Hydro-JULES
Reports

Hydro-JULES System Design

UK Centre for Ecology & Hydrology
Simon Dadson, Eleanor Blyth, Douglas Clark,
Richard Ellis, Matthew Fry, Jan Polcher

National Centre for Atmospheric Science
Thibault Hallouin, Bryan Lawrence, Grenville Lister

British Geological Survey
Andrew Hughes

Met Office
Martin Best

March 2021

UK Centre for Ecology & Hydrology

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

British Geological Survey

Introduction

Purpose

Scope

The Hydro-JULES programme is supported under NERC National Capability and delivered by CEH in partnership with NCAS and BGS. The purpose of this document is to describe the Hydro-JULES modelling framework, which has been informed by several stakeholder consultations undertaken during the requirements-gathering phase of the programme.

Hydro-JULES will produce an integrated modelling system for the research and academic community to: (i) address important science questions in the fields of land-atmosphere feedbacks, carbon and nutrient cycles, data science and integration with novel instrumentation and Earth observation technologies; (ii) quantify the risks of hydro-climatic extremes (e.g. floods and drought) in a changing environment to support long-range planning and policy decisions; and (iii) improve hydrological forecasting using new sensors and modelling technology.

Project aims and objectives

The major project outcome will be an integrated terrestrial hydrological model that is capable of using weather and rainfall information from global to local scales, and routing that through the terrestrial hydrological system to assess flood inundation and drought, and their consequent impacts. The Hydro-JULES model and its associated datasets will enable UK science to tackle research questions in hydrological science and will provide a national resource to support research both specific to the Hydro-JULES project and beyond. Hydro-JULES will provide the UK hydrological and land-surface communities with the modelling infrastructure to tackle pressing research questions in this field, which include:

- How do hydrological systems respond to present-day climate variability and how can the impacts of future climate change best be quantified in ungauged locations, in data-sparse regions and under non-stationary conditions?
- To what extent can new observational and modelling techniques improve our understanding of how extreme precipitation, especially high-intensity convective precipitation, drives flooding?
- How will changes in land-use and land management affect surface permeability, soil water storage, runoff, river flows and flood inundation?

- What are the combined probabilities of fluvial, pluvial, coastal and groundwater flooding in response to changes in climate, and can a coupled approach to flood risk estimation quantify those risks more effectively?
- How will biogeochemical and nutrient cycles respond to current and future variability in the hydrological cycle, especially under conditions of changing climate and land-cover?
- To what extent can assimilation of observed hydrological states and fluxes (e.g., soil moisture and stream flow) improve hydrological and meteorological predictions, and on what time-scales?
- Can uncertainty in large-scale hydrological predictions be attributed to specific hydrological processes in order to target future process-based research?
- What is the sensitivity of Earth system components (e.g., vegetation, carbon cycle, aerosols, land ice, sea ice, ocean circulation and biogeochemistry) to changes in the hydrological cycle; and can enhanced representation of terrestrial hydrology in Earth system models help constrain responses to such changes?

Programme delivery is via six work packages illustrated in Figure 1.

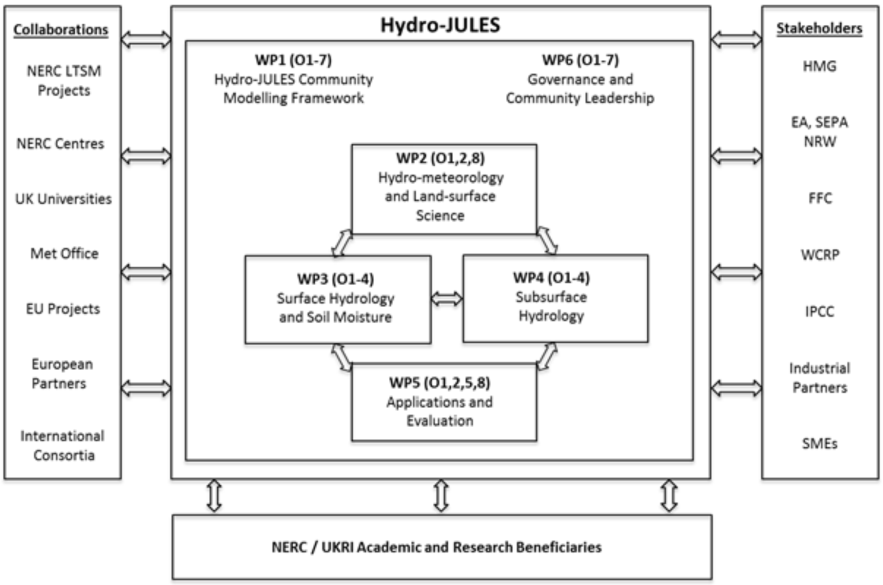


Figure 1 Schematic showing the key elements of an integrated flood risk approach. The arrows indicate status: implemented (solid green), partially implemented (dashed green) or currently experimental (dashed red) Source: National Flood Resilience Review, p.91

Consultation with the research and academic community

In March 2018, a period of consultation with members of the UK academic and research community indicated a shared desire to bring together the land-surface modelling community with the hydrological community for mutual benefit. Widespread support was received for the development of a community hydrological model alongside and in conjunction with the JULES[1] land-surface model. In September 2018 an open meeting was held to explore wider collaborative opportunities. In January 2019 a model integration workshop was held amongst a focussed group of domain-specific modelling experts in order to identify desirable properties of the resulting Hydro-JULES system. Between October 2018 and March 2019, a series of meetings were held with wider beneficiaries of the Hydro-JULES outputs.

The requirements document summarises the outputs of these community consultations and engagement activities in order to define key requirements of the Hydro-JULES system.

Perceptual models for Hydro-JULES

Physical understanding and predictive capability

Underpinning the design of a community model is a shared understanding of the important entities that must be represented and the process relations between them. In order to develop this idea further, a group of approximately 20 scientists attended a model integration workshop. The purpose of this workshop was to share each other's perceptual models of the terrestrial hydrological cycle. Through a series of breakout discussions, a common set of objectives emerged. Particular importance was attached to the need for agreed definitions of stores, fluxes and interfaces in this field.

A key part of the challenge in Hydro-JULES is to combine two distinct objectives: the first a scientific concern with development and refinement of hypotheses concerning the nature and functioning of hydrological systems, in order to arrive at physical (chemical, biological) understanding of the water cycle; the second a need to make accurate and timely predictions of water resources availability and hydrological extremes.

