OPEN ACCESS

University of BRISTOL

This electronic thesis or dissertation has been downloaded from Explore Bristol Research, http://research-information.bristol.ac.uk

Author:
Sarrazin, Fanny

Title:
Understanding the sensitivity of karst groundwater recharge to climate and land cover changes at a large-scale

General rights
Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy
Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

• Your contact details
• Bibliographic details for the item, including a URL
• An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

Understanding the sensitivity of karst groundwater recharge to climate and land cover changes at a large-scale

Fanny Sarrazin

Department of Civil Engineering

University of Bristol

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Engineering.

June 2018

Word count: 52153

Abstract

Karst aquifers are an important source of fresh water for agricultural activities and domestic use in many regions of the world. However, groundwater depletion has already started to occur in some karst regions, while future changes in climate, land cover and population will likely exacerbate this issue. Karst areas are highly permeable and produce large amounts of groundwater recharge. As a result, groundwater recharge in these systems may be particularly sensitive to climate and land cover changes compared to less permeable systems. Yet, little effort has been directed toward assessing the impact of climate and land cover change in karst areas at large-scales. This research aims to fill this gap by investigating and developing novel methodologies and a novel hydrological model that enable such an assessment and by analysing the sensitivity of simulated recharge over carbonate rock areas in Europe, the Middle East and Northern Africa.

Firstly, we examine the implementation of Global Sensitivity Analysis (GSA) that has been identified as a key methodology in the context of this thesis to support model development and to identify modelled controls. We introduce novel criteria to assess the robustness of GSA results. Secondly, we propose the first large-scale hydrological model including an explicit representation of vegetation and karst properties (V2Karst). We demonstrate the plausibility of the model predictions at carbonate rock FLUXNET sites using GSA. Thirdly, we apply the V2Karst model and GSA techniques to assess the relative sensitivity of recharge to climate and land cover change across Europe, the Middle East and Northern Africa. Our results reveal that the degree of subsurface heterogeneity of the karst system, the precipitation intensity and the land cover type are important controls of recharge that should be carefully considered in future modelling experiments and in future water and land cover management strategies.

Acknowledgments

I am deeply grateful to my two supervisors, Thorsten Wagener and Francesca Pianosi, for their support, their great availability, their commitment and their patience throughout the PhD. Working under their guidance has been an extremely fruitful experience. I also show considerable appreciation to Andreas Hartmann, who has been an 'unofficial' supervisor during the PhD, and whose enthusiasm for karst hydrology has inspired me to conduct this research.

I wish to thank Rafael Rosolem, for his valuable advice, specifically on evapotranspiration modelling, land surface modelling and data processing, not only during the yearly review discussions and but also on other numerous occasions.

This PhD was only possible thanks to a Scholarship Programme from the University of Bristol Development and Alumni Relations Office.

I acknowledge the contribution to the work presented in this thesis of Olkeba Tolessa Leta, Farkhondeh Khorashadi Zadeh, Ann van Griensven, Toby Dunne, Timothy Foster, Yoshihide Wada, Dimitri Rambourg, investigators of sites from the FLUXNET network (and in particular Martina Mund, Manfred Fink, Penelope Serrano Ortiz, Francisco Domingo Poveda, Andrew Kowalski, Richard Joffre, Serge Rambal, Guillaume Simioni) and Eleonora Canfora and Dario Papale from the FLUXNET European Fluxes Database Cluster.

I thank Justin Sheffield and Miguel Rico-Ramirez for the fruitful discussion during the defence.

I thank all my colleagues from the 'Woodland road office' that have created such a remarkable work environment, and in particular Valentina Noacco, Ludovica Beltrame, Susana Almeida, Joost Iwema, Jude Musuuza, Barney Dobson, Charles West, Wouter Knoben, Mostaquimur Rahman, Sherien Al-Azerji, Carolina Massman, Dongik Kim and Otto Chen for their help and friendship throughout the PhD.

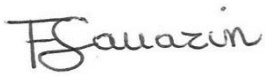
My gratitude also goes to my wonderful flatmates Beatriz Gallo Cordoba, Lina Stein, Chiara Amoretti, Maria Guerrero Farias, Luisa Zemp, Hui Xuan Teoh and Sabrina Fairchild who considerably helped me go through the PhD.

Finally, I wish to express my thanks to my friends and family and in particular to my parents for their unconditional support and encouragements.

Author’s declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED:



DATE: 28/06/2018

List of publications

Chapters 2 to 5 and Appendix A to C of this thesis are based on works which are either published, under review or in preparation as listed below.

Chapter 2 and Appendix A are adapted from:

Sarrazin, F., Pianosi, F. and Wagener, T.: An introduction to the SAFE Matlab Toolbox with practical examples and guidelines, in Sensitivity Analysis in Earth Observation Modelling, edited by G. Petropoulos and P. Srivastava, pp. 363–378, Elsevier Inc., 2017

Sarrazin, F., Pianosi, F. and Wagener, T.: Global Sensitivity Analysis of environmental models: Convergence and validation, Environmental Modelling and Software, 79, 135–152, doi:10.1016/j.envsoft.2016.02.005, 2016.

Chapter 3 and Appendix B are adapted from

Sarrazin, F., Hartmann, A., Pianosi, F. and Wagener, T.: V2Karst v1.0: A parsimonious large-scale integrated vegetation-recharge model to understand the impact of climate and land cover change in karst regions, Geoscientific Model Development Discussion, doi:10.5194/gmd-2017-315, in review.

Chapter 4 and Appendix C are adapted from:

Sarrazin F, Hartmann A, Pianosi F, Rosolem R, Wagener, T,. How do land cover and subsurface heterogeneity modulate climate change impacts on future groundwater recharge in karst areas?, in preparation.

Chapter 5 (Section 5.1) is adapted from:

The abstracts of the above-mentioned works.

Table of contents

List of figures

Page

List of tables

Page

List of acronyms

CART: Classification And Regression Trees

CDF: Cumulative Distribution Function

DES: Desert

EE: Elementary Effect

EET: Elemental Effect Test

ET: Evapotranspiration

EVI: Enhanced Vegetation Index

GCM: General Circulation Model

GLDAS: Global Land Data Assimilation System

GSA: Global Sensitivity Analysis

HRU: Hydrological Response Unit

HUM: Humid

IPCC: Intergovernmental Panel on Climate Change

ISI-MIP: Inter-Sectoral Impact Model Intercomparison Project

KS: Kolmogorov-Smirnov

LAI: Leaf Area Index

LHS: Latin Hypercube Sampling

LSA: Local Sensitivity Analysis

MED: Mediterranean

MTN: Mountain

NDVI: Normalized Difference Vegetation Index

NSE: Nash-Sutcliffe Efficiency

OAT: One-factor-At-the-Time

PET: Potential Evapotranspiration

RCP: Representative Concentration Pathway

RSA: Regional Sensitivity Analysis

SA: Sensitivity Analysis

SSP: Shared Socio-economic Pathway

VOD: Vegetation Optical Depth

VBM: Variance-Based Main effects

VBSA: Variance-Based Sensitivity Analysis

VBT: Variance-Based Total effects

Chapter 1. Introduction

1.1 Karst groundwater recharge: relevance and mechanisms

Aquifers are the largest store of fresh water of the Earth and significantly exceed the volume of surface water present in rivers, lakes and wetlands (Gleeson et al., 2016). Many regions of the world rely on groundwater extraction to support domestic fresh water use and agricultural activities (Giordano, 2009; Siebert et al., 2010). Aquifers are also key to maintaining water discharges to rivers, lakes and wetlands during drier periods, thereby sustaining human water consumption and use, as well as surface freshwater and riparian ecosystems' health (Giordano, 2009; Kløve et al., 2011; Taylor et al., 2013; Hendriks et al., 2014).

Aquifer recharge is a key component of the water balance. Long-term recharge characterises the volume of renewable groundwater available to meet human water consumption and ecosystems demand (Scanlon et al., 2006; Döll and Fiedler, 2008; Wada et al., 2012; Hartmann et al., 2017). When an aquifer is over-exploited, i.e. when groundwater extraction rate is large compared to average recharge rate, a persistent decrease in the volume of groundwater storage and in groundwater levels (groundwater depletion) may be observed (Custodio, 2002). Groundwater depletion has already been observed in many regions of the world, with often disastrous consequences for water security and biodiversity (Custodio, 2002; Bartolino and Cunningham, 2003; Ferguson and Gleeson, 2012; Wada, 2016; Rodell et al., 2018). As explained in these studies, groundwater depletion can lead to water scarcity, since extracting water from depleted aquifers to meet human water demand may become too expensive or technically unfeasible. It can also reduce surface water flows, in particular during drier periods, which endangers aquatic and riparian ecosystems. Other well-known detrimental effects of groundwater depletion mentioned in these studies are sea-level rise, salt-water contamination and land subsidence. Consequently, understanding groundwater recharge is crucially needed to inform the design of sustainable water management strategies (Gleeson et al., 2010; Aeschbach-Hertig and Gleeson, 2012).

Karst systems typically develop in carbonate rock areas, which cover about 10-15% of the world and as much as 14-29% of Europe land areas (Ford and Williams, 2007; Chen et al., 2017). Karst aquifers are particularly relevant for water management, since they are an important source of drinking water for 10-25% of the world population (Ford and Williams, 2007; Stevanović, 2018). Karst aquifers are also crucial in sustaining food production, since agricultural activities take place in many karst areas (Coxon,

2011). Karst groundwater resources are heavily used in Europe, the Middle East and Northern Africa. They supply fresh water to major cities such as Damascus in Syria, Rome, Naples and Bari in Italy, and Marseille and Montpelier in France, and they account for up to 50% of the total water supply in countries such as Austria and Slovenia (COST, 1995; Stevanović, 2018). Surface and underground karst systems also tend to have high ecological value. They can host a multitude of species, that strongly depend on water flow and groundwater recharge, among which rare and endangered species (Bonacci et al., 2009).

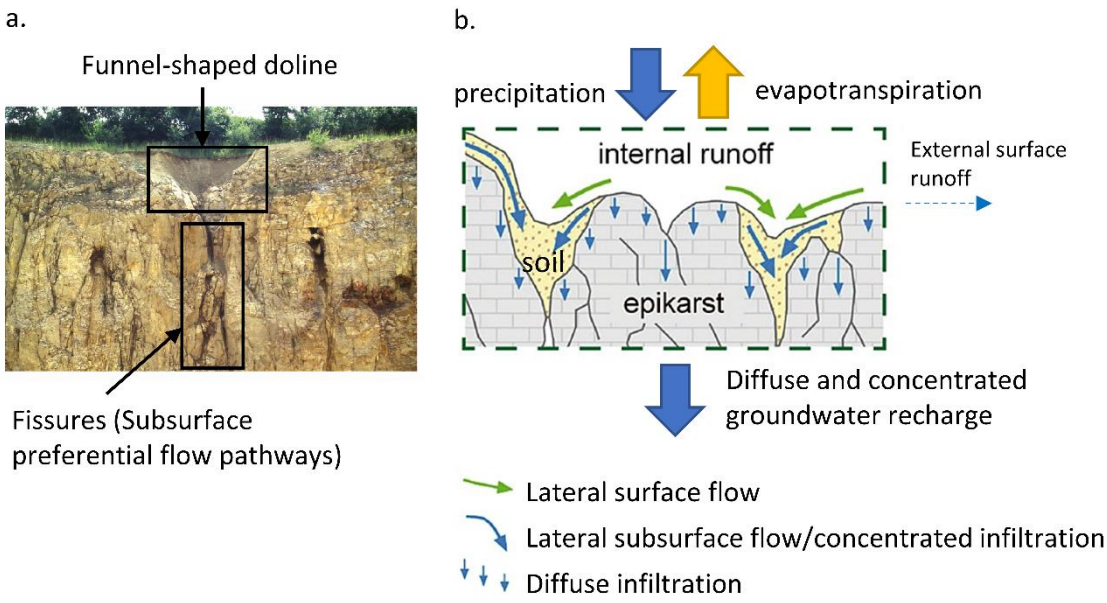


Figure 1.1 Preferential flow pathways and flow mechanisms in karst.
(a) Epikarst cross-section of a karst system in the Swabian Alb, Germany, showing a funnel-shaped doline (surface depression in which surface water can be concentrated) and fissures that can act as subsurface preferential flow pathways. (b) Conceptual representation of flow mechanisms in karst. High volumes of groundwater recharge can be produced due to the presence of higher permeability areas in which infiltrating water can be concentrated through lateral surface and subsurface flow. External surface runoff out of the karst system is typically negligible.
Source of the pictures: (a) photo by Nico Goldscheider. (b) scheme adapted from Figure 10 in Hartmann et al. (2014).

