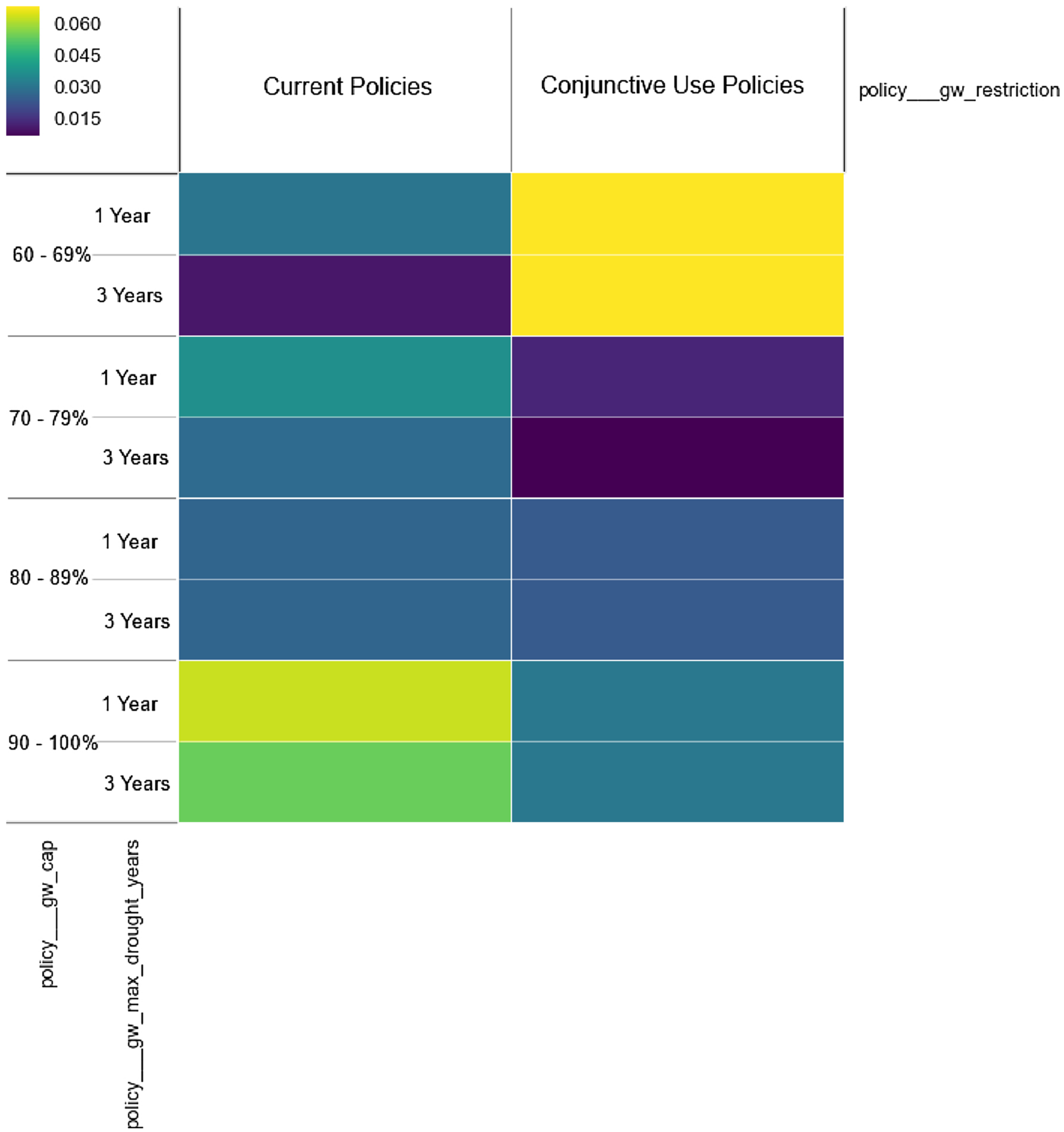


Fig. 17. Dimensional stack of current and conjunctive use policies under median Goulburn allocations with respect to considered farm groundwater use. A concentration of robust outcomes is found in scenarios wherein conjunctive use is allowed with groundwater use levels in line with what has been occurring historically. Without conjunctive use increased groundwater use (to 90–100 %) is necessary to achieve robust outcomes.

Each scenario explored was tied to three specific climate datasets which represent hypothetical shifts in aridity (i.e. “dry”, “usual”, and “wet” conditions). Limiting the climate scenarios to these three was done to keep the total number of possible scenario combinations to a manageable level as model runtime was a concern. An alternate approach is the use of multiple projections for each aridity scenario. This would more comprehensively address scenario uncertainty with respect to climate inputs as it would allow the influence of differing degrees of “dryness” to “wetness” to be explored.

Another source of uncertainty, rarely discussed, is *computational infrastructure uncertainty*. Here, we define computational infrastructure uncertainty as the uncertainty that arises in model interoperation and integration and application across various computational contexts. *Model technical uncertainty* (as defined in Refsgaard et al., 2007) is a related issue which we regard as being specific to the uncertainties that may arise from a model’s implementation. Computational infrastructure uncertainty is distinguishable from model technical uncertainty in that models that have identical implementations may yet exhibit different behaviour when applied on different infrastructure, such as operating systems (see for example, Bhandari Neupane et al., 2019), platforms (e.g. desktop vs supercomputer), interoperation of data via various means (e.g. local storages vs over a network) and formats and different

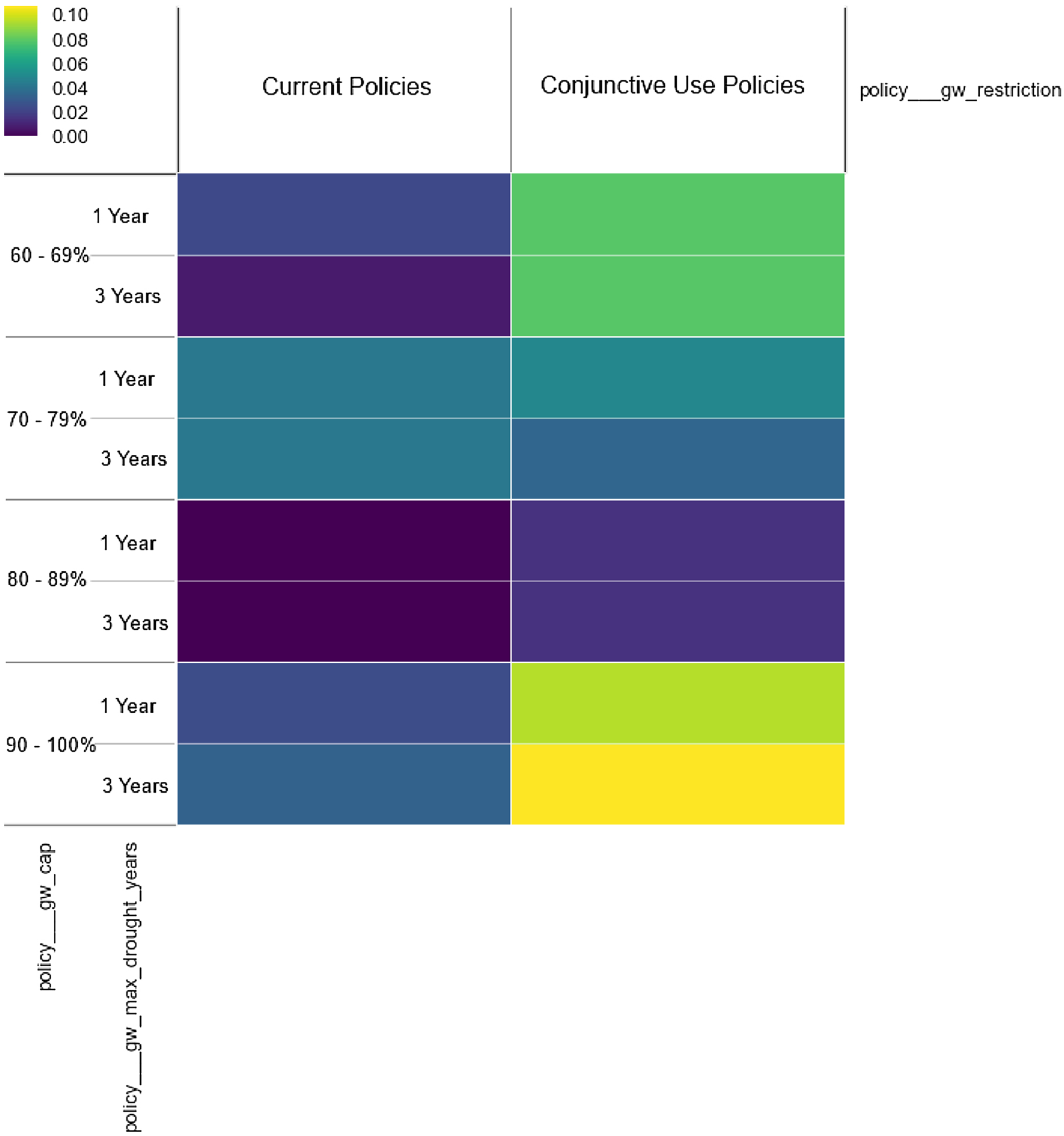


Fig. 18. Dimensional stack of current and conjunctive use policies under low Goulburn allocations. An increase to 90–100 % of groundwater use is necessary to increase the likelihood of having robust outcomes.

(programming) languages, or due to the use of different compilers (or different versions of the same compiler).

Computational infrastructure uncertainty was addressed by applying the model across various computational infrastructure including, but not limited to, both Windows and Linux operating systems, different (Fortran and C) compilers, and ensuring identical baseline results. A major error was discovered through this process due to the declaration of an uninitialized variable in the Fortran code, which was subsequently used in a later calculation. Depending on the compiler, the 'default' behaviour may be to initialize the undefined variable to 0.0 (and thereby will have no effect on the result) or may hold 'junk' values from its location in memory space, which subsequently propagate throughout the integrated model.

6. Limitations and future work

A difficult aspect to manage in the study was the determination of model scope. The collaborative modelling undertaken encompasses several disciplinary domains from economics and finance, bio-physical and social aspects and computational considerations. Limiting the scope to a manageable size often boiled down to making pragmatic compromises. Here we detail some known limitations and avenues for future work.

Climate data used in the modelling display little changes in evapotranspiration from scenario to scenario. Evapotranspiration is used as a reference value that informs crop water use in the model, and ultimately the frequency of irrigation throughout the growing season. Crop loss, due to extreme heat, pests, or other influences, are also not considered. Changing weather events due to a changing climate will also require the growing season to be shifted earlier or later (Prokopy et al., 2015; Wang et al., 2019), or timed to take advantage of forecasted rainfall, however these planting/harvest dates were constants in the modelling.

An avenue for further enhancement can be expanding the agricultural activities represented in the model. Dairying is a primary industry in the study area but is not explicitly represented in the model. Cropping was determined to be the common agricultural activity regardless of farm enterprise and so the decision was made to focus efforts towards representing farm behaviour in that context. While irrigation with groundwater was found to be an important aspect towards yielding robust outcomes, the model did not incorporate water quality aspects and so further in-depth investigation, particularly on salinity issues, are required.

Cropping enterprise profitability may reflect dairying (financial) performance and so beneficial model outcomes are used in the study to indicate beneficial outcomes for the catchment generally. As noted by stakeholders, the relationship between the enterprises is expected to be particularly poor in dry conditions as dairy farmers have the option of acquiring feedstock externally. Irrigators may also trade water (an activity not represented in the model) to cover shortfalls in water availability.

Averaging farm water orders across a 14-day time step means that the CIM does not represent high volume water orders that are released within a shorter time span. Consequently, the implications of high-volume water orders on ecological flow suitability indicator may not be adequately captured. On this note, the policy model currently includes environmental considerations based on legislated flow requirements but is not reactive to modelled streamflow/level or groundwater head, and effects on ecological indicators. Future work could then consider possible adaptive management processes in which water allocations and/or their releases are adjusted to meet environmental considerations.

Whilst uncertainty in the modelling is taken into account largely through the use of climate scenarios and sampling of parameter space, future work is envisaged to identify the most important sources of uncertainty in the modelling. In this endeavour, one way forward is to use a comprehensive qualitative and quantitative approach (e.g. Refsgaard et al., 2007), especially involving stakeholders and experts in helping to rank the criticality of the different sources as to their influence on model outcomes before embarking on a quantitative set of exercises. Nevertheless, we believe the work to this point is a valuable starting point for raising awareness and discussion amongst stakeholders as to opportunities for managing water more beneficially in the catchment.

One important social aspect not represented in the presented study is the cultural importance of the local flora and fauna and issues of cultural flows – the release of water to fulfil activities or conditions of cultural and social importance (Moggridge et al., 2019). Indeed, it has been increasingly acknowledged that water entitlements for cultural flows are not yet made a consistent part of Australian water management legislation, policies or guidelines despite being identified as a national priority area (Jackson et al., 2012; Williams et al., 2019). This limitation will be addressed in planned future work currently under discussion.

7. Conclusions on sustainable water management opportunities

This paper presents a component-based integrated environmental model developed to explore sustainable water management options within the Lower Campaspe sub-catchment of the Murray-Darling Basin. Stakeholder participation was critical to the model capabilities, as local stakeholders provided knowledge, feedback and data to assist in conceptualising the system. The participatory and model-based collaborative approach yields results that reveal opportunities to consider for achieving improved socio-environmental outcomes and water security, relative to the modelled historic baseline. Improvements at the farm level were found to be a prospective contribution towards this goal, as were farm (water) management and changes to governing policy rules. Specifically, conjunctive use of surface and groundwater resources and increased use of the latter was found to improve outcomes. Adoption of the most efficient irrigation systems considered (spray and pipe and riser) did not necessarily lead to desirable outcomes across all climate conditions. It then follows that simply improving irrigation water efficiency is not a sufficient course of action.

Declaration of Competing Interest

None to declare.

Acknowledgements

The primary author is supported through an Australian Government Research Training Program (AGRTP) Scholarship and a top-up scholarship from the ANU Hilda-John Endowment Fund. The work was carried out as part of the Murray-Darling Basin Authority's partnership with the National Centre for Groundwater Research and Training (NCGRT). The authors would like to thank representatives from the North Central Catchment Management Authority, the Australian Platypus Conservancy, Goulburn-Murray Water and the Victorian Department of Economic Development, Jobs, Transport and Resources for their valuable insights and assistance over the course of this work. The authors also gratefully thank Joseph Guillaume and Adrian Hindes for comments and suggestions prior to and throughout the review process, and the anonymous reviewers for their suggestions to improve the quality of this paper. The authors would also like to acknowledge and thank Wendy Merritt, Baihua Fu, and Yueqing Xie for early contributions to the body of work presented.

Appendix A. – Key Terms

Term	Description
Beneficial change/outcome	Positive change in the indicators Avg. Annual Profit, Ecology Index, GW Level Change, and Recreation Index compared to baseline (i.e. > 1.0). Negative changes to total surface and groundwater volume used (< = 1.0) and to Income Volatility (< 1.0) is preferred.
Desirable outcome	Scenarios which exhibit a beneficial change in system state across all indicators
Robust outcome	Scenarios which exhibit desirable outcomes across all climate conditions

Appendix B. – Farm Model Details

The farm model optimised irrigated area and source of water through linear programming, conducted with the OptLang python package (Jensen et al., 2017) in the form of:

$$\text{maximize}\,\pi,\ \pi = \sum_{i=1\ldots3} A_i(R_i - C_i)$$

s.t.

$$\begin{aligned}&\text{Rule}1.0 \le A_i \le L_i, L_i = A_T \,\text{if field is dryland only}\\&\text{Rule 2.} \sum A = A_T\\&\text{Rule 3.}0 \le A_{sw} + A_{gw} \le A_f\end{aligned} \tag{1}$$

where A indicates the area to be serviced by each water source, and where $i\,\varepsilon\,\{1,2,3\}$ represents the water source to be used: *sw* (surface water), *gw* (groundwater), or only rainfall (dryland, *nw*). R is the gross revenue per hectare that can be expected with the crop (sown for the season), irrigation system, and water source(s), C represents the costs incurred for the same. A is limited (L_i) by the available water resources and expected crop water demands. In the case of dryland, the total field area can be used (A_T). The values for L are calculated as $V_i/W_{c,ha}$ as in the volume of available water from a particular water source (V_i) divided by the expected seasonal (per hectare) crop water requirements ($W_{c,ha}$). In future modelling, irrigation areas could be informed based on recent Landsat imagery.

At the start of the season, only Rules 1 and 2 are used to determine the irrigated area for the growing season. The initially optimised area (A_f) is then locked for the rest of the growing season and is used to determine the proportional amount of water to be applied with surface and groundwater, such that in subsequent time steps Rule 3 is included in the formulation.

Dryland cropping is assumed to occur on non-irrigated areas, the area for which is calculated as $A_{nw} = A_T - (A_{gw} + A_{sw})$. Costs included in the calculation include the variable costs for the crop sown, maintenance of irrigation and pumping systems (if applicable), costs associated with licensing, water ordering, and pumping costs. The sum of these gives the dollar profit/income (π) for the farm/zone. In this manner the farm costs, expected yields and profit, and estimated crop water requirements play a role in scheduling irrigation events. The estimated total profit is necessarily an approximation. Profit after harvest is calculated directly from crop yield, as detailed below.

Irrigation Scheduling

Farmers will irrigate, ideally, when crops require additional water. Determining when these irrigations occur is referred to as irrigation scheduling. In this model it is assumed that farmers are monitoring soil moisture levels and have access to weather data, specifically pan evapotranspiration (ET_0). Soil water deficit (*SWD*) is a cumulative indicator of how dry soils can become before additional water is required to be applied to avoid crop losses. Soil water deficit worsens by subtracting crop evapotranspiration (ET_c) which is calculated by applying a scaling crop coefficient K_c (i.e. $ET_c = ET_0 \cdot K_c$), with each crop type having a corresponding K_c value. Once *SWD* reaches a refill point – commonly referred to as the Net Irrigation Depth (*NID*) – an irrigation event is scheduled and *SWD* reduced by the effective water applied. This approach is commonly applied on-farm and examples can be found in publications from State governments (see for example, Hughes, 1999; Qassim and Ashcroft, 2002).

The *NID* value itself is calculated by multiplying the effective root zone at a point in time ($D_{rz,t}$) with the possible Readily Available Water (*RAW*) for the given soil type. The effective root zone is the depth at which the crop gets much of its water via its roots and is dependent on soil type and crop properties (Baker and Ahern, 1989). Here, D_{rz} is assumed to be 55 % of root depth for the given crop type (as in Qassim and Ashcroft, 2002) where relevant information regarding root depths throughout the season and crop growth stages could not be obtained, and as such acts as a constant. If sufficient data were available, the alternative approach would be to calculate it as:

$$D_{rz,t} = RD_t \cdot ERD_t \tag{2}$$

where RD_t is the root depth for the stage of growth at time t, and ERD_t the effective root depth coefficient for the crop type (Maihol et al. in Itier et al., 1996). Nominal values for these parameters were obtained from FAO guidelines (Allen et al., 1998).

The Net Irrigation Depth can then be calculated as

$$NID_t = D_{rz,t} \cdot RAW,\ \ RAW = TAW \cdot p \tag{3}$$

Here, TAW corresponds to the Total Amount of Water a soil can hold, and p represents the crop soil water depletion fraction (as in Qassim and Ashcroft, 2002). The values used for TAW is discussed in the next section on soils. The soil water deficit at a point in time can be calculated as:

$$SWD_t = \min\{SWD_{t-1} - (ET_{c,t} - E_t + IW_t),\ 0\} \tag{4}$$

where

$E_t = \begin{cases} \sum_{t=1}^{n} P_t \text{ if month is June to August} \\ \sum_{t=1}^{n} \max\{P_t - 5,\ 0\} \text{ if other months} \end{cases}$ and represents the total effective rainfall that occurred within the (two week) time step (i.e. all winter rainfall is assumed to be effective rainfall)

IW_t denotes the irrigation water applied at the time step. No water may be applied in the time step in which case this coefficient will be 0.

$ET_{c,t}$ is the sum of crop evapotranspiration (ET_c) that occurred within the time step.

It should be noted here that the intention of this particular model is not to have an accurate representation of effective rainfall or water recharge/drainage processes. The model is, as mentioned above, based on the published advice for irrigators in Victoria and so represents the assumed behaviour of the irrigation process. Irrigation occurs to refill SWD once it reaches (or goes beyond) NID, and these are both represented as negative values (or else 0). The base volume of irrigation water (IW_b) to be applied is taken to be equal to SWD_t when it reaches this refill point. This volume is then adjusted to reflect the efficiency of the implemented irrigation system (IE) to arrive at the amount of irrigation water to be applied (IW_t).

$$IW_t = \frac{IW_{b,t}}{IE},\ IW_{b,t} = \max\{\mathrm{abs}(SWD_t),\ 0\}\ SWD_t \le NID_t \tag{5}$$

Irrigation efficiency refers to the percentage of water that reaches the crop root zone, allowing the crop ease of access to water. Water applied with less efficient irrigation systems are said to be "lost" for the purpose of contributing to crop growth. Therefore, more water is required if applied with irrigation systems of lesser efficiency for an equivalent effect on SWD. Irrigation systems considered in this study include gravity, pipe and riser, and spray with IE ratings of 0.5, 0.7 and 0.8 respectively. Returning, finally, to the proportional use of irrigation water, $A_{gw,sw}$ (the areas watered by ground and surface water respectively) are each then divided by A_f (total irrigated area) and constrained by the volume of water available. The optimal mix of water sources to use is then indicated by:

$$\min\left\{IW_t\left(\frac{A_{ws}}{A_f}\right),\ \frac{V/IW_t}{A_f}\right\} \tag{6}$$

where V is the volume of available water (in ML), and A_{ws} is the fixed area irrigated by the given water source (e.g. A_{sw} or A_{gw}). This process (depicted in Fig. 19) is repeated until the crop is harvested at the end of the growing season.

Estimated costs for each irrigation system were corroborated by a senior irrigation specialist with EcoDev (Maskey, 2016). The simplest irrigation system is "gravity" which relies on, as the name suggests, gravity to flood an area with water. Pipe and riser systems are similar in that it also 'floods' a field but instead uses a pressurized pipe system to transport water, increasing costs. Spray

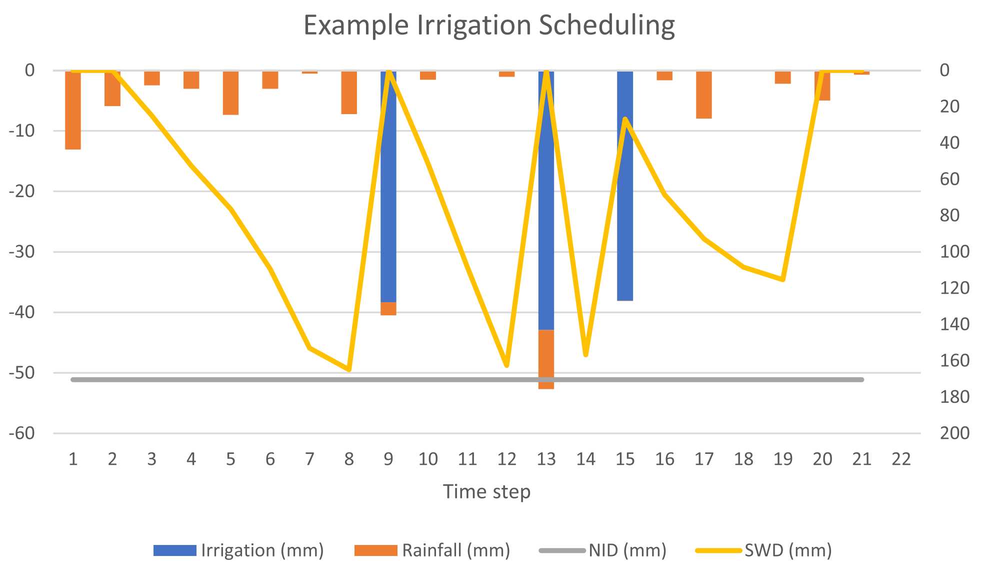


Fig. 19. Example depiction of irrigation scheduling.

Table 7
Implementation costs for each irrigation system considered and their efficiencies.

Variable	Irrigation	Nominal value (and range)	Description	Reference(s)
Cost ($/Ha)	Dryland Gravity Pipe and Riser Spray	$0 (relies on rainfall) $2000 ($2000 - $2500) $2500 ($2000 - $3000) $2500 ($2500 - $3500)	Cost of implementation (or replacement) per hectare of irrigated area in AUD ($/Ha)	(Laffan and Smith, 2015)
Irrigation Efficiency (%)	Gravity Pipe and Riser Spray	50 % (50 %–90 %) 70 % (60 %–90 %) 80 % (70 %–90 %)	Expected efficiency for each irrigation system	Finger and Morris (2005) Tennakoon et al. (2013)

is the most efficient but is also the most expensive to install and operate due to the fuel costs necessary to generate the pressure needed to move and apply water.

The upper limit for all irrigation systems was set to 90 % - which is possible for all systems depending on soil type, system set up and configuration, and additional work conducted to make the field more amenable for the chosen irrigation type (Finger and Morris, 2005). For example, the field could be laser graded to ensure a more consistent application of gravity fed irrigation water. Indeed, 77 % of those survey respondents reported having undertaken additional improvements to gravity irrigation such as laser grading and tail-water reuse (from excess water reaching the bottom of the field). Gravity irrigation was then modelled as being 50 % and 90 % efficient based on this information with the base efficiency set at 50 % representing the “usual” case (Table 7).

Soils

An early iteration of the model used the dominant soil type found in the Lower Campaspe area (sandy loam, see Fig. 21) as a representative surrogate. Stakeholders indicated that this approach may not adequately represent the importance of soil type and health in agricultural enterprises due to the diversity found at the smaller (zonal) scales. Stakeholders further indicated that farmers with lighter soils on their lands may find spray irrigation more attractive as light soils are not capable of holding as much water as heavy soils – the TAW value is comparatively less, influencing irrigation scheduling. Equally true is that farmers with heavier soils may not see a benefit from a move to spray. To reflect this, only farm zones that were identified as having light soils were modelled to have the option of changing irrigation systems to spray irrigation.

To better represent the conditions which impact irrigation scheduling and choice of irrigation system it is then necessary to represent the range of soil textures within the model farm zones. To achieve this, published TAW value ranges for various soil types (Allen et al., 1998; Qassim and Ashcroft, 2002) were used in conjunction with a soil map of the Campaspe catchment to create weighted zonal values (see Fig. 20). These values were based on the proportional area of soil types found within a zone. A weighted average median value was used as the nominal “best guess” value, with the weighted minimum and maximum values indicating the possible value bounds. The soil map was kindly provided by EcoDev, a Victorian State Government department.

Pumping

Pumping water for irrigation typically represents the largest operational costs for a farm (DEPI 2014a). Seasonal pumping costs were considered as this may vary depending on climate conditions and allocated water availability. This cost can itself vary depending on the type of pump and its configuration. Extracting groundwater decreases the groundwater level, thereby increasing pumping costs due to the greater distance and pressure required.

Pumping systems were simplified into two categories, indicating whether they are for shallow or deep bore pumping. The former represents pumping from an irrigation channel or shallow aquifer, with the latter used to represent groundwater pumping at a depth of 20 m or more. Stakeholders indicated that a mix of diesel and electric pumps are used in the study area (60 % and 40 % respectively). Electric pumping costs can range from 16.5c to 32.8c per kilowatt (kW), while electricity plans with a flat rate of 27c/kW can also be arranged (Bob Knowles, 8 Jan 2018, pers. comm). Diesel fuel was assumed to have a cost of $1.20 per litre, with a fuel volume per kW of 0.25 (Robinson, 2002), resolving to 30c/kW. A weighted average of these values was used to represent the mixed (60 %/40 %) use of both diesel and electric pumps (28.5c/kW). The kW cost of pumping is likely to change over time but was modelled as a constant under the assumption that the cost of upgrading infrastructure is cost prohibitive within the modelled time frame.

Installation of bores to access groundwater incur significant capital costs, ranging from $18,000 to $70,000 for a shallow bore and $90,000 to $320,000 for deep bores (Robinson, 2002). It is assumed here that such infrastructure is already in place and so no initial capital costs are considered. The average annual maintenance costs are included in the modelling, however, and are taken to be 5 % of capital costs every 5 years and 20 % every 15 years (minor and major maintenance respectively). Nominal values for these (shallow and deep bore) capital costs were $18,000 and $235,000 (taken from Robinson, 2002). The number and location of groundwater bores for each farm zone are indicated in Fig. 22.

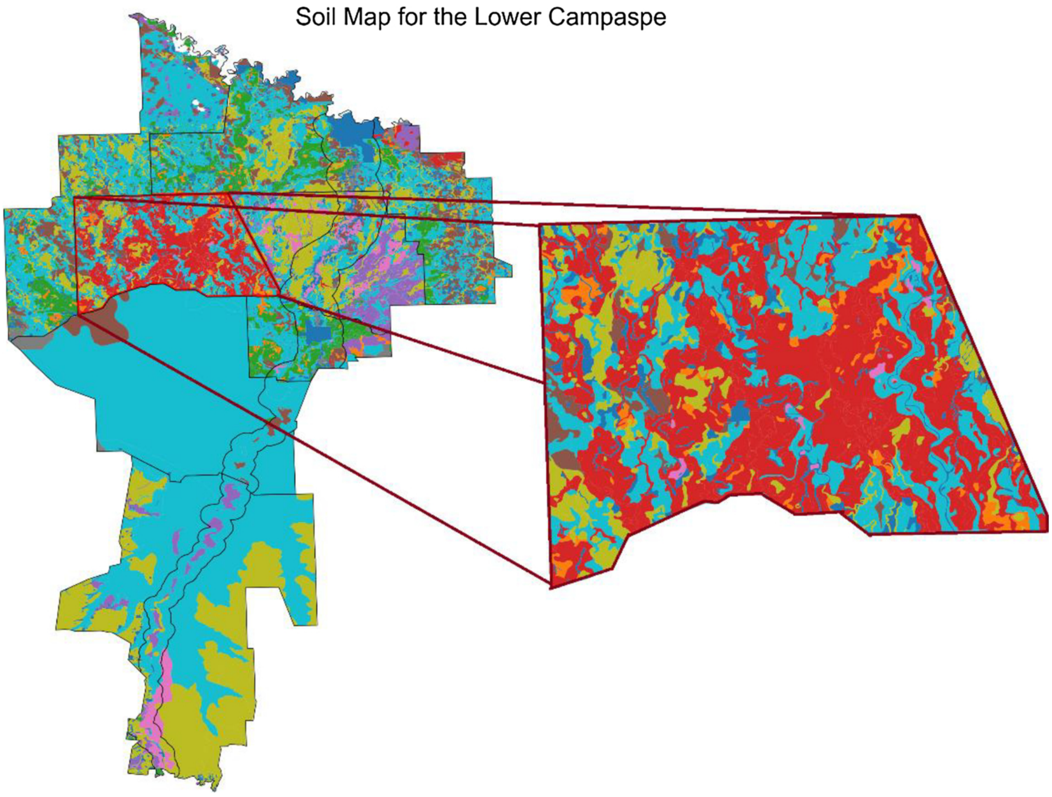


Fig. 20. Soil texture map for the Lower Campaspe farm zones with a zoomed close-up of Zone 7 showing the heterogeneity of surface soils. Soil map was obtained via EcoDev.

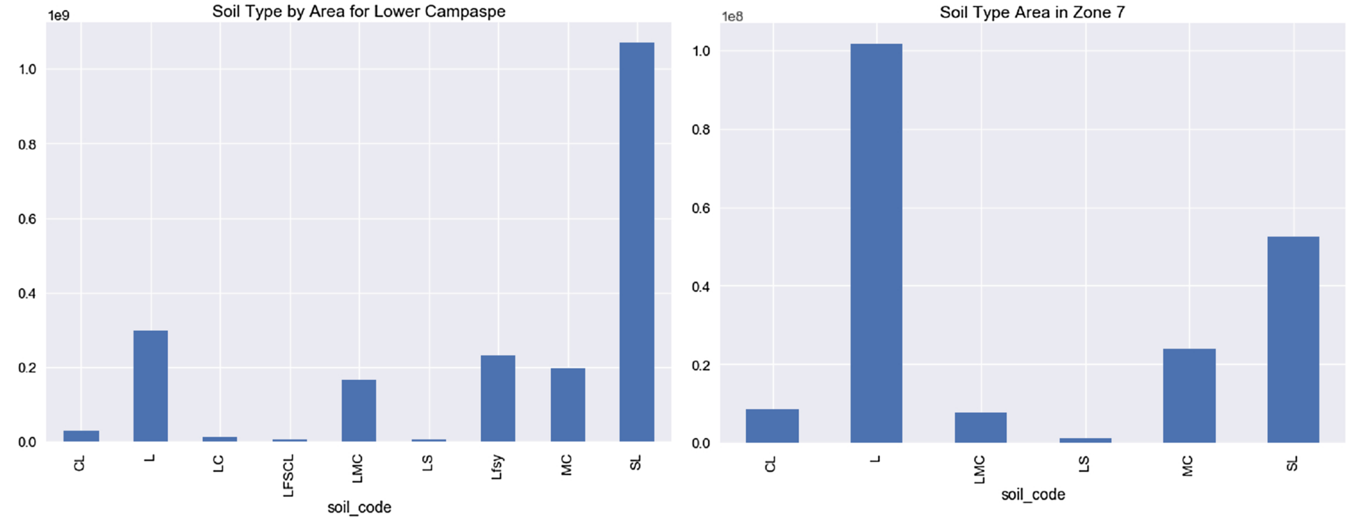


Fig. 21. Area for each soil type for the Lower Campaspe as a whole, and an example Zone (Zone 7 in the above). Sandy loam (soil code "SL") dominates the catchment area, but variations can be seen within specific Zones, precluding the use of a single representative soil type. A weighted TAW value based on the proportional area for each soil type found in the Zone was used instead.

Pumping costs throughout the irrigation season is then calculated as

$$C_{PML} = (C_{fuel}C_{kW})\cdot h_{ML} \tag{7}$$

Read as the cost per pumped ML being equal to the cost of fuel (per litre) by the cost per kilowatt (kW) hour, multiplied by the hours necessary to pump a megaliter of water. The total pumping cost can then be determined simply as

$$C_{pump} = IW\cdot C_{PML} \tag{8}$$

In the presented modelling C_{kW} is calculated in advance (28.5c/kW, as detailed earlier), however it can be calculated directly if needed and the necessary information on the pumping system is available. The calculation takes the form of:

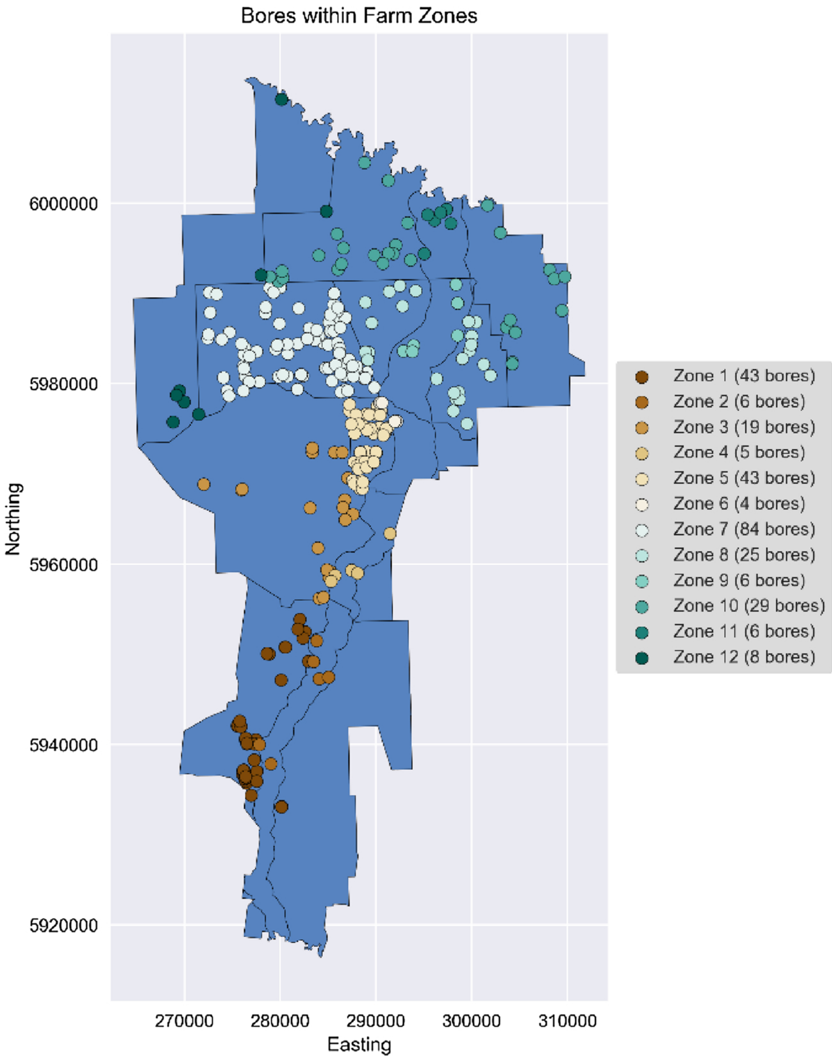


Fig. 22. Number of bores within each farm zone.

$$PUMP_{fuel} = P(kW) \cdot F \tag{9}$$

where

$$F = 0.25$$

$$P(kW) = \frac{(H \cdot f_c)}{(102 \cdot E_p)^* D}$$

The volume of diesel fuel ($PUMP_{fuel}$) is based on the required energy ($P(kW)$) and number of litres required to produce 1 kW ($F = 0.25$, as given in Robinson, 2002). The amount of kilowatt energy required is dependent on the total head pressure (in metres, H), the flow capacity of the irrigation system in use (f_c), pump efficiency (E_p) and derating factor (D).

Head pressure (H) is defined as the sum of 1) the discharge pressure head, 2) the friction head of water flow (i.e. friction loss) and 3) the height between the source of water and the discharge level (Robinson, 2002). The typical total pumping head for a given irrigation system supplied with surface water is taken from Smith (2015). These values range from 10 m head pressure for gravity to 60 m for spray irrigation. Although the values in Smith (2015) are intended for irrigators in New South Wales, the indicated pumping costs were within the value range suggested by a local irrigation specialist (Maskey, 2016) for a type of flood irrigation (pipe and riser, $8-15), and spray irrigation system (centre pivot, $30-50). Typical head and cost ranges are shown in Table 8. Head pressure is multiplied by the flow capacity of the irrigation system (f_c) the value of which is taken from literature regarding a farm in the study area, given as 138.88 litres/second, or 12 ML/Day (DEDJTR, 2015). The literature-derived values assume that the pumping system can operate at the desired head pressure and flow rate and that the pump itself is in good condition.

Pump efficiency (E_p) is the percent energy efficiency of the pump, representing the amount of energy imparted on the water. This

Table 8
Typical total head for each irrigation system, adapted from Smith (2015) and adjusted with input from Maskey (2016). The typical pumping costs indicated on the right-hand column were used to evaluate the pumping cost model.

Irrigation System	Total Head (m)	Pumping Cost ($/ML) @ $1.20/L diesel fuel
Gravity	10 - 15	8 - 15
Pipe and Riser	10 - 15	8 - 15
Spray	25 - 35	30 - 60

value is multiplied by a constant of 102 to convert unit of pressure (kPa) into metres (given in Faour 2001 in Robinson, 2002). The derating factor (D) accounts for efficiency losses between the total amount of energy required and the energy required at the pump shaft. The derating factor is said to be 0.75 for diesel pumps (Faour 2001 in Robinson, 2002). As pumping efficiencies may vary model evaluation was conducted with Ep set to a conservative value of 0.7, the suggested value to use when the pump configuration and efficiency are unknown (Vellotti and Kalogerinis, 2013). Other efficiency losses which influence total pumping head are not explicitly considered but are accounted for through the use of this conservative pump efficiency value (as in Vellotti and Kalogerinis, 2013).

Groundwater pumping costs may fluctuate due to the changes in height distance between water source and discharge point resulting in an increase in total pumping head. Such a decrease in water levels necessitates increased amounts of energy (and thus fuel) to pump water from the increased distance. To account for decreasing water levels the depth of groundwater is added to the head values given by Smith (2015) to allow consideration of the effect of fluctuating groundwater levels.

The time taken to pump a single ML of water (h_{ML}) is then determined by dividing the number of litres in a ML by the flow rate (Q, in litres per second). This resolves to the number of seconds required to pump 1 ML. Dividing this by the number of seconds in an hour (3600 or 60^2) results in the hour(s) required to pump 1 ML.

$$h_{ML} = \left(\frac{10^6}{Q}\right)/60^2, \ 10^6{=}1\mathrm{ML} \tag{10}$$

Crops

The crops represented include three cereal crops of wheat, barley and canola. These are applied as a three-year rotation; i.e. cultivating wheat one year, barley the next, and finally canola after which the rotation is repeated. An earlier version included tomato however this crop was removed as its widespread cultivation was described as unrealistic and highly improbable by stakeholders, as were horticultural crops in general.

Determining Seasonal Profit

Values for the expected revenue (R) and associated production costs (C) for each crop and irrigation type used in Eq. (1) are taken from various grey literature sources to determine the optimal irrigation volume throughout the growing season (or the irrigated area in the case of the first time step). Once the growing season has ended, however, a modified French-Schultz equation (Oliver et al., 2009; Whitbread and Hancock, 2008) is used to obtain the final crop yield (Table 9).

$$Y = \frac{((SSM + GSR + IW) - V)\cdot CWUE}{1000} \tag{11}$$

The modified French-Schultz equation above takes into account the stored soil water at the start of season (SSM), the effective rainfall that occurred during the growing season (GSR, which is the sum of E_t from Eq. (4)), the sum of any irrigation water applied (IW, which will be 0 for dryland crops), and the crop evaporation coefficient (V) which represents the required rainfall before a crop will yield. These are then adjusted by a Crop Water Use Efficiency Index ($CWUE$) to arrive at the per hectare crop yield (Y). The resulting value is then converted to tonnes per hectare by dividing by 1000.

Values for the French-Schultz equation were initially taken from published FAO guidelines (Allen et al., 1998), with SSM assumed to be 30 % of rainfall that occurred over February to April. These were subsequently calibrated against historic (per farm average) crop yield data for the North Central region obtained from the Australian Bureau of Agricultural and Resource Economics (ABARES).

Table 9
Growth stages for Winter Wheat, the length in days for each stage, and assumed planting date in month and day for each season.

Growth Stage	Duration (in days)	Crop Coefficient	Depletion Fraction	Source
Initial	30	0.4	0.6	Allen et al. (1998)
Development	140	0.4	0.6	
Mid-season	40	0.9	0.6	
Late	30	0.25	0.9	
Season Length	240 Days			
Assumed Plant Date	05-25 (MM-DD)			DEDJTR (2015b)

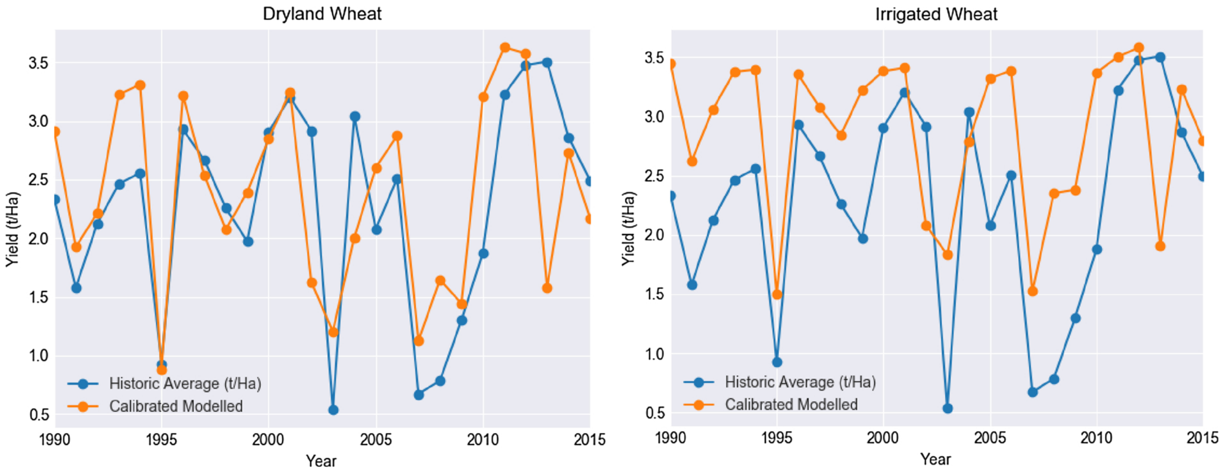


Fig. 23. Calibrated crop model results for (left to right) dryland and irrigated wheat. Long-term modelled yield was 2.37 t/ha compared to the historic average of 2.27 t/ha for dryland. The irrigated crop yields were adjusted to return higher yields compared to dryland production as the historic observations are farm averages. The Campaspe catchment is said to produce higher yields than regional averages. Results were deemed acceptable by stakeholders.

The North Central region represents an area significantly larger than the Lower Campaspe sub-catchment, however it is the most (spatially) relevant dataset available.

Calibration used the L-BFGS-B function minimization routine – implemented in the `scipy` Python package (Jones et al., 2001) – to reduce Root Square Mean Error (RMSE). This process achieved results with an overall RMSE of 0.68 and a long-term average yield comparable to what has occurred historically; 2.37 t/ha compared to the historic seasonal average of 2.27 t/ha for dryland wheat. The parameters for irrigated crops were adjusted to return higher yields (as is usual for crops under irrigation). The modelled results with calibrated values for both dryland and irrigation yields were deemed to be acceptable and reasonable by stakeholders. Further example calibration results are shown in Fig. 23.

The seasonal profit is calculated in the same manner as used to optimize water usage (see Eq. (1)), albeit with yield values (Y) replaced with those calculated by the French-Schultz equation, and the costs now representing the variable and fixed costs associated with the cropping enterprise, but also the costs accrued throughout the growing season, such as cost of pumping and water access fees. From here, the per hectare profit for a growing season (π_{ha}) and profit per ML of water used (π_W) can be determined as:

$$\pi_{ha} = \frac{\pi_T}{A_T} \tag{12}$$

$$\pi_w = \frac{\pi_T}{W_T} \tag{13}$$

where W_T is the total water volume applied for the season and A_T is the earlier defined total field area. Rather than model a global crop market, it is assumed here that the harvest for each year is sold.

Appendix C. Representing Goulburn Allocations

See Table 10 Fig. 24.

Table 10
Goulburn allocation scenarios and relationship with Campaspe surface water allocations.

Scenario	Equation
High	1.2525x + 48.541
Median	1.4005x + 5.3381
Low	max(0, (1.0116x – 3.2019))

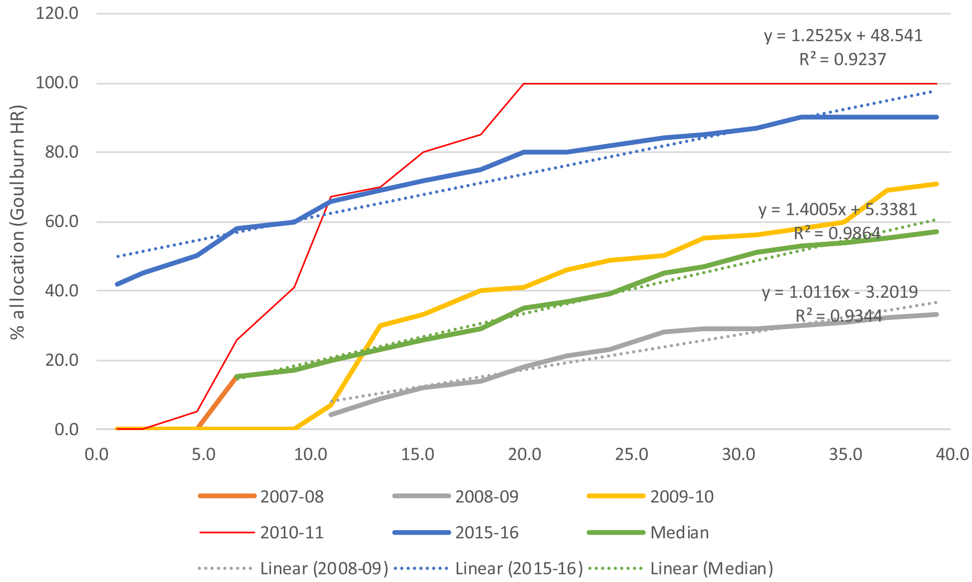


Fig. 24. Linear relationships between Campaspe allocations and allocations in the Goulburn catchment.

Appendix D. Calibration periods for IHACRES rainfall-runoff model

See Table 11.

Table 11
Time span of each period used to calibrate the IHACRES model.

Label	Historic Time Span	Daily Time Step Index
Pre-drought	1981-01-01 to 1995-03-13	1 – 5185
Start-drought	1995-03-14 to 2000-06-13	5186 – 7104
Early-drought	2000-06-14 to 2003-04-20	7105 – 8145
Mid-drought	2003-04-21 to 2005-11-21	8146 – 9091
Late-drought	2005-11-22 to 2010-06-20	9092 – 10763
Post-drought	2010-06-21 to 2016-12–31	10764 - End

Appendix E. Ecology and recreation index model detail

Platypus Index

Indicators developed for platypus are based on recommendations outlined in the Environmental Water Management Plan set out by the North Central Catchment Management Authority (North Central CMA, 2014). Following these recommendations streamflow is to be maintained to at least 10 ML/day in the summer, and 50 ML/day in winter to allow for movement and food supply (macro-invertebrate productivity and dispersal). Water releases in the summer and autumn ("freshes") are necessary to maintain food supply and, in the autumn, aid in the movement of platypus young. An additional "burrow flooding" index is used to represent prolonged high flow events during the nursing season which may flood platypus burrows and drown the young.

Fish Index

Fish represented in the ecology model are categorised by their lifespans (long and short-lived fish). Examples of long-lived fish include the Murray Cod and the Golden Perch, with their oldest recorded estimated ages being 48 years (Anderson et al., 1992) and 26 years (Mallen-Cooper and Stuart, 2003) respectively. Short-lived fish include the rainbowfish and carp gudgeons which have typical lifespans of 2–3 years. Species of long-lived fish have individual spawning preferences, e.g. Murray Cod is considered to spawn in low-flow conditions whereas Golden Perch prefer high-flow flood conditions. In contrast, short-lived fish generally prefer low-flow conditions (Ralph et al., 2010). Fish indicators represent suitable conditions during breeding and nesting seasons (low flow in summer and winter, e.g. 500–1000 ML/day), spring freshes to trigger spawning events for long-lived fish (at least 500 ML/day for two days), and summer and autumn freshes for dispersal for long-lived fish (e.g. 50 ML/day for at least two days).

River Red Gum (tree) Index

River red gums play vital roles in the maintenance of the aquatic and riparian ecosystem in the Campaspe River. The model provides indications of suitability for maintenance and regeneration (e.g. promotion of new growth) of River Red Gums which incorporates both groundwater and surface water flow regimes. Suitable groundwater conditions were taken from Roberts and Marston (2011) with established trees preferring the groundwater table to sit between 2 and 6 m below the ground. A linear reduction from 6 m to 12 m (where 12 m or more will produce a zero indicator value) is used in the model. Younger trees are modelled to prefer groundwater depths between 0.5 and 1 m, after which the index linearly deteriorates towards 2 m at which point the indicator gives a score of 0.

Recreation Index

Lake Eppalock is a popular destination for water-related recreational activities including boating, water skiing, and wind surfing, with associated economic benefits (City of Greater Bendigo, 2009). Falling water levels during the millennium drought and in recent years have led to increased concern for the continued viability of the dam for recreational purposes (ABC News, 2015; City of Greater Bendigo, 2009).

The volume of water in the dam is used to indicate the perceived suitability of the dam for recreational purposes where sufficiently high dam levels (> 65 % dam capacity) allows full enjoyment of recreational activities, whereas lower dam levels impede them. Collision with debris and fallen trees is a risk when the dam falls to 30 % capacity. Issues of crowding can occur at lower volumes as it equates to lower surface area for recreational use.

Appendix F. – Parameter Covariance Analysis

Fig. 25.

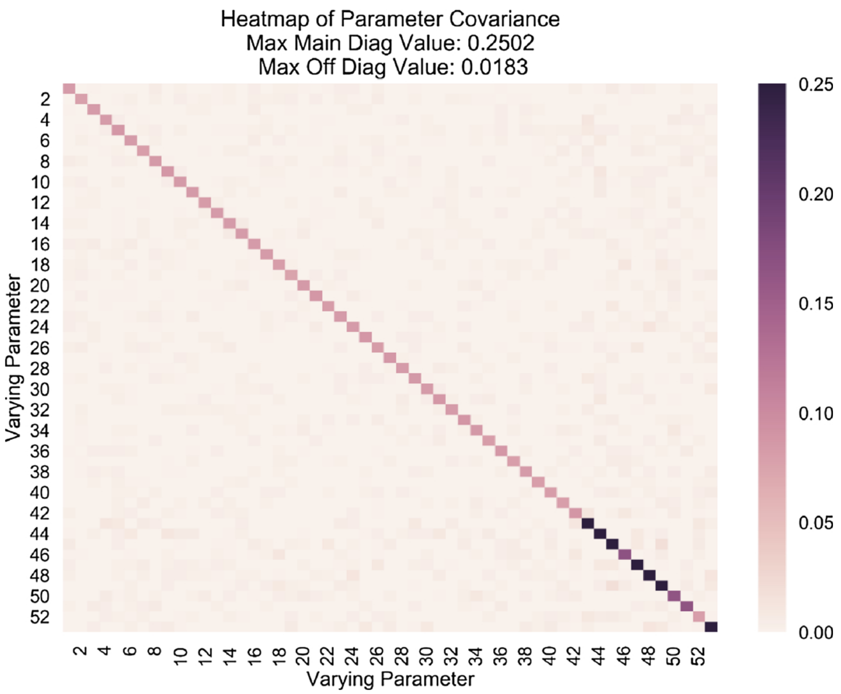


Fig. 25. Covariance matrix indicating (statistical) independence of model inputs. The number of scenarios to run (896) for each climate scenario was selected as a compromise between the time taken to run the model and the independence of inputs.

References

ABC News, 2015. Protesters Fear Falling Lake Eppalock Levels. [WWW Document]. ABC News. URL https://www.abc.net.au/news/2015-01-26/protesters-fear-falling-lake-eppalock-levels/6045852 (accessed 10.28.19). .

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Chapter 8 - ETc under soil water stress conditions. Crop Evapotranspiration – Guidelines for Computing Crop Water Requirements, FAO Irrigation and Drainage Paper. Food and Agriculture Organization of the United Nations (FAO).

Alston, M., Whittenbury, K., 2011. Climate change and water policy in Australia's irrigation areas: a lost opportunity for a partnership model of governance. Env. Polit. 20, 899–917. https://doi.org/10.1080/09644016.2011.617175.

Anderson, J.R., Morison, A.K., Ray, D.J., 1992. Age and growth of Murray Cod, Maccullochella peelii (Perciformes: Percichthyidae), in the Lower Murray-Darling Basin, Australia, from thin-sectioned Otoliths. Mar. Freshwater Res. 43, 983–1013. https://doi.org/10.1071/mf9920983.

Austin, E.K., Handley, T., Kiem, A.S., Rich, J.L., Lewin, T.J., Askland, H.H., Askarimarnani, S.S., Perkins, D.A., Kelly, B.J., 2018. Drought-related stress among farmers: findings from the australian rural mental health study. Med. J. Aust. 209, 159–165. https://doi.org/10.5694/mja17.01200.

Australian Department of Agriculture, 2019. AGSURF Data - Australian Bureau of Agricultural and Resource Economics and Sciences. [WWW Document]. AgSurf Data. URL http://apps.agriculture.gov.au/agsurf/agsurf.asp (accessed 7.10.19). .

Badham, J., Elsawah, S., Guillaume, J.H.A., Hamilton, S.H., Hunt, R.J., Jakeman, A.J., Pierce, S.A., Snow, V.O., Babbar-Sebens, M., Fu, B., Gober, P., Hill, M.C., Iwanaga, T., Loucks, D.P., Merritt, W.S., Peckham, S.D., Richmond, A.K., Zare, F., Ames, D., Bammer, G., 2019. Effective modeling for Integrated Water Resource Management: a guide to contextual practices by phases and steps and future opportunities. Environ. Model. Softw. 116, 40–56. https://doi.org/10.1016/j.envsoft.2019.02.013.

Baker, D.E., Ahern, C.R., 1989. Estimates of effective rooting depth for predicting available water capacity of Burdekin soils, Queensland. Aust. J. Soil Res (Australia). (Queensland D. of P.I.

Bakker, M., Post, V., Langevin, C.D., Hughes, J.D., White, J.T., Starn, J.J., Fienen, M.N., 2016. Scripting MODFLOW Model Development Using Python and FloPy. Groundwater 54, 733–739.

Beverly, C., Hocking, M., 2014. DEPI Northern Victoria Groundwater Model Conceptualisation and Calibration (No. 104220. Department of Environment and Primary Industries with Hocking et al. Pty Ltd, Rutherglen, Victoria, Australia.

Bhandari Neupane, J., Neupane, R.P., Luo, Y., Yoshida, W.Y., Sun, R., Williams, P.G., 2019. Characterization of Leptazolines A–D, Polar Oxazolines from the Cyanobacterium Leptolyngbya sp., Reveals a Glitch with the "Willoughby–Hoye" Scripts for Calculating NMR Chemical Shifts. Org. Lett. 21, 8449–8453. https://doi.org/10.1021/acs.orglett.9b03216.

Bowler, M., 2015. Report for the Department of Agriculture and Water Resources - Goulburn-murray Water Connections Project Stage 2 - Mid Term Review (No. 08/08200/07). GHD, Melbourne, Victoria, Australia 2015.

Bruyn, L.A.L., 2019. Learning opportunities: understanding farmers' soil testing practice through workshop activities to improve extension support for soil health management. Soil Use Manag. 35, 128–140. https://doi.org/10.1111/sum.12466.

Chiew, F.H.S., McMahon, T.A., O'Neill, I.C., 1992. Estimating groundwater recharge using an integrated surface and groundwater modelling approach. J. Hydrol. 131, 151–186. https://doi.org/10.1016/0022-1694(92)90216-I.

Chiew, F.H.S., McMahon, T.A., Dudding, M., Brinkley, A.J., 1995. Technical and economic evaluation of the conjunctive use of surface and groundwater in the Campaspe Valley, north-central Victoria. Australia. Water Resour Manage 9, 251–275. https://doi.org/10.1007/BF00872487.

City of Greater Bendigo, 2009. Rural Areas Strategy. City of Greater Bendigo - Strategy Unit.

CSIRO, 2017. Climate Change in Australia [WWW Document]. URL https://www.climatechangeinaustralia.gov.au/en/ (accessed 12.13.17). .

DEDJTR, 2015a. Irrigation Supply System Modernisation Enables Farm Irrigation Upgrades. Department of Economic Development, Jobs, Transport and Resources (DEDJTR), State of Victoria, Tatura, Victoria.

DEDJTR, 2015b. Victorian Winter Crop Summary 2015. Department of Economic Development, Jobs, Transport and Resources (DEDJTR), State of Victoria, Tatura, Victoria, Victoria.

Department of Economic Development, J, 2017. Farm Management [WWW Document]. URL http://agriculture.vic.gov.au/agriculture/farm-management (accessed 8.8.19).. .

DSE, 2012a. How carryover works on the Murray. Goulburn & Campaspe: Carryover Review Committee fact sheet 1.

DSE, 2012b. Limiting Carryover to Your Water Share Volume: Carryover Review Committee Fact Sheet. pp. 3.

DSE, 2012c. Victorian Aquifer Framework: Updates for Seamless Mapping of Aquifer Surfaces.

Dunford, R., Harrison, P.A., Rounsevell, M.D.A., 2015. Exploring scenario and model uncertainty in cross-sectoral integrated assessment approaches to climate change impacts. Clim. Change 132, 417–432. https://doi.org/10.1007/s10584-014-1211-3.

EcoDev, 2015. Victorian Winter Crop Summary 2015.

Elsawah, S., Filatova, T., Jakeman, A.J., Kettner, A.J., Zellner, M.L., Athanasiadis, I.N., Hamilton, S.H., Axtell, R.L., Brown, D.G., Gilligan, J.M., Janssen, M.A., Robinson, D.T., Rozenberg, J., Ullah, I.I.T., Lade, S.J., 2020a. Eight Grand Challenges in Socio-environmental Systems Modeling. pp. 16226. https://doi.org/10.18174/sesmo.2020a16226. 16226,1 2.

Finger, L., Morris, M., 2005. Water Management Options: Assisting Irrigators With Stream Flow Management Plan Implementation. Department of Primary Industries (DPI).

Fu, B., 2017. Modelling Ecological Indices to Evaluate Suitability of Surface Water and Groundwater Regimes in the Lower Campaspe: Draft Technical Report Submitted to the MDBA-NCGRT Partnership. Available on request. .

Gamo, M., Shinoda, M., Maeda, T., 2013. Classification of arid lands, including soil degradation and irrigated areas, based on vegetation and aridity indices. Int. J. Remote Sens. 34, 6701–6722. https://doi.org/10.1080/01431161.2013.805281.

GHD, 2015. Report for the Department of Agriculture and Water Resources – Goulburn-murray Water Connections Project Stage 2 - Mid Term Review.

GM-W, 2013. Lake Eppalock Land and On-Water Management Plan. Tatura, Victoria. .

Goode, A.M., Barnett, B.G., 2008. Southern riverine plains groundwater model calibration report. A Report to the Australian Government From the CSIRO Murray-darling Basin Sustainable Yields Project. CSIRO, Australia.

Goulburn-Murray Water, 2015. Lower Campaspe Valley WSPA Annual Report 2015 [WWW Document]. URL https://www.g-mwater.com.au/downloads/gmw/Groundwater/Lower_Campaspe_Valley_WSPA/30_Sept_2015-LOWER_CAMPASPE_VALLEY_WSPA_ANNUAL_REPORT_-_2014_15_-_SIGNED.pdf (accessed 7.4.19). .

Grafton, R.Q., Wheeler, S.A., 2018. Economics of water recovery in the Murray-Darling Basin, Australia. Annual Rev. Res. Eco. 10, 487–510. https://doi.org/10.1146/annurev-resource-100517-023039.

Haasnoot, M., Kwakkel, J.H., Walker, W.E., ter Maat, J., 2013. Dynamic adaptive policy pathways: a method for crafting robust decisions for a deeply uncertain world. Glob. Environ. Chang. Part A 23, 485–498. https://doi.org/10.1016/j.gloenvcha.2012.12.006.

Herman, J.D., Reed, P.M., Zeff, H.B., Characklis, G.W., 2015. How Should Robustness Be Defined for Water Systems Planning under Change? J. Water Resour. Plan. Manag. 141, 4015012. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000509.

Hughes, J.D., 1999. SOILpak - Southern Irrigators | NSW Department of Primary Industries [WWW Document]. URL http://archive.dpi.nsw.gov.au/content/land-and-natural-resources/soil-management/south-irrig (accessed 4.23.19). .

Hughes, L., Steffen, W., Rice, M., Pearce, A., 2015. Feeding a Hungry Nation: Climate Change, Food and Farming in Australia.

Itier, B., Maraux, F., Ruelle, P., Deumier, J.M., 1996. Applicability and limitations of irrigation scheduling methods and techniques (No. 8). Irrigation Scheduling: From Theory to Practice. Food and Agriculture Organization of the United Nations (FAO).

Ivkovic, K.M., Croke, B.F.W., Kelly, R.A., 2014. Overcoming the challenges of using a rainfall–runoff model to estimate the impacts of groundwater extraction on low flows in an ephemeral stream. Nord. Hydrol. 45, 58–72. https://doi.org/10.2166/nh.2013.204.

Iwanaga, T., Zare, F., Croke, B., Fu, B., Merritt, W., Partington, D., Ticehurst, J., Jakeman, A., 2018. Development of an integrated model for the campaspe catchment: a tool to help improve understanding of the interaction between society, policy, farming decision, ecology, hydrology and climate. In: Proceedings of the International Association of Hydrological Sciences. Presented at the Innovative Water Resources Management &Ndash; Understanding and Balancing Interactions between Humankind and Nature - 8th International Water Resources Management Conference of ICWRS, Beijing, China, 13-15 June 2018, Copernicus GmbH. pp. 1–12. https://doi.org/10.5194/piahs-379-1-2018.

Jackson, S., Tan, P.-L., Mooney, C., Hoverman, S., White, I., 2012. Principles and guidelines for good practice in Indigenous engagement in water planning. J. Hydrol. 474, 57–65. https://doi.org/10.1016/j.jhydrol.2011.12.015. Water Planning in Australia: meeting the challenges of incorporating social, cultural and scientific information into decision-making.

Janssen, S., Louhichi, K., Kanellopoulos, A., Zander, P., Flichman, G., Hengsdijk, H., Meuter, E., Andersen, E., Belhouchette, H., Blanco, M., Borkowski, N., Heckelei, T., Hecker, M., Li, H., Oude Lansink, A., Stokstad, G., Thorne, P., van Keulen, H., van Ittersum, M.K., 2010. A generic bio-economic farm model for environmental and economic assessment of agricultural systems. Environ. Manage. 46, 862–877. https://doi.org/10.1007/s00267-010-9588-x.
Jensen, K., Cardoso, J.G.R., Sonnenschein, N., 2017. Optlang: An algebraic modeling language for mathematical optimization [WWW Document]. J. Open Source Softw. https://doi.org/10.21105/joss.00139.
Jones, E., Oliphant, T., Peterson, P., et al., 2001. SciPy: Open Source Scientific Tools for Python.
Kelly, R.A., Jakeman, A.J., Barreteau, O., Borsuk, M.E., ElSawah, S., Hamilton, S.H., Henriksen, H.J., Kuikka, S., Maier, H.R., Rizzoli, A.E., van Delden, H., Voinov, A.A., 2013. Selecting among five common modelling approaches for integrated environmental assessment and management. Environ. Model. Softw. 47, 159–181. https://doi.org/10.1016/j.envsoft.2013.05.005.
Kwakkel, J.H., Haasnoot, M., Walker, W.E., 2016. Comparing Robust Decision-Making and Dynamic Adaptive Policy Pathways for model-based decision support under deep uncertainty. Environ. Model. Softw. 86, 168–183. https://doi.org/10.1016/j.envsoft.2016.09.017.
Laffan, J., Smith, P., 2015. Irrigation: system and pump selection. NSW Agriculture.
Letcher, R.A., Croke, B.F.W., Jakeman, A.J., 2007. Integrated assessment modelling for water resource allocation and management: A generalised conceptual framework. Environmental Modelling & Software, The Implications of Complexity for Integrated Resources 22, 733–742. https://doi.org/10.1016/j.envsoft.2005.12.014.
Mallen-Cooper, M., Stuart, I.G., 2003. Age, growth and non-flood recruitment of two potamodromous fishes in a large semi-arid/temperate river system. River Res. Appl. 19, 697–719. https://doi.org/10.1002/rra.714.
Maskey, R., 2016. Campaspe - Data Enquiry for Masters Project.
McKay, J., 2005. Water institutional reforms in Australia. Water Policy 7, 35–52. https://doi.org/10.1080/02508069908692147.
McPhail, C., Maier, H.R., Kwakkel, J.H., Giuliani, M., Castelletti, A., Westra, S., 2018. Robustness Metrics: How Are They Calculated, When Should They Be Used and Why Do They Give Different Results? Earths Future 6, 169–191. https://doi.org/10.1002/2017EF000649.
MDBA, 2012. Sustainable Rivers Audit 2: the Ecological Health of Rivers in the Murray–Darling Basin at the End of the Millennium Drought (2008–2010). Summary.
MDBA, 2019. Climate Change and the Murray-darling Basin Plan (No. 09/19). Murray–Darling Basin Authority. Canberra. .
Megdal, S.B., Eden, S., Shamir, E., 2017. Water governance, stakeholder engagement, and sustainable water resources management. Water 9, 190. https://doi.org/10.3390/w9030190.
Mitchell, T.D., 2003. Pattern scaling: an examination of the accuracy of the technique for describing future climates. Clim. Change 60, 217–242. https://doi.org/10.1023/A:1026035305597.
Moggridge, B.J., Betterridge, L., Thompson, R.M., 2019. Integrating Aboriginal cultural values into water planning: a case study from New South Wales, Australia. Australas. J. Environ. Manag. 26, 273–286. https://doi.org/10.1080/14486563.2019.1650837.
Niswonger, R.G., Panday, S., Ibaraki, M., 2011. MODFLOW-NWT, a Newton formulation for MODFLOW-2005. Section A, Groundwater; Book 6 Modeling Techniques. p. 44.
North Central CMA, 2014. Campaspe River Environmental Water Management Plan. North Central Catchment Management Authority, Huntly.
NVIRP, 2010. Majority of Campaspe Irrigators to Exit Permanently [WWW Document]. URL https://www.g-mwater.com.au/downloads/gmw/connections/Communication/Media/MediaReleases2010/Majority_of_Campaspe_irrigators_to_exit_permanently.pdf (accessed 2.26.18). .
NVIRP, 2011. NVIRP Annual Report 2010-2011.
Oliver, Y.M., Robertson, M.J., Stone, P.J., Whitbread, A., 2009. Improving estimates of water-limited yield of wheat by accounting for soil type and within-season rainfall. Crop Pasture Sci. 60, 1137–1146. https://doi.org/10.1071/CP09122.
Peckham, S.D., Hutton, E.W.H., Norris, B., 2013. A component-based approach to integrated modeling in the geosciences: the design of CSDMS. Computers & Geosciences, Modeling for Environmental Change 53, 3–12. https://doi.org/10.1016/j.cageo.2012.04.002.
Prokopy, L.S., Arbuckle, J.G., Barnes, A.P., Haden, V.R., Hogan, A., Niles, M.T., Tyndall, J., 2015. Farmers and climate change: a cross-national comparison of beliefs and risk perceptions in high-income countries. Environ. Manage. 56, 492–504. https://doi.org/10.1007/s00267-015-0504-2.
Pulido-Velazquez, D., Garrote, L., Andreu, J., Martin-Carrasco, F.-J., Iglesias, A., 2011. A methodology to diagnose the effect of climate change and to identify adaptive strategies to reduce its impacts in conjunctive-use systems at basin scale. J. Hydrol. 405, 110–122. https://doi.org/10.1016/j.jhydrol.2011.05.014.
Qassim, A., Ashcroft, B., 2002. Estimating Vegetable Crop Water Use [WWW Document]. Agriculture Victoria. URL http://agriculture.vic.gov.au/agriculture/horticulture/vegetables/vegetable-growing-and-management/estimating-vegetable-crop-water-use (accessed 4.23.19). .
Ralph, T., Spencer, J., Rayner, T., 2010. Chapter 4: Fish. In: Rogers, K., Ralph, T. (Eds.), Floodplain Wetland Biota in the Murray-Darling Basin. CSIRO Publishing.
Refsgaard, J.C., van der Sluijs, J.P., Højberg, A.L., Vanrolleghem, P.A., 2007. Uncertainty in the environmental modelling process – a framework and guidance. Environ. Model. Softw. 22, 1543–1556. https://doi.org/10.1016/j.envsoft.2007.02.004.
Reichert, P., Borsuk, M.E., 2005. Does high forecast uncertainty preclude effective decision support? Environmental Modelling & Software. Methods of Uncertainty Treatment in Environmental Models 20, 991–1001. https://doi.org/10.1016/j.envsoft.2004.10.005.
Roberts, J., Marston, F., 2011. Water regime for wetland and floodplain plants: a source book for the Murray-Darling Basin. National Water Commission, Canberra.
Robinson, D.W., 2002. Construction and operating costs of groundwater pumps for irrigation in the riverine Plain. CSIRO Land and Water Technical Report. CSIRO Land and Water, Canberra.
Sahin, S., 2012. An aridity index defined by precipitation and specific humidity. J. Hydrol. 444–445, 199–208. https://doi.org/10.1016/j.jhydrol.2012.04.019.
Schlüter, M., Haider, L., Lade, S., Lindkvist, E., Martin, R., Orach, K., Wijermans, N., Folke, C., 2019. Capturing emergent phenomena in social-ecological systems: an analytical framework. Ecol. Soc. 24. https://doi.org/10.5751/ES-11012-240311.
Smith, P., 2015. Is Your Diesel Pump Costing You Money?.
Steffen, W., Vertessy, R., Dean, A., Hughes, L., Bambrick, H., Gergis, J., Rice, M., 2018. Deluge & Drought: Water Security in a Changing Climate. Climate Council of Australia.
Storn, R., Price, K., 1997. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. J. Glob. Optim. 11, 341–359. https://doi.org/10.1023/A:1008202821328.
Suzuki, S., Stern, D., Manzocchi, T., 2015. Using association rule mining and High-dimensional visualization to explore the impact of geological features on dynamic flow behavior. In: Presented at the SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers. https://doi.org/10.2118/174774-MS.
Tennakoon, S., Richards, D., Milroy, S., Harris, G., 2013. Water use efficiency in the Australian cotton industry, in: Waterpak: A Guide for Irrigation Management in Cotton and Grain Farming Systems. Cotton Research and Development Corporation 22–27.
Ticehurst, J., Curtis, A., 2016. Improving On-farm Water Use Efficiency Amongst Irrigators in the Campaspe River System and the Rochester Irrigation Area. Wagga Wagga, NSW, Australia 2650.
Ticehurst, J., Curtis, A., 2017. Preliminary Assessment of Conjunctive Use Opportunities for in the Murray Darling Basin: a Case Study in the Lower Campaspe Catchment. Charles Sturt University, Wagga Wagga, NSW 2650.
van Dijk, A.I.J.M., Beck, H.E., Crosbie, R.S., de Jeu, R.A.M., Liu, Y.Y., Podger, G.M., Timbal, B., Viney, N.R., 2013v. The Millennium Drought in southeast Australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society. Water Resour. Res. 49, 1040–1057. https://doi.org/10.1002/wrcr.20123.
Vaze, J., Teng, J., Chiew, F.H.S., 2011. Assessment of GCM simulations of annual and seasonal rainfall and daily rainfall distribution across south-east Australia. Hydrol. Process. 25, 1486–1497. https://doi.org/10.1002/hyp.7916.
Vellotti, G., Kalogerinis, K., 2013. Calculating Energy Requirements of Irrigation Systems.
Voinov, A., Shugart, H.H., 2013. “Integronsters”, integral and integrated modeling. Environ. Model. Softw. 39, 149–158. https://doi.org/10.1016/j.envsoft.2012.05.014.
Wang, B., Feng, P., Chen, C., Liu, D.L., Waters, C., Yu, Q., 2019. Designing wheat ideotypes to cope with future changing climate in South-Eastern Australia. Agric. Syst. 170, 9–18. https://doi.org/10.1016/j.agsy.2018.12.005.