Components of the hydrological system where data are rich, boundaries are straightforward, and model parameters are known offer much potential for predictive modelling. One such problem is fluvial flood inundation. Here the governing equations are well established, and successful predictive modelling is limited mainly by the availability of good observations of system states, fluxes, parameters and boundary conditions, and by

---

[1] Joint UK Land Environment Simulator: https://jules.jchmr.org/

the computational power required to apply well-established principles at appropriate spatial and temporal scales.

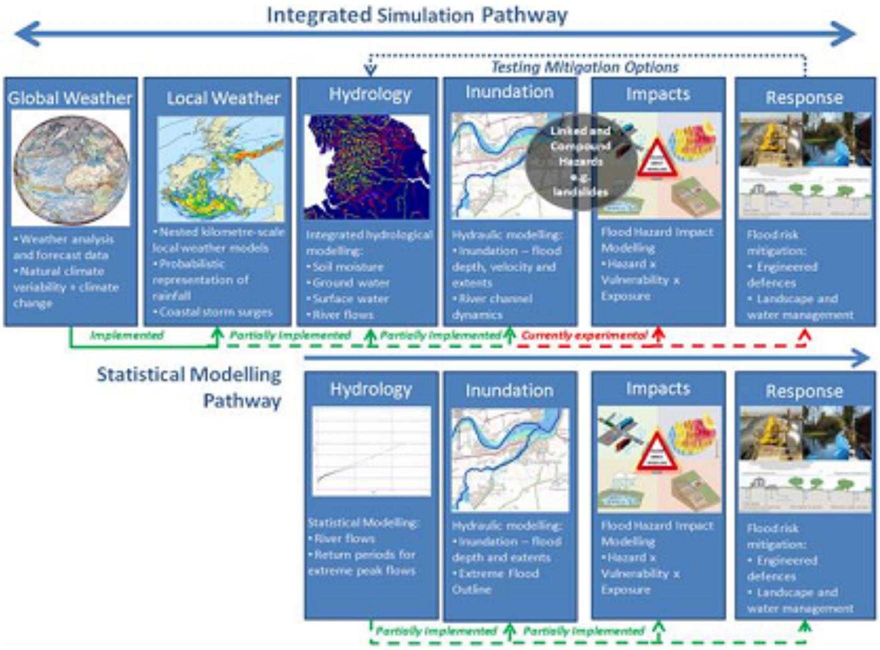


Figure 2 Schematic showing the key elements of an integrated flood risk approach. The arrows indicate status: implemented (solid green), partially implemented (dashed green) or currently experimental (dashed red) Source: National Flood Resilience Review, p.91.

By contrast, there are other areas of the terrestrial water cycle where perceptual models differ even amongst practitioners with the same end goals in mind. It may be that data availability is a constraint, but equally, progress may have been hampered by a lack of sound theoretical concepts, especially to link behaviour theorised at one scale with phenomena observed at another. An example of such a process is the flux of water through the soil moisture profile, where the variable of interest is difficult to measure and exhibits heterogeneity across a range of scales. Consequently, even though it may be possible to theorise that fluxes are driven by gradients, the gradients that drive the fluxes are rarely measured (or modelled) at appropriate scales to provide accurate predictions.

A new modelling framework for hydrology

Whilst workshop participants agreed on the importance of the problems articulated above, it was noted that this situation is a fundamental challenge for the science of hydrology – not one confined to Hydro-JULES alone. Nonetheless its practical upshot is that a new framework is required which can allow rapid progress in sub-disciplines where the ingredients for progress are abundant, yet which preserves the capability to build in later progress in parts of the field which will take longer to bear fruit.

Process

Design for a modular system

Definition of the user community

The approach to designing the Hydro-JULES system is divided into five stages: (i) requirements gathering, in consultation with the stakeholders in the research and academic community; (ii) design; (iii) implementation; (iv) testing and verification; (v) evaluation. In practice, elements of these activities will need to run concurrently and, in some cases, iteratively.

As a NERC National Capability programme, Hydro-JULES is intended to support the research and academic community with the research tools and infrastructure required to meet the scientific objectives outlined in section Functional requirements. The primary goal is to provide the underpinning scientific and technical capability to support this research. It is anticipated that users of the Hydro-JULES outputs will find opportunities to build additional innovations with applications beyond the R&D focus of the NC programme. Where such developments are in addition to the demonstrator applications funded under the programme it is acknowledged that additional funding or sponsorship must be provided in order to support these extension activities.

Science areas identified in consultation

Five science areas have been identified in the emerging consultation for initial prioritisation:

- Groundwater
- Soil moisture (including macropore flows)
- Fluvial inundation (including coastal, estuarine and shelf-seas interactions)
- Evaporation (including bare soil and evaporation from intercepted canopy water)
- Anthropogenic interventions (including irrigation, impoundments, inter-basin transfers)

Domains and resolutions

The modelling framework will support a set of model configurations centred on key meteorological datasets, as described in Table 1.

Table 1 Domains and resolutions for Hydro-JULES configurations

Resolution	Extent & projection	Meteorological data	Temporal coverage
UK			
1km	OSGB	CHESS-UK	1960-present
1km	OSGB	Future projections, initially derived from UKCP18	
Global			
0.5°	Global lat-lon	WFD-EI / WFDE5*	1979 – 2018
0.5°	Global lat-lon	GSWP3	1900 - 2014
0.5°	Global lat-lon	Future projections from commonly used global climate projections, e.g. ISIMIP2	

* These are both based on the 0.5° CRU grid but have different land-sea masks. Other inputs ("ancillary fields") will be provided on a common grid to support both driving datasets.

These will provide users with working baseline configurations and input data, but it should be emphasised that the modelling framework will also be capable of being used with other resolutions and grids. A list of initial candidate datasets supporting modelling over each of these domains is provided in Appendix C.

Benefits of using a modular structure

All stakeholder consultations emphasised the value of an efficient and robust modular structure in which interfaces are clearly specified and resilient to future development. The reasons cited included:

I. to enable efficient model development and project management – modularity allows innovations developed outside the scope of the initial funded programme to be incorporated, and presents a clear entry point to newcomers;
II. to enhance interoperability – interoperable components can be switched in and out as new science routines become available;
III. to improve testability and robustness - modules that can be subjected to systematic unit testing outside the coupled system will be more robust and their path towards integration will be smoother than had they been developed with ad hoc connections to an existing code base.