Karst systems are characterised by particularly high amounts of groundwater recharge due to their high infiltration capacities, while surface runoff is typically negligible (Fleury et al., 2007; Contreras et al., 2008; Hartmann et al., 2014). In fact, karst systems typically develop from the chemical weathering or dissolution of carbonate rocks by infiltrating water and atmospheric carbon dioxide CO_2 (Ford and Williams, 2007). This process, referred to as 'karstification', creates large dissolution conduits or fissures in the bedrock. Figure 1.1 shows an example of dissolution features and illustrates the flow mechanisms occurring in karst systems. Recharge and infiltration in karst can be fast and concentrated in the dissolution features that act as preferential flow pathways and slow and diffuse in the matrix (Maréchal et al., 2008; Hartmann and Baker, 2017). Lateral flow at the surface and in the subsurface is

an important process that concentrates the infiltrating water into the preferential flow pathways. Lateral flow at the land surface can occur along surface depressions that are characteristics of karst landscapes such as dolines (Fig. 1.1a) (Jeannin and Grasso, 1997; Hartmann et al., 2014). In the subsurface and more specifically in the epikarst, which is the upper layer of weathered carbonate rocks with the highest degree of dissolution and the highest storage capacity, pressure gradients can generate lateral flow towards large conduits or fissures (Williams, 1993, 2008).

Critically, the conjunction of changes in climate and excessive pumping has already led to groundwater depletion in some karst areas, and in particular to the drying of major karst springs such as Ra el Ain spring in the Khabour basin in Syria (Zakhem and Kattaa, 2017), the Faria spring in Palestine (Hartmann et al., 2012a), and the Jinci spring in the Shanxi Province in China (Li et al., 2012; Jia et al., 2017). Therefore, understanding future karst groundwater recharge is crucial to prevent further groundwater depletion and ensure water security in these regions.

1.2 Projected changes in climate, land cover and population in karst areas

Changes in the radiative budget and in climate are projected to occur worldwide over the 21st century. Climate models indicate, on average, a general increase in air temperature and a decrease in near-surface relative humidity over many land areas. The projected direction of change in precipitation, instead, varies tremendously across regions and seasons (Intergovernmental Panel on Climate Change IPCC, 2013, Fig.12.11, 12.21 and 12.22). It is also widely accepted that changes in land cover will happen. Anthropogenic changes in land cover are expected due to modification in socio-economic factors such as food and wood demand or agricultural yields, and due to the possible implementation of climate adaptation, climate mitigation and ecosystem conservation strategies (Hurtt et al., 2011; Lawrence et al., 2016; Holman et al., 2017). Natural changes in land cover and in vegetation characteristics, e.g. leaf area index, are likely to take place in response to changes in environmental conditions such as modifications in atmospheric carbon dioxide, nitrate deposition and climate, and in response to natural disturbances such as wildfire, storm, bark beetle (Seidl et al., 2014; Zhu et al., 2016). In addition to changes in climate and in land cover, population is expected to increase globally (United Nations, 2017) with commensurate increases in water demand.

However, future climate, land cover and population projections are determined through modelling experiments and are highly uncertain because of:

- The uncertainty in future scenarios, such as the emission and associated radiative forcing scenario (Representative Concentration Pathway, RCP, Van Vuuren et al., 2011), the demographic and socio-economic scenario (Shared Socio-economic Pathways, SSP, Van

Vuuren et al., 2014; United Nations, 2017) and the policy scenario (Shared climate Policy Assumptions, Van Vuuren et al., 2014).

- The availability of different model structures and parameter values for the projection models, namely General Circulation Models (GCMs) for climate projections (IPCC, 2013), land use/land cover projection models (Hurtt et al., 2011; Lawrence et al., 2016; Holman et al., 2017) and population projection models (United Nations, 2017);
- The multiplicity of downscaling techniques used to derive high spatial resolution projections at scales that are relevant for impact studies (Ekström et al., 2015, for climate downscaling; CIESIN, 2017a, for population downscaling).

Consequently, for a given spatial location, climate, land cover and population projection ensembles tend to provide a wide range of possible scenarios. For instance, Hattermann et al. (2018) analysed the change in annual precipitation over the period 2010-2099 for the highest radiative forcing scenario (RCP8.5). Their study found that different GCMs produce very different projections, and in particular that for 74.6% of the world continental area, less than 80% of GCMs agree in the sign of change. Still, projection ensembles are likely to underestimate the range of plausible futures because they do not exhaustively account for all sources of uncertainty along the modelling chain (Stainforth et al., 2007).

Figures 1.2-1.5 further exemplify the uncertainty in future annual precipitation, temperature, forest fraction and population projections for the carbonate rock areas in Europe, the Middle East and Northern Africa over a 0.5°x0.5° grid. Panels (a) in Figures 1.2-1.5 show the historical states, while panels (b-c) show the lower and upper bound of the projected changes between the historical period and the end of the 21st century. Future changes were assessed using twenty future climate scenarios derived from four RCPs and five GCMs (Inter-Sectoral Impact Model Inter-comparison Project, ISI-MIP, Warszawski et al., 2014), four future forest fraction scenarios (Harmonized Global Land Use database; Hurtt et al., 2011; Chini et al., 2014), and county-based population projections (United Nations, 2017) downscaled using a gridded population database (CIESIN, 2017a, 2017b). Figures 1.2-1.5 show that the magnitude and direction of change vary greatly in the projection ensembles. Projected precipitation, forest fraction and population could either increase or decrease for 90%, 17% and 57% of the cells respectively. We note that land cover projections do not indicate any change in the forest cover fraction for 80% of the cells (yellow areas in Fig. 1.3 b-c) due to specific modelling assumptions that were used to develop the land cover projections (i.e. it is considered that these areas cannot support forest land cover, see Hurtt et al., 2011). Mean annual temperature is projected to increase in all scenarios analysed, but the uncertainty range of the change in mean annual temperature is larger than 3.6°C for all cells and even larger than 5°C for 80% of the cells.

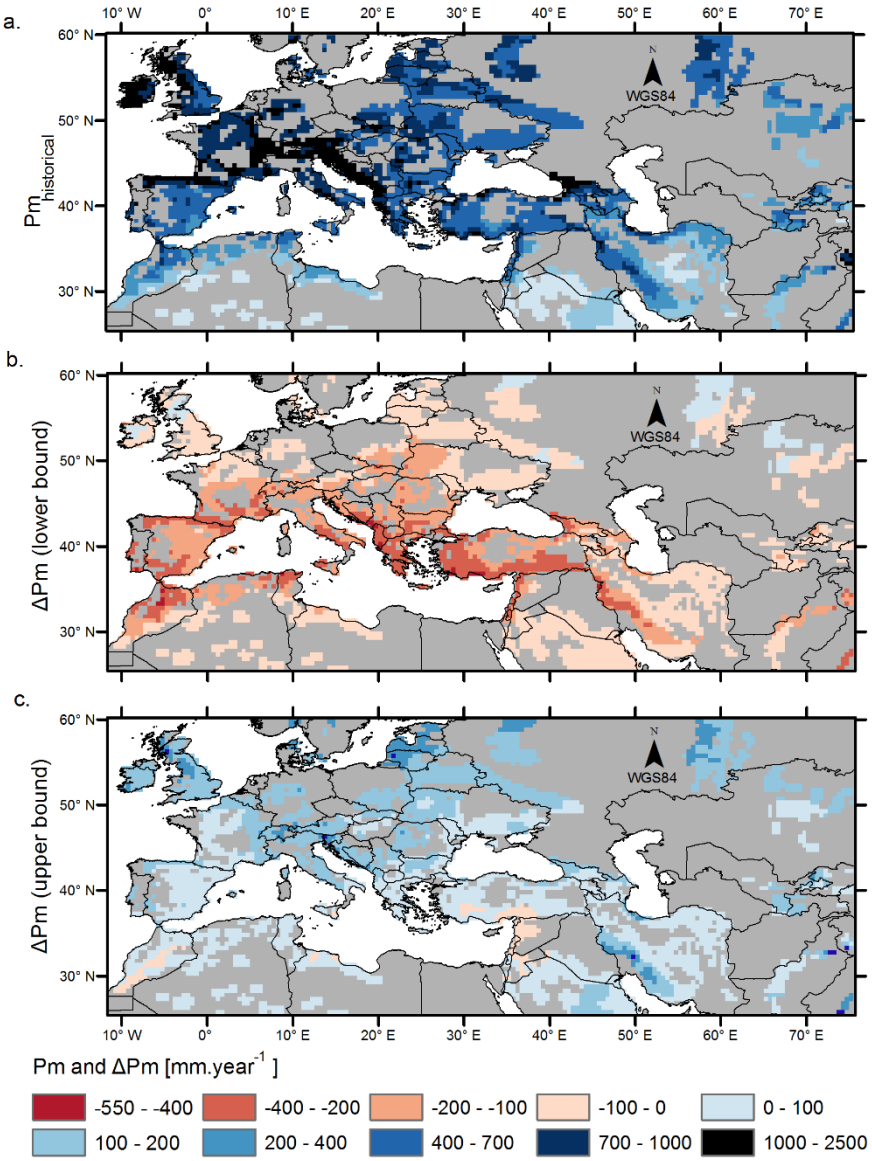


Figure 1.2 Historical mean annual precipitation for the period 1974-2004 (top) and projected changes from the period 1974-2004 to 2069-2099 in carbonate rock areas in Europe, the Middle East and Northern Africa.

(a) Historical mean annual precipitation $Pm_{historical}$ $[mm.year^{-1}]$, (b) minimum projected change in mean annual precipitation $\Delta Pm = Pm_{future} - Pm_{historical}$ $[mm.year^{-1}]$, (c) maximum projected change in mean annual precipitation. Data are at 0.5°x0.5° resolution.

Source of the data: Carbonate rock and country map from Williams and Ford (2006) and in particular the country map was obtained from Terraspace, Russian space agency. Gridded precipitation from the Inter-Sectoral Impact Model Inter-comparison Project (ISI-MIP, Warszawski et al., 2014). Historical and future mean annual precipitation were assessed over the 30-year period 1st October 1974 - 30th September 2004 and 1st October 2069 - 30th September 2099 respectively. The value of historical mean annual precipitation reported in (a) is the average value over five GCMs. Minimum and maximum projected changes reported in (b-c) were determined over twenty future scenarios (four RCPs and five GCMs).

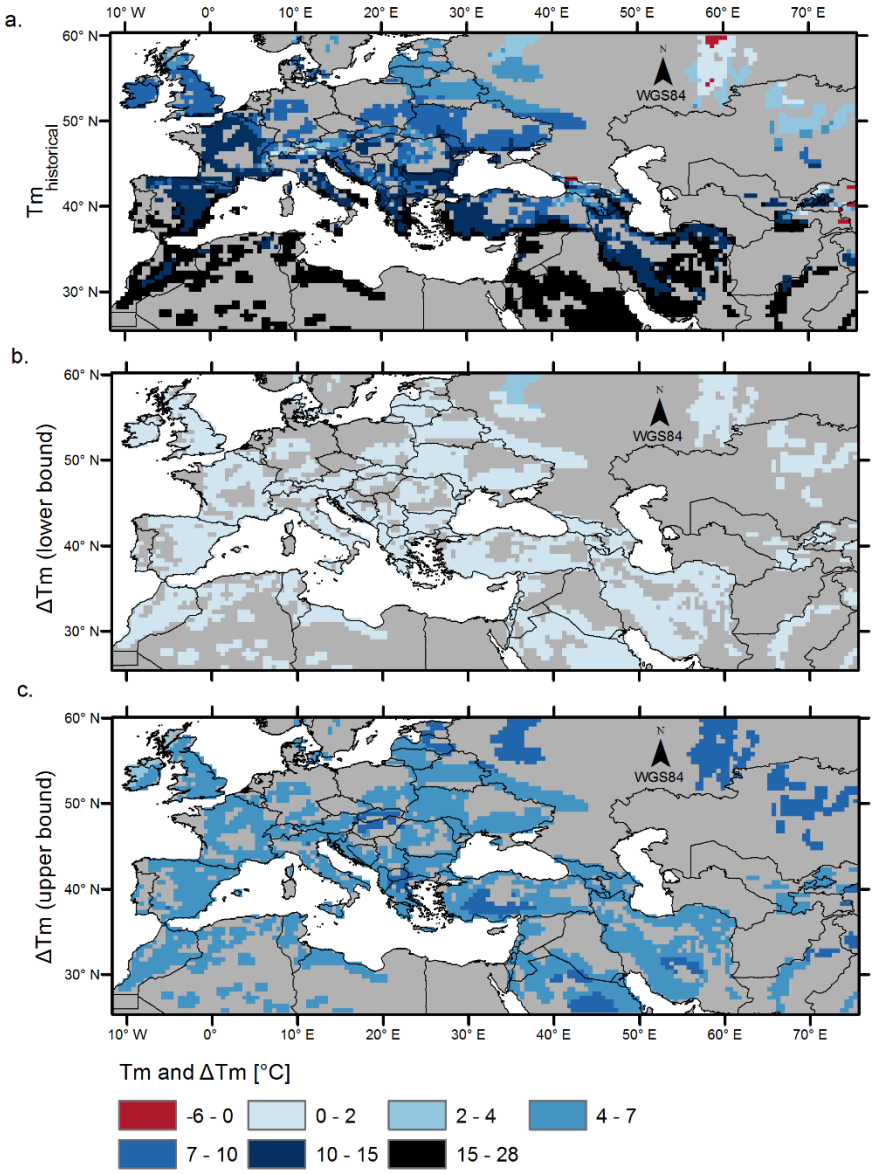


Figure 1.3 Historical mean annual temperature for the period 1974-2004 (top) and projected changes from the period 1974-2004 to 2069-2099 in carbonate rock areas in Europe, the Middle East and Northern Africa.