Wheeler, S.A., Cheeseman, J., 2013. Key Findings from a Survey of Sellers to the Restoring the Balance Programme. Economic Papers 32, 340–352. https://doi.org/10.1111/1759-3441.12038.

Whitbread, A., Hancock, J., 2008. Estimating grain yield with the French and schultz approaches vs simulating attainable yield with APSIM on the eyre peninsula. In: Australian Agronomy Conference. Presented at the 14th Australian Agronomy Conference, Adelaide, South Australia.

Williams, S., Connolly, D., Williams, A., 2019. The recognition of cultural water requirements in the montane rivers of the Snowy Mountains, Australia. Australas. J. Environ. Manag. 26, 255–272. https://doi.org/10.1080/14486563.2019.1652211.

Wines, M., 2015. Protesters voice water concerns [WWW document]. Bendigo Advertiser. URL https://www.bendigoadvertiser.com.au/story/2838129/protesters-voice-water-concerns/ (accessed 10.28.19).

Xie, Y., Crosbie, R., Simmons, C.T., Cook, P.G., Zhang, L., 2019. Uncertainty assessment of spatial-scale groundwater recharge estimated from unsaturated flow modelling. Hydrogeol. J. 27, 379–393. https://doi.org/10.1007/s10040-018-1840-0.

Chapter 4b: Supplement to Chapter 4

This supplement has been added to address concerns raised by a thesis Examiner on the calibration approach applied in Chapter 4. Specifically, the difficulty of calibrating for the entire time series of data may indicate an issue with model structure. The decision to break up the climate sequence into six periods and calibrating these (resulting in six different parameter sets) may produce misleading calibration results. The concern raised by the Examiner comes from a valid perspective – that poor model performance on non-stationary data can indicate that the model does not represent some feature of the system sufficiently well. Nevertheless, we believe that the approach applied in the paper is still justifiable despite limitations.

As a reminder of the context, the purpose of the surface water model was to provide indications of dam levels for the purpose of exploratory modelling. The crucial interaction to represent was the triggering of (hypothetical) water policy rulesets which become “active” when dam levels drop to/below a certain level. The focus then was on representing fluctuating dam level reductions sufficiently to trigger these policy changes. Climate conditions govern dam levels, as might be expected by the Examiner.

In Fowler et al., (2020), it is suggested that many rainfall-runoff models lack sufficient consideration of the long-term dynamics which affects their ability to represent streamflow across long time scales. Fowler et al. suggest that this may be at least in part because basins can switch between multiple semi-permanent states (due to climatic or other influences). These semi-permanent states represent different boundary conditions, and models calibrated for a given boundary condition may not be well suited to others (as discussed in Merz et al., 2011). In other words, the model structure issue raised by the Examiner is a recognized limitation across a wide range of conceptual rainfall-runoff models, including the IHACRES model adopted for use in the presented case study.

One approach to counteract this issue is to apply a multi-model ensemble, with many approaches to their development. The approach applied in the study was to develop a set of models that are individually concerned with a specific period. This approach is similar, but not identical, to that of Zhang et al., (2011) wherein multiple models are developed for identified hydroclimatic conditions across time (i.e., separate models for different stable states).

In the presented study, climate projection data applied for the paper were scaled historic data and so follow similar trends and conditions. Thus, the model as applied in the study is adequate for the purpose of providing indications of dam state, albeit only within the indicated simulation periods. We regard this as an acceptable limitation given the scope of the study.

A further step, not conducted here but a possibility, is to perform cross-validation of models to each period, which would indicate the extent to which each model is suited for the conditions represented by a temporal period. A single parameter set that is optimal for all periods could then be selected for use (as suggested in Gharari et al., 2013). It is noted here that further investigation and exploration of hydrological ensemble model development approaches to address known limitations is well outside the scope of the chapter and is likely a series of papers in and of itself. Multi-model ensemble approaches are not, at time of writing, widely used and remain an area ripe for further research (Sharma et al., 2019).

4b.1 References for Chapter 4b

Fowler, K., Knoben, W., Peel, M., Peterson, T., Ryu, D., Saft, M., Seo, K.-W., Western, A., 2020. Many Commonly Used Rainfall-Runoff Models Lack Long, Slow Dynamics: Implications for Runoff Projections. Water Resources Research 56, e2019WR025286. https://doi.org/10.1029/2019WR025286

Merz, R., Parajka, J., Blöschl, G., 2011. Time stability of catchment model parameters: Implications for climate impact analyses. Water Resources Research 47. https://doi.org/10.1029/2010WR009505

Sharma, S., Siddique, R., Reed, S., Ahnert, P., Mejia, A., 2019. Hydrological Model Diversity Enhances Streamflow Forecast Skill at Short- to Medium-Range Timescales. Water Resources Research 55, 1510–1530. https://doi.org/10.1029/2018WR023197

Zhang, H., Huang, G.H., Wang, D., Zhang, X., 2011. Multi-period calibration of a semi-distributed hydrological model based on hydroclimatic clustering. Advances in Water Resources 34, 1292–1303. https://doi.org/10.1016/j.advwatres.2011.06.005

Chapter 5: Property-based sensitivity analysis for timely model diagnostics

Development of IEMs is often made difficult due to the concurrent development of its constituent models. Each model may rapidly evolve throughout the modelling cycle. Errors, in terms of poor model implementation, performance or conceptual mismatch, and the subsequent correction of these, may necessitate changes in other constituent models. The cost of correcting errors may increase substantially as time goes by. Consequently, it is then desirable to identify issues of model behaviour in the integrated context as early in the modelling cycle as possible.

Analysis of model parameter sensitivities through global sensitivity analyses is commonly put forth as an approach to gauge "correctness" of model behaviour. As indicated in previous chapters, IEMs often exhibit long runtimes due to their complexity (in terms of implementation and dimensionality) such that the computational expense precludes a rigorous diagnostic sensitivity analysis.

In this chapter a practical approach to identifying problematic model behaviour through property-based sensitivity analysis is introduced and demonstrated. The approach can aid in the quick identification of issues for further investigation and complement subsequent global sensitivity analyses. This chapter was submitted to Environmental Modelling and Software and is currently under review.

Iwanaga, T.*, Sun, X., Wang, Q., Guillaume, J.H.A., Croke, B.F.W., Rahman, J., Jakeman, A.J., 2021. Property-based Sensitivity Analysis: An approach to identify model implementation and integration errors. Environmental Modelling and Software 139, 105013. https://doi.org/10.1016/j.envsoft.2021.105013

Data and Analysis:

Iwanaga, T.*, 2020. diagnostic-sa. https://doi.org/10.17605/OSF.IO/NUZAF

Iwanaga, T.*, Sun, X., Wang, Q., 2020. oat-diagnostic. OSF. https://doi.org/10.17605/OSF.IO/85EDC

Environmental Modelling and Software 139 (2021) 105013

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: http://www.elsevier.com/locate/envsoft

Property-based Sensitivity Analysis: An approach to identify model implementation and integration errors

Takuya Iwanaga [a,*], Xifu Sun [a,b], Qian Wang [a,c], Joseph H.A. Guillaume [a], Barry F.W. Croke [a,b], Joel Rahman [d], Anthony J. Jakeman [a]

[a] Institute for Water Futures and Fenner School of Environment and Society, The Australian National University, Canberra, Australia
[b] Mathematical Sciences Institute, The Australian National University, Canberra, Australia
[c] State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Center for Global Change and Water Cycle, Hohai University, Nanjing, China
[d] Flow Matters, Canberra, Australia

ARTICLE INFO

Keywords:
Integrated development cycle
Diagnostic testing
Sensitivity analysis
Parameter sensitivity
Parameter inactivity
Integrated environmental model

ABSTRACT

Diagnostic testing is an oft-recommended use of sensitivity analysis to assess correctness or plausibility of model behavior. In this paper we demonstrate the use of sensitivity analysis as a complementary first-pass software test for the validation of model behavior. Typical testing processes rely on comparing model outputs to results known to be correct. Such approaches are limited to specific model configurations and require that correct results be known in advance. Property-based Sensitivity Analysis (PbSA) examines model properties in terms of the behavior of parameter sensitivities to provide a line of evidence that the expected conceptual relationships between model factors and their effects are present. Unanticipated results can indicate issues to be corrected. The PbSA approach is also scalable as it can complement existing testing practices and be applied in conjunction with global sensitivity methods that can reuse existing model evaluations or are otherwise independent of the sampling scheme.

1. Introduction

Integrated Environmental Models (IEMs) are often developed to inform policy and management processes. In the problem realm of socio-environmental systems (SES), such integrated models account for multiple sectoral influences and their interactions, including the biophysical (e.g., hydrological, climate, ecological and agriculture) and socio-economic processes (e.g., human drivers, economy/market, policy and legislative interactions). Multiple models, both purpose-built and pre-existing (i.e., legacy models; Kelly (Letcher) et al., 2013), are often coupled to represent this system-of-systems.

Typical IEM development conceptualizes an iterative 'cyclic' process in which an interdisciplinary team (of teams) collaborates to appropriately represent the interactions across the SES being modeled (Hamilton et al., 2015; Little et al., 2019). The development process is such that the suite of models that constitute an IEM, and their coupling, are in a state of flux with each undergoing a separate iterative development cycle. Changes to one model component may necessitate changes in another, and there will be emergent behaviors that arise only when models are integrated. The modeler(s) responsible for integrating the disparate models involved is the foundation for ensuring that the constituent models and the resulting IEM are both technically and conceptually sound, lest usability of the IEM and confidence in the results be compromised (Voinov and Shugart, 2013).

Compounding matters is the fact that IEMs are increasingly being operated at grander 'scales', in terms of the number of systems represented, the breadth of researchers and interest groups involved, and consequently the required computational infrastructure, budget and time available (Elsawah et al., 2020; Little et al., 2019). The resulting IEM may have hundreds, possibly even thousands, of parameters. Models external to each discipline or sectoral component are often treated as black (or at best, gray) boxes given the spread of domain-specific knowledge required to understand, in full technical detail, the models representing the system-of-systems. Consequently, no single person is likely to have a full and complete understanding of the models involved.

Given the complex and complicated context of SES modeling and the pace at which IEM development occurs, the cost of correcting errors that

* Corresponding author.
E-mail address: iwanaga.takuya@anu.edu.au (T. Iwanaga).

https://doi.org/10.1016/j.envsoft.2021.105013

Available online 24 February 2021

may inadvertently creep in may increase as time progresses (Boehm, 1986). The associated opportunity cost may be substantial and so it is desirable for any issue to be identified and corrected as early in the development cycle as possible. Continuous and repeated testing of the models and their integration, therefore, plays an important part in the model development cycle.

In the development of software, "testing" is leveraged to gain confidence that the underlying code is working as intended and continues to do so throughout the rapid pace of iterative development (Danglot et al., 2020). A failing test then falsifies the assumption that the software is working correctly. Researchers in the field of Sensitivity Analysis (SA) have independently arrived at the idea of "diagnostic evaluation". Estimated sensitivities of parameters are used to provide some validation that the model behavior is in line with expectations (Campolongo et al., 2011; Gupta et al., 2008; Pianosi et al., 2016). Such diagnostic approaches have been recognized as vital for maximizing the capabilities of mathematical models (Rabitz, 1989). Due to the computational demands of IEMs, results from diagnostic SA may be effective as once-off analyses yet take an excessive amount of time relative to the computational time and budget available for continuous testing purposes.

Given the context of rapid iteration and high complexity of IEM development, there is a need for a diagnostic process that aids in the quick and early identification of issues throughout the integrated modeling process. In this paper, we showcase how a simple and computationally inexpensive SA, based on One-At-a-Time (OAT) sensitivity analyses, applied in the frame of software testing can be a complementary strategy in identifying model implementation and integration issues early in the modeling cycle. The approach, which we refer to as Property-based Sensitivity Analysis (PbSA), can help expose issues in the course of building or integrating models by exploiting expected and unexpected sensitivity of parameters. These are used as indicators to confirm the expected model behavior in areas of parameter space with known model behaviors.

In the following sections, we briefly introduce software testing practices contextualized by the integrated model development context (Section 2) and explore its conceptual linkages with diagnostic sensitivity analysis (Section 3 and 4). We then provide an illustrative example (in Section 5) using the Campaspe Integrated Model (CIM), an integrated model developed to explore sustainable water management futures within an agricultural setting in the Lower Campaspe catchment of Victoria, Australia (Iwanaga et al., 2020a). We then conclude in Section 6 with a discussion on directions for future research.

2. Software testing in integrated model development

Computational models are software in that they are implemented as code and are run on computers. Although there are clear similarities (perhaps even identicalities) between software and model development, model testing and development practices that are common in software production may not be readily adopted (Crouch et al., 2013; Hutton et al., 2016; Sletholt et al., 2012). In fact, publications have been retracted in the past for errors that software testing practices would have assisted in identifying (Ahalt et al., 2014; Bhandari Neupane et al., 2019; Kanewala and Bieman, 2014). In this section, we briefly introduce the concept of "unit testing" and the practice of "property-based" testing.

It has long been recognized that issues are easier and cheaper to address if they are identified earlier in the development process (Levin et al., 2019; Mossalam, 2018). A core aim of software testing is to reduce the time taken to reach a "stable" working piece of software (in this case, a model) by aiding in the identification of issues as early as possible in the development workflow (see Fig. 1). Developers write code to ensure the correct functionality of other code to accomplish this aim. Such code are referred to collectively as "tests". A common type of software test is referred to as a "unit test", as it tests an arbitrary but preferably small 'unit' of code against a specific known result (Sarma et al., 2016).

Unit tests support the development process by providing indications

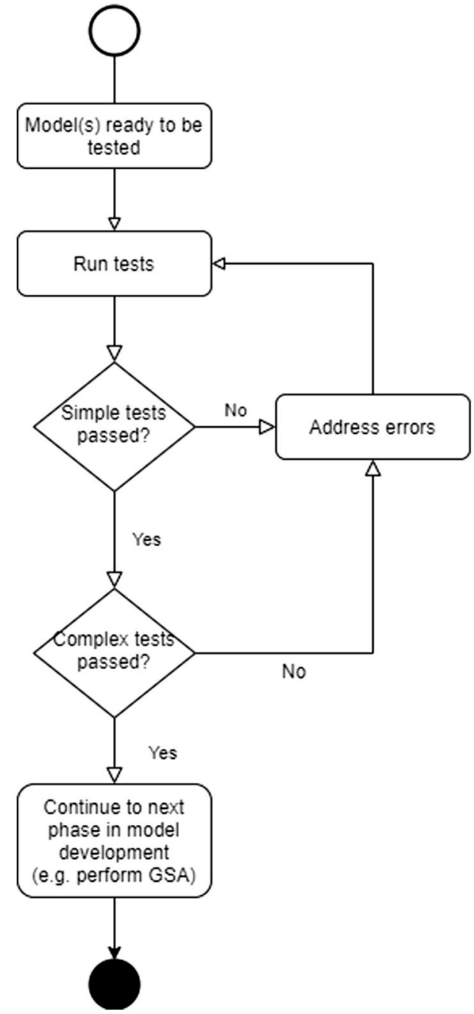


Fig. 1. Conceptual overview of the model testing workflow within the development process.

that the model is working in line with expectations. Frequent re-running of these tests (e.g., after every change) shorten the time between changes to the code and identification of issues, thereby smoothing the model development cycle. One issue is that identifying the "correct" behavior to test may be challenging in cases where the effects of model interactions may not be fully understood, as in the IEM context.

Running of tests can be automated (Verweij et al., 2010) such that a collection of unit (and property-based) tests could then form a regression and/or integration test suite. Regression tests help alert developers to the unintentional (re)introduction of issues that may have been previously addressed during model development (Huizinga and Kolawa, 2007; Yoo and Harman, 2012). Integration tests are those intended to ensure that the combined operation of multiple functions (e.g. model coupling) is both technically and conceptually sound and may also be continuously applied throughout the modeling process (Danglot et al., 2020; Laukkanen et al., 2017). Testing can uncover bugs or other issues that are "show stopping": high-priority issues that render further work inadvisable without them being addressed. From a Bayesian perspective, the more tests that are available (covering more of the codebase and the conditions of their use), the more confident modelers can be in the correct functionality of the model (Davidson-Pilon, 2016).

Although there is some evidence that software testing practices are being adopted within the computational sciences (Hannay et al., 2009; Sarma et al., 2016; Sletholt et al., 2012), to what extent is difficult to ascertain given the weak, albeit strengthening, norms requiring the provision of model code (Hutton et al., 2016). Adoption of software development practices such as testing is likely to be low given recent literature that encourage their adoption. Software development practices, in general, are also acknowledged to play a part in resolving issues with replicability and reproducibility of studies in environmental science and the computational sciences (Ahalt et al., 2014; Easterbrook, 2014; Gray and Marwick, 2019; Hut et al., 2017).

One possible reason for the sparsity of (reported) software testing is the lack of formal software development training for researchers (Hannay et al., 2009) and the reliance on mathematical or statistical rigor in model implementation. There is also an element of trust involved due to the variety of disciplines found within IEM development: as constituent models and their components are taken to function correctly in the *un*-integrated context, they are assumed to be correct in the integrated context. Regardless of the reasons, the consistent application of tests for environmental model quality assurance appears to still be in its infancy. The subsequent possibility of technical complications influencing model results (referred to as technical uncertainty; Walker et al., 2003) or as a consequence of conceptual mismatches across disciplinary specialists appears to be largely ignored.

2.1. Practical considerations of computational budget

It is important to recognize and consider the computational costs involved in the diagnostic context as every computational work is subject to a budget arising from the intertwined concerns of available time, computational power and monetary cost. These concerns are collectively referred to here as the computational budget. A hypothetical context, wherein a model integrator performs tests on a typical desktop computer, is described here to service the argument. Although dedicated infrastructure may be available (e.g., distributed or cloud-based platforms), they too would be constrained by the same or at least similar considerations regarding their computational budget.

As IEMs are often time-consuming to evaluate, model diagnostics may be scheduled to run overnight on a desktop computer (e.g., 5 p.m. to 9 a.m., or 16 h). Current typical development machines have 4 cores. If the model is estimated to take, on average, an hour to run, then 64 model evaluations may be conducted in the available time. In practice, model runtimes should be expected to be variable, and computational performance is unlikely to scale linearly with the number of cores due to the computational overheads involved. A rule of thumb to arrive at an estimate of runtime is given by:

$$c \cdot t \cdot (1 / r) \tag{1}$$

where c is the number of cores available, t is the time available in hours, and r is the estimated model run time in hours (assumed here to be $\leq t$). It is common practice to inflate the runtime estimate (r) by some degree (e.g., by 10%) based on prior empirical knowledge of the model's computational performance and requirements. Overestimating the runtime ensures that model evaluations complete within the defined available time given the variability of model runtime and computational overhead. Such considerations are also important in cases where cloud-based infrastructure is adopted as such services may charge by a unit of time (e.g., per minute). Description of the terms used throughout the paper are provided in Table 1 for ease of reference.

Running of tests can be structured such that they are run from the simplest (and least time-consuming) to the most complicated (and computationally intensive). Failure of a simple test may then negate the need to run a more computationally intensive test. In some cases, failure of any single test may preclude the necessity of running any other tests, as the model has been shown to have issues, or at least allow for a more targeted diagnostic to occur. Structuring tests in this manner aids in conserving available computational budget.

Table 1
List of terms and their definitions as used in this paper.

Variable	Definition
n	Number of sample repetitions
p	Number of model parameters/factors/inputs/dimensions
N	Total number of model evaluations
s	A targeted subset of model parameters, or groups of parameters, for analysis such that $s < p$
c	The number of available computer cores for the purpose of running a model for diagnostic sensitivity analysis
t	The time (in hours) available to conduct diagnostic sensitivity analysis
r	The estimated runtime of the model (in hours), assumed to be $\leq t$

2.2. Example unit and property-based testing

Box 1 shows an example unit test implemented in the Python programming language (with the 'pytest' framework; Krekel et al., 2004) for an example linear function (Case 1 in Li et al., 2010):

$$y = x_1 + x_2 + x_3 + x_4 + x_5 \tag{2}$$

This simple example illustrates unit tests that protect modelers from changes (inadvertent or otherwise) that may introduce errors that would otherwise go unnoticed, but only for a specific known result. One disadvantage of unit testing is the need for such specificities to be known, and for tests to be written for each. While requirements may be known in advance, particularly in "business-oriented" software development, it is less likely in research modeling contexts, and even less likely where the complex interactions between models are involved, as in IEMs. While it is possible to test that a known correct model output has not changed, such a test does not apply to new model configurations, as is common when integrating existing models.

To counter this limitation, modelers may adopt a property-based testing approach (Fink and Bishop, 1997), wherein the expected behavioral aspect of the software/model is tested, rather than a specific known output as with the regular unit testing approach. Sets of inputs to feed into the model would be automatically generated in a property-based testing approach. Property-based testing was perhaps popularized by the QuickCheck tool for the Haskell programming language (Claessen and Hughes, 2000), which sparked the development of similar tooling for other programming languages. Such testing frameworks can assist in determining the properties of failing tests themselves, helping to identify specific cases in which the model does not behave as expected (Löscher and Sagonas, 2017).

To give a specific example, one such test could serve to ensure a zero or positive valued output is obtained (i.e., ≥ 0) in cases where the sum of positive inputs is greater than the absolute sum of negative inputs, as this is an expected property of the model. Box 2 depicts an implementation of such a property-based test, along with its output indicating that the test failed as the model does not produce the expected behavior. Code for these examples are provided in Iwanaga (2020). On examination, we see that the model was incorrectly implemented (see Box 3) but, crucially, in a way the previous unit test shown in Box 1 would still pass. The results illustrated here should not be taken to mean that property-based testing supersedes unit testing as both are useful and can be leveraged in tandem to inform the level of confidence in the model implementation.

Modelers may find that property-based testing is somewhat analogous to pattern-oriented modeling (Grimm, 2005; Grimm and Railsback, 2012), although the focus of the latter is on model construction and calibration. There is a conceptual similarity in that both pattern-oriented and property-based approaches evaluate model "accuracy" against known (or desired) behavioral properties rather than evaluating against a single point of truth (i.e., a benchmark). Failure of a model to adhere to expected behavior then invalidates the assumption that the model is functioning correctly. It is, therefore, useful to test

Box 1
Example unit test which checks that the model run with all zero inputs returns 0.0 as its output. The output from the pytest framework is shown below indicating the test passed as expected.

```
def test_li_2010():
    expected_result = 0.0

    # Run the model with all zero inputs
    result = li_2010(0.0, 0.0, 0.0, 0.0, 0.0)

    assert result == expected_result, "Result should be zero"
```

```
                test_examples\test_li.py .                [100%]

                ============== 1 passed in 0.18s ==============
```

Box 2
Example property-based test. The sign of the result should be consistent with whether the sum of positive terms is greater or smaller than the sum of negative terms. The test generates 100 test cases with values for x between -100 and 100. Test results indicate that although the earlier unit test passed, the property-based test failed.

```
import numpy as np

def test_li_2010_property():
    inputs = np.random.randint(-100.0, 100.0, size=(100, 5))
    for i in inputs:
        result = li_2010(*i)
        if np.sum(i[i >= 0.0]) >= np.abs(np.sum(i[i < 0.0])):
            assert result >= 0.0, \
                f"Result should be positive for inputs {i}"
        else:
            assert result < 0.0, \
                f"Result should be negative for inputs {i}"
```

```
============================== FAILURES ==============================
___________________ test_li_2010_property_small ______________________

    def test_li_2010_property():
        inputs = np.random.randint(-100.0, 100.0, size=(100, 5))
        for i in inputs:
            result = li_2010(*i)
            if np.sum(i[i >= 0.0]) >= np.abs(np.sum(i[i < 0.0])):
>               assert result >= 0.0, \
                 f"Result should be positive for inputs {i}"
E               AssertionError: Result should be positive for
                  inputs [-29  88 -80   2  21]
E               assert -40 >= 0.0

test_examples\test_li.py:26: AssertionError
====================== short test summary info ======================
FAILED test_examples/test_li.py::test_li_2010_property_small - ...
==================== 1 failed, 1 passed in 0.34s ====================
```

against a broad range of expected behavioral patterns as models are modified and coupled, and to do so frequently throughout the modeling cycle.

3. SA in the evaluation process

Sensitivity analysis (SA) can play multiple roles in the model evaluation process. A common use of SA is to screen and rank factors

Box 3
The example function (#1) from Li et al. (2010) with an incorrect implementation. Note the subtraction of x_5 and compare with (Eq (2)).

```
def li_2010(x1, x2, x3, x4, x5):
    y = x1 + x2 + x3 + x4 - x5

    return y
```

(parameters and input variables) according to their influence on model outputs (Razavi et al., 2020; Saltelli et al., 2008). SA may also be used to analyze the bounds and uncertainties of a model's parameters and its predictions, and is valuable in assessing model identifiability (Guillaume et al., 2019; Shin et al., 2015). Model sensitivities have also been assessed as part of a diagnostic evaluation procedure, to aid in verifying models and their structure (Gupta et al., 2008; Pianosi et al., 2016; Sieber and Uhlenbrook, 2005). Typical applications of diagnostic SA concern themselves with the identification of model components or parameters that explain (or should explain) differences between simulated and observed system behavior (Gupta et al., 2008; Reiter, 1987; Saltelli et al., 2004). Diagnostic SA typically assumes that model development is complete. Rarely is it framed as an approach to test and validate model behavior throughout the model development cycle.

A key consideration in the selection of an SA method (or methods) is its appropriateness for the intended aim constrained by the available computational budget. Screening and ranking parameters, for example, requires substantially fewer model runs to accomplish compared to obtaining estimates of parameter sensitivities (Herman et al., 2013; Sarrazin et al., 2016). Screening for parameters on which to conduct further analysis is a common practice that aids in conserving computational budget (Cuntz et al., 2015; Mai and Cuntz, 2020). Fixing the resultant insensitive parameters constraints the number of parameter combinations to be run for later Global SA or Bayesian uncertainty analysis. The trade-off is a risk that fixing parameters may introduce large errors in the quantities of interest.

In the following subsections, we describe typical SA approaches and their suitability in the diagnostic context. For context, brief descriptions of the terms used are provided in Table 2.

3.1. Sensitivity analysis methods

In typical local sensitivity analysis (LSA), each model parameter is assigned a "best guess" baseline value and then changed ('perturbed') by setting to some pre-selected value or multiplying by some proportion and then returned to their baseline value whilst others remain fixed (Campolongo et al., 2011). The derivative is calculated for each change and the process repeated for each parameter one after the other, giving it its name "One-At-a-Time" (OAT). Any changes to the model output are thus attributable to the parameter that was perturbed. Such approaches are defined as "local" as they are only capable of providing indications of sensitivity at specific points in parameter space. In contrast to global sensitivity analysis (GSA), LSA cannot provide indications of interactions between parameters and their effect on model outputs (Saltelli et al., 2019; Wagener and Pianosi, 2019). The OAT approach described here is referred to as a 'pure OAT' to distinguish it from other (global) approaches, which may also vary parameters one-at-a-time.

There are other approaches to SA that do not rely on OAT. Variance-based methods are a commonly used class of GSA which involve the perturbation of parameters all-at-a-time (Douglas-Smith et al., 2020). Although more appropriate for parameter sensitivity estimation compared to pure OAT, variance-based approaches can be difficult to apply for early diagnosis of model issues where large numbers of parameters and long runtimes are involved. Sufficient samples are needed to obtain accurate sensitivity estimates, and this can increase exponentially with the number of parameters involved. There is, however, no

Table 2
Terms used to describe the role of sensitivity analysis.

Term	Description	Reference for more information (where applicable)
SA	Sensitivity analysis	–
LSA	Local sensitivity analysis	–
GSA	Global sensitivity analysis	–
OAT	One-At-a-Time analysis	–
R-OAT	Radial One-At-a-Time analysis	Campolongo et al. (2011)
PoI	Parameter of Interest	–
QoI	Quantity of Interest	–
Parameter sensitivity	Measures of sensitivity may have a direct interpretation, e.g., the magnitude of effect of an input on the output, or the variance attributable to an input.	(Hamby, 1994; Saltelli et al., 2008)
Screening	The identification of insensitive parameters; those that have little to no effect on model outputs. Screening may also be used as a diagnostic test for parameter inactivity (proposed in this paper).	(Herman et al., 2013; Saltelli et al., 2008; Sarrazin et al., 2016)
Ranking	The ordering of parameters by their influence on model results	(Pianosi et al., 2016; Saltelli et al., 2008)
Parameter identifiability and equifinality	"Parameter identifiability analysis assesses whether it is theoretically possible to estimate unique parameter values from data, given the quantities measured, conditions present in the forcing data, model structure (and objective function), and properties of errors in the model and observations." If the objective function is insensitive to a parameter, it means that the objective function is flat, and the parameter is not identifiable. A related concept is that of *equifinality*, which refers to the principle that the same output may be obtained using different methods, models, parameters, and combinations of parameter values with the same set of observations. In short, multiple conceptualizations may lead to equally acceptable outcomes.	Guillaume et al. (2019) Beven and Freer (2001)

universally applicable rule that provides a reliable estimation of the number of samples required, which changes from method to method, sampling regime, the number of parameters, the model itself and its quantities of predictive interest (Wagener and Pianosi, 2019).

Pure OAT is unsuitable for comprehensive analysis of sensitivities in complex models with non-linear behavior as wider areas of parameter space must be explored to capture global indications of parameter interactions (Razavi and Gupta, 2015; Saltelli and Annoni, 2010; Yang, 2011). Despite these shortcomings, OAT remains prevalent in model assessment against published advice (Ferretti et al., 2016), although the situation does appear to be slowly improving (Douglas-Smith et al., 2020). One clear advantage that OAT has, exploited in the PbSA approach of this paper, is its conceptual and computational simplicity relative to other methods.

The Morris method (Morris, 1991) is an extension of the OAT approach (Vanrolleghem et al., 2015), being capable of providing adequate indications of sensitivity for a variety of purposes in the context of complex nonlinear models (Sun et al., 2012). The Morris method changes parameter values one-at-a-time (and so is sometimes referred to as Morris One-At-a-Time) but does so in a stepwise manner, without dependence on nominal values, through a process known as trajectory sampling. Unlike the pure OAT approach, parameter values are not reset to their original start points and instead are kept until all parameters have been modified. The process is repeated n times so that the total number of model evaluations is $N = n{\cdot}(p + 1)$, where usually $n \approx p$ or less (Norton, 2009). Thus, the number of model evaluations increases quadratically with the number of parameters, unless $n \ll p$ in which case the increase is linear.

The sensitivity index produced by the Morris method indicates the relative change in the quantity of interest regarding the changed parameter value (the average elementary effect, μ), the average absolute change in parameter value, which accounts for the effect negative values may have (denoted as μ^*), and its standard deviation (σ), which indicates interaction and non-linear effects. A high σ indicates that a parameter is interacting with others (Braddock and Schreider, 2006; Pianosi et al., 2016; Saltelli et al., 2008). The Morris method is often recommended for screening and ranking purposes (Cuntz et al., 2015; Saltelli and Annoni, 2010) as it requires fewer model runs to arrive at an acceptable parameter rank or screening conclusion compared to other common SA approaches (see, for example, (Braddock and Schreider, 2006; Cuntz et al., 2015; Herman et al., 2013; Sun et al., 2012)).

An alternative to the Morris approach is the application of OAT with a “radial design”, wherein the pure OAT approach is repeatedly applied around different “start points” (Campolongo et al., 2011). In this Radial approach (referred to as R-OAT from hereon), the model is evaluated $n{\cdot}(p+1)$ times, where n is the number of repetitions. It is noted here that R-OAT transforms OAT from a local to global SA when $n > 1$. Thus R-OAT is equivalent to pure OAT when $n = 1$, and the total number of model evaluations is the same as with the Morris method. Unlike the Morris method, however, R-OAT does not require a specific sampling scheme and can leverage existing schemes such as Latin Hypercube, Sobol’ sequences, or even simple Monte Carlo to gain an indication of variance-based indices (Campolongo et al., 2011; Pianosi et al., 2016).

R-OAT is particularly appealing within the IEM context due to its simplicity and scalability, leading to its application being relevant throughout the model development life cycle. As suggested previously by Campolongo et al. (2011), a collection of SA results can be built up in stages where and when necessary. Smaller samples for diagnostic purposes can be built on, with additional samples added for screening and ranking. Larger samples can be used to obtain an indication of global effects via variance-based indices, assuming no implementation or integration errors are identified.

There are alternatives to variance-based approaches, such as moment-independent (also known as density-based) approaches from which usable indicators can be obtained with a reduced number of samples relative to variance-based approaches. The PAWN method (Pianosi and Wagener, 2015, 2018), for example, was found to be able to identify parameters of significance with 10% of the samples needed by the Sobol’ method for a 26-parameter hydrological model (200 compared to 2000 samples; Zadeh et al., 2017).

With the PAWN and Sobol’ methods, a dummy parameter can be used to obtain an indication of insensitive parameters. A dummy parameter is an inactive factor that does not have any influence on the behavior of the model (i.e., it is completely insensitive). Parameters that are awarded a sensitivity rank equal to or less than the dummy parameter are assumed to be insensitive. The focus therein, however, is on assessing parameter sensitivities rather than expected model behavior.

The use of emulators, which approximate the model response surface with an abstract formalism, is one oft-suggested approach to resolving issues of computational complexity and runtime (e.g. Yang et al., 2018), and could in principle be used to speed up testing. Developing emulators, however, requires sufficient areas of parameter space to be represented. The time taken to obtain the necessary samples for a complex model is typically prohibitive in the context of the model development cycle. By the time the emulator is ready, the model is likely to have undergone significant changes such that the emulator represents an obsolete version. A further consideration is that many methods require that the response surface have a level of smoothness for it to be approximated and that the parameterization of the original model is not exceedingly high (Oakley and O’Hagan, 2004; Sudret, 2008). The above criteria are often not met in the case of IEMs. The error in emulators also needs to be evaluated prior to use, as emulation of an IEM with conceptual or implementation issues renders any subsequent uses of the emulator beyond diagnostic tests inappropriate, making their development too costly for the sole purpose of obtaining indicative results.

3.2. Example diagnostic SA

To provide an illustrative, if simplistic, example of diagnostic SA within the development cycle, a hypothetical model developer could apply the Morris method (Morris, 1991) to gain an indication of the behavior of the (incorrectly implemented) model introduced above (see Box 2 and 3). The Morris sensitivity index indicates the relative change in the quantity of interest regarding the changed parameter value (μ), the average absolute change in parameter value (μ^*) which accounts for the effect negative values may have, and its standard deviation (σ) which indicates interaction and non-linear effects (Campolongo et al., 2011). The method as implemented in the SALib (Sensitivity Analysis Library; Herman and Usher, 2017) package for Python is used here for demonstration purposes, which applies the improved sampling method introduced in Ruano et al. (2012). Relevant code for this example may be found in Iwanaga (2020).

For the example linear function (Eq (2)), two properties are expected. First, the effect of each parameter is expected to be positive given the quantity of interest is the sum of all inputs. Second, the contribution of parameters to the quantity of interest is expected to be equal, again due to the linear nature of the model. Although the second property is satisfied, the results indicate that x_5 is having a negative effect due to the erroneous implementation. Fig. 2 depicts this unexpected result for the erroneously implemented example function (see also Eq (2) and Box 3).

In the software testing paradigm, diagnostic SA is a form of property-based test as the model property (i.e., its sensitivities) are being investigated and evaluated, although modelers usually apply this in a more ‘manual’ manner through the visualization and qualitative assessment of results. Diagnostic SA may be an effective complement to ‘traditional’ software development tests, particularly in complex integrated modeling contexts, as the correct functioning of code in isolation does not necessarily imply conceptually correct integrated model behavior

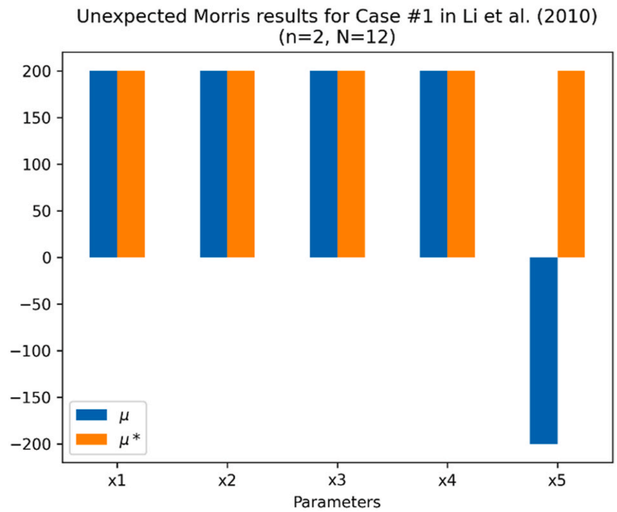


Fig. 2. Example of diagnostic sensitivity analysis using Morris. Identical, or near identical, positive effect (indicated by μ) would be expected for the example linear function. Diagnostic results instead show negative effect from x_5 given the incorrect implementation. The μ^* values indicate equal contributions from all parameters as is expected. The σ metric is not shown here to reduce clutter in the figure as it is unimportant for the purpose of this illustration. The number of grid levels for the Morris approach is set to 4 as suggested in the literature (Campolongo et al., 2007). Results were obtained with 12 model evaluations ($n = 2$).

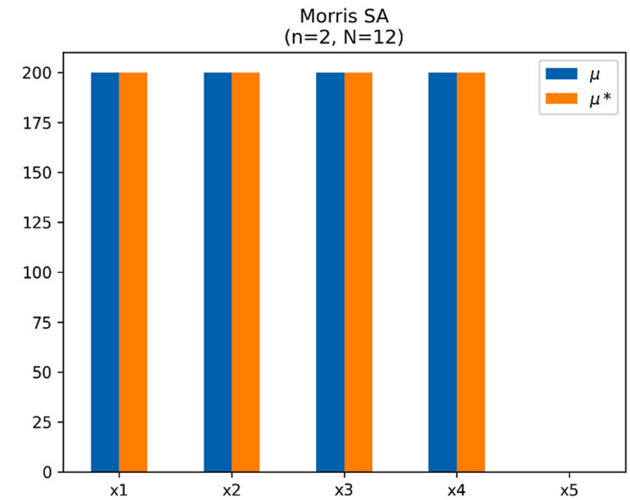


Fig. 3. Unexpected results (with Morris method) for the model with implementation error which renders x_5 inactive. Results obtained with 12 model evaluations.

(Voinov and Shugart, 2013). One barrier to the adoption of diagnostic SA is the reported lack of norms around investigating model sensitivities (Saltelli et al., 2019). Use of SA, in general, is reportedly low with the complexity and lack of understanding of recommended SA techniques being one suggested reason for the lack of uptake (Ferretti et al., 2016; Saltelli et al., 2019). There may also be resistance towards the adoption of the new and unfamiliar (known as the status quo bias; Samuelson and Zeckhauser, 1988) as hypothesized in Ferretti et al. (2016).

4. Parameter (in)sensitivity as a property to test

The IEM development context can involve multiple modeling paradigms, disciplinary/sectoral knowledge and feature the adoption of multiple technologies, including different computational infrastructure and programming languages (Hannay et al., 2009; Hut et al., 2017; Hutton et al., 2016; Sletholt et al., 2012). As mentioned in the Introduction, this leads to a situation in which no single modeler has a full and complete understanding of the models involved. Given the complex development context of IEMs, one fundamentally important property that IEM and other model developers can target for indicative assessment is the inappropriate sensitivity of parameters known to have cross-system influences.

The principal idea here is that parameter sensitivities are generally a robust property of model behavior that provide indications of correct model implementation and integration. Parameter activity or inactivity (i.e., complete insensitivity) is a property that will remain invariant even as the model itself changes and evolves through the model development cycle and as the precise model outputs change. Thus, diagnostic SA applied as a form of property-based test to regions of parameter space in which model behavior is expected to be sensitive (or insensitive) can then provide early confidence that other, more computationally demanding, processes can proceed without issue.

For IEMs, conceptual analysis of the relationships between the models can be invoked to identify parameters to test (an example is provided in Section 5). Quantitative assessment of SA results within the automated testing process could alert modelers to unintended changes that unknowingly affect model applications. Such tests may also guard against issues of technical uncertainty (specifically computational infrastructure uncertainty), as model behavior may differ under different computational contexts (Bhandari Neupane et al., 2019; Iwanaga et al., 2020a; Walker et al., 2003).

4.1. Testing for inactivity with Property-based SA

To provide a concrete example with the earlier example function, another error is introduced, perhaps in the process of correcting the earlier implementation issue (shown in Box 4), which cancels out the effect of parameter x_5. Diagnostic SA results with the Morris method are shown in Fig. 3, highlighting the issue for modelers to investigate. This simplistic example is intended to illustrate the concept; a more

Box 4
The example function with another bug introduced in the process of correcting the earlier issue shown in Box 3. Note that an addition of x_5 has been accidentally included rather than replacing the earlier subtraction of x_5.

```
def li_2010(x1, x2, x3, x4, x5):
    y = x1 + x2 + x3 + x4 - x5 + x5

    return y
```

expansive example is provided in Section 5.

Although the Morris method is applied in this specific diagnostic case, the same conclusion of inactivity can be reached with a purely One-At-a-Time (OAT) analysis with $p+1$ model evaluations in the worst case (i.e., $N=6$ for the example function). As indicated in Section 2.1, identified issues with the model implementation (or integration) through the failure of a test at any point may negate the need to run further tests. Failure of OAT to produce expected results negates the need to apply other, more computationally intensive, tests and SA. For this reason, the total number of evaluations to invalidate the model may be less than $p+1$. Hypothetically, it may not be necessary to test all parameters, as in cases where only a certain subset of parameters (denoted as s, where $s < p$) is expected to influence model interactions. In such cases, inappropriate model behavior could be determined with $N = 2, \ldots, s+1$ model evaluations.

Unexpected results should be investigated before any further analyses proceed. Use of a purely OAT approach is heavily discouraged in the literature (Saltelli et al., 2019; Saltelli and Annoni, 2010) and it is suggested as appropriate here only because of the expectation that other forms of tests and SA will be applied after property-based SA tests pass. It is stressed here that relying on OAT for purposes outside this first-pass diagnostic context is not encouraged. In the case of IEM development, property-based tests that focus on parameters that influence model interactions can be a computationally effective approach to obtaining a first-pass indication of correct conceptual and technical model integration.

4.2. Example Property-based SA for parameter inactivity

We devise two property-based testing strategies for the quick, first-pass, identification of model integration issues. The first is a form of OAT referred to as extremity testing, and the other follows a more usual sensitivity analysis approach using R-OAT. As noted previously, failure of these tests indicates the presence of issues that should be further investigated prior to the application of more computationally expensive diagnostics (e.g., a global sensitivity analysis) or operationalization of the model. Both approaches require that the conceptual relationships between parameters at their extremes and the targeted QoIs are known beforehand.

With extremity testing the model is run just twice (i.e., $N = 2$). One run is to be conducted with the targeted parameters perturbed to their lower extremes, and the other run with parameters set to their upper. An example is shown in Box 5. In practice, any sufficiently large perturbation should suffice, and the upper and lower extremes are suggested here for conceptual simplicity and ease of application. Under usual applications of SA, extremity testing comes with a risk of Type I and II error (a false positive or negative) due to non-monotonicity. The approach is applicable in this specific case as the conceptual relationship between QoIs and parameters is an *a priori* expectation. Diagnostics are being carried out in a restricted (local) area of parameter space where sensitivities are expected to exist. The primary concern is to determine whether the effect can be identified before the application of the model and more rigorous analysis such as with GSA. Note that diagnostics may also be carried out in regions of parameter space that are known to produce no effect, wherein larger than expected (i.e., non-zero) sensitivities can also indicate an issue.

PbSA, in this case using R-OAT, can be useful in identifying the conditions in which unexpected behavior occurs, thereby helping to avoid a potentially time-consuming debugging exercise. Two requirements can then be set for a GSA method to be a practical complement in the IEM development context. It would be desirable for any samples to be reusable in a later GSA if results are found to be acceptable. Another requirement is that the time taken to conduct such analyses should not exceed the available computational budget for such analyses to be timely and useful.

The illustrative examples provided in earlier sections showcase a diagnostic approach from both software development and SA perspectives. In these examples, however, the hypothetical modeler has sufficient understanding of the model and its implementation details to apply and evaluate results from both tests and diagnostic SA. In the context of integrated environmental modeling, this may be a luxury rather than a given due to the aforementioned interdisciplinary nature of IEM development (Iwanaga et al., 2021; Knapen et al., 2013). In Section 5 we describe the case of the Campaspe Integrated Model and the usefulness of PbSA with extremity testing as an indication of valid model integration.

5. An example with the Campaspe Integrated Model

The Campaspe Integrated Model (CIM) (Iwanaga et al, 2018, 2020a) is a hydro-environmental-economic model used to explore water management options. The CIM is highly complex, featuring interactions between six non-linear component models, each representing a specific system. It can be considered a system-of-systems model in which a representation of the socio-environmental system is built up from multiple independent and interacting constituent models (Little et al., 2019). In the hypothetical development context, individual model developers are disciplinarily diverse with their own traditions, practices and preferred modeling approaches. A common language and perspective of the modeling being conducted may still be developing (MacLeod and Nagatsu, 2018; Thomas and McDonagh, 2013). Modelers may also be geographically spread, inducing delays in communication that increase the risk of inadvertent errors being introduced.

To reflect this interdisciplinary context, the model is treated here as a gray-box for the purpose of the example. Modelers involved in the integration of constituent models may have working knowledge of the represented system and the operation of each model (e.g., implementation and usage), but are not necessarily disciplinary specialists themselves. Thus, the primary concern in the initial stage is to gain confidence that operation of the IEM is both conceptually and technically sound by testing the assumptions associated with the conceptual understanding (Iwanaga et al., 2020a; Wilson et al., 2017). Falsifying the assumption that the model is integrated correctly also helps to preserve available computational budget.

In the development IEMs, the relationships between all parameters and QoIs may not be fully known because of the complex model interactions that occur. Parameter activity/inactivity may be a proxy that indicates correct model integration. Testing for the "obvious" behavior (i.e., change of PoI have flow-on effects that should affect the QoI), and continual confirmation that the behavior is present throughout the development cycle is valuable in that errors or conceptual mismatches could be highlighted, and corrected, earlier in the modeling cycle. In other cases, the conceptual understanding that the model integrator has may not be complete and so the testing process could be helpful in improving modelers' understanding of the IEM. New knowledge or model configurations may invalidate previously "obvious/assumed" behavior; in which case the tests serve to alert modelers to a change in context. Change in context should subsequently be documented and the relevant tests updated to reflect this new understanding.

In this example, the model integrator is principally focused on the policy, surface water hydrology, and farming system models, however the entire model also includes representations of climate, groundwater and ecology. An example of the initially known interactions between the constituent models of interest is provided in Fig. 4. Further description of the CIM may be found in (Iwanaga et al., 2020a). Interactions between all models affect the main quantity of interest selected here; that is, the long-term surface water allocation index, which indicates the average volume of water made available to water users over the simulation period. The CIM has 53 parameters which may all be varied. Runtime of the model is variable depending on the scenario being run but typically takes 30 minutes.

The influence of a single PoI – "irrigation efficiency" – is investigated

Box 5
Example automation of extremity testing where the bounds of each parameter are set to −100 and 100. Output has been modified for clarity. In this specific case, checking for a lack of change in the 'result' variable relative to 'nominal_result' or a 'y_diff' equal to 0.0 also suffices. This test could be repeated with varying parameter values using a sampling scheme (e.g., Sobol', Latin Hypercube) or monte carlo.

```
============================= FAILURES ============================
_________________ test_li_2010_parameter_active __________________

    def test_li_2010_parameter_active():
        """Testing for parameter activity.

        This is an extremity test, a form of OAT sensitivity analysis.
        """
        np.random.seed(101)

        base = [-100.0, -100.0, -100.0, -100.0, -100.0]

        # run model at nominal position
        nominal_result = li_2010_case1_inactive(*base)

        for idx, x_i in enumerate([100.0, 100.0, 100.0, 100.0, 100.0]):
            tmp = base[:]  # copy nominal values
            tmp[idx] = x_i  # perturb parameter value
            result = li_2010_case1_inactive(*tmp)  # run model

            x_diff = np.abs(base[idx] - tmp[idx])
            y_diff = np.abs(nominal_result - result)

            # Test sensitivity measure
            # (should not be 0.0, or close to it)
>           assert not np.isclose(0.0, (y_diff / x_diff)), \
                f"Perturbing parameter x_{idx+1} should affect SA
                  metric!"
E           AssertionError: Perturbing parameter x_5 should affect SA
                            metric!
E           assert not True
E            +  where True = <function isclose at
                0x00000250546DD430>(0.0, (0.0 / 200.0))
E            +    where <function isclose at 0x00000250546DD430> =
                  np.isclose

tests\test_li.py:56: AssertionError
```

in this hypothetical context. The parameter relates to the efficiency of water application for "pipe and riser" irrigations, a common irrigation mode available to farmers in the Campaspe region. We consider the effect on modeled long-term average surface water allocations. Rendering the PoI static simulates an inadvertent change that introduced an implementation or integration error. Interactions between models are consequently inappropriately represented. Specific details of the PoI are provided in Table 3.

Irrigation efficiency relates to the proportion of water that reaches a crop's root zone. The higher the efficiency rating, the less water that is "wasted" or "lost" from a farmer's perspective to evaporation, run-off, or deep drainage (e.g., aquifer recharge). Hence, the more efficient an irrigation system, the less water required to maintain crop productivity for a given spatial area, and the less water extracted from the dam. Water is allocated each year to farms by the policy model. Excessive use of water by farmers in one year can reduce farm water availability in subsequent years, making efficient irrigations desirable.

Given the available computational budget (as contextualized in Section 2.1), overnight execution of tests between 5pm and 9am (i.e., 16 h) would allow < 128 model evaluations on a (currently) typical 4-core machine, using Eq (1) above, assuming a consistent 30-min runtime. A lesser number should be selected to ensure model runs resolve within the available time as runtime should not be expected to be consistent (as explained in Section 2.1). Relevant to the point here is the application of the Morris method for a REALM model as reported in Braddock and Schreider (2006). The REALM model is similar to the CIM in terms of geographic region (targeting the neighboring Goulburn catchment) and its use in water allocation modeling. Computational considerations which constrained the number of available model samples are highlighted therein. Use of cloud-based testing infrastructure is ignored for the purpose of illustration and may potentially be cost-prohibitive depending on the project budget. It is unlikely that indicative results would be obtained with GSA. Development and use of emulators are similarly precluded given their requirement for sufficient areas of parameter space to be represented, which is not possible within the allotted time.

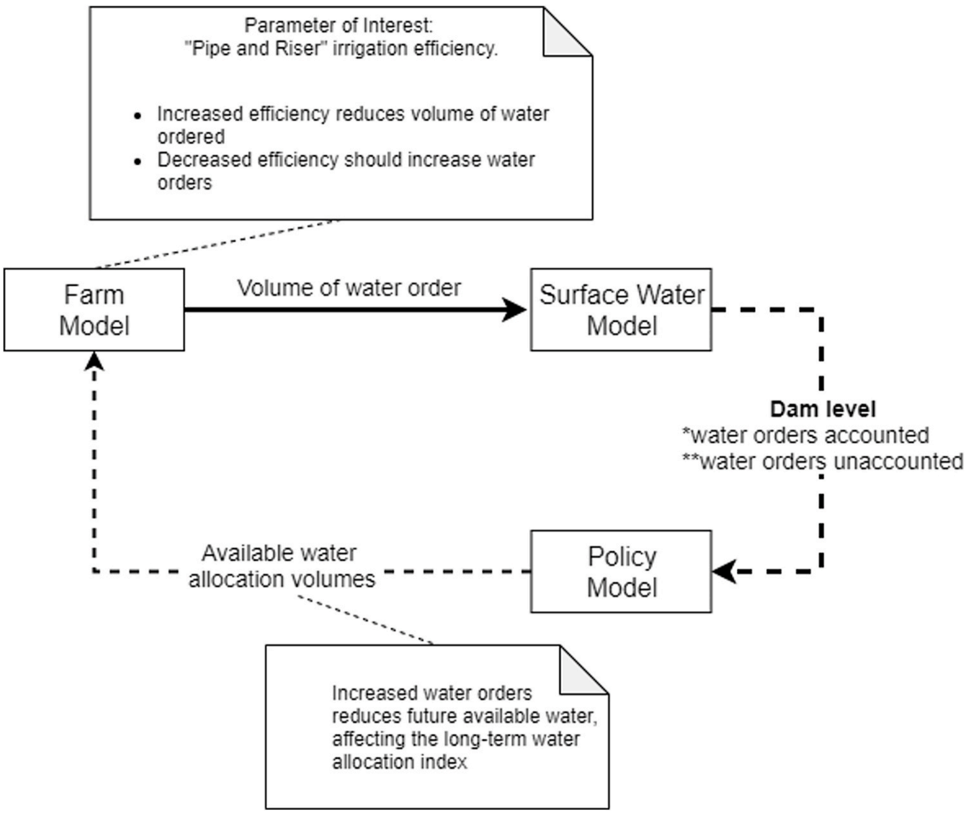


Fig. 4. A simplified component interaction diagram showcasing the feedback loop between constituent models, as initially envisioned by the model integrator. The farm model determines the volume of water to apply to satisfy crop needs. The volume of water required is dependent on the efficiency of the irrigation system (the Parameter of Interest). The water is extracted from the dam (represented in the surface water model) and the subsequent water levels inform the future volume of water allocated for agricultural use and thus, long-term surface water allocations (the Quantity of Interest). The inter-connection between the surface water and policy models is switched on (*) and off (**) to generate two case results for this study, simulating an inadvertent implementation and/or integration error.

Table 3
Description of the parameter of interest: irrigation water application efficiency.

Parameter of Interest	Nominal Value (and Range)	Reference(s)
Pipe and Riser Irrigation Efficiency (%)	60% (60–90%)	(Finger and Morris, 2005; Tennakoon et al., 2013)

5.1. *Extremity testing results*

To demonstrate the diagnostic use of OAT, interactions between the surface water and policy models are initially deactivated such that dam level calculations never account for the volume of water used by farms. Therefore, pipe and riser irrigation efficiency (the PoI) will not directly influence the long-term surface water allocation index (the QoI). Functionally, the deactivated parameter is equivalent to a dummy parameter.

In this example, an extremity test is applied by those integrating the models (the 'model integrators'). All 53 parameters are perturbed between their lowest and highest values with a cost of $N = 2$ model runs. Unexpected results (e.g., no change, or smaller-or-larger than expected change) indicate issues which modelers should investigate. Grouping parameters by their parent model components could also give at least an indication of which model in the IEM the issue stems from. The example results show that significantly higher volumes of water are allocated in the deactivated case, much more than what would be expected under "normal" circumstances (as shown in Fig. 5). The reason is that the farm water orders are never considered, and the dam is never depleted.

Such errors may inadvertently creep in during model development and integration. Examples include misunderstanding of the model interoperation (e.g., what outputs from one model relate to an input to another) implementation error (e.g., a bug in a model), or technical

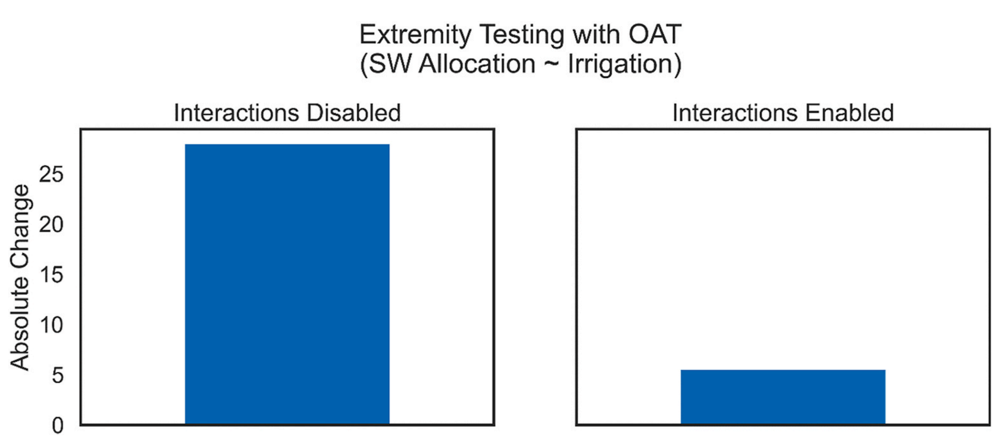


Fig. 5. Extremity testing results where all parameters are perturbed between their lower and upper bounds (i.e., $N = 2$). Change in surface water allocations in the disabled case far exceed what is possible from parameter perturbations alone.

issues (e.g., different compilers producing different machine code). It is acknowledged here that the presented approach is viable in cases where the rough order of magnitude effect is known. The results additionally indicate that the QoI will be affected even if the PoI is completely deactivated, suggesting that the QoI is affected by other factors. Thus, the conceptual understanding (depicted in Fig. 4) is not complete; there are other factors which influence the QoI. Example code, data, and figures presented in this paper are provided as supplementary material via the Open Science Framework (see Iwanaga, 2020; Iwanaga et al., 2020b). Further description of other model analyses conducted on the CIM can be found in Iwanaga et al. (2020a).

A single parameter can be targeted (i.e., $s = 1$), either after the above issue has been identified and further confirmation is desired, or where the relative change from perturbing all parameters is unknown. In this specific case, any perturbation of the PoI should be sufficient, as the behavioral property being tested for is the presence of change in the QoI when all other parameters are set to their nominal "best guess" values. An OAT test with a further two model runs is thus applied to the PoI (illustrated in Fig. 6). Because the interaction between models is disabled, the results show no change in long-term surface water allocation. It is re-emphasized here that the diagnostic property-based test targets areas of parameter space for which the PoI and QoI are expected to be sensitive, and that diagnostic applications of SA should be conducted alongside other testing processes.

5.2. A global approach to Property-based SA

In this example parameter activity/inactivity is used as a proxy to indicate correct model integration. The relationship between the PoI and QoI could be tested using R-OAT and Morris to confirm the presence of some sensitivity across parameter space. The R-OAT and Morris methods are applied here given that they are known to provide reliable indications with fewer samples compared to other GSAs (as noted in Section 2.1). Samples were generated by producing $n\cdot(p+1)$ parameter sets, such that n points in parameter space were sampled based on the targeting distribution.

As shown in Fig. 5, other factors may influence the QoI and so a non-zero sensitivity value is to be expected given that these GSA approaches report the average effect with parameter interactions. For this reason, we adapt the dummy threshold approach from Zadeh et al. (2017), wherein a parameter is considered insensitive if the reported sensitivity value is comparable to the sensitivities reported for the dummy parameter. In this case, we apply such a threshold to indicate an unexpected lack of activity: an "activity threshold".

An "activity threshold" of 0.1 is empirically set for this example, a value lower than expected sensitivities for the parameter in question for the available number of samples, but higher than typical sensitivity thresholds (e.g., 0.05; Sarrazin et al., 2016). As the PoI is expected to be active, its reported sensitivities should be above this threshold, and values lower than the threshold indicate a cause for concern. Testing for the property of parameter activity in this manner is more robust compared to searching for absolute inactivity as computational (precision) error – compounded as the models within the IEM continually interact – may introduce variability in results (Dunford et al., 2015). Unexpected interactions (based on modelers' current understanding of model interactions) may also cause non-zero sensitivities.

Such tests could be incorporated as part of an automated test suite. Existing property-based testing frameworks could also be leveraged to aid in pinpointing areas of parameter space wherein errors of concern occur (e.g., Löscher and Sagonas, 2017). The number of repetitions possible under the hypothetical 16-h time limit (i.e., 5pm to 9am) is $n = 2$, i.e., total number of possible model runs is $N = 108$ given that a model run takes roughly 30 min. Performing an additional repetition ($N = 162$ when $n = 3$) would exceed the available time limit, taking over 20 h. We take 540 model evaluations (i.e., $n = 10$) purely to illustrate response of μ^*.

In this example, indicative confirmation that the model is not behaving as expected could be obtained with $n \leq 2$ using R-OAT ($N = 108$, see the disabled case in Fig. 7). In general, how low n can be depends on the parameter and model context, and some initial experimentation is likely required. Similar results may be obtained with the Morris method (Fig. 8) in the disabled case, although in the active case concrete confirmation does not occur until $n = 9$ (i.e., $N = 486$). The results for both (i.e., no sensitivity at $N = 54$ for R-OAT and insufficient activity until $N = 486$ for Morris) suggest that the sampling scheme plays an important role in the efficacy of GSA methods for diagnostic purposes.

The diagnostic context severely limits the number of samples that could be obtained in a timely manner and, for this reason, other GSA methods were not wholly considered. Preliminary results are included in Appendix A for the Saltelli (2002), EASI (Plischke, 2010), and DMIM (Plischke et al., 2013) methods, which indicate the unreliability of GSA methods at such low sample sizes. These results indicate potential issues to be overcome if these methods are to be applied for diagnostic purposes in the context of rapid, iterative, model development and testing. In cases where no issues are identified, it is desirable for obtained samples to be reused to conserve computational budget.

6. Discussion and conclusions

This paper outlined a role SA can play in software testing practices in the IEM development process. Specifically, local OAT analyses coupled with R-OAT and/or Morris can provide a first-pass indication of the correctness of technical and conceptual integration of constituent

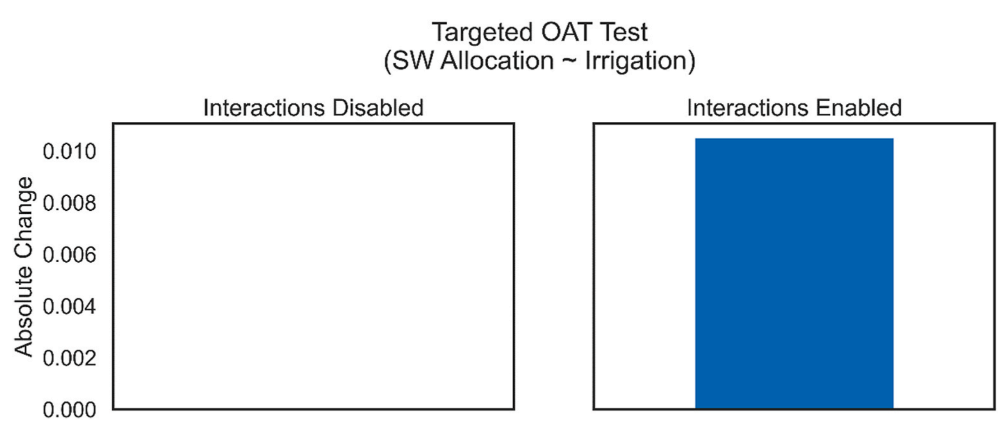


Fig. 6. Change in results from a targeted OAT testing ($N = 2$). Given the model context, we would expect to see even minor changes. Analysis of model runs with the surface water model interactions disabled (left hand side) correctly finds no parameter effect on surface water allocation.

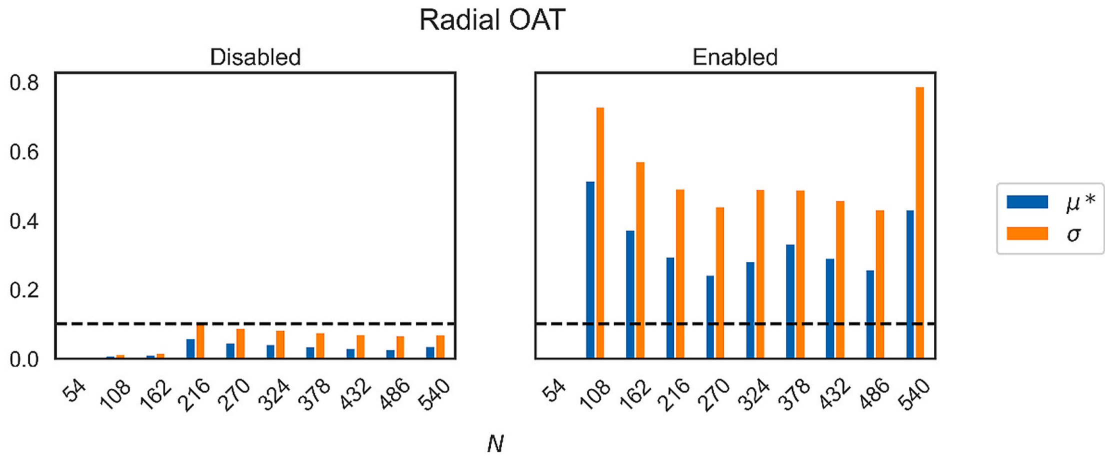


Fig. 7. Elementary Effects (EE) analysis on Radial OAT (R-OAT) samples demonstrating a lack of effect on the QoI (surface water allocations). X-axis refers to the total number of model evaluations. R-OAT was able to determine expected behavior at $N = 108$ (i.e., $n = 2$) model evaluations for both active and inactive cases. Dashed line indicates the parameter activity threshold of 0.1.

models, particularly in terms of checking the effects of active/inactive parameters on (expected) model behavior. Including such property-based diagnostics as part of an automated test suite can aid in conserving a limited computational budget, which is often desirable even in cases where there is an abundance of computational time available.

With the Campaspe Integrated Model used as an example, the lack of an expected relationship between a PoI and QoI could be identified using extremity parameter value testing with just $N = 2$ runs when an activity threshold is applied. Although global sensitivity analysis methods can be computationally demanding, the use of R-OAT is shown to be a computationally efficient approach to assessing expected behavior. Although the Morris method was able to identify inactivity of the parameter of interest, it required more model runs to do so, at least in the presented example case. It is emphasized here that for the purpose of integrated model testing, the number of parameters to be perturbed could be further reduced in all approaches described through qualitative assessment that identifies which parameters would have system-level (or inter-system) implications. Targeting these parameters, or organizing these into groups, such that the number of perturbations is much fewer than the total number of model parameters would reduce the overall computational effort involved to gain an indicative result.

Additionally, a failing test may negate the need to conduct further diagnostics as the assumption that the model is operating correctly is falsified, thereby aiding in conserving computational budget. More complete analyses could follow in cases where no Property-based SA tests fail. In a "full" property-based testing approach, the framework applied would generate a random set of inputs and iteratively narrow the parameter space to specific areas that cause unexpected model behavior (Löscher and Sagonas, 2017). Such tests could be augmented to use GSA methods that require comparatively limited number of samples, such as R-OAT. Use of 'given data' methods such as PAWN (Pianosi and Wagener, 2015), HDMR (Li et al., 2002) or methods with flexible sampling requirements such as STAR-VARS (Razavi et al., 2019) could also be explored to identify potential advantages and limitations (e.g., Puy et al., 2020).

These "given data" methods may be more suitable in the IEM context due to their ability to leverage available samples, and may also be used to complement any diagnostic analyses conducted towards a comprehensive GSA (Mora et al., 2019). The use of dummy parameters in combination with extremity testing under conditions in which model parameters (or targeted subset of parameters) are known to be active could also be explored. Alternate OAT-based global analyses that are potentially more efficient for obtaining indications of parameter interaction (e.g., Borgonovo, 2010), may also be beneficial.

There are many approaches to validating computational models. Model developers can adopt a mix of testing practices from both software engineering and statistical/mathematical analysis to cover the range of issues that may occur during model development. Ideally, modelers would not restrict themselves to techniques found in one

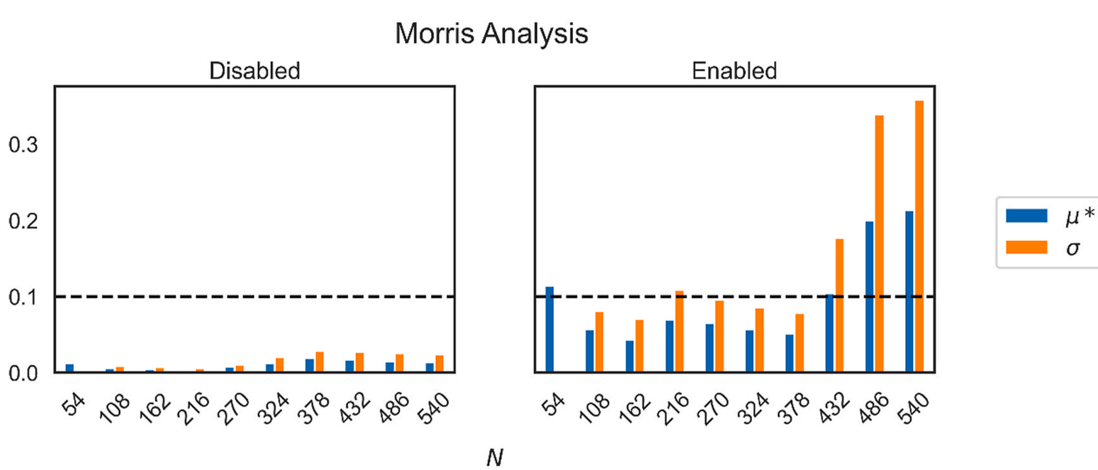


Fig. 8. Similar to R-OAT, the Morris method is capable of determining insensitivity in the disabled case but required 486 evaluations (i.e., $n = 9$) to confirm correct behavior in the enabled case. Dashed line indicates the parameter activity threshold of 0.1.

discipline over the other. There are, however, barriers to the adoption of this hybrid approach. For one, it requires the technical knowledge and capacity of modelers to develop and maintain tests, including the application of relevant SA techniques.