A strong argument can be made that the overhead of modularity is outweighed by these benefits. Nonetheless, it is not the case that all hydrological problems can be solved better within a land-surface modelling framework. This is particularly true in cases where the physically-based models cannot be constrained with good empirical data or parameter estimates. Therefore, an ability to construct a simplified workflow from modules linked together in a high-level programming language like Python is in demand – and it is anticipated that this facility would further increase the uptake and utility of the combined system.

Coupled evaluation

Notwithstanding the advantages of a modular system, a comprehensive programme of evaluation is needed to identify combinations of individually tried and tested modules which work well together to simulate the hydrological cycle in coupled experiments where the terrestrial model is coupled with atmospheric and ocean models, and in uncoupled experiments where the terrestrial model is run standalone.

The motivation for taking such an approach is both practical – it is well known that complex modelling software requires skill and expertise to configure well – and philosophical – if we view a numerical model as the concatenation of multiple hypotheses then, in the presence of uncertainty, we need to test not only the individual hypotheses but also the performance of the coupled system because there may be many arrangements of interacting parts that result in the same observable outcome. It is desirable that strategies to diagnose and take account of equifinality be incorporated into the research design from the outset.

System requirements

Functional requirements

The following functional requirements were set out by NERC Council (December 2016).

- **O1.** Integration of the 3-dimensional terrestrial water cycle in such a way as to ensure consistency across space and timescales, in the horizontal and vertical, and ensuring conservation of water across the various components.
- **O2.** Integration with the atmosphere to ensure consistency in evapo-transpiration across the land-atmosphere interface, and full representation of the space/time heterogeneity of precipitation, which is fundamentally important for surface water flooding and the functioning of the terrestrial water cycle.
- **O3.** Ability to be integrated with models of the coastal ocean to represent river outflows (including nutrients), estuaries and the effects of tides and storm surges.
- **O4.** Integration of the water cycle with other key terrestrial cycles to ensure consistency in the treatment of water, heat, carbon and nitrogen cycles, which are critical for modelling and understanding the Earth system.
- **O5.** Flexibility to constrain the terrestrial system model with observations through stand-alone testing and through the process of data assimilation.
- **O6.** Flexibility within the terrestrial system model to allow testing and inter-comparison of different approaches to specific components.
- **O7.** Flexibility of the terrestrial system model to permit its coupling to atmospheric and other model components through an appropriate standardised interface, and to permit its coupling to the Met Office Unified Model.
- **O8.** Formal processes for quantification of uncertainty, including the propagation from driving meteorological variables through the hydrological system.

Additional requirements

Performance requirements

- Be open source and have an open development process
- Have clear goals, scope, and where appropriate, deliver stable software interfaces, particularly to the land surface model JULES

- Provide compatibility with developments necessary for JULES as part of its integration in LFRic[2], the future modelling system of the UK Met Office for weather and climate predictions
- Provide compatibility with the OASIS coupler[3]
- Minimally disruptive to existing solutions

Usability requirements

- Be accompanied by documentation including a user manual and technical reference
- Allow flexible data input
- Provide accessible visualisation tools
- Offer technical support (including training) subject to the scope of future funding

Maintainability requirements

- System for management of configurations and simulation results for reproducibility
- Managed access to supported configurations via JASMIN[4], the cloud computing and data storage delivered by the Science & Technology Facilities Council
- Subject to version control system
- Have clear coding standards

Data requirements

Provision of accessible and well-ordered datasets for modelling will enhance ease of use, and hence uptake, of the modelling system, and also support model comparisons and incremental model developments.

Data will be provided for the different domains defined above, initially UK at 1 km and globally at 0.5°. Data should be fully documented, placed within an accessible archive, and given DOIs to support robust science and scientific reproducibility. The granularity of this information is important, and individual datasets within a model run should be identified and documented individually, rather than at the “model run” level. This supports a modular model development approach, and also interchange and improvement of datasets themselves.

---

[2] https://www.metoffice.gov.uk/research/approach/modelling-systems/lfric
[3] https://portal.enes.org/oasis
[4] https://jasmin.ac.uk/

The requirement for dataset documentation applies to driving data, supporting (ancillary) data, and also validation datasets (from in situ monitoring but also Earth Observation). Whilst it may be useful to provide publicly accessible archived (with a DOI) versions of model run outputs where these outputs form the basis of wider analysis (e.g. of UKCP18 runs), model outputs should be uniquely defined by the datasets, model version, and model "configuration", and therefore readily reproducible.

Origins and derivations of datasets are useful in order to understand limitations, and tools and datasets used in production of Hydro-JULES datasets should, ideally, also be documented and accessible for reference.

Datasets themselves should contain appropriate metadata in order to enhance understanding and use. NetCDF formats are proposed as they enable rich metadata to be incorporated within files, and metadata will conform to Climate and Forecast (CF) conventions[5] and be further enhanced with global metadata from the Attribute Convention for Data Discovery (ACDD[6]). NetCDF file structures should also conform to CF conventions to enable interoperability with common code libraries and processing and visualisation software.

The system should enable and encourage data sharing following this approach. Principles and practices should be set out within the system documentation, through this document and others (fair share policies, data licences, data protocols and guidance, example data formats, potentially tools for creating data to meet these requirements). These will need to evolve through time to cope with changes in technology, formats, etc. The datasets for the specified domains will provide an example of the data needed to support a reproducible modular modelling approach. The wider community should be encouraged to follow this best practice and build equivalent data repositories where the model is to be applied, e.g. publishing data for regional applications or different scales.

NERC Data Centres, such as the Environmental Information Data Centre (EIDC), should be used for archiving of datasets, and the CEDA archive provides the additional benefit of enabling archived datasets to be directly accessed by model code running within the JASMIN infrastructure, without replication of datasets.

New input and underpinning datasets for newly developed modules should be incorporated into the core datasets provided with the modelling system, as they are

---

[5] https://cfconventions.org/

[6] http://wiki.esipfed.org/index.php/Attribute_Convention_for_Data_Discovery_1-3

developed, and documented and archived alongside existing data. The Hydro-JULES data should be catalogued to ensure that the datasets available are readily identified and understood by end users.