(a) Historical mean annual temperature $Tm_{historical}$ [$°C$], (b) Minimum projected change in mean annual temperature $\Delta Tm = Tm_{future} - Tm_{historical}$ [$°C$], (c) Maximum projected change in mean annual temperature. Data are at 0.5°x0.5° resolution.

Source of the data: Carbonate rock and country map from Williams and Ford (2006) and in particular the country map was obtained from Terraspace, Russian space agency. Gridded precipitation from the Inter-Sectoral Impact Model Inter-comparison Project (ISI-MIP, Warszawski et al., 2014). Historical and future mean annual temperature were assessed over the 30-year period 1st October 1974 - 30th September 2004 and 1st October 2069 - 30th September 2099 respectively. The value of historical mean annual temperature reported in (a) is the average value over five GCMs. Minimum and maximum projected changes reported in (b-c) were determined over twenty future scenarios (four RCPs and five GCMs).

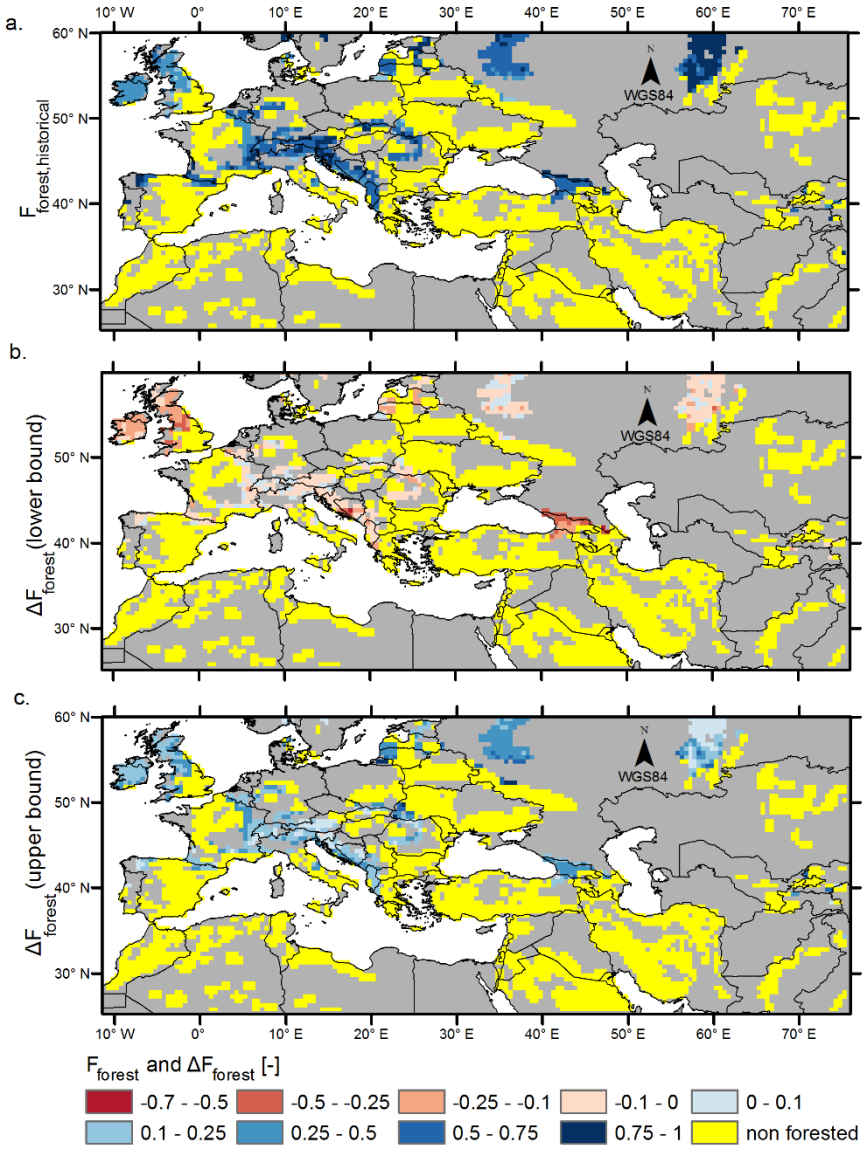


Figure 1.4 Historical forest cover fraction for year 2004 (top) and projected changes from 2004 to 2099 in carbonate rock areas in Europe, the Middle East and Northern Africa.

(a) Historical forest cover fraction $F_{forest,historical}[-]$ assessed for year 2004, (b) Minimum projected change in forest cover fraction $\Delta F_{forest} = F_{forest,future} - F_{forest,historical}$ $[-]$, (c) Maximum projected change in forest fraction. It is assumed that areas marked in yellow cannot support forest land cover over time (Hurtt et al., 2011). Data are at 0.5°x0.5° resolution.

Source of the data: Carbonate rock and country map from Williams and Ford (2006) and in particular the country map was obtained from Terraspace, Russian space agency. Gridded forest fraction and non-forested areas from Harmonized Global Land Use database V1.0 (Chini et al., 2014). Minimum and maximum projected changes reported in (b-c) were determined over four future scenarios corresponding to four RCPs.

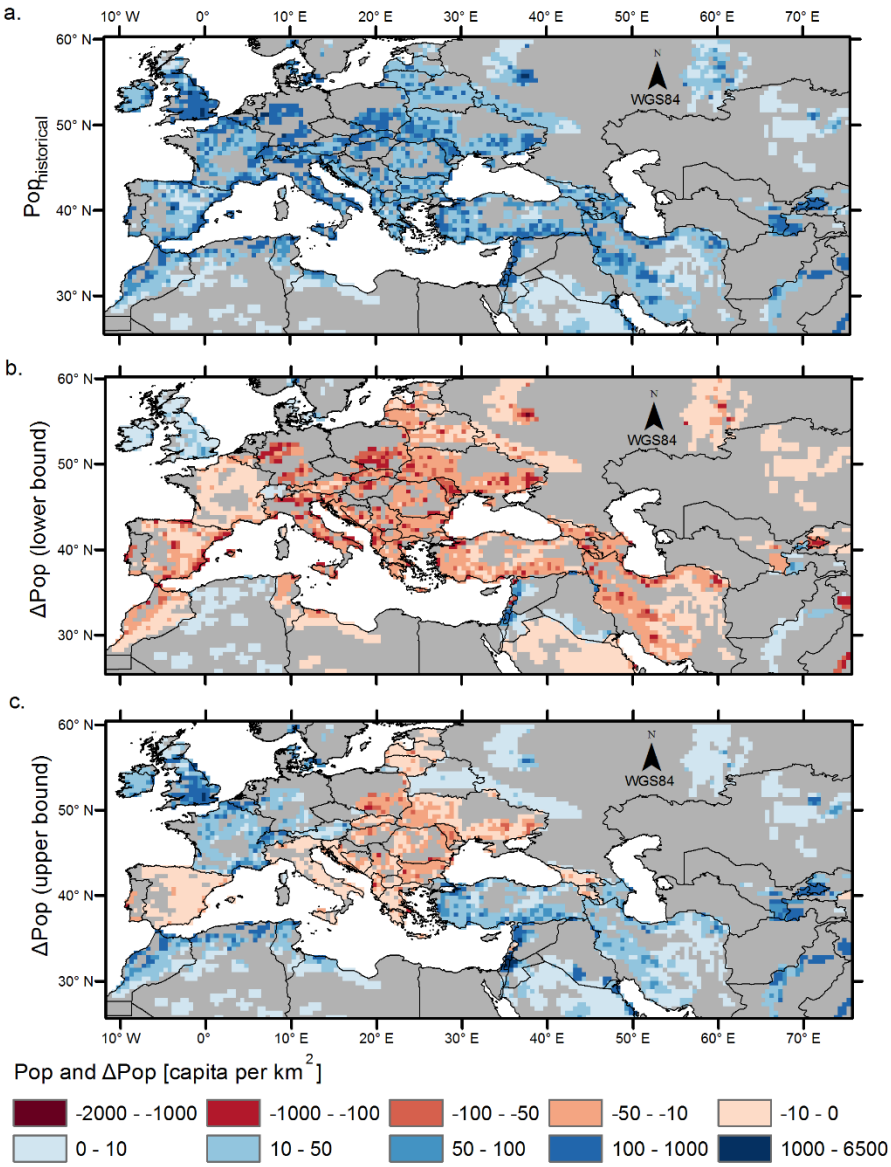


Figure 1.5 Historical population density for year 2015 (top) and projected changes from 2015 to 2100 in carbonate rock areas in Europe, the Middle East and Northern Africa.

(a) Historical population density $Pop_{historical}$ $[capita\ per\ km^2]$, (b) Lower bound of the 95% confidence interval of the projected change in population density $\Delta Pop = Pop_{future} - Pop_{historical}$ $[capita\ per\ km^2]$, (c) Upper bound of the 95% confidence interval of the projected change in population density. Data are at 0.5°x0.5° resolution.

Source of the data: Carbonate rock and country map from Williams and Ford (2006) and in particular the country map was obtained from Terraspace, Russian space agency. Gridded population density from (CIESIN, 2017b); projected population per country from (United Nations, 2017). To produce gridded projections of population density, the population projections per country were downscaled using the gridded population density database and assuming a spatially uniform percentage change within each country. Minimum and maximum projected changes reported in (b-c) are the 95% confidence intervals as indicated by (United Nations, 2017)

In summary, future projections suggest that climate, and in particular mean annual precipitation and temperature, land cover and population will change in karst areas, but the magnitude and direction of change are highly uncertain.

1.3 Modelling groundwater recharge at a large-scale

It is widely understood that changes in climate, land cover and population have an impact on water scarcity, since changes in population can modify water demand (Vörösmarty, 2002; Wada et al., 2012), while changes in climate and land cover can affect water availability. Changes in precipitation alter the moisture supply to the system, while changes in other climate factors (e.g. temperature, humidity) influence the evaporative demand and therefore the amount of water which is lost back to the atmosphere (Monteith, 1995; Shuttleworth, 2012). Changes in land cover further affect the partitioning of precipitation between green water, i.e. evapotranspiration losses, and blue water, i.e. water that could potentially be used to meet human water demand, namely groundwater recharge and runoff (Falkenmark and Rockström, 2006). Forest land cover typically leads to higher evapotranspiration losses and lower water yield compared to shorter vegetation (Brown et al., 2005; Gordon et al., 2005). While karst systems produce particularly high amounts of recharge compared to other less permeable systems, their hydrology is also likely to be more sensitive to changing conditions at the land surface (climate and land cover) because preferential flow pathways strongly connect the land surface and the groundwater (Hartmann et al., 2017).

Hydrological models allow for a quantitative assessment of the impact of changing climate and land cover on water availability. However, this assessment involves a number modelling choices, such as the choice of the processes that are included in the model and the choice of the specific equations and parameter values to represent the selected processes. These choices tend to be poorly constrained for large-scale applications because of:

- a lack of observations of hydrological variables to compare with model predictions. For instance, evapotranspiration measurements are limited in number and generally do not separate the different evapotranspiration components, i.e. vegetation transpiration, soil evaporation and evaporation from canopy interception (Baldocchi et al., 2001; Miralles et al., 2016);
- a lack of observations of vegetation and soil properties to constrain model parameters. For instance measurement of some vegetation properties such as stomatal resistance or canopy interception capacity are scarce and affected by large uncertainties (Körner, 1995; Eckhardt et al., 2003; Van Dijk et al., 2015);
- an incommensurability between model parameters and physical properties, as shown for instance in Hogue et al. (2006) and in Rosero et al. (2010). As discussed in Beven (2001) and

Kirchner et al. (2006), this incommensurability can be explained by (1) the scale mismatch between physical variables that are measured in the field, often at point scale, and model parameters that represent average spatial properties over the model's spatial unit, which typically ranges from 9 to 111 km^2 for large-scale applications (Bierkens, 2015), and (2) the fact that governing equations that describe the behaviour of the system over small scales may not be valid for model applications over larger domains.

These modelling choices have been shown to have a significant impact on hydrological projections and they introduce further uncertainties besides the expected uncertainties in future climate and land cover projections discussed in Sect. 1.2. A large number of model inter-comparison studies have been carried out to assess the impact of climate change on different hydrological aspects (different runoff quantiles in Vetter et al., 2015; low flows in Marx et al., 2018; annual distribution of river discharge in Melsen et al., 2018; river floods and high flows in Thober et al., 2018). These studies have established that future hydrological projections are highly sensitive not only to the choice of the climate scenario, but also to the choice of the hydrological model. More specifically, Sheffield et al. (2012) found that different potential evapotranspiration equations can produce very different magnitudes and even directions of change in drought severity. Other studies have revealed that the choice of parameter values and parameter estimation strategy can alter the projected hydrological impacts of climate change (Mendoza et al., 2015b, 2016; Melsen et al., 2018) and of land cover/land use change (Eckhardt et al., 2003; White et al., 2017).