It is demonstrated here that a diagnostic property-based testing approach with SA methods is a useful, pragmatic, and computationally efficient approach to providing a line of evidence that the model parameters are, in fact, having an (expected) effect. In the IEM development context, any single model may itself require teams of domain specialists to fully understand, and no single person can be expected to grasp, all aspects, especially in cases where legacy models are adopted. Assessing the expected behavioral properties of a model could be leveraged to reduce the time taken to identify and correct model implementation and integration errors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors acknowledge support from the Australian Government Research Training Program (Takuya Iwanaga), ANU Hilda John Endowment fund (Takuya Iwanaga, Xifu Sun and Barry Croke), China Scholarship Council (Qian Wang) and the Australian Research Council Discovery Early Career Researcher Award (Joseph Guillaume; Project No. DE190100317). The Campaspe Integrated Model was developed as part of the Murray-Darling Basin Authority's partnership with the National Centre for Groundwater Research and Training (NCGRT) under Contract No. MD2594. The authors would like to thank the reviewers for their comments and insight, which greatly improved the quality and clarity of the paper.

Appendix A

Here, we showcase some preliminary results with three global sensitivity analysis approaches, namely Saltelli (2010), EASI (2010), and DMIM (2013). Unreliable results are obtained for all approaches at these relatively low sample sizes. EASI and DMIM are "given data" approaches, for which Morris samples are used. For the Saltelli method, only first order indices ($S1$) are shown (using the approach described in Saltelli et al., 2010) estimated with a cost of $n \cdot (p+2)$. In practice, total, first and second order indices may be estimated at a cost of $n \cdot (2p+2)$ runs (Saltelli, 2002). The results for the Saltelli analysis include negative values (Fig. 9), which indicate an insufficient number of samples (Saltelli, 2008; Sharifi et al., 2019), which is to be expected given the known high sampling requirements of Sobol'-based approaches (Razavi and Gupta, 2015).

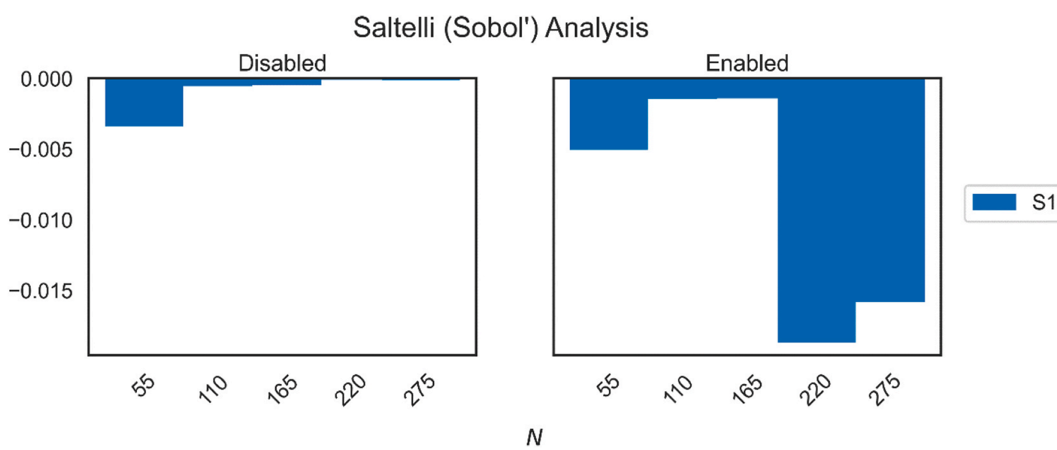


Fig. 9. Negative first-order sensitivity values (S1) from the Saltelli method analysis for both inactive and active cases.

The EASI analysis technique indicates that an effect is occurring when interactions are disabled (i.e., Type I error, as shown in Fig. 10). While the EASI approach does not require a specific sampling scheme (Plischke et al., 2013), the results produced may be sensitive to the sampling approach. Very little difference was found between disabled and enabled cases when results were obtained with Morris sampling (see Fig. 11). In both cases EASI was unable to distinguish the (lack of) effect of an inactive parameter at these low sample sizes.

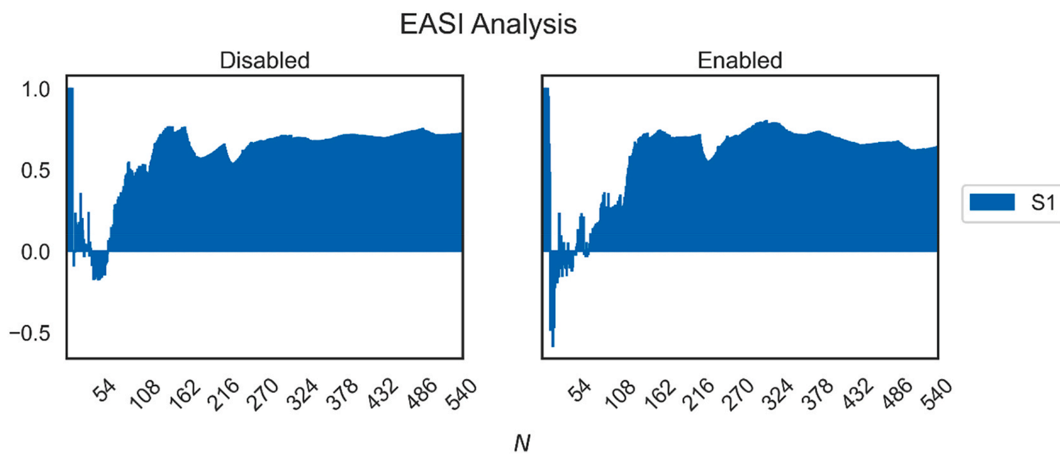


Fig. 10. First order effect (S1) of irrigation efficiency on surface water allocations using OAT samples. EASI analysis indicates the parameter is sensitive where model interactions are disabled.

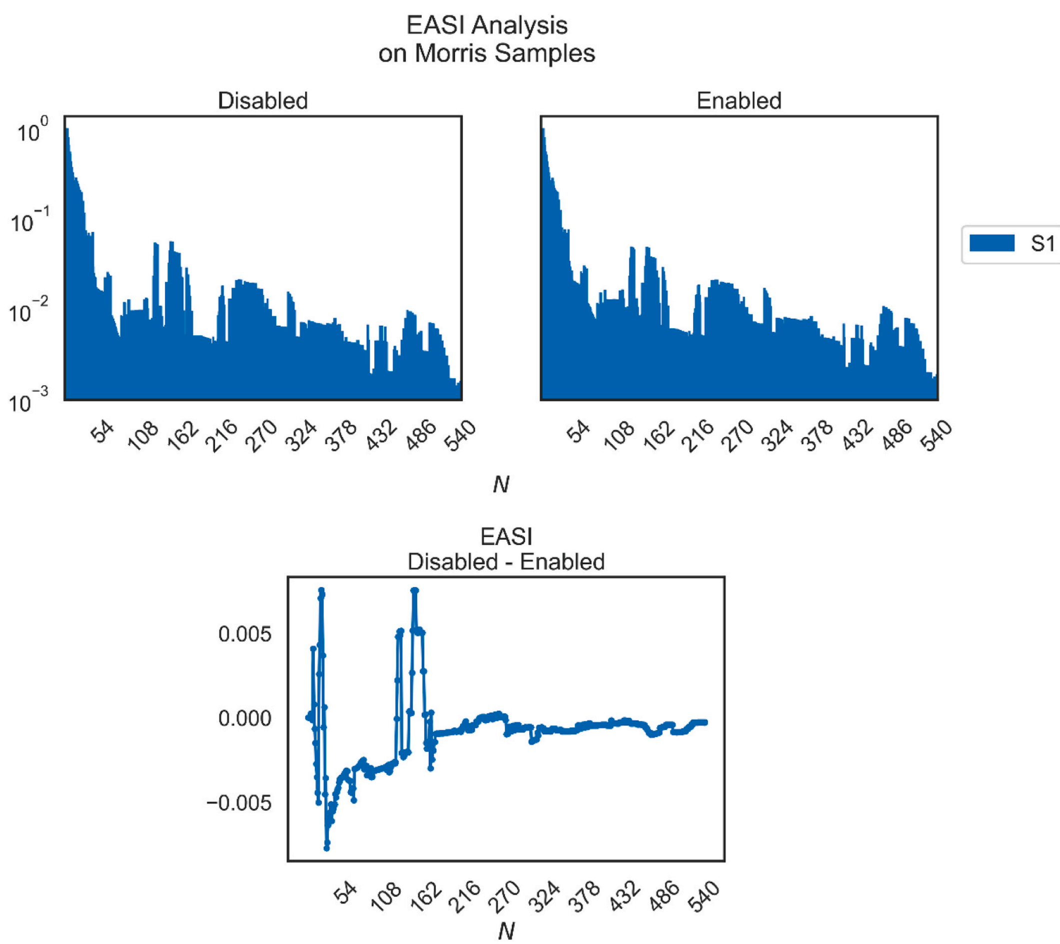


Fig. 11. Example of EASI analysis on results taken with Morris sampling. Results of first-order sensitivities (S1) appear near identical indicating larger Morris samples are necessary for a distinction to be made with EASI. Note that the y-axis for the top panel is in log scale.

DMIM was unable determine the lack of influence from the PoI (see Fig. 12 and Fig. 13) with similar issues.

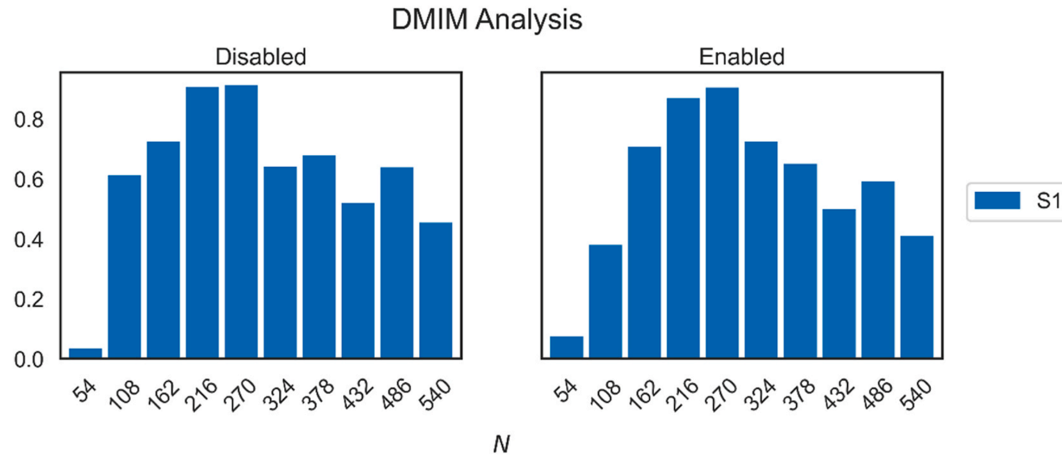


Fig. 12. DMIM analysis on OAT samples.

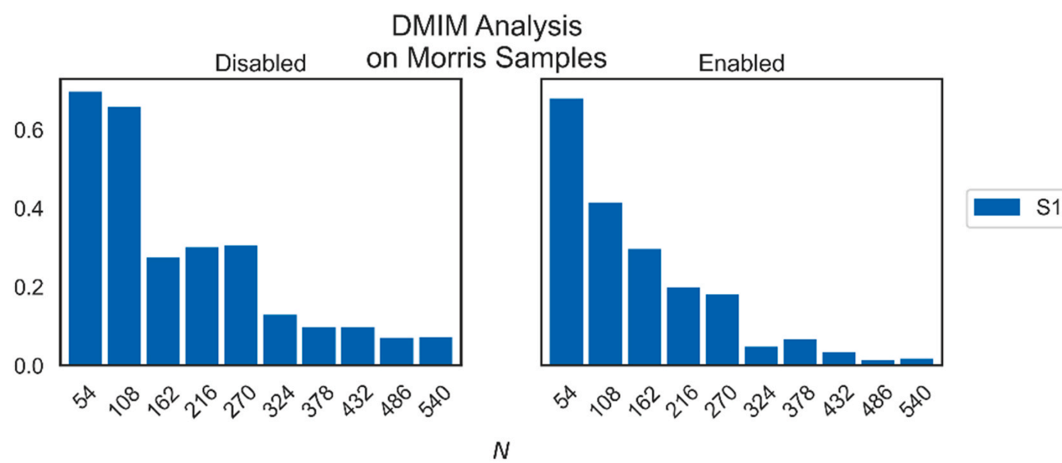


Fig. 13. DMIM analysis on Morris samples.

When compared to the results of OAT-based GSA (Morris and R-OAT in the main text), it appears that methods which indirectly estimate first-order sensitivity while varying multiple parameters at once do not correctly identify inactive parameters, at least at the given sample sizes. Furthermore, a significantly larger number of model evaluations may be required to achieve convergence which may exceed the available time for diagnostic testing, particularly in the case of complex and highly parameterized IEMs.

References

Ahalt, S., Band, L., Christopherson, L., Idaszak, R., Lenhardt, C., Minsker, B., Palmer, M., Shelley, M., Tiemann, M., Zimmerman, A., 2014. Water science software institute: agile and open source scientific software development. Comput. Sci. Eng. 16, 18–26. https://doi.org/10.1109/MCSE.2014.5.

Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J. Hydrol. 249, 11–29. https://doi.org/10.1016/S0022-1694(01)00421-8.

Bhandari Neupane, J., Neupane, R.P., Luo, Y., Yoshida, W.Y., Sun, R., Williams, P.G., 2019. Characterization of leptazolines A–D, polar oxazolines from the Cyanobacterium leptolyngbya sp., reveals a glitch with the "Willoughby–Hoye" scripts for calculating NMR chemical shifts. Org. Lett. 21, 8449–8453. https://doi.org/10.1021/acs.orglett.9b03216.

Boehm, B., 1986. A spiral model of software development and enhancement. SIGSOFT Softw. Eng. Notes 11, 14–24. https://doi.org/10.1145/12944.12948.

Borgonovo, E., 2010. Sensitivity analysis with finite changes: an application to modified EOQ models. Eur. J. Oper. Res. 200, 127–138. https://doi.org/10.1016/j.ejor.2008.12.025.

Braddock, R.D., Schreider, S.Yu, 2006. Application of the Morris algorithm for sensitivity analysis of the REALM model for the Goulburn irrigation system. Environ. Model. Assess. 11, 297–313. https://doi.org/10.1007/s10666-005-9029-z.

Campolongo, F., Cariboni, J., Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. Environ. Modell. Softw. Model. Comput. Assist. Simul. Map. Danger. Phenom. Hazard Assess. 22, 1509–1518. https://doi.org/10.1016/j.envsoft.2006.10.004.

Campolongo, F., Saltelli, A., Cariboni, J., 2011. From screening to quantitative sensitivity analysis. A unified approach. Comput. Phys. Commun. 182, 978–988. https://doi.org/10.1016/j.cpc.2010.12.039.

Claessen, K., Hughes, J., 2000. QuickCheck: a lightweight tool for random testing of Haskell programs. In: Proceedings of the Fifth ACM SIGPLAN International Conference on Functional Programming, ICFP '00. Association for Computing Machinery, New York, NY, USA, pp. 268–279. https://doi.org/10.1145/351240.351266.

Crouch, S., Hong, N.C., Hettrick, S., Jackson, M., Pawlik, A., Sufi, S., Carr, L., Roure, D. D., Goble, C., Parsons, M., 2013. The software sustainability institute: changing research software attitudes and practices. Comput. Sci. Eng. 15, 74–80. https://doi.org/10.1109/MCSE.2013.133.

Cuntz, M., Mai, J., Zink, M., Thober, S., Kumar, R., Schäfer, D., Schrön, M., Craven, J., Rakovec, O., Spieler, D., Prykhodko, V., Dalmasso, G., Musuuza, J., Langenberg, B., Attinger, S., Samaniego, L., 2015. Computationally inexpensive identification of noninformative model parameters by sequential screening. Water Resour. Res. 51, 6417–6441. https://doi.org/10.1002/2015WR016907.

Danglot, B., Monperrus, M., Rudametkin, W., Baudry, B., 2020. An approach and benchmark to detect behavioral changes of commits in continuous integration. Empir. Softw. Eng. https://doi.org/10.1007/s10664-019-09794-7.

Davidson-Pilon, C., 2016. Bayesian Methods for Hackers: Probabilistic Programming and Bayesian Inference. Addison-Wesley, New York.

Douglas-Smith, D., Iwanaga, T., Croke, B.F.W., Jakeman, A.J., 2020. Certain trends in uncertainty and sensitivity analysis: an overview of software tools and techniques. Environ. Model. Software 124, 104588. https://doi.org/10.1016/j.envsoft.2019.104588.

Dunford, R., Harrison, P.A., Rounsevell, M.D.A., 2015. Exploring scenario and model uncertainty in cross-sectoral integrated assessment approaches to climate change impacts. Climatic Change 132, 417–432. https://doi.org/10.1007/s10584-014-1211-3.

Easterbrook, S.M., 2014. Open code for open science? Nat. Geosci. 7, 779–781. https://doi.org/10.1038/ngeo2283.

Elsawah, S., Filatova, T., Jakeman, A.J., Kettner, A.J., Zellner, M.L., Athanasiadis, I.N., Hamilton, S.H., Axtell, R.L., Brown, D.G., Gilligan, J.M., Janssen, M.A., Robinson, D. T., Rozenberg, J., Ullah, I.I.T., Lade, S.J., 2020. Eight Grand Challenges in Socio-Environmental Systems Modeling, 1 2. https://doi.org/10.18174/sesmo.2020a16226, 16226–16226.

Ferretti, F., Saltelli, A., Tarantola, S., 2016. Trends in sensitivity analysis practice in the last decade. Sci. Total Environ. 568, 666–670. https://doi.org/10.1016/j.scitotenv.2016.02.133.

Finger, L., Morris, M., 2005. Water Management Options: Assisting Irrigators with Stream Flow Management Plan Implementation. Department of Primary Industries (DPI).

Fink, G., Bishop, M., 1997. Property-based testing: a new approach to testing for assurance. SIGSOFT Softw. Eng. Notes 22, 74–80. https://doi.org/10.1145/263244.263267.

Gray, C.T., Marwick, B., 2019. Truth, proof, and reproducibility: there's No counter-attack for the codeless. In: Nguyen, H. (Ed.), Statistics and Data Science, Communications in Computer and Information Science. Springer, Singapore, pp. 111–129. https://doi.org/10.1007/978-981-15-1960-4_8.

Grimm, V., 2005. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science 310, 987–991. https://doi.org/10.1126/science.1116681.

Grimm, V., Railsback, S.F., 2012. Pattern-oriented modelling: a 'multi-scope' for predictive systems ecology. Phil. Trans. R. Soc. B 367, 298–310. https://doi.org/10.1098/rstb.2011.0180.

Guillaume, J.H.A., Jakeman, J.D., Marsili-Libelli, S., Asher, M., Brunner, P., Croke, B., Hill, M.C., Jakeman, A.J., Keesman, K.J., Razavi, S., Stigter, J.D., 2019. Introductory overview of identifiability analysis: a guide to evaluating whether you have the right type of data for your modeling purpose. Environ. Model. Software 119, 418–432. https://doi.org/10.1016/j.envsoft.2019.07.007.

Gupta, H.V., Wagener, T., Liu, Y., 2008. Reconciling theory with observations: elements of a diagnostic approach to model evaluation. Hydrol. Process. 22, 3802–3813. https://doi.org/10.1002/hyp.6989.

Hamby, D.M., 1994. A review of techniques for parameter sensitivity analysis of environmental models. Environ. Monit. Assess. 32, 135–154. https://doi.org/10.1007/BF00547132.

Hamilton, S.H., ElSawah, S., Guillaume, J.H.A., Jakeman, A.J., Pierce, S.A., 2015. Integrated assessment and modelling: overview and synthesis of salient dimensions. Environ. Model. Software 64, 215–229. https://doi.org/10.1016/j.envsoft.2014.12.005.

Hannay, J.E., MacLeod, C., Singer, J., Langtangen, H.P., Pfahl, D., Wilson, G., 2009. How do scientists develop and use scientific software?. In: 2009 ICSE Workshop on Software Engineering for Computational Science and Engineering. IEEE, pp. 1–8. https://doi.org/10.1109/SECSE.2009.5069155.

Herman, J., Usher, W., 2017. SALib: an open-source Python library for Sensitivity Analysis. J. Open Source Software 2, 97. https://doi.org/10.21105/joss.00097.

Herman, J.D., Kollat, J.B., Reed, P.M., Wagener, T., 2013. Technical Note: method of Morris effectively reduces the computational demands of global sensitivity analysis for distributed watershed models. Hydrol. Earth Syst. Sci. 17, 2893–2903. https://doi.org/10.5194/hess-17-2893-2013.

Huizinga, D., Kolawa, A., 2007. Automated Defect Prevention: Best Practices in Software Management.

Hut, R.W., van de Giesen, N.C., Drost, N., 2017. Comment on "Most computational hydrology is not reproducible, so is it really science?" by Christopher Hutton et al.: let hydrologists learn the latest computer science by working with Research Software Engineers (RSEs) and not reinvent the waterwheel our. Water Resour. https://doi.org/10.1002/2017WR020665.

Hutton, C., Wagener, T., Freer, J., Han, D., Duffy, C., Arheimer, B., 2016. Most computational hydrology is not reproducible, so is it really science? Water Resour. Res. 52, 7548–7555. https://doi.org/10.1002/2016WR019285.

Iwanaga, T., 2020. diagnostic-sa. https://doi.org/10.17605/OSF.IO/NUZAF.

Iwanaga, T., Partington, D., Ticehurst, J., Croke, B.F.W., Jakeman, A.J., 2020a. A socio-environmental model for exploring sustainable water management futures: participatory and collaborative modelling in the Lower Campaspe catchment. J. Hydrol.: Reg. Stud. 28, 100669. https://doi.org/10.1016/j.ejrh.2020.100669.

Iwanaga, T., Sun, X., Wang, Q., 2020b. Oat-Diagnostic. OSF. https://doi.org/10.17605/OSF.IO/85EDC.

Iwanaga, T., Wang, H.-H., Hamilton, S.H., Grimm, V., Koralewski, T.E., Salado, A., Elsawah, S., Razavi, S., Yang, J., Glynn, P., Badham, J., Voinov, A., Chen, M., Grant, W.E., Peterson, T.R., Frank, K., Shenk, G., Barton, C.M., Jakeman, A.J., Little, J.C., 2021. Socio-technical scales in socio-environmental modeling: managing a system-of-systems modeling approach. Environ. Model. Software 104885. https://doi.org/10.1016/j.envsoft.2020.104885.

Iwanaga, T., Zare, F., Croke, B., Fu, B., Merritt, W., Partington, D., Ticehurst, J., Jakeman, A., 2018. Development of an integrated model for the Campaspe catchment: a tool to help improve understanding of the interaction between society, policy, farming decision, ecology, hydrology and climate. In: Proceedings of the International Association of Hydrological Sciences. Presented at the Innovative Water Resources Management - Understanding and Balancing Interactions between Humankind and Nature - 8th International Water Resources Management Conference of ICWRS, Beijing, China, 13-15 June 2018, Copernicus GmbH, pp. 1–12. https://doi.org/10.5194/piahs-379-1-2018.

Kanewala, U., Bieman, J.M., 2014. Testing scientific software: a systematic literature review. Inf. Software Technol. 56, 1219–1232. https://doi.org/10.1016/j.infsof.2014.05.006.

Kelly Letcher, R.A., Jakeman, A.J., Barreteau, O., Borsuk, M.E., ElSawah, S., Hamilton, S. H., Henriksen, H.J., Kuikka, S., Maier, H.R., Rizzoli, A.E., van Delden, H., Voinov, A. A., 2013. Selecting among five common modelling approaches for integrated environmental assessment and management. Environ. Model. Software 47, 159–181. https://doi.org/10.1016/j.envsoft.2013.05.005.

Krekel, H., Oliveira, B., Pfannschmidt, R., Bruynooghe, F., Laugher, B., Bruhin, F., 2004. pytest 6.1.2. https://github.com/pytest-dev/pytest.

Knapen, R., Janssen, S., Roosenschoon, O., Verweij, P., de Winter, W., Uiterwijk, M., Wien, J.E., 2013. Evaluating OpenMI as a model integration platform across disciplines. Environ. Model. Software 39, 274–282. https://doi.org/10.1016/j.envsoft.2012.06.011.

Laukkanen, E., Itkonen, J., Lassenius, C., 2017. Problems, causes and solutions when adopting continuous delivery—a systematic literature review. Inf. Software Technol. 82, 55–79. https://doi.org/10.1016/j.infsof.2016.10.001.

Levin, M.A., Kalal, T.T., Rodin, J., 2019. Improving Product Reliability and Software Quality: Strategies, Tools, Process and Implementation. John Wiley & Sons.

Li, G., Rabitz, H., Yelvington, P.E., Oluwole, O.O., Bacon, F., Kolb, C.E., Schoendorf, J., 2010. Global sensitivity analysis for systems with independent and/or correlated inputs. J. Phys. Chem. A 114, 6022–6032. https://doi.org/10.1021/jp9096919.

Li, G., Wang, S.-W., Rabitz, H., Wang, S., Jaffé, P., 2002. Global uncertainty assessments by high dimensional model representations (HDMR). Chem. Eng. Sci. 57, 4445–4460. https://doi.org/10.1016/S0009-2509(02)00417-7.

Little, J.C., Hester, E.T., Elsawah, S., Filz, G.M., Sandu, A., Carey, C.C., Iwanaga, T., Jakeman, A.J., 2019. A tiered, system-of-systems modeling framework for resolving complex socio-environmental policy issues. Environ. Model. Software 112, 82–94. https://doi.org/10.1016/j.envsoft.2018.11.011.

Löscher, A., Sagonas, K., 2017. Targeted property-based testing. In: Proceedings of the 26th ACM SIGSOFT International Symposium on Software Testing and Analysis, ISSTA 2017. Association for Computing Machinery, New York, NY, USA, pp. 46–56. https://doi.org/10.1145/3092703.3092711.

MacLeod, M., Nagatsu, M., 2018. What does interdisciplinarity look like in practice: mapping interdisciplinarity and its limits in the environmental sciences. Stud. Hist. Philos. Sci. 67, 74–84. https://doi.org/10.1016/j.shpsa.2018.01.001.

Mai, J., Cuntz, M., 2020. Computationally inexpensive identification of noninformative model parameters by sequential screening: efficient elementary effects (EEE) (v1.0). Zenodo. https://doi.org/10.5281/ZENODO.3620895.

Mora, E.B., Spelling, J., van der Weijde, A.H., 2019. Benchmarking the PAWN distribution-based method against the variance-based method in global sensitivity analysis: empirical results. Environ. Model. Software 122, 104556. https://doi.org/10.1016/j.envsoft.2019.104556.

Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments. Technometrics 33, 161–174. https://doi.org/10.1080/00401706.1991.10484804.

Mossalam, A., 2018. Projects' issue management. HBRC J. 14, 400–407. https://doi.org/10.1016/j.hbrcj.2017.12.001.

Norton, J.P., 2009. Selection of Morris trajectories for initial sensitivity analysis. IFAC Proceedings Volumes, 15th IFAC Symposium on System Identification 42, 670–674. https://doi.org/10.3182/20090706-3-FR-2004.00111.

Oakley, J.E., O'Hagan, A., 2004. Probabilistic sensitivity analysis of complex models: a Bayesian approach. J. Roy. Stat. Soc. B 66, 751–769.

Pianosi, F., Beven, K., Freer, J., Hall, J.W., Rougier, J., Stephenson, D.B., Wagener, T., 2016. Sensitivity analysis of environmental models: a systematic review with practical workflow. Environ. Model. Software 79, 214–232. https://doi.org/10.1016/j.envsoft.2016.02.008.

Pianosi, F., Wagener, T., 2018. Distribution-based sensitivity analysis from a generic input-output sample. Environ. Model. Software 108, 197–207. https://doi.org/10.1016/j.envsoft.2018.07.019.

Pianosi, F., Wagener, T., 2015. A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. Environ. Model. Software 67, 1–11. https://doi.org/10.1016/j.envsoft.2015.01.004.

Plischke, E., 2010. An effective algorithm for computing global sensitivity indices (EASI). Reliab. Eng. Syst. Saf. 95, 354–360. https://doi.org/10.1016/j.ress.2009.11.005.

Plischke, E., Borgonovo, E., Smith, C.L., 2013. Global sensitivity measures from given data. Eur. J. Oper. Res. 226, 536–550. https://doi.org/10.1016/j.ejor.2012.11.047.

Puy, A., Lo Piano, S., Saltelli, A., 2020. A sensitivity analysis of the PAWN sensitivity index. Environ. Model. Software 127, 104679. https://doi.org/10.1016/j.envsoft.2020.104679.

Rabitz, H., 1989. Systems analysis at the molecular scale. Science 246, 221–226. https://doi.org/10.1126/science.246.4927.221.

Razavi, S., Gupta, H.V., 2015. What do we mean by sensitivity analysis? The need for comprehensive characterization of "global" sensitivity in earth and environmental systems models: a critical look at sensitivity analysis. Water Resour. Res. 51, 3070–3092. https://doi.org/10.1002/2014WR016527.

Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., Plischke, E., Lo Piano, S., Iwanaga, T., Becker, W., Tarantola, S., Guillaume, J.H.A., Jakeman, J., Gupta, H., Melillo, N., Rabitti, G., Chabridon, V., Duan, Q., Sun, X., Smith, S., Sheikholeslami, R., Hosseini, N., Asadzadeh, M., Puy, A., Kucherenko, S., Maier, H. R., 2020. The future of sensitivity analysis: an essential discipline for systems modeling and policy support. Environ. Model. Software 104954. https://doi.org/10.1016/j.envsoft.2020.104954.

Razavi, S., Sheikholeslami, R., Gupta, H.V., Haghnegahdar, A., 2019. VARS-TOOL: a toolbox for comprehensive, efficient, and robust sensitivity and uncertainty analysis. Environ. Model. Software 112, 95–107. https://doi.org/10.1016/j.envsoft.2018.10.005.

Reiter, R., 1987. A theory of diagnosis from first principles. Artif. Intell. 32, 57–95. https://doi.org/10.1016/0004-3702(87)90062-2.

Ruano, M.V., Ribes, J., Seco, A., Ferrer, J., 2012. An improved sampling strategy based on trajectory design for application of the Morris method to systems with many input factors. Environ. Model. Software 37, 103–109. https://doi.org/10.1016/j.envsoft.2012.03.008.

Saltelli, A. (Ed.), 2008. Sensitivity Analysis, Paperback ed. Wiley paperback series. Wiley, Chichester.

Saltelli, A., 2002. Making best use of model evaluations to compute sensitivity indices. Comput. Phys. Commun. 145, 280–297. https://doi.org/10.1016/S0010-4655(02)00280-1.

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., Wu, Q., 2019. Why so many published sensitivity analyses are false: a systematic review of sensitivity analysis practices. Environ. Model. Software 114, 29–39.

Saltelli, A., Annoni, P., 2010. How to avoid a perfunctory sensitivity analysis. Environ. Model. Software 25, 1508–1517. https://doi.org/10.1016/j.envsoft.2010.04.012.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S., 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comput. Phys. Commun. 181, 259–270. https://doi.org/10.1016/j.cpc.2009.09.018.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. Global Sensitivity Analysis. The Primer. Wiley, West Sussex, U. K.

Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., 2004. Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models. Halsted Press, New York, NY, USA.

Samuelson, W., Zeckhauser, R., 1988. Status quo bias in decision making. J. Risk Uncertain. 1, 7–59. https://doi.org/10.1007/BF00055564.

Sarma, G.P., Jacobs, T.W., Watts, M.D., Ghayoomie, S.V., Larson, S.D., Gerkin, R.C., 2016. Unit Testing, Model Validation, and Biological Simulation. F1000Res 5. https://doi.org/10.12688/f1000research.9315.1.

Sarrazin, F., Pianosi, F., Wagener, T., 2016. Global sensitivity analysis of environmental models: convergence and validation. Environ. Model. Software 79, 135–152. https://doi.org/10.1016/j.envsoft.2016.02.005.

Sharifi, A., Ahmadi, M., Badfar, H., Hosseini, M., 2019. Modeling and sensitivity analysis of NOx emissions and mechanical efficiency for diesel engine. Environ. Sci. Pollut. Res. 26, 25190–25207. https://doi.org/10.1007/s11356-019-05613-0.

Shin, M.-J., Guillaume, J.H.A., Croke, B.F.W., Jakeman, A.J., 2015. A review of foundational methods for checking the structural identifiability of models: results for rainfall-runoff. J. Hydrol. 520, 1–16. https://doi.org/10.1016/j.jhydrol.2014.11.040.

Sieber, A., Uhlenbrook, S., 2005. Sensitivity analyses of a distributed catchment model to verify the model structure. J. Hydrol. 310, 216–235. https://doi.org/10.1016/j.jhydrol.2005.01.004.

Sletholt, M.T., Hannay, J.E., Pfahl, D., Langtangen, H.P., 2012. What Do We Know about Scientific Software Development's Agile Practices?, pp. 24–36. https://doi.org/10.1109/MCSE.2011.113.

Sudret, B., 2008. Global sensitivity analysis using polynomial chaos expansions. Reliab. Eng. Syst. Safety Bayesian Netw. Depend. 93, 964–979. https://doi.org/10.1016/j.ress.2007.04.002.

Sun, X.Y., Newham, L.T.H., Croke, B.F.W., Norton, J.P., 2012. Three complementary methods for sensitivity analysis of a water quality model. Environ. Model. Software 37, 19–29. https://doi.org/10.1016/j.envsoft.2012.04.010.

Tennakoon, S., Richards, D., Milroy, S., Harris, G., 2013. Water use efficiency in the Australian cotton industry. In: Waterpak: A Guide for Irrigation Management in Cotton and Grain Farming Systems. Cotton Research and Development Corporation, pp. 22–27.

Thomas, J., McDonagh, D., 2013. Shared language:Towards more effective communication. Australas. Med. J. 6, 46–54. https://doi.org/10.4066/AMJ.2013.1596.

Vanrolleghem, P.A., Mannina, G., Cosenza, A., Neumann, M.B., 2015. Global sensitivity analysis for urban water quality modelling: terminology, convergence and comparison of different methods. J. Hydrol. 522, 339–352. https://doi.org/10.1016/j.jhydrol.2014.12.056.

Verweij, P.J.F.M., Knapen, M.J.R., de Winter, W.P., Wien, J.J.F., te Roller, J.A., Sieber, S., Jansen, J.M.L., 2010. An IT perspective on integrated environmental modelling: the SIAT case. Ecol. Model. 221, 2167–2176. https://doi.org/10.1016/J.ECOLMODEL.2010.01.006.

Voinov, A., Shugart, H.H., 2013. "Integronsters", integral and integrated modeling. Environ. Model. Software 39, 149–158. https://doi.org/10.1016/j.envsoft.2012.05.014.

Wagener, T., Pianosi, F., 2019. What has Global Sensitivity Analysis ever done for us? A systematic review to support scientific advancement and to inform policy-making in earth system modelling. Earth Sci. Rev. 194, 1–18. https://doi.org/10.1016/j.earscirev.2019.04.006.

Walker, W.E., Harremoës, P., Rotmans, J., Sluijs, J.P. van der, van Asselt, M.B.A., Janssen, P., Krauss, von, M.P.K., 2003. Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. Integrated Assess. 4, 5–17. https://doi.org/10.1076/iaij.4.1.5.16466.

Wilson, C., Kriegler, E., van Vuuren, D.P., Guivarch, C., Frame, D., Krey, V., Osborn, T.J., Schwanitz, V.J., Thompson, E.L., 2017. Evaluating process-based integrated assessment models of climate change mitigation [WWW Document]. URL. http://pure.iiasa.ac.at/id/eprint/14502/. (Accessed 16 December 2019).

Yang, J., 2011. Convergence and uncertainty analyses in Monte-Carlo based sensitivity analysis. Environ. Model. Software 26, 444–457. https://doi.org/10.1016/j.envsoft.2010.10.007.

Yang, J., Jakeman, A., Fang, G., Chen, X., 2018. Uncertainty analysis of a semi-distributed hydrologic model based on a Gaussian Process emulator. Environ. Model. Software 101, 289–300. https://doi.org/10.1016/j.envsoft.2017.11.037.

Yoo, S., Harman, M., 2012. Regression testing minimization, selection and prioritization: a survey. Softw. Test. Verif. Reliab. 22, 67–120. https://doi.org/10.1002/stvr.430.

Zadeh, F.K., Nossent, J., Sarrazin, F., Pianosi, F., van Griensven, A., Wagener, T., Bauwens, W., 2017. Comparison of variance-based and moment-independent global sensitivity analysis approaches by application to the SWAT model. Environ. Model. Software 91, 210–222. https://doi.org/10.1016/j.envsoft.2017.02.001.

Chapter 6: Socio-technical scales in socio-environmental modelling

This chapter explores issues related to scale in the modelling of SES. A multi-disciplinary socio-technical lens is used to identify and articulate the practices, issues and challenges that arise when dealing with the various influences and effects of scale. The term “system-of-systems” is more readily adopted as part of the framing in lieu of the usual terms found in the environmental modelling disciplines, acting as a bridge to the systems engineering perspective.

This chapter was published in Environmental Modelling and Software, also as a Position Paper, after undergoing review by four reviewers.

Iwanaga, T.*, Wang, H.-H., Hamilton, S.H., Grimm, V., Koralewski, T.E., Salado, A., Elsawah, S., Razavi, S., Yang, J., Glynn, P., Badham, J., Voinov, A., Chen, M., Grant, W.E., Peterson, T.R., Frank, K., Shenk, G., Barton, C.M., Jakeman, A.J., Little, J.C., 2021. Socio-technical scales in socio-environmental modeling: managing a system-of-systems modeling approach. Environmental Modelling & Software 104885. https://doi.org/10.1016/j.envsoft.2020.104885

Environmental Modelling and Software 135 (2021) 104885

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: http://www.elsevier.com/locate/envsoft

Socio-technical scales in socio-environmental modeling: Managing a system-of-systems modeling approach

Takuya Iwanaga [a,*], Hsiao-Hsuan Wang [b], Serena H. Hamilton [a,c], Volker Grimm [d,e], Tomasz E. Koralewski [b], Alejandro Salado [f], Sondoss Elsawah [a,g], Saman Razavi [h], Jing Yang [i], Pierre Glynn [j], Jennifer Badham [k], Alexey Voinov [l,m], Min Chen [n], William E. Grant [b], Tarla Rai Peterson [o], Karin Frank [d], Gary Shenk [p], C. Michael Barton [q], Anthony J. Jakeman [a], John C. Little [r]

[a] *Institute for Water Futures and Fenner School of Environment and Society, The Australian National University, Canberra, Australia*
[b] *Ecological Systems Laboratory, Department of Ecology and Conservation Biology, Texas A&M University, College Station, TX, 77843, USA*
[c] *CSIRO Land & Water, Canberra, Australia*
[d] *Helmholtz Centre for Environmental Research - UFZ, Department of Ecological Modelling, Leipzig, Germany*
[e] *University of Potsdam, Plant Ecology and Nature Conservation, Potsdam, Germany*
[f] *Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA, 24061, USA*
[g] *School of Electrical Engineering and Information Technology, University of New South Wales, Australian Defence Force Academy, Canberra, ACT, Australia*
[h] *Global Institute for Water Security, School of Environment and Sustainability, Department of Civil, Geological, and Environmental Engineering, University of Saskatchewan, Saskatoon, Saskatchewan, Canada*
[i] *National Institute of Water and Atmospheric Research, New Zealand*
[j] *U.S. Department of the Interior, U.S. Geological Survey, Reston, VA, USA*
[k] *Centre for Research in Social Simulation, University of Surrey, Guildford, GU2 7XH, United Kingdom*
[l] *Center on Persuasive Systems for Wise Adaptive Living (PERSWADE), Faculty of Engineering & IT, University of Technology, Sydney, Australia*
[m] *Faculty of Engineering Technology, University of Twente, Netherlands*
[n] *Key Laboratory of Virtual Geographic Environment (Ministry of Education of PRC), Nanjing Normal University, Nanjing, 210023, China*
[o] *Environmental Science and Engineering Program, University of Texas at El Paso, El Paso, TX, 79968, USA*
[p] *U.S Geological Survey, Chesapeake Bay Program, Annapolis, MD, 21403, USA*
[q] *Center for Social Dynamics & Complexity, School of Human Evolution & Social Change, Arizona State University, Tempe, AZ, USA*
[r] *Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA, 24061, USA*

A R T I C L E I N F O

Keywords:
Social-ecological modeling
Interdisciplinary modeling
Integrated modeling
Scale issues
System-of-systems approach

A B S T R A C T

System-of-systems approaches for integrated assessments have become prevalent in recent years. Such approaches integrate a variety of models from different disciplines and modeling paradigms to represent a socio-environmental (or social-ecological) system aiming to holistically inform policy and decision-making processes. Central to the system-of-systems approaches is the representation of systems in a multi-tier framework with nested scales. Current modeling paradigms, however, have disciplinary-specific lineage, leading to inconsistencies in the conceptualization and integration of socio-environmental systems. In this paper, a multidisciplinary team of researchers, from engineering, natural and social sciences, have come together to detail socio-technical practices and challenges that arise in the consideration of scale throughout the socio-environmental modeling process. We identify key paths forward, focused on explicit consideration of scale and uncertainty, strengthening interdisciplinary communication, and improvement of the documentation process. We call for a grand vision (and commensurate funding) for holistic system-of-systems research that engages researchers, stakeholders, and policy makers in a multi-tiered process for the co-creation of knowledge and solutions to major socio-environmental problems.

* Corresponding author.
E-mail addresses: iwanaga.takuya@anu.edu.au (T. Iwanaga), hsuan006@tamu.edu (H.-H. Wang), serena.hamilton@anu.edu.au (S.H. Hamilton), volker.grimm@ufz.de (V. Grimm), tkoral@tamu.edu (T.E. Koralewski), asalado@vt.edu (A. Salado), s.elsawah@adfa.edu.au (S. Elsawah), saman.razavi@usask.ca (S. Razavi), jing.yang@niwa.co.nz (J. Yang), pglynn@usgs.gov (P. Glynn), research@criticalconnections.com.au (J. Badham), alexey.voinov@uts.edu.au (A. Voinov), chenmin0902@163.com (M. Chen), wegrant@tamu.edu (W.E. Grant), trpeterson@utep.edu (T.R. Peterson), karin.frank@ufz.de (K. Frank), GShenk@chesapeakebay.net (G. Shenk), Michael.Barton@asu.edu (C.M. Barton), tony.jakeman@anu.edu.au (A.J. Jakeman), jcl@vt.edu (J.C. Little).

https://doi.org/10.1016/j.envsoft.2020.104885
Accepted 29 September 2020
Available online 6 October 2020

1. Introduction

Socio-environmental systems (SES) function across a range of interrelated scales that collectively represent a system of systems (SoS). The term SoS has been used since the 1950s and various definitions exist (Nielsen et al., 2015). In this paper, we distinguish between an SoS as a collection of human and natural systems, and SoS models which are engineered representations of an SoS. The former is defined as an interconnected collection of multiple, heterogeneous, distributed systems that collectively may give rise to emergent behavior, where each system represents a process or set of processes. In the modeling of SoS, we follow Little et al. (2019) who define SoS models as "a collection of independent constituent systems, in which each fulfills its own purpose while acting jointly towards a common goal." (p. 84). In environmental modeling, SoS models may take the form of Integrated Assessment Models (IAMs) or, more generally, Integrated Environmental Models (IEMs), which are commonly applied to inform environmental management processes (Ewert et al., 2011; Iwanaga et al., 2020; Letcher (Kelly) et al., 2013; Matott et al., 2009).

Central to SoS modeling is the view of system representations as a multi-tier structure with different levels of abstraction, where systems and indicators at lower levels can be scaled up to higher levels. These representations capture processes that operate at different scales (e.g. temporal, spatial, organizational) in contrast to 'single-system' approaches, which assume such drivers to be exogenous and, crucially, do not account for any feedback mechanisms between the represented systems. This view also sets the focus on how to integrate knowledge from the different disciplines involved and coordinate information exchange among these in a consistent and meaningful way. Knowledge integration is not limited to the technical coupling of models, but to integration among multi-scale stakeholder and expert processes. This combined socio-technical focus makes scale issues and their treatment a core consideration of SoS modeling.

1.1. The need for a holistic treatment of scale

A crucial ingredient in SoS modeling is attending to the socio-technical processes involved. Representation of scales is defined by modelers for a particular purpose and is ultimately subject to human processes (Meadows, 2008). Accordingly, the representation of an SoS is the end-product of what the people involved implicitly or explicitly have chosen to represent, and how they implemented their choices. These then influence the model structure and uncertainties embedded, and the consideration of its different dimensions, analyses conducted, and data and methods used (Glynn et al., 2017; Gorddard et al., 2016; Voinov et al., 2018). Such choices are subject to the available knowledge, experiences, biases, beliefs, heuristics and social values, as well as the perceived purpose(s) of the modeling.

A key scale issue in SoS modeling is the development of a consistent and defensible characterization of scale (Elsawah et al., 2020). Existing systems analysis and modeling approaches tend to come from entrenched disciplinary paradigms and so with a specific focus on their scales and facets, and embedded language and terms. Inconsistencies then manifest in the conceptualization and treatment of scale in SoS approaches, which prevent researchers from: (1) understanding the implications of scale choices; (2) formulating, implementing and validating models that are relevant to the questions of interest; (3) predicting future SoS responses in support of decision making (Elsawah et al., 2020; Little et al., 2019; Razavi et al., 2020); and (4) communicating modeling results in ways that help identify trade-offs and synergies within an SoS and among the systems under investigation (Fridman and Kissinger, 2019; Miyasaka et al., 2017). Addressing issues that arise from the conceptualization and representation of multiple scales are often omitted or left for future discussion (Ayllón et al., 2018).

Discrepancies in the treatment of scale can be addressed firstly by developing a shared understanding of the system(s) being analyzed through a holistic interdisciplinary process (Thompson, 2009; White et al., 2019). There is increasing recognition that holistic approaches are necessary to enable an integrated assessment of scale issues in socio-environmental (social-ecological) systems (Schlüter et al., 2019a, 2019b; Hoekstra et al., 2014). The rise of inter/multidisciplinary fields, such as socio-hydrology (Elshafei et al., 2014; Sivapalan et al., 2012) and eco-hydrology (Hannah et al., 2004; Porporato and Rodriguez-Iturbe, 2002), gives further credence to this need. For SoSs in particular, it is necessary to additionally acknowledge the socio-technical influences on their modeling. Explicit inclusion of the socio-technical perspective pushes beyond traditional modeling approaches, as it advocates assimilation of not only the data and mechanistic processes across different systems, but also includes the knowledge and information held in the social institutions involved in the modeling.

1.2. Purpose

The purpose of this paper is to advance knowledge and implementation of interdisciplinary SoS modeling by identifying and articulating the practices, issues and challenges involved with respect to issues of scale. Central to this interdisciplinary lens is making concrete the multidimensional nature of scale issues and the interplay among these. Here, the term "interdisciplinary" is favored over trans- or multidisciplinary as the focus is on the "blending" of disciplinary knowledge (White et al., 2019).

The primary audience of the paper is modelers, albeit in different domains and scientific disciplines with an interest in adopting an SoS approach as a methodological framework in SES modeling. In the following (Section 2), we first provide definitions for the key terminology used throughout this paper. These definitions are not intended to be universal but are provided to contextualize and aid in communication given the range of disciplines involved in SES modeling. In Section 3, we explore issues of scale which need to be considered throughout the modeling. We then describe in Section 4 the long-term challenges towards resolving such scale issues and suggest paths to be taken in the shorter-term.

2. Concepts and definitions of scale

2.1. The process of defining scales

SoS models principally provide a representation of the interactions that occur between the systems involved. Holistic integration of knowledge from the various disciplines involved is necessary so that the implications of the different methodological choices on scale can be understood (Elsawah et al., 2020). To this end, a three-day workshop was held in October 2019 in which a culturally and disciplinary diverse group of 20 participants convened to share their knowledge. An additional 3 contributed in complementary ways to the drafting of this paper. Contributors originated from Europe, North America and the Asia-Pacific and included engineers, economists, social scientists, mathematicians, physicists, hydrologists, computer scientists and ecologists.

To prevent miscommunication, we developed a set of terms (outlined in Section 2.2) to build a shared language (Rubin et al., 2010; Spitzberg and Cupach, 1989; Thompson, 2009). Although prior definitions of "scale" are available (see for example Cash et al., 2006; Gibson et al., 2000), it was considered useful to develop a shared, empathetic understanding of each other's perspectives (Banerjee et al., 2019; Thomas and McDonagh, 2013). The process additionally served to break down cognitive constraints (MacLeod and Nagatsu, 2018), which may otherwise blind researchers to relevant notions of scale allowing disciplinary bias to creep in and knowledge gaps to form. The range of disciplines involved in SES modeling often makes addressing cognitive constraints difficult, as there are different notions of scale, and related terms are

Table 1
Brief descriptions of the primary terms defined in this paper and relevant literature. Where no references are provided, the terms are assumed to be generic and widely known.

Term	Definition	Relevant Literature
Spatial/temporal	Spatial and temporal aspects define, respectively, the bounds or horizons over the space and time frame of the events and processes of interest as well as their discretization in a model.	N/A
Multi-system model	A catch-all term referring to any model that represents multiple systems.	N/A
Emergence or emergent behavior/simplicity/ complexity	Here, emergence relates to the behavior of the system and can span from simple to complex. Emergent complexity describes the complex, possibly chaotic, behavior that arises from the collective interactions of simple constituent systems, whereas emergent simplicity is the opposite.	Bar-Yam (1997)
System and System of systems	At its core a "system" refers to a collection of processes and mechanisms that may interact depending on context. A system of systems is represented as a collection of autonomous constituent systems that give rise to collective behavior. A constituent model may, itself, be a system-of-systems model. A system-of-systems model then is an interconnected, tiered, network of models.	(Eusgeld et al., 2011; Little et al., 2019; Tranquillo, 2019)
Integrated model	A model which consists of two or more separate and separable models, connected through a common computational framework to allow automated interactions between models to occur.	(van Ittersum et al., 2008; Voinov and Shugart, 2013; Whelan et al., 2014)
Resolution/Granularity	The represented unit of scale at which a system component is modeled (e.g. unit of distance, volume, time, social unit, etc.)	(Ewert et al., 2011; Groen et al., 2019; Neumann et al., 2019)
Actor	Actors are entities, both human and non-human (e.g. objects, biota, flora and fauna, institutions, and organizations), which influence the modeling, the pathways taken throughout the modeling process, and their representations within a model. Actors may themselves be composed of actors, such that a system is an actor within a larger system (e.g. engine in a car, team within a company, etc.). Actors may influence one another through a network of relationships and be modeled as such. Actors may embody collective culture and personalities, as may be the case with teams and organizations.	(Cresswell et al., 2010; Macy and Willer, 2002; Tate, 2013; Hobday et al., 2018; Schneider et al., 2013)
Hierarchy/Level	The ordered linkage crossing scales, which may be spatial/temporal (neighborhood to city) or virtual/conceptual (employee and employer), and these may be nested within one another.	(Ostrom, 2007; Schweiger et al., 2020; Steinhardt and Volk, 2001)

used in different ways depending on context. This variance has been observed in the use of common terms with conflicting definitions between (and sometimes within) disciplinary fields (Bridle et al., 2013).

2.2. Scale terminology in SoS modeling

Defining the terminology associated with scales was an arduous process at first, owing to the diversity amongst workshop participants. A brief overview of the resulting primary terms used in this paper is provided in Table 1. For the discussion here, "scale" is taken to have an expansive definition, covering the scope of work to be conducted in the treatment and representation of system processes. Aspects of scale that had unanimous consensus included the *commensurability* of the choice of scale within the *purpose of the modeling,* and the consistency of *spatial* and *temporal* scales across models. It was also acknowledged that scale can mean many things beyond the spatial and temporal, for example the less tangible such as treatment of ethical considerations within the modeling process (e.g. Häyhä et al., 2016). Regardless of definitions, treatment of scales - and the choices made in this treatment - influences the model uncertainties and the outcomes of the modeling.

Commensurability refers to the appropriateness of the selected approaches and methods for the SoS modeling purpose. Broadly speaking, these approaches can be described as being subject to *socio-technical* considerations, which are the focus of the discussion in this paper. The social (human) aspect of modeling includes the circumstances of collaboration, project management and participatory processes, as well as those settings influencing the technical aspects, including modeling and computational considerations.

The *spatial* and *temporal* features of a system are often the primary aspects around which scale is traditionally considered and framed. These define the time and space of interest (both their horizons and discretization) and the events and processes that are considered important to represent (Cash et al., 2006). The spatial scales selected may be influenced by the temporal scales of interest, and vice versa. Their dependence can be intensified by the fact that spatio-temporal scales are often influenced by factors outside their defined boundaries. Such influences may be important but may not be well understood or ignored (Zhang et al., 2014b, 2014a).

Resolution defines the *granularity* of system representation and refers to the unit of spatial/temporal scale represented in each system. Resolution may be spatial or temporal in nature but extends in other ways such as to social units (individuals to families to communities, etc.) and thus may be represented so as to conform to a semantic or conceptual *hierarchy* (Cash et al., 2006). Choice of resolution is highly dependent on the modeling context, generally informed by the availability of data, the needs of the model (including for numerical stability, sensitivity and model identifiability), and model purpose.

Hierarchy and their respective *levels of organization* relate to the representation of nested relationships among systems (Ostrom, 2007). For example, various governance systems may co-exist at a range of scales with separate administrative or institutional concerns (Daniell and Barreteau, 2014). Team-based organizations are one example where the hierarchical scales may not be constrained to specific locations, with members performing a variety of roles within an organization that may be geographically spread across different time zones.

Actors influence and define the aspects of scale that are considered and may be both human and non-human entities which affect or influence one another. The term has its roots in the social sciences (an example may be found in Wessells, 2007). Actors have roles and carry out one or more activities in the system and can be represented individually or collectively. Human actors have attributes such as values, goals and mental models, which influence their behavior (Pahl-Wostl, 2007). Non-human actors are defined here in a literal sense (i.e. not an individual biological person) such that organizations, flora and fauna are non-human actors but may still exhibit collective culture and personalities (Hobday et al., 2018; Schneider et al., 2013). A system can encapsulate many actors and may be an actor itself.

The different types of system modeling encompass many terms that are often used interchangeably across the sciences. As alluded to in the introduction we are guided by, but do not directly adopt, definitions as applied in system-of-systems engineering (cf. Dahmann and Baldwin, 2008). Here, a single-system model targets a specific system, for instance an agricultural system without explicit representation of the hydrological dynamics or climatic influences. Consequently, single-system models constrain themselves to the concerns and considerations of a single sector. Models concerned with a single system may, of course, use several models internally (e.g. crop growth, soil water properties, etc.) and these are referred to here as *component models*.

A direct approach to representing additional systems can be accomplished by applying, albeit separately, a selection of single-system models for a given problem domain. In such cases, knowledge gained in the application of a model may inform the use of another. Data from one model may be fed into another, and vice versa, typically via manual processes. For example, a weather forecast model may be used to provide inputs to an agricultural model to determine seasonal effects on crops, and the agricultural model may provide land surface boundaries to the weather forecast model.

Multi-system representations can be *integrated* by *coupling* models together such that data interoperation occurs in an automated fashion. Individual "system level" models are then referred to as *constituent models*. The advantage of multi-system models over their single system relatives is that the impacts and feedback mechanisms can be represented across/between their individual scales (Elag et al., 2011; Tscheikner-Gratl et al., 2019; Wang et al., 2019). Multi-system models, with their explicit representation of system interactions, are therefore capable of providing more holistic assessment compared to the use of individual models in isolation (Kelly (Letcher) et al., 2013). Component-based modeling stems from Component-Based Software Engineering (Vale et al., 2016; Hutton et al., 2020) and common usage in environmental modeling typically makes no distinction between constituent and component models (e.g. Malard et al., 2017). A conscious decision has been made here to adopt the term "constituent" from the systems engineering field (Nielsen et al., 2015) to convey this distinction.

It is important to note that "integrated" and "multi-system" models could then equally apply to both single-system models with several component or constituent models. The requirement for a model to be regarded as "integrated" is that its (component or constituent) models are coupled together through the use of a common automated infrastructure to facilitate data interoperation (see for example, Malard et al., 2017; Whelan et al., 2014). By necessity, multi-system integrated models are more complex and may involve a variety of modeling paradigms (e.g. Bayesian networks, agent-based, system dynamics, etc.) and their combinations.

An *SoS model* is then regarded here as an integrated model with constituent models. Each constituent model may be a single-system or another SoS model such that a tiered network of relationships between models is formed, with each representing a layer of abstraction. In SoS modeling, each constituent model may operate across different spatial/temporal scales, hierarchical levels, and resolutions to incorporate multiple aspects of distinctly separate (disciplinary or sectoral) domains and modeling paradigms. An SoS perspective allows, but does not prescribe, consideration of complex system properties including nonlinearities, interdependencies, feedback loops, thresholds and emergence.

3. Scale issues to consider

Models are developed through a life cycle of various phases, each with specific considerations and steps (the "modeling cycle"; Grimm and Railsback, 2012; Hamilton et al., 2015; Jakeman et al., 2006). SoS modeling is more complex compared to 'single-system' models due to the number of people and disciplines involved as well as the dependencies between the constituent models. Similarly, management of the modeling process is made more complex, as there is not a single modeling cycle, but multiple cycles occurring asynchronously. Each actor and model may have separate objectives and purposes, priorities and differing levels of available resources not to mention the need to consider the availability of resources for the SoS modeling as a whole.

The sections below are adapted from the modeling phases identified in Badham et al. (2019) and Hamilton et al. (2015), wherein the actions undertaken in each modeling phase are described. In contrast, we identify the relevant phases within an SoS context and outline the considerations with respect to scale issues. Fig. 1 depicts the high-level considerations/objectives within each phase. While the sections below are presented in a sequential manner, we stress that modeling is an iterative and concurrent process.

3.1. Scoping phase

In this phase, the objectives of the modeling are clarified by defining the problem and how modeling is intended to address it. Examples of model (or modeling) purpose could be to fill gaps in knowledge, to support learning and communication processes, to validate current understandings and assumptions, to predict what might happen in the future, or to carry out scenario analysis (Badham et al., 2019; Kelly (Letcher) et al., 2013). Ideally, this scoping phase results in a clear understanding of the model types and components that need to be developed or, in later iterations, their limitations with respect to the model purpose and how to address these.

3.1.1. Problem definition and scoping

While the overarching purpose of the SoS model may be known, the specifics may be less clear at the outset. Development of a consistent and shared view of the scales to be considered involves communication of the scope and interactions across the constituent systems between all involved (see Fig. 2). This process can aid in identifying and addressing areas that require reconciliation of different views that often exist across the stakeholders. Awareness of the scale issues will likely evolve as the modeling progresses through the iterations. The choice of modeling pathways and methodological framework employed is heavily informed by this awareness (MacLeod and Nagatsu, 2018).

Involvement of stakeholders, including domain experts, through participatory processes can inform the identification of relevant scales in the face of uncertainty and (poor) data availability (Hamilton et al., 2015; Kragt et al., 2013). Stakeholders can also play a role in selecting

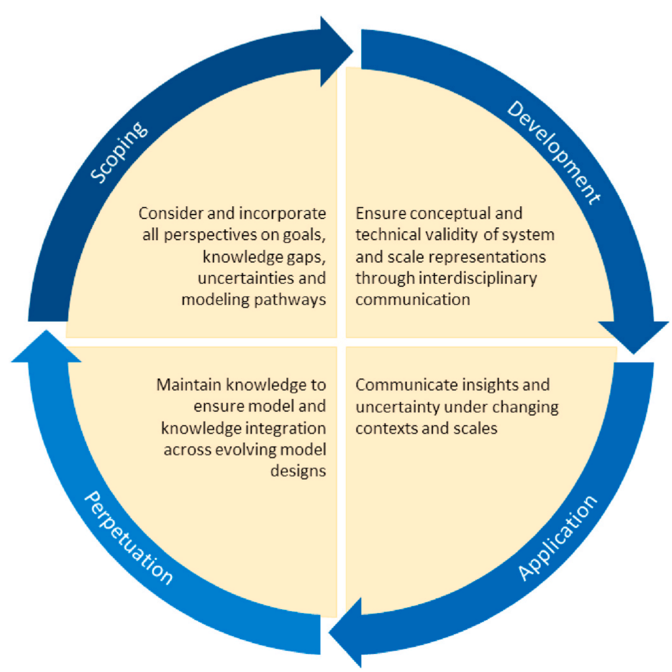


Fig. 1. The phases in the modeling cycle (adapted from Badham et al., 2019, and Hamilton et al., 2015) with key considerations within each phase.

and combining data, furthering holistic consideration of system actors and aid in developing the model purpose. The relationship between actors and their roles in framing the scale, scope and purpose of the modeling has been previously recognized (Kragt et al., 2013; Refsgaard et al., 2007) and is further explored in the next subsection.

Insufficient consideration or agreement regarding the overarching purpose of the SoS model may ultimately affect model performance and outcomes (Connor et al., 2019). The higher number of actors in SoS modeling increases the difficulty in reconciling different or mismatched perspectives, requirements and purposes. This is a "problem of heterogeneity" (O'Connell and Todini, 1996) and is not restricted to any single discipline. Often, and by necessity, the scale of the modeling is to be commensurate with its purpose, including the level of certainty being sought, and the available resources.

Purpose and use of constituent models may be mismatched if conflicting perspectives over the scope of the modeling are not addressed. Modelers that have different goals in mind may only consider scales relevant to their immediate (and often discipline-specific) concerns, leading to an improper selection of constituent models. There is potential for a high degree of mismatch between constituent models even if modelers coordinate their efforts. Unexpected cascades of effects through scales is commonplace in complex systems (Tranquillo, 2019), and could arguably be taken as the rule rather than the exception.

Change in scale may also occur during the modeling process, due to new information that triggers a necessary change in model context. The scale of model interactions to be represented can also influence the number and type of constituent models included, and overall system complexity. The choices regarding scale then have implications for how well interactions among systems can be represented with respect to the model purpose. Scope creep, wherein the scale of the modeling is continually extended to cover contexts not originally envisioned (cf. Barton and Shan, 2017), may eventually compromise modeling efforts, as available resources get stretched too thinly to achieve sufficient progress (Sarosa and Tatnall, 2015).

Choice of scales is further compounded in cases where system bounds cannot be clearly and definitively defined. Coastal zones, atmospheric systems, and natural resource management systems are examples of systems with ambiguous system boundaries. Social systems and their dynamic structures are another example that do not have clear boundaries yet place important, even governing, conditions on system behavior. Such social systems, and their influences, are so far underrepresented in current integrated assessment efforts (Zare et al., 2017). The lack of clear boundaries of such systems are often considered to be part of the problem (Voinov and Bousquet, 2010).

Reconciling conceptual differences and perspectives between human actors can be demanding but not insurmountable. There are various methods available for group decision making, such as the Delphi technique (Gokhale, 2001), which can be used to help the group reach agreement on the definition of the problem and/or the system boundaries. The subsequent modeling itself can be used to combine and reconcile different views among stakeholders, and may be useful in cross-cultural or particularly contentious settings (cf. Potter et al., 2016). The influence of modeler and stakeholder bias can also be constrained such as by using numerical optimization and/or exploratory modeling processes (Martin et al., 2017; Reichert, 2020). The influence of personal preferences is restricted by using the exploratory approach as it focuses on identifying the relevant scales and conditions (or combinations of conditions) that normally lead to desirable outcomes.

3.1.2. Stakeholder planning

Here, "stakeholder" refers to the individual or groups that may affect or be affected by the modeling or have an interest in its outcomes (Freeman, 2010). Thus, in this context, the modelers (and teams of modelers) are also stakeholders. There is a plethora of stakeholder-focused approaches (e.g. in integrated modeling, participatory modeling), but these methodologies are still limited in their capacity to deal with scale-specific questions and challenges brought by SoS modeling (Jordan et al., 2018). Generally, participatory approaches aim to bring together the multiple goals, issues, and concerns of interest from multiple scales and governance systems by developing a mutually beneficial relationship between stakeholders (Thompson, 2009). Thoughtful consideration of transparency, traceability and governance issues in engagement and participatory processes (Cockerill et al., 2019; Glynn et al., 2017) will be essential for optimizing saliency, legitimacy, and credibility of the SoS modeling (Cash et al., 2003).

The participation of a higher diversity of stakeholders in such processes allows for a more holistic representation to be developed, covering potential 'blind-spots' in the system conceptualization and avoiding the "siloing" of knowledge (Hoekstra et al., 2014). Including further perspectives may increase the complexity of the modeling, and so requires careful management of individual expectations and biases (Martin et al., 2017). Management of an SoS may at times be predicated on effective management of stakeholders and their level (and capacity) of involvement (Ostrom, 2007; Boone and Fragaszy, 2018).

Increases in the variety of perspectives also increases potential for conflict - defined here as disagreements of any degree - between teams, team members and/or stakeholders. On the one hand, there is evidence that conflict plays a positive role in learning and effective teamwork (Tjosvold et al., 2003). Such positive benefits, however, may only occur in cases where there are high levels of pre-existing trust within the

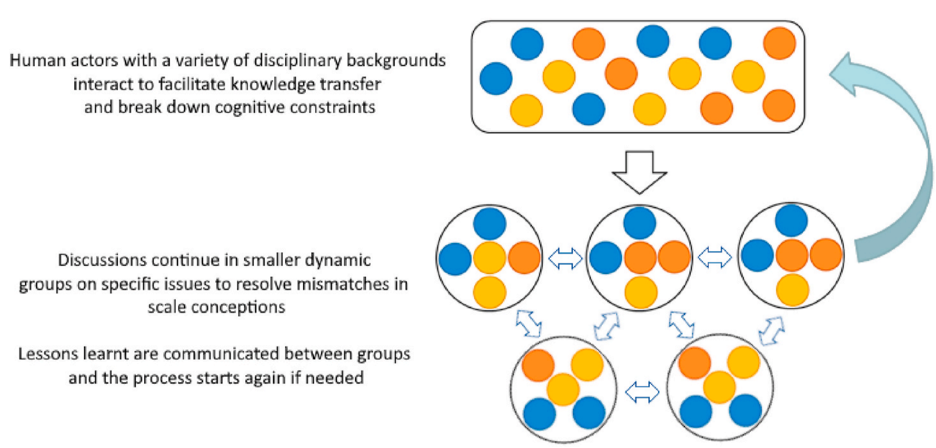


Fig. 2. Continuous and repeated interactions between human actors (domain experts, stakeholders, modelers, etc. represented by the different colored circles), and between their social groups, are necessary throughout the modeling process to ensure mismatches in system conceptualization and constituent model scales are avoided.

group, and when the conflict is task-related rather than interpersonal (De Dreu, 2008). Power dynamics within teams and stakeholders therefore need to be considered (National Research Council, 2013). Identification and focus on objectives that require participants to work together (known as goal interdependence) is an identified foundation towards project success and may additionally help in avoiding conflict (Knight et al., 2001; Lee et al., 2015; Tjosvold et al., 2003). Careful design and management of interactions between teams and stakeholders requires an explicit consideration of how the multiple, and at times contradictory, objectives might align or connect. Approaches to conflict resolution and prevention (e.g. boundary critiquing, Midgley and Pinzón, 2011) are promising, but still under-utilized techniques.

Effective stakeholder engagement will in practice be impacted by geographic spread (Allen and Henn, 2006), as the realities of scheduling rarely allow all stakeholders to be engaged at the same time and place. Additionally, a diversity of stakeholders (e.g. policy makers, scientists, and the public) mean material and modes of communication may need to be tailored for each. Online participation platforms and technologies extends the reach to participants and are appealing for their asynchronous and distributed modes of engagement (Yearworth and White, 2018). These relatively new technologies are simply tools, however, and a capacity to both use and leverage their advantages is also required (Cooke et al., 2015). Regardless of how interactions are to occur, without documenting a Record of Engagement and Decision-making (RoED, Cockerill et al., 2019), the original purpose, assumptions, and social and biophysical context of the engagement and resulting model choices might be lost, leading to mismatches in understanding, conceptualization, and implementation. The literature is still limited on the effectiveness of using different participatory methods for different purposes and audiences (Voinov et al., 2018). Nevertheless, plans for stakeholder engagement for SoS modeling should explicitly consider the scaling challenges, and devise strategies to deal with these.

3.1.3. Preliminary conceptual model

The preliminary conceptual model represents the current understanding of the system and the relationship between constituents, including identification of key drivers, interactions and outputs of interest (Badham et al., 2019). In describing and capturing the essence of the system, development of the conceptual model helps with the design of the subsequent (computational) model as well as making concrete the model purpose. Two scale-specific aspects are to be considered here: the approach used for conceptual model development (see Table 2 for a general overview) and the formal representation (e.g. equations, technical specifications, etc.). The processes that are included or excluded based on actors' perceptions, priorities, beliefs, and values under the SoS context will inevitably influence the data leveraged, the properties of the computational model, and therefore the paths taken.

Few mapping techniques exist that focus on illustrating multi-scale representations. Scale separation maps (Hoekstra et al., 2007) or Stommel diagrams (Scholes et al., 2013) represent the scales of the constituent systems on a two dimensional space-time map. System diagrams, such as the representations used in van Delden et al. (2011) and Oxley and ApSimon (2007), organize the system components according to their spatial and/or temporal scales, and show the interactions between these components. On the other hand, coupling diagrams (Falcone et al., 2010) show the flow of data between models.

A further approach is to use the ODD protocol, named after its three blocks: Overview, Design concepts, and Details (Grimm et al., 2006). The original purpose of the ODD protocol was to describe and enable transparent communication of agent-based models (ABMs) to ensure their replication and the reproducibility of results based solely on the model description (Grimm et al., 2020). The conceptualization involved in the Overview block mandates identifying the scales of the processes or system components to ensure a shared understanding of the system being modeled. This is further complemented with the identification of relevant resolutions and spatial/temporal bounds. At this stage, the bounds can be vaguely defined (e.g. local, regional, global). This initial assessment of the scales involved may be revised throughout the modeling process as understanding improves. The ODD protocol is under continual development, and planned additions extend its consideration and applicability of use to other areas not previously considered (as outlined in Grimm et al., 2020).

Table 2
Description of the general approaches in the development of multi-scale models, adapted from Ingram et al. (2004).