Stakeholder requirements

This project requires close and meaningful collaboration with the wider NERC academic and research community in order to realise the development and deployment of a world-leading terrestrial hydrological model and to ensure that the model is comprehensively applied to outstanding hydrological and Earth system science questions. The design of the modelling system itself and the engagement and community-building strategy must take account of these requirements at all stages. It remains important to note that the target audience for Hydro-JULES is the research and academic community. Any effort to derive commercial or regulatory benefit from this work whilst not precluded is likely a research endeavour in itself and therefore not within the scope of this document or the NERC-funded programme.

Design principles

Design of modules and components

The framework provides a model for the terrestrial water cycle. The model is composed of five components – three internal components (surface layer, sub-surface, and open water; see Figure 3 below) and two boundary components (atmosphere and ocean) – together these components simulate the whole water cycle. The framework is used to choose from a collection of component instances, representing variations of the science (i.e. processes to consider) for each component category – together these choices form the model configuration.

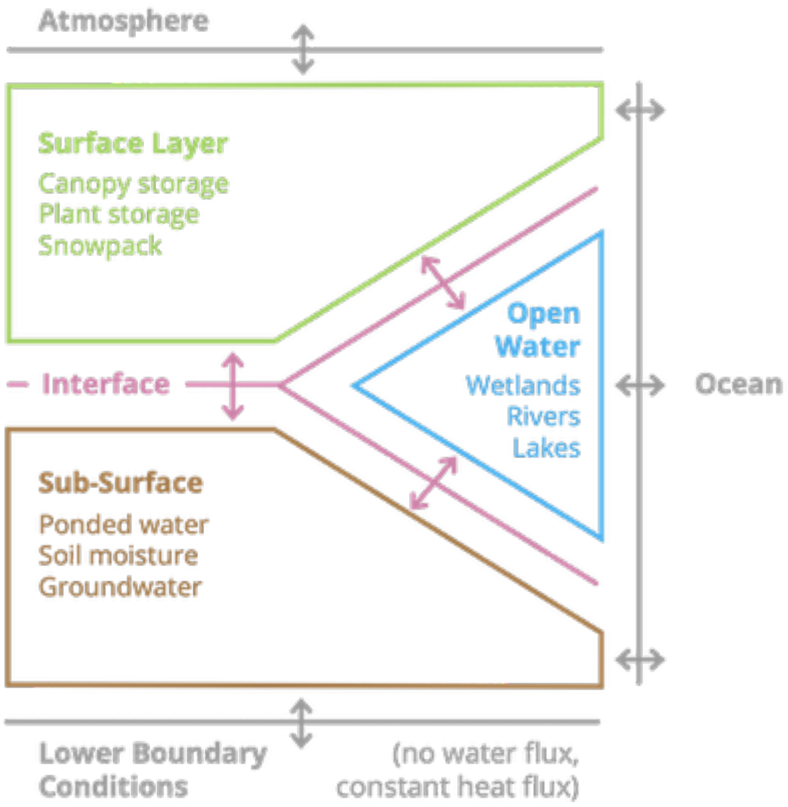


Figure 3 Blueprint of the framework

Each component can be a computational component (e.g. code computing the runoff), a dataset derived from another simulation or from observations (e.g. a file containing the runoff) or a null component (e.g. producing zero runoff). The atmosphere and ocean components can be used to link to driving datasets or to other models through coupling.

The thought processes and discussions behind the choice of the three components will be presented in further documentation but to summarise: the surface component represents fast processes dominated by one-dimensional vertical transfer and closely connected to the overlying atmosphere (e.g. evaporation); the sub-surface is dominated by slower two- or three-dimensional exchanges (e.g. soil moisture and groundwater); and the open water domain is dominated by one- or two- dimensional processes with information exchanged laterally (e.g. shallow water flow). To illustrate the possibilities, three examples are given below.

In the first example, the three internal components are all computational and driven by atmospheric and oceanic boundary conditions supplied from data files.

Example 1 – land only	
Atmosphere	Dataset containing meteorological driving variables compatible with the surface component
Surface	Computational component calculating evaporation and throughfall
Sub-surface component	Calculating soil moisture, runoff and groundwater
Open water component	Calculating river flow and inundation
Ocean component	Dataset providing tide heights

In the second example, only the open water component performs calculations; these are driven by a dataset given for the sub-surface component.

Example 2 – offline routing	
Atmosphere	Null component
Surface	Null component
Sub-surface component	Dataset with runoff and groundwater states
Open water component	Calculating river flow and inundation
Ocean component	Null component

In this third example, all components are actual modelling components forming a complete Earth System consideration for the water cycle.

Example 3 – fully coupled atmosphere-land-ocean models	
Atmosphere	Calculating atmospheric humidity and precipitation
Surface	Computational component calculating evaporation and throughfall
Sub-surface component	Calculating soil moisture, runoff and groundwater
Open water component	Calculating river flow and inundation
Ocean component	Calculating tides and storm surges

Each component instance will be configured using Python objects specific to the framework describing the time domain, spatial domain, parameters, and driving and ancillary data required by the component. Each component instance is callable by the framework through its initialise-run-finalise methods, within which calls to external code (e.g. Fortran or C/C++ libraries) can be made.

The components are coupled through transfers of information that are expected to be fluxes or states – together these transfers form the interface of the framework (see Figure 4 below). Further details on the options for the implementation of this interface in the section Model coupling technologies.