When developing a model of groundwater recharge in karst areas, additional properties and processes need to be taken into account, such as the heterogeneity in soil and epikarst water storage capacity and permeability and the process of concentration flow (see Sect. 2.1). Most current large-scale hydrological models do not explicitly include these processes and properties. Some large-scale models have a homogeneous representation of subsurface sub-grid properties, namely WBM (Vörösmarty et al., 1998), LaD (Milly and Shmakin, 2002), LPJ (Gerten et al., 2004) and GLEAM (Martens et al., 2017; Miralles et al., 2011). Other hydrological models use a statistical distribution of sub-grid storage or infiltration capacity in their runoff generation routine, but consider homogeneous soil properties to assess evapotranspiration and infiltration below the soil, namely WaterGap (Döll et al., 2003; Müller Schmied et al., 2014), Mac-PDM (Arnell, 1999; Gosling and Arnell, 2011), PCR-GLOBWB (Van Beek and Bierkens, 2008) and VIC (Bohn and Vivoni, 2016; Liang et al., 1994). The VarKarst model (Hartmann et al., 2015) is the only large-scale hydrological model that was specifically developed for karst areas. VarKarst has an explicit representation of sub-grid subsurface heterogeneity, as it explicitly evaluates the water balance over a number of vertical compartments with different storage capacities and permeability and it includes the process of concentration flow. VarKarst has been shown to produce

overall larger and more realistic recharge estimates compared to other large-scale model representations (Hartmann et al., 2015, 2017). Importantly, the application of VarKarst over carbonate rocks areas in Europe, the Middle East and Northern Africa, revealed different sensitivities of recharge to changes in climate, and in particular produced much larger absolute changes in recharge compared to a non-karst model representation (Hartmann et al., 2017).

Yet, the VarKarst model cannot be used for land cover change impact assessment because it does not include an explicit representation of vegetation properties. No study of the combined impact of climate and land cover change has indeed been conducted so far on karst groundwater recharge. Therefore, a new model is needed to perform such assessment at a large-scale, and the design of that model will require careful consideration of the specific challenges associated with large-scale model development and application with sparse and uncertain observations.

1.4 Research questions

The overall goal of this thesis is to improve our understanding of the impact of changes in climate and land cover on karst groundwater recharge at large-scales and over large domains, ultimately informing the design of sustainable water management strategies in karst areas. In pursuing this goal, the thesis seeks to use strategies that consider the large uncertainties inherent in climate and land cover projections and in large-scale hydrological models.

The thesis investigates potential groundwater recharge, which is defined as the vertical water flux leaving the bottom of the weathered bedrock (epikarst) and does not account for groundwater storage and groundwater flow. It examines water quantity only and not water quality. It considers the impact of changes in land cover, which are the physical characteristics of the land surface (e.g. grass, forest, bare soil), and does not account for land use and management practices. As a first large-scale study of the combined effect of changes in climate and land cover in karst areas, this thesis does not examine the effect of changes in atmospheric CO_2 concentration on the vegetation and on the degree of karstification of the karst systems. Such in-depth investigation is left for future research and is further discussed in Sect. 3.2.2 and 5.2.

Carbonate rock areas in Europe, the Middle East and Northern Africa have been selected as case study, given their large extent and the importance of karst groundwater resources in this region (Sect. 1.2), the large changes in climate and land cover projected for this region (Fig. 1.2-1.4), and finally the fact that water demand may significantly increase in part of this region because of population growth (Fig. 1.5c).

Specifically, the thesis focuses on the following research questions:

1. *How can we assess the sensitivity of groundwater recharge to climate properties and land cover type in karst areas?*
2. *How sensitive is groundwater recharge in karst areas across Europe, the Middle East and Northern Africa to changes in climate and land cover?*

Regarding (1), we first aim to develop a hydrological model that allows to simulate groundwater recharge at a large-scale in karst areas and under changing climate and land cover. Such model does not exist at present as discussed in Sect. 1.3. The specificity of the approach used in this research is that it seeks to include a parsimonious representation of the model processes, given the limited data availability to constrain model simulations at a large-scale, and to explicitly account for parameter uncertainty.

To address (1), we also need to identify a strategy to test the adequacy of the newly developed model and the plausibility of its predictions, and to analyse the simulated groundwater recharge under changing climate and land cover so to maximise the amount of information that is extracted from the simulation dataset. Global Sensitivity Analysis (GSA) has been chosen as a key methodology for these purposes in the context of this thesis. GSA is a set of statistical techniques that aim to apportion the uncertainty in the output of a model (e.g. groundwater recharge) among its different uncertain input factors (e.g. climate property, land cover type, model parameters) (Saltelli et al., 2008; Norton, 2015; Pianosi et al., 2016). This methodology is very useful to guide model development and calibration, since it can inform the choice of model structure and parameter values (Hartmann et al., 2013b; Vanuytrecht et al., 2014; Mockler et al., 2015; Cuntz et al., 2016). Additionally, GSA can be used to derive crucial information on the relationship between the output of a model and its input factors, and in particular it can support the identification of the dominant controls of the model's behaviour (Van Werkhoven et al., 2008; Anderson et al., 2014; Singh et al., 2014). However, the implementation of GSA faces a range of challenges and needs to be careful examined to ensure that the results of GSA are robust and credible (Shin et al., 2013; Pianosi et al., 2016). Consequently, the initial part of the thesis is dedicated to the presentation and investigation of GSA techniques, and offers new solutions to enable their robust implementation.

With respect to (2), we use the newly developed large-scale hydrological model and a set of selected GSA techniques to assess the sensitivity of groundwater recharge to changes in climate and land cover in carbonate rock areas in Europe, the Middle East and Northern Africa. So far, climate change impact studies have been generally focused on the forward propagation of uncertainties (arising from the choice of the climate and land cover scenarios and the hydrological model structure and parameter values) into the hydrological projections. This approach typically results in a very large range of possible future

hydrological conditions and therefore tends to provide limited actionable information to decision makers (Wilby and Dessai, 2010; Kundzewicz et al., 2018). The difference in the research presented here is that it also aims to identify the controls of karst groundwater recharge by uncovering the relationship between the simulated groundwater recharge and a range of climate properties (e.g. precipitation intensity, precipitation seasonality, mean temperature) and types of land cover (e.g. forest or grass), while considering the variability in the degree of subsurface heterogeneity observed in karst systems and parameter uncertainty. A similar approach to assessing the hydrological impact of climate and land cover change has been pursued in a few previous studies, such as the one by Singh et al. (2014) that analysed the effect of changing mean precipitation, mean temperature and land cover on streamflow and the one by Guo et al. (2017) that examined the impact of a range of climate properties on runoff.

1.5 Thesis outline

The thesis is structured as follows:

Chapter 2 reviews GSA techniques and their possible uses, focusing on the assessment of the convergence and validation of GSA results and introducing novel criteria and guidance to support the implementation of GSA.

Chapter 3 introduces a large-scale integrated vegetation-recharge model to simulate the impact of climate and land cover change in karst regions. The plausibility of model predictions is tested at four European carbonate rock sites for which some flux observations are available through the FLUXNET project (Baldocchi et al., 2001). The chapter also investigates the mechanisms of recharge production in the model and its sensitivity to climate properties and land cover.

Chapter 4 examines the sensitivity of groundwater recharge to climate, land cover type and degree of subsurface heterogeneity in carbonate rock areas in Europe, the Middle East and Northern Africa. In this analysis, we use climate projections from the Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP, Warszawski et al., 2014) ensemble and selected land cover scenarios.

Finally, Chapter 5 summarises the findings and contributions of the thesis and proposes directions for future research.

Chapter 2. Global Sensitivity Analysis: a methodology for guiding model development and identifying modelled controls

This chapter is adapted from a published work that has been partially modified to improve the consistency throughout the thesis:

1. *Sarrazin, F., Pianosi, F. and Wagener, T.: Global Sensitivity Analysis of environmental models: Convergence and validation, Environmental Modelling and Software, 79, 135–152, doi:10.1016/j.envsoft.2016.02.005, 2016.*

Sections 2.2 and 2.8 use the text from an additional published work:

2. *Sarrazin, F., Pianosi, F. and Wagener, T.: An introduction to the SAFE Matlab Toolbox with practical examples and guidelines, in Sensitivity Analysis in Earth Observation Modelling, edited by G. Petropoulos and P. Srivastava, pp. 363–378, Elsevier Inc., 2017.*

For these two works, partial support to Francesca Pianosi and Thorsten Wagener was provided by the UK Natural Environment Research Council [Consortium on Risk in the Environment: Diagnostics, Integration, Benchmarking, Learning and Elicitation (CREDIBLE); grant number NE/J017450/1]. Work (1) has also benefited from the contributions of Olkeba Tolessa Leta who set up the SWAT model for the Zenne River basin, Farkhondeh Khorashadi Zadeh and Ann van Griensven who provided support for the implementation of the SWAT model and Toby Dunne who provided computer support. Three anonymous referees also contributed to the improvement of work (1) thanks to their many useful suggestions.

2.1 Definition and objectives of Global Sensitivity Analysis

Sensitivity Analysis (SA) aims to characterise the impact that changes in the input factors of a model have on the model output. SA can be used to investigate the sensitivity of any type of numerical or non-numerical input factors such as model parameters (e.g. Vanuytrecht et al., 2014; Hamdia et al., 2015), model structures (e.g. Baroni and Tarantola, 2014; Dai and Ye, 2015; Dai et al., 2017), initial states (e.g. Pianosi et al., 2017), input data (e.g. Singh et al., 2014; Pianosi and Wagener, 2016), output measurements (e.g. Baroni and Tarantola, 2014; Pianosi and Wagener, 2016), or time/spatial resolution grid (e.g. Savage et al., 2016). The model output can be any scalar variable, e.g. an objective function (e.g Root Mean Squared Error) or a summary statistic of the simulated model output for temporally or spatially distribution inputs (e.g. average or maximum simulated streamflow).

Sensitivity Analysis can be either *local* or *global*. Local Sensitivity Analysis (LSA) investigates the model output sensitivity under a limited range of possible input conditions, typically around a model reference setting (Hill and Tiedeman, 2007, for an example application to groundwater model; see e.g. Ljung, 1999, for a general link between LSA and model calibration). Importantly, LSA typically does not account for interactions among input factors. In contrast, Global Sensitivity Analysis (GSA) aims to explore the entire space of variability of the model input factors. The present thesis focuses on GSA techniques, since they permit a more comprehensive characterisation of model output sensitivity.

GSA is a diagnostic tool that can guide model parameter estimation, model verification, support the prioritization of efforts for uncertainty reduction, or help with model-based decision-making (Norton, 2015; Song et al., 2015; Pianosi et al., 2016). Such purposes are generally implemented as four different settings of GSA: (1) screening (or Factor Fixing), (2) ranking (or Factor Prioritization), (3) Variance Cutting, and (4) Factor Mapping (Saltelli et al., 2008; Pianosi et al., 2016).

Screening:
the identification of those input factors, if any, which have no influence on the model output and therefore can be fixed to any value within their feasible space of variability with negligible implications on the output. Screening can be used as a preliminary step to inform a subsequent calibration, which is tailored to the subset of influential parameters, while non-influential parameters can be set to default values (e.g. Kannan et al., 2007; Vanuytrecht et al., 2014; Cuntz et al., 2016). Screening can also be used to guide model development. It reveals those model structure that could be simplified because they have a large number of non-influential or non-identifiable parameters given the state of data availability, and those model structures that may instead be most appropriate because most of their parameters are influential (e.g. Rosolem et al., 2012; Hartmann et al., 2013b; Mockler et al., 2015).

Ranking:

the ordering of the input factors according to their relative influence on the model output. It can enhance our understanding of the model and to identify dominant controls of the model's behaviour (e.g. Van Werkhoven et al., 2008a; Guo et al., 2017) as well as to prioritise efforts for uncertainty reduction (e.g. Sin et al., 2011; Savage et al., 2016). Ranking has been used in numerous climate change impact studies to assess the relative impact that the different uncertain components of the modelling chain (such as Representative Concentration Pathway, Global Circulation Model, downscaling technique or hydrological model, as discussed in Sect. 1.2) have on the hydrological projections (e.g. Déqué et al., 2007; Vetter et al., 2015; Giuntoli et al., 2015; Vidal et al., 2016; Mishra et al., 2017; Hattermann et al., 2018). The aim was not only to prioritise efforts for uncertainty reduction, but also to guide the design of future impact assessment studies, i.e. to identify which sources of uncertainty should be included in future studies. Ranking can also be used to perform a model diagnostic evaluation (verification of model structures) in that it allows to check the consistency of the model sensitivities with expectations (e.g. Rosero et al., 2010) and, when output observations are available, to identify the model components that are responsible for the model poor performance (e.g. Reusser and Zehe, 2011).