Approach	Description
Top-down	Creation of a coarse generalized model which is then progressively refined to an appropriate mix of scales.
Bottom-up	Models are developed at the smallest resolution initially conceptualized to be necessary and are then expanded to encompass scales as further information becomes available.
Middle-out	Development of the SoS model begins at the scale richest in data or information, working "outwards" towards smaller and larger scale models, as necessary. In SoS modeling, what is "richest" is likely to be subjective to each discipline and available understanding.
Concurrent	The process of constructing models to represent all hierarchical levels at the same time.

If differences in conceptual understanding of the scales and their interactions cannot be reconciled at this stage, it is possible to create multiple alternative models representing the different hypotheses which can be tested in later stages of the modeling process. Such an approach can also assist in assessing uncertainty rooted in model building choices, as the treatment of scale may affect model outputs and outcomes (further discussed in Section 3.2.4). Although conceptual diagrams can be developed without specifying the scales involved, explicit consideration of scale is valuable for avoiding misinterpretation of the conceptualization and ensuring key variables and processes are included. A useful reflexive exercise, not usually reported but aiding transparency, is to identify what alternative approaches were considered, or could have been considered, and how these may have affected results and outcomes, if adopted.

3.2. Development phase

3.2.1. Collecting data, information, and knowledge

Data, information and knowledge for each constituent model may come from the field or through literature, solicited through expert and stakeholder engagement, or collected through analysis. Considerations towards data collection in the integrated setting have been previously explored in Badham et al. (2019). Correctly communicating and interpreting data across heterogeneous systems, however, requires that the data are interoperated between constituent models and that model behavior across scales remains valid and meaningful (Renner, 2001). For this purpose, metadata serves an essential role.

Transparency in the collection process and approval from those involved in the modeling are necessary to ensure that collected data remain conceptually relevant across scales. Furthermore, transparency in the context of data collection and usage is a key factor to develop trust among stakeholders and model users, and future adoption of the constituent models (Barba, 2019; Gray and Marwick, 2019). Data may need to be transformed to be fully relevant for the context of its intended use, such as up-or-downscaling to ensure compatibility with other processes. Ideally, metadata would include information on the data collection, uncertainty and transformation process, which aids in determining the appropriateness of data for the SoS model. Explicit descriptors of both input and output data can assist in identifying the commensurate level of data collection with respect to available resources.

Modeler bias can have a compounding effect as the choice of data collection, as well as the metadata that describes the data, influences how system interactions are perceived, and thus conceptualized (Bhattacherjee et al., 2008). What may be considered irrelevant in one field may dictate modeling pathways in another. In an SoS setting there are many more participants involved and so there is a high degree of uncertainty stemming from the decisions made as a result.

Data quality and informativeness (e.g. accuracy or precision) provided by constituent models may also be diverse. Diversity of data obtained from a diversity of sources, however, runs the risk of conflicting information (Gray et al., 2012). Modelers from different disciplines may also utilize different scales for the same process, resulting in inconsistencies, and thus errors, the sources of which are difficult to identify. In this regard, non-quantitative sources of information, gathered from literature and/or through stakeholder engagement, may become key assets that resolve such issues (Grant and Swannack, 2007). In cases where data describing a particular linkage in an SoS model are not available, theoretical relationships, generally applicable empirical relationships, or model process and output can be useful representations for the purpose of the SoS model (Rai et al., 2002). The documentation developed in the Scoping phase can be leveraged to ensure applicability and validity with regard to the model purpose.

3.2.2. Construction

Construction of computational SoS models requires the marrying of domain expertise from across the various disciplines involved with technical software development knowledge. While the overarching context may be well-defined within the scoping phase, it is in this Construction step that the individual components, and the scales they represent, are developed, and coupled, tested and validated. Here, existing models may be repurposed or new models developed. The specifics of their initialization, interoperation, method of execution and management of the data involved are to be determined and prototyped in this phase (Igamberdiev et al., 2018; Madni and Sievers, 2014).

A balanced approach is needed in SoS model development that takes several factors into account. There is a danger that the models themselves become treated as pieces of software that merely require connection, ignoring the socio-technical context for their intended use (Voinov and Shugart, 2013). Another issue is the overparameterization of constituent and component models (Brun et al., 2001; Nossent and Bauwens, 2012), as simply integrating these models to form an SoS model exacerbates issues of uncertainty and identifiability (considerations of which are explored in the following sections). At the same time, ignoring the technical considerations of integration is also inadvisable (Verweij et al., 2010). Mitigating the issues that consequently arise becomes increasingly difficult as more systems and scales are included (Voinov and Shugart, 2013; Wirtz and Nowak, 2017).

Requisite systems could be represented at the level of detail necessary for the SoS model purpose through a tiered modeling structure (Little et al., 2019). Implementation of such a tiered approach can involve developing metamodels or entirely different system models. Metamodels being simplified representations of more complex models (revisited in Section 3.3). Two pertinent issues in SoS model construction are the focus below: managing the conceptual inter-connection between models, and the process of integration.

3.2.2.1. Conceptual integration. Conceptual integration of constituent models can benefit from requiring that constituent models be mechanistic as opposed to black boxes. When a model is implemented as a black box, it becomes difficult to evaluate and understand (Lorek and Sonnenschein, 1999). SoS modeling may make use of pre-existing models which constitutes re-purposing, implying the transference of the model assumptions, limitations, and scale to a new context. It is emphasized here that model suitability within its original context is not necessarily applicable to the new context (Ayllón et al., 2018; Belete et al., 2017; Voinov and Shugart, 2013). Availability of code alone, for example, does not imply transparency. What is important is the contextual information that is necessary to assess the suitability of the model purpose and functionality.

A key challenge then is ensuring the box remains open and transparent rather than closed and opaque. Opaque development can be attributed to the modular nature of constituent model development, with the teams working separately - both conceptually and geographically - and often split along disciplinary lines. Such teams can be described as self-organizing (Sletholt et al., 2012) but may lack cross-disciplinary knowledge (cross-functionality, as in Hidalgo, 2019; Hoda et al., 2013). The lack of interdisciplinary communication between teams then results in black, or at best gray, box models to those not involved in their development.

What is important in this interdisciplinary context is clear documentation and an organizational culture that supports the perpetuation of the relevant contextual knowledge. As previously mentioned in Section 3.1.3, describing the model and its conceptual linkages in a single canonical document via the ODD Protocol (introduced in Section 3.1.3) is one approach that could be leveraged. Furthermore, a "nested ODD" approach may be adopted in the case of complex SoS models wherein the constituent models may be another SoS model.

3.2.2.2. Technical integration. Technical integration refers to the correctness of model interactions, recognizing the distinction between conceptual or abstract representation (e.g. an equation or flow diagram) and its implementation as software. Successful technical integration of computational models requires the necessary engineering expertise to be available (Knapen et al., 2013). Crucial considerations are that constituent models interact and accordingly that errors will propagate (cf. Dunford et al., 2015), and that each constituent model may undergo its own separate development cycle which invariably necessitates continual adjustments to be made.

Flexibility of integration is often desirable as it allows the model to be resilient against changes in the modeling scope. Flexibility facilitates investigations into model structure (of both constituent and component models) and the technical design considerations that lead to flexibility allows for the composition of different combinations of relevant code and data represented through a nested hierarchy (e.g. 'loose coupling'; Elag et al., 2011; Vale et al., 2016; Whelan et al., 2014). Use of integration frameworks are helpful in that they allow the treatment of individual models as loose, composable, modules that provide some flexibility in dealing with the range of scales involved.

Current integration frameworks typically have their roots in specific disciplines and tend to focus on physical processes (cf. Ayllón et al., 2018). The Open Modeling Interface (OpenMI, Moore and Tindall, 2005), for example, has had to evolve from its initial focus in the hydrological sciences to accommodate an interdisciplinary modeling process (Buahin and Horsburgh, 2018). Thus, while the processes and requirements of such frameworks may be generally applicable, there remains some difficulty in their generic implementation and adoption within the interdisciplinary context of SoS modeling.

In some cases, such frameworks may be overly complex or otherwise unsuitable for the purpose and context in which the modeling is being conducted. Such difficulties may be resolved in the future as improvements to these frameworks are ongoing (Voinov and Shugart, 2013). Often modelers adopt a less formalized approach to avoid an inappropriate or constraining framework. In either case, ensuring semantic and conceptual correctness between models is typically left to the modelers themselves (cf. Hutton et al., 2020). Direct, manual, "tight-coupling" of models without the use of integration frameworks is still very much the norm.

More recent efforts include a collaborative web-based platform through which the conceptual, semantic and technical integration occurs (OpenGMS, in Chen et al., 2019; Chen et al., 2020). Faster feedback between participants then allows identified issues to be addressed earlier. Other approaches provide a curated ontological set of descriptors for common phenomena of interest (e.g. snowmelt or rainfall). These can be referred to as "system variables" (as in Pacheco-Romero et al., 2020) and efforts to record their quantities (e.g. centimetre, grams, etc.) and relevant operators in a specific metadata format have also been undertaken (e.g., the Standard Names, in Hobley et al., 2017). Having the inputs and outputs described and documented in such a way aids in reducing potential mismatches in later (re)use and could be used to enable later automated model coupling. Frameworks do not yet fully automate conversions or identify incompatible or inconsistent usage (e.g. litres per second to degrees Celsius) although this is likely to change in the near future.

Both the selected framework and constituent models may change over the course of the modeling cycle along with the scales represented. Such changes may affect its appropriateness with respect to the model purpose. For example, adoption of a particular framework or model may increase the computational requirements or necessitate changes to constituent models to allow interoperation. Inadequate consideration of the concerns and requirements of the modeling as a whole may occur in cases where cognitive constraints are still in place. The modeling process may be smoothed if requirements of the later phases are kept in mind during the design, construction (or selection) of models, and the resources allocated – including the availability of expertise – to each of these activities.

3.2.3. Model calibration

Calibration is the process of tuning parameters or altering the functional forms of equations or relations to achieve desired model behavior (Bennett et al., 2013). In SoS modeling, issues such as non-identifiability and equifinality (Beven and Freer, 2001; Guillaume et al., 2019), curse of dimensionality (Bellman, 2015), computational burden (Razavi et al., 2010), and data representativeness (Beven and Westerberg, 2011; Singh and Bárdossy, 2012) may all be amplified.

Calibration implies the existence of appropriate and sufficient data to calibrate models against. Availability of data relevant for the modeling purpose is a requirement no matter how perfect the model may be. Conversely, a lack of data does not imply subsequent modeling is not useful. A model with high uncertainty may still characterize uncertainty in a way that is meaningful to decision makers, for example indicating the comparative tradeoffs between available management options (Reichert and Borsuk, 2005). Assessment of uncertainty can be helpful in determining the relative "worth" of data to be collected to better characterize uncertainty and inform future modeling or research (López-Fidalgo and Tommasi, 2018; Partington et al., 2020). Such optimal experiment design approaches may also be leveraged to maximize the use of available data (Bandara et al., 2009; López–Fidalgo and Tommasi, 2018; Vanlier et al., 2014).

Arguably, model calibration within the SoS paradigm can take three general approaches: (1) calibration of each constituent model independently before integration, (2) calibration of all models together after integration, or (3) a combination thereof. The first approach is the simplest and most straightforward as each constituent model would be calibrated within its own domain (Phillips et al., 2001). While pragmatic, it ignores the effect of representing different scales across the represented SoS and system-system interactions, which in turn affects model behavior and performance of the individual constituent model. If a model is considered "calibrated" when both an acceptable level of fit and reasonable parameter values are found (as in Anderson et al., 2015), calibration in the *dis*integrated context does not necessarily transfer to the integrated context. In other words, what is "reasonable" in one context may not be so in another, and the selected parameter values may not be robust to the change in context that integration brings due to the different scales, interactions and data space involved.

The second approach is seemingly the most comprehensive approach to model calibration, as every possible interaction between models could be present in the process of model calibration (Huang et al., 2013). Interdisciplinary knowledge is leveraged to ensure calibrated values are both reasonable for the expanded operationalization. This then enriches the data space for individual constituent models and improves their performance (Jones et al., 2017). The approach, however, has the following major barriers:

- The search space for model calibration will be excessively large (Ling et al., 2012). In addition, new (possibly erroneous) interaction effects might emerge between the parameters of one model with those of another model, especially with different scales of information, which makes the response surface extremely complex for model calibration. The calibration process might then become computationally cumbersome and/or infeasible.
- The available data with different scales may not be enough to properly constrain the model in the process of calibration (Ingwersen et al., 2018), as it is not identifiable from the data (Guillaume et al., 2019). There is a risk of overfitting as well, as the available data might be insufficient to produce a generalized model that covers the integrated domain.
- Expert knowledge for each model may have scale constraints and may not be easily transferable to the full SoS domain (Howard and Derek, 2016).

In the third approach, models are integrated one-at-a-time, incrementally adding complexity so that the influence of each constituent model can be directly attributed and subsequent issues can be addressed. This approach may include modifying the conceptualization as necessary and sequentially calibrating the resulting integrated configurations (Duchin, 2016; Duchin and Levine, 2019). While this approach may be as pragmatic as the first, and perhaps as comprehensive as the second, the disadvantage is the time and computational cost to perform sequential coupling and calibration. Such an approach would seem more practical in cases where there is little disciplinary friction and a relatively small number of models to be integrated.

In all approaches above, the role of expert knowledge in determining the acceptability of the calibration cannot be understated. In management contexts, for example, change in policy (e.g. the governing rulesets) may impart shifts in system behavior that may be hard to discern by examining quantitative data alone, and even more difficult to represent. Machine learning approaches may assist in identifying and representing non-stationary system behavior (e.g. Rui Wu et al., 2019; Razavi and Tolson, 2013) but still require intensive data for training and validation by experts where possible (Razavi and Tolson, 2013), and scale issues still exist between different single-system models or different levels of model integration. Such information in one system may have implications for how other constituent models are calibrated, and so interdisciplinary communication, awareness and consideration of the intertwining issues is necessary to safeguard against mismatches.

A calibration method which seems not to have been used explicitly for SoS models is pattern-oriented modeling (Grimm and Railsback, 2012; Railsback and Grimm, 2019; Wiegand et al., 2004, 2003). Here, a set of patterns observed at different scales and levels of organization is used to reject, as a set of filters, unsuitable parameter combinations and process representations, and may be closely related to the use of hydrologic signatures for (hydrological) model calibration and testing (Gupta et al., 2008). As for parameters, this approach corresponds to the rejection method in Approximate Bayesian Computing (van der Vaart et al., 2016). The basic idea is that a combination of "weak" patterns, which by themselves do not contain much information and thus would not reject many parameter combinations, can be as efficient as using a "strong" pattern, which is highly distinctive, but might not be available. For models with multiple scales, this approach holds high potential as it would help to keep both the SoS and constituent models within realistic operation spaces.

3.2.4. Uncertainty analysis

SoS models often target large problem domains which necessitate complex models for their assessment and by their nature have a high degree of uncertainty. For the discussion here, we speak to the quantitative and qualitative aspects of uncertainty, which may be further classified based on their source or primary influence. Prior literature, for example, speaks of model structure, technical, parameter, scenario, contextual and predictive uncertainty (for further description, see Beven, 2009; Pianosi et al., 2016; Walker et al., 2003).

Quantitative approaches aim to measure the effect of uncertainty in a specific parameter, input or assumption on an output and allow the numerical characterization of the output distribution and therefore model behavior (Saltelli et al., 2019; Zimmermann, 2000). Qualitative uncertainty, however, cannot be characterized with a value and arises from sources such as the biases and subjective beliefs of human actors (Chen et al., 2007). Qualitative uncertainty can also arise from the modelers' subjective judgment, linguistic imprecision and disagreement across actors involved (Linkov and Burmistrov, 2003; Refsgaard et al., 2007).

One reason for increased model uncertainty in SoS modeling is the complexity that is largely a result of the increased scope of modeling, which comes with a larger number of models and people (and their perspectives) involved. The increase in the number of actors typically results in an increase in the overall number of parameters and their possible interactions (Oreskes, 2003), the number of possible decision pathways in the modeling process (Lahtinen et al., 2017), and the level of stakeholder influence at each decision fork (Ostrom, 2007).

Increasing model complexity allows for a higher-fidelity model, but can also increase the perceived uncertainty in a traditional sense; known as the complexity paradox (Oreskes, 2003). Characterizing "true" uncertainty in an SoS model, however, is impossible as it requires a model that represents everything perfectly including unknown unknowns (Hunt, 2017). Uncertainty may then compound with each interaction across constituent models in the SoS framework, propagating some amount of error (Dunford et al., 2015). Thus, it becomes progressively difficult to gain insights as to what effect and influence the combinations of these have (structural and parameter identifiability as in Bellman and Åström, 1970; Guillaume et al., 2019). High levels of model uncertainty need not be a barrier to effective decision support, however, and is ameliorated by providing estimates or assessments of such uncertainties (Reichert and Borsuk, 2005), both quantitative and qualitative. Different strategies and further considerations for uncertainty assessment are needed in SoS modeling compared to single-system modeling.

One commonly suggested approach to restricting model complexity (and possibly runtime) is to screen for insensitive parameters (Pianosi et al., 2016). Such parameters are said to have negligible influence on model output and so may be "fixed", i.e., made static in subsequent analyses, or otherwise removed from the model. Another is to "tie" related parameters so that they may be represented by a single "hyperparameter" (Raick et al., 2006). Reducing the number of parameters, however, does not necessarily equate to a reduction in uncertainty. Rather, it may simply mean consideration of an uncertainty source is determined to be unimportant for a given context or purpose (Pianosi et al., 2016), and doing so may trade off model fidelity under new unseen conditions.

Use of a constituent model within an SoS model as opposed to its individual operation, or its modification or simplification through parameter screening and tying constitutes a change in context. Therefore, parameters initially found to be influential might become inactive and non-influential (and vice versa), or the relationships that led to parameters being tied may change. The change of context also changes the relevance of the assumptions and objectives, and what constitutes an appropriate uncertainty analysis (Song et al., 2015). Uncertainty analysis conducted in one context is not valid across all scales. Thus, premature model simplification may ultimately affect the appropriateness of the SoS model for its overarching purpose. A comprehensive sensitivity analysis under current and possibly alternative conditions can provide valuable insights into a key question: "*when and how does uncertainty matter?*", as discussed in Razavi et al. (2019). An alternate view is that, given the likelihood of limited computational resources, efforts to characterize and communicate uncertainties to stakeholders may be more beneficial than an exhaustive sensitivity analysis (Reichert, 2020; Anderson et al., 2015).

An additional consideration is that a constituent model may be a legacy or third-party model that cannot be modified (e.g., due to lack of access to the underlying code). This would introduce some hidden or uncharacterized uncertainty into the SoS modeling. In this case, metamodeling (expanded on in the next subsection) might provide some help in simplifying the model.

Explicit documentation of the criteria used for each constituent model can ensure relevance of its application and reduce contextual uncertainty (see Walker et al., 2003) across all the scales involved. Accordingly, in the recent update of the ODD protocol (Grimm et al., 2020), a standard format for describing models, the element "Purpose" has been changed to "Purpose and patterns", with patterns being the multiple criteria for ensuring a model's structural realism, as defined in the "pattern-oriented" modeling strategy (Grimm, 2005; Grimm and Railsback, 2012). The effect and relative importance of model structure uncertainty may be assessed through expert and stakeholder knowledge of alternate models (van der Sluijs, 2007) and Bayesian approaches could be applied to characterize the known unknowns (Clark, 2005). Uncertainty matrices have also been suggested as a tool to qualitatively identify and document the source, type and nature of uncertainty and assess its relative priority in a table-like format (see Refsgaard et al., 2007; Koo et al., 2020).

Increased consideration of technical uncertainty (adopting the term from Walker et al., 2003) is another area which warrants further consideration in the SoS modeling context. Choice of what infrastructure and technologies to use is likely to stem from the prior experiences of the team(s) involved. Constituent models may be run on different infrastructure than was originally intended, especially as issues around computational reproducibility are addressed (Barba, 2019; Hutton et al., 2016). Identical code run under different computational environments may produce different results (see for example Bhandari Neupane et al., 2019). Such infrastructure may differ in physical or virtual architecture (e.g., laptop, supercomputer, or operating systems) or method of generating/interpreting code (e.g., different languages, compilers, package versions). Various combinations of these may be used and may also differ in the development and application phases. For these reasons the influences of different and interoperating infrastructure are important considerations (Iwanaga et al., 2020).

Correlation between parameters is another issue that is often ignored in the characterization and attribution of uncertainty (Do and Razavi, 2020). Correlation refers to statistical dependency between parameters. It is different from interaction effects which refer to the presence of non-additive operations among two or more factors embedded in constitutive equations of the model. In SoS modeling the issue is further escalated as possible correlations between the factors of different models needs to be accounted for. Ignoring correlations can falsify any estimation of uncertainty (Do and Razavi, 2020).

3.2.5. Testing and evaluation

Testing and evaluation can assist in the assessment of the ramifications of scale choice. In this step reasonableness of model structure and interpretability of relationships within models are assessed along with the traditional analysis of model behavior. Not all outputs produced by the constituent models may be relevant for the SoS model purpose and the validity of their outputs are affected due to the integrated nature of SoS modeling. For any evaluation to be effective, the specific model outputs of interest that are relevant for the model purpose must be well understood. Outputs may be at a particular spatio-temporal scale, for instance a long-term average of a model output over a large spatial domain or an extreme event at a specific point location. Issues may also stem from the conceptual suitability of constituent models as uncertainty may be propagated throughout and may compound as more models are integrated (Dunford et al., 2015). Thus, the first step in testing and evaluation involves attempting to refute aspects of SoS model structure and functional relationships within the model based on their lack of correspondence with the represented system and the model outputs. Stakeholders could be leveraged to evaluate the conceptual alignment and appropriateness of the SoS representation at the selected scales.

Evaluation of the behavioral relationships at the integrated level is similar to scientific hypothesis testing (Wilson et al., 2017) or "conceptual testing" (Iwanaga et al., 2020) wherein functional relationships within the SoS model are examined. Such tests may be especially useful in cases where the internal workings of a model are inaccessible or otherwise unknown but expected behavior of the constituent model in the integrated context can be characterized (Iwanaga et al., 2020). These approaches can be used to identify impossible or implausible aspects of the SoS model output. If any aspect of model structure or any functional relationship within the model can be shown to be an inadequate representation of the corresponding aspects of the real system, then that particular portion of the model is refuted (Li et al., 2016). Examination of model behavior over a range of inputs will also help to expose additional inadequacies in the model (Bennett et al., 2013).

The interesting aspect in this regard is that successful testing and evaluation of the constituent models does not guarantee correctness of the SoS model and vice versa. Testing and evaluation may happen at different scale levels, and acceptable model behavior depends on the model purpose and consequent measures or indicators of interest. Model behavior of constituent models could be examined quantitatively through assessment of the intermediate data in the models to ensure their behavior is consistent with a priori expectations.

It is necessary to test the software used to interoperate data across the different hierarchical levels using relevant testing approaches. These include checking the mapping of input-outputs between models, conversion of units, use of metadata to perform semantic operations, and translation of spatial temporal dimensions (Ayllón et al., 2018; Belete et al., 2017; Voinov and Shugart, 2013). Testing processes found in software engineering may additionally aid in conducting such checks (see for example, Laukkanen et al., 2017; Verweij et al., 2010; Yoo and Harman, 2012).

It may also be possible that some data gaps or uncertainties from constituent models have a lesser or negligible effect on the SoS model depending on how the constituent model is leveraged at the SoS level. Furthermore, constituent models may present overlapping and/or conflicting data or assumptions that will only be revealed when testing and evaluating their integration. A common example is *double counting* uncertainty due to embedded assumptions in the model or failure to detect correlated variables with a common cause.

The next step focuses more specifically on the correspondence between model projections and observed data. Strictly speaking, data used in model testing and evaluation must be independent of data used to develop the model (Raick et al., 2006). A variety of visual, statistical, and machine learning methods are widely used to evaluate SoS models. The choice of method, however, should be based on the fundamental questions of what scenarios and observations to use in the evaluation. Evaluation of models under the range of conditions similar to those of interest can aid in identifying limitations of the model (Ramaswami et al., 2005).

Sensitivity analysis is now regarded as standard practice in modeling (Norton, 2015; Pianosi et al., 2016; Razavi and Gupta, 2015). The sensitivity of SoS model behavior to changes to its constituents and their interactions is the target of the assessment (Moriasi et al., 2007). An issue stemming from the likely overparameterization of constituent models is equifinality and the lack of identifiability. Equifinality refers to the phenomenon of different implementations or combinations of model structure, parameter values, and their interactions producing equally acceptable results (Wagener et al., 2003; Beven, 2006). Identifiability then refers to the ability to attribute the influence on model outputs to unique model parameters or structure (Muñoz et al., 2014; Guillaume et al., 2019). Therefore, the greater the number of parameters, the less identifiable the model becomes.

Sensitivities are assessed as part of identifiability analysis, typically by ranking parameters based on their influence on outputs which can aid in determining what parameters require focused efforts to reduce uncertainty or improve identifiability (e.g. Factor Prioritization; Nossent and Bauwens, 2012). Information from sensitivity and identifiability analysis can then aid in simplifying the model (as discussed in the previous section). Similar to what was noted in Section 3.2.3, naively applying sensitivity and identifiability analysis without consideration of the SoS context may adversely affect modeling outcomes.

Assessment of sensitivities would ideally rely on global, rather than local analyses for reasons that have been expounded in prior literature (see for example Pianosi et al., 2016; Saltelli and Annoni, 2010). Use of global sensitivity analyses in model assessment has seen increasing use, despite the lack of uptake or reported use of available software tools to conduct such analyses (Douglas-Smith et al., 2020). Still, the importance of such analyses tends to be under-appreciated (Saltelli et al., 2019).

One practical reason for the lack of global sensitivity analyses is that they are typically computationally expensive to perform and the SoS models themselves typically exhibit long runtimes. Dependencies and correlations between parameters across constituent models and their respective scales pose another challenge (Koo et al., 2020). Metamodeling (expanded on in the next section) along with recently developed sampling and analysis methods may be more amenable to the SoS context. Examples of such methods that warrant further investigation include moment-independent methods (such as PAWN; Pianosi and Wagener, 2015) which can be applied independent of the sampling scheme used, and variogram-based approaches (e.g. STAR-VARS; Razavi and Gupta, 2015) which can reportedly account for temporal and spatial correlations. Adaptive sampling of the parameter space, through sparse-grids for example, in combination with these analysis techniques, may also aid in reducing the computational costs associated with sensitivity and uncertainty analyses (Buzzard and Xiu, 2011; Xiong et al., 2010).

3.3. Application phase

A critical aspect in the *application* of SoS models is that constituent models evolve independently. Development of each constituent model, by necessity, is led by disciplinary experts and undergoes separate, asynchronous, development cycles. As each model may come from different paradigms and sources of knowledge, the implementation may be adjusted over time or even replaced in response to newly acquired knowledge. Advancing towards trial model applications using the expected type and volume of data as early, quickly and often as possible allows modelers to encounter issues in the model application earlier in the process (Warren, 2014). Experience gained with each iteration subsequently serves to rectify and protect against future application challenges. Application of the model then requires monitoring and scrutinizing to ensure the underlying models (including their metadata, represented knowledge and application context) remain current and appropriate.

When models are integrated, the runtime may prevent practical application for its primary purpose, such as social learning through interactive use with stakeholders, or for global sensitivity analyses. One option to overcome this problem is to simplify the constituent models for the specific purpose. Doing so requires a high degree of knowledge of the constituent models, however, and may not be practical in cases where legacy models are used. Spatially explicit models can especially be a problem in regard to runtime, and a solution for reduction in computational burden may be achieved through aggregating grid cells into similar zones (e.g. groundwater model aggregated into hydraulic conductivity zones; Elsawah et al., 2017).

In cases of high runtime, replacing the most computationally expensive constituent models with metamodels may be a viable option.

Metamodels approximate the input-output behavior of the original model (Castelletti et al., 2012; Christelis and Hughes, 2018; Pietzsch et al., 2020) and therefore provide simplified representation(s) of more complex models (Asher et al., 2015; Razavi et al., 2012). Metamodels leverage the emergent simplicity of complex systems and although there are a variety of methods available to accomplish this, generally metamodels require the complex models (i.e. the original constituent models) to be available beforehand. Metamodels, being approximations of an original model's response surface, are most relevant to the conditions existing in the datasets upon which they are tuned, so care needs to be taken if using them under conditions that transcend those extant in the data. System forcing data beyond that experienced, such as climate change or groundwater extractions, are of particular concern in this regard. If possible, simply allocating more computational resources (e.g. supercomputers) may be the most pragmatic and resource efficient alternative, especially considering the time taken to investigate and implement the options listed above. It is acknowledged, however, that more computational capacity may not be available.

3.3.1. Analysis and visualization

In the management context, where SoS models are typically applied, there is a need to adequately describe the level of uncertainties in the SoS model and its predictions. Individual stakeholders may react differently to uncertainties and levels of uncertainty (Cockerill et al., 2019). Presenting scenario results relative to the modeled baseline neatly reduces the inherent biases that come with relying on stakeholder preferences to inform desirable thresholds, as would usually occur in multi-criteria, or multi-objective, analysis approaches (Maier et al., 2016; Martin et al., 2017; Reichert and Borsuk, 2005). With such an approach, the acceptability of a (possible) maximum or minimum relative change becomes the focus of stakeholder discussion.

Software tooling for supporting analyses of model results (including sensitivity and uncertainty analyses) typically necessitates interaction between the analysis software and the model(s), which may require the development of additional interfaces (i.e. code or supporting software). Due to the number of models involved, the associated parameters, and the possibly dynamic model structure (Wirtz and Nowak, 2017), maintaining these interfaces in the SoS context may quickly become unwieldy. Additionally, it may be desirable to replace entire models to analyze the influence of model structure and the scales they represent (Ewert et al., 2011), thus potentially rendering existing interfaces obsolete. Recent efforts circumvent this issue by supporting the near-seamless transition between the nested hierarchical representation common in SoS design to the conceptually simpler "flat" structure expected in typical analyses (e.g. Schouten and Deits, 2020). An example of nested and flattened representations of a node network is provided in Appendix 1.

A common requirement shared with tooling for conducting analyses (e.g. for sensitivity and uncertainty analysis, and exploratory modeling) is the provision and definition of parameter values. These may consist of a "default" value, a range within which values may vary, whether these values are categorical, scalar, or regarded as constants (examples may be found in Adams et al., 2014; Kwakkel, 2017; Pianosi et al., 2015; Razavi et al., 2019). Categorical values may indicate substitution with other data types or a collection of data types (e.g. rasters, climate sequences, etc.). Such information may be the minimum necessary to conduct such analyses, to reproduce and replicate results, and to support later automation of these activities. Parameter values in effect represent dimensions of scale and the inappropriate selection of their values and ranges may result in misleading results (Shin et al., 2013; Wagener and Pianosi, 2019).

3.4. Perpetuation phase

As in Badham et al. (2019), *perpetuation* is about the intended influence the modeling is to have into the future. The focus here is on the scale of documentation and process evaluation in SoS modeling which is informed by the level of consensus among stakeholders and modelers as to its purpose. In the research context, for example, there is a newfound expectation that the model be developed and provided in a manner that supports reproducibility and replicability. Reproducibility is the ability to recreate results, whereas replicability captures the ability of the model to generate new but consistent data in other applications (Patil et al., 2016).

Where SoS models are used by external stakeholders, some amount of technical support is likely expected. Without this, use of the model and thus its impact is likely to be minimal. Computational models are software in that they are made of code, and so continued use comes with a baseline cost to cover maintenance, improvements, and updating of documentation. Such capacity is crucial in contexts where long-term management and decision support is an acknowledged requirement. In such cases the design, implementation and documentation of the model should plan for these long-term activities from the beginning. In the SoS context this implies retaining the interdisciplinary knowledge within a team or organization (e.g. Cockerill et al., 2019; Kragt et al., 2013).

3.4.1. Documentation

Whereas earlier sections spoke to the content of documentation, this section focuses on the role of documentation in an interdisciplinary setting such as SoS modeling. Documentation is a conduit through which information and knowledge are propagated and provides the necessary context for model evaluation (Cockerill et al., 2019). Without sufficient documentation, it is difficult to understand the context that led to any specific issue, including mismatches between constituent models. Lack of context then affects the perceived validity of the model conceptualization, restricts model use, rendering the model inappropriate or invalid for its purpose.

The act of documenting itself allows for reflexive and transparent communication and for new insights to be gained. Undocumented assumptions regarding scale and their influence may compromise other constituent models, thus holistic awareness of the SoS issues can be obstructed by a lack of documentation. Long-term maintenance and use of the model may also be impeded (Ahalt et al., 2014). No individual holds the knowledge and awareness of the modeling details in their entirety, let alone the effects of interactions between models. It is therefore important to recognize that writing and maintaining documentation should be a team effort, and a culture to support this should be fostered.

In practice there are few incentives for documenting models to such an extent. A key problem in SoS model documentation is that details of the constituent models important for the SoS team may be considered unnecessary for the teams developing the constituent models. Once again, this stems from potential disconnects between the purpose of the SoS model and the individual (or original) objectives of each constituent model. In the sciences the focus is often on the publication of papers at the expense of ensuring model reuse or reproducibility and replicability (Easterbrook, 2014; Joppa et al., 2013; Peng, 2011; Schnell, 2018). There is an increasing push to change the culture surrounding the publication process, however, to better recognize, credit and incentivize model code publication. For example, a number of organizations have begun supporting "Open Code Badges" to highlight reproducible work (https://www.comses.net/resources/open-code-badge/).

3.4.2. Process evaluation

The extent to which the modeling has achieved its overarching purpose is evaluated in this step (Badham et al., 2019). This evaluation extends beyond the technical performance of the SoS model (Bennett et al., 2013) to consider outcomes of modeling as a social process. Success of a model depends on the beliefs and expectations of the intended users and in their satisfaction with the model and its results (Hamilton et al., 2019). It may also depend on the biases and beliefs of the model creators (Glynn et al., 2017) and in an alignment of expectations between creators and users (Sterling et al., 2019). The suitability of the success criteria is dependent on the context of the project, including not only the model purpose, but also the characteristics of the problem, such as its complexity and the resources that were available (Hamilton et al., 2019).

Process evaluation in SoS focuses on two facets: achievement of goals and longevity of the models. In terms of goal achievement, process evaluation considers whether the goals of the SoS model were supported by its constituent models and, where applicable, whether constituent models achieved their own goals. Although satisfying the goals of the constituent models may seem an indirect path to satisfying the goals of the SoS model, this interpretation is misleading. An SoS approach to modeling, instead of simply a multi-modeling approach, leverages the autonomy and independence of the constituent models. Constituent models still need to be capable of yielding their own outcomes, regardless of how those models are used in the context of the SoS model (Salado, 2015).

Evaluation of the longevity of the SoS model, referring to the ability to leverage or reuse the SoS model over time, requires the development and assessment of a targeted plan for its sustainment that includes: (1) monitoring the evolution of the constituent models; (2) identifying alternatives for models that may cease their validity, availability or accessibility during the lifetime of the SoS model; (3) establishing a strategy for the continued evolution of the SoS model, including the development of potential transformation frameworks and implementations; and (4) identifying opportunities to facilitate the sustainment of constituent systems aligned with the sustainment of the SoS model.

Process evaluation for SoS models may consider adopting a reflexive process in which questions are asked of those involved in the modeling, such as 'did the modeling process help to improve understanding of the system/problem?' or 'did the modeling process help facilitate communication between stakeholders?' (Hamilton et al., 2019). The line of questioning can then leverage input from the various perspectives available, including those of experts and stakeholders for the different constituent systems of an SoS. Bias in the model, such as whether their respective positions were adequately represented, may then be assessed. Alternative conceptions and processes of the system and their scales could also be assessed at this stage (Voinov et al., 2016).

4. The paths forward

4.1. A grander vision and commensurate funding

Addressing all the scale-related issues outlined in the paper requires a level of cooperation and concerted integrative effort that is by and large not possible given the usual short-term funding of the sciences (e.g. Saltelli, 2018). Recent publications have also brought attention to deficiencies in the current science resourcing structure, characterized in part by competition over limited funding and an emphasis on (number and citation counts of) publications. Existing funding mechanisms may well be detrimental to the quality of science produced (Binswanger, 2014; Sandström and Besselaar, 2018).

Limited resourcing is one reason for the multiple, albeit siloed, efforts with a focus on single case studies (Pulver et al., 2018; Hoekstra et al., 2014), and the necessity of excluding salient aspects of the modeling (such as adequate participatory processes; Eker et al., 2018) or making less than ideal choices about the model or data (e.g. using existing coarser scale data rather than collecting new data at a finer scale). Commentary by researchers highlight the importance of interdisciplinary work (Kretser et al., 2019; Meirmans et al., 2019), which is typically not funded to the same extent as monodisciplinary efforts (Kwon et al., 2017; Bromham et al., 2016). Regardless of the importance of such holistic assessments these real-world constraints essentially make holistic SoS modeling and analyses unrealistic.

On the other hand, examples of large concerted efforts can be found, such as in astronomy and physics which have produced groundbreaking work with the Event Horizon Telescope (e.g. first photograph of a blackhole, Akiyama, 2019) and the Large Hadron Collider (e.g. discovery of the Higgs boson, Aad et al., 2012). These resource intensive projects are important and could substantially influence future societal development. At the same time, lesser importance is placed by funding organizations on interdisciplinary socio-environmental works which arguably have a more immediate impact and benefit to society.

A grander vision for SoS research, in line with large-scale collaborations in other fields, is vital to achieve a truly holistic consideration of SoS modeling for resolving socio-environmental issues. Realizing this vision itself requires fundamental shifts in how such interdisciplinary work, and associated expertise, are viewed and funded (Elsawah et al., 2020). Greater funding focused on education and training of interdisciplinary system practitioners is fundamental for greater cohesion and consensus in the socio-environmental sciences (Little et al., 2019). While alternative funding models have been suggested for the sciences (see for example Meirmans et al., 2019; Higginson and Munafò, 2016), the current state of affairs is unlikely to change in the near future. Thus, any benefits from a systemic change, if they occur at all, will be experienced only in the long-term.

Although disciplinary experts may collaborate, pool resources, engage with stakeholders and gain experience in interdisciplinary work in the process of investigating a socio-environmental issue, this is not an effective way forward. In the medium-term, existing case studies could be leveraged to perform a comparative meta-analysis to determine the level of influence system connections have, and the scales at which such connections matter (Pulver et al., 2018). Such meta-analyses could extend to the practices used to manage the socio-technical influences in the modeling process. Shifts towards leveraging collections of studies for meta-analyses are emerging in fields such as psychology to allow for what is known as "statistical objectivity" towards reported findings in the literature (Freese and Peterson, 2018). Although the focus there is in resolving issues of replicability, the same approach can be additionally leveraged to characterize scale commonalities.

We conclude here by re-emphasizing three key considerations which can reinforce current SoS modeling efforts in a move towards the larger consensus needed for this grander vision.

4.2. Strengthen interdisciplinary communication

Here lies the crux of the challenge in developing a tiered SoS model. It is not only necessary for the science and engineering to mesh together appropriately, but it is fundamental that the modeling process also consider and embed the socio-technical considerations. While we as modelers struggle with the former, the latter is too often ignored. As there are a variety of participants, and therefore disciplinary perspectives involved, a key set of considerations are in the social dimensions that provide the interface between modeling efforts.

Integrating multiple perspectives requires an integrative approach which is ultimately necessary to navigate towards a beneficial system change (why else do we model?). Choices made in the treatment of scale are unavoidable and may result in conflicting decisions with separate implications. Just to name one, members of teams may have a path preselected without full consideration of the implications on the system representations, leading to further issues when such decisions are not communicated.

The next generation of systems modelers would ideally embody a culture that is cognizant of the socio-technical issues, considerations, and their influences throughout the modeling process (e.g. Little et al., 2019). Such a systemic cultural shift can only be developed in the longer term, however, and so in the meantime clearer communication requires adequate resourcing for documenting decisions made, and code and data used, including their maintenance. Practices for the co-production of knowledge to fulfill the needs and requirements of the modeling is necessary for advances to be made (Norström et al., 2020).

There is often a preference for face-to-face meetings to facilitate the necessary level of communication but that may not always be possible. Geographic distance, scheduling conflicts, travel restrictions and other factors may preclude such activities. Communication technologies play a critical role in mitigating some aspects of the issue. For example, travel and social distancing restrictions during the COVID-19 pandemic has prohibited many teams from meeting in person, forcing reliance on technologies such as video conferencing. Regardless of the mode of communication, a team and organizational culture of consistent and continual communication is one necessity repeatedly highlighted to resolve a variety of scale issues and the conflict that may arise between actors throughout the modeling process. Incorporating knowledge beyond the bounds of one's own disciplinary training is crucial to the holistic attention to and incorporation of scales and to avoid the siloing of information and knowledge, and to break down cognitive constraints.

4.3. Improve documentation processes

The importance of documentation is another aspect that was repeatedly raised throughout this paper. Documentation of the modeling process communicates, and makes accessible, the decisions, actions, the context of those decisions and actions, and reflection on those choices to those who may or may not have been active participants in their making. Insufficient documentation affects many aspects from the pace of model development throughout the modeling cycle, quality of model integration especially across disciplinary boundaries, and the perceived quality of the modeling conducted. A lack of documentation accessibility additionally affects the (re)use and maintenance of the SoS model (or its constituents) and so could lead to duplication of effort across those involved in modeling SESs.

One approach to ensure that documentation is made a priority is to adopt a documentation-driven development and design approach (Heeager, 2012). Such approaches are exemplified by the ODD Protocol (Grimm et al., 2020, 2014, 2010). In this paradigm, documentation is developed first, serving as a vehicle for discussion, ideally prior to any model development (Heeager, 2012). Ambiguities in the documentation (and thus the modeling) may be addressed earlier in the process as a result, and documentation could be iteratively revised, commensurate with any changes to modeling scale. Furthermore, maintaining Records of Engagement and Decision-making (RoED, Cockerill et al., 2019) to document the process and pathway decisions were made in a context-appropriate manner may be crucial to ensuring conceptual and technical validity throughout the modeling cycle. Sufficient, rather than exhaustive, documentation to describe model context would be preferred (Ambler, 2002; Cockerill et al., 2019).

4.4. Explicit consideration of scale and uncertainty

There is an increasing expectation that SoS models can more completely represent processes within an SES, however, it is impossible to model everything for all purposes. Further explicit consideration of the inter-relationships between scales, choices made in representing scale, and their influence on uncertainty is paramount in the SoS context. Identifying, managing and reconciling the disparate treatment of scale is a key step towards a holistic approach, as opposed to the concurrent, but separate, processes currently applied (Cheong et al., 2012; Elsawah et al., 2020).

As noted several times throughout this paper, the socio-technical context has an inordinate influence on uncertainty. In addition to the communication and documentation considerations outlined above, an avenue for a more holistic assessment of uncertainty includes the use of robustness analysis (Grimm and Berger, 2016). In such analysis, a model with multiple systems is systematically deconstructed through forceful changes to the model parameters, structure, and process representations within each system to assess uncertainty. Use of these approaches with pattern-oriented modeling processes, which filter unsuitable representations across scales, may also be helpful in this regard (Grimm and Railsback, 2012; Gupta et al., 2008).

Additionally, qualitative and quantitative uncertainties could be jointly assessed through the representation of multiple plausible futures that stem from different sets of assumptions through exploratory approaches (Maier et al., 2016; Roberts et al., 2018; Rounsevell and Metzger, 2010). A related approach is a multi-model approach wherein an ensemble of equally plausible models are applied to identify the influence of structural and qualitative uncertainty (Matott et al., 2009; Tebaldi and Knutti, 2007; Uusitalo et al., 2015). Using an ensemble of estimates (such as the average or median of model outputs) may have the benefit of providing more robust and accurate forecasts (Willcock et al., 2020). Applying these on different computational platforms may additionally assist in identifying technical uncertainties (Iwanaga et al., 2020).

It was noted throughout this paper that the scale of the modeling itself should be commensurate with the available resources and purpose. A holistic SoS model may not be entirely possible given resource constraints, however relationships between systems can still be acknowledged and represented (albeit simplistically). Doing so allows some assessment of the uncertainties at least, and constitutes a step towards holistic SoS modeling so long as the underlying assumptions are explicitly documented (e.g. Kloprogge et al., 2011).

Declaration of competing interest

None to declare.

Acknowledgements

This work was supported by the National Socio-Environmental Synthesis Center (SESYNC) under funding received from the National Science Foundation DBI-1639145. The primary author (Takuya Iwanaga) is supported through an Australian Government Research Training Program (AGRTP) Scholarship and a top-up scholarship from the ANU Hilda-John Endowment Fund. Hsiao-Hsuan Wang and Tomasz E. Koralewski acknowledge partial support from USDA, ARS Agreement No. 58-3091-6-035 with Texas A&M AgriLife Research, titled 'Areawide pest management of the invasive sugarcane aphid in grain sorghum, regional population monitoring and forecasting.' Min Chen is supported by the Key Program of NSF of China (No. 41930648). John Little acknowledges partial support from NSF Award EEC 1937012. The authors would like to thank the three anonymous reviewers and Prof. Randall Hunt (USGS) for their constructive feedback and comments. The authors additionally thank Faye Duchin and Adrian Hindes for comments provided on an earlier draft.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Appendix 1

Example of hypothetical model inputs for a hydrological routing model provided in a nested data structure (left column) compared to a more traditional "flat" format (right column). Nested structures are arguably better suited for representing collections of data structures and their relationships (e.g. a network or graph structure) and, pragmatically, are typically more amenable to the inclusion of comments and multiple values associated with specific parameters, reducing cognitive overhead. While perhaps more readable, a disadvantage of nested representations is the additional complexity that may be perceived.

Nested

```
999002:
  node_type: "StreamNode"
  # Interpret as links to
  # other nodes
  prev_node:
    - 999000
    - 999001
  next_node: 999003
  formula_type: 1
  node_params:
    # Default, Min, Max values
    # (assume constant if scalar)
    d: 200.0, 150.0, 225.0
    d2: 2.0, 1.5, 2.2
    e: 1.0
    f: 1.4, 1.2, 1.5
    alpha: 0.95
    a: 0.9
    b: 0.1
  initial_storage: 0.0
  storage_coef: 2.9
  area: 452.22
```

Flat, table-like

```
ID, node_type, prev_node, next_node, d, d_min,
d_max, d2, d2_min, d2_max, e, e_min, e_max, f,
f_min, f_max, alpha, alpha_min, alpha_max, a,
a_min, a_max, b, b_min, b_max,
initial_storage, storage_coef, area
999002, "StreamNode", [999000, 999001],
999003, 200.0, 150.0, 225.0, 2.0, 1.5, 2.2,
1.0, 1.0, 1.0, 1.4, 1.2, 1.5, 0.95, 0.9, 0.9,
0.9, 0.1, 0.1, 0.1, 0.0, 2.9, 452.22
```

References

Aad, G., et al., 2012. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. Phys. Lett. B 716, 1–29. https://doi.org/10.1016/j.physletb.2012.08.020.

Adams, B.M., Ebeida, M.S., Eldred, M.S., Jakeman, J.D., Swiler, L.P., Stephens, J.A., Vigil, D.M., Wildey, T.M., Bohnhoff, W.J., Dalbey, K.R., 2014. Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.0 Developers Manual.

Ahalt, S., Band, L., Christopherson, L., Idaszak, R., Lenhardt, C., Minsker, B., Palmer, M., Shelley, M., Tiemann, M., Zimmerman, A., 2014. Water science software institute: agile and open source scientific software development. Comput. Sci. Eng. 16, 18–26. https://doi.org/10.1109/MCSE.2014.5.

Akiyama, K., et al., 2019. First M87 event horizon telescope results. IV. Imaging the central supermassive black hole. The Astrophysical Journal Letters 875, L4. https://doi.org/10.3847/2041-8213/ab0e85.

Allen, T.J., Henn, G.W., 2006. The Organization and Architecture of Innovation: Managing the Flow of Technology. Routledge.

Ambler, S., 2002. Agile Modeling: Effective Practices for eXtreme Programming and the Unified Process. John Wiley & Sons, Inc., New York.

Anderson, M.P., Woessner, W.W., Hunt, R.J., 2015. Chapter 9 - model calibration: assessing performance. In: Anderson, M.P., Woessner, W.W., Hunt, R.J. (Eds.), Applied Groundwater Modeling, second ed. Academic Press, San Diego, pp. 375–441. https://doi.org/10.1016/B978-0-08-091638-5.00009-2.

Asher, M.J., Croke, B.F.W., Jakeman, A.J., Peeters, L.J.M., 2015. A review of surrogate models and their application to groundwater modeling. Water Resour. Res. 51, 5957–5973. https://doi.org/10.1002/2015WR016967.

Ayllón, D., Grimm, V., Attinger, S., Hauhs, M., Simmer, C., Vereecken, H., Lischeid, G., 2018. Cross-disciplinary links in environmental systems science: current state and claimed needs identified in a meta-review of process models. Sci. Total Environ. 622–623, 954–973. https://doi.org/10.1016/j.scitotenv.2017.12.007.

Badham, J., Elsawah, S., Guillaume, J.H.A., Hamilton, S.H., Hunt, R.J., Jakeman, A.J., Pierce, S.A., Snow, V.O., Babbar-Sebens, M., Fu, B., Gober, P., Hill, M.C., Iwanaga, T., Loucks, D.P., Merritt, W.S., Peckham, S.D., Richmond, A.K., Zare, F., Ames, D., Bammer, G., 2019. Effective modeling for Integrated Water Resource Management: a guide to contextual practices by phases and steps and future opportunities. Environ. Model. Software 116, 40–56. https://doi.org/10.1016/j.envsoft.2019.02.013.

Bandara, S., Schlöder, J.P., Eils, R., Bock, H.G., Meyer, T., 2009. Optimal experimental design for parameter estimation of a cell signaling model. PLoS Comput. Biol. 5, e1000558 https://doi.org/10.1371/journal.pcbi.1000558.

Banerjee, P., Wang, H.-H., Peterson, M.J., Grant, W.E., Peterson, T.R., 2019. Collaborative modeling and social learning in the context of joint forest management in east Sikkim, India. Front. Environ. Sci. 7 https://doi.org/10.3389/fenvs.2019.00154.

Bar-Yam, Y., 1997. Dynamics of Complex Systems. Perseus Books, USA.

Barba, L.A., 2019. Praxis of reproducible computational science. Comput. Sci. Eng. 21, 73–78. https://doi.org/10.1109/MCSE.2018.2881905.

Barton, W., Shan, X., 2017. Environmental, health, and safety department reorganization and prioritization: using analytical hierarchy process to mitigate scope creep. Environ. Qual. Manag. 26, 31–45. https://doi.org/10.1002/tqem.21505.

Belete, G.F., Voinov, A., Laniak, G.F., 2017. An overview of the model integration process: from pre-integration assessment to testing. Environ. Model. Software 87, 49–63. https://doi.org/10.1016/j.envsoft.2016.10.013.

Bellman, R.E., 2015. Adaptive Control Processes. Princeton University Press, Princeton.

Bellman, R., Åström, K.J., 1970. On structural identifiability. Math. Biosci. 7, 329–339. https://doi.org/10.1016/0025-5564(70)90132-X.

Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S. A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising performance of environmental models. Environ. Model. Software 40, 1–20.

Beven, K., 2006. A manifesto for the equifinality thesis. Journal of Hydrology, The model parameter estimation experiment 320, 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007.

Beven, K.J., 2009. Environmental Modelling: an Uncertain Future?: an Introduction to Techniques for Uncertainty Estimation in Environmental Prediction. Routledge, London; New York.

Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J. Hydrol. 249, 11–29. https://doi.org/10.1016/S0022-1694(01)00421-8.

Beven, K., Westerberg, I., 2011. On red herrings and real herrings: disinformation and information in hydrological inference. Hydrol. Process. 25, 1676–1680. https://doi.org/10.1002/hyp.7963.

Bhandari Neupane, J., Neupane, R.P., Luo, Y., Yoshida, W.Y., Sun, R., Williams, P.G., 2019. Characterization of leptazolines A–D, polar oxazolines from the cyanobacterium leptolyngbya sp., reveals a glitch with the "willoughby–hoye" scripts for calculating NMR chemical shifts. Org. Lett. 21, 8449–8453. https://doi.org/10.1021/acs.orglett.9b03216.

Bhattacherjee, A., Perols, J., Sanford, C., 2008. Information technology continuance: a theoretic extension and empirical test. J. Comput. Inf. Syst. 49, 17–26. https://doi.org/10.1080/08874417.2008.11645302.

Binswanger, M., 2014. Excellence by nonsense: the competition for publications in modern science. In: Bartling, S., Friesike, S. (Eds.), Opening Science: the Evolving Guide on How the Internet Is Changing Research, Collaboration and Scholarly Publishing. Springer International Publishing, Cham, pp. 49–72. https://doi.org/10.1007/978-3-319-00026-8_3.

Boone, S., Fragaszy, S., 2018. Emerging scarcity and emerging commons: Water management groups and groundwater governance in Aotearoa New Zealand. Water Alternatives 11, 795–823.

Bridle, H., Vrieling, A., Cardillo, M., Araya, Y., Hinojosa, L., 2013. Preparing for an interdisciplinary future: a perspective from early-career researchers. Futures 53, 22–32. https://doi.org/10.1016/j.futures.2013.09.003.

Bromham, L., Dinnage, R., Hua, X., 2016. Interdisciplinary research has consistently lower funding success. Nature 534, 684–687. https://doi.org/10.1038/nature18315.

Brun, R., Reichert, P., Künsch, H.R., 2001. Practical identifiability analysis of large environmental simulation models. Water Resour. Res. 37, 1015–1030. https://doi.org/10.1029/2000WR900350.

Buahin, C.A., Horsburgh, J.S., 2018. Advancing the open modeling interface (OpenMI) for integrated water resources modeling. Environ. Model. Software 108, 133–153. https://doi.org/10.1016/j.envsoft.2018.07.015.

Buzzard, G.T., Xiu, D., 2011. Variance-based global sensitivity analysis via sparse-grid interpolation and cubature. Commun. Comput. Phys. 9, 542–567. https://doi.org/10.4208/cicp.230909.160310s.

Cash, D.W., Clark, W.C., Alcock, F., Dickson, N.M., Eckley, N., Guston, D.H., Jäger, J., Mitchell, R.B., 2003. Knowledge systems for sustainable development. Proc. Natl. Acad. Sci. Unit. States Am. 100, 8086–8091. https://doi.org/10.1073/pnas.1231332100.

Cash, D., Adger, W.N., Berkes, F., Garden, P., Lebel, L., Olsson, P., Pritchard, L., Young, O., 2006. Scale and cross-scale dynamics: governance and information in a multilevel world. Ecol. Soc. 11 https://doi.org/10.5751/ES-01759-110208.

Castelletti, A., Galelli, S., Ratto, M., Soncini-Sessa, R., Young, P.C., 2012. A general framework for Dynamic Emulation Modelling in environmental problems. Environmental Modelling & Software, Emulation techniques for the reduction and sensitivity analysis of complex environmental models 34, 5–18. https://doi.org/10.1016/j.envsoft.2012.01.002.

Chen, C.-F., Ma, H., Reckhow, K.H., 2007. Assessment of water quality management with a systematic qualitative uncertainty analysis. Sci. Total Environ. 374, 13–25. https://doi.org/10.1016/j.scitotenv.2006.12.027.

Chen, M., Voinov, A., Ames, D.P., Kettner, A.J., Goodall, J.L., Jakeman, A.J., Barton, M.C., Harpham, Q., Cuddy, S.M., DeLuca, C., Yue, S., Wang, J., Zhang, F., Wen, Y., Lü, G., 2020. Position paper: Open web-distributed integrated geographic modelling and simulation to enable broader participation and applications. Earth Sci. Rev. 207, 103223. https://doi.org/10.1016/j.earscirev.2020.103223.

Chen, M., Yue, S., Lü, G., Lin, H., Yang, C., Wen, Y., Hou, T., Xiao, D., Jiang, H., 2019. Teamwork-oriented integrated modeling method for geo-problem solving. Environ. Model. Software 119, 111–123. https://doi.org/10.1016/j.envsoft.2019.05.015.

Cheong, S.-M., Brown, D.G., Kok, K., Lopez-Carr, D., 2012. Mixed methods in land change research: towards integration. Trans. Inst. Br. Geogr. 37, 8–12. https://doi.org/10.1111/j.1475-5661.2011.00482.x.

Christelis, V., Hughes, A.G., 2018. Metamodel-assisted analysis of an integrated model composition: an example using linked surface water – groundwater models. Environ. Model. Software 107, 298–306. https://doi.org/10.1016/j.envsoft.2018.05.004.

Clark, J.S., 2005. Why environmental scientists are becoming Bayesians. Ecol. Lett. 8, 2–14.

Cockerill, K., Glynn, P., Chabay, I., Farooque, M., Hämäläinen, R.P., Miyamoto, B., McKay, P., 2019. Records of engagement and decision making for environmental and socio-ecological challenges. EURO J Decis Process. https://doi.org/10.1007/s40070-019-00104-6.

Connor, T.D., Vin'a, A.R., Winkler, J.A., Hull, V., Tang, Y., Shortridge, A.M., Yang, H., Zhao, Z., Wang, F., Zhang, J., Zhang, Z., Zhou, C.-Q., Bai, W., Liu, J., 2019. Interactive spatial scale effects on species distribution modeling: the case of the giant panda. Sci. Rep. 9 https://doi.org/10.1038/s41598-019-50953-z.

Cooke, N.J., Hilton, M.L., Science, C., on the S. of T., Board on Behavioral, C., Education, D. of B. and S.S. and, Council, N.R, 2015. Supporting Virtual Collaboration, Enhancing the Effectiveness of Team Science. National Academies Press (US).

Cresswell, K.M., Worth, A., Sheikh, A., 2010. Actor-Network Theory and its role in understanding the implementation of information technology developments in healthcare. BMC Med. Inf. Decis. Making 10, 67. https://doi.org/10.1186/1472-6947-10-67.

Dahmann, J.S., Baldwin, K.J., 2008. Understanding the current state of US defense systems of systems and the implications for systems engineering. In: 2008 2nd Annual IEEE Systems Conference. Presented at the 2008 2nd Annual IEEE Systems Conference, pp. 1–7. https://doi.org/10.1109/SYSTEMS.2008.4518994.

Daniell, K.A., Barreteau, O., 2014. Water governance across competing scales: coupling land and water management. Journal of Hydrology, Water governance across competing scales: Coupling land and water management 519, 2367–2380. https://doi.org/10.1016/j.jhydrol.2014.10.055.

De Dreu, C.K.W., 2008. The virtue and vice of workplace conflict: food for (pessimistic) thought. J. Organ. Behav. 29, 5–18. https://doi.org/10.1002/job.474.

Do, N.C., Razavi, S., 2020. Correlation effects? A major but often neglected component in sensitivity and uncertainty analysis. Water Resour. Res. 56 https://doi.org/10.1029/2019WR025436 e2019WR025436.

Douglas-Smith, D., Iwanaga, T., Croke, B.F.W., Jakeman, A.J., 2020. Certain trends in uncertainty and sensitivity analysis: an overview of software tools and techniques. Environ. Model. Software 124, 104588. https://doi.org/10.1016/j.envsoft.2019.104588.

Duchin, F., 2016. A global case-study framework applied to water supply and sanitation. J. Ind. Ecol. 20, 387–395. https://doi.org/10.1111/jiec.12462.

Duchin, F., Levine, S.H., 2019. The recovery of products and materials for reuse: the global context of resource management. Resour. Conserv. Recycl. 145, 422–447. https://doi.org/10.1016/j.resconrec.2018.10.028.

Dunford, R., Harrison, P.A., Rounsevell, M.D.A., 2015. Exploring scenario and model uncertainty in cross-sectoral integrated assessment approaches to climate change impacts. Climatic Change 132, 417–432. https://doi.org/10.1007/s10584-014-1211-3.

Easterbrook, S.M., 2014. Open code for open science? Nat. Geosci. 7, 779–781. https://doi.org/10.1038/ngeo2283.

Eker, S., Rovenskaya, E., Obersteiner, M., Langan, S., 2018. Practice and perspectives in the validation of resource management models. Nat. Commun. 9, 5359. https://doi.org/10.1038/s41467-018-07811-9.

Elag, M.M., Goodall, J.L., Castronova, A.M., 2011. Feedback loops and temporal misalignment in component-based hydrologic modeling. Water Resour. Res. 47 https://doi.org/10.1029/2011WR010792 10.1016/j.envsoft.2010.01.003.

Elsawah, S., Pierce, S.A., Hamilton, S.H., van Delden, H., Haase, D., Elmahdi, A., Jakeman, A.J., 2017. An overview of the system dynamics process for integrated modelling of socio-ecological systems: lessons on good modelling practice from five case studies. Environ. Model. Software 93, 127–145. https://doi.org/10.1016/j.envsoft.2017.03.001.

Elsawah, S., Filatova, T., Jakeman, A.J., Kettner, A.J., Zellner, M.L., Athanasiadis, I.N., Hamilton, S.H., Axtell, R.L., Brown, D.G., Gilligan, J.M., Janssen, M.A., Robinson, D.T., Rozenberg, J., Ullah, I.I.T., Lade, S.J., 2020. Eight grand challenges in socio-environmental systems modeling, 16226–16226. https://doi.org/10.18174/sesmo.2020a16226, 12.

Elshafei, Y., Sivapalan, M., Tonts, M., Hipsey, M.R., 2014. A prototype framework for models of socio-hydrology: identification of key feedback loops and parameterisation approach. Hydrol. Earth Syst. Sci. 18, 2141–2166. https://doi.org/10.5194/hess-18-2141-2014.

Eusgeld, I., Nan, C., Dietz, S., 2011. "System-of-systems" approach for interdependent critical infrastructures. Reliability Engineering & System Safety. ESREL 2009 Special Issue 96, 679–686. https://doi.org/10.1016/j.ress.2010.12.010.

Ewert, F., van Ittersum, M.K., Heckelei, T., Therond, O., Bezlepkina, I., Andersen, E., 2011. Scale changes and model linking methods for integrated assessment of agri-environmental systems. Agriculture, Ecosystems & Environment. Scaling methods in integrated assessment of agricultural systems 142, 6–17. https://doi.org/10.1016/j.agee.2011.05.016.

Falcone, J.-L., Chopard, B., Hoekstra, A., 2010. MML: towards a multiscale modeling language. Procedia Computer Science, ICCS 1, 819–826. https://doi.org/10.1016/j.procs.2010.04.089, 2010.

Freeman, R.E., 2010. Strategic Management: a Stakeholder Approach, Reissue. Cambridge University Press, Cambridge New York Melbourne Madrid Cape Town Singapore.

Freese, J., Peterson, D., 2018. The emergence of statistical objectivity: changing ideas of epistemic vice and virtue in science: sociological theory. https://doi.org/10.1177/0735275118794987.

Fridman, D., Kissinger, M., 2019. A multi-scale analysis of interregional sustainability: applied to Israel's food supply. Sci. Total Environ. 676, 524–534. https://doi.org/10.1016/j.scitotenv.2019.04.054.

Gibson, C.C., Ostrom, E., Ahn, T.K., 2000. The concept of scale and the human dimensions of global change: a survey. Ecol. Econ. 32, 217–239. https://doi.org/10.1016/S0921-8009(99)00092-0.

Glynn, P.D., Voinov, A.A., Shapiro, C.D., White, P.A., 2017. From data to decisions: processing information, biases, and beliefs for improved management of natural resources and environments. Earth's Future 5, 356–378. https://doi.org/10.1002/2016EF000487.

Gokhale, A.A., 2001. Environmental initiative prioritization with a Delphi approach: a case study. Environ. Manag. 28, 187–193. https://doi.org/10.1007/s002670010217.

Gorddard, R., Colloff, M.J., Wise, R.M., Ware, D., Dunlop, M., 2016. Values, rules and knowledge: adaptation as change in the decision context. Environ. Sci. Pol. 57, 60–69. https://doi.org/10.1016/j.envsci.2015.12.004.

Grant, W.E., Swannack, T.M., 2007. Ecological Modeling: A Common-Sense Approach to Theory and Practice. Wiley-Blackwell, Malden, Massachusetts.

Gray, C.T., Marwick, B., 2019. Truth, proof, and reproducibility: there's No counter-attack for the codeless. In: Nguyen, H. (Ed.), Statistics and Data Science, Communications in Computer and Information Science. Springer, Singapore, pp. 111–129. https://doi.org/10.1007/978-981-15-1960-4_8.

Gray, S., Chan, A., Clark, D., Jordan, R., 2012. Modeling the integration of stakeholder knowledge in social–ecological decision-making: benefits and limitations to knowledge diversity. Ecological Modelling, Modeling Human Decisions 229, 88–96. https://doi.org/10.1016/j.ecolmodel.2011.09.011.

Grimm, V., Berger, U., 2016. Robustness analysis: deconstructing computational models for ecological theory and applications. Ecological Modelling, Next generation ecological modelling, concepts, and theory: structural realism, emergence, and predictions 326, 162–167. https://doi.org/10.1016/j.ecolmodel.2015.07.018.

Grimm, V., Railsback, S.F., 2012. Pattern-oriented modelling: a 'multi-scope' for predictive systems ecology. Phil. Trans. R. Soc. B 367, 298–310. https://doi.org/10.1098/rstb.2011.0180.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S.K., Huse, G., Huth, A., Jepsen, J.U., Jørgensen, C., Mooij, W.M., Müller, B., Pe'er, G., Piou, C., Railsback, S.F., Robbins, A.M., Robbins, M.M., Rossmanith, E., Rüger, N., Strand, E., Souissi, S., Stillman, R.A., Vabø, R., Visser, U., DeAngelis, D.L., 2006. A standard protocol for describing individual-based and agent-based models. Ecol. Model. 198, 115–126. https://doi.org/10.1016/j.ecolmodel.2006.04.023.

Grimm, V., Berger, U., DeAngelis, D.L., Polhill, J.G., Giske, J., Railsback, S.F., 2010. The ODD protocol: a review and first update. Ecol. Model. 221, 2760–2768. https://doi.org/10.1016/j.ecolmodel.2010.08.019.

Grimm, V., 2005. Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology. Science 310, 987–991. https://doi.org/10.1126/science.1116681.

Grimm, V., Augusiak, J., Focks, A., Frank, B.M., Gabsi, F., Johnston, A.S.A., Liu, C., Martin, B.T., Meli, M., Radchuk, V., Thorbek, P., Railsback, S.F., 2014. Towards better modelling and decision support: documenting model development, testing, and analysis using TRACE. Ecological Modelling, Population Models for Ecological Risk Assessment of Chemicals 280, 129–139. https://doi.org/10.1016/j.ecolmodel.2014.01.018.

Grimm, V., Railsback, S.F., Vincenot, C.E., Berger, U., Gallagher, C., DeAngelis, D.L., Edmonds, B., Ge, J., Giske, J., Groeneveld, J., Johnston, A.S.A., Milles, A., Nabe-Nielsen, J., Polhill, J.G., Radchuk, V., Rohwäder, M.-S., Stillman, R.A., Thiele, J.C., Ayllón, D., 2020. The ODD protocol for describing agent-based and other simulation models: a second update to improve clarity, replication, and structural realism. JASSS 23, 7. https://doi.org/10.18564/jasss.4259.

Groen, D., Knap, J., Neumann, P., Suleimenova, D., Lourens, V., Kenneth, L., 2019. Mastering the scales: a survey on the benefits of multiscale computing software. Phil. Trans. Math. Phys. Eng. Sci. 377, 20180147. https://doi.org/10.1098/rsta.2018.0147.

Guillaume, J.H.A., Jakeman, J.D., Marsili-Libelli, S., Asher, M., Brunner, P., Croke, B., Hill, M.C., Jakeman, A.J., Keesman, K.J., Razavi, S., Stigter, J.D., 2019. Introductory overview of identifiability analysis: a guide to evaluating whether you have the right type of data for your modeling purpose. Environ. Model. Software 119, 418–432. https://doi.org/10.1016/j.envsoft.2019.07.007.

Gupta, H.V., Wagener, T., Liu, Y., 2008. Reconciling theory with observations: elements of a diagnostic approach to model evaluation. Hydrol. Process. 22, 3802–3813. https://doi.org/10.1002/hyp.6989.

Hamilton, S.H., ElSawah, S., Guillaume, J.H.A., Jakeman, A.J., Pierce, S.A., 2015. Integrated assessment and modelling: overview and synthesis of salient dimensions. Environ. Model. Software 64, 215–229. https://doi.org/10.1016/j.envsoft.2014.12.005.

Hamilton, S.H., Fu, B., Guillaume, J.H.A., Badham, J., Elsawah, S., Gober, P., Hunt, R.J., Iwanaga, T., Jakeman, A.J., Ames, D.P., Curtis, A., Hill, M.C., Pierce, S.A., Zare, F., 2019. A framework for characterising and evaluating the effectiveness of environmental modelling. Environ. Model. Software 118, 83–98. https://doi.org/10.1016/j.envsoft.2019.04.008.

Hannah, D.M., Wood, P.J., Sadler, J.P., 2004. Ecohydrology and hydroecology: a 'new paradigm'? Hydrol. Process. 18, 3439–3445. https://doi.org/10.1002/hyp.5761.

Häyhä, T., Lucas, P.L., van Vuuren, D.P., Cornell, S.E., Hoff, H., 2016. From Planetary Boundaries to national fair shares of the global safe operating space — how can the scales be bridged? Global Environ. Change 40, 60–72. https://doi.org/10.1016/j.gloenvcha.2016.06.008.

Heeager, L.T., 2012. Introducing agile practices in a documentation-driven software development practice: a case study. Journal of Information Technology Case and Application Research 14, 3–24. https://doi.org/10.1080/15228053.2012.10845693.

Hidalgo, E.S., 2019. Adapting the scrum framework for agile project management in science: case study of a distributed research initiative. Heliyon 5, e01447. https://doi.org/10.1016/j.heliyon.2019.e01447.

Higginson, A.D., Munafò, M.R., 2016. Current incentives for scientists lead to underpowered studies with erroneous conclusions. PLoS Biol. 14, e2000995 https://doi.org/10.1371/journal.pbio.2000995.

Hobday, A.J., Doerr, V.A.J., Marshall, N.A., Cvitanovic, C., Lim-Camacho, L., 2018. Adapting to climate change: the role of organisational personalities in natural resource management. Reg. Environ. Change 18, 561–571. https://doi.org/10.1007/s10113-017-1227-0.

Hobley, D.E.J., Adams, J.M., Nudurupati, S.S., Hutton, E.W.H., Gasparini, N.M., Istanbulluoglu, E., Tucker, G.E., 2017. Creative computing with Landlab: an open-source toolkit for building, coupling, and exploring two-dimensional numerical models of Earth-surface dynamics. Earth Surface Dynamics 5, 21–46. https://doi.org/10.5194/esurf-5-21-2017.

Hoekstra, A.G., Lorenz, E., Falcone, J.-L., Chopard, B., 2007. Towards a complex automata framework for multi-scale modeling: formalism and the scale separation map. In: Shi, Y., van Albada, G.D., Dongarra, J., Sloot, P.M.A. (Eds.), Computational Science – ICCS 2007, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 922–930.

Hoda, R., Noble, J., Marshall, S., 2013. Self-Organizing Roles on Agile Software Development Teams. IEEE Transactions on Software Engineering 39, 422–444. https://doi.org/10.1109/TSE.2012.30.

Hoekstra, A., Bastien, Chopard, Peter, Coveney, 2014. Multiscale modelling and simulation: a position paper. Phil. Trans. Math. Phys. Eng. Sci. 372, 20130377. https://doi.org/10.1098/rsta.2013.0377.