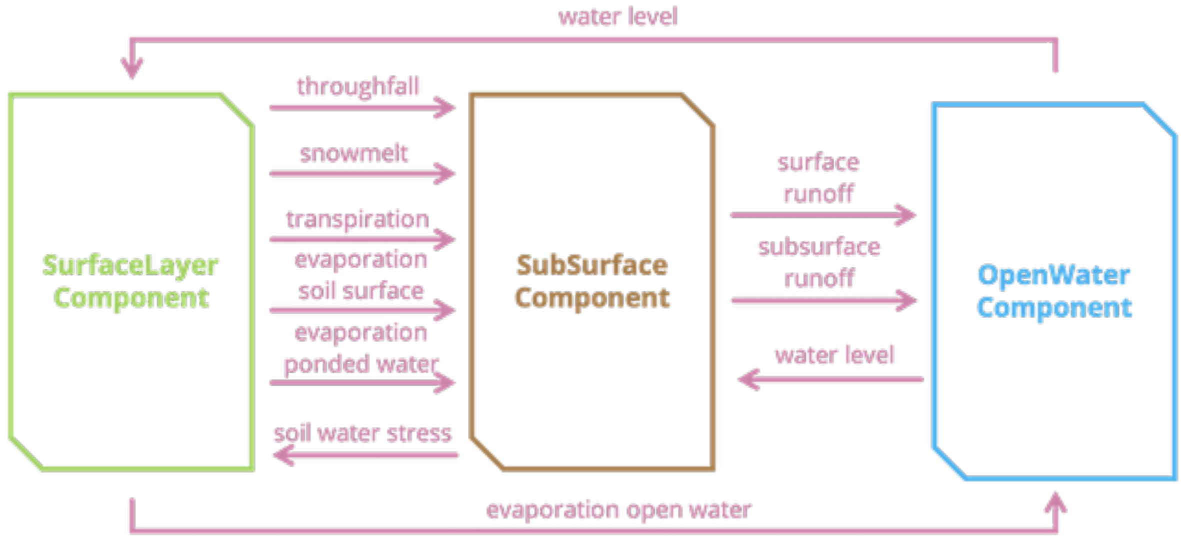


Figure 4 *Illustration of the information transferred between components through the framework interface*

Each component will use its own solver. That is, each component will have its own space and time domains and discretisations which will be common to all processes represented in that component. It is not necessary for the space discretisation to be regularly shaped, whereas the time discretisation must be regularly spaced. These features must be used consistently within a given component but can differ between components.

Model configurations will be stored in the YAML format, chosen over XML or JSON formats because YAML is intrinsically more human readable, it follows the same indentation as Python, and the format is increasingly used as the go-to format for configurations, e.g. GitHub Actions workflows[7], conda environments[8], ESMValTool recipes[9]. These files will provide a simple way to share configurations of the framework, which is crucial for collaboration and reproducibility.

Supporting datasets – standard formats

Supporting datasets are required to specify model parameters and must be provided in CF-compliant netCDF formats served via recognised, maintained data centres with documented provenance in order to preserve reproducibility. For new science developments which create new datasets it will be necessary to create versions at the resolutions listed in Table 1, or to provide a mechanism to re-grid as required (cf. Met Office CAP utility and its descendants). Licensing requirements must be considered at the outset, with the aim being to use freely available datasets where possible (see section Licensing and fair use).

Data dictionary

Metadata dictionaries will be created for common terms in use within the modelling framework, platform and wider project. Even apparently uncontroversial definitions can be contested, and without detailed definition there can be a great deal of supposition about the meaning of terms. The principal reasons for using such dictionaries are to:

- Provide a single, web-accessible point of access for definitive lists of terms
- Provide a mechanism for management of terms and definitions, e.g. version control

---

[7] https://github.com/features/actions

[8] https://docs.conda.io/projects/conda/en/latest/user-guide/tasks/manage-environments.html#sharing-an-environment

[9] https://docs.esmvaltool.org/en/latest/develop/recipe.html

- Provie a persistent identifier for defined terms and definitions that can be widely used e.g. embedded in datasets
- Enable machine-actionable use of terms, including discovery of terms and equivalents
- Enable links to equivalent terms in other systems, and describing this equivalence (e.g. exact match or a broader / narrower definition)

The principal focus of these dictionaries will be on modelled variables, in particular those used as inputs and outputs of models, and those defined at the model interfaces (including conceptual variables). Where possible standard names from the CF conventions will be used for variables in datasets and additional hydrological variables will be added to the standard names list. Comparable terms from other initiatives within the hydrological and land surface modelling domain (e.g. CSDMS, see below) will be identified and linked.

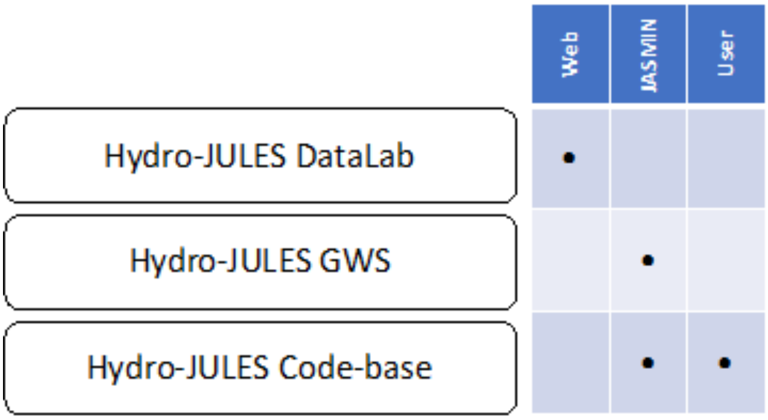
Examples of such dictionaries include:

- JULES output variables: http://jules-lsm.github.io/latest/output-variables.html
- NERC Vocab Server standard names: http://vocab.nerc.ac.uk/collection/P07/current/ including e.g. precipitation flux: http://vocab.nerc.ac.uk/collection/P07/current/CFSN0453/2/
- CSDMS standard names: https://csdms.colorado.edu/wiki/CSDMS_Standard_Names
- CF conventions standard names: http://cfconventions.org/standard-names.html

Implementation

Three tiers of access

It is envisaged that three tiers of access will be necessary in order for users with varying degrees of engagement or expertise to use and interact with Hydro-JULES. These are illustrated in Figure 5.



	Web	JASMIN	User
Hydro-JULES DataLab	●		
Hydro-JULES GWS		●	
Hydro-JULES Code-base		●	●

Figure 5 *Tiers of access to Hydro-JULES models and datasets*

The Hydro-JULES DataLab is a web-based service built using NERC Data Labs Infrastructure which allows a beginner user access to pre-computed datasets to conduct simple analyses, subset data, plot results and share via Jupyter notebooks.

The Hydro-JULES Group Workspace (GWS)[10,11], is a shared area accessed via JASMIN which allows a community of users to share data and code and to collaborate, including by running Hydro-JULES modules and JULES configurations on the LOTUS cluster.

The Hydro-JULES code-base is a maintained repository of Fortran90 and Python which provides access to the underlying subroutines used to implement process models in the Hydro-JULES system. Core subroutines will be shared with the JULES code management system.