Variance Cutting:

the reduction of the output variance to a value below a user chosen tolerance. It aims at obtaining specific sensitivities for the different input factors and is, for example, applied in reliability and risk assessment (Saltelli and Tarantola, 2002).

Mapping:

the identification of the conditions (e.g. sub-ranges of input factors like parameters or forcing inputs) that produce specific values of the output that are of interest to the GSA user. It can be used to enhance model understanding (e.g. Spear and Hornberger, 1980). It has been applied in climate and land cover change impact studies to support robust decision-making, and in particular to identify adverse thresholds on model inputs beyond which hydrological variables take critical values (Singh et al., 2014; Almeida et al., 2017).

2.2 Performing Global Sensitivity Analysis

2.2.2 Workflow for application of Global Sensitivity Analysis

Application of GSA often consists in the calculation of indices that measure the sensitivity of the input factors or in the visual qualitative inspection of the model output. Generally, the implementation of GSA is sampling-based and the value of the sensitivity indices are approximated using Monte Carlo (MC) simulations, apart from rare cases in which an analytical computation of the sensitivity indices is

possible, for instance for benchmark functions widely used in the GSA literature (e.g. Ishigami-Homma function, Ishigami and Homma, 1990; Sobol' g-function, Saltelli et al., 2008). GSA is performed following three basic steps, namely (1) sampling of the input factor space, (2) model evaluation (the most computationally demanding step) and (3) post processing of the input-output samples by different qualitative and quantitative GSA methods.

For each of these three steps, the GSA user needs to make a number of choices that can significantly affect the results. (1) The sampling step requires defining the inputs that will be subject to GSA, the input probability distributions (and input ranges), the sampling strategy to explore the input factor space and the sample size. (2) The model evaluation step necessitates, among others, selecting the simulation time period, when models are evaluated using time series of input data (e.g. hydrological models), and the model output(s) to be analysed. (3) The post-processing step involves choosing the GSA methods (among the visual or quantitative methods available) and possibly performing some filtering of the input-output sample. These choices are further discussed in Table 2.1, while the choice of the sample size will be examined later in this Chapter (Sect. 2.3-2.6). Importantly, Table 2.1 reveals that these choices depend on the objective of the analysis and on the specific case study, and therefore they need to be examined on a case-by-case basis. Testing and revising these choices is an important part of sensitivity analysis, to ensure the robustness of the results. In particular, previous studies have recommended to perform multi-method and multi-criteria analyses for a more exhaustive characterisation of model sensitivities (Rosolem et al., 2012; Borgonovo et al., 2017; Pianosi et al., 2017) and to examine model sensitivities under different environmental conditions (Van Werkhoven et al., 2008a; Vanuytrecht et al., 2014).

The three basis steps of GSA are implemented for instance in the open source software SAFE (Sensitivity Analysis For Everybody; Pianosi et al., 2015; Sarrazin et al., 2017; available at https://www.safetoolbox.info/), that includes a range of quantitative GSA methods and visualisation tools. The Matlab SAFE toolbox has been used to perform all sensitivity analyses in this thesis.

Step 1. Sampling
Q: How many input factors can be handled by GSA? A: *There is no upper limit to the number of input factors handled by GSA. However, when the number is large, the computational cost of the analysis may be prohibitive as discussed in Sect. 2.3-2.6.*
Q: How can the input probability distributions (and input ranges) be defined? *A: Input distributions quantify the uncertainty in the input factors and can be defined based on information available prior to performing GSA. When inputs are model parameters, uncertainty can be characterised based on their physical meaning, experimental data (when applicable) or previous studies. When inputs are forcing data, uncertainty can be quantified based on the characteristics of the data measurement and data processing techniques. Distributions and ranges can also be elicited from experts. When no specific information is available that suggests otherwise, then uniform distributions are typically chosen. A more detailed discussion of this issue can be found in (Pianosi et al., 2016). Sometimes, the assumed distribution can be tested and refined using site-specific data. For example, for model parameters, the comparison of model output simulations with observations (when available) or with the modeller's expectation can lead to removing sub-ranges that produce implausible model behaviour (Pianosi et al., 2017). If the definition of input distributions is not univocal, then we recommend testing the impact of different choices on GSA results as done for instance in Wang et al. (2013).*
Q: Should very wide input ranges be chosen to be sure that all possible input values are included? *A: No, input ranges should realistically characterize the uncertainty in the input factors. The existence of implausible values in the input ranges (and thus in the model response) could bias GSA results. If the input ranges are poorly known, then it might be necessary to refine them in a stepwise manner to avoid that unreasonable values control the study result (see for example Kelleher et al., 2013).*
Q: Which sampling strategy should be used to explore the input space? *A: Some GSA methods (e.g. Elementary Effect Test or Variance-Based Sensitivity Analysis presented in Sect. 2.2.2) may require a tailored sampling strategy. Other methods (e.g. Regional Sensitivity Analysis presented in Sect. 2.2.2) can be applied to input/output datasets generated by any sampling techniques. We note that recent studies have discussed the possibility of using a generic sample for any GSA method (Pianosi et al., 2017; Stanfill et al., 2015). When using random sampling, it may happen that several samples be clustered in one region of the sampling space while other regions might not be covered at all. More sophisticated strategies have been developed to provide a more uniform coverage of the input space, such as quasi-random sampling or Latin-Hypercube sampling. In particular, maximin Latin-Hypercube samples are built so to maximise the minimum inter-point distance between any two sampled points, or in other words the spread between the points. Sampling techniques are further discussed in Forrester et al. (2008) and Press et al. (1992).*

Table 2.1 Discussion of the choices regarding the implementation of the three main steps of GSA

Step 2. Model evaluation
Q: When models are evaluated using time series of input data, how long should the simulation time horizon be? *A: The choice of the simulation time horizon depends on the information content of the input time series. Since different input factors may have an influence on the output over different types of event (e.g. dry or wet period), the input time series should include a sufficient variety of events. The adequacy of the length of input time series can be assessed by analysing the variability of the GSA results over different data periods as proposed in Shin et al. (2013).*
Q: Are observations of the model output needed to perform a GSA? *A: Observations of model output are not required to perform GSA. The scalar model output used for GSA can be either a statistic of the simulated time series (e.g. Hartmann et al., 2015) or an objective function when output observations are available (e.g. van Werkhoven et al. 2008a). However, it is good practice to use observations of model output when available. In fact, they can provide information on the plausibility of the different input sets and can allow to filter out poor performing input sets as discussed in Step 3 below.*
Q: Which output should be analysed? *A: The choice of the output should be consistent with the purpose of the analysis. If needed, multiple outputs can be analysed, either by examining different outputs individually (e.g. Van Werkhoven et al., 2008a; Pianosi et al., 2017) or by directly integrating GSA with a multi-criteria analysis (e.g.* Rosolem *et al., 2012).*

Step 3. Post-Processing
Q: Which GSA method should be used? A: *Many different quantitative and qualitative GSA methods are available in the literature (Saltelli et al., 2008; Norton, 2015; Pianosi et al., 2016). The choice of the most appropriate method must consider three criteria:* 1. *the available computational resources, which will be further examined in Sect. 2.3-2.6;* 2. *the objective(s) of the analysis, as discussed in Sect. 2.2.2;* 3. *the fit to the method's underlying assumptions. In fact, some GSA methods can be applied to any case studies. Others assume, for instance, a particular shape of the output distribution, such as a normal distribution in the case of Variance-Based Sensitivity Analysis (see Sect. 2.2.2), or a linear model. The issue of credibility of the GSA results arises when the method's underlying assumptions are not satisfied (Pianosi et al., 2016).* *Qualitative methods (such as parallel coordinate plots discussed in Sect. 2.2.2 or scatter plots used e.g. in Prudhomme et al., 2013) are particularly useful to visualise the model response surface. They can be used, for instance, to perform a preliminary assessment of input factor sensitivity, often using only a limited number of model evaluations. They are particularly useful in the early phases of model development or of getting to know a new model. Modelers can use them, for example, for code debugging to check that the model's behaviour is consistent with their expectations, before applying more rigorous and computationally demanding GSA methods.*
Q: Should the sensitivity indices be computed over all sampled input sets or over selected sets only (e.g. sets for which the output value is considered to be physically plausible or for which the output value is close to observations)? *A: This choice depends on the objective of the GSA. It is important to bear in mind that filtering out some input sets can significantly impact the GSA results (e.g. Pianosi et al., 2017).*

Table 2.1 continued

2.2.3 Three well-established quantitative Global Sensitivity Analysis methods

In this section we present three widely used GSA methods that rely on different assumptions and rationales, namely Variance-Based Sensitivity Analysis (VBSA), Elementary Effect Test (EET, or method of Morris) and Regional Sensitivity Analysis (RSA). These three methods are implemented in the SAFE toolbox presented in Sect. 2.2.1. We denote the model output as y, the vector of input factors subject to GSA as X, the number of input factors as M, the total number of model evaluations required to perform GSA as N and the base sample size that must be specified by the SA user during the sampling step (step 1 in Sect. 2.2.1) by n. We further discuss the choice of the sample size n in Sect. 2.3-2.6.

Variance-Based Sensitivity Analysis (VBSA)

Variance-Based SA (VBSA) is based on the variance decomposition first proposed by Sobol' (1990). Two sensitivity indices are commonly used in GSA applications (Saltelli et al., 2008), the Main effect (VBM) index S_i^{VBM} and the Total effect (VBT) index S_i^{VBT}, which includes the main effect and interactions. The two sensitivity indices are expressed as follows (Saltelli et al., 2008):

$$S_i^{VBM} = \frac{V_{Xi}\left(E_{X\sim i}(y|X_i)\right)}{V(y)}, \tag{2.1}$$

$$S_i^{VBT} = \frac{E_{X\sim i}(V_{Xi}(y|X_{\sim i}))}{V(y)}, \tag{2.2}$$

where X_i is the i-th input factor, $X_{\sim i}$ denotes the vector of all input factors but the i-th one, E is the expected value and V is the variance. Both sensitivity indices can be used for ranking the input factors depending on whether the GSA user is interested in main effect only or in main effect and interactions. However, only the total effect index is suitable for screening because it accounts for input factor interactions as well as individual impact on its own.

The indices S_i^{VBM} and S_i^{VBT} can be estimated following a method proposed by Saltelli (2002) when the input factors are independent. First, two independent input samples X_A and X_B are built (each being a matrix of dimension (n, M)). Then, a matrix X_C of dimension $(n.M, M)$ is generated by recombination of the samples in X_A and X_B: X_C is composed of M blocks X_{Ci} $(i = 1, \dots, M)$, each block being a (n, M) matrix whose columns are all taken from X_B exception made for the i-th column, which is taken from X_A. We denote the three corresponding sets of model outputs as y_A, y_B and y_C. Then, S_i^{VBM} and S_i^{VBT} are computed as follows:

$$S_i^{VBM} = \frac{\frac{1}{n}\sum_{j=1}^{n} y_A^{(j)} y_{Ci}^{(j)} - \left(\frac{1}{n}\sum_{j=1}^{n} y_A^{(j)}\right)^2}{\frac{1}{n}\sum_{j=1}^{n}\left(y_A^{(j)}\right)^2 - \left(\frac{1}{n}\sum_{j=1}^{n} y_A^{(j)}\right)^2}, \tag{2.3}$$

$$S_i^{VBT} = \frac{\frac{1}{n}\sum_{j=1}^{n} y_B^{(j)} y_{Ci}^{(j)} - \left(\frac{1}{n}\sum_{j=1}^{n} y_A^{(j)}\right)^2}{\frac{1}{n}\sum_{j=1}^{n}\left(y_A^{(j)}\right)^2 - \left(\frac{1}{n}\sum_{j=1}^{n} y_A^{(j)}\right)^2}. \tag{2.4}$$

The total number of model evaluations for the approximation of both the main and total effects indices is therefore equal to:

$$N = n(M+2). \tag{2.5}$$

To maximize the coverage of the input factor space, for each given sample size, maximin (maximisation of the minimum inter-point Euclidean distance between any two sampled points) Latin Hypercube Sampling (LHS) can be used as sampling strategy to build the matrices X_A and X_B (we refer to Table 2.1 for a discussion on the choice of sampling strategy).

VBSA is deemed to provide reliable results for screening and ranking purposes and can be used to perform Variance Cutting (Saltelli et al., 2008). It is often considered as a benchmark to assess the credibility of other GSA methods (e.g. Yang, 2011). The sensitivity indices are expressed in terms of fraction of the output variance and always take values between 0 and 1.

Elementary Effect Test (EET) or method of Morris

The Elementary Effect Test (EET; Saltelli et al., 2008) or Method of Morris (Morris, 1991) is a understood as being a less computationally expensive method than VBSA, and therefore as being most suitable when dealing with time-consuming models (Saltelli et al., 2008). Campolongo et al. (2007) demonstrated empirically that the sensitivity measure produced by EET could be used as a proxy of the total effect index produced by VBSA, and therefore EET is particularly suitable for screening. We will further investigate the two above-mentioned interesting properties of the EET method (computational requirements and similarity to VBSA) in Sect. 2.3-2.6.