Howard, P.H., Derek, S., 2016. The Wisdom of the Economic Crowd: Calibrating Integrated Assessment Models Using Consensus (2016 Annual Meeting, July 31-August 2, Boston, Massachusetts No. 235639). Agricultural and Applied Economics Association. https://doi.org/10.22004/ag.econ.235639.

Huang, Y., Yu, D., Gong, Y., Liu, C., 2013. Semi-supervised GMM and DNN acoustic model training with multi-system combination and confidence re-calibration. In: INTERSPEECH.

Hunt, R.J., 2017. Applied uncertainty. Groundwater 55, 771–772. https://doi.org/10.1111/gwat.12604.

Hutton, C., Wagener, T., Freer, J., Han, D., Duffy, C., Arheimer, B., 2016. Most computational hydrology is not reproducible, so is it really science? Water Resour. Res. 52, 7548–7555. https://doi.org/10.1002/2016WR019285.

Hutton, E.W.H., Piper, M.D., Tucker, G.E., 2020. The Basic Model Interface 2.0: a standard interface for coupling numerical models in the geosciences. Journal of Open Source Software 1, 2317. https://doi.org/10.21105/joss.02317.

Igamberdiev, M., Grossmann, G., Selway, M., Stumptner, M., 2018. An integrated multi-level modeling approach for industrial-scale data interoperability. Software Syst. Model 17, 269–294. https://doi.org/10.1007/s10270-016-0520-6.

Ingram, G.D., Cameron, I.T., Hangos, K.M., 2004. Classification and analysis of integrating frameworks in multiscale modelling. Chem. Eng. Sci. 59, 2171–2187. https://doi.org/10.1016/j.ces.2004.02.010.

Ingwersen, J., Högy, P., Wizemann, H.D., Warrach-Sagi, K., Streck, T., 2018. Coupling the land surface model Noah-MP with the generic crop growth model Gecros: model description, calibration and validation. Agric. For. Meteorol. 262, 322–339. https://doi.org/10.1016/j.agrformet.2018.06.023.

Iwanaga, T., Partington, D., Ticehurst, J., Croke, B.F.W., Jakeman, A.J., 2020. A socio-environmental model for exploring sustainable water management futures: participatory and collaborative modelling in the Lower Campaspe catchment. J. Hydrol.: Reg. Stud. 28, 100669. https://doi.org/10.1016/j.ejrh.2020.100669.

Jakeman, A.J., Letcher, R.A., Norton, J.P., 2006. Ten iterative steps in development and evaluation of environmental models. Environ. Model. Software 21, 602–614. https://doi.org/10.1016/j.envsoft.2006.01.004.

Jones, J.W., Antle, J.M., Basso, B., Boote, K.J., Conant, R.T., Foster, I., Godfray, H.C.J., Herrero, M., Howitt, R.E., Janssen, S., Keating, B.A., Munoz-Carpena, R., Porter, C. H., Rosenzweig, C., Wheeler, T.R., 2017. Toward a new generation of agricultural system data, models, and knowledge products: state of agricultural systems science. Agric. Syst. 155, 269–288. https://doi.org/10.1016/j.agsy.2016.09.021.

Joppa, L.N., McInerny, G., Harper, R., Salido, L., Takeda, K., O'Hara, K., Gavaghan, D., Emmott, S., 2013. Troubling trends in scientific software use. Science 340, 814–815. https://doi.org/10.1126/science.1231535.

Jordan, R., Gray, S., Zellner, M., Glynn, P.D., Voinov, A., Hedelin, B., Sterling, E.J., Leong, K., Olabisi, L.S., Hubacek, K., Bommel, P., BenDor, T.K., Jetter, A.J., Laursen, B., Singer, A., Giabbanelli, P.J., Kolagani, N., Carrera, L.B., Jenni, K., Prell, C., 2018. Twelve questions for the participatory modeling community. Earth's Future 6, 1046–1057. https://doi.org/10.1029/2018EF000841.

Kelly Letcher, R.A., Jakeman, A.J., Barreteau, O., Borsuk, M.E., ElSawah, S., Hamilton, S. H., Henriksen, H.J., Kuikka, S., Maier, H.R., Rizzoli, A.E., van Delden, H., Voinov, A. A., 2013. Selecting among five common modelling approaches for integrated environmental assessment and management. Environ. Model. Software 47, 159–181. https://doi.org/10.1016/j.envsoft.2013.05.005.

Kloprogge, P., van der Sluijs, J.P., Petersen, A.C., 2011. A method for the analysis of assumptions in model-based environmental assessments. Environmental Modelling & Software, Thematic issue on the assessment and evaluation of environmental models and software 26, 289–301. https://doi.org/10.1016/j.envsoft.2009.06.009.

Knapen, R., Janssen, S., Roosenschoon, O., Verweij, P., de Winter, W., Uiterwijk, M., Wien, J.-E., 2013. Evaluating OpenMI as a model integration platform across disciplines. Environmental Modelling & Software, Thematic Issue on the Future of Integrated Modeling Science and Technology 39, 274–282. https://doi.org/10.1016/j.envsoft.2012.06.011.

Knight, D., Durham, C.C., Locke, E.A., 2001. The relationship of team goals, incentives, and efficacy to strategic risk, tactical implementation, and performance. Acad. Manag. J. 44, 326–338. https://doi.org/10.2307/3069459.

Koo, H, Iwanaga, T, Croke, BFW, Jakeman, AJ, Yang, J, Wang, H-H, Sun, X, Lü, G, Li, X, Yue, T, Yuan, W, Liu, X, Chen, M., 2020. Position paper: Sensitivity analysis of spatially distributed environmental models- a pragmatic framework for the exploration of uncertainty sources. Environ. Model. Software 134, 104857. https://doi.org/10.1016/j.envsoft.2020.104857.

Kragt, M.E., Robson, B.J., Macleod, C.J.A., 2013. Modellers' roles in structuring integrative research projects. Environmental Modelling & Software, Thematic Issue on the Future of Integrated Modeling Science and Technology 39, 322–330. https://doi.org/10.1016/j.envsoft.2012.06.015.

Kretser, A., Murphy, D., Bertuzzi, S., Abraham, T., Allison, D.B., Boor, K.J., Dwyer, J., Grantham, A., Harris, L.J., Hollander, R., Jacobs-Young, C., Rovito, S., Vafiadis, D., Woteki, C., Wyndham, J., Yada, R., 2019. Scientific integrity principles and best practices: recommendations from a scientific integrity consortium. Sci. Eng. Ethics 25, 327–355. https://doi.org/10.1007/s11948-019-00094-3.

Kwakkel, J.H., 2017. The Exploratory Modeling Workbench: an open source toolkit for exploratory modeling, scenario discovery, and (multi-objective) robust decision making. Environ. Model. Software 96, 239–250. https://doi.org/10.1016/j.envsoft.2017.06.054.

Kwon, S., Solomon, G.E.A., Youtie, J., Porter, A.L., 2017. A measure of knowledge flow between specific fields: implications of interdisciplinarity for impact and funding. PloS One 12, e0185583. https://doi.org/10.1371/journal.pone.0185583.

Lahtinen, T.J., Guillaume, J.H.A., Hmlinen, R.P., 2017. Why pay attention to paths in the practice of environmental modelling? Environ. Model. Software 92, 74–81. https://doi.org/10.1016/j.envsoft.2017.02.019.

Laukkanen, E., Itkonen, J., Lassenius, C., 2017. Problems, causes and solutions when adopting continuous delivery—a systematic literature review. Inf. Software Technol. 82, 55–79. https://doi.org/10.1016/j.infsof.2016.10.001.

Lee, C., Lin, Y., Huan, H., Huang, W., Teng, H., 2015. The effects of task interdependence, team cooperation, and team conflict on job performance. SBP (Soc. Behav. Pers.): Int. J. 43, 529–536. https://doi.org/10.2224/sbp.2015.43.4.529.

Li, G., Bie, Z., Kou, Y., Jiang, J., Bettinelli, M., 2016. Reliability evaluation of integrated energy systems based on smart agent communication. Appl. Energy 167, 397–406. https://doi.org/10.1016/j.apenergy.2015.11.033.

Ling, Y., Mullins, J., Mahadevan, S., 2012. Calibration of Multi-Physics Computational Models Using Bayesian Networks.

Linkov, I., Burmistrov, D., 2003. Model uncertainty and choices made by modelers: lessons learned from the international atomic energy agency model intercomparisons†. Risk Anal. 23, 1297–1308. https://doi.org/10.1111/j.0272-4332.2003.00402.x.

Little, J.C., Hester, E.T., Elsawah, S., Filz, G.M., Sandu, A., Carey, C.C., Iwanaga, T., Jakeman, A.J., 2019. A tiered, system-of-systems modeling framework for resolving complex socio-environmental policy issues. Environ. Model. Software 112, 82–94. https://doi.org/10.1016/j.envsoft.2018.11.011.

López-Fidalgo, J., Tommasi, C., 2018. Optimal experimental design for model selection: a partial review. In: Gil, E., Gil, Eva, Gil, J., Gil, M.Á. (Eds.), The Mathematics of the Uncertain: A Tribute to Pedro Gil, Studies in Systems, Decision and Control. Springer International Publishing, Cham, pp. 253–263. https://doi.org/10.1007/978-3-319-73848-2_24.

Lorek, H., Sonnenschein, M., 1999. Modelling and simulation software to support individual-based ecological modelling. Ecol. Model. 115, 199–216. https://doi.org/10.1016/S0304-3800(98)00193-8.

MacLeod, M., Nagatsu, M., 2018. What does interdisciplinarity look like in practice: mapping interdisciplinarity and its limits in the environmental sciences. Stud. Hist. Philos. Sci. 67, 74–84. https://doi.org/10.1016/j.shpsa.2018.01.001.

Macy, M.W., Willer, R., 2002. From factors to actors: computational sociology and agent-based modeling. Annu. Rev. Sociol. 28, 143–166. https://doi.org/10.1146/annurev.soc.28.110601.141117.

Madni, A.M., Sievers, M., 2014. Systems integration: key perspectives, experiences, and challenges. Syst. Eng. 17, 37–51. https://doi.org/10.1002/sys.21249.

Maier, H.R., Guillaume, J.H.A., van Delden, H., Riddell, G.A., Haasnoot, M., Kwakkel, J. H., 2016. An uncertain future, deep uncertainty, scenarios, robustness and adaptation: how do they fit together? Environ. Model. Software 81, 154–164. https://doi.org/10.1016/j.envsoft.2016.03.014.

Malard, J.J., Inam, A., Hassanzadeh, E., Adamowski, J., Tuy, H.A., Melgar-Quiñonez, H., 2017. Development of a software tool for rapid, reproducible, and stakeholder-friendly dynamic coupling of system dynamics and physically-based models. Environ. Model. Software 96, 410–420. https://doi.org/10.1016/J.ENVSOFT.2017.06.053.

Martin, D.M., Powell, S.J., Webb, J.A., Nichols, S.J., Poff, N.L., 2017. An objective method to prioritize socio-environmental water management tradeoffs using multi-criteria decision analysis. River Res. Appl. 33, 586–596. https://doi.org/10.1002/rra.3103.

Matott, L.S., Babendreier, J.E., Purucker, S.T., 2009. Evaluating uncertainty in integrated environmental models: a review of concepts and tools. Water Resour. Res. 45 https://doi.org/10.1029/2008WR007301.

Meadows, D.H., 2008. Thinking in Systems: A Primer. Chelsea Green Publishing.

Meirmans, S., Butlin, R.K., Charmantier, A., Engelstädter, J., Groot, A.T., King, K.C., Kokko, H., Reid, J.M., Neiman, M., 2019. Science policies: how should science funding be allocated? An evolutionary biologists' perspective. J. Evol. Biol. 32, 754–768. https://doi.org/10.1111/jeb.13497.

Midgley, G., Pinzón, L.A., 2011. Boundary critique and its implications for conflict prevention. J. Oper. Res. Soc. 62, 1543–1554. https://doi.org/10.1057/jors.2010.76.

Miyasaka, T., Le, Q.B., Okuro, T., Zhao, X., Takeuchi, K., 2017. Agent-based modeling of complex social–ecological feedback loops to assess multi-dimensional trade-offs in dryland ecosystem services. Landsc. Ecol. 32, 707–727. https://doi.org/10.1007/s10980-017-0495-x.

Moore, R.V., Tindall, C.I., 2005. An overview of the open modelling interface and environment (the OpenMI). Environmental Science & Policy, Research & Technology Integration in Support of the European Union Water Framework Directive 8, 279–286. https://doi.org/10.1016/j.envsci.2005.03.009.

Moriasi, D.N., Arnold, J.G., Liew, M.W.V., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE 50, 885–900. https://doi.org/10.13031/2013.23153.

Muñoz, E., Rivera, D., Vergara, F., Tume, P., Arumí, J.L., 2014. Identifiability analysis: towards constrained equifinality and reduced uncertainty in a conceptual model. Hydrol. Sci. J. 59, 1690–1703. https://doi.org/10.1080/02626667.2014.892205.

National Research Council, 2013. Sustainability for the nation: resource connections and governance linkages. https://doi.org/10.17226/13471.

Neumann, P., Düben, P., Adamidis, P., Bauer, P., Brück, Kornblueh, L., Klocke, D., Stevens, B., Wedi, N., Biercamp, J., 2019. Assessing the scales in numerical weather and climate predictions: will exascale be the rescue? Phil. Trans. Math. Phys. Eng. Sci. 377, 20180148. https://doi.org/10.1098/rsta.2018.0148.

Nielsen, C.B., Larsen, P.G., Fitzgerald, J., Woodcock, J., Peleska, J., 2015. Systems of systems engineering: basic concepts, model-based techniques, and research directions. ACM Comput. Surv. 48, 18. https://doi.org/10.1145/2794381, 1–18:41.

Norström, A.V., Cvitanovic, C., Löf, M.F., West, S., Wyborn, C., Balvanera, P., Bednarek, A.T., Bennett, E.M., Biggs, R., de Bremond, A., Campbell, B.M., Canadell, J.G., Carpenter, S.R., Folke, C., Fulton, E.A., Gaffney, O., Gelcich, S., Jouffray, J.-B., Leach, M., Le Tissier, M., Martín-López, B., Louder, E., Loutre, M.-F., Meadow, A.M., Nagendra, H., Payne, D., Peterson, G.D., Reyers, B., Scholes, R., Speranza, C.I., Spierenburg, M., Stafford-Smith, M., Tengö, M., van der Hel, S., van Putten, I., Österblom, H., 2020. Principles for knowledge co-production in sustainability research. Nature Sustainability 3, 182–190. https://doi.org/10.1038/s41893-019-0448-2.

Norton, J., 2015. An introduction to sensitivity assessment of simulation models. Environ. Model. Software 69, 166–174. https://doi.org/10.1016/j.envsoft.2015.03.020.

Nossent, J., Bauwens, W., 2012. Multi-variable sensitivity and identifiability analysis for a complex environmental model in view of integrated water quantity and water quality modeling. Water Sci. Technol. 65, 539–549. https://doi.org/10.2166/wst.2012.884.

Oreskes, N., 2003. The role of quantitative models in science. In: Canham, C.D., Cole, J. J., Lauenroth, W.K. (Eds.), Models in Ecosystem Science. Princeton University Press, pp. 13–31.

Ostrom, E., 2007. A diagnostic approach for going beyond panaceas. Proc. Natl. Acad. Sci. U.S.A. 104, 15181. https://doi.org/10.1073/pnas.0702288104.

Oxley, T., ApSimon, H.M., 2007. Space, time and nesting integrated assessment models. Environ. Model. Software 22, 1732–1749. https://doi.org/10.1016/j.envsoft.2007.02.002.

O'Connell, P.E., Todini, E., 1996. Modelling of rainfall, flow and mass transport in hydrological systems: an overview. J. Hydrol. 175, 3–16. https://doi.org/10.1016/S0022-1694(96)80003-5.

Pacheco-Romero, M., Alcaraz-Segura, D., Vallejos, M., Cabello, J., 2020. An expert-based reference list of variables for characterizing and monitoring social-ecological systems. Ecol. Soc. 25 https://doi.org/10.5751/ES-11676-250301.

Pahl-Wostl, C., 2007. Transitions towards adaptive management of water facing climate and global change. Water Resour. Manag. 21, 49–62. https://doi.org/10.1007/s11269-006-9040-4.

Partington, D., Knowling, M.J., Simmons, C.T., Cook, P.G., Xie, Y., Iwanaga, T., Bouchez, C., 2020. Worth of hydraulic and water chemistry observation data in terms of the reliability of surface water-groundwater exchange flux predictions under varied flow conditions. J. Hydrol. 590, 125441. https://doi.org/10.1016/j.jhydrol.2020.125441.

Patil, P., Peng, R.D., Leek, J.T., 2016. A Statistical Definition for Reproducibility and Replicability. bioRxiv. https://doi.org/10.1101/066803, 066803.

Peng, R.D., 2011. Reproducible research in computational science. Science 334, 1226–1227. https://doi.org/10.1126/science.1213847.

Phillips, S.D., Estler, W.T., Doiron, T., Eberhardt, K.R., Levenson, M.S., 2001. A careful consideration of the calibration concept. Journal of research of the National Institute of Standards and Technology 106. https://doi.org/10.6028/jres.106.014.

Pianosi, F., Wagener, T., 2015. A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. Environ. Model. Software 67, 1–11. https://doi.org/10.1016/j.envsoft.2015.01.004.

Pianosi, F., Sarrazin, F., Wagener, T., 2015. A matlab toolbox for global sensitivity analysis. Environ. Model. Software 70, 80–85. https://doi.org/10.1016/j.envsoft.2015.04.009.

Pianosi, F., Beven, K., Freer, J., Hall, J.W., Rougier, J., Stephenson, D.B., Wagener, T., 2016. Sensitivity analysis of environmental models: a systematic review with practical workflow. Environ. Model. Software 79, 214–232. https://doi.org/10.1016/j.envsoft.2016.02.008.

Pietzsch, B., Fiedler, S., Mertens, K.G., Richter, M., Scherer, C., Widyastuti, K., Wimmler, M.-C., Zakharova, L., Berger, U., 2020. Metamodels for evaluating, calibrating and applying agent-based models: a review. JASSS 23, 9. https://doi.org/10.18564/jasss.4274.

Porporato, A., Rodriguez-Iturbe, I., 2002. Ecohydrology-a challenging multidisciplinary research perspective/Ecohydrologie: une perspective stimulante de recherche multidisciplinaire. Hydrol. Sci. J. 47, 811–821. https://doi.org/10.1080/02626660209492985.

Potter, S., Doran, B., Mathews, D., 2016. Modelling collective Yawuru values along the foreshore of Roebuck Bay, Western Australia using fuzzy logic. Appl. Geogr. 77, 8–19. https://doi.org/10.1016/j.apgeog.2016.09.016.

Pulver, S., Ulibarri, N., Sobocinski, K., Alexander, S., Johnson, M., McCord, P., Dell'Angelo, J., 2018. Frontiers in socio-environmental research: components, connections, scale, and context. Ecol. Soc. 23 https://doi.org/10.5751/ES-10280-230323.

Rai, A., Lang, S.S., Welker, R.B., 2002. Assessing the validity of IS success models: an empirical test and theoretical analysis. Inf. Syst. Res. 13, 50–69. https://doi.org/10.1287/isre.13.1.50.96.

Raick, C., Soetaert, K., Grégoire, M., 2006. Model complexity and performance: how far can we simplify? Prog. Oceanogr. 70, 27–57. https://doi.org/10.1016/j.pocean.2006.03.001.

Railsback, S.F., Grimm, V., 2019. Agent-based and individual-based modeling: a practical introduction, 2nd edition. ed. Princeton University Press, Princeton, NJ.

Ramaswami, A., Milford, J.B., Small, M.J., 2005. Integrated Environmental Modeling: Pollutant Transport, Fate, and Risk in the Environment. John Wiley & Sons Inc., Hoboken, NJ.

Razavi, S., Gupta, H.V., 2015. What do we mean by sensitivity analysis? The need for comprehensive characterization of "global" sensitivity in Earth and Environmental systems models. Water Resour. Res. 51, 3070–3092. https://doi.org/10.1002/2014WR016527.

Razavi, S., Tolson, B.A., 2013. An efficient framework for hydrologic model calibration on long data periods. Water Resour. Res. 49, 8418–8431. https://doi.org/10.1002/2012WR013442.

Razavi, S., Tolson, B.A., Matott, L.S., Thomson, N.R., MacLean, A., Seglenieks, F.R., 2010. Reducing the computational cost of automatic calibration through model preemption. Water Resour. Res. 46 https://doi.org/10.1029/2009WR008957.

Razavi, S., Tolson, B.A., Burn, D.H., 2012. Review of surrogate modeling in water resources. Water Resour. Res. 48 https://doi.org/10.1029/2011WR011527.

Razavi, S., Sheikholeslami, R., Gupta, H.V., Haghnegahdar, A., 2019. VARS-TOOL: a toolbox for comprehensive, efficient, and robust sensitivity and uncertainty analysis. Environ. Model. Software 112, 95–107. https://doi.org/10.1016/j.envsoft.2018.10.005.

Razavi, S., Gober, P., Maier, H.R., Brouwer, R., Wheater, H., 2020. Anthropocene flooding: challenges for science and society. Hydrol. Process. 34, 1996–2000. https://doi.org/10.1002/hyp.13723.

Refsgaard, J.C., van der Sluijs, J.P., Højberg, A.L., Vanrolleghem, P.A., 2007. Uncertainty in the environmental modelling process – a framework and guidance. Environ. Model. Software 22, 1543–1556. https://doi.org/10.1016/j.envsoft.2007.02.004.
Reichert, P., 2020. Towards a comprehensive uncertainty assessment in environmental research and decision support. Water Sci. Technol. https://doi.org/10.2166/wst.2020.032.
Reichert, P., Borsuk, M.E., 2005. Does high forecast uncertainty preclude effective decision support? Environmental Modelling & Software, Methods of Uncertainty Treatment in Environmental Models 20, 991–1001. https://doi.org/10.1016/j.envsoft.2004.10.005.
Renner, S.A., 2001. A "Community of Interest" Approach to Data Interoperability.
Roberts, J.J., Marotta Cassula, A., Silveira, J.L., da Costa Bortoni, E., Mendiburu, A.Z., 2018. Robust multi-objective optimization of a renewable based hybrid power system. Appl. Energy 223, 52–68. https://doi.org/10.1016/j.apenergy.2018.04.032.
Rounsevell, M.D.A., Metzger, M.J., 2010. Developing qualitative scenario storylines for environmental change assessment. WIREs Climate Change 1, 606–619. https://doi.org/10.1002/wcc.63.
Rubin, R.B., Rubin, A.M., Graham, E., Perse, E.M., Seibold, D., 2010. Communication Research Measures II: A Sourcebook. Routledge.
Salado, A., 2015. Abandonment: a natural consequence of autonomy and belonging in systems-of-systems. In: 2015 10th System of Systems Engineering Conference (SoSE). Presented at the 2015 10th System of Systems Engineering Conference (SoSE), pp. 352–357. https://doi.org/10.1109/SYSOSE.2015.7151936.
Saltelli, A., 2018. Why science's crisis should not become a political battling ground. Futures 104, 85–90. https://doi.org/10.1016/j.futures.2018.07.006.
Saltelli, A., Annoni, P., 2010. How to avoid a perfunctory sensitivity analysis. Environ. Model. Software 25, 1508–1517. https://doi.org/10.1016/j.envsoft.2010.04.012.
Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., Wu, Q., 2019. Why so many published sensitivity analyses are false: a systematic review of sensitivity analysis practices. Environ. Model. Software 114, 29–39.
Sandström, U., Van den Besselaar, P., 2018. Funding, evaluation, and the performance of national research systems. Journal of Informetrics 12, 365–384. https://doi.org/10.1016/j.joi.2018.01.007.
Sarosa, S., Tatnall, A., 2015. Failure to launch: scope creep and other causes of failure from an actor-network theory perspective. Int. J. Actor-Network Theory Technol. Innovation (IJANTTI) 7, 1–13. https://doi.org/10.4018/ijantti.2015100101.
Schlüter, M., Haider, L., Lade, S., Lindkvist, E., Martin, R., Orach, K., Wijermans, N., Folke, C., 2019a. Capturing emergent phenomena in social-ecological systems: an analytical framework. Ecol. Soc. 24 https://doi.org/10.5751/ES-11012-240311.
Schlüter, M., Müller, B., Frank, K., 2019b. The potential of models and modeling for social-ecological systems research: the reference frame ModSES. Ecol. Soc. 24 https://doi.org/10.5751/ES-10716-240131.
Schneider, B., Ehrhart, M.G., Macey, W.H., 2013. Organizational climate and culture. Annu. Rev. Psychol. 64, 361–388. https://doi.org/10.1146/annurev-psych-113011-143809.
Schnell, S., 2018. "Reproducible" research in mathematical sciences requires changes in our peer review culture and modernization of our current publication approach. Bull. Math. Biol. 80, 3095–3105. https://doi.org/10.1007/s11538-018-0500-9.
Scholes, R.J., Reyers, B., Biggs, R., Spierenburg, M.J., Duriappah, A., 2013. Multi-scale and cross-scale assessments of social–ecological systems and their ecosystem services. Current Opinion in Environmental Sustainability 5, 16–25. https://doi.org/10.1016/j.cosust.2013.01.004.
Schouten, R., Deits, R., 2020. Flatten.jl. Zenodo. http://doi.org/10.5281/zenodo.3599802.
Schweiger, G., Nilsson, H., Schoeggl, J., Birk, W., Posch, A., 2020. Modeling and simulation of large-scale systems: a systematic comparison of modeling paradigms. Appl. Math. Comput. 365, 124713. https://doi.org/10.1016/j.amc.2019.124713.
Shin, M.-J., Guillaume, J.H.A., Croke, B.F.W., Jakeman, A.J., 2013. Addressing ten questions about conceptual rainfall–runoff models with global sensitivity analyses in R. J. Hydrol. 503, 135–152. https://doi.org/10.1016/j.jhydrol.2013.08.047.
Singh, S.K., Bárdossy, A., 2012. Calibration of hydrological models on hydrologically unusual events. Adv. Water Resour. 38, 81–91. https://doi.org/10.1016/j.advwatres.2011.12.006.
Sivapalan, M., Savenije, H.H.G., Blöschl, G., 2012. Socio-hydrology: a new science of people and water. Hydrol. Process. 26, 1270–1276. https://doi.org/10.1002/hyp.8426.
Sletholt, M.T., Hannay, J.E., Pfahl, D., Langtangen, H.P., 2012. What do we know about scientific software development's agile practices? https://doi.org/10.1109/MCSE.2011.113, 24-36.
Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., Xu, C., 2015. Global sensitivity analysis in hydrological modeling: review of concepts, methods, theoretical framework, and applications. J. Hydrol. 523, 739–757. https://doi.org/10.1016/j.jhydrol.2015.02.013.
Spitzberg, B.H., Cupach, W.R., 1989. Handbook of Interpersonal Competence Research. Springer-Verlag.
Steinhardt, U., Volk, M., 2001. Scales and spatio-temporal dimensions in landscape research. In: Krönert, R., Steinhardt, U., Volk, M. (Eds.), Landscape Balance and Landscape Assessment. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 137–162. https://doi.org/10.1007/978-3-662-04532-9_6.
Sterling, E.J., Zellner, M., Jenni, K.E., Leong, K., Glynn, P.D., BenDor, T.K., Bommel, P., Hubacek, K., Jetter, A.J., Jordan, R., Olabisi, L.S., Paolisso, M., Gray, S., 2019. Try, try again: lessons learned from success and failure in participatory modeling. Elem Sci Anth 7, 9. https://doi.org/10.1525/elementa.347.
Tate, L.E., 2013. Growth-management implementation in metropolitan vancouver: lessons from actor-network theory. Environ. Plann. Plann. Des. 40, 783–800. https://doi.org/10.1068/b37170.
Tebaldi, C., Knutti, R., 2007. The use of the multi-model ensemble in probabilistic climate projections. Philos Trans A Math Phys Eng Sci 365, 2053–2075. https://doi.org/10.1098/rsta.2007.2076.
Thomas, J., McDonagh, D., 2013. Shared language:Towards more effective communication. Australas. Med. J. 6, 46–54. https://doi.org/10.4066/AMJ.2013.1596.
Thompson, J.L., 2009. Building collective communication competence in interdisciplinary research teams. J. Appl. Commun. Res. 37, 278–297. https://doi.org/10.1080/00909880903025911.
Tjosvold, D., Hui, C., Ding, D.Z., Hu, J., 2003. Conflict values and team relationships: conflict's contribution to team effectiveness and citizenship in China. J. Organ. Behav. 24, 69–88. https://doi.org/10.1002/job.180.
Tranquillo, J., 2019. An Introduction to Complex Systems: Making Sense of a Changing World. Springer Nature.
Tscheikner-Gratl, F., Bellos, V., Schellart, A., Moreno-Rodenas, A., Muthusamy, M., Langeveld, J., Clemens, F., Benedetti, L., Rico-Ramirez, M.A., de Carvalho, R.F., Breuer, L., Shucksmith, J., Heuvelink, G.B.M., Tait, S., 2019. Recent insights on uncertainties present in integrated catchment water quality modelling. Water Res. 150, 368–379. https://doi.org/10.1016/j.watres.2018.11.079.
Uusitalo, L., Lehikoinen, A., Helle, I., Myrberg, K., 2015. An overview of methods to evaluate uncertainty of deterministic models in decision support. Environ. Model. Software 63, 24–31. https://doi.org/10.1016/j.envsoft.2014.09.017.
Vale, T., Crnkovic, I., De Almeida, E.S., Silveira Neto, P.A.D.M., Cavalcanti, Y.C., Meira, S.R.D.L., 2016. Twenty-eight years of component-based software engineering. J. Syst. Software 111, 128–148. https://doi.org/10.1016/j.jss.2015.09.019.
van Delden, H., van Vliet, J., Rutledge, D.T., Kirkby, M.J., 2011. Comparison of scale and scaling issues in integrated land-use models for policy support. Agriculture, Ecosystems & Environment, Scaling methods in integrated assessment of agricultural systems 142, 18–28. https://doi.org/10.1016/j.agee.2011.03.005.
van der Sluijs, J.P., 2007. Uncertainty and precaution in environmental management: insights from the UPEM conference. Environmental Modelling & Software, The Implications of Complexity for Integrated Resources 22, 590–598. https://doi.org/10.1016/j.envsoft.2005.12.020.
van der Vaart, E., Johnston, A.S.A., Sibly, R.M., 2016. Predicting how many animals will be where: how to build, calibrate and evaluate individual-based models. Ecol. Model. 326, 113–123. https://doi.org/10.1016/j.ecolmodel.2015.08.012.
van Ittersum, M.K., Ewert, F., Heckelei, T., Wery, J., Alkan Olsson, J., Andersen, E., Bezlepkina, I., Brouwer, F., Donatelli, M., Flichman, G., Olsson, L., Rizzoli, A.E., van der Wal, T., Wien, J.E., Wolf, J., 2008. Integrated assessment of agricultural systems – a component-based framework for the European Union (SEAMLESS). Agric. Syst. 96, 150–165. https://doi.org/10.1016/j.agsy.2007.07.009.
Vanlier, J., Tiemann, C.A., Hilbers, P.A., van Riel, N.A., 2014. Optimal experiment design for model selection in biochemical networks. BMC Syst. Biol. 8, 20. https://doi.org/10.1186/1752-0509-8-20.
Verweij, P.J.F.M., Knapen, M.J.R., de Winter, W.P., Wien, J.J.F., te Roller, J.A., Sieber, S., Jansen, J.M.L., 2010. An IT perspective on integrated environmental modelling: the SIAT case. Ecological Modelling, Model-based Systems to Support Impact Assessment - Methods, Tools and Applications 221, 2167–2176. https://doi.org/10.1016/j.ecolmodel.2010.01.006.
Voinov, A., Bousquet, F., 2010. Modelling with stakeholders. Environmental Modelling & Software, Thematic Issue - Modelling with Stakeholders 25, 1268–1281. https://doi.org/10.1016/j.envsoft.2010.03.007.
Voinov, A., Shugart, H.H., 2013. "Integronsters", integral and integrated modeling. Environ. Model. Software 39, 149–158. https://doi.org/10.1016/j.envsoft.2012.05.014.
Voinov, A., Kolagani, N., McCall, M.K., Glynn, P.D., Kragt, M.E., Ostermann, F.O., Pierce, S.A., Ramu, P., 2016. Modelling with stakeholders – next generation. Environ. Model. Software 77, 196–220. https://doi.org/10.1016/j.envsoft.2015.11.016.
Voinov, A., Jenni, K., Gray, S., Kolagani, N., Glynn, P.D., Bommel, P., Prell, C., Zellner, M., Paolisso, M., Jordan, R., Sterling, E., Schmitt Olabisi, L., Giabbanelli, P. J., Sun, Z., Le Page, C., Elsawah, S., BenDor, T.K., Hubacek, K., Laursen, B.K., Jetter, A., Basco-Carrera, L., Singer, A., Young, L., Brunacini, J., Smajgl, A., 2018. Tools and methods in participatory modeling: selecting the right tool for the job. Environ. Model. Software 109, 232–255. https://doi.org/10.1016/j.envsoft.2018.08.028.
Wagener, T., McIntyre, N., Lees, M.J., Wheater, H.S., Gupta, H.V., 2003. Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. Hydrol. Process. 17, 455–476. https://doi.org/10.1002/hyp.1135.
Wagener, T., Pianosi, F., 2019. What has Global Sensitivity Analysis ever done for us? A systematic review to support scientific advancement and to inform policy-making in earth system modelling. Earth Sci . Rev. 194, 1–18. https://doi.org/10.1016/j.earscirev.2019.04.006.
Walker, W.E., Harremoës, P., Rotmans, J., Van der Sluijs, J.P., Van Asselt, M.B.A., Janssen, P., Krayer von Krauss, M.P., 2003. Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. Integrated Assessment 4, 5–17. https://doi.org/10.1076/iaij.4.1.5.16466.
Wang, K., Davies, E.G.R., Liu, J., 2019. Integrated water resources management and modeling: a case study of Bow river basin, Canada. J. Clean. Prod. 240, 118242. https://doi.org/10.1016/j.jclepro.2019.118242.
Warren, K., 2014. Agile SD: fast, effective, reliable. In: Proceedings of the 32nd International Conference of the System Dynamics Society. Presented at the Conference of the System Dynamics Society. Systems Dynamics Society, Delft, Netherlands.

Wessells, A.T., 2007. A review of: "reassembling the social: an introduction to actor-network-theory". Int. Publ. Manag. J. 10, 351–356. https://doi.org/10.1080/10967490701515606.

Whelan, G., Kim, K., Pelton, M.A., Castleton, K.J., Laniak, G.F., Wolfe, K., Parmar, R., Babendreier, J., Galvin, M., 2014. Design of a component-based integrated environmental modeling framework. Environ. Model. Software 55, 1–24. https://doi.org/10.1016/j.envsoft.2014.01.016.

White, D.D., Lawless, K.L., Vivoni, E.R., Mascaro, G., Pahle, R., Kumar, I., Coli, P., Castillo, R.M., Moreda, F., Asfora, M., 2019. Co-producing interdisciplinary knowledge and action for sustainable water governance: lessons from the development of a water resources decision support system in pernambuco, Brazil. Global Challenges 3, 1800012. https://doi.org/10.1002/gch2.201800012.

Wiegand, T., Jeltsch, F., Hanski, I., Grimm, V., 2003. Using pattern-oriented modeling for revealing hidden information: a key for reconciling ecological theory and application. Oikos 100, 209–222. https://doi.org/10.1034/j.1600-0706.2003.12027.x.

Wiegand, T., Revilla, E., Knauer, F., 2004. Dealing with uncertainty in spatially explicit population models. Biodivers. Conserv. 13, 53–78. https://doi.org/10.1023/B:BIOC.0000004313.86836.ab.

Willcock, S., Hooftman, D.A.P., Blanchard, R., Dawson, T.P., Hickler, T., Lindeskog, M., Martinez-Lopez, J., Reyers, B., Watts, S.M., Eigenbrod, F., Bullock, J.M., 2020. Ensembles of ecosystem service models can improve accuracy and indicate uncertainty. Sci. Total Environ. 141006. https://doi.org/10.1016/j.scitotenv.2020.141006.

Wilson, C., Kriegler, E., van Vuuren, D.P., Guivarch, C., Frame, D., Krey, V., Osborn, T.J., Schwanitz, V.J., Thompson, E.L., 2017. Evaluating process-based integrated assessment models of climate change mitigation [WWW Document]. URL. http://pure.iiasa.ac.at/id/eprint/14502/. accessed 12.16.19.

Wirtz, D., Nowak, W., 2017. The rocky road to extended simulation frameworks covering uncertainty, inversion, optimization and control. Environ. Model. Software 93, 180–192. https://doi.org/10.1016/j.envsoft.2016.10.003.

Wu, R., Yang, L., Chen, C., Ahmad, S., Dascalu, S.M., Harris Jr., F.C., 2019. MELPF version 1: modeling error learning based post-processor framework for hydrologic models accuracy improvement. Geosci. Model Dev. (GMD) 12, 4115–4131. https://doi.org/10.5194/gmd-12-4115-2019.

Xiong, F., Xiong, Y., Greene, S., Chen, W., Yang, S., 2010. A new sparse grid based method for uncertainty propagation. In: Presented at the ASME 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. American Society of Mechanical Engineers Digital Collection, pp. 1205–1215. https://doi.org/10.1115/DETC2009-87430.

Yearworth, M., White, L., 2018. Spontaneous emergence of Community OR: self-initiating, self-organising problem structuring mediated by social media. European Journal of Operational Research, Community Operational Research: Innovations, internationalization and agenda-setting applications 268, 809–824. https://doi.org/10.1016/j.ejor.2018.01.024.

Yoo, S., Harman, M., 2012. Regression testing minimization, selection and prioritization: a survey. Softw. Test. Verif. Reliab. 22, 67–120. https://doi.org/10.1002/stvr.430.

Zare, F., Elsawah, S., Iwanaga, T., Jakeman, A.J., Pierce, S.A., 2017. Integrated water assessment and modelling: a bibliometric analysis of trends in the water resource sector. J. Hydrol. 552, 765–778. https://doi.org/10.1016/j.jhydrol.2017.07.031.

Zhang, C., Lin, H., Chen, M., Li, R., Zeng, Z., 2014a. Scale compatibility analysis in geographic process research: a case study of a meteorological simulation in Hong Kong. Appl. Geogr. 52, 135–143. https://doi.org/10.1016/j.apgeog.2014.05.005.

Zhang, C., Lin, H., Chen, M., Yang, L., 2014b. Scale matching of multiscale digital elevation model (DEM) data and the Weather Research and Forecasting (WRF) model: a case study of meteorological simulation in Hong Kong. Arab J Geosci 7, 2215–2223. https://doi.org/10.1007/s12517-014-1273-6.

Zimmermann, H.-J., 2000. An application-oriented view of modeling uncertainty. Eur. J. Oper. Res. 122, 190–198. https://doi.org/10.1016/S0377-2217(99)00228-3.

Chapter 7: Reflexive lessons for effective interdisciplinarity

This chapter explores the interplay between teams and scale within the context of SoS modelling for SES and its subsequent influence on (model) uncertainty and complexity. The exploration is conducted through reflexive analysis of the modelling processes within two, unrelated, case studies to draw out lessons across five fundamental themes that are especially applicable within the interdisciplinary system-of-systems modelling context. As a publication, it serves as a companion to Chapter 6. The chapter was submitted to Elementa: Science of the Anthropocene and has been accepted for publication.

Iwanaga, T.*, Wang, H-H., Koralewski, T. E., Jakeman A. J., Little, J. C., 2021. Towards a complete interdisciplinary treatment of scale: reflexive lessons from socio-environmental systems modeling. Elementa: Science of the Anthropocene 9(1), https://doi.org/10.1525/elementa.2020.00182

Iwanaga, T, et al. 2021. Toward a complete interdisciplinary treatment of scale: Reflexive lessons from socioenvironmental systems modeling. *Elem Sci Anth*, 9: 1. DOI: https://doi.org/10.1525/elementa.2020.00182

PRACTICE BRIDGE

Toward a complete interdisciplinary treatment of scale: Reflexive lessons from socioenvironmental systems modeling

Takuya Iwanaga[1,*], Hsiao-Hsuan Wang[2], Tomasz E. Koralewski[2], William E. Grant[2], Anthony J. Jakeman[1], and John C. Little[3]

The pathways taken throughout any model-based process are undoubtedly influenced by the modeling team involved and the decision choices they make. For interconnected socioenvironmental systems (SES), such teams are increasingly interdisciplinary to enable a more expansive and holistic treatment that captures the purpose, the relevant disciplines and sectors, and other contextual settings. In practice, such interdisciplinarity increases the scope of what is considered, thereby increasing choices around model complexity and their effects on uncertainty. Nonetheless, the consideration of scale issues is one critical lens through which to view and question decision choices in the modeling cycle. But separation between team members, both geographically and by discipline, can make the scales involved more arduous to conceptualize, discuss, and treat. In this article, the practices, decisions, and workflow that influence the consideration of scale in SESs modeling are explored through reflexive accounts of two case studies. Through this process and an appreciation of past literature, we draw out several lessons under the following themes: (1) the fostering of collaborative learning and reflection, (2) documenting and justifying the rationale for modeling scale choices, some of which can be equally plausible (a perfect model is not possible), (3) acknowledging that causality is defined subjectively, (4) embracing change and reflection throughout the iterative modeling cycle, and (5) regularly testing the model integration to draw out issues that would otherwise be unnoticeable.

Keywords: Reflexive analysis, Integrated assessment and modeling, System-of-Systems, Socioenvironmental modeling, Interdisciplinary teams, Uncertainty

1. Introduction

Consideration of scale is a common activity in all system-of-systems (SoS) modeling approaches involving the integration of multiple models when representing any complex socioenvironmental system (SES) of interest. Unfortunately, such consideration is all too often conducted tacitly, or at best minimally, and recently has been considered a grand challenge in SES modeling (Elsawah et al., 2020). Scale underlies many modeling concerns including how to address model complexity, conceptual mismatches, and uncertainty. In short, explicit consideration of scale issues provides a valuable, and indeed critical, lens to view the decisions made in any SES modeling activity.

This article follows an earlier publication (Iwanaga et al., 2021b) in which the current practices, issues, and challenges with respect to scale were explored through a sociotechnical lens. Scale can thus be characterized as an expansive term relating not just to the properties of the system under investigation but also the interplay between the social and technical dimensions. These influence what is considered, what is not, and what is eventually included in the modeling. A crucial aspect is the influence of the people involved and the subsequent technical processes and decisions that produce a model for a given purpose. These underlying influences, including scale decisions taken, often remain implicit and are not explicitly discussed. But for reasons of saliency, legitimacy, and transparency, they are best appreciated and considered by team members in as complete a sense as possible, albeit taking resources and time available into account.

Interdisciplinarity is now recognized as a crucial necessity in understanding and dealing with the complexity of socioenvironmental interactions (Hall et al., 2012; Saltelli

[1] Institute for Water Futures, Fenner School of Environment & Society, The Australian National University, Canberra, Australia

[2] Ecological Systems Laboratory, Department of Ecology and Conservation Biology, Texas A&M University, College Station, TX, USA

[3] Department of Civil & Environmental Engineering, Virginia Tech, Blacksburg, VA, USA

* Corresponding author:
Email: iwanaga.takuya@anu.edu.au

and Funtowicz, 2017; MacLeod and Nagatsu, 2018; Sterling et al., 2019). Challenges to a successful modeling process and set of outcomes necessitate effective collaboration, teamwork, and cross-disciplinary communication and discussion of a high order among team members (Nancarrow et al., 2013; Hall et al., 2018). Technological solutions cannot resolve mismatches in understanding among people, although they can facilitate and prompt discussions. Thus, there is a need to examine the practices and decision choices we make in any modeling activity, but especially so for one as complex as in SES. Acknowledgment of the human aspects that influence the choice and treatment of scales in modeling, and their implications, is therefore crucial in moving beyond the status quo. A key activity then is identifying the practices and approaches that facilitate and promote effective interdisciplinary cohesion among and within modeling teams.

The treatment of scale in the modeling process is an essential and valuable activity for focusing the attention of modelers on many of their key decisions. But that treatment can be affected by the level of cohesion and reflexivity within the collaborative process, which in turn may have a substantial influence on modeling outcomes, especially with greater interdisciplinarity of the modeling issue being addressed (Jones et al., 2011; Lahtinen et al., 2017). The level of inclusivity in communication that leads to interdisciplinary considerations and participation of all stakeholders, where stakeholders may include modelers themselves (following the definition in Freeman, 2010), are then also issues of scale to be explored. Recent publications espouse similar positions in recognizing the role that researchers play in shaping the scientific and policy discourse (Crouzat et al., 2018; Connolly, 2020; Walsh et al., 2020). We, as researchers, are perhaps coming to the full realization that "the technique is never neutral" (Saltelli et al., 2020) and that we cannot divorce ourselves from the influence we have on processes we take part in (Glynn et al., 2017; Cockerill et al., 2019).

1.1. The reflexive approach

A (more) reflexive approach to interdisciplinarity has been suggested over the years to aid in bridging the gap in understanding between the research that is conducted and the interdisciplinary processes that produce research outcomes (Finlay, 2002; Preston et al., 2015; Lahtinen et al., 2017). As with many terms that cross disciplinary bounds, "reflexivity" has several meanings with different practitioners holding differing views on its definition. The term "reflexive" is adopted here to convey a more transformative intent; the goal is to improve future practice through reflection on the seemingly self-evident choices and influences in the activities undertaken, the underlying assumptions, the role one played in the decisions, and the broader context in which these choices occur (Preston et al., 2015; May and Perry, 2017; Bolton and Delderfield, 2018). Reflexive evaluation is therefore one approach to considering the implication of scale choices, a practice which can aid in identifying the lessons learnt that are of benefit to future research (Krueger et al., 2016; Montana et al., 2020).

In this article, we draw five lessons through the reflexive accounts of the treatment of scale across two interdisciplinary socioenvironmental modeling case studies, also drawing upon diverse literature, where appropriate, to corroborate our experiences. As noted by others, the reflexive approach is highly situation-specific, such that there is no "one" approach to reflexivity (e.g., Montana et al., 2020). The reflexive process applied here was, however, informed by descriptions of reflexivity given in Finlay (2002) and May and Perry (2017), alongside accounts provided by Krueger et al. (2016) and Preston et al. (2015).

The described approaches involve critical self-analysis, which we define in this context as analyzing one's own influence on the modeling process, and a process of joint discussions to form reflexive accounts of our experiences. The choices made in the modeling and their implications were analyzed as part of the reflexive process to elicit the how and why of the modeling and their influence on outcomes. The adopted approach also involved a third party who acted to provide an external viewpoint to elicit further reflection and pushed forward the reflexive process. The approach aided in drawing out the successes and the struggles encountered when working within an interdisciplinary context. It is acknowledged here that the described approach is subject to some uncertainty as not all those involved in the original case studies could participate (due to availability and the necessary time commitment) and so may not include their valuable insights and perspectives (a matter revisited in Section 2).

The reflexive approach encompasses not just the "technical" decisions made (such as what models to use and the scope of stakeholder engagement) but also acknowledges that the modeling teams form a social system in its own right with their own complex interactions which influence the path taken. Model outcomes are therefore heavily influenced by the social context of the modeling process as well as the technical decisions made therein. Future efforts can be improved by concretely acknowledging this interplay (Catalano et al., 2019; Sterling et al., 2019; Montana et al., 2020). A sociotechnical view was taken to elicit these aspects in the reflexive process.

In the next section, we briefly detail the modeling conducted for the two case studies alongside the reflexive accounts of the choices made in the consideration of modeling scale, the team processes involved, and the decisions made. Both studies employed an SoS approach involving the integration of multiple models to represent the SES of interest. The fundamental need to consider these scale aspects has been previously articulated in Elsawah et al. (2020), Little et al. (2019), Badham et al. (2019), and Hamilton et al. (2015), albeit from different perspectives. We then synthesize the five main lessons learnt from the case studies, which we hope might enhance future SES modeling activities.

2. The case studies

The two case studies represent different facets of the issues that SES modelers face within an SoS context. A reflexive account for each case study is provided below and is aligned with the basic steps in the modeling

Table 1. Overview of each case study including the team context, socioenvironmental systems (SESs) involved, and purpose of the modeling. DOI: https://doi.org/10.1525/elementa.2020.00182.t1

Case Study	Team Context	SESs Involved	Time Steps	Purpose
Sugarcane aphids in Great Plains	Interdisciplinary group including experts in areawide pest management, entomology, and ecological modeling located in several states and employed by federal, state, and private institutions The core modeling team consisted of three ecological modelers, an areawide pest manager, an entomologist, and a meteorologist/aeroecologist	Four in total: agroecological systems (sorghum growth, aphid life cycle, and crop management); meteorological system (airborne aphid dispersal)	Once per model run: crop management model Daily: sorghum growth model, aphid life-cycle model Hourly: meteorological dispersal model	Forecasting sugarcane aphid infestations of sorghum fields within an areawide pest management program, providing infestation forecasts to areawide pest managers and sorghum producers
Campaspe	Large group of participants across different disciplines (>10) geographically spread across many institutions (>6). The team included modeling specialists across five systems, and one generalist who developed the farm model, aided in integrated design and development, and led the integration of models	Seven in total: agricultural, hydrological (surface and groundwater), ecological, climatic variability, policy, and recreational suitability	Daily: surface and groundwater, climate Two weekly: agriculture and policy Once per model run at end of scenario: ecology and recreational water suitability index	Knowledge integration and stakeholder discussion of the range of impacts that changing climatic and policy contexts have on water-related farm and environmental concerns

process. The subsections are not organized identically, however, owing to the different experiences encountered and the focus on providing a reflexive account. Key information is briefly summarized here with an overview provided in **Table 1**, and readers who feel sufficiently informed may skip ahead to Section 3 (Lessons learnt).

Both models are of the SoS type as they leverage constituent models that individually represent separate systems wherein each model could, potentially, be applied separately. As is typical of SoS approaches, each case study (1) considered different time frames and spatial/temporal granularities, (2) spanned multiple systems, and (3) involved multiple disciplines and stakeholders. An aspect of scale to be considered is the process of deciding which representations are to be included or excluded and how they are to be represented in terms of the scale of the modeling to be conducted. Modeling scale therefore includes all aspects of the modeling process including the conceptualization of the model, the relationships between constituent models, model structures, boundaries, parameterizations, implementation approach, and the decisions that underpin each of these. These decisions may be influenced by factors external to the modeling concerns, such as the available resources or imposed legacy software, but are also influenced by the disciplinary representation within the team, the interests represented by stakeholders, and the level of interdisciplinary cohesiveness.

The first case study, referred to as the sugarcane aphids in Great Plains (GPSCA) case study, focuses on areawide integrated pest management of aphids infesting grain sorghum fields across a large spatial area, incorporating local- and regional-scale dynamics. More expansive descriptions for each case study are available—Wang et al. (2019) and Koralewski et al. (2019, 2020a) for the GPSCA study and Iwanaga et al. (2018, 2020) for the Campaspe case study. The GPSCA case study emphasizes modeling choices from a technical point of view, while the Campaspe case study offers a description of the team processes, which influenced decisions during model development.

2.1. The GPSCA case study

The sugarcane aphid is an economic pest of sorghum worldwide (Singh et al., 2004), and outbreaks in U.S. sorghum fields have been recurring annually since 2013. Economic losses result from direct feeding, compromised harvesting efficiency, and damage to harvesting equipment and may exceed 50% of the yield (Bowling et al., 2016). Aphids are highly prolific and disperse with wind over long distances within the prairie-steppe region of the North American Great Plains, which is the principal sorghum production area in the United States (van Rensburg, 1973; Singh et al., 2004; USDA-NASS, 2010; Bowling et al., 2016).

Two key tactics within an areawide integrated pest management program for cereal aphids include deployment of aphid-resistant sorghum cultivars and selective use of insecticides (Elliott et al., 2008; Giles et al., 2008; Brewer et al., 2019). The model of Wang et al. (2019) was developed to support wise use of these management

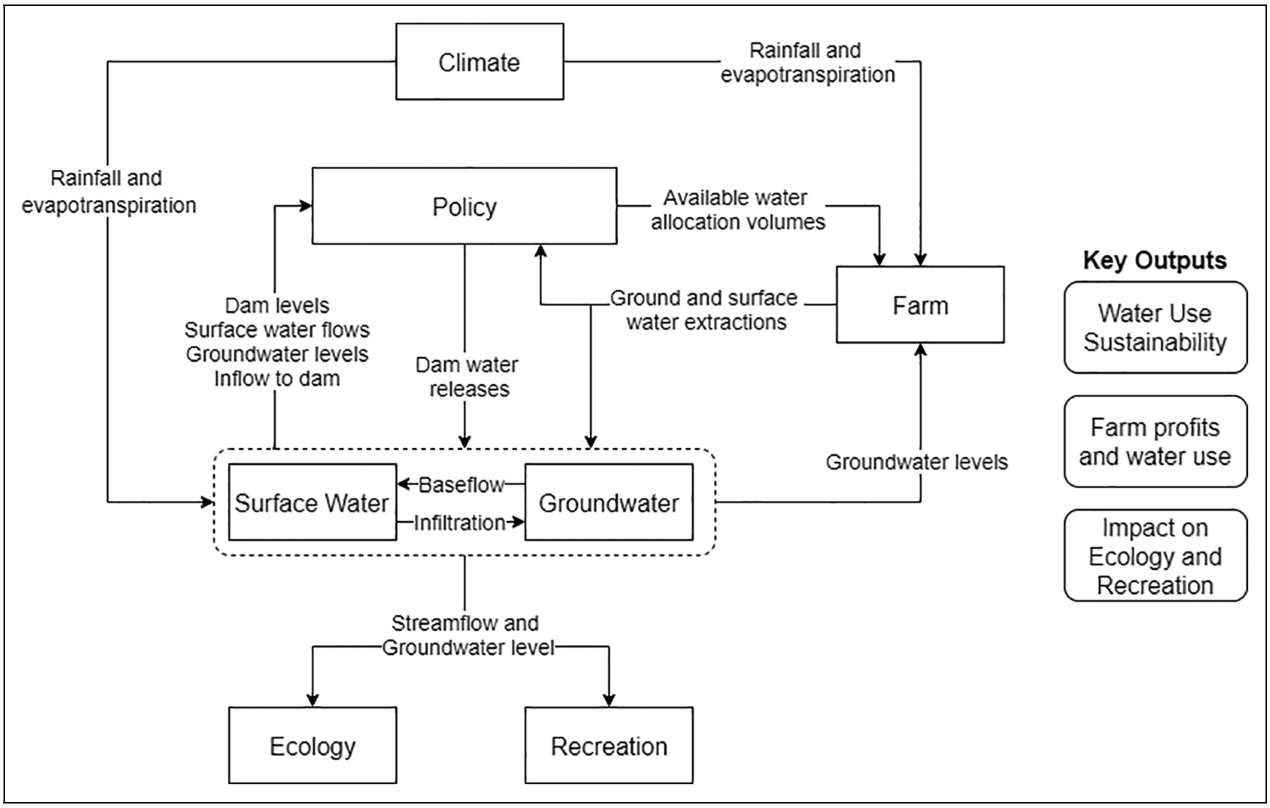


Figure 1. Relationship and interactions between constituent models and the key outputs of the Campaspe integrated model (Iwanaga et al., 2020). Each box represents a constituent model. Dashed line around surface and groundwater models is to simplify the diagram and does not signify a separate model. Arrowed lines indicate the process of data interoperation; the direction of interaction and the data communicated between models. DOI: https://doi.org/10.1525/elementa.2020.00182.f1

tactics for sugarcane aphids. The purpose of the model was to simulate areawide spatiotemporal patterns of sugarcane aphid infestations of sorghum fields, with a focus on timing of initial infestations. Near-real-time model forecasts could inform growing season activities such as timing of field monitoring to detect aphids and optimal insecticide use (Koralewski et al., 2020b). Model outputs also could be useful for region-scale management recommendations such as deployment of aphid-resistant sorghum cultivars (Koralewski et al., 2020a).

2.1.1. Conceptualization

The core modeling team that developed the conceptual model consisted of three ecological modelers: an areawide pest manager, an entomologist, and a meteorologist/aeroecologist. Although we all conceptualized the SoS as consisting of linked agroecological and meteorological systems, agreement on representation of causal processes operating within, and especially between, these two constituent systems was achieved only after considerable debate. The debate was centered explicitly on our choice of appropriate temporal and spatial scales at which to represent system processes. However, implicitly, we were debating the level of causality to include in the representation of those processes. That is, did we require our model, or parts thereof, to be interpretable as embodying cause–effect relationships or did we require only that model outputs correspond well with available real-world observations. Below, we describe the details of the conceptual model of the integrated SoS that emerged as a shared understanding of the modeling team (see also figure 1 in Wang et al., 2019).

Important processes modeled in the agroecological system included sorghum growth, aphid development, and crop management. Both sorghum growth (through phenological stages) and aphid development (through life-cycle stages) were modeled primarily as a function of environmental temperature. Crop management (i.e., decisions to plant aphid-resistant sorghum cultivars and rules for insecticide use) was modeled as a set of external variables. Important processes modeled in the meteorological system included emigration (time and location of aphid "takeoff"), wind-borne aphid migration, and immigration (time and location of aphid "landing").

Migration was modeled primarily as a function of wind velocity and direction and flight duration. The processes of emigration and immigration linked the agroecological and meteorological systems, with emigration initiated in the agroecological model (based on sorghum phenological stage and aphid life stage) and immigration initiated in the meteorological model (based on the deposition pattern of aphids). Additional conceptual details on linkage of the meteorological and agroecological components are provided in Koralewski et al. (2019; figure 1 therein).

The spatial extent of the model included the primary sorghum-producing areas in the United States, including Kansas, Oklahoma, and Texas. The spatial resolution was set to 0.5° latitude by 0.5° longitude, which resulted in about 700 georeferenced landscape cells of approximately 2,500 km^2. This coarse-grained scale facilitated linkage of the agroecological system of the SoS to an existing atmospheric particle trajectory model (HYSPLIT; Stein et al., 2015; see Section 2.1.3.2 on model construction). The number of cells may vary annually depending on whether or not sorghum is present in the cell during a given year. The temporal scale was 1 year, which allowed for encompassing one sorghum growing season, and thus the maximum potential extent of aphid infestations, with resolution of a daily step to capture important details related to phenological development of sorghum and population dynamics of aphids. Immigration of aphids from outside of the southern boundary was approximated as occurring from the Rio Grande Valley along the border between Texas and Mexico based on reported field detection of aphids (Bowling et al., 2016). Aphids arrive in the Valley from Mexico "unannounced" because producer reports of infestations are not available from Mexico on a regular basis. We elaborate our rationale for scale choices in the subsection below.

However, before proceeding further, we reflect for a moment on our modeling team for this case study. Our team did not include a social scientist. This was not by design, in the sense of an explicit decision in favor of exclusion. Rather, it resulted from funding priorities within the overall project and the associated restriction on the number of modeling team members. Our team also did not include sorghum producers. Our entomologist maintained close ties with numerous producers via his agricultural extension activities and could explain their perspectives and main interests with relative confidence. Nonetheless, sorghum producers did not participate directly in discussions among the members of the modeling team.

Ideally, a social scientist and at least one sorghum producer would have been members of the core modeling team from the beginning. These "social voices" would have enriched our shared understanding of the SES and, arguably, could have modified the course of model development. For example, one of our livelier debates during model development, which we describe below in Section 2.1.2.2, undoubtedly would have included a more detailed discussion of the guidelines that have been developed for sorghum planting dates and subsequent management activities. We currently are exploring ways to quantify producer decision making within predominantly biophysical models (e.g., Wang et al., 2020b), which may be applicable to the GPSCA model. But perhaps more challenging than the quantitative details involved are the financial and logistical problems that impede direct long-term involvement of stakeholders in the model development process.

2.1.2. Scale choices

The multifarious reasoning that led to our choice of scales is not intuitively obvious. In principle, we based our determination of spatial and temporal scales (outlined in Table 1) on model objectives, the ecology of the organisms involved, the level of detail contained in information available from literature and from stakeholders (sorghum producers), and computational considerations. Sorghum producers did not participate directly in discussions among members of the core modeling team. However, our entomologist maintained close ties with numerous producers via his agricultural extension activities and could represent with confidence their perspectives and main interests. Spatial and temporal scales both spanned several orders of magnitude. The spatial scale of interest ranged from the regional management perspective (approximately 1.75 million km^2 of modeled area) to that of the sorghum producers' and field scientists', focused on a single sorghum leaf which, for practical purposes, encompasses an aphid colony. Temporal scales of interest ranged from an approximately 9-month period of sorghum availability in the region (for regional managers) to a "near-real-time" estimation of aphid density in a sorghum field (for sorghum producers and field scientists).

Scale choices were complicated further because aphids are small (approximately 0.05 mg) and prolific (population doubling time as short as a few days). Non-winged (apterous) morphs are relatively sedentary, whereas wind-borne dispersal can carry winged (alate) morphs over long distances (hundreds of kilometer). Thus, densities of local colonies can exceed 1,000 individuals per sorghum leaf, while emigrants from a single colony can be dispersed over thousands of square kilometers. Important life processes occurring during the terrestrial portion of the aphid life cycle are commonly measured in terms of daily rates, and the most common metric used to record field measurements of aphid densities is individuals per sorghum leaf. Sorghum development through phenological stages also is measured in terms of days (or "degree-days") per stage. However, important dynamics occurring during wind-borne aphid migration result from physical environmental conditions (wind velocities and directions) that are highly variable over the entire U.S. Great Plains.

Reflecting on these various considerations, we needed to "scale up" spatially and temporally from representation of the agroecological processes occurring at the individual aphid/sorghum leaf interface to generate seasonally variable regional patterns of aphid infestations of sorghum of interest at the areawide pest management level. Placing our model objectives within the context of Levins's (1966) classical modeling trade-offs (precision vs. generality vs. realism), it also seemed clear that our priority was realism. That is, we wanted to explicitly consider the agroecological characteristics specific to the south-central U.S. Great Plains.

In addition, we wanted to explicitly consider stochastic effects on these agroecological processes that are dependent on meteorological conditions. Infestation forecasts needed to be probabilistic. Within this context, the inherent stochasticity of the SoS and the parametric uncertainty associated with representations of system processes shaped our scale decisions. To provide some insight into our thought processes, we initially focused on the model output of most interest to end users and worked our way back to sets of modeled processes that might generate that output, noting the relative level of detail included in representation of the various processes (Wang and Grant, 2021).

2.1.2.1. Model output

The model output of most interest to end users (sorghum producers) was a set of calendar dates indicating when aphids were most likely to first infest their sorghum fields. We began by conceptually bounding the level of detail at one end with a deterministic, static, correlative model that estimates a mean date of the first infestation of the south-central U.S. Great Plains based on observed first infestation dates (which date back to 2013). At the other end, we conceptually bounded the level of detail with a dynamic, spatially explicit, individual-based model that represents all of the individual sorghum leaves in the south-central U.S. Great Plains and all of the individual aphids that might infest them.

Given the purpose, a useful model needed to be probabilistic, dynamic, and spatially explicit. Thus, regarding spatial and temporal scales, we divided the south-central U.S. Great Plains into smaller-sized areas and the approximately 9-month period into shorter time steps. Furthermore, we knew that producers were most interested in their sorghum fields and in associated management activities (e.g., planting, monitoring for aphids, pesticide applications), which might be shifted by a few days or weeks. Areawide pest managers were interested in helping individual producers make such decisions, but via more synoptic infestation forecasts, which could be individualized by local agricultural advisors (e.g., in the United States, agricultural extension agents working at the county level). Thus, for end users, the model needed to provide daily forecasts that could be interpreted at farm-level and regional-level spatial scales.

2.1.2.2. Process representation

System processes needing to be explicitly modeled included those at the sorghum/aphid/crop management interface. As we mentioned in Section 2.1.1 on model conceptualization, our main debates about scale choices were primarily debates about the level of causality to include in the representation of SoS processes. In particular, we debated whether our model, or parts thereof, needed to be interpretable as embodying cause–effect relationships. Below, to avoid an overly confusing description of process representation, we first present our final shared understanding of the appropriate scales to use. We then conclude this section with an attempt to provide some insight into the sorts of debates that led to that shared understanding.

Guidelines have been developed for sorghum planting dates and subsequent management activities in terms of latitudinal differences in weather patterns during the growing season. Population dynamics of sugarcane aphids on grain sorghum have been widely studied over the past several years, although our ability to quantify with confidence the effects of aphid-resistant sorghum cultivars, natural aphid enemies, and proximate causes of emigration remains quite limited. The fact that migrating aphids are dispersed by the wind as essentially inert particles above the flight boundary layer (i.e., a few meters above ground level) allows representation of migration via the use of well-developed meteorological particle dispersion models but also results in the uncertainty necessarily associated with weather forecasts.

Thinking about positioning our representations of these processes at the sorghum/aphid/crop management interface with regard to the level of detail included in the representations, it seemed that the modeled processes should meet two criteria. They should generate output directly comparable to personal observations commonly made by end users, and they should be viewed by research scientists as being acceptable mechanistic representations. The most common observational metric used by producers and field biologists was the number of aphids on a sorghum leaf. Usually, several leaves per plant and several plants per field were sampled on a given day, with results accumulated over time and summarized at field-, farm-, county-, and regional-level spatial scales. Regarding mechanistic (cause–effect) representation, we emphasized the term "acceptable" to acknowledge that causality is defined subjectively. The requisite level of detail to claim that a process is represented mechanistically is to a large degree problem-specific.

There was a reasonably narrow range of defensible levels of detail to consider for the model to be viewed as mechanistic. Aphid development, reproduction, mortality, and emigration, as well as processes affecting the quality of sorghum leaf (sorghum phenological development), were represented as functions of environmental temperature modified by aphid density and seemed a defensible "mark" along the level of detail scale for the agroecological model. One step toward the more detailed representation might be marked by a representation of the processes just mentioned explicitly in terms of the physiology involved in sorghum and aphid development and the frequency of physical contact among aphids. One step toward a less detailed representation might be marked by an implicit representation of these processes in terms of sorghum phenological stage and aphid population density as functions of days since planting and days since initial infestation, respectively, and emigration as a function of population density per se.

The level of detail for representation of agroecological processes that met the two criteria just described suggested a sorghum leaf and a day as appropriate spatial and temporal scales. This left us with two final considerations related to scale choice. One involved summarizing numerically the results of mechanistically modeled daily processes occurring on individual sorghum leaves in terms of a set of calendar dates indicating when aphids were most likely to first infest sorghum fields in the south-central U.S. Great Plains. The other involved dealing with potential phase shifts along the level of detail continuum that might be needed when passing information about migrating aphids between the agroecological and meteorological models.

The first step in summarizing results from individual sorghum leaves involved deciding how many leaves we needed to represent explicitly, how they might differ from one another, and how aphids on one leaf might affect aphids on another leaf. There is, however, relatively large observed variation in aphids/leaf on a single plant, aphids/plant within a single field, and aphid densities among neighboring fields, as well as spatial variation in environmental temperatures to which leaves (and the aphids on them) were exposed. We felt comfortable, therefore, letting a single sorghum leaf represents a mean-field approximation of the conditions of sorghum leaves over an area large enough to be of interest from the synoptic perspective of areawide managers.

We felt that forecasts summarized probabilistically from this synoptic perspective also would be interpretable at the farm level by producers. Since we would be executing sets of Monte Carlo simulations to make infestation forecasts, which would encompass the environmental stochasticity inherent in the modeled system, they could be interpreted in a similar manner to local weather forecasts. Producers were accustomed to inferring probable future weather conditions for their specific location based on weather forecasts for areas much larger than their sorghum fields. They also were accustomed to interpreting field-based observations of aphid infestations summarized at the county level in terms of infestation likelihoods for their fields. The final detail involved in summarizing results based on dynamics occurring on single sorghum leaves simply involved making the required unit conversions. For this, we had estimates of mean number of leaves per sorghum plant, mean number of sorghum plants per hectare, and number of hectares of sorghum within various-sized areas of the south-central U.S. Great Plains.

Regarding potential phase shifts along the level of detail continuum that were needed when passing information between agroecological and meteorological models, we identified two. One was conceptual and one was quantitative. Conceptually, aphids were treated as inert particles in the meteorological model as they are weak flyers. Within the meteorological model, particle depositions were updated hourly (during the 12-h migration time), but deposited particles (immigrating aphids) were passed back to the agroecological model as daily cohorts.

Quantitatively, aphids underwent a phase shift within the meteorological model in that we severed the numerical connection between the number of aphids emigrating and the number of aphids immigrating by placing an arbitrarily small number of (super-) aphids on each sorghum leaf receiving immigrants. Although not ideal, we felt this phase shift did not compromise the forecasting ability of the model. Given the variable size of emigration events, the lack of data on mortality rates during migration, and the dependency of successful colonization on the time lag between arrival of immigrants and arrival of natural enemies, we felt colonization could be represented appropriately as a stochastic event occurring within any landscape cell in the agroecological model (Wang et al., 2020a).

Having presented our final shared understanding of appropriate scales, we now attempt to provide some insight into one of the livelier scale debates with regard to the level of detail with which to represent SoS processes. As we described above, our final decision with regard to aphid development, reproduction, mortality, and emigration was to represent these processes as functions of environmental temperature modified by aphid density. Our meteorologist/aeroecologist would have been satisfied with a "causal" representation of aphid population dynamics that represented population density as a function of number of days since initial infestation and emigration as a function of population density. Such a representation was perceived as unacceptably phenomenological by our entomologist. Our entomologist initially proposed a more mechanistic representation of the aphid life cycle, which included, among other things, mortality due to natural enemies (predators and parasites). Arguably, aphid population growth depends on timing and magnitude of mortality imposed by their natural enemies, which depends on species composition of the community of natural enemies, which depends on the characteristics of the landscape surrounding a sorghum field. However, in view of (1) the site-specificity of such relationships, (2) the fact that connection of the terrestrial portion of the SoS model with the aeroecological portion required just a single number of aphids emigrating from each of the approximately 2,500 km^2 landscape cells, and (3) the fact that the purpose of the model was to simulate areawide spatiotemporal patterns of aphid infestations, our entomologist agreed to a simpler "causal" representation of the aphid life cycle. The simpler representation upon which we finally agreed was acknowledged as acceptably "causal" by our entomologist because of the general acceptance among subject-matter experts of the temperature dependency of insect reproduction and development and the density dependency of aphid emigration. Our meteorologist/aeroecologist doubted that model output would be improved by this, from his perspective, more complicated representation but acknowledged the benefits in terms of increasing model credibility.

2.1.3. Development

The integrated SoS model was built for use specifically within the context of the areawide pest management program for sugarcane aphids in the south-central U.S. Great Plains. It was developed by the three ecological modelers, all of whom worked at the same physical location. The modelers maintained frequent direct communication with the areawide pest manager, the entomologist, and the meteorologist/aeroecologist, each of whom facilitated indirect communication with a broad array of specific subject-matter specialists, as well as sorghum producers throughout the south-central U.S. Great Plains.

2.1.3.1. Collecting data, information, and knowledge

Several important processes included in the agroecological model had been studied extensively. Data representing growth of sorghum and development of sugarcane aphids to environmental temperature were available in the scientific literature. Information on crop management (e.g., guidelines for planting and harvesting) for sorghum in the U.S. Great Plains had been summarized and was easily accessible. Other important processes, while generally understood conceptually, could not be quantified based on available data. Proximate causes of aphid mortality and emigration remained topics of debate among subject-matter specialists. We drew upon the knowledge of the core modeling team, supplemented by the array of subject-matter specialists with whom we communicated, to quantify these processes.