---

[10] Marthews TR & Fry M (2019). *Options for running Hydro-JULES models on JASMIN*. Hydro-JULES Report D1.4.1. (available on request via hydrojules@ceh.ac.uk)

[11] Marthews TR, Fry M & Clark D (2020). *A post-processing environment for Hydro-JULES models on JASMIN using Jupyter notebooks*. Hydro-JULES Report D1.4.5a. (available on request via hydrojules@ceh.ac.uk)

Tier I: Hydro-JULES Datalab

Tier 1 will provide managed access to pre-computed fields for subsequent online analysis via the NERC DataLabs service. Using this system, it will be possible for the beginner user to retrieve, analyse and plot data held on JASMIN from a series of pre-determined experiments. An audit trail of the analysis will be available via Jupyter notebooks and an ability to archive and share notebooks in support of publications and for subsequent analysis. Training materials will be produced for a range of typical users (see section Training requirements). Other data requirements include, e.g. flow gauge data, water quality data, Earth Observation data.

Tier II: Hydro-JULES group workspace

A second level of access is via shared Group Workspaces on JASMIN with access to the Hydro-JULES modular libraries via a version-controlled system. This mechanism will enable open development and will be implemented using GitHub, and the ability to call individual Hydro-JULES modules via a Python wrapper will be provided.

The intermediate model user will also be able to access validated, published Hydro-JULES configurations and deploy them via LOTUS on the JASMIN system. The ability to modify experiments to run with user-supplied driving data, ancillary datasets, and model parameter settings will be provided.

Tier III: Hydro-JULES code base

The third tier of access will draw on the existing JULES code management system and will continue to be the way in which advanced users and developers contribute coupled model functionality. It is envisaged that hydrological code will be developed, tested and evaluated first in the Hydro-JULES repository and then when validated in its modular form may be brought into the JULES managed system.

The existing JULES configuration management system will remain in place and will provide a mechanism to ensure reproducibility of experiments. A mechanism to ensure that the two repositories remain synchronised is necessary.

Supporting utilities

The availability of a range of modular options to combine in multiple valid ways provides flexibility to isolate particular model components for testing purposes and to construct simpler workflows. Appropriate utility functions are likely to be required for re-gridding, time-stepping, basic calculations and plotting, many of which would also be available via the NERC DataLabs interface.

Model coupling technologies

In designing the framework, a number of issues must be considered: (i) What coupling functionality is required? (ii) How much refactoring will be required to get code into the framework? (iii) How intrusive will the framework be? (iv) How will flexibility and ease of development trade-off against computational efficiency? In some cases (e.g., where there is no change in domain) it may suffice to pass information by using a function or subroutine call. In other situations, re-gridding functions or a coupler like OASIS may be required. In some cases the information passing between components may need to be transformed to conform to the requirements of the framework.

In the first instance a Python framework is being developed with the coupling functionality provided by the re-gridding features of cf-python (a Python package making use of the remapping functionality of the Earth System Modelling Framework), along with developer-written transformation functions where the requirements of the framework differ. A predefined interface between modules will be provided for developers to design new components, along with a number of example transformation functions and a tutorial based on the development of these functions.

The framework requires that the components of the model cannot be constrained to have their domain discretisation defined simply by the meteorological driving data. To implement this feature information will be passed between two model components with different domain discretisations using an intermediate supermesh[12]. This supermesh represents the nodes of each element in both domains thus allowing the information to be passed conserving the quantity in question. The weighting files to allow this can be created in either OASIS or ESMF and must be calculated for each supported domain. Additionally there will, in effect, be a supermesh in the temporal dimension. This will be defined by the

---

[12] P.E. Farrell, M.D. Piggott, C.C. Pain, G.J. Gorman, C.R. Wilson, 2009. Conservative interpolation between unstructured meshes via supermesh construction. *Computer Methods in Applied Mechanics and Engineering* 198(33–36): 2632–2642. doi:10.1016/j.cma.2009.03.004

fastest model component. Slower components will be required to work at an integer product of the fastest component.

There are situations where not all the information in an interface will be passed, for example one way coupling or coupling between components that do not have the required complexity. In these cases, zeros or null values may be passed. The components would be required to be able to receive but ignore this information.

Governing collaborative science

Licensing and fair use

Open source and open development

Open source modelling software relies on contributions from the community in order to support its development. In the present context, open source modelling software offers potential benefits to its developers, who seek recognition for their work in the form of academic reputation, publications, and citations, and to other researchers and academic users, who benefit from interdisciplinary collaboration and reduced costs of developing individual solutions to common problems. It is a requirement that Hydro-JULES adheres to NERC's modelling guidelines, which are stated definitively here: https://nerc.ukri.org/research/sites/data/policy/.

Software licensing

Of specific importance are the requirements for code to be developed: (i) in an open-source, non-proprietary environment; (ii) with source-code management; and (iii) with adequate documentation, in accordance with NERC's metadata standards given here: http://www.bgs.ac.uk/data/nercmodelmetadata/NERCmmgdv101.pdf.

Software components and documentation will be released under the terms of an open-source Berkeley Software Distribution 3-clause licence (BSD-3), as follows:

Copyright © 2020 UK Centre for Ecology and Hydrology, All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Important features of this licence are that it allows licensees to distribute derivative copies of the software for any use, including commercial, providing the copyright notice is preserved. Derived products cannot be endorsed with the name of the original copyright holder without permission. Liability is restricted and no warranties are given. It is intended that a similar licence will be used for components of JULES owned by the Met Office.

Data licensing

Model configuration files and forcing data will be provided in standard formats and adequately documented, where necessary in NERC Data Centres, to support specific publications. In general, it is considered that the Open Government Licence v3.0 is the most appropriate for use in Hydro-JULES and where possible this licence will be used for project-generated datasets and new model code. The OGL allows copying, publication, transmission and adaptation for any use providing the source of the data is acknowledged. The text of this license is available here: http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/.

It is acknowledged that some third-party datasets or model components may have licence conditions attached to them which prohibit onward dissemination. In those cases, it will be the responsibility of model users to obtain additional enhanced proprietary supporting datasets if they wish to use them.

Fair use policy

In addition to the licensing arrangement, a fair use policy is required in order to establish what constitutes appropriate community use of the modules and modelling system in order that proper recognition can be sought for work done. This is particularly important in relation to ensuring true collaboration where it is appropriate. The Fair Use Policy in Appendix A has been drafted and agreed by the Hydro-JULES team and a derivative of it has since been adopted for the whole of JULES[13]. A publication policy is given in Appendix B.

Management

---

[13] https://jules.jchmr.org/sites/default/files/JULES%20Fair%20Use%20%26%20Publication%20Policy_v5_FINAL.docx

Documentation

The requirement for good documentation is essential. The basic units of documentation are at the level of the statement, subroutine, interface, data structure or class, and module. It is also necessary to document key procedures which involve applications of these components to particular datasets.