EET is a global extension of One-factor-At-the-Time (OAT) Local SA methods. It is based on the computation of several Elementary Effects (EEs). Specifically, the EE of the i-th input factor x_i at given baseline point X^j and for a predefined perturbation Δ is given by:

$$EE_i^j = \frac{y\left(x_1^j, x_2^j, \dots, x_{i-1}^j, x_i^j + \Delta, \dots x_M^j\right) - y\left(x_1^j, x_2^j, \dots, x_{i-1}^j, x_i^j, \dots x_M^j\right)}{\Delta}. \tag{2.6}$$

For each input factor, EEs are computed at n randomly chosen baseline points across the input factor space. Specifically, each baseline point is associated with an auxiliary point and the perturbation Δ of Eq. (2.6) is computed as the difference between the i-th coordinate of the auxiliary and baseline point. The estimated mean (μ_i) of the EEs is taken as a measure of the total effects of the i-th input factor. The standard deviation (σ_i) of the EEs can be interpreted as a combined measure the intensity of the interactions of the i-th input factor with other input factors and the degree of non-linearity in the model response to variations in the ith input factor. To avoid compensations between EEs of opposite sign, the mean of the absolute values of the EEs (μ_i^*) can be used, as first suggested by Campolongo et al. (2007), i.e.:

$$\mu_i^* = \frac{1}{n}\sum_{j=1}^{n}\left|EE_i^j\right| . \tag{2.7}$$

The sensitivity index of Eq. (2.7) provides a semi-quantitative measure of sensitivity, particularly suitable to rank the factors on an interval scale (Saltelli et al., 2008). To define baseline and auxiliary points, the radial design strategy proposed by Campolongo et al. (2011) was shown to be computationally efficient. In this approach, n baseline points can be sampled across the input factor space, and associated with other n auxiliary points, also randomly chosen. As for VBSA, maximin Latin Hypercube sampling strategy can be used so to maximize the coverage of the input factor space. The total number of model evaluations required to compute the mean EEs for all input factors is equal to:

$$N = n\,(M+1) . \tag{2.8}$$

We note that the value of μ_i^* has no specific meaning per se, as it depends on the scale and units of measurements of the model output y. To allow to allow comparison between different case studies, in the following a normalized mean of the EEs will be used as sensitivity index. This is defined as the ratio between μ_i^* and the maximum value of the mean EEs across all the input factors:

$$S_i^{EET} = \frac{\mu_i^*}{\max_k \mu_k^*} . \tag{2.9}$$

The sensitivity index of Eq. (2.9) now takes values between 0 and 1 regardless of the units of measurement of y, and it expresses input factor sensitivity as a fraction of the sensitivity for the most influential input factor. The index still provides a semi-quantitative measure of sensitivity.

Regional Sensitivity Analysis (RSA)

While we previously discussed GSA methods linked to the screening, ranking and Variance Cutting settings only (VBSA and EET), here we present a technique for Factor Mapping.

Regional Sensitivity Analysis (RSA; Young et al., 1978; Spear and Hornberger, 1980) is a method for Factor Mapping that is widely used because of its ease of implementation (e.g. Freer et al., 1996; Sieber and Uhlenbrook, 2005; Wagener et al., 2001). It is however of limited use for screening, since it does not detect interactions between input factors (for instance, factors combined as products or quotients may compensate, see p.190 in Saltelli et al., 2008), and therefore a zero-value sensitivity index produced by RSA is a necessary, but not sufficient condition for an input factor to be non-influential. The method first decomposes the set of input factor samples into two groups, depending on whether their associated output exceeds a prescribed threshold value (e.g. a certain level of performance). The two marginal Cumulative Distribution Functions (CDF) $F_i^B(x_i)$ and $F_i^{\bar{B}}(x_i)$ for the two groups (B and $\bar{B}$), for instance behavioural (acceptable) and non-behavioural (poor) model predictions, are then derived and compared. The discrepancy between the behavioural and the non-behavioural CDFs can be quantified by means of the Kolmogorov-Smirnov statistic (Kolmogorov, 1933; Smirnov, 1939), which can be used for ranking the input factors (Pianosi et al., 2016). The sensitivity index for the i-th input factor x_i is then expressed as follows:

$$S_i^{RSA} = \max_{x_i} \left| F_i^B(x_i) - F_i^{\bar{B}}(x_i) \right| . \tag{2.10}$$

The sensitivity index of Eq. (2.10) varies between 0 and 1 and is semi-quantitative.

The choice of the output threshold value is critical when applying RSA. The threshold is commonly chosen so to separate 'behavioural' input sets (that produce simulations consistent with observations when available or with the modeller's expectation) from 'non-behavioural'. For instance, the threshold separates minor from excessive growth of an alga species in Spear and Hornberger (1980). When no clear threshold value can be easily identified, it can be constructed by trial and error. For instance, Hartmann et al. (2015) used the median of the simulated groundwater recharge. An alternative approach is to perform RSA based on grouping as proposed in Wagener et al. (2001) which does not require the selection of a single threshold on output values. This variant consists of splitting the input factor sample into a given number of groups (e.g. ten) according to the associated output value (e.g. ten intervals of increasing output value, designed so to have an equal number of samples in each group) and to derive the corresponding ten CDFs for the different input factors. Influential input factors are identified when differences are observed between the ten CDFs. When instead input factors are not influential, all distributions are very close to the distribution from which the inputs were sampled.

Other methods for Factor Mapping

Classification and Regression Trees (CART) is a machine learning technique that has been proposed to build non-linear prediction models from data (Breiman et al., 1984). CART is constructed following a binary recursive algorithm that partitions the space of a set of independent variables (e.g. climate and land cover properties) based on the value of a dependent variable (e.g. groundwater recharge). In the case of Classification Trees, each subset of the partition corresponds to a specific output class. CART can take the form of a binary tree composed of nodes and branches. Each internal node of the tree corresponds to a logical expression based on the value of a specific independent variable (e.g. "mean annual precipitation above or below 300 mm") and splits into two branches that correspond to the two possible outcomes (i.e. logical expression is "true" or "false"). Predicting the value of the dependent variable (e.g. groundwater recharge) using the tree consists of moving down the branches from the root node to the leaf nodes that indicate the value of the dependent variables.

In the context of GSA, the independent variables that appear in the internal node of the tree can be interpreted as being influential with respect to the dependent variables, and the higher they appear in the tree the higher their sensitivity. Moreover, CART can identify critical thresholds on the dependent variables that produce unwanted outcomes (e.g. low groundwater recharge values). Classification Tree was applied in Singh et al. (2014) to identify climate and land cover properties that produce lower/higher change in runoff and in Almeida et al. (2017) to reveal the precipitation conditions and slope properties that can trigger a landslide. CART is implemented in the Matlab Statistics and Machine Learning Toolbox (Mathworks, 2016).

Additionally, parallel coordinate plot (e.g. Inselberg, 2009; Kasprzyk et al., 2013) is a useful tool that allows visualisation of Factor Mapping results. A parallel coordinate plot is a two-dimensional plot that represents a multidimensional space (e.g. the space of the model input-output). The plot can also provide insight into factor interactions. Parallel coordinate plots are implemented for instance in the SAFE toolbox and in the Matlab Statistics and Machine Learning Toolbox (function "parallelcoords").

2.3 Convergence and validation of Global Sensitivity Analysis results

GSA explores more exhaustively the input space compared to LSA (see Sect. 2.1) and therefore it typically requires larger computational resources than LSA. A critical step of sampling-based GSA is the choice of the sample size to run the MC experiment. If the sample is too small to adequately cover the input space, the analysis may not provide robust results. On the other hand, for very large sample sizes the computational cost may become very high while not improving the precision of the results significantly. In environmental applications, where models are often complex and simulations

expensive, an acceptable trade-off has to be found between the need to obtain robust results and the need to limit computational cost.

The total number of model evaluations (N) used in GSA typically increases with the number of model input factors (M). For some GSA methods such as VBSA and EET (Sect. 2.2.2), depending on the methodology used to derive the estimates of the sensitivity indices, N is expressed as a function of M and of a base sample size (n) that must be specified by the user (i.e. $N = f(n, M)$). Thus, choosing the value of the total number of model evaluations (N) comes down to choosing the value of the base sample size (n). For other methods such as RSA (Sect. 2.2.2), no explicit expression relates N to M and therefore N is directly chosen by the GSA user ($N = n$). Suggestions for the choice of n can be found in the literature for several GSA methods. For instance, Saltelli et al. (2008, Table 6.9) provide typical values of n for EET, RSA and VBSA. However, a relatively limited number of studies actually focus on a rigorous assessment of the convergence of GSA results. Figure 2.1 reports several examples taken from the literature regarding the relationship between N and M for EET, RSA and VBSA. From these studies, we make three observations:

1. Previous convergence studies assessed different types of convergence, namely convergence of the sensitivity indices, of the screening results, and of the ranking. This lack of uniformity in the definition of 'convergence' makes it difficult to consistently compare the results obtained for models of different complexities when using different GSA methods. However, a preliminary conclusion that seems to emerge from these studies is that different sample sizes are required for different types of convergence. For instance, in the case of EET, (Vanuytrecht et al., 2014) highlighted that while a low sample size (n =25) can be suitable for screening, it can be insufficient for factor ranking. Nossent et al. (2011) showed that a base sample size of 12,000 is needed to ensure the convergence of Variance-Based sensitivity indices in their specific case study, however, a much smaller sample size (n <2000) was found to be sufficient if one is only interested in ranking the most important input factors.
2. Within a given type of convergence, different values of the base sample size are found for the same method when applied to different models. For instance to ensure convergence of the value of Variance-Based sensitivity indices (Fig. 2.1 bottom left panel), Tang et al. (2007) used a base sample size n of 8192 (for a case study with 18 input factors), while Yang (2011) used n equal to 3000 (for a case study with 5 input factors). This suggests that the base sample size may also be a function of the number of input factors or of other characteristics of the model or of the case study. It is also worth noting that these studies show that convergence is often reached using a base sample size significantly larger than the values suggested in Saltelli et al. (2008).

3. Convergence is generally assessed based on a visual analysis of the stability of the results for increasing sample size. Some authors used the confidence intervals of the sensitivity indices for a more quantitative assessment of their convergence (Campolongo and Saltelli, 1997; Nossent et al., 2011). However, they did not explicitly define a convergence criterion. Herman et al. (2013) and Vanrolleghem et al. (2015) both introduced a quantitative criterion to measure the convergence of the sensitivity indices values (that will be discussed in Sect. 2.4.1), but they do not consider the convergence of ranking or screening.

Another issue for GSA is the choice of the screening threshold i.e. a threshold value for the sensitivity indices below which factors are classified as no-influential (more details in Sect. 2.4.2). In this respect, the following can be learned from existing studies:

1. For VBSA, the input factors that have a sensitivity index below 0.01 are often considered non-influential (Tang et al., 2007; Sin et al., 2011; Cosenza et al., 2013; Vanrolleghem et al., 2015). The adequacy of this screening threshold was tested in Tang et al. (2007), however the validation strategy used in that work (based on a visual approach introduced by Andres, 1997) has some limitations that we discuss and overcome here (more details in Sect. 2.4.2). Nossent et al. (2011) considered a screening threshold value of zero. They identified as statistically significant any input factor for which the lower bound of the confidence interval on the sensitivity index is positive. This method is quite conservative since, in our experience, a sensitivity index could have positive confidence bounds, and therefore a non-zero value, even if the input factor has negligible effect on the output.
2. EET, which is widely used for screening purpose, provides a relative measure of sensitivity that has a different meaning and range of variation depending on the model output definition in the particular case under study (Eq. (2.7)). Therefore, case-specific threshold values are usually taken (e.g. Vanuytrecht et al., 2014) and little guidance exists in the literature on this topic. Cosenza et al. (2013) and Vanrolleghem et al. (2015) presented an attempt at defining an absolute value for the screening threshold for EET. However, they did not validate the adequacy of their proposed threshold values.