Most of the important processes needed in the meteorological model had been incorporated into an existing, readily available, atmospheric particle trajectory model (see next section), which we used to simulate wind-borne aphid migration and subsequent immigration (particle deposition; aphids are weak flyers and, once airborne, are dispersed essentially as inert particles).

Specifically, the agroecological component uses data on air temperature at the soil surface and at 2 m above the soil surface, sorghum planting and harvest dates, and percentage of land on which sorghum was grown. Published information was used to model sorghum growth stages (Gerik et al., 2003), sorghum leaf area (Roozeboom and Prasad, 2019), sorghum harvest dates (USDA-NASS, 2010), aphid life stages (Davidson, 1944; Poché et al., 2016), aphid reproduction (Brewer et al., 2017; Hinson, 2017), and density-dependent reduction of aphid population size (Brewer et al., 2017). EDAS 40-km resolution data (National Oceanographic and Atmospheric Administration, 2019) were used as input for the atmospheric dispersion model HYSPLIT (Stein et al., 2015). HYSPLIT also received georeferenced information on emigrating aphids from the ecological component of the model. References for data and other sources of other information used to parameterize the agroecological and meteorological models are available in Wang et al. (2019).

Documentation to support interdisciplinary cohesion followed established standards for documenting individual-based (or agent-based) models in the field of ecological modeling, including the overview, design concepts, and details (ODD) protocol (Grimm et al., 2006, 2010).

2.1.3.2. Construction

The agroecological component of the integrated model was constructed using the individual-based modeling framework NetLogo (Wilensky, 1999). The need to model aphid life-cycle processes at an acceptably "causal" scale (see Section 2.1.2.2) prompted our choice of an individual-based model. Our choice of NetLogo over other types of modeling platforms within which individual-based models can be developed (e.g., see Ch. 8 in Grimm and Railsback, 2005) was based on our familiarity with NetLogo, its wide acceptance for individual-based modeling in ecology (Grimm et al., 2020), and its facilitation of model documentation via the ODD protocol. Our choice of NetLogo imposed computational limitations with regard to the number of individual entities that could be represented explicitly during simulations, as we describe below. The meteorological component was constructed using the established and widely used atmospheric particle trajectory model HYSPLIT (Stein et al., 2015), which computes airborne dispersal of aphids as inert particles. The NetLogo and HYSPLIT components were connected computationally with a custom-developed algorithm "Link" (Koralewski et al., 2019), with data exchange possible at a daily time step. The NetLogo platform is often used for individual-based ecological models (see, e.g., Thiele et al., 2014).

Two HYSPLIT input files EMITIMES and CONTROL are used to pass georeferenced information on emigrants from the agroecological component of the model. HYSPLIT estimates synoptic dispersal of aphids aloft. The georeferenced information on aphid immigrants is passed back to the agroecological component of the model, and subsequent updates of landscape cell states follow. Considering the spatial resolution and the regional scale, and to reduce the overall computational cost, a cohort of aphids is represented by a collective super-aphid (Scheffer et al., 1995).

An individual-based modeling approach allowed explicit representation and customization of the stage- and morph-specific reaction of sugarcane aphids to changing environmental conditions (e.g., sorghum phenological stage and environmental temperature). These reactions, or behavioral responses, of individual aphids were programmed in NetLogo via sets of equations, often embedded within logical statements. The rules were realistic, that is, they were interpretable in terms of sugarcane aphid physiology and ecology on grain sorghum in the south-central U.S. Great Plains. Population-level phenomena of interest (e.g., migration events) then emerged as the cumulative result of understandable cause–effect reactions of individuals rather than as a correlate of an arbitrary index, such as calendar date.

The conceptual basis for our choice to use an existing atmospheric model was the universal applicability of the laws of fluid mechanics upon which such models are founded. Thus, our need for a realistic integrated model, which required a "custom-built" agroecological model to accommodate the unique biological characteristics of the organisms involved, was not compromised by the generality of a model based in the physical sciences; of course, as per Levins (1966), we necessarily sacrificed precision in the sense that any realistic ecological model will contain stochastic effects, which will inevitably reduce precision (Evans, 2012). As noted earlier, aphids were treated as inert particles during the migration phase. Parameterization of the particle dispersion model required specification of the point sources (latitude and longitude) of particle emission (aphid emigration), number of particles (aphids) emitted, altitudes (meters above ground level) at which particles are dispersed (migration altitudes), and duration (hours) of dispersal events (migration duration).

Computational considerations limited the number of entities that could be dealt with numerically during simulations. We reduced the number of entities involved in calculations by simulating the phenological development of only one sorghum plant within each ≈2,500 km^2 landscape cell and the population dynamics of the aphids on only one leaf on each plant. That is, each aphid population consisted of a series of daily cohorts, with each cohort (superindividual) representing a variable number of identical aphids. The number of aphids represented by a superindividual was initialized by a reproduction or immigration event and subsequently reduced by mortality and emigration events. Each simulation, which forecasted spatiotemporal patterns of aphid infestations of sorghum during one growing season, required less than an hour of runtime on a desktop personal computer, and the necessary data input files for the meteorological model fitted comfortably within available data storage space.

Worthy of comment here is the fact that we did not face model construction problems related to concurrent development of the agroecological and meteorological models. The following case study describes communication problems, both human and computational, associated with the integration of models that were being developed concurrently (see Section 2.2.3.3). Although we needed to develop a customized algorithm ("Link") to connect NetLogo and HYSPLIT, the information passed between the two models (aphids treated as inert particles) did not change as a result of coding changes in NetLogo during the development of the agroecological model.

2.1.3.3. Model calibration

Model calibration was twofold. First, sorghum development was calibrated to adjust simulated sorghum harvest dates and number of days from planting to harvest to those reported by USDA-NASS (2010). Second, the regional migration of aphids was calibrated to adjust the simulated spatiotemporal pattern of infestations to field data from sorghum producers in Texas in 2017. This step was accomplished by adjusting colonization probabilities and did not require changes to the meteorological component of the integrated model.

2.1.4. Uncertainty analysis

The primary source of uncertainty in the integrated SoS model arose at the intersection of aphid terrestrial ecology and airborne aphid dispersal. At the time this study was published, we based this assessment on an informal sensitivity analysis that consisted of qualitative analyses of aphid infestation maps (based on expert opinion) produced by simulations with a variety of different iterations of parameters in the agroecological and aeroecological portions of the model (the maps analyzed were analogous to those in figure 8 of Wang et al., 2019). We describe the manner in which we conducted this initial, and a subsequent, sensitivity analysis in the next section on model testing and evaluation. Initiation of emigration from local populations likely depends on (1) host plant growth stage, (2) pest density and (3) developmental stage, and (4) weather or some combination thereof (Parry, 2013 and references therein). There also was uncertainty regarding duration of migration events, mortality while aloft (and thus also vigor upon landing), and aphid responses to meteorological factors in general while aloft (Eagles et al., 2013). Since processes governing initiation of emigration were modeled at the surface of a sorghum leaf, whereas processes governing airborne migration were modeled over the entire south-central U.S. Great Plains, scale issues pervaded uncertainty analysis. Furthermore, end users of the model fell into two groups with different spatiotemporal perspectives on system uncertainty.

Model purpose dictated that uncertainty analysis be focused primarily on forecasts of timing of initial aphid infestations of sorghum fields. Day-of-year of initial infestation is a common metric used by both areawide pest managers and sorghum producers to analyze and discuss infestation dynamics. However, a statement that an infestation may occur sometime during a 10-day period is likely to be interpreted quite differently by an areawide manager compared to a producer. From the spatiotemporal perspective of an areawide manager, a 10-day window of uncertainty associated with the northward advance of an aphid infestation front over the south-central U.S. Great Plains during the sorghum growing season may provide useful planning information. But from the spatiotemporal perspective of a producer, such a window of uncertainty associated with the first appearance of aphids in their sorghum field may be less useful. Likewise, a forecasted infestation front advancing via 2,500 km^2 "footsteps" may provide useful areawide management information but be less useful to a producer with a few thousand hectares of sorghum. Nonetheless, although synoptic areawide forecasts may not contain the specificity desired by producers, they do contain useful information if the forecast uncertainty is interpreted within the appropriate spatiotemporal context. Analogous to regional weather forecasts, uncertainty inevitably increases with decreasing spatial scale. SoS modelers might make more effective use of this analogy when interpreting their uncertainty analyses to end users.

2.1.5. Testing and evaluation

The initial assessment of model structure, linkages between model components, and overall model function was performed to verify overall correspondence with model purpose and to identify potentially missing components. Model behavior was then evaluated regarding the ability to produce the general south-to-north temporal trend in emergence of sorghum and the subsequent infestation of sorghum fields by sugarcane aphids.

Simulated and observed spatiotemporal patterns of aphid infestations were then compared to validate the model. The simulated data were based on 10 replicate stochastic simulations. The field data were collected in Texas, Oklahoma, and Kansas during 2017 and were not used in model development. The average simulated dates of first aphid infestations were within the range of observed dates of first infestations in four of the five sorghum growing regions (Wang et al., 2019; figure 5). The ranges of observed dates were narrower than the corresponding simulated ones, which was attributed to the fact that all simulated infestations were detected whereas field data were limited by temporal and spatial field sampling constraints. Initial testing and evaluation details are available in Wang et al. (2019).

After publication of the work reported above, in which model testing was limited by the ever-present combination of limited funding and impending deadlines, we were fortunate to have the opportunity to extend our testing in two areas of particular interest. The testing was basically a sensitivity analysis that consisted of varying the value of one parameter at a time in either the agroecological model or the aeroecological model and qualitatively assessing the effects on SoS model outputs. Both involved aphid migration, the key process (which includes the processes of immigration and emigration) connecting aphid terrestrial ecology and airborne aphid dispersal. We were interested particularly in evaluating more formally the uncertainty in model outputs describing spatiotemporal infestation trends resulting from uncertainty in the values of key parameters affecting migration. First, we evaluated the effects of altering timing of first appearance of aphids in the southernmost U.S. Great Plains. The first appearance of aphids is an initial condition of the agroecological model representing immigration from an external source (Mexico). Next, we evaluated effects of altering dispersal duration, minimum dispersal height (meters above ground level), and maximum dispersal height. Dispersal duration and heights are parameters controlling airborne dispersal in the aeroecological model. Results of these new tests indicated alteration of the timing of first appearance of aphids in the southernmost U.S. Great Plains affected forecasted spatiotemporal patterns of infestation (as indicated by georeferenced probabilities of first infestations) throughout the entire south-central Great Plains region (Koralewski et al., 2020a, 2020b). However, alteration of the three dispersal parameters, over the 63 combinations of values tested, had little effect on georeferenced probabilities of first infestations.

These new results more clearly identified the timing of first aphid infestations in landscape cells as the primary source of forecasting uncertainty in the integrated SoS model. They also suggested some rescaling of modeled processes that would be interesting to examine from the standpoint of increasing utility of infestation forecasts for sorghum producers, specifically, reducing the level of detail with which we represent processes in the agroecological model and increasing the spatial resolution with which we represent migration in the aeroecological model. We have conducted a series of thought experiments, which suggests accurate forecasting of timing of initial infestations is more important than accurate forecasting of magnitudes of migrations and initial infestations within the context of areawide pest management (Wang et al., 2020a). Given the high fecundity and rapid development of aphids at temperatures characteristic of the sorghum growing season, time lags between initial infestation, and the presence of potential emigrants is only a few days. Aphid colony growth versus local extinction depends on interaction of myriad processes (see Section 2.1.2.2) that can be aggregated into a single stochastic variable without increasing the level of uncertainty associated with colony survival and production of emigrants. However, increasing the spatial resolution of simulated immigration points poses a technical problem. Although there is increasing availability of high-resolution atmospheric data and increasing sophistication of atmospheric particle trajectory models, it is unlikely that data supporting field validation of fine-scale immigration forecasts will be available in the foreseeable future.

2.2. The Campaspe case study

The Campaspe study focused on the long-term management of water resources between agroeconomic and environmental concerns at a regional scale, under a backdrop of uncertain future climate and policy conditions. The study area, the Lower Campaspe subcatchment, is in South-East Australia and part of the Southern Murray–Darling Basin. The area is of ecological, socioeconomic, and agricultural importance. Increasing agricultural and environmental concerns and the impact of recent droughts (e.g., the Millennium Drought, 1996–2010; Kendall, 2013) have spurred a series of hotly contested water policy reforms. Regionally, riverine health is said to be poor (Murray–Darling Basin Authority, 2012; North Central CMA, 2014) and is set to become increasingly challenging, especially under uncertain climate conditions (Dey et al., 2019) that are likely to exacerbate water availability. The Campaspe integrated model (CIM) was developed to facilitate discussion among stakeholders of the long-term implications of water management decisions and potential policy changes, including conjunctive use of surface and groundwater, under a range of uncertain futures.

The interplay between the scale decisions made by the team and the implications regarding modeling scale and treatment thereof is explored here. Some context on the team and the model development approach is first provided (in Sections 2.2.1 and 2.2.2), followed by an exploration of the scale issues, and the decisions in their treatment in Section 2.2.3. The team aspects and decision choices from a scale perspective are the focus of the exploration.

2.2.1. The team context

The team consisted of specialists and research students across the fields of ground and surface water hydrology, the social sciences, software engineering, economics, systems analysis, and uncertainty assessment. Local specialists in water management, agricultural and ecological matters were engaged as part of the project. Organizationally, the team spanned six Australian institutions. Subgroups within the team each focused on an aspect of the SES. The bulk of the team had prior working relationships conducting integrated assessments, but this was the first time their models were so intimately integrated and in a manner that accounted for feedbacks between systems.

The team previously underwent a self-reflection process using a survey-based approach (discussed in Zare et al., 2021). The "Monitoring and Evaluation" process described therein aided in identifying opportunities for improvement of practices that could better structure the modeling processes and enhance team efficiency. The account provided here differs from the first in that the focus here is on the issues of scale that arise throughout rather than demonstrating the value of self-reflection in the modeling process. Common experiences then inform the lessons learnt (discussed in Section 3).

2.2.2. Development and application

The CIM was developed to represent the spatiotemporal forcing and system interactions that changing climatic, market, and policy contexts have on water-related farm decisions and profits, as well as catchment-scale groundwater and ecological concerns. Team members self-organized to develop constituent models for this SoS model and, at least initially, focused on the processes and issues of concern specific to their system of interest. The approach, and the number of people involved, then had interrelated implications regarding the treatment of scale issues and the decisions therein (which are explored in Section 2.2.3). Here, the approach to construction and simulation of the model is described to provide some context.

2.2.2.1. Construction

To address the spatiotemporal forcing and system interactions that changing climatic and policy contexts have on water-related farm and environmental concerns, an integrated model built from a collection of system-specific models was developed. Having experience in integrated assessment, modelers were aware that models would be dependent on data interoperated between models. A practical approach was taken in integrating these models, and so the CIM operates on a linear feed-forward concept where outputs from one model are fed into other models with which it has a direct relationship (see **Figure 1**). Interoperation of data occur at a daily time step for surface and groundwater models and a two-weekly step for policy and agricultural models. Feedback between models occurs once at their respective time steps, except for the two indicator models (i.e., ecology and recreation impact evaluation) that are run at the end of a scenario. Further detail on the models is provided in Appendix A.

It was known and expected early in the modeling process that the constituent models were to be developed in a variety of approaches and programming languages. Different development environments (e.g., laptop vs. supercomputer) would have to be accommodated. Technical integration of the constituent models was achieved through a purpose-built (software) framework developed in Python. The primary reason for Python is that it is cross-platform and is popular within the sciences as a "glue" between models (Muller et al., 2015; Dysarz, 2018).

2.2.2.2. Simulation approach

Exploratory scenario modeling (ESM) was the selected approach in simulating outputs with the CIM as it allows for the consideration of a multitude of plausible futures in conjunction with scenario, model, and decision uncertainty (Maier et al., 2016; Horne et al., 2019). Certainly, the involvement of researchers with a history and expertise in uncertainty assessment brought considerations of uncertainty to the forefront. Another key reason for the adoption of ESM was to better enable the communication of the scale of uncertainty to local stakeholders, which may influence the decisions enacted (Maier et al., 2016; Little et al., 2019).

Exploratory approaches involve many model runs, with each run representing a possible plausible future (i.e., a scenario) under a variety of conditions. With the CIM, these include hypothetical policy changes (e.g., conjunctive use of surface and groundwater resources), changing climate conditions, market prices for commodities and input costs, and on-farm management options to allow assessment of impacts on the agricultural, groundwater, recreational, and ecological systems.

2.2.3. Scaling issues

The scales to be represented in the CIM were identified through analyzing the needs and purpose of the individual systems of interest as well as the intersystem relationships that needed to be represented. These included the agricultural, hydrological (surface and groundwater), ecological, climatic variability, policy, and recreational systems. Specific aspects of these to be represented by models were informed by the range of local stakeholder interests and concerns. Interactions between the seven systems then enhance or degrade the ability to meet the needs of all water users over time. From these, the spatial and temporal scales (including extent and granularity) that were amenable to the context and purpose of the model were identified.

Nominally, each model was informed by both the natural and anthropogenic properties of the catchment. These included water management zones (i.e., areas subject to differing policies), aquifer boundaries, the hydrologic subbasins in the study area, and the available data. The spatial area represented by the surface water, groundwater, and farm models is depicted in **Figure 2**, and a summary of the spatial and temporal scales internal to each model is provided in Appendix A. Further details on the modeling context and findings are available in Iwanaga et al. (2018, 2020).

Because of the number of models and disciplinary experts involved and some geographic dispersion between the team members, maintaining a high degree of cohesion throughout the modeling process was challenging. In the subsections below, a reflexive account is given of the considerations of the SoS approach on the level of detail, participation, interdisciplinarity and team cohesion, and subsequent implications encountered in practice.

2.2.3.1. Scale of detail

Identifying and representing the systems of interest at a level of detail commensurate with the modeling purpose is one challenging aspect that leads to multiple, equally plausible system representations. The farming system, for

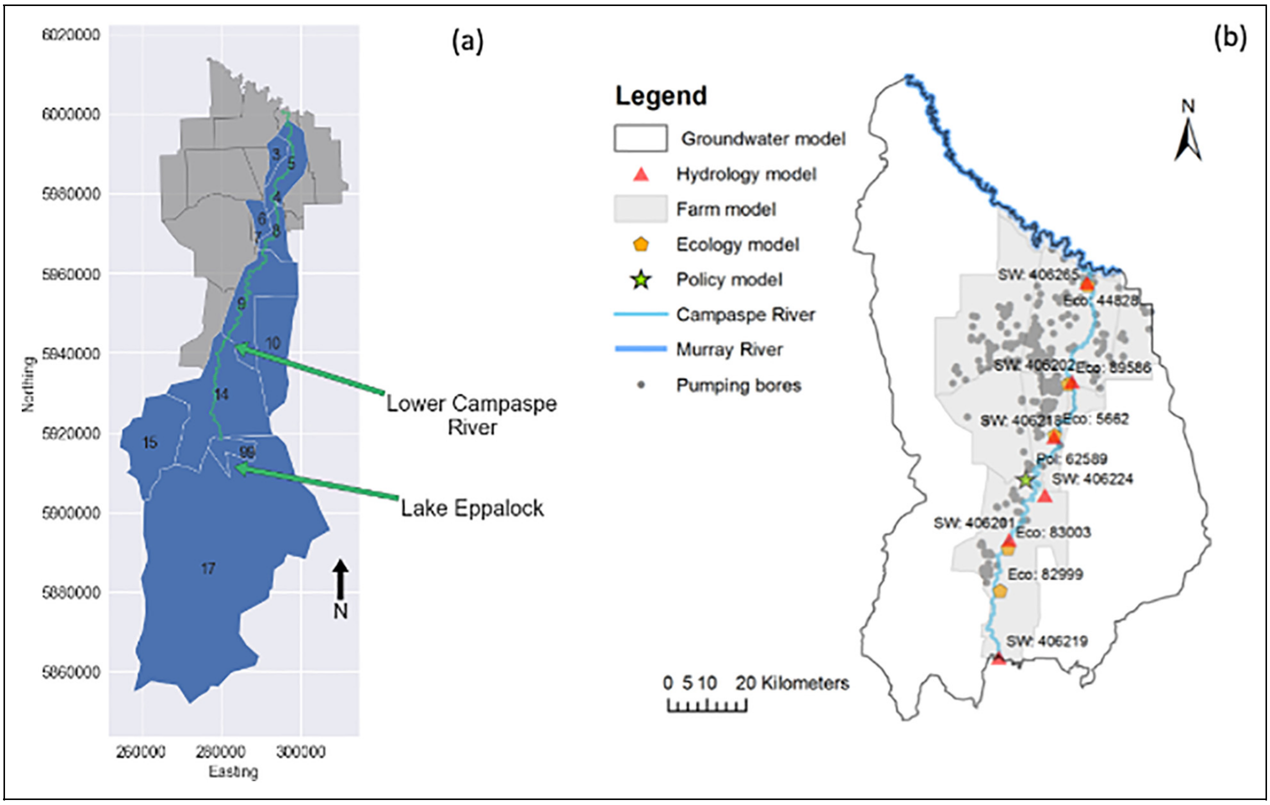


Figure 2. Surface water area (Panel a, left-hand side with subcatchment identifiers annotated) and groundwater area (Panel b, right-hand side) with farm management zones (semitransparent gray areas in both panels). Surface water area extends further south compared to the other models, whereas the represented groundwater area extends further east and west. Figure adapted from Iwanaga et al. (2018). DOI: https://doi.org/10.1525/elementa.2020.00182.f2

example, was represented as a collection of 12 spatially lumped zones primarily determined by local planning areas (known as the Goulburn–Murray Water Supply Protection Areas). Its model additionally operates on a 2-week time step to match the typical irrigation time frame considered by farmers. In other words, anthropogenic considerations (governance boundaries and water use behavior) influenced the representation more so than biophysical concerns (e.g., soil attributes).

As the quantities of interest were to be predicted primarily at the catchment level, it may have been possible to aggregate some representations to a coarser level without compromising the modeled outcomes. Toward one extreme, the catchment could be represented as a single spatial zone in the farm model rather than the adopted 12 zones. On the other end, the ecological indicator model provides a long-term indication of the average suitability of streamflow for ecological purposes (e.g., averaged value over decades). Expanding scale considerations to holistically capture the temporal dynamics, its influence on the constituent systems, and how these may adapt and evolve (e.g., adaptive management of stochastic environmental flow) may influence modeled outcomes (Horne et al., 2019; John et al., 2020). Further research is necessary to determine whether increased or decreased detail is in fact appropriate for the context in which the systems are represented. A move toward a finer level of detail than that chosen, however, would require more data at the farm and field level (e.g., long-term groundwater pumping and irrigation usage) that were not available.

In an ideal setting with more time and resources, one would undertake some analysis of the possible alternative scale assumptions to explore their effects on model outputs. In this way, one could decide on the trade-offs among different scale choices regarding improved model performance versus resources required to implement them. At the very least, for transparency, the scale choices would be documented, and the ones selected for the modeling justified with a narrative that captures the decision context, the decision, and the known implications and consequences of those decisions. There are many "good practices" for documentation in both software and model development. Software practices include the Architectural Decision Records (Emery and Hilliard, 2008; Zdun et al., 2014), which advocate storing such documentation alongside code in version control. Likewise, the TRACE documentation framework suggests keeping "computational notebooks" in version control as a complement to traditional "pen-and-paper" notebooks with similar aims of documenting decisions made throughout the modeling (Ayllón et al., 2021).

Processes and phenomena couched in ambiguous or disciplinary-specific (or context-specific) terms may drive misconceptualizations of the constituent models. For example, surface and farm models both applied separate representations of "effective rainfall." Although the surface water model provides a physically based estimation of effective rainfall at a subcatchment level (see Croke and Jakeman, 2004; Ivkovic et al., 2014), the farm model applies a soil moisture accounting method that is recommended to farmers in the region for each of the 12 farming zones represented (Iwanaga et al., 2020). The moisture accounting approach informs irrigation schedules, helping farmers determine the timing and volume of irrigation, but is not a physically based estimation, to the surprise of some. Considerations around the scale of interdisciplinary communication are explored in Section 2.2.3.3.

Often in SoS modeling, the appropriate level of detail is not readily apparent, particularly during the earlier modeling phases when model development tends to focus on higher level considerations. Choices of scale are often framed by one's disciplinary focus, and individual preferences may result in decisions that lead down unintended pathways (Lahtinen et al., 2017). Modeled scales, and their most appropriate level of representation, are often not readily apparent and could be construed to be somewhat arbitrary, but not senseless, for example, when being constrained by "real-world" considerations. Insufficient consideration of the interdisciplinary aspect and challenges in cross-disciplinary communication may then have implications in the testing, evaluation, and application of the model (i.e., different paths are taken, as in Lahtinen et al., 2017), particularly in the (disaggregated) model development phases (revisited in Sections 2.2.3.4, 2.2.3.5, and 2.2.3.6).

2.2.3.2. Scale of participation

A catchment-wide survey of farmers, a series of workshops with local experts, and targeted engagement with ecologists and those representing recreational interests were among the participatory processes used to collect expert knowledge and perspectives. Furthermore, scenarios of interest were identified and co-developed through stakeholder engagement. In effect, system experts and stakeholders act as representatives of the systems under consideration including the issues and concerns that are most pertinent with respect to the modeling. The participatory process aided in constraining the overarching scenarios to those that were deemed both technically plausible and socially acceptable regarding agricultural water use (Ticehurst and Curtis, 2016, 2017).

Aside from the usual budgetary considerations (of time, money, and personnel), timing was a crucial factor in terms of the social (local stakeholder) engagement process. Not all system experts and stakeholders could be expected to attend face-to-face meetings due to timing and scheduling conflicts and the limited resourcing available. Those involved ultimately had the available time, inclination, and goodwill to participate in the time frame selected and required by researchers. This is also true in the context of writing this reflexive account as not all involved in the original case study could contribute (as noted in the Introduction).

A strong focus on the agricultural system and related water management (albeit underpinned by surface and groundwater modeling) is therefore evident in the model conceptualization as most stakeholders were linked to the agricultural and water sector. A consequence is that the model does not consider certain sociocultural values such as those held by local indigenous peoples. Potential adaptive management processes wherein water use policies change in response to improving or deteriorating ecological flow suitability were also not considered to be in scope (see description of the ecology model in Section 2.2.3.1).

Although not an active or conscious decision, the consequential filtering of participants in this manner may have introduced a self-selection bias in the sample of local stakeholders that took part in discussions. Commensurate with the specified scope—one of investigating and discussing water management and policy changes under uncertainty—future work building on this case study will likely feature a greater emphasis on the social dynamics. Incorporating reflexivity as part of the modeling can aid in managing the scale of participation and recognizing when/where the bounds may not suit objectives. In the grander scheme of things, however, enabling such work requires that commensurate funding be available to enable greater levels of participation (Iwanaga et al., 2021b) and to capture lessons learnt through reflexivity (Montana et al., 2020).

2.2.3.3. Scale of interdisciplinarity and communication

Interdisciplinary work at the heart of SoS modeling comes with unique challenges not found in single-system contexts. Many of these are detailed by Iwanaga et al. (2021b), but key to the discussion here is that in SoS, there are several sectors and disciplines involved with associated systems and models being concurrently developed and ultimately integrated. Changes to one model, because of new information or simply because of continual improvements, may necessitate changes to another model. A continual challenge throughout the project lifecycle was effectively scaling communication and participation to an appropriate level to facilitate a deeper understanding of the SES being modeled. Modelers self-organized into subteams to accomplish goals but were, for the most part, focused on their sectoral concerns. Separate and mismatched conceptualizations of the modeling arose throughout the modeling cycle, in part due to this partitioning.

Members of the team can take the role of a mediator, resolving or otherwise addressing inconsistencies and mismatches. Methodological conflict can be addressed at the technical level via model interfaces, which translate one conceptualization to another. In the CIM, for example, lumped 2-weekly farm water extractions were translated into daily averages for the ground and surface water models. Mediators may also handle task-related and interpersonal conflict (De Dreu, 2008) but may only be effective in cases where the role is assumed by someone with sufficient standing within the team and/or a cooperative team culture exists (De Dreu, 2008; Gren and Lenberg, 2018; Hidalgo, 2019).

Certainly, those managing self-organizing teams can guide interdisciplinary communication by holding regular meetings or team bonding activities (as suggested in Zare et al., 2021). Prior research suggests goal interdependence—where the success of one is contingent on the success of another—can improve team performance by setting the stage for effective collaboration (Knight et al., 2001; Tjosvold and Yu, 2004; Lee et al., 2015), particularly where flexibility and rapid response to complex and emergent issues are important (cf. Hansen et al., 2020). Effectiveness of such management strategies is likely to be highly dependent on team context, however. Depending on the larger cultural context, it may be preferable to allow (or guide) team cultures to evolve organically without direct intervention on the frequency and scale of team interactions (e.g., by mandate from management).

In the case of the CIM, each system of interest had different—but at times overlapping—concerns and issues (with some examples provided in Section 2.2.3.1). Close coordination between collaborators was needed to avoid conceptual mismatches in the models and their coupling, given the variety of scales involved and the separate, but interdependent, development paths for each model. Maintaining a high frequency of face-to-face meetings between team members was problematic because of the geographic spread of participants and financial constraints limiting travel, with the default mechanisms being emails and phone calls between individuals and within subgroups. In retrospect, more regular virtual meetings with the whole team may have helped in the longer run, particularly around technical scaling issues.

It is now seen by the team that the use of technologies and practices available to ease the burden of maintaining communication and documentation of decisions would be valuable (Zare et al., 2021). Certainly, there was a preference toward established, often disciplinary-specific, workflows rather than approaches that are perhaps more suitable for the interdisciplinary SoS context wherein team members are also geographically dispersed. For example, most modelers involved in writing code did not actively use version control, making difficult the review and dissemination of code, changing code, and documenting the reasons behind those changes. Code was instead often shared via email. Given the evolving needs and requirements of both the modeling and interdisciplinary context, it is expected that new skills and approaches should be progressively tried and, where found applicable, incorporated into the modeling workflow (Knapen et al., 2013; Hidalgo, 2019).

Separate and mismatched conceptualizations and expectations (forewarned in Knapen et al., 2013; Kragt et al., 2013; Verweij et al., 2010) of model components arose through insufficient communication. The issues that consequently arose were challenging, not to mention time consuming, to identify and correct. To give examples, in one case, numerical values were hardcoded into a model with the expectation that they would be changed manually for every run; an approach that is inappropriate given that the exploratory approach requires hundreds to thousands of model runs. In another case, input fed in from another model was found not to affect any calculations, as the integrated context was not considered.

It is worth noting that commonly suggested solutions to the above, such as adopting "advanced" communication platforms or increasing the frequency of communication, are tools and strategies that can help maintain existing interdisciplinary foundations (to paraphrase Heffernan, 2011). Care should be taken as use of such communication technologies should not be conflated with, nor a replacement for, interdisciplinarity itself. Recent research suggests continual monitoring, regulation, and a collaborative team culture are ideal, lest discrepancies affect overall team efficacy and performance (Driskell et al., 2020). Supporting lines of evidence show that a level of empathy and receptiveness to the experiences and knowledge outside of one's own ("social intelligence" in Woolley and Malone, 2011) is also needed to effectively leverage the diverse abilities found within interdisciplinary teams (Thomas, 2012; Thomas and McDonagh, 2013). This suggests that it is the culture of empathetic open-mindedness, inclusivity, and a motivation to achieve team goals that likely drives communication and the cross-pollination of interdisciplinary ideas, more so than the method and scale of communication.

2.2.3.4. Computational scalability

The computational approach is a pertinent scale consideration, especially when uncertainty primarily involves running many scenarios. In this respect, the computational scalability of the CIM became a concern to manage, mainly due to the combined runtime of the constituent models and overhead associated with their interactions. A major decision taken was to run the SoS model on a 5-km square grid rather than the initially chosen 1-km grid. Even then, a single run of the CIM could take 30 min or more, with initial implementations prior to optimizations exceeding an hour. Runtime was not an obvious issue during the disaggregated development of constituent models, even when partially integrated, especially early in the development process when the full scale and number of interactions was neither apparent nor known.

One technical barrier to increased computational performance was the use of files as an intermediary format to interoperate between models. This decision was somewhat imposed rather than selected due to the use of legacy models. Using the MODFLOW implementation for the groundwater model component as an example, computer memory (i.e., RAM), was far more limited and expensive at the time of MODFLOW's development in the 1970s (McDonald and Harbaugh, 2003). Consequently, intermediate results and parameter values between time steps (for the purpose of the CIM) were required to be written out to files rather than kept in memory. Although this process was automated through the FloPy package (Bakker et al., 2016), the comparatively high cost of file read/write activity was unavoidable and constrained the possible avenues for optimizing runtime performance. The issue was sidestepped by using a high-performance (at least at the time of writing) workstation with 32 cores, running thousands of simulations over a period of days to obtain results. This, however, is not ideal and may not be a viable solution for many.

Use of Python itself became an issue as the number, and complexity, of the models that were coupled increased. Python cannot achieve the same level of computational performance as lower-level languages (e.g., Julia, C, Fortran). The same is true for any high-level dynamic and interpreted programming language. Under usual circumstances, this is not a big issue as Python is used to leverage libraries and methods written in lower-level languages (see, e.g., NumPy; Harris et al., 2020), or otherwise "slow" parts of a Python program can be abstracted away into a lower-level language (usually Cython or C). Both strategies were taken with the farm model to improve computational performance. In the case of the CIM, Python handled the interoperation of data between models and so computational performance could not be improved without significant overhaul of the design and structure of the interfacing code, which was not possible in the available time.

As noted earlier (in Section 2.2.2.1), Python was selected for its common use as a "glue" language in the expectation that a variety of languages and approaches would be adopted by the team. It is also well-regarded as a platform for rapid prototyping. In future, an alternative language that is as flexible as Python but is more efficient computationally could well be sought as a replacement. High-performance integration necessitates a high-performance language. The Julia language (Bezanson et al., 2017), a recent addition to the scientific programming landscape, is one promising avenue in this regard.

2.2.3.5. Testing and evaluation

One salient issue that arose in the development of the CIM was the difficulty in assessing the behavior and performance of the integrated model. Calibration of models all together throughout their development was not possible as each model component was at a separate stage in the model cycle. It is acknowledged that models that are calibrated separately may exhibit unexpected behavior when integrated. Model behavior, both in the integrated and *dis*integrated context, was therefore evaluated against available observations and through stakeholder engagement.

Additional concerns revolved around uncertainties that will propagate and compound. Conceptual (or hypothesis) testing was one approach applied to address such concerns. This testing approach involved the identification of questions with a known range of acceptable answers and the subsequent testing of these against the model. The conceptual testing approach is adjustable to the available data and is especially useful in data poor contexts. Framing the context surrounding expected model behavior provides a high-level check of conditions, which can indicate the model is not fit for purpose and that changes are required. The greater the comprehensiveness of such tests, the higher the confidence that the integrated model is fit for purpose (Davidson-Pilon, 2016).

One form of conceptual testing applied was property-based sensitivity analysis. The property-based approach attempts to falsify the conceptual integrity of the integrated model by the sensitivity of model parameters within a restricted area of parameter space (Iwanaga et al., 2021a). Unexpected sensitivity results (e.g., too high, too low, or no sensitivity) then indicate an issue with the model implementation or integration, such as the inadvertent absence of model coupling. Failure of a model to conform to expected/known behavior can then falsify the assumption that the model is functioning correctly or alert to a change of context that invalidates previous understanding of the model (Claessen and Hughes, 2000). Failure of a test then avoids the computational expense of conducting a larger scale global analyses, which, due to the presence of errors, would return misleading and unreliable results.

2.2.3.6. Complexity and model uncertainty

A central challenge in the development of the CIM was determining an appropriate level of complexity while also considering its influence on (model) uncertainty. Complexity of the CIM arose from the variety of workflows, terminologies, expected spatial/temporal scales, and requirements both to individual constituent models and those pertaining to the SoS model and context. Modeler experience (and thus preferences) and available data informed several considerations throughout the modeling process.

As an example, the ground and surface water models were implemented through modifications of existing models, a decision based on prior modeler experience. These were MODFLOW-NWT with FloPy (Bakker et al., 2016) and IHACRES_GW (Ivkovic et al., 2014), respectively. Modified implementations of the groundwater model were additionally applied for other studies that were occurring concurrently (e.g., Partington et al., 2020). Available climate data were at a 5-km grid resolution, which was then the minimum granularity possible, without using interpolation, for the operation of the groundwater model. These models provided inputs for the policy, farm, and ecological indicator model with data upscaled or downscaled as appropriate for their respective purposes (see Appendix A).

The number of models involved and their structure, parameters, resolution/granularity, and data (and sources of data) were all sources of complexity. Increased complexity through the inclusion of additional systems, their interactions, and computational infrastructure generally results in compounding uncertainty (Dunford et al., 2015). This is the uncertainty that arises from the interactions between constituent models with the possibility of each interaction introducing, and propagating, some error (Refsgaard et al., 2007; Dunford et al., 2015). The error propagated may differ depending on what computational platform is in use (Iwanaga et al., 2020).

Additional model complexity allows for further investigation into the possible sources of uncertainty to be considered. Reduction of model complexity and uncertainty is often conflated with reducing its parameterization (or dimensionality), which facilitates the apportioning of parameter uncertainty to a smaller number of (considered) uncertainty sources. Reducing complexity via constraining the number of parameters does not, however, reduce uncertainty in the sense that the effect of random influences or incomplete knowledge is reduced (aleatory or epistemic uncertainty, respectively, as defined in Beven, 2009). On the other hand, model parameterization can be reduced where sensitivity and/or other analysis show that quantities of predictive interest are not influenced by certain choices. These sources of uncertainty can be explicitly documented following processes and considerations as described in Refsgaard et al. (2006, 2007), van der Sluijs (2007), and Reichert (2020).

The decision to adopt established disciplinary-specific models (e.g., MODFLOW-NWT) did quicken model development compared to starting from scratch but introduced additional complexity and considerations. For one, the MODFLOW-based model was to serve multiple purposes (across multiple studies), and so infrastructure to support the generic application and data processing was developed. Use of MODFLOW in this context is one example of a constituent model that is amenable to the overarching modeling purpose, but not necessarily complementary to it. Other constituent models of the CIM required indications of average depth to groundwater for both general and specific locations, whereas MODFLOW operates on a grid-cell (or mesh). Given MODFLOW's computational expense and additional complexity involved, it may have been worthwhile to develop a bespoke model specific to the Campaspe context of lesser complexity. Both approaches are arguably acceptable.

The question then is what level of complexity is warranted for the purpose and context of the model, recognizing constraints due to resources and legacy issues. In the context of the CIM, different scenarios to be explored required different model structures and formulations. Constituent models that could generically represent system behavior across the range of scenarios were considered a necessity. This contrasts with the development of several models specialized for each scenario context, for example, separate models for wet climate conditions, enactment of conjunctive water use policies, and so on. Considerations external to the SoS modeling exercise, as well as prior modeler experience, were additional factors that influenced the choice of constituent models, their implementation, and the process of modeling. Choice of preexisting models arguably allowed models to be developed more quickly, but at the cost of adding model complexity.

A point of interest here is that such considerations regarding the model complexity and uncertainty and their effect on quantities of interest cannot be known in advance, at least not without significant experience with the specific set of constituent models that make up the SoS model. In the context of model development, changes to constituent models invariably happen, which may sufficiently change the context of their application. Prematurely attempting to reduce model complexity and uncertainty before the full context is known (e.g., prior to model integration) is therefore inadvisable (as alluded to in Section 2.2.3.3).

3. Lessons learnt

We conclude our reflexive exercise on two SoS case studies with a synthesis of lessons across five fundamental themes elicited through reflexive self-analysis and discussions between and across the teams involved and supported by corroborating experiences drawn from existing literature. We, at least, would take these lessons forward and incorporate into future SES modeling activities. Although these lessons are also somewhat applicable to single system modeling, we believe they become especially important in the interdisciplinary SoS modeling context. It is acknowledged again here that although efforts toward discussions with team members were made, not all were able to contribute to the reflexive accounts presented. Certainly, availability and the necessary time commitment placed a limit on the scale of participation (as in Section 2.2.3.2).

3.1. Foster constant collaborative learning and reflection

The two case studies detailed in this article both featured a wide variety of disciplinary experts working together. One challenge is a risk that interdisciplinarity can be eroded as researchers gravitate toward the systems that they are familiar with. In the GPSCA case study, even though team members may have initially viewed the problem at hand through different disciplinary lenses, team members shared certain fundamental concepts. In the Campaspe case study, some conceptual mismatches arose that led to problematic issues in the integration of models, lengthening the development/modeling cycle. In our experience, the most efficient way to move an interdisciplinary conversation forward is to look backward in search of those shared concepts (Banerjee et al., 2019). Once we have found common ground, we can move the conversation forward along diverse paths under the guidance of experts who then can explain where they are leading us and why via timely additions of new concepts to our common knowledge base. Common roots were found in the concepts of system dynamics (e.g., Forrester, 1961, is a seminal work in industrial dynamics and is well-known to systems ecologists) and general systems theory. An open attitude and commitment to continual learning, both individually and as a group, is necessary for these guided paths between disciplinary domains to appear (empathetic horizons in Thomas and McDonagh, 2013) and break down disciplinary barriers (MacLeod and Nagatsu, 2018). At the very least, shared concepts avoid potential mismatches in modeler understanding.

Communication among interdisciplinary team members is crucial toward the development of a cohesive systems representation, and its importance cannot be understated. One strategy is to adopt documentation practices to ensure the existence of a collective, and cohesive, body of knowledge (Cockburn and Highsmith, 2001; Kragt et al., 2013). Specific to scale choices, the level of shared understanding and other major considerations could be explicitly catalogued in a "core" table. This table would detail the spatial and temporal scales (Koo et al., 2020), knowledge sources (Kragt et al., 2013), expected computational requirements, major uncertainty sources (Refsgaard et al., 2007; van der Sluijs, 2007; Reichert, 2020), the relevant system(s) affected, and the modeling process (Hutton et al., 2016; Ayllón et al., 2021).

The ODD protocol (Grimm et al., 2006, 2010) was used to capture these considerations in the GPSCA study, adoption of which mandates that pertinent aspects of scale and their representations are documented. The common team goals and the minimum skills/knowledge needed to achieve those goals (e.g., specific expertise in aspects of software and model development) could advantageously be made explicit as part of this process as well. Moreover, such a table is recommended here to be continually updated to consider new information and lessons learnt throughout the modeling cycle.

Others have suggested increasing the number of meetings on the progress of the modeling and to incorporate reflexive evaluation of the team (Preston et al., 2015; Dongen et al., 2018; Delice et al., 2019; Gool et al., 2019). Increased frequency and number of meetings (whether face-to-face or virtual) in effect raises the minimum number of interactions between team members so that knowledge sharing can occur. Contextual examples of how these may be helpful with regard to teams are discussed elsewhere (see Kragt et al., 2013; Cockerill et al., 2019; Zare et al., 2021); however, support for reflexive activities must be available at the organizational level (Salas et al., 2018).

What is perhaps more important than meetings, however, is a team (and organizational) *culture* that allows for empathetic and inclusive communication to occur. Team members may speak different languages or at least adopt heavy disciplinary accents. Preferring one language or dialect at the expense of a "shared language" (Thomas and McDonagh, 2013) could lead to a disregard of relevant knowledge no matter the number, length, format or medium of meetings, or how expansive the documentation (as alluded in Section 2.2.3.3). An overreliance on technological solutions to communication without acknowledging the role of team and organizational culture may lead to more, rather than fewer, misunderstandings (cf. Andres, 2012; Benishek and Lazzara, 2019).

In addition to the reflexive monitoring and evaluation of team processes (as in Driskell et al., 2020; Zare et al., 2021), we recommend that such processes additionally account for the culture that underpins knowledge sharing and communication. Ignoring the role of team and organizational culture risks naturalizing the intuitions of its most privileged members (cf. James, 2014). An open attitude and commitment to continual and collaborative learning, both individually and as a group, is necessary for disciplinary barriers to be broken down and perspectives to be embraced (Woolley and Malone, 2011; Thomas and McDonagh, 2013; MacLeod and Nagatsu, 2018). In essence, teams would ideally culturally evolve throughout the modeling cycle toward more effective models of (inter-disciplinary) cooperation (cf. Wilson and Wilson, 2007).

3.2. Document the rationale and reasons for scale choices

Debates about appropriate scales at which to represent structures and processes in multidisciplinary models should pervade discussions among modeling team members, particularly during conceptual model formulation and initial attempts to quantify linkages among model components. Most commonly, however, we begin model formulation with preconceived notions about the appropriate scales with which to represent the structures and processes in those parts of the system with which we are familiar, framed by workflows with which we are accustomed to. These preconceived notions typically are based on the way we have found most useful to think about such structures and processes in the past. Thus, the conceptualizations are coherent from a disciplinary perspective, but the cohesion breaks down when encountering other disciplines.

Our perceived usefulness of system representations is biased by our disciplinary training and experience (Huutoniemi et al., 2010). Such preconceived notions may blind us to alternate, yet still valid, representations or otherwise cause their dismissal as being of little use or simply incorrect. For example, the choice of a daily time step in the GPSCA study was informed by a shared familiarity with daily weather reports and the concept of degree-days of development of plants and insects. A 2-day time step may have been considered, thus cutting computing time in half, arguably without sacrificing usefulness of model output to end users. But a 2-day time step never crossed our minds. With the Campaspe case study, the primary focus on water-related agricultural concerns is partly a result of the level of engagement with agricultural experts (see Section 2.2.3.2), but also that the agency requirements for assessing the instream and riparian ecological impacts were quite modest. Consequently, the possibility of representing adaptive management processes of ecological issues was not actively considered (described in Section 2.2.3.1).

In building a shared understanding to develop a cohesive and complete treatment of scale, it may be more productive to agree to disagree on certain scaling issues that are particularly problematic during conceptual model formulation. Issues that are virtually impossible to resolve conceptually were almost always, in the case of the GPSCA study, clarified via quantification of the factors involved. The issues were clarified in the sense that differences in model output resulting from the use of different scales are made precise. Another reason to move on is that scale transitions that seem easy to accomplish when described in narrative form may be surprisingly difficult to accomplish computationally and which may require modifications that obviate the initially identified scale problems.

The choices made regarding scale were therefore influenced by the people involved and of course their perspectives and judgments. A "perfect" model is not possible, so we choose scales, which we believe best represent the system given "real-world" constraints. These choices are a series of subjective decisions involving consideration of model objectives and available information and resources at the time. A different group of people may arrive at a completely different, and perhaps equally plausible, valuable, and useful, model. The considerations and choices in the treatment of scale should be documented and made transparent for this reason. Such documentation allows researchers external to the process (and their future selves) to better understand the sociotechnical context in which the modeling decisions were made, the reasoning behind the decisions, and any implications or consequences from those decisions. Thus, documentation of the process helps illuminate model limitations and uncertainty (Refsgaard et al., 2006; van der Sluijs, 2007; Reichert, 2020).

We offer a final comment regarding the paucity of documentation available describing the debates preceding final SoS scaling decisions. For example, the ODD protocol, which is widely used to document agent-based models in ecology and which we used to document the GPSCA model, begins with a statement of model purpose followed by a second section that defines model entities (agents), state variables (attributes of agents), and (temporal and spatial) scales. Although this second section requires a justification of the final scale choices for each model component, it does not require documentation of the pros and cons of the alternative scales that were debated over time. Thus, a rich source of information defining the larger context of the modeling decisions, which would be particularly useful when contemplating reuse of the model, often is lost.

3.3. Acknowledge that causality is defined subjectively

When we described process representation in our GPSCA study (Section 2.1.2.2), we referred to the concept of a continuum of levels of perceived causality, of "subject-matter interpretability," extending in a theoretical sense from purely phenomenological/correlative to entirely mechanistic/explanatory. In practice, how different people perceive the representation of any given process in an SoS model will almost surely differ. In terms of model credibility, the important point is that all stakeholders, and here we include members of the modeling team as well as end-users of the model, perceive that the model behavior of most interest to them results from processes represented at an acceptable level of causality, at an acceptable level of subject-matter interpretability. Of overriding importance is that end users can explain, and hence understand, model output in cause–effect terms meaningful to them. But it also is important that members of the modeling team perceive the representations of processes in their areas of expertise as scientifically credible, given the objectives of the integrated SoS model. The cause–effect relationships responsible for output of the integrated model may be explained acceptably to end users in highly aggregated terms, whereas subject-matter specialists may require relatively detailed representations of some modeled processes in order for them to acknowledge those representations as causal.

Debates related to scale decisions in integrated SoS modeling are inextricably related to perceptions of causality. Scale decisions include not only those associated with defining temporal and spatial scales per se but also decisions associated with identifying which components and processes in the real system to include in the model and deciding at what level of detail to represent them. In our GPSCA case study, such debates arose regarding the level of detail with which to represent processes related to the aphid life cycle and the phenological development of sorghum. As described in Section 2.1.2.2, the final decision, which resulted from a lively debate among modeling team members, was to represent these processes as a function of environmental temperature modified by aphid density. Our meteorologist/aeroecologist would have been satisfied with a "causal" representation of aphid population dynamics that represented population density as a function of number of days since initial infestation and emigration as a function of density. Such a representation was perceived as unacceptably phenomenological by our entomologist. Our entomologist initially proposed a more mechanistic representation of the aphid life cycle, which included, among other things, mortality due to natural enemies (predators and parasites). However, in view of the site specificity of such relationships and the fact that the purpose of the integrated SoS model was to simulate areawide spatiotemporal patterns of aphid infestations, our entomologist agreed to a simpler "causal" representation of the aphid life cycle.

As mentioned in Section 2.2.3.1, there were several approaches to represent the spatial areas for the various models in the Campaspe case study. Each were arguably plausible, and objections could be raised depending on modeler perspectives and understanding of the modeling context. Here, we remind modelers that representing greater detail may not be appropriate given the model purpose and context. The "bigger picture" should be kept in mind.

The lesson learnt is that it would serve modeling teams well if their members explicitly acknowledged the subjective nature of their perception of causality at the very beginning of the modeling process. A discussion focused on the concept of a continuum of levels of perceived causality would be time well spent. The initial response to such a discussion most likely would be "everyone already knows that," which probably is true enough if viewed as an abstract concept. But based on our experience, we are quite sure that if early discussions among modeling team members were documented and reexamined, it would be obvious that the subjective nature of defining causality is seldom recognized in practice.

3.4. Embrace change and reflect throughout the iterative modeling cycle

The modeling process is commonly described as undergoing a "cycle" of iterations of a set of (concurrent) phases and steps. Although the number of steps and activities conducted may differ depending on purpose and conceptualization of the cycle (Boehm, 1986; Jakeman et al., 2006; Pianosi et al., 2016; Badham et al., 2019; Arnold et al., 2020; Zare et al., 2021), each step is intended to be revisited as often as needed to incorporate newly discovered or available knowledge, or ideas generated on deep reflection, as "[t]he first model is rarely the best model" (Sterling et al., 2019). It may at times be necessary to abandon an iteration and start over.

Arguably, recognizing and embracing the need for change is fundamental to the flexibility that iterative approaches afford (Dingsøyr et al., 2012; Strode et al., 2012). In the SoS context, the modeling process may have to be restarted due to discovery or incorporation of new knowledge *for another constituent system*, necessitating changes to one's own constituent model or even the modeling process. A shift in scales may be a (pragmatic) necessity to accommodate the integration of constituent models and such a decision may be governed, or have implications toward, data availability/requirement, computational capacity, and model purpose.

Change is inevitable due to the complexity of the systems being studied and the speed at which new information may come to light. Where team members are more accustomed to single-system investigations, a *cultural* shift in thinking may be required to enable flexible response to the (continuous) adjustment of scale, in all its forms. New information may necessitate skills to be acquired or adapted to an unfamiliar modeling context (Knapen et al., 2013; Voznesenskaya et al., 2019). As noted in Section 3.1, being overly tied to a single disciplinary perspective results in an inflexible system conceptualization that is resistant to "new" knowledge or perspectives. The adoption of new practices, technologies, and workflows more amenable to the new modeling context is therefore restricted and hampers team productivity (Cockburn and Highsmith, 2001; Hoda et al., 2013). The lack of version control of model code and data and the consequent effect in the development of the CIM was given as an example in Section 2.2.3.3. Change should be embraced for the lessons learnt to be effectively carried over between iterations and for knowledge to be cross-pollinated between team members (Knight et al., 2001; Lee et al., 2015).

3.5. Regularly test the integration

The reality of iterative development means that (1) constituent models may be of varying complexity and developed against different schedules, (2) changes made in one model may necessitate changes in another, (3) the necessary computational requirements and available computational infrastructure may preclude the possibility of calibrating all models at once, and (4) issues may only become a highlighted concern in the integrated context as the implications of the scale and volume of interactions may not be apparent until all models are coupled.

A somewhat naive view is that any topically relevant sectoral model can be coupled and applied to represent an SoS. This may be true at a technical level but without regard for its conceptual, and contextual, appropriateness, the resulting model is likely to be unwieldy, overly complex, and unsuited for a given purpose (e.g., Voinov and Shugart, 2013). In addressing water resource management problems, for example, Croke et al. (2014) argue that hydrological model choice requires engagement with appropriate concepts, model structures, scales of analyses, performance evaluation, and communication. Again, such issues may not be evident until the scale of the modeling becomes sufficiently expansive. Thus, the relevance of any constituent model to the integrated model's purpose and the propagation of uncertainty needs serious evaluation.

Specific to model coupling, future work could investigate a typology of design elements, which make models more amenable for their use in SoS modeling contexts and classify system models along those lines. In the short-to-medium term, strategies and plans to address or mitigate the impact of a constituent model that turns out to be not wholly suited to the SoS modeling context, such as when the scales of the problem involved increase could be explored. In the case of the CIM, computational performance became a concern as the scale of the modeling increased. One approach would have been to develop a model specifically for the integrated context, as opposed to the (continued) use of a legacy model. In the end, the issue was sidestepped by leveraging high-performance computing infrastructure. Ideally, such considerations would be considered and planned for early in the modeling process.

Methodologically, conceptual-or-hypothesis testing is one (but not the only) approach that may be applied to address concerns around the structure of the SoS model (Wilson et al., 2017; Iwanaga et al., 2020). Such testing approaches involve the identification of questions with a known range of acceptable answers that the SoS model can produce. The greater the number of such tests that can cover the range of possible realities being simulated by the model, the more confident modelers can be that the integrated model is functioning correctly, both technically and conceptually. Property-based sensitivity analysis is one approach (of many) leveraged in the development of the CIM to alert modelers of technical and conceptual issues in model integration (Iwanaga et al., 2021a). Continual testing and integration throughout the modeling process could then highlight context change (e.g., cases wherein previous understandings are falsified) and facilitate understanding of model structure and behavior (Iwanaga et al., 2021a).

In this manner, conceptual tests frame the context for incorrect model behavior. Frequent integration and testing, even at this highly aggregated level, is likely to highlight conceptual mismatches between the knowledge of disciplinary experts and model implementations. Testing of the models and their integration throughout the development cycle then plays an important role in ensuring issues are identified earlier in development (Warren, 2014). Earlier correction of issues helps to avoid "wasteful" model runs and quickens the pace through the modeling cycle. It would be beneficial if all modelers involved strive to enable repeated, and frequent, integration and testing.

Appendix A

Table A1. Individual systems represented in the Campaspe integrated model and their spatial, temporal, and data aspects. DOI: https://doi.org/10.1525/elementa.2020.00182.t2

Constituent System	Spatial	Temporal	Metrics/Data
Climate	5 km grid (0.05°, interpolated) matching the groundwater area	Daily time step. Available data constrained the time frames considered	Data represented differing levels of aridity ranging from extreme dry to "wet" over a 30-year time frame. Data sourced via Climate Change in Australia (CSIRO, 2020)
Groundwater, implemented with MODFLOW-NWT with FloPy interface (Bakker et al., 2016)	5 km grid, seven layers of variable thickness based on hydrogeologic units Higher spatial resolutions were impractical due to the long runtime of MODFLOW-NWT Largest spatial extent, extending further west than other models. Covers 4,896 km^2. Assumes irrigation events are uniformly applied across farm zone areas	Daily time step Assumes irrigation input from the farm model is to be uniformly disaggregated across 14 days	Estimates distance to water table, which influences farm groundwater pumping costs (farmer decisions) and groundwater allocations (policy) Provides estimations of surface–groundwater exchange along the river
Surface water, implemented with IHACRES_GW (Ivkovic et al., 2014)	Lumped, node-based routing model. Nodes represent subcatchments. Covers 3,518 km^2 Extends further south compared to the groundwater model to estimate inflows to the dam Assumes irrigation events are uniformly applied across farm zone areas	Daily time step Assumes irrigation input from the farm model is to be uniformly disaggregated across 14 days	Calculates dam levels, influencing water allocations for both environmental and agricultural users and perceived recreational value Stage height along the river is also provided for policy and ecology models
Farm model	Lumped, zone-based. Each zone represents farming areas of variable size. Covers 2,154 km^2	2 week time step, indicated to be the usual time frame in which irrigation decisions are made (Xie et al., 2019) Total volume of rainfall over the previous 14 days is used to determine irrigation schedule	Crop yield, farm profit estimations, water use (in ml) Incorporated data from farmer surveys
Policy model	Regional/catchment-wide	2 week time step Temporally, the model operates on a 14-day time step matching that of the farm model. In reality, such allocations are announced every 6 weeks and so constitutes	Surface water allocations, determined by dam levels. Groundwater allocations determined by groundwater level at two bores (one in the south, one in the north)

(continued)

TABLE A1. *(continued)*

Constituent System	Spatial	Temporal	Metrics/Data
		a finer grain regulation of water availability	Hypothetical conjunctive use policies allow further extraction of groundwater under periods of low surface water allocations and restrict these in times where surface water allocations are met
Recreational suitability index model	Dam	Lumped—Result is the average index value for the modeled time frame	Suitability of dam for recreational use, tied to the water level in the dam (Lake Eppalock). Generally speaking, higher dam levels allow a higher level of enjoyment to be experienced by recreational users
Ecological suitability index model	River and groundwater levels	Lumped—Result is the average index value for the modeled time frame	Suitability of flow at key gauges to support ecological activities for platypus, fish, and river red gums (iconic trees found in the area). These are lumped together into a single metric

Data accessibility statement

Data sharing is not applicable to this article as the data analyzed were our reflexive experiences, reported within this article.

Acknowledgments

The U.S. National Oceanic and Atmospheric Administration Air Resources Laboratory (ARL) provided access to the HYSPLIT transport and dispersion model. The authors additionally acknowledge past contributions of colleagues and local knowledge holders toward the sugarcane aphids in Great Plains (area wide pest management) and Campaspe case studies without which this work would not have been possible. The authors finally would like to thank the two anonymous reviewers and the Associate Editor, Prof. Zavaleta (University of California, Santa Cruz), for their time and effort. This article was greatly improved as a result of their comments.

Funding

The primary author (Takuya Iwanaga) is supported through an Australian Government Research Training Program Scholarship and a top-up scholarship from the Australian National University Hilda-John Endowment Fund. Hsiao-Hsuan Wang and Tomasz E. Koralewski acknowledge partial support from the USDA Agricultural Research Service provided through the Areawide Pest Management Program, "Areawide Pest Management of the Invasive Sugarcane Aphid in Grain Sorghum," project number 3072-22000-017-07-S. The Campaspe Integrated Model was developed as part of the Murray–Darling Basin Authority's partnership with the National Centre for Groundwater Research and Training under Contract No. MD2594. John Little acknowledges support from National Science Foundation (NSF) Award EEC 1937012. This work was also supported by the National Socio-Environmental Synthesis Center under funding received from the NSF DBI-1639145.

Competing interests

There were no competing interests with regard to this article.

Author contributions

Contributed to the conception and approach: TI, H-HW.

Contributed to the reflexive analysis: TI, H-HW, TEK, WEG, AJJ.

Additional perspectives and commentary to enhance reflexive analysis: JCL.

Drafted and/or revised this article: TI, H-HW.

Additional contributions to manuscript and revisions: TEK, WEG, AJJ, JCL.

References

Andres, HP. 2012. Technology-mediated collaboration, shared mental model and task performance. *Journal of Organizational and End User Computing* **24**: 18. DOI: http://dx.doi.org/10.4018/joeuc.2012010104.

Arnold, T, Guillaume, JHA, Lahtinen, TJ, Vervoort, RW, 2020. From ad-hoc modelling to strategic infrastructure: A manifesto for model management. *Environmental Modelling & Software* **123**: 104563. DOI: http://dx.doi.org/10.1016/j.envsoft.2019.104563.

Ayllón, D, Railsback, SF, Gallagher, C, Augusiak, J, Baveco, H, Berger, U, Charles, S, Martin, R, Focks, A, Galic, N, Liu, C, van Loon, EE, Nabe-Nielsen, J, Piou, C, Polhill, JG, Preuss, TG, Radchuk, V, Schmolke, A, Stadnicka-Michalak, J, Thorbek, P, Grimm, V. 2021. Keeping modelling notebooks with TRACE: Good for you and good for environmental research and management support. *Environmental Modelling & Software* **136**. DOI: http://dx.doi.org/10.1016/j.envsoft.2020.104932.

Badham, J, Elsawah, S, Guillaume, JHA, Hamilton, SH, Hunt, RJ, Jakeman, AJ, Pierce, SA., Snow, VO, Babbar-Sebens, M, Fu, B, Gober, P, Hill, MC, Iwanaga, T, Loucks, DP, Merritt, WS, Peckham, SD, Richmond, AK, Zare, F, Ames, D, Bammer, G. 2019. Effective modeling for Integrated Water Resource Management: A guide to contextual practices by phases and steps and future opportunities. *Environmental Modelling & Software* **116**: 40–56. DOI: http://dx.doi.org/10.1016/j.envsoft.2019.02.013.

Bakker, M, Post, V, Langevin, CD, Hughes, JD, White, JT, Starn, JJ, Fienen, MN. 2016. Scripting MODFLOW model development using Python and FloPy. *Groundwater* **54**: 733–739. DOI: http://dx.doi.org/10.1111/gwat.12413.

Banerjee, P, Wang, H-H, Peterson, MJ, Grant, WE, Peterson, TR. 2019. Collaborative modeling and social learning in the context of joint forest management in east Sikkim, India. *Frontiers of Environmental Science* **7**. DOI: http://dx.doi.org/10.3389/fenvs.2019.00154.

Benishek, LE., Lazzara, EH. 2019. Teams in a new era: Some considerations and implications. *Frontiers in Psychology* **10**. DOI: http://dx.doi.org/10.3389/fpsyg.2019.01006.

Beven, KJ. 2009. *Environmental modelling: An uncertain future? An introduction to techniques for uncertainty estimation in environmental prediction*. London, UK: Routledge.

Bezanson, J, Edelman, A, Karpinski, S, Shah, VB. 2017. Julia: A fresh approach to numerical computing. *SIAM Review* **59**: 65–98. DOI: http://dx.doi.org/10.1137/141000671.

Boehm, B. 1986. A spiral model of software development and enhancement. *SIGSOFT Software Engineering Notes* **11**: 14–24. DOI: http://dx.doi.org/10.1145/12944.12948.

Bolton, G., Delderfield, R. 2018. *Reflective practice: Writing and professional development* (5th ed.). Los Angeles, CA: SAGE.

Bowling, RD, Brewer, MJ, Kerns, DL, Gordy, J, Seiter, N, Elliott, NE, Buntin, GD, Way, MO, Royer, TA, Biles, S, Maxson, E. 2016. Sugarcane Aphid (Hemiptera: Aphididae): A new pest on sorghum in North America. *Journal of Integrated Pest Management* **7**. DOI: http://dx.doi.org/10.1093/jipm/pmw011.

Brewer, MJ, Gordy, JW, Kerns, DL, Woolley, JB, Rooney, WL, Bowling, RD. 2017. Sugarcane aphid population growth, plant injury, and natural enemies on selected grain sorghum hybrids in Texas and Louisiana. *Journal of Economic Entomology* **110**: 2109–2118. DOI: http://dx.doi.org/10.1093/jee/tox204.

Brewer, MJ, Peairs, FB, Elliott, NC. 2019. Invasive cereal aphids of North America: Ecology and pest management. *Annual Review of Entomology* **64**: 73–93. DOI: http://dx.doi.org/10.1146/annurev-ento-011118-111838.

Catalano, AS, Lyons-White, J, Mills, MM, Knight, AT. 2019. Learning from published project failures in conservation. *Biological Conservation* **238**. DOI: http://dx.doi.org/10.1016/j.biocon.2019.108223.

Claessen, K, Hughes, J. 2000. *QuickCheck: A lightweight tool for random testing of Haskell programs*, in *Proceedings of the Fifth ACM SIGPLAN International Conference on Functional Programming*, ICFP '00. New York, NY: Association for Computing Machinery: 268–279. DOI: http://dx.doi.org/10.1145/351240.351266.

Cockburn, A, Highsmith, J, 2001. Agile software development, the people factor. *Computer* **34**: 131–133. DOI: http://dx.doi.org/10.1109/2.963450.

Cockerill, K, Glynn, P, Chabay, I, Farooque, M, Hämäläinen, RP, Miyamoto, B, McKay, P. 2019. Records of engagement and decision making for environmental and socio-ecological challenges. *Euro Journal on Decision Processes*. DOI: http://dx.doi.org/10.1007/s40070-019-00104-6.

Connolly, R., 2020. Why computing belongs within the social sciences. *Communications of the ACM* **63**: 54–59. DOI: http://dx.doi.org/10.1145/3383444.

Croke, BFW, Blakers, RS, El Sawah, S, Fu, B, Guillaume, JHA, Kelly, RA, Patrick, MJ, Ross, A, Ticehurst, J, Barthel, R, Jakeman, AJ. 2014 June 4–6. Marrying hydrological modelling and integrated assessment for the needs of water resource management, in *Proceedings of the International Association of Hydrological Sciences. Presented at the Proceedings of ICWRS2014—6th IAHS-EGU International Symposium on Integrated Water Resources Management*. Bologna, Italy: Copernicus GmbH: 351–356. DOI: http://dx.doi.org/10.5194/piahs-364-351-2014.

Croke, BFW, Jakeman, AJ. 2004. A catchment moisture deficit module for the IHACRES rainfall-runoff model. *Environmental Modelling & Software* **19**: 1–5. DOI: http://dx.doi.org/10.1016/j.envsoft.2003.09.001.

Crouzat, E, Arpin, I, Brunet, L, Colloff, MJ, Turkelboom, F, Lavorel, S. 2018. Researchers must be aware of their roles at the interface of ecosystem services science and policy. *Ambio* **47**: 97–105. DOI: http://dx.doi.org/10.1007/s13280-017-0939-1.

CSIRO. 2020. Climate Change in Australia. Climate Change in Australia. Available at https://climatechangeinaustralia.gov.au/en/. Accessed 24 February 2020.

Davidson, J. 1944. On the relationship between temperature and rate of development of insects at constant temperatures. *Journal of Animal Ecology* **13**: 26–38. DOI: http://dx.doi.org/10.2307/1326.

Davidson-Pilon, C. 2016. *Bayesian methods for hackers: Probabilistic programming and Bayesian inference*. New York: Addison-Wesley.

De Dreu, CKW. 2008. The virtue and vice of workplace conflict: Food for (pessimistic) thought. *Journal of Organizational Behavior* **29**: 5–18. DOI: http://dx.doi.org/10.1002/job.474.

Delice, F, Rousseau, M, Feitosa, J, 2019. Advancing teams research: What, when, and how to measure team dynamics over time. *Frontiers in Psychology*

10. DOI: http://dx.doi.org/10.3389/fpsyg.2019.01324.

Dey, R, Lewis, SC, Arblaster, JM, Abram, NJ. 2019. A review of past and projected changes in Australia's rainfall. *WIREs Climate Change* **10**: e577. DOI: http://dx.doi.org/10.1002/wcc.577.

Dingsøyr, T, Nerur, S, Balijepally, V, Moe, NB. 2012. A decade of agile methodologies: Towards explaining agile software development. *Journal of Systems and Software, Special Issue: Agile Development* **85**: 1213–1221. DOI: http://dx.doi.org/10.1016/j.jss.2012.02.033.

Driskell, JE, Salas, E, Driskell, T. 2020. Foundations of teamwork and collaboration. *American Psychologist* **73**: 334–348.

Dunford, R, Harrison, PA, Rounsevell, MDA. 2015. Exploring scenario and model uncertainty in cross-sectoral integrated assessment approaches to climate change impacts. *Climatic Change* **132**: 417–432. DOI: http://dx.doi.org/10.1007/s10584-014-1211-3.

Dysarz, T. 2018. Application of Python scripting techniques for control and automation of HEC-RAS simulations. *Water* **10**: 1382. DOI: http://dx.doi.org/10.3390/w10101382.

Eagles, D, Walker, PJ, Zalucki, MP, Durr, PA. 2013. Modelling spatio-temporal patterns of long-distance Culicoides dispersal into northern Australia. *Preventive Veterinary Medicine* **110**: 312–322. DOI: http://dx.doi.org/10.1016/j.prevetmed.2013.02.022.

Elliott, NC, Onstad, DW, Brewer, MJ. 2008. History and ecological basis for areawide pest management, in Koul, O, Cuperus, GW, Elliott, N eds., *Areawide pest management: Theory and implementation*. Wallingford, UK: CABI: 15–33.

Elsawah, S, Filatova, T, Jakeman, AJ, Kettner, AJ, Zellner, ML, Athanasiadis, IN, Hamilton, SH, Axtell, RL, Brown, DG, Gilligan, JM, Janssen, MA, Robinson, DT, Rozenberg, J, Ullah, IIT, Lade, SJ. 2020. Eight grand challenges in socio-environmental systems modeling. *Socio-Environmental Systems Modeling* **2**: 16226. DOI: http://dx.doi.org/10.18174/sesmo.2020a16226.

Emery, D, Hilliard, R. 2008. Updating IEEE 1471: Architecture Frameworks and Other Topics, in *Presented at the seventh working IEEE/IFIP conference on Software Architecture (WICSA 2008)*: 303–306. DOI: http://dx.doi.org/10.1109/WICSA.2008.32.

Evans, MR. 2012. Modelling ecological systems in a changing world. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**: 181–190. DOI: http://dx.doi.org/10.1098/rstb.2011.0172.

Finlay, L. 2002. "Outing" the researcher: The provenance, process, and practice of reflexivity. *Qualitative Health Research* **12**. DOI: http://dx.doi.org/10.1177/104973202129120052.

Forrester, JW. 1961. *Industrial dynamics*. First edition. Cambridge: MIT Press.

Freeman, RE. 2010. *Strategic management: A stakeholder approach, Reissue*. Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore: Cambridge University Press.

Gerik, T, Bean, BW, Vanderlip, R. 2003. *Sorghum growth and development, Texas FARMER Collection*. College Station, TX: Texas AgriLife Extension Service.

Giles, K., Hein, G., Peairs, F., 2008. Areawide pest management of cereal aphids in dryland wheat systems of the Great Plains, USA, in Koul, O, Cuperus, GW, Elliott, N eds., *Areawide pest management: Theory and implementation*. Wallingford, UK: CABI: 441–466.

Glynn, PD, Voinov, AA, Shapiro, CD, White, PA. 2017. From data to decisions: Processing information, biases, and beliefs for improved management of natural resources and environments. *Earth's Future* **5**: 356–378. DOI: http://dx.doi.org/10.1002/2016EF000487.