A coding standards manual will be kept giving details of documentation requirements. The maintenance requirements that run alongside particular components of the model will also need to be specified and updated frequently.

Documentation of datasets and other ancillary items will be required, in keeping with the standards established elsewhere.

Training requirements

Training requirements are mapped onto each of the tiers of access – it is expected that:

- A two-day course will be developed to showcase the abilities of the Hydro-JULES DataLab. Attendees will have no background knowledge beyond GCSE maths and an ability to use a web-browser to plot pre-computed datasets, to extract subset regions and compute time series and to perform basic calculations relating to climate and land cover scenarios in current and future conditions.
- A week-long course (Summer School planned from 2021 onwards) is envisaged for those with some data analysis and modelling experience as pre-requisites, based on the existing JULES course with a hydrological flavour and examples. This will develop an ability, using Python and Fortran, to run pre-packaged model components including JULES and standalone modules, and to interact with data in shared workspaces, to record and automate workflows and to share, document and archive model configurations and output in support of publications.
- An “extension” course, to include a session on how to add new modules and routines to the Hydro-JULES repository and to JULES.

Project monitoring and evaluation

The project steering committee comprises the Programme Director, Project Manager, Work Package leaders and at least one member from each of the partner institutes. It meets monthly with responsibility for reviewing progress, monitoring performance, forward planning and resolving any problems or delays. For consultative purposes, the project provides regular updates to the UK CEH Science Board, and an external science advisory panel. Policies are being instituted to set out good practice for interaction between stakeholders, including a fair-use policy, communications strategy, open access policy, policies for the administration of working groups, formal structures to organise interactions between stakeholders and to set out links with related programmes including JULES, JWCRP[14], and other UKRI-funded programmes. The programme reports to NERC biannually and will be the subject of a mid-term review at the end of its third year.

Project sustainability and risk management

Initial funding has been received to implement Hydro-JULES but does not include funding for its ongoing maintenance; nor does it consider risks associated with the delivery of some components in collaboration with other programmes.

Technical

Testing protocols

The source code for the community model can be considered as two main separate parts: the framework and its functionalities itself, and the suite of science models it hosts.
The framework itself is expected to feature a range of functionalities, e.g. input-output, defining spatial and temporal resolutions for the science models, re-gridding data exchanged between components, exchanging data between components. Testing the framework is the responsibility of the Hydro-JULES developers. Each part of the framework will come with its own "unit test" to guarantee it behaves as expected when used independently from any other functionality, and that its behaviour is sustained over time as the framework evolves.

Integration tests in the form of workflow scripts will also be written to test whether the functionalities used together are working as expected. The notion of "coverage" will be of prime importance for the integration tests - if a coverage of 100% is achieved, it indicates that every line of code is executed during the tests.

In combination with the version control system (in this case git with GitHub), a continuous integration solution (e.g. Travis CI, AppVeyor, GitHub Actions) will be used to automate workflows running the ensemble of unit tests and integration tests on different combinations of operating systems (Linux, Windows, macOS) and different Python versions (e.g. Python 3.7, 3.8, etc.). This will ensure that the framework will run locally in a range of

---

[14] Joint Weather and Climate Research Programme: http://www.jwcrp.org.uk/

environments. In addition, all tests will be run on JASMIN, in serial mode and in parallel mode using LOTUS.

It is expected that to test the entirety of the framework (system testing), components (i.e. science models) will be required. However, it is important to dissociate the third-party model source code (which is not the responsibility of the Hydro-JULES developers per se) and the framework source code. So, for each component, a "dummy" model working with a "dummy" dataset will need to be provided. In this way, the expected model outputs will be simple to determine, and the framework can be tested independently of any third-party model source code.

Finally, system testing will be carried out with a suite of supported models to make sure that a simulation run with inputs, supporting datasets and configuration options produces the same outputs (with some accuracy tolerance) as the model used in standalone mode. The scripts and the data required for these tests will be stored alongside the model source code in the Hydro-JULES repository, but they will not be part of the tests run with the continuous integration solution as they are likely to be too computationally demanding.

Coding standards

All Fortran programs hosted in the Hydro-JULES repository must follow JULES coding standards for Fortran[15]. Coding standards must also be followed to guarantee the proper interaction between Fortran programs and Python scripts, through the use of the f2py interface generator[16].

All Python scripts hosted in the Hydro-JULES repository must follow the style guide detailed in the Python Enhancement Proposal (PEP) 8[17]. All objects forming the Application Programming Interface (API) must feature docstrings to explicitly document their behaviour. All science models included in the framework must follow a particular template to comply with the set interface between components. In addition, all inputs, parameters, and outputs must be provided with explicit names and units, and as far as possible, these should be compliant with the Climate and Forecast (CF) conventions.

---

[15] https://jules-lsm.github.io/coding_standards/

[16] https://numpy.org/doc/stable/f2py/

[17] https://www.python.org/dev/peps/pep-0008/

Conclusions

This document is provided to the research and academic community and to wider project stakeholders under Hydro-JULES for a six-week consultation period during Autumn 2020. Any comments or feedback should be sent to hydrojules@ceh.ac.uk.

Appendix A

What is Hydro-JULES?

Hydro-JULES is a research programme funded by the UK's Natural Environment Research Council. It will produce a fully integrated open-source model of the terrestrial hydrological cycle. Hydro-JULES will provide integrated modelling capability for the UK research and academic community in hydrology and land-surface science. It will enable scientific researchers to collaborate more effectively on hydrological problems which link the atmosphere, land-surface, groundwater, cryosphere and oceans.

Hydro-JULES will produce a model of the terrestrial hydrological cycle which will be capable of running in a standalone configuration. Model sub-components will be provided in a modular framework so that other models will be able to link to them. Where appropriate, some modules will be linked to JULES[18], which is the UK's land-surface model used in weather forecasting and climate research.

What does the fair use policy apply to?

The fair use policy applies to all uses of Hydro-JULES products, including but not limited to data and computer code, for research, teaching or commercial applications. It is not intended to restrict what can be done with Hydro-JULES products, rather to ensure appropriate acknowledgement and communication between users and developers. This policy will be updated at least once a year and was last updated in March 2019.

Why do we have a fair use policy?