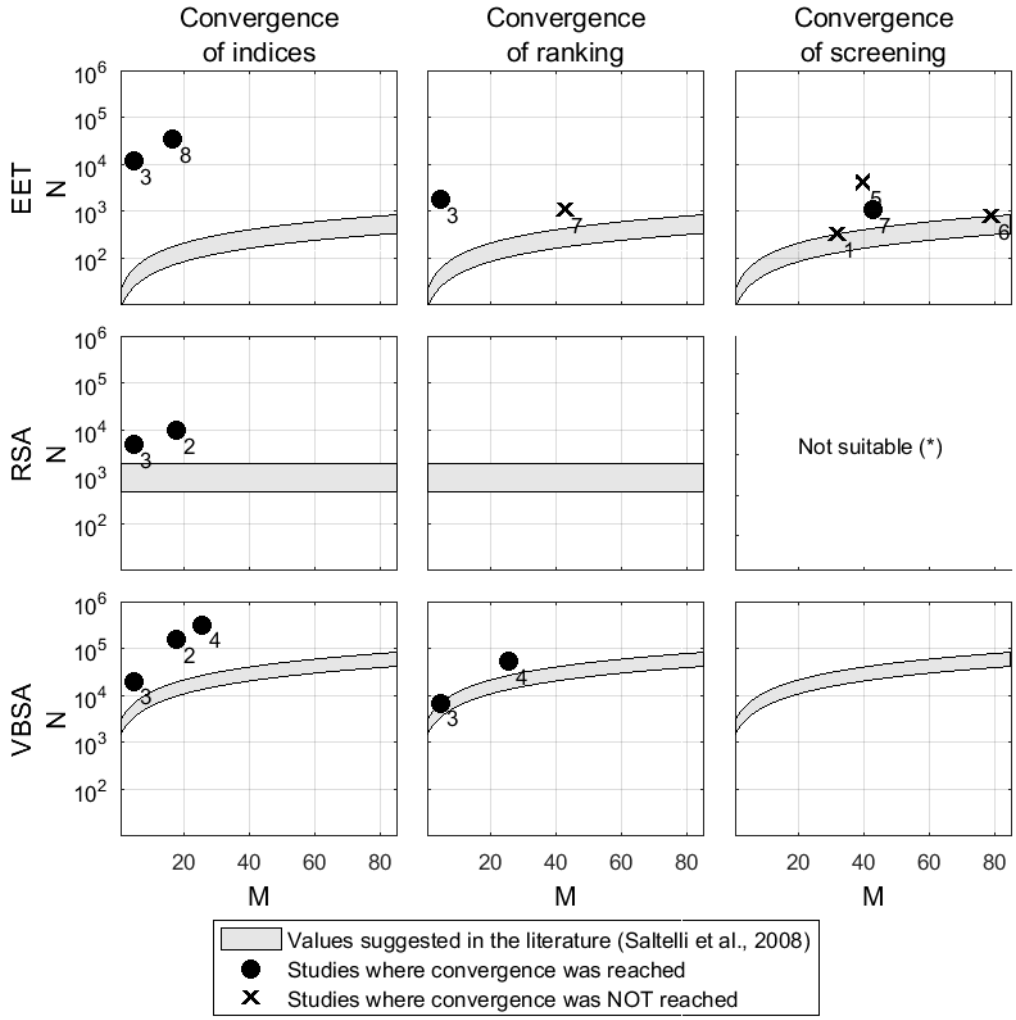


Figure 2.1 Number of model evaluations used in GSA against number of input factors from several examples reported in the literature.

N number of model evaluations, *M* number of input factors. Circles indicate that convergence is reached (respectively convergence of sensitivity indices, ranking and screening) and crosses indicate that convergence is not reached. The filled grey areas correspond to values of the sample size suggested in the literature for the different GSA methods (Saltelli et al., 2008, Table 6.9). Many GSA studies adopted these suggested values, without checking the convergence of the GSA results.

N is computed as follows: $N = n(M + 1)$ for applications of Elementary Effect Test (EET), $N = n$ for applications of Regional Sensitivity Analysis (RSA), $N = n(M + 2)$ for applications of Variance-Based method (VBSA), where n is a base sample size chosen by the GSA user.

The results are taken from [1]Campolongo and Saltelli. (1997), [2]Tang et al. (2007), [3]Yang (2011), [4]Nossent et al. (2011), [5]Nossent and Bauwens (2012), [6]Cosenza et al. (2013), [7]Vanuytrecht et al. (2014), [8]Vanrolleghem et al. (2015).

(*) RSA is not used for screening since it neglects parameter interactions.

Based on this literature review, we believe that there is a lack of guidance to support GSA users in the choice of an adequate sample size and in the definition of a screening threshold, while there is an opportunity for improving current approaches to the validation of GSA results. Thus, the remaining of the chapter has the following objectives:

1. To define quantitative criteria to assess different types of convergence of GSA results, i.e. convergence of sensitivity indices, ranking and screening.
2. Based on these quantitative convergence measures, to investigate the convergence of three widely used GSA methods and to assess whether it is possible to give general guidelines for an adequate choice of the base sample size.
3. To develop a methodology to quantitatively validate screening results and therefore to formally investigate the adequacy of different choices for the screening threshold.

Here, we consider three widely used GSA methods presented in Sect. 2.2.2 (VBSA, EET and RSA) implemented in the SAFE toolbox. We apply GSA to three hydrological models of increasing complexity (HyMod, HBV and SWAT). The input factors are the model parameters and the output is the model accuracy. However, our approach could equally be applied to other GSA methods or models, and with different experimental set-ups, i.e. different definition of the model output and of the input factors subject to GSA (e.g. boundary conditions, errors in input forcing data, model resolution, etc.). In Section 2.4 we define the convergence criteria and the validation procedure for the screening results, we describe the workflow adopted for the experiments and we briefly present the three case studies analysed. We then report the results obtained for convergence and for screening validation in Sect. 2.5. We discuss meaning, implications and limitations of these results in Sect. 2.6. Sect. 2.7 summarises the results of this chapter and explains how GSA will be used in the subsequent chapters.

2.4 A new methodology for assessing convergence and validating GSA results

2.4.1 Three novel convergence criteria

In this section we provide three definitions of ‘convergence' of GSA results and we propose criteria to quantitatively assess the different types of convergence. By ‘convergence’ we mean here the fact that GSA results do not change (or change to a limited degree) when using a different sample of model evaluations (of equal or larger size). We suggest distinguishing three different types of convergence (Fig. 2.2):

1. Convergence of the sensitivity indices, which is reached if the values of the indices remain stable;
2. Convergence of ranking, which is achieved if the ordering between the parameters remains stable;
3. Convergence of screening, which is reached if the partitioning between sensitive and insensitive parameters remains stable.

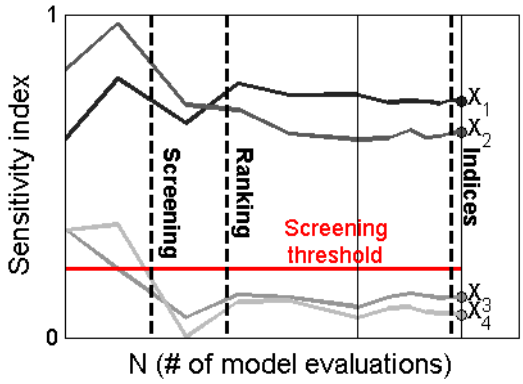


Figure 2.2 Definition of convergence for the three GSA objectives.
The figure reports the value of the sensitivity index against the number of model evaluations (N) in a hypothetical example with four parameters. Vertical lines indicate the convergence of the screening, ranking and indices. The screening converges when the partitioning between influential parameters and non-influential parameters (indices below the screening threshold) stabilizes. The ranking converges when the ordering among the parameters stabilizes. The sensitivity indices converge when their value stabilizes.

We propose three indicators that can be used to assess the three types of convergence defined above. All three indicators satisfy the following properties:

a. They are quantitative indicators, i.e. they are computed through a numerical, reproducible procedure;
b. They are efficient, i.e. the numerical procedure for their computation does not require additional model evaluations;
c. They are easy to interpret and they allow for comparison across case studies and GSA methods.

The convergence indicators are described in the following paragraphs. They all measure the degree of uncertainty in GSA results, which is estimated via the bootstrap technique (Efron and Tibshirani, 1994; Archer et al., 1997). In bootstrapping, many different resamples are constructed by drawing randomly with replacement from the original sample of the model input/output so that no additional model runs

are required and property (b) above is respected. The drawback is that resamples are obviously not independent from each other. A discussion regarding the quality of the bootstrap can be found in other studies, e.g. for estimating the mean of a distribution (Yang, 2011) and for its quantiles (Romano and Shaikh, 2012). Under certain conditions, the reliability of the bootstrap technique may be questioned, for instance when the sample size is small (Isaksson et al., 2008). We elaborate on this issue in Sect. 2.6.3.

Convergence of the sensitivity indices value

To assess the convergence of the sensitivity indices, we compute the width of the 95% confidence intervals (5% significance level) of the index distribution obtained by bootstrapping. We use the maximum width of the confidence intervals across all the model input factors as a summary statistic:

$$Stat_{indices} = \max_{i=1\ldots M}(S_i^{ub} - S_i^{lb}), \tag{2.11}$$

where S_i^{ub} and S_i^{lb} are the upper and lower bounds of the sensitivity index of the i-th input factor while M is the number of input factors. A value of the width of the confidence interval close to zero indicates that the sensitivity index has converged. Since in our study we use normalised sensitivity indices that vary between 0 and 1 (as reported in Sect. 2.4.4), we could define an absolute threshold value for $Stat_{indices}$, below which we considered that convergence is reached. We selected a threshold values of 0.05, which allows a 5% error on the value of the sensitivity indices. Other threshold values could be considered, for instance a percentage of the sensitivity index value of the most influential input factor as in Herman et al. (2013).

Our convergence criterion is quite different from the one by Vanrolleghem et al. (2015). In that study, the authors measure the variability of the sum of the sensitivity indices between two random samples of different size. Convergence is reached when this variability is low. We believe that their criterion is a necessary, but not sufficient condition for convergence for two reasons. First, it does not ensure that the sensitivity indices for all input factors have converged individually. Second, it assesses the variability between two random samples only and therefore it could happen that this variability is low even if convergence is not actually reached (the two samples can give similar results 'by chance'). Our criterion instead is based on the statistic of Eq. (2.11), which measures the variability of the sensitivity estimates of all individual input factors and across multiple resamples.

Convergence of input factor ranking

To assess the convergence of the ranking process, we use a rank correlation coefficient that quantifies the agreement between pairs of rankings generated using different bootstrap resamples. We initially considered and compared different rank correlation coefficients, starting with Spearman's rank correlation coefficient (Spearman, 1904). Its main limitation, when used in our context, is that it gives the same importance to rank differences occurring in the higher positions of the ranking (highly sensitive input factors) as in the lowest (insensitive input factors). However, rank reversals between insensitive input factors are typically of scarce interest since the main aim of ranking is to separate out and sort the most sensitive input factors. This limitation of the Spearman rank correlation coefficient is described by Iman and Conover (1987) and shown in Appendix A.3 using a theoretical example. Therefore, we decided that this indicator is unsuitable for our purposes.

Other studies attempted to deal with the limitation of the Spearman rank correlation coefficient. Iman and Conover (1987) proposed to compute a correlation coefficient based on Savage scores (Savage, 1956) instead of ranks. Dancelli et al. (2013) introduced weights in the Spearman rank correlation coefficient formula (these weights are decreasing functions of the ranks). These two studies partially overcame the limitation of the un-weighted Spearman coefficient by giving more weight to rank reversals occurring at the top of the ranking (most sensitive input factors). However, rank reversals of low-sensitivity input factors can still have a significant impact when a model has a large number of low-sensitivity input factors and therefore some of them are in relatively high ranked positions despite having a small sensitivity index (see again Appendix A.3 for further details). We tackle this situation by proposing the following adjusted and weighted rank correlation coefficient:

$$\rho_{s,j,k} = \sum_{i=1}^{M} \left| R_i^j - R_i^k \right| \frac{\max_{j,k}\left(S_i^j, S_i^k\right)^2}{\sum_{i=1}^{M} \max_{j,k}\left(S_i^j, S_i^k\right)^2}, \tag{2.12}$$

where S_i^j and S_i^k are the values of the sensitivity index and R_i^j and R_i^k are the ranks of the i-th input factor, estimated using the j-th and the k-th bootstrap resample respectively. This indicator emphasises the disagreements in the ranking for the most sensitive input factors, while neglecting the disagreements for the least sensitive input factors by directly using the sensitivity values to weight rank reversals. As a weight function we choose the squared maximum sensitivity index value between the two samples. This indicator has a clearly interpretable meaning: it represents the (weighted) average distance in the input factor ranks, obtained over two different bootstrap samples. The choice of the convergence statistic for ranking is further analysed in Appendix A.3. To aggregate the rank coefficients obtained over all possible pairs of bootstrap resamples, we use the 95% quantile value (5% significance level):

$$Stat_{ranking} = Q_{\substack{0.95 \\ j,k}}(\rho_{s,j,k}), \tag{2.13}$$

We consider that the convergence of the ranking is reached when the statistic of Eq. (2.13) falls below 1. This choice is motivated by the fact that a value of the weighted and adjusted rank correlation coefficient (Eq. (2.12)) equal to 1 means that, on average, the differences in the ranking for the most sensitive input factors are less than one position.

Convergence of input factor screening

Screening the model input factors (in our application the model parameters) consists in separating them in two groups, the influential and the non-influential (insensitive) ones.