Gren, L, Lenberg, P. 2018. The importance of conflict resolution techniques in autonomous agile teams, in *Proceedings of the 19th International Conference on Agile Software Development: Companion*, XP '18. New York, NY: Association for Computing Machinery: 1–4. DOI: http://dx.doi.org/10.1145/3234152.3234185.

Grimm, V, Berger, U, Bastiansen, F, Eliassen, S, Ginot, V, Giske, J, Goss-Custard, J, Grand, T, Heinz, SK, Huse, G, Huth, A, Jepsen, JU, Jørgensen, C, Mooij, WM, Müller, B, Pe'er, G, Piou, C, Railsback, SF, Robbins, AM, Robbins, MM, Rossmanith, E, Rüger, N, Strand, E, Souissi, S, Stillman, RA, Vabø, R, Visser, U, DeAngelis, DL. 2006. A standard protocol for describing individual-based and agent-based models. *Ecological Modelling* **198**: 115–126. DOI: http://dx.doi.org/10.1016/j.ecolmodel.2006.04.023.

Grimm, V, Berger, U, DeAngelis, DL, Polhill, JG, Giske, J, Railsback, SF, 2010. The ODD protocol: A review and first update. *Ecological Modelling* **221**: 2760–2768. DOI: http://dx.doi.org/10.1016/j.ecolmodel.2010.08.019.

Grimm, V, Railsback, SF. 2005. *Individual-based modeling and ecology*. STU-Student. Princeton, NJ: Princeton University Press.

Grimm, V, Railsback, SF, Vincenot, CE, Berger, U, Gallagher, C, DeAngelis, DL, Edmonds, B, Ge, J, Giske, J, Groeneveld, J, Johnston, ASA, Milles, A, Nabe-Nielsen, J, Polhill, JG, Radchuk, V, Rohwäder, M-S, Stillman, RA, Thiele, JC, Ayllón, D. 2020. The ODD protocol for describing agent-based and other simulation models: A second update to improve clarity, replication, and structural realism. *Journal of Artificial Societies and Social Simulation* **23**: 7. DOI: http://dx.doi.org/10.18564/jasss.4259.

Hall, KL, Vogel, AL, Huang, GC, Serrano, KJ, Rice, EL, Tsakraklides, SP, Fiore, SM. 2018. The science of team science: A review of the empirical evidence and research gaps on collaboration in science. *American Psychologist* **73**: 532–548. DOI: http://dx.doi.org/10.1037/amp0000319.

Hall, KL, Vogel, AL, Stipelman, BA, Stokols, D, Morgan, G, Gehlert, S. 2012. A four-phase model of transdisciplinary team-based research: Goals, team processes, and strategies. *Translational Behavioral Medicine* **2**: 415–430. DOI: http://dx.doi.org/10.1007/s13142-012-0167-y.

Hamilton, SH, ElSawah, S, Guillaume, JHA, Jakeman, AJ, Pierce, SA. 2015. Integrated assessment and modelling: Overview and synthesis of salient dimensions. *Environmental Modelling & Software* **64**: 215–229. DOI: http://dx.doi.org/10.1016/j.envsoft.2014.12.005.

Hansen, MJ, Vaagen, H, van Oorschot, K. 2020. Team collective intelligence in dynamically complex projects—A shipbuilding case. *Project Management Journal*. DOI: http://dx.doi.org/10.1177/8756972820928695.

Harris, CR, Millman, KJ, van der Walt, SJ, Gommers, R, Virtanen, P, Cournapeau, D, Wieser, E, Taylor, J, Berg, S, Smith, NJ, Kern, R, Picus, M, Hoyer, S, van Kerkwijk, MH, Brett, M, Haldane, A, del Río, JF, Wiebe, M, Peterson, P, Gérard-Marchant, P, Sheppard, K, Reddy, T, Weckesser, W, Abbasi, H, Gohlke, C, Oliphant, TE. 2020. Array programming with NumPy. *Nature* **585**: 357–362. DOI: http://dx.doi.org/10.1038/s41586-020-2649-2.

Heffernan, M. 2011. *Willful blindness: Why we ignore the obvious*. First U.S. edition. New York, NY: Walker & Co.

Hidalgo, ES. 2019. Adapting the scrum framework for agile project management in science: Case study of a distributed research initiative. *Heliyon* **5**: e01447. DOI: http://dx.doi.org/10.1016/j.heliyon.2019.e01447.

Hinson, P. 2017. Effect of temperature on the development of sugarcane aphid, *Melanaphis sacchari*, on Sorghum [Thesis]. West Texas A&M University. Available at http://hdl.handle.net/11310/140. Accessed 2 June 2021.

Hoda, R, Noble, J, Marshall, S. 2013. Self-organizing roles on agile software development teams. *IEEE Transactions on Software Engineering* **39**: 422–444. DOI: http://dx.doi.org/10.1109/TSE.2012.30.

Horne, AC, Nathan, R, Poff, NL, Bond, NR, Webb, JA, Wang, J, John, A. 2019. Modeling flow-ecology responses in the Anthropocene: Challenges for sustainable riverine management. *BioScience* **69**: 789–799. DOI: http://dx.doi.org/10.1093/biosci/biz087.

Hutton, C, Wagener, T, Freer, J, Han, D, Duffy, C, Arheimer, B. 2016. Most computational hydrology is not reproducible, so is it really science? *Water Resources Research* **52**: 7548–7555. DOI: http://dx.doi.org/10.1002/2016WR019285.

Huutoniemi, K, Klein, JT, Bruun, H. Hukkinen, J. 2010. Analyzing interdisciplinarity: Typology and indicators. *Research Policy* **39**: 79–88. DOI: http://dx.doi.org/10.1016/j.respol.2009.09.011.

Ivkovic, KM, Croke, BFW, Kelly, RA. 2014. Overcoming the challenges of using a rainfall–runoff model to estimate the impacts of groundwater extraction on low flows in an ephemeral stream. *Hydrology Research* **45**: 58–72. DOI: http://dx.doi.org/10.2166/nh.2013.204.

Iwanaga, T, Partington, D, Ticehurst, J, Croke, BFW, Jakeman, AJ. 2020. A socio-environmental model for exploring sustainable water management futures: Participatory and collaborative modelling in the lower Campaspe catchment. *Journal of Hydrology: Regional Studies* **28**. DOI: http://dx.doi.org/10.1016/j.ejrh.2020.100669.

Iwanaga, T, Sun, X, Wang, Q, Guillaume, JHA, Croke, BFW, Rahman, J, Jakeman, AJ. 2021a. Property-based sensitivity analysis: An approach to identify model implementation and integration errors. *Environmental Modelling & Software* **139**. DOI: http://dx.doi.org/10.1016/j.envsoft.2021.105013.

Iwanaga, T, Wang, H-H, Hamilton, SH, Grimm, V, Koralewski, TE, Salado, A, Elsawah, S, Razavi, S, Yang, J, Glynn, P, Badham, J, Voinov, A, Chen, M, Grant, WE, Peterson, TR, Frank, K, Shenk, G, Barton, CM, Jakeman, AJ, Little, JC. 2021b. Socio-technical scales in socio-environmental modeling: Managing a system-of-systems modeling approach. *Environmental Modelling & Software*. DOI: http://dx.doi.org/10.1016/j.envsoft.2020.104885.

Iwanaga, T, Zare, F, Croke, B, Fu, B, Merritt, W, Partington, D, Ticehurst, J, Jakeman, A, 2018 June 13–15. Development of an integrated model for the Campaspe catchment: A tool to help improve understanding of the interaction between society, policy, farming decision, ecology, hydrology and climate, in *Proceedings of the International Association of Hydrological Sciences. Presented at the Innovative Water Resources Management—Understanding And Balancing Interactions between Humankind And Nature—8th International Water Resources Management Conference of ICWRS*. Beijing, China: Copernicus GmbH: 1–12. DOI: http://dx.doi.org/10.5194/piahs-379-1-2018.

Jakeman, AJ, Letcher, RA, Norton, JP. 2006. Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling and Software* **21**: 602–614. DOI: http://dx.doi.org/10.1016/j.envsoft.2006.01.004.

James, R. 2014. What we can learn about Philosophy's diversity problems by comparing ourselves to Music Theory. Available at https://www.its-her-factory.com/2014/10/what-we-can-learn-about-philosophys-diversity-problems-by-comparing-ourselves-to-music-theory/. Accessed 29 October 2020.

John, A, Horne, A, Nathan, R, Stewardson, M, Webb, JA, Wang, J, Poff, NL. 2020. Climate change and freshwater ecology: Hydrological and ecological methods of comparable complexity are needed to predict risk. *WIREs Climate Change*: e692. DOI: http://dx.doi.org/10.1002/wcc.692.

Jones, NA, Ross, H, Lynam, T, Perez, P, Leitch, A. 2011. Mental models: An interdisciplinary synthesis of theory and methods. *Ecology and Society* **16**(1): 46.

Kendall, M. 2013. Drought and its role in shaping water policy in Australia, in Schwabe, K, Albiac, J, Connor, JD, Hassan, RM, Meza González, L eds., *Drought in arid and semi-arid regions: A multi-disciplinary and cross-country perspective*. Dordrecht, the Netherlands: Springer: 451–467. DOI: http://dx.doi.org/10.1007/978-94-007-6636-5_26.

Knapen, R, Janssen, S, Roosenschoon, O, Verweij, P, de Winter, W, Uiterwijk, M, Wien, J-E. 2013. Evaluating OpenMI as a model integration platform across disciplines. *Environmental Modelling & Software, Thematic Issue on the Future of Integrated Modeling Science and Technology* **39**: 274–282. DOI: http://dx.doi.org/10.1016/j.envsoft.2012.06.011.

Knight, D, Durham, CC, Locke, EA. 2001. The relationship of team goals, incentives, and efficacy to strategic risk, tactical implementation, and performance. *Academy of Management Journal* **44**: 326–338. DOI: http://dx.doi.org/10.5465/3069459.

Koo, H, Iwanaga, T, Croke, BFW, Jakeman, AJ, Yang, J, Wang, H-H, Sun, X, Lü, G, Li, X, Yue, T, Yuan, W, Liu, X, Chen, M. 2020. Position paper: Sensitivity analysis of spatially distributed environmental models- a pragmatic framework for the exploration of uncertainty sources. *Environmental Modelling & Software*. DOI: http://dx.doi.org/10.1016/j.envsoft.2020.104857.

Koralewski, TE, Wang, H-H, Grant, WE, Brewer, MJ, Elliott, NC, Westbrook, JK, Szczepaniec, A, Knutson, A, Giles, KL, Michaud, JP. 2020a. Integrating models of atmospheric dispersion and crop-pest dynamics: Linking detection of local aphid infestations to forecasts of region-wide invasion of cereal crops. *Annals of the Entomological Society of America* **113**: 79–87. DOI: http://dx.doi.org/10.1093/aesa/saz047.

Koralewski, TE, Wang, H-H, Grant, WE, LaForest, JH, Brewer, MJ, Elliott, NC, Westbrook, JK. 2020b. Toward near-real-time forecasts of airborne crop pests: Aphid invasions of cereal grains in North America. *Computers and Electronics in Agriculture* **179**: 105861. DOI: http://dx.doi.org/10.1016/j.compag.2020.105861.

Koralewski, TE, Westbrook, JK, Grant, WE, Wang, H-H. 2019. Coupling general physical environmental process models with specific question-driven ecological simulation models. *Ecological Modelling* **405**: 102–105. DOI: http://dx.doi.org/10.1016/j.ecolmodel.2019.02.004.

Kragt, ME, Robson, BJ, Macleod, CJA. 2013. Modellers' roles in structuring integrative research projects. *Environmental Modelling & Software* **39**: 322–330. DOI: http://dx.doi.org/10.1016/j.envsoft.2012.06.015.

Krueger, T, Maynard, C, Carr, G, Bruns, A, Mueller, EN, Lane, S. 2016. A transdisciplinary account of water research. *WIREs Water* **3**: 369–389. DOI: http://dx.doi.org/10.1002/wat2.1132.

Lahtinen, TJ, Guillaume, JHA, Hmlinen, RP. 2017. Why pay attention to paths in the practice of environmental modelling? *Environmental Modelling and Software* **92**: 74–81. DOI: http://dx.doi.org/10.1016/j.envsoft.2017.02.019.

Lee, C, Lin, Y, Huan, H, Huang, W, Teng, H. 2015. The effects of task interdependence, team cooperation, and team conflict on job performance. *Social Behavior and Personality: An International Journal* **43**: 529–536. DOI: http://dx.doi.org/10.2224/sbp.2015.43.4.529.

Levins, R. 1966. The strategy of model building in population biology. *American Scientist* **54**: 421–430.

Little, JC, Hester, ET, Elsawah, S, Filz, GM, Sandu, A, Carey, CC, Iwanaga, T, Jakeman, AJ. 2019. A tiered, system-of-systems modeling framework for resolving complex socio-environmental policy issues. *Environmental Modelling & Software* **112**: 82–94. DOI: http://dx.doi.org/10.1016/j.envsoft.2018.11.011.

MacLeod, M, Nagatsu, M. 2018. What does interdisciplinarity look like in practice: Mapping interdisciplinarity and its limits in the environmental sciences. *Studies in History and Philosophy of Science: Part A* **67**: 74–84. DOI: http://dx.doi.org/10.1016/j.shpsa.2018.01.001.

Maier, HR, Guillaume, JHA, van Delden, H, Riddell, GA, Haasnoot, M, Kwakkel, JH. 2016. An uncertain future, deep uncertainty, scenarios, robustness and adaptation: How do they fit together? *Environmental Modelling & Software* **81**: 154–164. DOI: http://dx.doi.org/10.1016/j.envsoft.2016.03.014.

May, T, Perry, B. 2017. *Reflexivity: The essential guide* (1st ed.). Thousand Oaks, CA: SAGE.

McDonald, MG, Harbaugh, AW. 2003. The history of MODFLOW. *Groundwater* **41**: 280–283. DOI: https://doi.org/10.1111/j.1745-6584.2003.tb02591.x.

Montana, J, Elliott, L, Ryan, M, Wyborn, C. 2020. The need for improved reflexivity in conservation science. *Environmental Conservation* 1–3. DOI: http://dx.doi.org/10.1017/S0376892920000326.

Muller, E, Bednar, JA, Diesmann, M, Gewaltig, M-O, Hines, M, Davison, AP. 2015. Python in neuroscience. *Frontiers in Neuroinformatics* **9**. DOI: http://dx.doi.org/10.3389/fninf.2015.00011.

Murray–Darling Basin Authority. 2012. *Sustainable Rivers Audit 2: The ecological health of rivers in the Murray–Darling Basin at the end of the millennium drought (2008–2010).* Summary. (No. 75/12). Canberra, ACT, Australia: MDBA.

Nancarrow, SA, Booth, A, Ariss, S, Smith, T, Enderby, P, Roots, A. 2013. Ten principles of good interdisciplinary team work. *Human Resources for Health* **11**: 19. DOI: http://dx.doi.org/10.1186/1478-4491-11-19.

National Oceanographic and Atmospheric Administration. 2019. Air Resources Laboratory—EDAS 40 km Data Archive. Available at https://ready.arl.noaa.gov/edas40.php. Accessed 21 December 2020.

North Central CMA. 2014. *Campaspe river resource guide.* Huntly, UK: North Central Catchment Management Authority.

Parry, HR. 2013. Cereal aphid movement: General principles and simulation modelling. *Movement Ecology* **1**: 14. DOI: http://dx.doi.org/10.1186/2051-3933-1-14.

Partington, D, Knowling, MJ, Simmons, CT, Cook, PG, Xie, Y, Iwanaga, T, Bouchez, C, 2020. Worth of hydraulic and water chemistry observation data in terms of the reliability of surface water-groundwater exchange flux predictions under varied flow conditions. *Journal of Hydrology* **590**: 125441. DOI: http://dx.doi.org/10.1016/j.jhydrol.2020.125441.

Pianosi, F, Beven, K, Freer, J, Hall, JW, Rougier, J, Stephenson, DB, Wagener, T. 2016. Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software* **79**: 214–232. DOI: http://dx.doi.org/10.1016/j.envsoft.2016.02.008.

Poché, DM, Grant, WE, Wang, H-H. 2016. Visceral Leishmaniasis on the Indian subcontinent: Modelling the dynamic relationship between vector control schemes and vector life cycles. *PLOS Neglected Tropical Diseases* **10**: e0004868. DOI: http://dx.doi.org/10.1371/journal.pntd.0004868.

Preston, BL, Rickards, L, Fünfgeld, H, Keenan, RJ. 2015. Toward reflexive climate adaptation research. *Current Opinion in Environmental Sustainability, Open Issue* **14**: 127–135. DOI: http://dx.doi.org/10.1016/j.cosust.2015.05.002.

Refsgaard, JC, van der Sluijs, JP, Brown, J, van der Keur, P. 2006. A framework for dealing with uncertainty due to model structure error. *Advances in Water Resources* **29**: 1586–1597. DOI: http://dx.doi.org/10.1016/j.advwatres.2005.11.013.

Refsgaard, JC, van der Sluijs, JP, Højberg, AL, Vanrolleghem, PA. 2007. Uncertainty in the environmental modelling process—A framework and guidance. *Environmental Modelling & Software* **22**: 1543–1556. DOI: http://dx.doi.org/10.1016/j.envsoft.2007.02.004.

Reichert, P. 2020. Towards a comprehensive uncertainty assessment in environmental research and decision support. *Water Science and Technology* **81**: 1588–1596. DOI: http://dx.doi.org/10.2166/wst.2020.032

Roozeboom, KL, Prasad, PVV. 2019. Sorghum growth and development, in *Sorghum*. Wiley & Sons: 155–172. DOI: http://dx.doi.org/10.2134/agronmonogr58.c8.

Salas, E, Reyes, DL, McDaniel, SH. 2018. The science of teamwork: Progress, reflections, and the road ahead. *American Psychologist* **73**: 593–600. DOI: http://dx.doi.org/10.1037/amp0000334.

Saltelli, A, Benini, L, Funtowicz, S, Giampietro, M, Kaiser, M, Reinert, E, van der Sluijs, JP. 2020. The technique is never neutral. How methodological choices condition the generation of narratives for sustainability. *Environmental Science & Policy* **106**: 87–98. DOI: http://dx.doi.org/10.1016/j.envsci.2020.01.008.

Saltelli, A, Funtowicz, S. 2017. What is science's crisis really about? *Futures, Post-Normal Science in Practice* **91**: 5–11. DOI: http://dx.doi.org/10.1016/j.futures.2017.05.010.

Scheffer, M, Baveco, JM, DeAngelis, DL, Rose, KA, van Nes, EH. 1995. Super-individuals, a simple solution for modelling large populations on an individual basis. *Ecological Modelling* **80**: 161–170. DOI: http://dx.doi.org/10.1016/0304-3800(94)00055-M.

Singh, BU, Padmaja, PG, Seetharama, N. 2004. Biology and management of the sugarcane aphid, *Melanaphis sacchari* (Zehntner) (Homoptera: Aphididae), in sorghum: A review. *Crop Protection* **23**: 739–755. DOI: http://dx.doi.org/10.1016/j.cropro.2004.01.004.

Stein, AF, Draxler, RR, Rolph, GD, Stunder, BJB, Cohen, MD, Ngan, F. 2015. NOAA's HYSPLIT atmospheric transport and dispersion modeling system. *Bulletin of the American Meteorological Society* **96**: 2059–2077. DOI: http://dx.doi.org/10.1175/BAMS-D-14-00110.1.

Sterling, EJ, Zellner, M, Jenni, KE, Leong, K, Glynn, PD, BenDor, TK, Bommel, P, Hubacek, K, Jetter, AJ, Jordan, R, Olabisi, LS, Paolisso, M, Gray, S. 2019. Try, try again: Lessons learned from success and failure in participatory modeling. *Elementa: Science of the Anthropocene* **7**: 9. DOI: http://dx.doi.org/10.1525/elementa.347.

Strode, DE, Huff, SL, Hope, B, Link, S. 2012. Coordination in co-located agile software development projects. *Journal of Systems and Software, Special Issue: Agile Development* **85**: 1222–1238. DOI: http://dx.doi.org/10.1016/j.jss.2012.02.017.

Thiele, JC, Kurth, W, Grimm, V. 2014. Facilitating Parameter estimation and sensitivity analysis of agent-based models: A cookbook using NetLogo and R. *JASSS* **17**: 11.

Thomas, J. 2012. Power and illusion of self-organizing teams. Presented at the PMI® Global Congress 2012, PA: Project Management Institute, North America, Vancouver, BC, Canada. Newtown Square.

Thomas, J. McDonagh, D. 2013. Shared language: Towards more effective communication. *Australasian Medical Journal* **6**: 46–54. DOI: http://dx.doi.org/10.4066/AMJ.2013.1596.

Ticehurst, J, Curtis, A. 2016. *Improving on-farm water use efficiency amongst irrigators in the Campaspe River system and the Rochester Irrigation Area.* Wagga Wagga, Australia: EH Graham Centre for Agricultural Innovation, Charles Sturt University.

Ticehurst, J, Curtis, A. 2017. *Preliminary assessment of conjunctive use opportunities in the Murray-Darling Basin: A case study in the lower Campaspe catchment.* Canberra, Australia: Murray-Darling Basin Authority; Adelaide, Australia: The National Centre for Groundwater Research and Training.

Tjosvold, D, Yu, Z. 2004. Goal Interdependence and applying abilities for team in-role and extra-role performance in China. *Group Dynamics: Theory, Research, and Practice* **8**: 98–111. DOI: http://dx.doi.org/10.1037/1089-2699.8.2.98.

USDA-NASS. 2010. *Field crops usual planting and harvesting dates, Agricultural Handbook Number 628.* Washington, DC: USDA.

van der Sluijs, JP. 2007. Uncertainty and precaution in environmental management: Insights from the UPEM conference. *Environmental Modelling & Software, The Implications of Complexity for Integrated Resources* **22**: 590–598. DOI: http://dx.doi.org/10.1016/j.envsoft.2005.12.020.

van Dongen, MJ, van Bokhoven, MA, Goossens, WN, Daniëls, R, van der Weijden, T, Beurskens, A. 2018. Development of a customizable programme for improving interprofessional team meetings: An action research approach. *International Journal of Integrated Care* **18**: 8. DOI: http://dx.doi.org/10.5334/ijic.3076.

van Gool, FW, Bierbooms, JJ, Bongers, IM, Janssen, RT. 2019. Co-creating a program for teams to maintain and reflect on their flexibility. *The International Journal of Health Planning and Management* **34**: e1937–e1947. DOI: http://dx.doi.org/10.1002/hpm.2855.

van Rensburg, NJ. 1973. Notes on the occurrence and biology of the sorghum aphid in South Africa. *Journal of the Entomological Society of Southern Africa* **36**: 293–298.

Verweij, PJFM, Knapen, MJR, de Winter, WP, Wien, JJF, te Roller, JA, Sieber, S, Jansen, JML. 2010. An IT perspective on integrated environmental modelling: The SIAT case. *Ecological Modelling* **221:** 2167–2176. DOI: https://doi.org/10.1016/j.ecolmodel.2010.01.006.

Voinov, A, Shugart, HH. 2013. "Integronsters," integral and integrated modeling. *Environmental Modelling and Software* **39**: 149–158. DOI: http://dx.doi.org/10.1016/j.envsoft.2012.05.014.

Voznesenskaya, T, Krasnov, F, Yavorsky, R, Chesnokova, P. 2019. Modeling self-organizing teams in a research environment. *Business Informatics* **13**: 7–17. DOI: http://dx.doi.org/10.17323/1998-0663.2019.2.7.17.

Walsh, Z, Böhme, J, Wamsler, C. 2020. Towards a relational paradigm in sustainability research, practice, and education. *Ambio*. DOI: http://dx.doi.org/10.1007/s13280-020-01322-y.

Wang, H-H, Grant, WE. 2021. Reflections of two systems ecologists on modelling coupled human and natural (socio-ecological, socio-environmental) systems. *Ecological Modelling* **440**. DOI: http://dx.doi.org/10.1016/j.ecolmodel.2020.109403.

Wang, H-H, Grant, WE, Elliott, NC, Brewer, MJ, Koralewski, TE, Westbrook, JK, Alves, TM, Sword, GA. 2019. Integrated modelling of the life cycle and aeroecology of wind-borne pests in temporally-variable spatially-heterogeneous environment. *Ecological Modelling* **399**: 23–38. DOI: http://dx.doi.org/10.1016/j.ecolmodel.2019.02.014.

Wang, H-H, Grant, WE, Koralewski, TE, Elliott, NC, Brewer, MJ, Westbrook, JK. 2020a. *Where do all the aphids go? A series of thought experiments within the context of area-wide pest management*. Agricultural Systems.

Wang, H-H, Grant, WE, Teague, R. 2020b. Modeling rangelands as spatially-explicit complex adaptive systems. *Journal of Environmental Management* **269**: 110762. DOI: http://dx.doi.org/10.1016/j.jenvman.2020.110762.

Warren, K. 2014. Agile SD: Fast, Effective, Reliable. Proceedings of the 32nd International Conference of the System Dynamics Society. Presented at the Conference of the System Dynamics Society, Systems Dynamics Society; Delft, the Netherlands.

Wilensky, U. 1999. *NetLogo (and NetLogo user manual)*. Evanston, IL: Center for Connected Learning and Computer-Based Modeling, Northwestern University.

Wilson, C, Kriegler, E, van Vuuren, DP, Guivarch, C, Frame, D, Krey, V, Osborn, TJ, Schwanitz, VJ, Thompson, EL. 2017. Evaluating process-based integrated assessment models of climate change mitigation. Available at http://pure.iiasa.ac.at/id/eprint/14502/. Accessed 16 December 2019.

Wilson, DS, Wilson, EO. 2007. Rethinking the theoretical foundation of sociobiology. *The Quarterly Review of Biology* **82**: 327–348. DOI: http://dx.doi.org/10.1086/522809.

Woolley, A, Malone, T, 2011. What makes a team smarter? More women. *Harvard Business Review* **89**: 32–33.

Xie, Y, Crosbie, R, Simmons, CT, Cook, PG, Zhang, L. 2019. Uncertainty assessment of spatial-scale groundwater recharge estimated from unsaturated flow modelling. *Hydrogeology Journal* **27**: 379–393. DOI: http://dx.doi.org/10.1007/s10040-018-1840-0.

Zare, F, Guillaume, JHA, ElSawah, S, Croke, B, Fu, B, Iwanaga, T, Merritt, W, Partington, D, Ticehurst, J, Jakeman, AJ. 2021. A formative and self-reflective approach to monitoring and evaluation of interdisciplinary team research: An integrated water resource modelling application in Australia. *Journal of Hydrology* **596**: 126070. DOI: http://dx.doi.org/10.1016/j.jhydrol.2021.126070.

Zdun, U, Tran, H, Capilla, R, Zimmerman, O. 2014. Sustainable architectural design decisions. *InfoQ*. Available at https://www.infoq.com/articles/sustainable-architectural-design-decisions/. Accessed 14 December 2020.

How to cite this article: Iwanaga, T, Wang, H-H, Koralewski, TE, Grant, WE, Jakeman, AJ, Little, JC. 2021. Toward a complete interdisciplinary treatment of scale: Reflexive lessons from socioenvironmental systems modeling. *Elementa: Science of the Anthropocene* 9(1). DOI: https://doi.org/10.1525/elementa.2020.00182

Domain Editor-in-Chief: Steven Allison, University of California, Irvine, CA, USA

Associate Editor: Erika Zavaleta, University of California, Santa Cruz, CA, USA

Knowledge Domain: Ecology and Earth Systems

Published: June 11, 2021 **Accepted:** April 30, 2021 **Submitted:** December 22, 2020

Chapter 8: Conclusion

There are many challenges in integrated environmental modelling (IEM) and the effective and cohesive coupling of social, technical, and scientific knowledge remains a significant area of research. There are many perspectives, practices and the underlying lessons learnt from across the various fields involved in IEM development such that a considered incorporation of these is likely to be constructive and beneficial towards improving the modelling process and associated outcomes. The work presented in this thesis uses an example of such a multi-faceted approach that brings to bear perspectives from software development, systems engineering, and environmental modelling, shedding new light on the interdisciplinary processes that underlie IEM. The key contributions and findings are outlined below.

1. **Uncertainty and sensitivity analysis tooling must be easy to use, not just available**

Chapter 2 presents findings from an expansive hybrid bibliometric analysis of 11,625 papers. Publications on Uncertainty Analysis (UA) and Sensitivity Analysis (SA) in environmental modelling were analysed to identify common software tools as used in the environmental sciences for the purpose of UA/SA. An overview of common and emerging analysis approaches and terminologies were also synthesised.

Available literature acknowledges and emphasizes the importance of applying uncertainty and sensitivity analyses. There are now many tools available in support of such activities. The uptake of tooling, however, does not appear commensurate with the number of papers published detailing UA/SA applications, despite most software being open-source and freely available. The issue of longevity and usability is also raised as many tools appear to be unmaintained, lack sufficient documentation, and may be difficult to quickly incorporate into the typical modelling workflow. We therefore conclude that tooling must be operationalised by improving its accessibility for modeller use, rather than simply available, to ease the burden on modellers in the consideration of uncertainty and model sensitivities. For example, little guidance is offered to users on the suitability of methods (which can depend on the modelling context) or the interpretation of UA/SA results. An open development process and improving the level of accessible documentation and availability of user-centric interfaces and workflows would increase the uptake, and therefore efficacy, of UA/SA methods.

2. **A detailed assessment of a socio-environmental system through an exploratory approach**

Chapter 3 and 4 detail the modelling process and pathways undertaken for a case study in the Lower Campaspe catchment, North-Central Victoria. An IEM is developed to represent the systems of interest in the study area. The model represents six systems involving climatic influences, agricultural processes including use of water resources (i.e., farmer irrigation scheduling), recreational values, ecological suitability, as well as ground and surface water dynamics. The model explored opportunities and vulnerabilities using an exploratory approach that identified conditions that lead to beneficial outcomes relative to modelled baselines. These identified conditions are then regarded as “robust” pathways which enhance (or at least mitigate losses to) farm profitability, and recreational/ecological outcomes across a range of climatic and policy contexts. Adoption of conjunctive use (of ground and surface water) policies was found to improve the likelihood of robust outcomes. Perhaps unsurprisingly, farm level knowledge and management were also significant factors towards experiencing robust futures.

3. Appropriate model analyses are tied to development and application context

Integrated models that represent a system-of-systems investigation are often complex but ensuring or checking for sensibility with respect to their correct and expected function does not always require equally complex approaches. In Chapter 5, a Property-based Sensitivity Analysis (PbSA) approach is proposed to identify problematic (i.e., incorrect, or unexpected) integrated model behaviour as early in the development cycle as possible. Applying PbSA to complement traditional software testing approaches is useful as it is difficult, and not always productive, to create test cases for IEMs. The approach is demonstrated to help verify model behaviour in the model development context in which the constituent model(s) of an IEM may be undergoing frequent and rapid change, and the number of samples available may be restricted by the limited computational budget.

Additional guidance through a comprehensive framework for conducting sensitivity analysis on Spatially Distributed Environmental Models (SDEMs) are included in the Addendum. SDEMs are commonly required for IEMs, especially for hydrological constituent models. The framework consists of four broad steps, involving the identification of uncertainty sources and their strengths, the selection of SA method(s) and quantities of interest(s) appropriate to the spatial context, propagation, and finally evaluation and post-processing of SA results including visualization and reliability tests.

4. Modellers are human and so socio-technical processes must be considered in system-of-systems modelling

Holistic assessment of the interconnected socio-environmental systems necessitates the holistic consideration of the modelling process. Those conducting the modelling are human and dictate the scale of the modelling conducted. Thus, the social processes underlying the modelling should additionally be considered to better manage socio-technical concerns and issues. Chapter 6 explores and details the influence of the socio-technical domain on the treatment of scale and uncertainty in system-of-systems modelling. Chapter 7 expands on this through reflexive accounts of two case studies to draw out lessons learnt on managing interdisciplinary teams in the context of SoS modelling.

It is argued therein that considerations of the underlying socio-technical concerns should be explicitly incorporated for a more holistic approach to be realised. These include consistent and continual communication between those involved in the modelling, improved documentation practices to propagate understanding of the assumptions, decisions, and reflexive lessons and the reasons underlying those, and for discussions around scale and its influence on uncertainty to be explicit considerations in the modelling.

In summary, perspectives from environmental modelling, systems engineering, and software development are embedded in the thesis to synthesize and incorporate the wider and valuable contributions that can be made to the multifarious aspects of integrated environmental modelling. These include approaches in the development, testing, and application of models, and consideration of the socio-technical issues that underlie the modelling process. As integrated modelling is, by necessity, an interdisciplinary process, the research presented in this thesis helps clear a path towards more integrative and holistic approaches that better enable environmental systems modelling to achieve its purposes.

Addendum: Considerations for the sensitivity analysis of spatially distributed environmental models

Sensitivity analysis is a common analysis conducted to assess model complexity, behaviour, and the potential sources of uncertainty regarding quantities of interest that are output from a model. This chapter details the considerations in the application of sensitivity analysis specific to spatially distributed environmental models, through a pragmatic step-by-step framework. The framework guides modelers through the sensitivity analysis process when working with spatially distributed environmental models (SDEMs) with an emphasis on addressing sources of uncertainty related to raster and vector spatial datasets. This addendum was ultimately selected by the Editor-in-Chief of Environmental Modelling and Software to be a Position Paper in that journal. It was reviewed by two anonymous reviewers before publication.

The content of this addendum was originally included as Chapter 6 but has since been appended instead to conform with Section 5 of ANU procedure 003405 (https://policies.anu.edu.au/ppl/document/ANUP_003405).

Koo, H., **Iwanaga, T.**, Croke, B.F.W., Jakeman, A.J., Yang, J., Wang, H-H., Sun, X., Lü, G., Li, X., Yue, T., Yuan, W., Liu, X., Chen, M.*, 2020. Sensitivity analysis of spatially distributed environmental models - a pragmatic framework for the exploration of uncertainty, Environmental Modelling and Software 134, 104857. https://doi.org/10.1016/j.envsoft.2020.104857

Environmental Modelling and Software 134 (2020) 104857

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: http://www.elsevier.com/locate/envsoft

Position paper: Sensitivity analysis of spatially distributed environmental models- a pragmatic framework for the exploration of uncertainty sources

Hyeongmo Koo [a,i,j], Takuya Iwanaga [b], Barry F.W. Croke [b,k,a], Anthony J. Jakeman [b], Jing Yang [c], Hsiao-Hsuan Wang [d], Xifu Sun [b,k], Guonian Lü [a,i,j], Xin Li [e,l], Tianxiang Yue [f], Wenping Yuan [g], Xintao Liu [h], Min Chen [a,i,j,*]

[a] Key Laboratory of Virtual Geographic Environment (Ministry of Education of PR China), Nanjing Normal University, Nanjing, China
[b] Institute for Water Futures, Fenner School of Environment and Society, The Australian National University, Canberra, Australia
[c] National Institute of Water and Atmospheric Research, Christchurch, New Zealand
[d] Ecological Systems Laboratory, Department of Ecology and Conservation Biology, Texas A&M University, College Station, Texas, USA
[e] National Tibetan Plateau Data Center, Key Laboratory of Tibetan Environmental Changes and Land Surface Processes, Institute of Tibetan Plateau Research Chinese Academy of Sciences, Beijing, China
[f] State Key Laboratory of Resources and Environment Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China
[g] Guangdong Province Key Laboratory for Climate Change and Natural Disaster Studies, School of Atmospheric Sciences, Sun Yat-Sen University, Guangdong, China
[h] Department of Land Surveying and Geo-informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China
[i] State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing, 210023, China
[j] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, 210023, China
[k] Mathematical Sciences Institute, The Australian National University, Canberra, Australia
[l] CAS Center for Excellence in Tibetan Plateau Earth Sciences, Chinese Academy of Sciences, Beijing, China

ARTICLE INFO

Keywords:
Sensitivity analysis
Spatially distributed environmental model
Uncertainty
SWAT
Environmental modeling

ABSTRACT

Sensitivity analysis (SA) has been used to evaluate the behavior and quality of environmental models by estimating the contributions of potential uncertainty sources to quantities of interest (QoI) in the model output. Although there is an increasing literature on applying SA in environmental modeling, a pragmatic and specific framework for spatially distributed environmental models (SD-EMs) is lacking and remains a challenge. This article reviews the SA literature for the purposes of providing a step-by-step pragmatic framework to guide SA, with an emphasis on addressing potential uncertainty sources related to spatial datasets and the consequent impact on model predictive uncertainty in SD-EMs. The framework includes: identifying potential uncertainty sources; selecting appropriate SA methods and QoI in prediction according to SA purposes and SD-EM properties; propagating perturbations of the selected potential uncertainty sources by considering the spatial structure; and verifying the SA measures based on post-processing. The proposed framework was applied to a SWAT (Soil and Water Assessment Tool) application to demonstrate the sensitivities of the selected QoI to spatial inputs, including both raster and vector datasets - for example, DEM and meteorological information - and SWAT (sub) model parameters. The framework should benefit SA users not only in environmental modeling areas but in other modeling domains such as those embraced by geographical information system communities.

1. Introduction

Sensitivity analysis (SA) and uncertainty analysis (UA) are important tools for investigating model behavior, testing model hypotheses, and exploring the potential for simplifying models (Wagener and Pianosi, 2019). Uncertainty is intrinsic to all modeling work that involves representing natural processes and/or human behavior. Sources of uncertainty that need to be considered in such exercises are model input datasets, model structure, and model parameters. SA studies the influence of input factors (e.g., parameters, forcing, initial value of model states, model resolution, and model structure such as different parameterization schemes of a model or submodel) on model outputs. It is

* Corresponding author. Key Laboratory of Virtual Geographic Environment (Ministry of Education of PR China), Nanjing Normal University, Nanjing, China.
E-mail address: chenmin0902@163.com (M. Chen).

https://doi.org/10.1016/j.envsoft.2020.104857
Accepted 26 August 2020
Available online 1 September 2020

considered a key practice in the assessment of environmental models (Chen et al., 2019; Gan et al., 2014; Jakeman et al., 2006; Matott et al., 2009; Oakley and O'Hagan, 2004; Pianosi et al., 2016; Yue et al., 2020). In comparison, UA quantifies the uncertainty of model outputs from input datasets and model parameters, typically characterized by empirical probability distributions and/or confidence bounds for the model parameters and outputs. UA can be considered an extension of SA with the uncertainty distributions for the input factors being used as the perturbations. Therefore, SA can be used to indicate when uncertainty in input factors matters in terms of the impact on the uncertainty in the outputs. Care must however be taken in the interpretation of SA results as sensitivities can be dependent on parameter ranges selected, model structure assumed, length of data period examined and its climatic forcing (Shin et al., 2013).

The use of SA and UA in environmental modeling has become of particular importance due to the highly complex nature of environmental systems, and the attendant complexity of models typically invoked to represent them. This is especially the case for spatially distributed environmental models (referred to from hereon as SD-EMs), where there tends to be a considerable number of model parameters due to their spatially variant nature, and substantial uncertainty in the model and its predictions. Uncertainty and sensitivity related studies in environmental modeling are rising in popularity because of the growing awareness of the importance of models in supporting informed decision making (Douglas-Smith et al., 2020), coupled with the fact that current process-based environmental models are typically, and perhaps necessarily, deterministic in their representation (Farmer and Vogel, 2016; Uusitalo et al., 2015).

This paper focuses on Monte Carlo simulation-based SA of SD-EMs, which is a valuable tool per se and one that can also inform uncertainty analyses. Monte Carlo simulation-based approaches are widely applied due to their ease of implementation, yet there is a lack of a comprehensive pragmatic framework for conducting such approaches for SD-EMs (Yang et al., 2018). An SD-EM is intrinsically tied to the spatial dimensions of producing and utilizing data that represent the spatially distributed nature of the modeling context. Grid-based digital elevation models (DEMs), site-specific point measurements, and remotely sensed images are examples of such data. However, SAs are rarely conducted for DEM and DEM-derived parameters even though the inherent scale and errors of a spatial dataset and/or of the whole environmental model can have a significant impact on model outputs (Tran et al., 2018). A crucial issue to take into account regarding spatial datasets is the spatial structure of their uncertainty. Generally, spatial datasets are characterized by spatial dependence (i.e., spatial coherence), and their uncertainties are also spatially autocorrelated (Oksanen and Sarjakoski, 2005a; Wechsler, 2007). Thus, ignoring such characteristics can lead to erroneous estimation of sensitivity measurements. Moreover, because spatial datasets often determine the uncertainty in model resolution and structures through their boundaries, discretization and scale, exploring uncertainty related to spatial datasets can partly account for model uncertainty in SD-EMs.

This article introduces a pragmatic framework for the application of SA to an SD-EM, using a scenario/simulation-based approach to investigate the significance of potential uncertainties in the model inputs, which can not only explore model and data assumptions transparently but also be an informative precursor to a more thorough UA. The objectives of the framework are to provide sufficient information and background in order to guide the selection of more appropriate choices at each step of the SA process: potential uncertainty source identification; selection of SA method(s) and quantities of interest (QoI); perturbation propagation; and SA evaluation and post-processing. The framework emphasizes the following aspects: it attempts to address potential uncertainty sources related to spatial datasets; and assists in propagating the potential uncertainty sources by considering their likely spatial structure. Therefore, the framework helps to explore the impact of potential uncertainty of spatial datasets in an SD-EM, and to compare their relative impacts with the usual factors in SA (e.g., model parameters). The framework is intended to benefit both non-experts and SA users in environmental modeling and geographical information system (GIS) communities.

The remainder of this article is organized as follows. Section 2 broadly introduces the pragmatic framework for applying SA to SD-EMs, covering potential uncertainty source identification, selection of SA method and QoI, perturbation propagation, and SA evaluation and post-processing. Then, from Sections 3 to 6, the detailed steps and their corresponding considerations are discussed. Section 7 provides a concise example of the SA framework. The article concludes in Section 8 with a discussion of future needs and opportunities.

2. A pragmatic SA framework for SD-EMs

The presented framework prescribes sequential steps in which important considerations are highlighted to guide modelers towards the selection of appropriate choices for the pragmatic application of SA to uncertainty exploration in SD-EMs. The overarching steps and the corresponding considerations are depicted in Fig. 1. The main purpose of the framework is to identify the contributions of potential uncertainty sources to the selected QoI. This section introduces the pragmatic framework to provide a broad guideline for SA users, while the following sections detail the considerations within each step.

The first step is to identify potential sources of uncertainty (Section 3). Numerous studies have investigated uncertainty sources in the context of environmental modeling, and classified them in their own schemes (Matott et al., 2009; Refsgaard et al., 2007). Understanding these general classification schemes and the uncertainty sources involved assists in identifying the sources of uncertainty related to a specific application. In particular, this article discusses potential uncertainty sources not only in model parameters and model uncertainty, but also in spatial datasets which are used as direct input(s) and/or to derive parameters to describe the underlying spatially distributed structure of SD-EMs (e.g., DEM).

The second step is the selection of SA methods and QoI (Section 4). This selection primarily depends on the purposes of SA (e.g., screening and ranking) and the characteristics of the SD-EM. The characteristics can include the model complexity, and/or computational cost. As this framework is intended for Monte Carlo simulation-based SA, applying SA methods that require a large number of model evaluations to determine SA measures might not be feasible for a computationally expensive model. This article broadly categorizes the most frequently-used SA methods for environmental modeling based on their purposes and characteristics, and synthesizes them to assist SA users and communities in selecting appropriate ones for a pragmatic SA application. Here, we provide a general description, and several previous studies provide complementary explanations for further SA methods (Pianosi et al., 2016; Borgonovo and Plischke, 2016; Sarrazin et al., 2016). For selecting QoI, only scalar outputs are generally utilized as QoI in SA of environmental models, which often requires aggregating spatially and/or temporally distributed outputs into a scalar function (Pianosi et al., 2016). However, since potential uncertainty sources in SD-EMs have different impacts on spatially, and also temporally, distributed scalar outputs (Pappenberger et al., 2008; Wang et al., 2013), preserving a spatial distribution of scalar outputs might be useful to understand the underlying spatial processes within SD-EMs, even if that requires managing vast computing power.

The third step in the framework is the perturbation propagation of the identified potential uncertainty sources (Section 5). Generally, local and global SA methods require different types of perturbation propagation methods. Thus, local SA utilizes the neighborhood values of the nominal value, and global SA generally requires the input variability space via a probability distribution (Borgonovo and Plischke, 2016). For all SA methods, perturbation propagation with appropriate distributions is crucial because the representativeness of uncertainty sources

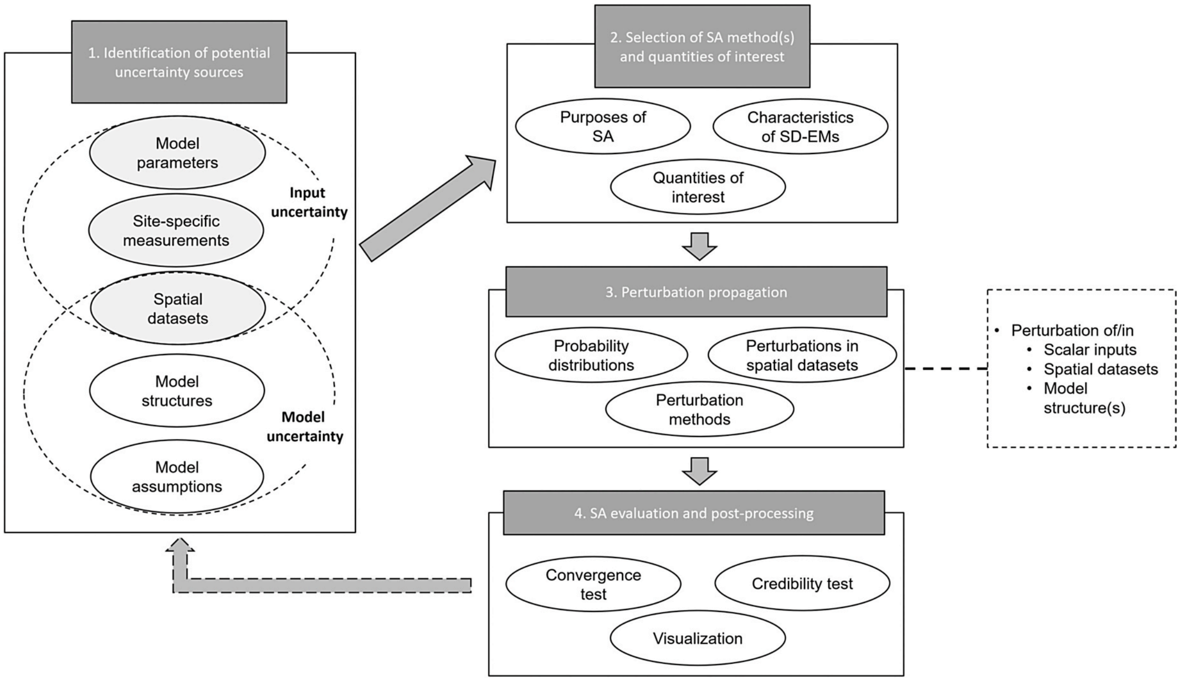


Fig. 1. The steps and considerations in the application of the SA framework.

primarily determines the SA results. Different types of uncertainties also need different types of perturbation propagations. For example, the propagation of perturbations in a model parameter that is represented as a scalar random variable could be performed using a probability distribution (e.g., normal distribution), while more complex perturbation propagation methods are necessary for characterizing the uncertainty of input datasets (e.g., spatial datasets) by simultaneously taking account of their characteristics (e.g., spatial autocorrelation) (Wechsler, 2007; Crosetto and Tarantola, 2001). Nonetheless, perturbation propagation of model uncertainty is still an ongoing research subject and remains a fruitful area of investigation (Matott et al., 2009; Uusitalo et al., 2015; O'Hagan, 2012).

The final step of the framework is evaluating the results of the SA, which includes post-processing of analysis results (Section 6). In evaluating SA results, reliability and convergence aspects should be assessed as a verification step. SA measures or metrics vary with sample size, thereby requiring a convergence test to check if the metric(s) of choice is converging and its confidence bounds acceptable for the purpose (Yang, 2011). In addition, different SA methods are based on different premises, may produce different metrics, and hence produce different outputs. Thus, SA with different methods can increase confidence in the reliability or interpretation of SA outputs. Finally, this step also includes credibility assessment for SA outputs. If unexpected SA outputs are obtained, these outputs can lead to new indications of uncertainty in model behavior (Pappenberger et al., 2008), or indicate issues with the model implementation (Pianosi et al., 2016). Otherwise, these outputs can assist with revising the SA steps, such as identifying missing uncertainty sources and redefining the perturbation propagation approach. Because SA results are generally associated with large sets of potential uncertainty sources, visualization methods of SA results are useful to identify critical uncertainty sources and to compare their importance. Thus, this step includes the descriptions of specific visualization methods with their corresponding SA methods, conventional scientific visualization techniques for SA outputs (Kelleher and Wagener, 2011), and geographical visualization and analysis for the representation of spatially distributed SA measures (Feick and Hall, 2004).

3. Identification of potential uncertainty sources

3.1. Classification of uncertainty sources

This initial step involves the identification of potential uncertainty sources associated with the model's input factors that influence the selected outputs of an environmental model, or functions of those outputs (i.e., QoI). Various types of uncertainty sources could influence the outputs of the models, and numerous classification schemes for uncertainty sources have been introduced to categorize them (Matott et al., 2009; Refsgaard et al., 2007; Beck, 1987; Linkov and Burmistrov, 2003). Because the classification schemes commonly share two fundamental uncertainty sources (i.e., input and model uncertainties), this article discusses the various uncertainty sources in these two broad categories.

3.2. Input uncertainty

Input uncertainty is associated with model parameters and input datasets. Among various sources of uncertainty, model parameter uncertainty is the most commonly considered source (Setegn et al., 2010; Wu and Chen, 2015; Wu and Liu, 2012; Berzaghi et al., 2019; Porada et al., 2018) and can be controlled to some extent through calibration processes (Zhao et al., 2018). Therefore, uncertainty in model parameters is often considered to be reducible, and it has been argued that model parameters can be carefully tailored to reduce that uncertainty related to model outputs and to improve model performance (Matott et al., 2009). However, if globally optimized parameters are obtained through a calibration process, they would also be affected by other sources of uncertainty, including input data uncertainty, model uncertainty, and also model parameter uncertainty, and these might lead to "equifinality" (Zhao et al., 2018; Beven and Freer, 2001).

Input datasets are usually assumed to be accurate, that is, effectively without uncertainty. However, this is incorrect as all data have inherent uncertainties (Chrisman, 1991). Uncertainties related to input datasets can be irreducible (Matott et al., 2009), and thus they are often ignored in uncertainty related studies. Moreover, uncertainty in spatial datasets involves spatial autocorrelation (Griffith, 2008; Koo et al., 2018c). In the SD-EM context, spatial datasets include maps and site-specific measurements. For example, temperature and precipitation surfaces would have strong positive spatial autocorrelation, and soil and land use-land cover (LULC) datasets possess complex spatial autocorrelation (Legendre, 1993). Site-specific measurements (e.g., meteorological data) have spatial autocorrelation as well as temporal autocorrelation. Therefore, if an environmental modeling analysis does not address the independent inputs and relationships among them, an incomplete understanding of the uncertainty in the model will result, leading to a biased estimation of the confidence in the model outputs.

The uncertainty in spatial datasets is generally caused by five fundamental components: lineage, positional accuracy, attribute accuracy, logical consistency and completeness (ANSI, 1998; Koo et al., 2020). Briefly, lineage relates to the description of spatial data sources (e.g., dates and reference systems), and logical consistency describes the fidelity of spatial data structure (e.g., topology). Completeness refers to selection criteria of spatial entities, for example, geometric thresholds such as minimum width and area of spatial features. Positional and attribute accuracies literally refer to uncertainties respectively in position (i.e., location) and attribute information of spatial datasets. Among these components, positional and attribute accuracies are closely related to uncertainties in SD-EMs, and they show different aspects depending on the type of spatial datasets. Spatial datasets can be broadly divided into raster and vector datasets.

A typical example of a raster dataset is a DEM, where scale (i.e., resolution), random and systematic measurement uncertainties resulting from attribute accuracy are the major uncertainty sources (Hengl et al., 2010). The scale issue is relatively well discussed as a source of uncertainty in DEMs (Chaubey et al., 2005; Dixon and Earls, 2009; Lin et al., 2013c; Shen et al., 2013). Together with scale, measurement uncertainty is known to also have an impact on watershed delineation (Oksanen and Sarjakoski, 2005a; Wu et al., 2008), stream network extraction (Hengl et al., 2010), and derivation of other topographic parameters (Wechsler, 2007). As systematic measurement uncertainty generally shows a fixed pattern stemming from DEM generation processes (e.g., blunders), if the cause of the uncertainty is known, it could be reduced (Wechsler, 2007). However, random measurement uncertainty still remains after reducing systematic uncertainty. For example, the Shuttle Radar Topography Mission (SRTM) v.4.1 dataset has a vertical accuracy of $\pm$16 m at a 95% confidence level (Mukul et al., 2017). Other widely used raster datasets are LULC and soil datasets, which possess uncertainty related to positional uncertainty and scale issues (Koo et al., 2020).

Vector datasets are typically used to define the boundary of a study area, and describe topographic and/or environmental features such as stream networks. In addition, site-specific measurements are handled as a type of vector dataset, generally point features that have attributes on specific locations, for example, measurements of precipitation, temperature, wind speed, humidity and solar radiance. Even though some raster datasets (e.g., precipitation and temperature surfaces) are converted from site-specific measurements, their uncertainty sources can mainly be explained by uncertainty in vector datasets. Vector datasets typically include two main uncertainty sources, which are positional and attribute uncertainties (Koo et al., 2018a). Positional uncertainty refers to the uncertainty of geographical features in vector datasets, which often results from a global positioning system (GPS), geocoding and digitizing errors. Attribute uncertainty describing non-spatial properties of geographical features in vector datasets are generally estimated from their sampling processes (Aouissi et al., 2013; Strauch et al., 2012; Tasdighi et al., 2018; Bárdossy and Das, 2008; Chaplot et al., 2005; Cho et al., 2009; Gong et al., 2012; Masih et al., 2011) and measurements (Shen et al., 2015; Li, 2014). In addition, attribute uncertainty often includes spatial autocorrelation. When attribute uncertainty contains temporally varying quantities, they also need to consider the information lost in converting to discrete-time (Littlewood and Croke, 2013).

3.3. Model uncertainty

Model uncertainty results from the inability of a model to mimic the real-world (Yen et al., 2014). Model uncertainty might be subdivided into the effects of model structure, model resolution, and model integration uncertainties (Matott et al., 2009; Voinov and Shugart, 2013). First of all, model structure uncertainty is caused by a model structure that imperfectly represents underlying environmental processes in a model (Yen et al., 2014). Numerous alternative model structures (e.g., scientific hypotheses and equations) might be proffered in a model, which could adversely impact model outputs. A related consideration is the issue of the identifiability of the model structure (Guillaume et al., 2019; Shin et al., 2015), which largely means the data available are insufficiently informative to identify unique values of some of its parameters. SA methods are often used to determine the insensitive/non-identifiable parameters so that focus for calibration and/or uncertainty analysis can then be turned to the most sensitive ones. Second, model resolution uncertainty is due to uncertainties in the spatio-temporal discretization, boundary specification, and scale dependence of a model (Matott et al., 2009). In an SD-EM, spatial discretization, boundary and scale are often determined by the available spatial datasets, and model resolution uncertainty can then be partially explained through exploring uncertainty of the spatial datasets (Koo et al., 2020; Trusel et al., 2015) (Fig. 1).

Another aspect of model uncertainty arises from model integration processes (Chen et al., 2020). Currently, environmental models become more complex by integrating multiple models (Lin et al., 2013a, 2013b; Lu et al., 2019). The integration processes yield uncertainty from skewed space (e.g., difference in spatial resolution), mismatched measurement scales, and confusion of linguistic representations (Voinov and Shugart, 2013). Particularly, in an SD-EM, when sub-models with different spatial and temporal scales are integrated without a solid design, the uncertainty of an integrated model could become large and undetectable (Tscheikner-Gratl et al., 2019).

4. Selection of SA method(s) and quantities of interest

This step firstly provides guidance on SA method selection based on two main criteria: the purposes of the SA and the characteristics of the SD-EM. This guidance includes only two fundamental SA purposes (i.e., ranking and screening), but SA can have additional purposes such as factor mapping that provides further descriptions for the input space related to QoI (Saltelli et al., 2008). SA methods for ranking generate the order of input factors based on their relevant influence on QoI, and screening methods identify input factors with significant or negligible influence on model output (Pianosi et al., 2016). SA methods for each purpose can be applied sequentially, such that model results from screening methods are leveraged to reduce the number of input factors and are followed by SA for the purpose of ranking, thus reducing the overall number of model evaluations (Saltelli et al., 2004; Sun et al., 2012). Secondly, two major characteristics of an environmental model are necessary to consider in selecting appropriate SA methods: model complexity and interdependency between input factors (Saltelli, 2002). Here we briefly discuss widely used SA methods, including local SA, the Morris method, correlation and regression, and variance-based SA methods according to the two major criteria. Fig. 2 classifies these SA methods, where the positions of each SA method relate to SA purpose and model complexity and their outlines represent interdependency. In addition, the advantages of an emulator and its consideration for dealing with spatially distributed outputs are also briefly discussed.

Local SA is the simplest SA method, and is often conducted through one-at-a-time (OAT) perturbation of input factors around their nominal values to determine the response of model outputs (Sun et al., 2012; Campolongo and Saltelli, 2000). A formal approach for local SA involves using partial derivatives (Helton, 1993). Partial derivatives can provide SA measures/metrics for both ranking and screening; however, they

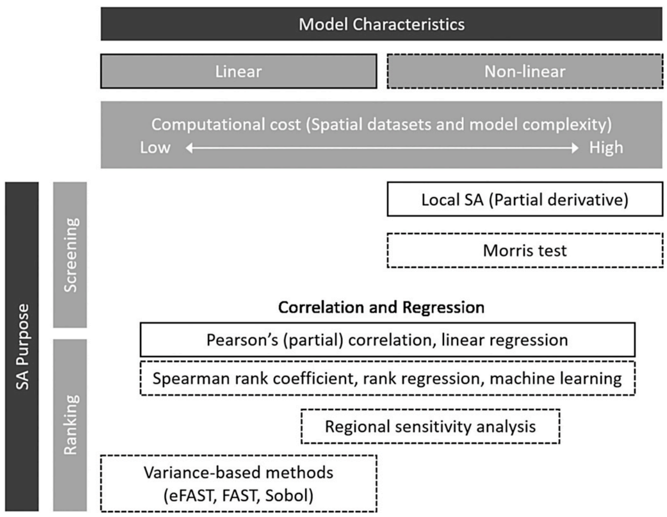


Fig. 2. The criteria for SA method selection. Appropriateness of SA method depends on model characteristics including linearity (solid box outline) or non-linearity (dashed box outline), computational cost and SA purpose (position of the boxes).

should be rescaled and applied to several locations in factor space in order to reveal the global effects of input factors with different measurement units (Borgonovo and Plischke, 2016; Campolongo et al., 2011).

Purely OAT analyses are, however, typically inappropriate for determining sensitivity estimates. Such analyses do not consider interactions among input factors (Borgonovo and Plischke, 2016), and while OAT can investigate non-linearities if input factors are independent (Newham et al., 2003), typical applications are unable to do so due to the use of a single perturbation (Sun et al., 2012; Saltelli and Annoni, 2010). Importantly, because local SA evaluates sensitivity at a specific location of input factors, rather than over their plausible ranges as global SA does, it might provide a limited indication of model behavior (Sun et al., 2012; Sobol, 2001) although pure OAT analyses have the advantage of reduced computational time over a more substantive global analysis. An initial indication of model behavior can be gained with just N+1 model evaluations for N input factors (Pianosi et al., 2016), or fewer if groups of input factors are perturbed together through group sampling (Sobol', 2001). Improper model behavior caught at this stage indicates errors in the model implementation to be addressed before global analyses are applied.

A simple global extension of local SA is the Morris method (Morris, 1991), also known as the Elementary Effect Test (Saltelli et al., 2008). The Morris method generally requires much lower numbers of model evaluations than other global SA methods for the purpose of screening, and thus the Morris method is appropriate for computationally complex models and/or models with a large number of input factors (Campolongo et al., 2007; Herman et al., 2013). A drawback of the Morris method is that it gives a poor measure of the relative importance between factors, and can be considered as offering qualitative sensitivity measures only (Brockmann and Morgenroth, 2007). Besides the average elementary effects, it does provide the standard deviations of the elementary effects, which are beneficial for identifying interaction effects among input factors (Norton, 2015).

Regional sensitivity analysis (Spear and Hornberger, 1980) typically divides input factors into two or more groups depending on a prescribed threshold of model output, and then studies the difference in their empirical cumulative distribution functions (CDF) for each input factor. The Kolmogorov-Smirnov (K–S) statistic quantifies the divergence between the CDF and serves as a common sensitivity measure. Thus, if the K–S statistic is high (i.e, the CDF of one group differs from the other), the input factor has a significant influence on model output (Pianosi and Wagener, 2015). K–S statistics are mainly utilized for ranking input factors. However, the K–S statistic is inappropriate for screening because it is only applicable to the same groups (Saltelli et al., 2008). The advantage of regional sensitivity analysis is that it is applicable for any type of splittable model outputs (e.g., Futter et al., 2007; Whitehead et al., 2015; Whitehead and Hornberger, 1984). However, if the grouping (i.e., splitting) criterion is not clear (i.e., a model does not have meaningful model output values to describe model behavior), regional sensitivity analysis would be inappropriate.

Various correlation and regression methods are also extensively used to measure sensitivities. These methods basically obtain SA measures based on different statistics (i.e., correlation and regression coefficients) between input factors and QoI generated from a Monte Carlo simulation (Pianosi et al., 2016; Helton et al., 2006). Specifically, for correlation coefficient estimations, various types of correlation coefficients are selected mainly based on the linearity between input factors and model outputs. When they have a linear relationship, Pearson and partial correlation coefficients are appropriate methods (Saltelli and Marivoet, 1990). If the relationship is non-linear, Spearman and partial rank correlation coefficients can be used as alternatives (Pastres et al., 1999). Furthermore, if SA methods simultaneously are to take account of multiple relationships for multiple outputs, a canonical correlation analysis provides an additional option (Minunno et al., 2013).

Regression methods obtain sensitivity measures by estimating regression coefficients, which are commonly standardized. Regression methods are often superior to correlation methods in deriving sensitivity measures, especially when a large number of input factors are considered since regression methods can obtain SA measures of all input factors at once. However, while linear regression is the simplest and most widely used SA method (Iman and Helton, 1988), it is not suitable if there is a non-linear or non-monotonic relationship in the model response, and a high level of interaction among factors also makes linear regression act poorly (Yang, 2011). When a non-linear relationship exists, rank regression (Storlie et al., 2009) and machine learning techniques, such as decision trees (Singh et al., 2014), are appropriate. Regression and correlation methods are commonly utilized for both screening and ranking purposes.

Variance-based methods produce sensitivities by decomposing the variance of a model output into the contributions from input factors. The contributions can be defined according to different indices, for example, first-order and total indices (Saltelli et al., 2008). The first-order index quantifies the contribution of a specific input factor to the variance of the selected QoI, while the total index measures the total contribution of an input factor to the variance of the QoI, including those due to its interactions with other input factors. The first-order index is usually used to rank input factors when interactions are not significant. With total indices, variance-based SA methods are able to address non-linearity of model responses to input factors. Factors with a total index close to zero can be considered negligible and screened out (Pianosi et al., 2016). Often, these negligible factors are made constant, a practice referred to as "factor fixing".

Variance-based SA methods can be challenging for computationally intensive complex models because they take a relatively large number of factor samples and related model evaluations to obtain reasonably accurate and stable indices (Gan et al., 2014). However, several approaches, such as the Sobol' method (Sobol, 2001), Fourier Amplitude Sensitivity Test (FAST) (Cukier et al., 1973), and extended FAST (Saltelli et al., 1999), have been proposed to more efficiently estimate main and total effects (Paleari and Confalonieri, 2016). In practice, however, the computational requirement is still a burden for these SA methods. Moreover, common variance-based approaches operate on a number of assumptions, including that: the variance of model outputs resulting from the prior input distribution is indicative of input factor sensitivities; inputs are independent (Saltelli and Tarantola, 2002); and the distribution of the sampled model outputs, often estimated through kernel density estimation approaches, are unimodal. Misleading SA results may be produced if model outputs do not conform to these assumptions (Pianosi et al., 2016).

For computationally intensive models, SA (e.g., variance-based methods) can be implemented using an emulator, which is a statistical approximation of the output response surface of the original environmental model (O'Hagan, 2012). A simple approach for building an emulator is through use of Gaussian processes (Oakley and O'Hagan, 2004), though other options exist, such as polynomial chaos expansions (Sudret, 2008), statistical emulators (Young and Ratto, 2011) and machine learning-based emulators (e.g., random forest and gradient boosting) (Storlie et al., 2009). However, an emulator might be inappropriate to evaluate a large number of input factors because it suffers from estimation inefficiency and inaccuracy due to the curse of dimensionality (Storlie et al., 2009; Li et al., 2020). This can be resolved by screening out negligible input factors, or by applying an emulator that includes a procedure for input factor selection (Yang, 2011).

The selection of QoI (sometimes embodied in an objective function or loss function) is also crucial to reflect the modeling purposes, as different modeling purposes lead to different sensitivity measures in the input factors. For example, rainfall intensity yields more sensitivity to stream flow peak than to baseflow. Based on modeling purpose, a large number of QoI have been used in hydrological models. The most frequently used include the Nash-Sutcliffe coefficient, root-mean-square-error (RMSE), and differences in the flow duration curve (e.g., between simulated and observed flows, and total nitrate). Although SA tends to select only single scalar QoI, SD-EMs often need to explore spatially distributed sensitivities of input factors on multiple- and multi-dimensional outputs (Gupta and Razavi, 2018; Pappenberger et al., 2008), which can involve a colossal computational burden. Emulators may be developed to circumvent the issue of computational cost. A typical practice, however, is to build separate emulators for individual outputs (Ryan et al., 2018), which may impose an additional computational cost. Implementing other types of SA methods may be useful to mitigate this computational burden, including a separate generalized additive model (Mara and Tarantola, 2008), partial least squares (Sobie, 2009), multi-fidelity polynomial chaos expansions (Palar et al., 2018), and global sensitivity matrix approaches (Razavi and Gupta, 2019). Moreover, decreasing the dimensionality of outputs by using a principal component analysis and grouping factors based on bootstrap-based clustering (Sheikholeslami et al., 2019) can provide another solution (Gómez-Dans et al., 2016).

5. Perturbation propagation

Monte Carlo simulation-based SA needs to propagate the perturbations of input factors through the model to analyze the sensitivity of model outputs and their QoI to those input factors (Saltelli and Tarantola, 2002). Proper selection of the perturbations within plausible ranges and distributional assumptions is a crucial step in SA because the perturbation attempts to reflect the degree of uncertainty in input factors. This section introduces useful methods for perturbation propagation of the corresponding uncertainty sources, including input parameters, spatial and point datasets.

5.1. Model parameters

Model parameters in environmental models are typically represented as scalar variables, and often treated as random variables with prior probability distributions (Borgonovo and Plischke, 2016). Specifically, samples of individual model parameters are obtained from their corresponding probability distributions, and then SA evaluates model responses based on the samples. The interactions between model parameters can be represented using a covariance matrix, for example, the Cholesky decomposition (Xiu and Karniadakis, 2003). Because defining appropriate probability distributions with plausible ranges are also crucial for evaluating model parameter sensitivities, taking all available information on individual model parameters is necessary for the generation of those probability distributions (e.g., expert opinion) (Crosetto and Tarantola, 2001).

5.2. Raster datasets

An SD-EM normally uses various types of spatial datasets, including spatially distributed input datasets and site-specific measurements. Because spatial datasets have various forms (e.g., vector and raster datasets), different types of perturbation propagation are required (Crosetto and Tarantola, 2001). Furthermore, the propagation for spatial datasets should consider the characteristics of those datasets, especially spatial autocorrelation (Temme et al., 2009). Table 1 demonstrates applicable perturbation propagation methods for the various types of spatial datasets.

Raster datasets are either generally subdivided into categorical (e.g., LULC and soil datasets) or quantitative (e.g., DEM, temperature and precipitation surfaces) rasters, requiring different perturbation propagation methods (Heuvelink, 1998). In quantitative rasters, individual cell values can be treated as individual random variables with their own probability distributions, which means an observed quantitative raster is just one rendering of all possible realizations. However, this assumption ignores the spatial structure of uncertainty in quantitative rasters. Random fields are widely used to represent uncertainty in quantitative rasters. Random fields comprise a surface of random values that estimates uncertainty magnitude, variance, and spatial variability, where each value represents potential uncertainty at a specific location of the grid cell (Wechsler, 2007). Furthermore, random fields are applicable for regularly discretized space-time voxels in 3D rasters (Pebesma et al., 2007). Like other perturbation propagation methods, random fields require a definition of the appropriate potential uncertainty level (i.e., plausible ranges) and their spatial structure. If the information for the

Table 1
Perturbation propagation for spatial datasets and their applications.

Spatial dataset type		Examples	Perturbation methods
Raster	Quantitative	DEM, and temperature surface	Spatial moving average (Wechsler and Kroll, 2006) Pixel swapping (Fisher, 1991a) Spatial autoregressive models (Hunter and Goodchild, 1996) Sequential Gaussian simulation (Aerts et al., 2003) Three parameter method (Ehlschlaeger et al., 1997)
	Categorical	LULC, and soil datasets	Using fuzzy classification information (Lucieer and Kraak, 2004) Using the confusion matrix (Fisher, 1991b) Applying models for vector datasets (Kiiveri, 1997; Shi, 1998)
Vector	Point	Site-specific measurements	Error ellipse (Dutton, 1992) Spatial autoregressive models (Hunter and Goodchild, 1996)
	Line	Boundary of study area, and stream network	Epsilon band (Crosetto and Tarantola, 2001) Stochastic process-based models (Shi and Liu, 2000) Entropy-based models (Gong and Li, 2011) Statistical simulation error models (Tong et al., 2013)

ranges and structures are not available, random fields are often estimated using the accuracy statistics of a target raster dataset (e.g., root mean square error). If the information is obtained from a survey and other methods, this can be used for estimating parameters of a random field generation.

The simplest method for random field generation is using a normal distribution with a mean of zero and standard deviation derived from the accuracy statistics of a quantitative raster dataset. However, potential uncertainty in spatial datasets has spatial structure, including spatial autocorrelation (Wechsler, 2007; Oksanen and Sarjakoski, 2005b), which should be considered in a random field generation. The following methods are the typical methods accounting for spatial autocorrelation in random field generation. The first method is the spatial moving average (Wechsler and Kroll, 2006), which applies a low pass filter to a random field generated from a simple probability distribution (e.g., normal distribution) without considering spatial autocorrelation. The size of a low pass filter determines the level of spatial autocorrelation, which often covers from 3 by 3 grids to the grid size that is computed from the range of a semi-variogram in a target quantitative raster.