Most Hydro-JULES products are open-source and free for anyone to use for any purpose. The reason for making them available in this way is to foster collaboration amongst the research and academic community and to enable a wider range of applications in the public interest than would otherwise be possible.

The formal, legal conditions that govern the use of any Hydro-JULES product are set out in the licence attached to the product itself. Alongside the legal conditions of the Hydro-JULES licence, this Fair Use Policy sets out how members of the community should collaborate with one another in order to recognise the intellectual contribution of those whose work

---

[18] Joint UK Land Environment Simulator (https://jules.jchmr.org/)

has made the existence of Hydro-JULES possible. Doing so strengthens the case for further investment in those developments.

What do you need to do if you want to work with Hydro-JULES products?

You need to (i) register your use of Hydro-JULES products; (ii) ensure that your use of Hydro-JULES products is acknowledged in any publications, reports, presentations, derived software, and datasets; (iii) where appropriate offer the developers of the Hydro-JULES products co-authorship of any publications or other outputs that arise. These steps are outlined below.

How to register your use of Hydro-JULES products

When Hydro-JULES products become available, registration of their use will be possible at www.hydro-jules.org/register/.

How to acknowledge Hydro-JULES in reports, presentations, and publications?

If you have received funding via Hydro-JULES you should include the following acknowledgement: "This research was funded by National Environment Research Council award number NE/S017380/1 as part of the Hydro-JULES Programme". It is a requirement that papers be published in open-access journals and archived to NORA[19].

Co-authorship policy

Where extensive use of Hydro-JULES products has been made, it is expected that users will contact the developers to agree co-authorship or acknowledgement in publications. The developers will be able to tell you when co-authorship or acknowledgement is expected. A detailed publication policy is given in Appendix B.

---

[19] NERC Open Research Archive: http://nora.nerc.ac.uk/

Appendix B – Publication policy

Publication policy for Hydro-JULES

In common with all NERC-funded projects, Hydro-JULES is required to make outputs open-access via in accordance with the policy listed here: https://nerc.ukri.org/funding/next/publicationofwork/.

The following text should be used in the acknowledgements of all publications to indicate the nature of our funding: This work was supported by the Natural Environment Research Council [grant number NE/S017380/1]. Where appropriate, 'Hydro-JULES' should be mentioned in the acknowledgements.

Once accepted for publication, the lead author should send a copy of the author-accepted version to the Project Manager and ensure that it is uploaded to the relevant online research archive (e.g. NORA) accordingly.

Where relevant this policy should be read alongside the JULES Publication and Fair Use Policy https://jules.jchmr.org/content/home.

Guidance on co-authorship

The Hydro-JULES project is the result of a team of people and as such, the desire is for the whole team's efforts to be reflected in as many publications as is warranted.

This draft publication protocol is a working document, reflecting our current understanding of how we would like things to proceed, however, it does not cover all circumstances, hence, this should be used as a starting point for the discussion of authorship on a case-by-case basis. For further guidance, see the ICMJE Recommendations on the appropriate use of co-authorship (http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html).

- Authorship should be offered to those Hydro-JULES team members that have played a significant role in the:
    - conception/design/analysis/interpretation of data, and/or
    - drafting and critically analysing the paper.
- Authorship must reflect relative contribution to writing the paper or provision of content, with the primary authorship given to the primary writer. All other authors

should be listed in alphabetical order. Last authorship can be designated to project leaders or those necessitating special recognition.

- All authors must have read and commented on the paper prior to submission.
- Authors should process papers quickly.
- Authors should be available to address reviewer comments in a prompt manner.
- Corresponding authorship will default to the primary author but can be given to another author at the discretion of the primary author.

Appendix C – Candidate datasets

An initial list of candidate datasets for modelling over the domains described in Table 1 is given below.

UK, 1km

Meteorological data

CHESS UK driving data (full set of driving meteorological variables required by the JULES land surface model) from 1960-2017 (updated annually) https://doi.org/10.5285/2ab15bf0-ad08-415c-ba64-831168be7293

CEH-GEAR-1 hour. 1km gridded hourly rainfall dataset, 1990 – 2014 currently and to be updated through the course of the project https://doi.org/10.5285/d4ddc781-25f3-423a-bba0-747cc82dc6fa

CHESS PET, produced by the JULES land surface model, 1960-2017.
https://doi.org/10.5285/9116e565-2c0a-455b-9c68-558fdd9179ad

Future climate projections will be produced by downscaling and bias-correction of selected members of the current UKCP18 12km transient runs for the 21st Century.

Supporting data

Topography based on the CEH Integrated Hydrological DTM
(https://www.ceh.ac.uk/services/integrated-hydrological-digital-terrain-model).

Land cover fractions of 8 different categories: broadleaf trees, needleleaf trees, grass, crops, shrub, water, bare soil and urban, based on Land Cover Map 2000.

Soil ancillaries for the JULES land surface model, derived from the Harmonised World Soils Database, specific parameters defined at: http://jules-lsm.github.io/latest/namelists/ancillaries.nml.html#list-of-soil-params

Human influences e.g. groundwater and river abstractions and inputs.

Global, 0.5 degree

Meteorological data

GSWP3: Daily-resolution observed climate data on a global (land and ocean) 0.5°x0.5° lat-lon grid from the Global Soil Wetness Project Phase 3 (GSWP3), based on the reanalysis dataset 20CR. The dataset covers the period 1901-2010.

ERA5 / WFDE5: bias-corrected reconstruction of near-surface meteorological variables derived from the fifth generation of the European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalyses (ERA5) from 1979 - 2018. https://doi.org/10.24381/cds.20d54e34.

ERA5-Land dataset will also be considered. This dataset is produced by replaying the land component of the ECMWF ERA5 climate reanalysis, combining model data with observations from across the world into a globally complete and consistent dataset. It is currently available from 1981 – present, updated continually with a 5-day lag, with plans to extend back in time to 1950. https://doi.org/10.24381/cds.e2161bac

Future projections will be provided through selection of widely supported global climate model runs, e.g. ISIMIP2b. https://www.isimip.org/protocol/

Ancillary data

A consistent set of global ancillary datasets of topography, soil parameters, vegetation and LAI will be produced for use with the JULES land surface model and archived within the project and NERC Environmental Information Data Centre. Additional ancillaries will be added as required for new model components.

Output variables

A core set of output variables will be defined to capture essential aspects of the hydrological cycle, including the variables defined at the interfaces of the modelling components.