Pixel swapping (Fisher, 1991a; Goodchild and Openshaw, 1980) and a spatial autoregressive model (Hunter and Goodchild, 1996; Koo et al., 2019; Anselin, 1995) are other methods for random field generation that considers spatial autocorrelation. They can be utilized for quantitative raster datasets and attribute uncertainty in vector datasets. Pixel swapping is developed based on the concept of simulated annealing, where two cells in random fields are continuously swapped until spatial autocorrelation in the random fields achieves its threshold level derived from spatial autocorrelation level in a target raster dataset. This method has the advantage that it is a simple process, but it is difficult to implement for a large spatial dataset due to the slowness of the procedure (Oksanen and Sarjakoski, 2005a). A spatial autoregressive model produces random fields based on the following equation:

$$Y = (I - \rho W)^{-1}\varepsilon$$

where I denotes an identity matrix, and ε is a vector that is generated generally from a probability distribution. By using an autoregressive parameter (ρ), this model effectively determines the level of spatial autocorrelation. W is a spatial weights matrix, which is also useful for specifying various types of neighborhood definitions (e.g., contiguity and k nearest neighbors). Because it is an n by n matrix and tends to be sparse, computing its inverse often requires a lot of resources for large datasets (Anselin, 2005).

Sequential Gaussian simulation is a widely applicable method for random field generation (Koo et al., 2020; Aerts et al., 2003), which is developed based on a geostatistical approach with a normality assumption for potential uncertainty (Goovaerts, 1997). The basic steps for sequential Gaussian simulation are as follows: first, random paths are generated in a field, and each node in the path is sequentially visited; second, at each node, descriptive statistics for a local conditional probability function are estimated based on surrounding values using kriging; and finally, a random value is generated from the local conditional probability function. If survey samples for uncertainty exist, the values at the sample nodes are maintained with the original sample value, which has been referred to as conditional Gaussian simulation (Aerts et al., 2003). A data transformation process will be required if a dataset does not follow a Gaussian distribution because sequential Gaussian simulation is only applicable to a dataset that follows a Gaussian distribution (Deutsch and Journel, 1998).

Additionally, if prior knowledge for the spatial structure of uncertainty in a quantitative raster dataset can be obtained from surveys or higher accuracy datasets, the three-parameter method (Ehlschlaeger et al., 1997) is useful to reflect this prior knowledge. Whereas only one parameter is available in pixel swapping and a spatial autoregressive model, this three-parameter method is superior for generating random fields with the mean and standard deviation under Gaussian distribution, and spatial autocorrelation (Wechsler, 2007). Other studies (Hengl et al., 2010; Fisher, 1998; Holmes et al., 2000) discuss further perturbation propagation for quantitative raster datasets and have implemented them in real-world applications.

Perturbation propagation for categorical raster datasets has received less attention than those for quantitative raster datasets (Crosetto and Tarantola, 2001). Generally, the selection of the method for categorical raster datasets relies on the method of dataset generation. When a categorical raster dataset (e.g., for LULC and soil data) is generated using a fuzzy classification method of satellite images, individual pixels include uncertainty information of the classification result such as probability or membership vectors (Lucieer and Kraak, 2004). This type of uncertainty information can directly be applied to Monte Carlo simulation-based SA. When a conventional classification method is used for raster data generation, the confusion matrix (i.e., error matrix) is available for SA, which is a cross-tabulation of the classified raster against reference samples to estimate classification accuracy. The confusion matrix based perturbation propagation is fully discussed in (Heuvelink, 1998; Fisher, 1991b). However, the confusion matrix might be limited in representing the spatial structure of uncertainty (Comber et al., 2012). In addition, perturbation propagation for vector datasets can also be applied for categorical raster datasets, mainly to represent their positional uncertainty (Kiiveri, 1997; Shi, 1998). The details of perturbation propagation will be discussed with that for vector datasets in Section 5.3.

5.3. Vector datasets

Vector datasets, including the case of site-specific measurements, mainly have two major potential uncertainty sources - attribute and positional uncertainties (Koo et al., 2018a; ANSI, 1998). Attribute uncertainty results from sampling and measurement errors. Thus, perturbation propagation can be simply a probability distribution function derived from the descriptive statistics of potential attribute uncertainty. If the attributes of geographical features in vector datasets are spatially autocorrelated, pixel swapping (Goodchild and Openshaw, 1980) and a spatial autoregressive model (Anselin, 1995) are useful to describe the spatial structure of attribute uncertainty. For example, with a spatial autoregressive model, a random vector (i.e., ε) is generated from the descriptive statistics of attribute uncertainty, and spatial structures are described by a spatial weights matrix (i.e., W) of geographical features and predefined spatial autocorrelation level (i.e., ρ) as we described in the previous Section 5.2.

Propagating perturbations for positional uncertainty in vector datasets has been one of the major focuses in spatial data quality research groups (Devillers et al., 2010). The geometric features of vector datasets are generally classified as point, line and polygon. However, in the context of modeling positional uncertainty, a polygon is considered as a closed line (Shi, 1998). According to the latter, perturbation propagation is discussed in two general feature types - point and line.

Error ellipses constitute a common method for propagating perturbations of point features (Dutton, 1992; Stanislawski et al., 1996; Goodchild, 1991), where x-y coordinates follow a two-dimensional extension of a probability distribution function on individual points. For positional uncertainty representation, a normal distribution can be simply used for the probability distribution (Goodchild, 1991; Wolf and Ghilani, 1997), but a log-normal distribution, a mixture of bivariate t distributions (Zimmerman et al., 2007), and a chi-square distribution (Griffith et al., 2007) have also been suggested (Karimi et al., 2004; Koo et al., 2018c). Positional uncertainties of point features are usually considered independent, though possibly spatially autocorrelated. If independent, error ellipses can be directly applicable to their selected probability distributions. However, if positional uncertainties of individual point features are dependent and spatially autocorrelated, additional stochastic techniques (e.g., a spatial autoregressive model) are required to represent their dependence structure. In particular, the source of point features (e.g., GPS, geocoding, and LiDAR) is the important criterion to select a proper probability distribution and independence between positional uncertainty (Zandbergen, 2008). For example, positional inaccuracy and its empirical distribution of GPS have been regularly reported by the Federal Aviation Administration, and its accuracy is generally reliable and independent regardless of study areas. However, the positional uncertainty of geocoded points varies and is dependent on the locations of geocoded points (Koo et al., 2018c).

Perturbation propagation for positional uncertainty of line features is more complex than that of point features because the former consists of a set of the propagations for individual point features (Shi, 1998; Shi et al., 2014). Similar to error ellipses for point features, the epsilon band is a typical method to represent positional uncertainty of line features (Crosetto and Tarantola, 2001; Shi, 1998), which defines an error band of a boundary using a constant distance for both sides of lines. However, the epsilon band does not represent spatial structures of positional uncertainty between individual point features on a line feature. Specifically, point features that are located relatively midway on a line segment have generally smaller positional uncertainty than that of endpoints (Shi, 1998), and positional uncertainty of individual point features are also often spatially autocorrelated (Tong et al., 2013). Thus, under the assumption that the positional uncertainties at both endpoints are independent, and the amount of positional uncertainty varies by their location, perturbation propagation is offered (Shi, 1998). Furthermore, a stochastic process-based model has been proposed for considering spatial autocorrelation between points (Shi and Liu, 2000). Recently, entropy-based models (Gong and Li, 2011) and the statistical simulation error model (Tong et al., 2013) have offered alternatives for perturbation propagation for positional uncertainty of line features.

5.4. Model uncertainty

Quantification of model uncertainty is challenging and is a subject of ongoing research (O'Hagan, 2012). Additionally, quantitative methods alone cannot address all aspects of model uncertainty as there are qualitative sources that cannot be quantified and arise from the subjective judgment and biases of the modelers (and stakeholders) (Chen et al., 2007). Thus, model uncertainty can also be explained based on qualitative (e.g., expert assessment) rather than quantitative evaluation (Uusitalo et al., 2015). That said, a qualitative evaluation may be sufficient for the model purpose when its evaluators are well-informed. For instance, several solver types are compared by adjusting convergence criteria to evaluate the impact of model solution precision (Ahlfeld and Hoque, 2008). Similarly, different model formulations, resolutions and solvers are explored for model uncertainty investigation (Farthing et al., 2012). Evaluating the uncertainty caused by model integration is also difficult using quantitative methods due to a lack of proper perturbation propagation (Voinov and Shugart, 2013; Tscheikner-Gratl et al., 2019; Voinov and Cerco, 2010).

The resolution of spatial datasets has been used as one source to explore model uncertainty (Trusel et al., 2015). An SD-EM generally uses predefined sources of spatial datasets, for example, SRTM for DEM, and the impact of spatial dataset resolution is often investigated by comparing several spatial datasets with different resolutions. Thus, the resolutions of DEM, LULC and soil datasets can show a significant impact on SD-EM outputs (Chaubey et al., 2005; Dixon and Earls, 2009; Lin et al., 2013c; Kumar and Merwade, 2009). Generally, finer resolutions of spatial datasets lead to more accurate model outputs (Shen et al., 2013) but longer model evaluation time, and furthermore, the uncertainties arising from the resolutions of spatial datasets can be compensated for when different types of spatial datasets are utilized (Shen et al., 2015).

However, Monte Carlo simulation-based SA requires propagating perturbations for individual input factors with their plausible ranges and sometimes probability distribution assumptions. Adding a systematic or random model error at model runtime is suggested to assess perturbation propagation (Marin et al., 2003); however, the method could not differentiate the sources of uncertainties because the added errors include both input and model uncertainties. Adding systematic errors directly in the model structure might be an alternative to adding errors into model runtime. For example, general perturbation propagation is provided for state-parameter estimation based on recursive and batch estimation (Beck, 1987). Another frequently used strategy for model uncertainty exploration is adjusting parameters that relate to model structures (Koo et al., 2020; Yen et al., 2014; Wagener et al., 2003). However, this strategy is not applicable when the model parameters are discrete and have limited options. Recently, Koo et al. (2020) explored model structure uncertainty using SA by adjusting the level of spatial discretization, which could provide another solution for taking model uncertainty into account in SA.

6. SA evaluation and post-processing

6.1. Assessing convergence and credibility

The convergence of SA measures needs to be assessed because SA measures sometimes are not constant and vary with sample sizes, especially when they are obtained from smaller sample sizes than required sizes suggested in the literature (Sarrazin et al., 2016; Vanrolleghem et al., 2015). Two methods are generally utilized to evaluate the convergence of SA measures, which are based on the central limit theorem (CLT) and the bootstrapping technique (Yang, 2011). According to the CLT, the sample mean of a distribution with mean (μ) and standard deviation (σ) approaches a normal distribution with mean μ and standard deviation σ/n with increasing sample size (n). The CLT based-method calculates SA measures R times using different sets of sub-samples, and compares its mean and standard deviation using gradually increasing sizes of sub-samples. The convergence of an SA measure could be regarded as achieved when the coefficient of variation (σ/μ) does not show a significant change. Bootstrapping uses sub-samples from the original samples, and then compares SA measures derived from the sub-samples to the original SA measures. The advantage of the bootstrapping technique is that there is no requirement for additional simulation, but the convergence rate could be overestimated (i.e., underestimation of uncertainty) when the sub-samples are strongly dependent on the original samples. In addition, a convergence test could be performed by analyzing SA measures obtained from different numbers of Monte Carlo simulations (Vanrolleghem et al., 2015).

The number of required samples for a convergence test differs according to the purpose, type(s) of SA applied and the characteristics of the environmental model (Sarrazin et al., 2016). SA for screening purposes generally requires a smaller sample size for a convergence test than that for ranking. Similarly, the required number of samples for convergence is typically the largest in variance-based SA, and significantly smaller for the Morris method and local SA methods (Campolongo et al., 2007). In variance-based SA methods, SA measures for the main effect converge faster than those for the total effect (Sarrazin et al., 2016; Nossent et al., 2011). The characteristics of the environmental models, and its related study area processes and data, also influence convergence rate so that there is no clear relationship between the number of input factors and the required sample size (Sarrazin et al., 2016).

Finally, the reliability and credibility of SA measures should be assessed. The reliability and credibility are obtained by verifying that the underlying assumptions and conditions of the SA are satisfied in a target environmental model (Pianosi et al., 2016). For example, linear regression assumes a linear relationship in model response, which might well be inappropriate for the model with a non-linear response. Additionally, the SA results obtained could be biased due to use of implausible perturbation propagation and missing input factors. Another check is conducting the SA with different SA methods. The reliability of SA results could be regarded as being enhanced if there is a consensus among different methods. If the SA results are contradictory, it encourages further investigations to discover various aspects of model behaviors that are captured from different SA methods (Pappenberger et al., 2008; Paleari and Confalonieri, 2016).

6.2. Visualizing SA measures

Effective visualization methods help to increase understanding and interpretation of SA measures and their relationships to input factors, being especially valuable when SA measures are associated with a large number of input factors. Specifically, visualizing SA results helps in achieving the general purposes of SA by finding and ordering the critical input factors. Visualization also supports the discovery of counterintuitive SA results, which could lead to unearthing new aspects of model behavior, or revising our SA processes, for example, adding missing uncertainty sources and re-propagating perturbation methods. When revising SA processes, the visualization helps to compare SA results under various conditions, such as different ranges and theoretical assumptions of perturbation propagation. Furthermore, in SD-EMs, spatial and temporal patterns in SA measures could be revealed through visualization methods (Pianosi et al., 2016).

Simply, the relationship between input factors and their corresponding outputs are typically visualized using scatter plots, colored scatter plots, and parallel coordinate plots. Specifically, scatter plots demonstrate the relationship between model output with one input (Fig. 3-A), while colored scatter plots typically show the relationship of model output with two input factors (Fig. 3-B). Thus, scatter plots and colored scatter plots are useful for screening and ranking input factors. Parallel coordinate plots show distributions of input factors and outputs (Fig. 3-C), which can highlight patterns using colors and/or dynamic linking and brushing (Ge et al., 2009; Koo et al., 2018b; Symanzik et al., 2000). Violin plots are also useful to visualize the distributions of input factors and outputs (Hintze and Nelson, 1998) whose relationship could be emphasized using dynamic linking and brushings. Further examples of general scientific visualization methods for SA are suggested in Pianosi et al. (2016) and Kelleher and Wagener (2011).

Some SA methods are effectively represented using their own specific visualization methods. For instance, Morris method results are represented using scatter plots of the absolute means of elementary effects against their standard deviations for individual input factors (Fig. 3-D), where both the relative importance of each input factor and their interactions are highlighted. Local SA, and also correlation and regression-based SA methods, utilize general scatter plots of input factors against outputs. Additionally, a regression coefficient plot can be applicable for regression-based SA results (Fig. 3-E). A regression coefficient plot is a scatter plot of an estimated coefficient with lines indicating standard errors, which effectively show the relative importance of individual input factors and compare changes of regression coefficients in SA results under different conditions. Visualization of variance-based SA methods is sometimes difficult because it requires a simultaneous representation of multiple SA measures for main and total effects. A simple and general method for such visualization is using a stacked bar plot for individual input factors with main and total effects (Fig. 3-F). Recently, Circos (Kelleher et al., 2013) and radial convergence diagrams (Butler et al., 2014) were developed to effectively visualize the main and total effects of multiple SA measures.

In an SD-EM, because all the uncertainty sources could make different contributions to the spatially varying outputs, geographical visualization of spatially variable SA measures and their spatial analysis can enhance understanding of spatial aspects in SA measures (Chen et al., 2010). When SA measures are represented in a discrete object approach (e.g., vector datasets), they can be simply displayed by using pie-chart data series in multivariate map compositions (Feick and Hall, 2004) and effectively visualize the relative sensitivities of all associated

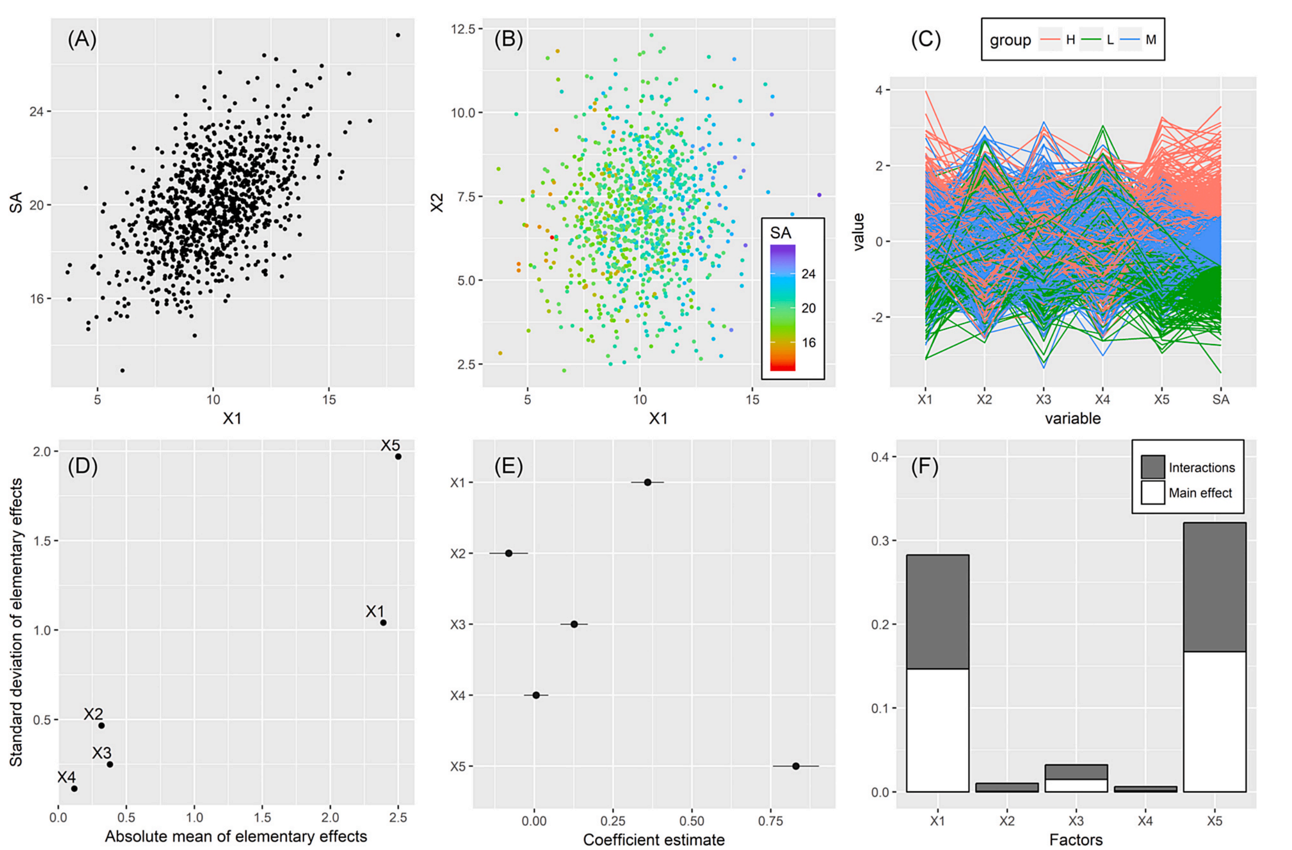


Fig. 3. Visualization methods for SA ($X1$ to $X5$ denotes input factors and SA signals a certain SA measure). (A) Scatter plot, (B) colored scatter plot, (C) parallel coordinate plot, (D) Morris plot, (E) regression coefficient plot, and (F) stacked bar plot.

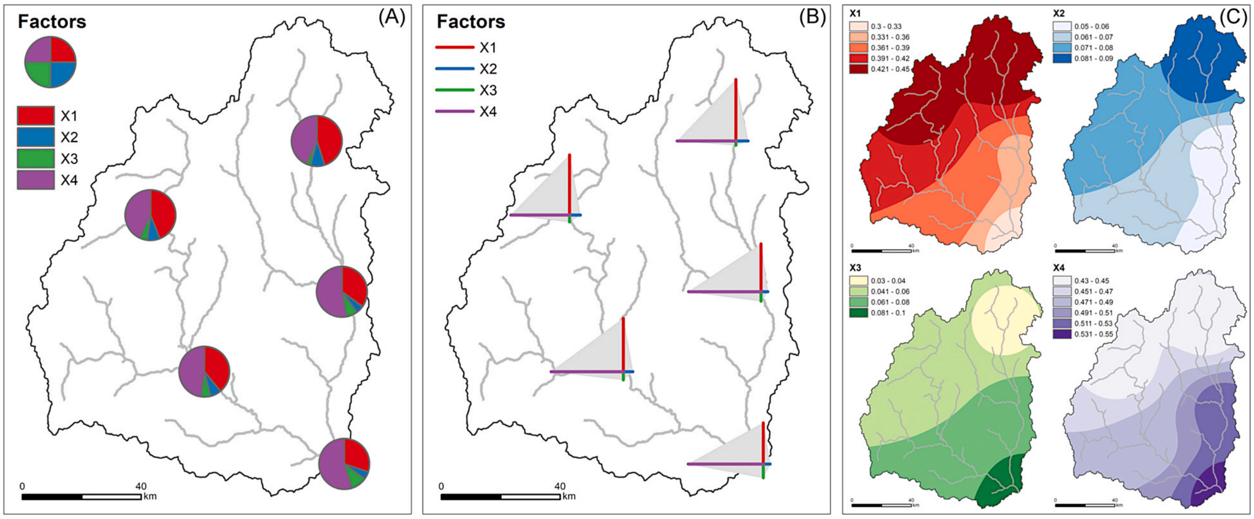


Fig. 4. Geographical visualization of spatially variable SA measures (A) pie-charts on a map, (B) snowflake, and (C) multiple choropleth maps.

input factors at different locations (Fig. 4-A). Additionally, other multivariate mapping techniques, for example, bar chart, ray-glyphs, and snowflake (Fig. 4-B) (Slocum et al., 2009), can also be applicable for geographical visualization of SA measures. If SA measures are associated with a continuous field approach (e.g., raster datasets), using multiple choropleth maps (i.e., small multiple for multivariate) of sensitivities for individual input factors is a useful approach (Chen et al., 2010; Xu and Zhang, 2013) (Fig. 4-C). However, the classification schemes of multiple choropleth maps should be carefully selected to compare input factors on different choropleth maps (Slocum et al., 2009). Furthermore, spatial analysis of SA measures, for example, applying univariate spatial autocorrelation measures and/or semi-variograms for individual input factors, or multivariate spatial autocorrelation measures (Anselin, 2019), may offer further elucidations for the spatial distribution of SA measures.

7. An example of the framework with SWAT

This section illustrates the SA framework applied to a widely used SD-EM, the Soil and Water Assessment Tool (SWAT) (Neitsch et al., 2011; Zhang et al., 2019), based on an extension of the SA applied in (Koo et al., 2020). Following the framework, the first step is uncertainty source identification. The application of SWAT can be divided into three sub-models: the watershed delineation model, the HRU (hydrological response units) generation model for preprocessing, and the SWAT model for the prediction of water quantity and quality.

Most studies focus on the model parameters related to the execution of SWAT, as these are often considered the main sources of uncertainty (Yang et al., 2015). There are, however, other parameters associated with the preprocessing of SWAT submodels to be considered, which may have profound effects on the spatial discretization and resolution (watershed and HRUs) used in the model (Ray, 2018). These preprocessing parameters should be treated as input factors in an SA, which then aids in identifying, for example, minimum percentages of LULC and soil classes in order to eliminate inappropriately small HRUs. Exploring uncertainty in these parameters partially addresses uncertainties in model structure and resolution.

SWAT requires input raster datasets of DEMs, LULC, soil datasets, and vector datasets of meteorological information on monitoring stations, and optionally predefined stream networks. These datasets provide fundamental information for describing the characteristics of an underlying watershed (Shen et al., 2015). As discussed in Section 3.2, DEMs involve both resolution and random and systematic measurement uncertainties. LULC and soil datasets possess resolution and positional uncertainty. Meteorological information, being site-specific, can have positional uncertainty, as well as attribute uncertainty in its measurement. Importantly, measurement uncertainty in both raster and vector datasets should take account of its spatial structure (i.e., spatial autocorrelation). Although stream networks used in SWAT do not include attribute information and its related uncertainty, the scale and positional uncertainty of a stream network would have significant impacts on the scale and shape of a watershed, respectively, and should be considered in SA (Koo et al., 2020).

The second step is selection of the SA method based on the purposes of the SA and the characteristics of the SWAT model. The purpose of SA can be different for different users, but generally SWAT applications include many uncertainty sources (e.g., hundreds of SWAT model parameters). Thus, reducing the number of input factors through screening methods is recommended prior to ranking. Related to the SWAT characteristics, local SA, Pearson’s correlation, and linear regression-based SA methods are inadequate because input factors in SWAT are interdependent and generally interact with one other. Variance-based SA methods are often utilized in SWAT applications (Zadeh et al., 2017). If holistic uncertainty sources in SWAT are to be evaluated through SA then variance-based SA methods may be unsuitable due to their relatively high computational costs (Razavi and Gupta, 2015). Use of the Morris method and rank regression, or implementing an emulator instead of SWAT (Yang et al., 2015) would therefore be recommended.

The third step is perturbation propagation for individual input factors. Plausible ranges and assumed probability distributions for SWAT model parameters can be found in the SWAT model calibration literature (Yang et al., 2018; Abbaspour, 2015). As the DEM consists of a huge number of grids, pixel swapping and a spatial autoregressive model might be undesirable due to their high computational costs. Thus, spatial moving average and sequential Gaussian simulation are suggested approaches for propagate perturbations. For LULC and soil datasets, if their confusion matrices exist, using these matrices would provide credible values to propagate, although they cannot represent the spatial structure of uncertainty. If confusion matrices do not exist, simple epsilon bands or other methods for a line feature type could be applicable.

Meteorological information on monitoring stations and predefined stream networks could utilize perturbation propagation for vector datasets. The perturbation of position in meteorological information can be propagated using error ellipses, and its attributes can be propagated using a spatial autoregressive model. Positional uncertainty of stream networks can be propagated through simple epsilon bands or other methods for a line feature type. If the precision of stream networks is subjected to SA, a stream network can be generated directly from a DEM using a GIS operation, and its impact can be evaluated by adjusting a model parameter for the GIS operation to designate drainage to a stream network (Koo et al., 2020).

The final step is SA evaluation and post-processing. In this example, the Morris method and rank regression were the SA methods applied to the SWAT model (in the second step). Thus, the SA result of the Morris method could be visualized via scatter plots for absolute means of elementary effects against their standard deviations, and rank regression could be represented using scatter plots and regression coefficient plots. If spatially varying SA results are of interest, for example, QoIs at the outlet of sub-watersheds, pie charts could be used to display the proportional influence of input factors at different locations (See Fig. 4-A) (Feick and Hall, 2004). The convergence of the SA measures should additionally be assessed, as should any dependence of sensitivity results on the climatic forcing as it changes through the period on interest.

8. Conclusions

This article presents a pragmatic framework for the application of sensitivity analysis (SA) to a spatially distributed environmental model (SD-EM). The suggested framework for SA consists of four general steps: potential uncertainty source identification, selection of SA method and predictive quantities of interest, perturbation propagation, and SA evaluation and post-processing. This framework also provides useful background and general guidance on applying SA to other areas of environmental modeling and related GIS communities. More specifically, the framework can be used to assist in identifying potential uncertainty sources and their corresponding uncertainty classification scheme, choosing an appropriate SA method according to the SA purpose(s) and model characteristics, propagating perturbations with plausible ranges and assumptions, and verifying SA measures using visualization methods, convergence and reliability tests.

In particular, it provides guidance on the assessment and treatment of uncertainty sources related to spatial datasets, including positional and attribute uncertainty, and on widely utilized perturbation propagation methods that take account of the spatial structure of potential uncertainty in those spatial datasets. The framework includes guidance on methods for analyzing SA results involving multiple outputs and their visualization, which could offer efficient ways to handle spatially distributed SA measures. The framework should, therefore, be helpful in incorporating the uncertainty of spatial components in SD-EMs into a general SA process, along with the usual model parameters and other input factors that SA commonly evaluates. Furthermore, we expect that model structure uncertainty related to the scales, boundaries, and discretization of spatial datasets could be addressed through this framework, and provide concrete support for further uncertainty analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The primary idea of this manuscript was discussed during two workshops in Nanjing (May 2019) and Canberra (December 2019), and the manuscript was drafted through several Zoom meetings. This work was supported by the Key Project of NSF of China (Grant 41930648), the NSF for Excellent Young Scholars of China (Grant 41622108), National Key Research and Development Program of China (Grant 2017YFB0503500), the Australian Government Research Training Program (AGRTP) Scholarship and a top-up scholarship from the ANU Hilda-John Endowment Fund, and the Priority Academic Program Development of Jiangsu Higher Education Institutions (Grant 164320H116).

References

Abbaspour, K.C., 2015. SWAT Calibration and Uncertainty Programs—A User Manual. Swiss Federal Institute of Aquatic Science and Technology: Eawag, Switzerland.
Aerts, J.C., Goodchild, M.F., Heuvelink, G.B., 2003. Accounting for spatial uncertainty in optimization with spatial decision support systems. Trans. GIS 7 (2), 211–230.
Ahlfeld, D.P., Hoque, Y., 2008. Impact of simulation model solver performance on ground water management problems. Ground Water 46 (5), 716–726.
Anselin, L., 1995. Local indicators of spatial association—LISA. Geogr. Anal. 27 (2), 93–115.
Anselin, L., 2005. Spatial Regression Analysis in R: A Workbook. University of Illinois, Urbana, IL.
Anselin, L., 2019. A local indicator of multivariate spatial association: extending Geary's C. Geogr. Anal. 51 (2), 133–150.
Ansi (American National Standards Institute), 1998. Spatial Data Transfer Standard (SDTS) - Part 1. Logical Specifications. ANSI NCITS 320-1998. Washington, DC.
Aouissi, J., Benabdallah, S., Chabaâne, Z.L., Cudennec, C., 2013. Sensitivity analysis of SWAT model to the spatial rainfall distribution and watershed subdivision in streamflow simulations in the Mediterranean context: a case study in the Joumine watershed. In: 5th International Conference on Modeling, Simulation and Applied Optimization (ICMSAO), Hammamet, Tunisia.
Bárdossy, A., Das, T., 2008. Influence of rainfall observation network on model calibration and application. Hydrol. Earth Syst. Sci. 12 (1), 77–89.
Beck, M.B., 1987. Water quality modeling: a review of the analysis of uncertainty. Water Resour. Res. 23 (8), 1393–1442.
Berzaghi, F., Longo, M., Ciais, P., Blake, S., Bretagnolle, F., Vieira, S., Scaranello, M., Scarascia-Mugnozza, G., Doughty, C.E., 2019. Carbon stocks in central African forests enhanced by elephant disturbance. Nat. Geosci. 12 (9), 725–729.
Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J. Hydrol. 249 (1–4), 11–29.
Borgonovo, E., Plischke, E., 2016. Sensitivity analysis: a review of recent advances. Eur. J. Oper. Res. 248 (3), 869–887.
Brockmann, D., Morgenroth, E., 2007. Comparing global sensitivity analysis for a biofilm model for two-step nitrification using the qualitative screening method of Morris or the quantitative variance-based Fourier Amplitude Sensitivity Test (FAST). Water Sci. Technol. 56 (8), 85–93.
Butler, M.P., Reed, P.M., Fisher-Vanden, K., Keller, K., Wagener, T., 2014. Identifying parametric controls and dependencies in integrated assessment models using global sensitivity analysis. Environ. Model. Software 59, 10–29.
Campolongo, F., Cariboni, J., Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. Environ. Model. Software 22 (10), 1509–1518.
Campolongo, F., Saltelli, A., 2000. Comparing different sensitivity analysis methods on a chemical reactions model. In: Saltelli, A., Chan, K., Scott, E.M. (Eds.), Sensitivity Analysis. John Wiley & Sons, Chichester, England, pp. 355–365.
Campolongo, F., Saltelli, A., Cariboni, J., 2011. From screening to quantitative sensitivity analysis. A unified approach. Comput. Phys. Commun. 182 (4), 978–988.
Chaplot, V., Saleh, A., Jaynes, D.B., 2005. Effect of the accuracy of spatial rainfall information on the modeling of water, sediment, and NO3-N loads at the watershed level. J. Hydrol. 312 (1–4), 223–234.
Chaubey, I., Cotter, A.S., Costello, T.A., Soerens, T.S., 2005. Effect of DEM data resolution on SWAT output. Hydrol. Process. 19 (3), 621–628.
Chen, C.F., Ma, H.W., Reckhow, K.H., 2007. Assessment of water quality management with a systematic qualitative uncertainty analysis. Sci. Total Environ. 374 (1), 13–25.
Chen, M., Voinov, A., Ames, D.P., Kettner, A.J., Goodall, J.L., Jakeman, A.J., Barton, M. C., Harpham, Q., Cuddy, S.M., DeLuca, C., Yue, S., Wang, J., Zhang, F., Wen, Y., Lü, G., 2020. Position paper: open web-distributed integrated geographic modelling and simulation to enable broader participation and applications. Earth Sci. Rev. 207, 103223.
Chen, M., Yue, S., Lü, G., Lin, H., Yang, C., Wen, Y., Hou, T., Xiao, D., Jiang, H., 2019. Teamwork-oriented integrated modeling method for geo-problem solving. Environ. Model. Software 119, 111–123.
Chen, Y., Yu, J., Khan, S., 2010. Spatial sensitivity analysis of multi-criteria weights in GIS-based land suitability evaluation. Environ. Model. Software 25 (12), 1582–1591.
Cho, J., Bosch, D., Lowrance, R., Strickland, T., Vellidis, G., 2009. Effect of spatial distribution of rainfall on temporal and spatial uncertainty of SWAT output. Trans. ASABE (Am. Soc. Agric. Biol. Eng.) 52 (5), 1545–1556.
Chrisman, N.R., 1991. The error component in spatial data. In: Maguire, D.J., Goodchild, M.F., Rhind, D.W. (Eds.), Geographical Information Systems: Principles and Applications, vol. 1. Longman Scientific and Technical, Harlow, pp. 165–174, 1 (12).
Comber, A., Fisher, P., Brunsdon, C., Khmag, A., 2012. Spatial analysis of remote sensing image classification accuracy. Remote Sens. Environ. 127, 237–246.
Crosetto, M., Tarantola, S., 2001. Uncertainty and sensitivity analysis: tools for GIS-based model implementation. Int. J. Geogr. Inf. Sci. 15 (5), 415–437.
Cukier, R.I., Fortuin, C.M., Shuler, K.E., Petschek, A.G., Schaibly, J.H., 1973. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. J. Chem. Phys. 59 (8), 3873–3878.
Deutsch, C.V., Journel, A.G., 1998. GSLIB Geostatistical Software Library and User's Guide, second ed. Oxford University Press.

Devillers, R., Stein, A., Bédard, Y., Chrisman, N., Fisher, P., Shi, W., 2010. Thirty years of research on spatial data quality: achievements, failures, and opportunities. Trans. GIS 14 (4), 387–400.
Dixon, B., Earls, J., 2009. Resample or not?! Effects of resolution of DEMs in watershed modeling. Hydrol. Process. 23 (12), 1714–1724.
Douglas-Smith, D., Iwanaga, T., Croke, B.F., Jakeman, A.J., 2020. Certain trends in uncertainty and sensitivity analysis: an overview of software tools and techniques. Environ. Model. Software 124, 104588.
Dutton, G., 1992. Handling positional uncertainty in spatial databases. In: Proceedings 5th International Symposium on Spatial Data Handling, pp. 460–469.
Ehlschlaeger, C.R., Shortridge, A.M., Goodchild, M.F., 1997. Visualizing spatial data uncertainty using animation. Comput. Geosci. 23 (4), 387–395.
Farmer, W.H., Vogel, R.M., 2016. On the deterministic and stochastic use of hydrologic models. Water Resour. Res. 52 (7), 5619–5633.
Farthing, M.W., Fowler, K.R., Fu, X., Davis, A., Miller, C.T., 2012. Effects of model resolution on optimal design of subsurface flow and transport problems. Adv. Water Resour. 38, 27–37.
Feick, R., Hall, B., 2004. A method for examining the spatial dimension of multi-criteria weight sensitivity. Int. J. Geogr. Inf. Sci. 18 (8), 815–840.
Fisher, P.F., 1991a. First experiments in viewshed uncertainty: the accuracy of the viewshed area. Photogramm. Eng. Rem. Sens. 57 (10), 1321–1327.
Fisher, P.F., 1991b. Modelling soil map-unit inclusions by Monte Carlo simulation. Int. J. Geogr. Inf. 5 (2), 193–208.
Fisher, P., 1998. Improved modeling of elevation error with geostatistics. GeoInformatica 2 (3), 215–233.
Futter, M.N., Butterfield, D., Cosby, B.J., Dillon, P.J., Wade, A.J., Whitehead, P.G., 2007. Modeling the mechanisms that control in-stream dissolved organic carbon dynamics in upland and forested catchments. Water Resour. Res. 43 (2), 1–16.
Gan, Y., Duan, Q., Gong, W., Tong, C., Sun, Y., Chu, W., Ye, A., Miao, C., Di, Z., 2014. A comprehensive evaluation of various sensitivity analysis methods: a case study with a hydrological model. Environ. Model. Software 51, 269–285.
Ge, Y., Li, S., Lakhan, V.C., Lucieer, A., 2009. Exploring uncertainty in remotely sensed data with parallel coordinate plots. Int. J. Appl. Earth Obs. Geoinf. 11 (6), 413–422.
Gómez-Dans, J.L., Lewis, P.E., Disney, M., 2016. Efficient emulation of radiative transfer codes using Gaussian processes and application to land surface parameter inferences. Rem. Sens. 8, 1–32.
Gong, J., Li, D., 2011. Entropy-based models for positional uncertainty of line segments in GIS. Surv. Rev. 43 (322), 390–401.
Gong, Y., Shen, Z., Liu, R., Hong, Q., Wu, X., 2012. A comparison of single and muti-gauge based calibrations for hydrological modelling of the Upper Daning River Watershed in China's Three Gorges Reservoir Region. Hydrol. Res. 43 (6), 822–832.
Goodchild, M.F., Openshaw, S., 1980. Algorithm 9: simulation of autocorrelation for aggregate data. Environ. Plann. 12 (9), 1073–1081.
Goodchild, M., 1991. Issues of quality and uncertainty. In: Müller, J.C. (Ed.), In Advances in Cartography. Elsevier Science Publication Ltd, London and New York, pp. 113–139.
Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation (Applied Geostatistics Series). Oxford University Press, New York.
Griffith, D.A., 2008. Geographic sampling of urban soils for contaminant mapping: how many samples and from where. Environ. Geochem. Health 30 (6), 495–509.
Griffith, D.A., Millones, M., Vincent, M., Johnson, D.L., Hunt, A., 2007. Impacts of positional error on spatial regression analysis: a case study of address locations in Syracuse, New York. Trans. GIS 11 (5), 655–679.
Guillaume, J.H., Jakeman, J.D., Marsili-Libelli, S., Asher, M., Brunner, P., Croke, B., Hill, M.C., Jakeman, A.J., Keesman, K.J., Razavi, S., Stigter, J.D., 2019. Introductory overview of identifiability analysis: a guide to evaluating whether you have the right type of data for your modeling purpose. Environ. Model. Software 119, 418–432.
Gupta, Hoshin V., Razavi, Saman, 2018. Revisiting the basis of sensitivity analysis for dynamical earth system models. Water Resources Res. 54 (11), 8692–8717. https://doi.org/10.1029/2018WR022668.
Helton, J.C., 1993. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. Reliab. Eng. Syst. Saf. 42 (2–3), 327–367.
Helton, J.C., Johnson, J.D., Sallaberry, C.J., Storlie, C.B., 2006. Survey of sampling-based methods for uncertainty and sensitivity analysis. Reliab. Eng. Syst. Saf. 91 (10–11), 1175–1209.
Hengl, T., Heuvelink, G.B.M., Van Loon, E.E., 2010. On the uncertainty of stream networks derived from elevation data: the error propagation approach. Hydrol. Earth Syst. Sci. 14 (7), 1153–1165.
Herman, J.D., Kollat, J.B., Reed, P.M., Wagener, T., 2013. Technical Note: method of Morris effectively reduces the computational demands of global sensitivity analysis for distributed watershed models. Hydrol. Earth Syst. Sci. 17 (7), 2893–2903.
Heuvelink, G.B., 1998. Error Propagation in Environmental Modelling with GIS. Taylor & Francis, London, UK.
Hintze, J.L., Nelson, R.D., 1998. Violin plots: a box plot-density trace synergism. Am. Statistician 52 (2), 181–184.
Holmes, K.W., Chadwick, O.A., Kyriakidis, P.C., 2000. Error in a USGS 30-meter digital elevation model and its impact on terrain modeling. J. Hydrol. 233 (1–4), 154–173.
Hunter, G.J., Goodchild, M.F., 1996. A new model for handling vector data uncertainty in geographic information systems. J. Urban Reg. Inf. Syst. Assoc. 8, 51–57.
Iman, R.L., Helton, J.C., 1988. An investigation of uncertainty and sensitivity analysis techniques for computer models. Risk Anal. 8 (1), 71–90.
Jakeman, A.J., Letcher, R.A., Norton, J.P., 2006. Ten iterative steps in development and evaluation of environmental models. Environ. Model. Software 21 (5), 602–614.
Karimi, H.A., Durcik, M., Rasdorf, W., 2004. Evaluation of uncertainties associated with geocoding techniques. Comput. Civ. Infrastruct. Eng. 19 (3), 170–185.
Kelleher, C., Wagener, T., 2011. Ten guidelines for effective data visualization in scientific publications. Environ. Model. Software 26 (6), 822–827.
Kelleher, C., Wagener, T., McGlynn, B., Ward, A.S., Gooseff, M.N., Payn, R.A., 2013. Identifiability of transient storage model parameters along a mountain stream. Water Resour. Res. 49 (9), 5290–5306.
Kiiveri, H.T., 1997. Assessing, representing and transmitting positional uncertainty in maps. Int. J. Geogr. Inf. Sci. 11 (1), 33–52.
Koo, H., Chen, M., Jakeman, A.J., Zhang, F., 2020. A global sensitivity analysis approach for identifying critical sources of uncertainty in non-identifiable spatially distributed environmental models: a holistic analysis applied to SWAT for input datasets and model parameters. Environ. Model. Software 127.
Koo, H., Chun, Y., Griffith, D.A., 2018a. Geovisualizing attribute uncertainty of interval and ratio variables: a framework and an implementation for vector data. J. Vis. Lang. Comput. 44, 89–96.
Koo, H., Chun, Y., Griffith, D.A., 2018b. Integrating spatial data analysis functionalities in a GIS environment: spatial Analysis using ArcGIS Engine and R (SAAR). Trans. GIS 22 (3), 721–736.
Koo, H., Chun, Y., Griffith, D.A., 2018c. Modeling positional uncertainty acquired through street geocoding. Int. J. Appl. Geospatial Res. (IJAGR) 9 (4), 1–22.
Koo, H., Wong, D.W., Chun, Y., 2019. Measuring global spatial autocorrelation with data reliability information. Prof. Geogr. 71 (3), 551–565.
Kumar, S., Merwade, V., 2009. Impact of watershed subdivision and soil data resolution on SWAT model calibration and parameter uncertainty. J. Am. Water Resour. Assoc. 45 (5), 1179–1196.
Legendre, P., 1993. Spatial autocorrelation: problem or new paradigm. Ecology 74 (6), 1659–1673.
Li, X., 2014. Characterization, controlling, and reduction of uncertainties in the modeling and observation of land-surface systems. Sci. China Earth Sci. 57 (1), 80–87.
Li, X., Liu, F., Fang, M., 2020. Harmonizing models and observations: data assimilation for Earth System Science. Sci. China Earth Sci. 63, 1059–1068.
Lin, H., Chen, M., Lv, G.N., 2013a. Virtual geographic environment: a workspace for computer-aided geographic experiments. Ann. Assoc. Am. Geogr. 103 (3), 465–482.
Lin, H., Chen, M., Lu, G.N., Zhu, Q., Gong, J.H., You, X., Wen, Y.N., Xu, B.L., Hu, M.Y., 2013b. Virtual geographic environments (VGEs): a new generation of geographic analysis tool. Earth Sci. Rev. 126, 74–84.
Lin, S., Jing, C., Coles, N.A., Chaplot, V., Moore, N.J., Wu, J., 2013c. Evaluating DEM source and resolution uncertainties in the soil and water assessment tool. Stoch. Environ. Res. Risk Assess. 27 (1), 209–221.
Linkov, I., Burmistrov, D., 2003. Model uncertainty and choices made by modelers: lessons learned from the international atomic energy agency model intercomparisons. Risk Anal. 23 (6), 1297–1308.
Littlewood, I.G., Croke, B.F., 2013. Effects of data time-step on the accuracy of calibrated rainfall - stream flow model parameters: practical aspects of uncertainty reduction. Hydrol. Res. 44 (3), 430–440.
Lu, G.N., Batty, M., Strobl, J., Lin, H., Zhu, A.X., Chen, M., 2019. Reflections and speculations on the progress in Geographic Information Systems (GIS): a geographic perspective. Int. J. Geogr. Inf. Sci. 33 (2), 346–367.
Lucieer, A., Kraak, M.J., 2004. Interactive and visual fuzzy classification of remotely sensed imagery for exploration of uncertainty. Int. J. Geogr. Inf. Sci. 18 (5), 491–512.
Mara, T.A., Tarantola, S., 2008. Application of global sensitivity analysis of model output to building thermal simulations. Build. Simulat. 1 (4), 290–302.
Marin, C.M., Guvanasen, V., Saleem, Z.A., 2003. The 3MRA risk assessment framework—a flexible approach for performing multimedia, multipathway, and multireceptor risk assessments under uncertainty. Hum. Ecol. Risk Assess. 9 (7), 1655–1677.
Masih, I., Maskey, S., Uhlenbrook, S., Smakhtin, V., 2011. Assessing the impact of areal precipitation input on streamflow simulations using the SWAT model. J. Am. Water Resour. Assoc. 47 (1), 179–195.
Matott, L.S., Babendreier, J.E., Purucker, S.T., 2009. Evaluating uncertainty in integrated environmental models: a review of concepts and tools. Water Resour. Res. 45 (6), 1–14.
Minunno, F., van Oijen, M., Cameron, D.R., Pereira, J.S., 2013. Selecting parameters for Bayesian calibration of a process-based model: a methodology based on canonical correlation analysis. SIAM/ASA J. Uncertain. Quantification 1 (1), 370–385.
Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments. Technometrics 33 (2), 161–174.
Mukul, M., Srivastava, V., Jade, S., Mukul, M., 2017. Uncertainties in the Shuttle radar Topography mission (SRTM) heights: insights from the Indian Himalaya and Peninsula. Sci. Rep. 7, 1–10.
Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R., 2011. Soil and Water Assessment Tool Theoretical Documentation Version 2009. Texas Water Resources Institute.
Newham, L.T.H., Norton, J.P., Prosser, I.P., Croke, B.F.W., Jakeman, A.J., 2003. Sensitivity analysis for assessing the behaviour of a landscape-based sediment source and transport model. Environ. Model. Software 18 (8–9), 741–751.
Norton, J., 2015. An introduction to sensitivity assessment of simulation models. Environ. Model. Software 69, 166–174.
Nossent, J., Elsen, P., Bauwens, W., 2011. Sobol'sensitivity analysis of a complex environmental model. Environ. Model. Software 26 (12), 1515–1525.
O'Hagan, A., 2012. Probabilistic uncertainty specification: overview, elaboration techniques and their application to a mechanistic model of carbon flux. Environ. Model. Software 36, 35–48.
Oakley, J.E., O'Hagan, A., 2004. Probabilistic sensitivity analysis of complex models: a Bayesian approach. J. R. Stat. Soc. Ser. B. Stat. Methodol. 66 (3), 751–769.
Oksanen, J., Sarjakoski, T., 2005a. Error propagation analysis of DEM-based drainage basin delineation. Int. J. Rem. Sens. 26 (14), 3085–3102.

Oksanen, J., Sarjakoski, T., 2005b. Error propagation of DEM-based surface derivatives. Comput. Geosci. 31 (8), 1015–1027.
Palar, P.S., Zuhal, L.R., Shimoyama, K., Tsuchiya, T., 2018. Global sensitivity analysis via multi-fidelity polynomial chaos expansion. Reliab. Eng. Syst. Saf. 170, 175–190.
Paleari, L., Confalonieri, R., 2016. Sensitivity analysis of a sensitivity analysis: we are likely overlooking the impact of distributional assumptions. Ecol. Model. 340, 57–63.
Pappenberger, F., Beven, K.J., Ratto, M., Matgen, P., 2008. Multi-method global sensitivity analysis of flood inundation models. Adv. Water Resour. 31 (1), 1–14.
Pastres, R., Chan, K., Solidoro, C., Dejak, C., 1999. Global sensitivity analysis of a shallow-water 3D eutrophication model. Comput. Phys. Commun. 117 (1–2), 62–74.
Pebesma, E.J., de Jong, K., Briggs, D., 2007. Interactive visualization of uncertain spatial and spatio-temporal data under different scenarios: an air quality example. Int. J. Geogr. Inf. Sci. 21 (5), 515–527.
Pianosi, F., Beven, K., Freer, J., Hall, J.W., Rougier, J., Stephenson, D.B., Wagener, T., 2016. Sensitivity analysis of environmental models: a systematic review with practical workflow. Environ. Model. Software 79, 214–232.
Pianosi, F., Wagener, T., 2015. A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. Environ. Model. Software 67, 1–11.
Porada, P., Van Stan, J.T., Kleidon, A., 2018. Significant contribution of non-vascular vegetation to global rainfall interception. Nat. Geosci. 11 (8), 563–567.
Razavi, S., Gupta, H.V., 2015. What do we mean by sensitivity analysis? The need for comprehensive characterization of "global" sensitivity in Earth and Environmental systems models. Water Resour. Res. 51, 3070–3092.
Ray, L.K., 2018. Limitation of automatic watershed delineation tools in coastal region. Ann. GIS 24 (4), 261–274.
Razavi, Saman, Gupta, Hoshin V., 2019. A multi-method Generalized Global Sensitivity Matrix approach to accounting for the dynamical nature of earth and environmental systems models. Environ. Modell. Softw. 114, 1–11. https://doi.org/10.1016/j.envsoft.2018.12.002.
Refsgaard, J.C., van der Sluijs, J.P., Højberg, A.L., Vanrolleghem, P.A., 2007. Uncertainty in the environmental modelling process - a framework and guidance. Environ. Model. Software 22 (11), 1543–1556.
Ryan, E., Wild, O., Voulgarakis, A., Lee, L., 2018. Fast sensitivity analysis methods for computationally expensive models with multi-dimensional output. Geosci. Model Dev. (GMD) 11 (8), 3131–3146.
Saltelli, A., 2002. Sensitivity analysis for importance assessment. Risk Anal. 22 (3), 579–590.
Saltelli, A., Annoni, P., 2010. How to avoid a perfunctory sensitivity analysis. Environ. Model. Software 25 (12), 1508–1517.
Saltelli, A., Marivoet, J., 1990. Non-parametric statistics in sensitivity analysis for model output: a comparison of selected techniques. Reliab. Eng. Syst. Saf. 28 (2), 229–253.
Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. Global Sensitivity Analysis: the Primer. John Wiley & Sons, Chichester, England.
Saltelli, A., Tarantola, S., 2002. On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal. J. Am. Stat. Assoc. 97 (459), 702–709.
Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., 2004. Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models. Wiley Online Library, Chichester, England.
Saltelli, A., Tarantola, S., Chan, K.S., 1999. A quantitative model-independent method for global sensitivity analysis of model output. Technometrics 41 (1), 39–56.
Sarrazin, F., Pianosi, F., Wagener, T., 2016. Global sensitivity analysis of environmental models: convergence and validation. Environ. Model. Software 79, 135–152.
Setegn, S.G., Srinivasan, R., Melesse, A.M., Dargahi, B., 2010. SWAT model application and prediction uncertainty analysis in the Lake Tana Basin, Ethiopia. Hydrol. Process. 24 (3), 357–367.
Sheikholeslami, Razi, Ravazi, Saman, Gupta, Hoshin V., Becker, William, Haghnegahdar, Amin, 2019. Global sensitivity analysis for high-dimensional problems: How to objectively group factors and measure robustness and convergence while reducing computational cost. Environ. Modell. Softw. 111, 282–299. https://doi.org/10.1016/j.envsoft.2018.09.002.
Shen, Z.Y., Chen, L., Liao, Q., 2015. Effect of rainfall measurement errors on nonpoint-source pollution model uncertainty. J. Environ. Inf. 26 (1), 14–26.
Shen, Z.Y., Chen, L., Liao, Q., Liu, R.M., Huang, Q., 2013. A comprehensive study of the effect of GIS data on hydrology and non-point source pollution modeling. Agric. Water Manag. 118, 93–102.
Shi, W., 1998. A generic statistical approach for modelling error of geometric features in GIS. Int. J. Geogr. Inf. Sci. 12 (2), 131–143.
Shi, W., Liu, W., 2000. A stochastic process-based model for the positional error of line segments in GIS. Int. J. Geogr. Inf. Sci. 14 (1), 51–66.
Shi, W., You, Y., Shao, P., Wan, Y., 2014. Standard deviation of line objects in geographic information science. Ann. GIS 20 (1), 39–51.
Shin, M.J., Guillaume, J.H., Croke, B.F., Jakeman, A.J., 2013. Addressing ten questions about conceptual rainfall-runoff models with global sensitivity analyses in R. J. Hydrol. 503, 135–152.
Shin, M.J., Guillaume, J.H., Croke, B.F., Jakeman, A.J., 2015. A review of foundational methods for checking the structural identifiability of models: results for rainfall-runoff. J. Hydrol. 520, 1–16.
Singh, R., Wagener, T., Crane, R., Mann, M.E., Ning, L., 2014. A vulnerability driven approach to identify adverse climate and land use change combinations for critical hydrologic indicator thresholds: application to a watershed in Pennsylvania, USA. Water Resour. Res. 50 (4), 3409–3427.
Slocum, T.A., McMaster, R.B., Kessler, F.C., Howard, H.H., 2009. Thematic Cartography and Geovisualization, third ed. Prentice Hall, Upper Saddle River, New Jersey.
Sobie, E.A., 2009. Parameter sensitivity analysis in electrophysiological models using multivariable regression. Biophys. J. 96 (4), 1264–1274.
Sobol, I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math. Comput. Simulat. 55 (1–3), 271–280.
Spear, R., Hornberger, G., 1980. Eutrophication in peel inlet. II. Identification of critical uncertainties via generalized sensitivity analysis. Water Res. 14 (1), 43–49.
Stanislawski, L.V., Dewitt, B.A., Shrestha, R.L., 1996. Estimating positional accuracy of data layers within a GIS through error propagation. Photogramm. Eng. Rem. Sens. 62 (4), 429–433.
Storlie, C.B., Swiler, L.P., Helton, J.C., Sallaberry, C.J., 2009. Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. Reliab. Eng. Syst. Saf. 94 (11), 1735–1763.
Strauch, M., Bernhofer, C., Koide, S., Volk, M., Lorz, C., Makeschin, F., 2012. Using precipitation data ensemble for uncertainty analysis in SWAT streamflow simulation. J. Hydrol. 414–415, 413–424.
Sudret, B., 2008. Global sensitivity analysis using polynomial chaos expansions. Reliab. Eng. Syst. Saf. 93 (7), 964–979.
Sun, X.Y., Newham, L.T., Croke, B.F., Norton, J.P., 2012. Three complementary methods for sensitivity analysis of a water quality model. Environ. Model. Software 37, 19–29.
Symanzik, J., Cook, D., Lewin-Koh, N., Majure, J.J., Megretskaia, I., 2000. Linking ArcView™ and XGobi: insight behind the front end. J. Comput. Graph Stat. 9 (3), 470–490.
Tasdighi, A., Arabi, M., Harmel, D., Line, D., 2018. A Bayesian total uncertainty analysis framework for assessment of management practices using watershed models. Environ. Model. Software 108, 240–252.
Temme, A.J.A.M., Heuvelink, G.B.M., Schoorl, J.M., Claessens, L., 2009. Geostatistical simulation and error propagation in geomorphometry. Dev. Soil Sci. 33, 121–140.
Tong, X., Sun, T., Fan, J., Goodchild, M.F., Shi, W., 2013. A statistical simulation model for positional error of line features in Geographic Information Systems (GIS). Int. J. Appl. Earth Obs. Geoinf. 21, 136–148.
Tran, Q.Q., De Niel, J., Willems, P., 2018. Spatially distributed conceptual hydrological model building: a generic top-down approach starting from lumped models. Water Resour. Res. 54 (10), 8064–8085.
Trusel, L.D., Frey, K.E., Das, S.B., Karnauskas, K.B., Munneke, P.K., Van Meijgaard, E., Van Den Broeke, M.R., 2015. Divergent trajectories of Antarctic surface melt under two twenty-first-century climate scenarios. Nat. Geosci. 8 (12), 927–932.
Tscheikner-Gratl, F., Bellos, V., Schellart, A., Moreno-Rodenas, A., Muthusamy, M., Langeveld, J., Clemens, F., Benedetti, L., Rico-Ramirez, M.A., Carvalho, R.F., Breuer, L., Shucksmith, J., Heuvelink, G.B.M., Tait, S., 2019. Recent insights on uncertainties present in integrated catchment water quality modelling. Water Res. 150, 368–379.
Uusitalo, L., Lehikoinen, A., Helle, I., Myrberg, K., 2015. An overview of methods to evaluate uncertainty of deterministic models in decision support. Environ. Model. Software 63, 24–31.
Vanrolleghem, P.A., Mannina, G., Cosenza, A., Neumann, M.B., 2015. Global sensitivity analysis for urban water quality modelling: terminology, convergence and comparison of different methods. J. Hydrol. 522, 339–352.
Voinov, A., Cerco, C., 2010. Model integration and the role of data. Environ. Model. Software 25 (8), 965–969.
Voinov, A., Shugart, H.H., 2013. "Integronsters", integral and integrated modeling. Environ. Model. Software 39, 149–158.
Wagener, T., McIntyre, N., Lees, M.J., Wheater, H.S., Gupta, H.V., 2003. Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. Hydrol. Process. 17 (2), 455–476.
Wagener, T., Pianosi, F., 2019. What has Global Sensitivity Analysis ever done for us? A systematic review to support scientific advancement and to inform policy-making in earth system modelling. Earth Sci. Rev. 194, 1–18.
Wang, J., Li, X., Lu, L., Fang, F., 2013. Parameter sensitivity analysis of crop growth models based on the extended Fourier Amplitude Sensitivity Test method. Environ. Model. Software 48, 171–182.
Wechsler, S.P., 2007. Uncertainties associated with digital elevation models for hydrologic applications: a review. Hydrol. Earth Syst. Sci. Discuss. 3 (4), 1481–1500.
Wechsler, S.P., Kroll, C.N., 2006. Quantifying DEM uncertainty and its effect on topographic parameters. Photogramm. Eng. Rem. Sens. 72 (9), 1081–1090.
Whitehead, P.G., Bussi, G., Bowes, M.J., Read, D.S., Hutchins, M.G., Elliott, J.A., Dadson, S.J., 2015. Dynamic modelling of multiple phytoplankton groups in rivers with an application to the Thames River system in the UK. Environ. Model. Software 74, 75–91.
Whitehead, P.G., Hornberger, G.M., 1984. Modelling algal behaviour in the river thames. Water Res. 18 (8), 945–953.
Wolf, P.R., Ghilani, C.D., 1997. Adjustment Computations: Statistics and Least Squares in Surveying and GIS, third ed. John Wiley & Sons, New York.
Wu, H., Chen, B., 2015. Evaluating uncertainty estimates in distributed hydrological modeling for the Wenjing River watershed in China by GLUE, SUFI-2, and ParaSol methods. Ecol. Eng. 76, 110–121.
Wu, S., Li, J., Huang, G.H., 2008. Characterization and evaluation of elevation data uncertainty in water resources modeling with GIS. Water Resour. Manag. 22 (8), 959–972.
Wu, Y., Liu, S., 2012. Automating calibration, sensitivity and uncertainty analysis of complex models using the R package Flexible Modeling Environment (FME): SWAT as an example. Environ. Model. Software 31, 99–109.
Xiu, D., Karniadakis, G.E., 2003. Modeling uncertainty in flow simulations via generalized polynomial chaos. J. Comput. Phys. 187 (1), 137–167.

Xu, E., Zhang, H., 2013. Spatially-explicit sensitivity analysis for land suitability evaluation. Appl. Geogr. 45, 1–9.
Yang, J., 2011. Convergence and uncertainty analyses in Monte-Carlo based sensitivity analysis. Environ. Model. Software 26 (4), 444–457.
Yang, J., Jakeman, A., Fang, G., Chen, X., 2018. Uncertainty analysis of a semi-distributed hydrologic model based on a Gaussian Process emulator. Environ. Model. Software 101, 289–300.
Yen, H., Wang, X., Fontane, D.G., Harmel, R.D., Arabi, M., 2014. A framework for propagation of uncertainty contributed by parameterization, input data, model structure, and calibration/validation data in watershed modeling. Environ. Model. Software 54, 211–221.
Young, P.C., Ratto, M., 2011. Statistical emulation of large linear dynamic models. Technometrics 53 (1), 29–43.
Yue, S., Chen, M., Song, J., Yuan, W., Chen, T., Lü, G., Chen, C., Ma, Z., Xu, K., Wen, Y., Song, H., 2020. Participatory intercomparison strategy for terrestrial carbon cycle models based on a service-oriented architecture. Future Generat. Comput. Syst. 112, 449–466.
Zadeh, F.K., Nossent, J., Sarrazin, F., Pianosi, F., van Griensven, A., Wagener, T., Bauwens, W., 2017. Comparison of variance-based and moment-independent global sensitivity analysis approaches by application to the SWAT model. Environ. Model. Software 91, 210–222.
Zandbergen, P.A., 2008. Positional accuracy of spatial data: non-normal distributions and a critique of the national standard for spatial data accuracy. Trans. GIS 12 (1), 103–130.
Zhang, F., Chen, M., Ames, D.P., Shen, C., Yue, S., Wen, Y., Lü, G., 2019. Design and development of a service-oriented wrapper system for sharing and reusing distributed geoanalysis models on the web. Environ. Model. Software 111, 498–509.
Zhao, F., Wu, Y., Qiu, L., Sun, Y., Sun, L., Li, Q., Niu, J., Wang, G., 2018. Parameter uncertainty analysis of the SWAT model in a mountain-loess transitional watershed on the Chinese Loess Plateau. Water (Switzerland) 10 (6), 1–16.
Zimmerman, D.L., Fang, X., Mazumdar, S., Rushton, G., 2007. Modeling the probability distribution of positional errors incurred by residential address geocoding. Int. J. Health Geogr. 6 (1), 1.

Appendix 1: Statement of Contributions

This thesis has been approved for submission as a Thesis by Compilation under ANU Guideline 266/2013, as posted at https://policies.anu.edu.au/ppl/document/ANUP_003405. For each constituent publication, the title, authorship, publication outlet, current status and extent of contribution are given below.

Australian National University

Statement of Contribution

This thesis is submitted as a Thesis by Compilation in accordance with https://policies.anu.edu.au/ppl/document/ANUP_003405

I declare that the research presented in this Thesis represents original work that I carried out during my candidature at the Australian National University, except for contributions to multi-author papers incorporated in the Thesis where my contributions are specified in this Statement of Contribution.

Title: Certain trends in uncertainty and sensitivity analysis: An overview of software tools and techniques

Authors: Douglas-Smith, D., **Iwanaga, T.*,** Croke, B.F.W., Jakeman, A.J.

Publication outlet: Environmental Modelling & Software

Current status of paper: Published

Contribution to paper: Conceptualisation, new software development, supporting analysis, writing, and editing by Takuya Iwanaga. Majority of initial draft written by Takuya Iwanaga. Further analysis, writing and editing by Dominique Douglas-Smith. Further document structure and editing by Barry F.W. Croke and Anthony J. Jakeman.

Senior author or collaborating authors endorsement: DDS. (Miss Dominique Douglas-Smith)

Candidate – Print Name	Signature	Date

Endorsed

Anthony J. Jakeman	A. J. Jakeman	
Primary Supervisor – Print Name	Signature	Date
Philip Gibbons	Philip Gibbons	21/6/2021
Delegated Authority – Print Name	Signature	Date

Australian National University

Statement of Contribution

This thesis is submitted as a Thesis by Compilation in accordance with https://policies.anu.edu.au/ppl/document/ANUP_003405

I declare that the research presented in this Thesis represents original work that I carried out during my candidature at the Australian National University, except for contributions to multi-author papers incorporated in the Thesis where my contributions are specified in this Statement of Contribution.

Title: Development of an integrated model for the Campaspe catchment: a tool to help improve understanding of the interaction between society, policy, farming decision, ecology, hydrology and climate

Authors: **Iwanaga, T.***, Zare, F., Croke, B., Fu, B., Merritt, W., Partington, D., Ticehurst, J., Jakeman, A.

Publication outlet: Proceedings of the International Association of Hydrological Sciences

Current status of paper: Published

Contribution to paper: Conceptualisation and writing by Iwanaga and Croke. Additional writing and editing by Merritt, Partington, Fu, Ticehurst and Jakeman. Iwanaga developed the (software) framework to enable model integration, led model development and integration liaising with component model developers where necessary. Sensitivity analysis and model testing to ensure model quality also by Iwanaga.

Senior author or collaborating authors endorsement: (Prof. Barry F. W. Croke)

Takuya Iwanaga		30/12/2020
Candidate – Print Name	Signature	Date

Endorsed

Anthony J. Jakeman	A. J. Jakeman	30/12/2020
Primary Supervisor – Print Name	Signature	Date
Philip Gibbons	Philip Gibbons	21/06/2021
Delegated Authority – Print Name	Signature	Date

Australian National University

Statement of Contribution

This thesis is submitted as a Thesis by Compilation in accordance with https://policies.anu.edu.au/ppl/document/ANUP_003405

I declare that the research presented in this Thesis represents original work that I carried out during my candidature at the Australian National University, except for contributions to multi-author papers incorporated in the Thesis where my contributions are specified in this Statement of Contribution.

Title: A socio-environmental model for exploring sustainable water management futures: Participatory and collaborative modelling in the Lower Campaspe catchment

Authors: **Iwanaga, T.***, Partington, D., Ticehurst, J., Croke, B.F.W., Jakeman, A.J.

Publication outlet: Journal of Hydrology: Regional Studies

Current status of paper: Published

Contribution to paper: Farmer decision model and implementations of irrigation scheduling and crop yield models by Takuya Iwanaga. Integration of models and data, data analysis and visualization, document structure and writing by Takuya Iwanaga. Social stakeholder engagement and additional writing by Jenifer Ticehurst. Further writing and editing by Barry F.W. Croke, and Anthony J. Jakeman. Iwanaga developed the (software) framework to enable model integration, led model development and integration liaising with component model developers where necessary. Sensitivity analysis and model testing to ensure model quality also by Iwanaga.

Senior author or collaborating authors endorsement: Daniel Partington (Dr. Daniel Partington)

Takuya Iwanaga		30/12/2020
Candidate – Print Name	Signature	Date

Endorsed

Anthony J. Jakeman	A. J. Jakeman	30/12/2020
Primary Supervisor – Print Name	Signature	Date
Philip Gibbons	Philip Gibbons	21/06/2021
Delegated Authority – Print Name	Signature	Date

Australian National University

Statement of Contribution

This thesis is submitted as a Thesis by Compilation in accordance with https://policies.anu.edu.au/ppl/document/ANUP_003405

I declare that the research presented in this Thesis represents original work that I carried out during my candidature at the Australian National University, except for contributions to multi-author papers incorporated in the Thesis where my contributions are specified in this Statement of Contribution.

Title: Property-based sensitivity analysis: an approach to identify model implementation and integration errors

Authors: **Iwanaga, T.***, Sun, X., Wang, Q., Guillaume, J.H.A, Croke, B.F.W., Rahman, J., Jakeman, A.J.

Publication outlet: Environmental Modelling and Software

Current status of paper: Submitted

Contribution to paper: Initial concept, model development and implementation, analysis and writing by Takuya Iwanaga. Further scoping of the analysis and structuring of the paper with Joseph Guillaume and Xifu Sun. Additional programming and code review by Xifu Sun and Qian Wang. Additional writing and editing by Joseph Guillaume, Qian Wang, Xifu Sun, Barry F.W. Croke, and Anthony J. Jakeman.

Senior author or collaborating authors endorsement: Xifu Sun (Mr. Xifu Sun)

Takuya Iwanaga		30/12/2020
Candidate – Print Name	Signature	Date

Endorsed

Anthony J. Jakeman	A. J. Jakeman	30/12/2020
Primary Supervisor – Print Name	Signature	Date
Philip Gibbons	Philip Gibbons	21/06/2021
Delegated Authority – Print Name	Signature	Date

Statement of Contribution

This thesis is submitted as a Thesis by Compilation in accordance with https://policies.anu.edu.au/ppl/document/ANUP_003405

I declare that the research presented in this Thesis represents original work that I carried out during my candidature at the Australian National University, except for contributions to multi-author papers incorporated in the Thesis where my contributions are specified in this Statement of Contribution.

Title: Socio-technical scales in socio-environmental modeling: managing a system-of-systems modeling approach

Authors: **Iwanaga, T.***, Wang, H-H., Hamilton, S. H., Grimm, V., Koralewski, T. E., Salado, A., Elsawah, S., Razavi, S., Yang, J., Glynn, P., Badham, J., Voinov, A., Chen, M., Grant, W. E., Peterson, T. R., Frank, K., Shenk, G., C Barton, M., Jakeman A. J., Little, J. C.,

Publication outlet: Environmental Modelling and Software

Current status of paper: Published

Contribution to paper: Initial idea and concept led by Little and Jakeman. Literature review and subsequent focus on the socio-technical aspects of scale, writing and editing by Iwanaga, Wang, and Hamilton. Significant, important discussions and input provided by other co-authors. Coordination of contribution and further editing by Iwanaga.

Senior author or collaborating authors endorsement:

Hsiaohsuan Wang (Dr. Hsiao-Hsuan Wang)

(Dr. Serena Hamilton)

Takuya Iwanaga		30/12/2020
Candidate – Print Name	Signature	Date

Endorsed

Anthony J. Jakeman	A. J. Jakeman	30/12/2020
Primary Supervisor – Print Name	Signature	Date
Philip Gibbons	Philip Gibbons	21/06/2021
Delegated Authority – Print Name	Signature	Date

Australian National University

Statement of Contribution

This thesis is submitted as a Thesis by Compilation in accordance with https://policies.anu.edu.au/ppl/document/ANUP_003405

I declare that the research presented in this Thesis represents original work that I carried out during my candidature at the Australian National University, except for contributions to multi-author papers incorporated in the Thesis where my contributions are specified in this Statement of Contribution.

Title: Sensitivity analysis of spatially distributed environmental models- a pragmatic framework for the exploration of uncertainty sources

Authors: Koo, H., **Iwanaga, T.**, Croke, B.F.W., Jakeman, A.J., Yang, J., Wang, H-H., Sun, X., Lü, G., Li, Xin, Yue, T., Yuan, W., Liu, X., Chen, M.*

Publication outlet: Environmental Software and Modelling

Current status of paper: Published

Contribution to paper: Initial concept and writing by Hyeongmo Koo (first author) and Min Chen (corresponding author). Further intellectual contribution and editing support by all co-authors in authorship order.

Senior author or collaborating authors endorsement:

Hyeongmo (Dr. Hyeongmo Koo)

Min chen (Prof. Min Chen)

Takuya Iwanaga		30/12/2020
Candidate – Print Name	Signature	Date

Endorsed

Anthony J. Jakeman	A. J. Jakeman	30/12/2020
Primary Supervisor – Print Name	Signature	Date
Philip Gibbons	Philip Gibbons	21/06/2021
Delegated Authority – Print Name	Signature	Date