Theoretically, input factors are completely insensitive when the corresponding sensitivity index is equal to zero. In practice, when sensitivity indices are approximated via MC simulation, they are likely to take a very small positive value even when their (unknown) exact value is zero. Moreover, the objective of screening is often to identify not only the parameters that are completely insensitive, but also the ones that have a small but negligible effect. For these reasons, it is common practice to assume a threshold value T for the sensitivity index below which the input factors are considered as insensitive (e.g. Tang et al., 2007; Sin et al., 2011; Cosenza et al., 2013; Vanrolleghem et al., 2015). For a given value of the screening threshold T, the corresponding subset of insensitive input factors X_0 is defined as follows:

$$X_0 = \{x_i \ when \ S_i < T\}, \tag{2.14}$$

where x_i is the i-th input factor and S_i is the sensitivity index (bootstrap mean) for the i-th input factor. In principle, the screening convergence might be assessed by measuring the stability in the partitioning as defined by Eq. (2.14). However, results would be highly dependent on the choice of the screening threshold T, whose exact value is not known a priori. Here, we therefore use a proxy measure for the screening convergence. We set the threshold in Eq. (2.14) to a relatively high value (T = 0.05) so that X_0 takes the meaning of set of "lower-sensitivity" input factors rather than set of "insensitive" input factors. Then, we use as a summary statistic the maximum width of the 95% confidence intervals across the lower-sensitivity input factors in X_0:

$$Stat_{screening} = \max_{x_i \in X_0}(S_i^{ub} - S_i^{lb}), \tag{2.15}$$

Similar to the convergence of the sensitivity indices (Eq. (2.11)), we consider that screening convergence is reached when $Stat_{screening}$ is below a value of 0.05. In other words, we assume that screening convergence has been reached when the sensitivity indices for the lower-sensitivity input factors have converged. We can then investigate whether all input factors in X_0 are actually insensitive

using the validation test presented in the next section. This test is also a tool to determine a posteriori the value of the screening threshold T that would identify insensitive input factors in the case under study.

2.4.2 A Validation procedure for screening results

In this section we discuss and review two methods that can be used to validate the screening results obtained by GSA. Both methods aim to detect possible effects of the input factors classified as insensitive to avoid classifying influential input factors as insensitive. We denote the model output as y and the vector of input factors subject to GSA as X.

Andres (1997) proposed a method to evaluate the accuracy of the set X_0 defined in Eq. (2.14) and obtained by a generic GSA approach. In Andres' test, three sets of samples are generated. Set 1 is obtained by sampling the entire input factor space. In set 2 only the non-influential input factors (X_0) are allowed to vary while the influential input factors (denoted hereon by $\overline{X_0}$) are fixed to a prescribed value (e.g. the default parameter value from literature, or the mean of the assumed distribution). Finally, in set 3, the influential input factors are sampled within their feasible range, while the non-influential input factors are kept fixed. The actual value we fix the input factors at should not matter if the input factors are indeed insensitive. Three sets of model output samples are then obtained through MC simulations: the set of unconditional outputs $\{y\}$ (obtained from input factor set 1) and the two sets of conditional outputs $\{y|\overline{X_0}\}$ (from set 2) and $\{y|X_0\}$ (from set 3). The original test consists of a visual analysis of the two scatter plots that are obtained by plotting the unconditional output samples $\{y\}$ against the two conditional sets $\{y|\overline{X_0}\}$ and $\{y|X_0\}$. The input factors in X_0 are confirmed to be non-influential when in the first plot the conditional samples $y|\overline{X_0}$ align along a horizontal line (i.e. the output does not vary if varying the input factors in X_0 only) and when in the second plot the conditional samples $\{y|X_0\}$ align along a 45° line (i.e. the output variability when varying all input factors but those in X_0 is the same as when varying them all). Tang et al. (2007) and Nossent et al. (2011) used this approach to validate their screening results. These authors also proposed to quantify the satisfaction of this screening test by computing the correlation coefficient of the scatter plots. However, the correlation coefficient also takes a value close to one even when the points align along a straight line that does not coincide with a 45° line. Therefore, a high value of the correlation coefficient does not necessarily indicate that the input factors in X_0 are insensitive.

In this study, we use a variation of the original Andres' test, first introduced by Pianosi and Wagener (2015). It is based on the computation of the Kolmogorov-Smirnov (KS) statistic (Kolmogorov, 1933; Smirnov, 1939; see Wall, 1996 for a general introduction) to estimate the discrepancy between the sets

of unconditional and conditional outputs. Specifically, in this test, the first step is to compute the empirical unconditional Cumulative Distribution Function (CDF) $F_y(y)$ and the empirical conditional CDF $F_{y|X_0}(y)$ of the model output. Then, a two-sample KS-test can be applied to test the null hypothesis that the two CDFs ($F_y(y)$ and $F_{y|X_0}(y)$) are drawn from the same distribution. To this end, the KS-statistic is computed:

$$\widehat{KS}(X_0) = \max_y \left| F_y(y) - F_{y|X_0}(y) \right|, \tag{2.16}$$

The null hypothesis is rejected if the KS-statistic between the two CDFs is above a critical value KS_{crit}. For a given significance level of the test α_c, $KS_{crit}(\alpha_c)$ is computed as:

$$KS_{crit}(\alpha_c) = c(\alpha_c)\sqrt{\frac{N_u + N_c}{N_u N_c}}, \tag{2.17}$$

where N_u is the size of the unconditional sample used to build the empirical CDF $F_y(y)$, N_c is the size of the conditional sample used to build the empirical CDF $F_{y|X_0}(y)$ and the critical value $c(\alpha_c)$ can be found in the literature (e.g. Wall, 1996).

The KS statistic of Eq. (2.16) depends on the conditioning (prescribed) value attributed to the input factors in X_0 and used to estimate the conditional CDF. Therefore, the test should be repeated at different conditioning values, thus obtaining a set of KS values, which are then aggregated using a summary statistic. Here, we derive a number n_c of conditional samples, each of them is of size N_c and each of them uses a different conditioning value for the input factors in X_0. We aggregate the n_c corresponding values of the KS statistic (calculated using Eq. (2.16)) by taking the maximum $\widehat{KS}_{max}$ over the n_c conditioning values $X_{0,1}, \dots, X_{0,n_c}$:

$$\widehat{KS}_{max} = \max_{X_{0,1} \dots X_{0,nc}} \left(\widehat{KS}(X_{0,i}) \right), \tag{2.18}$$

The validation test we apply in our study compares the statistic $\widehat{KS}_{max}$ of Eq. (2.18) with the critical value $KS_{crit}(\alpha_c)$ of Eq. (2.17). It is worth noting that unlike the original two-sample KS-test, which consists in the comparison of two CDFs, our validation test consists in the comparison of n_c CDFs (conditional CDFs) to a reference CDF (unconditional CDF). Therefore, given the value of the significance level α_c used to compute the critical value of the KS-statistic of Eq. (2.17), the significance level of our validation test, denoted as α_t, is higher than α_c. In particular, if the n_c KS-statistics are considered to be independent, it can be shown that $\alpha_t = 1 - (1 - \alpha_c)^{n_c}$.

The choice of the sample sizes N_u, N_c and n_c must ensure a sufficient coverage of the input factor space and the convergence and robustness of the results of the KS-test. To set the value of the sample size for conditional (N_c) and unconditional (N_u) outputs, we assessed the convergence of the results of the KS-test. The results and methodology are presented in detail in Appendix A.4. From our analyses, the KS-test appears to be very sensitive in that it can detect small deviations between two CDFs. We choose a significance level α_c equal to 0.001 (minimum value of the significance level given in the tables see for instance Wall, 1996). In this way, input factors with very small but non-zero sensitivity are more likely to be detected as insensitive by the KS-test. For the purpose of screening, we believe it is appropriate to identify not only the input factors that are completely insensitive, but also the input factors that have a very small influence on the output, otherwise the screening would be too strict.

We note that a different summary statistic could be chosen instead of the maximum, for instance the mean or the median. The same level of confidence of the validation test α_t can be obtained using any of these summary statistics, provided that an appropriate value of the significance level α_c is chosen to compute the critical KS-statistic. Given the value of α_t, the value of α_c is lower for the maximum than for the median or the mean.

2.4.3 Procedure for testing the proposed techniques

Figure 2.3 presents the workflow of the analysis that we conduct to investigate and compare the convergence of several GSA methods. First, for each GSA method, we build a dataset of N input/output samples by MC random sampling and model evaluation. We estimate the sensitivity indices and their bootstrapping confidence intervals by resampling with replacement (Fig. 2.3.a). We can then compute the three statistics of Eq. (2.11), (2.13) and (2.15) and verify whether convergence of the indices, ranking and screening has been reached according to the criteria introduced in Sect. 2.4.1. Computations can be repeated using sub-samples of reduced size. A visual summary of the values of sensitivity indices and their uncertainty at different sample size is given by the convergence plot as Fig. 2.3.b. At the sample size when screening convergence is reached, we also apply the validation procedure defined in Sect. 4.2.2. For a given value of the assumed screening threshold, we obtain the set of insensitive input factors, compute the KS-statistic (Fig. 2.3.c) and apply the KS-test. We repeat the test for increasing values of the assumed screening threshold and obtain Fig. 2.3.d.

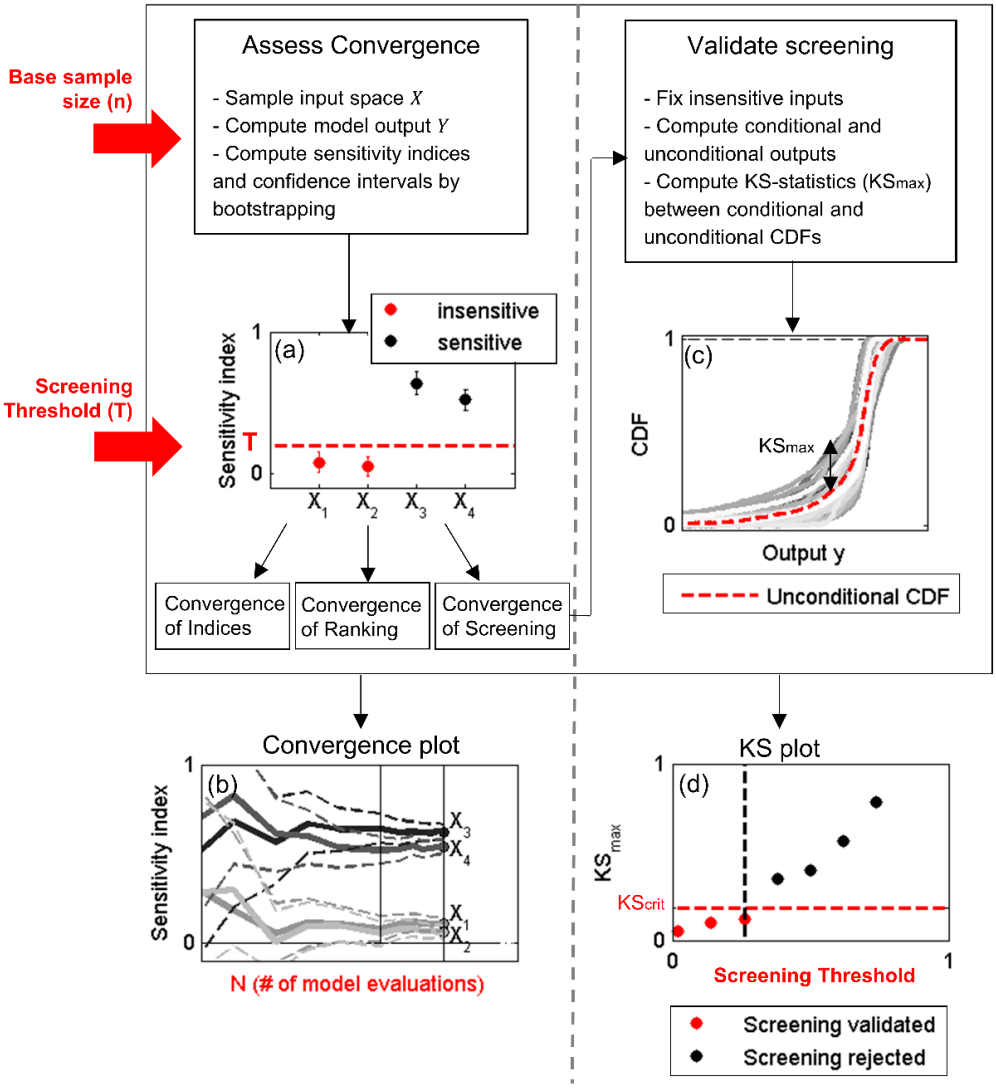


Figure 2.3 Workflow for the convergence analysis of GSA and for the screening validation. We investigate the influence of the choice of base sample size n and screening threshold T. We keep the other characteristics of the experimental setup unchanged. We repeat the experiments for different GSA methods. We note that the number of model evaluations N is a function of the base sample size n and the number of input factors M.

2.4.4 Test case studies

Three hydrological conceptual-type models of varying complexity are investigated in this chapter (HyMod, HBV and SWAT) and are applied to three different catchments. In our application, the input factors are the model parameters. Figure 2.4 presents the available datasets for the three case studies and Appendix A.1 provides a brief description of the model parameters.

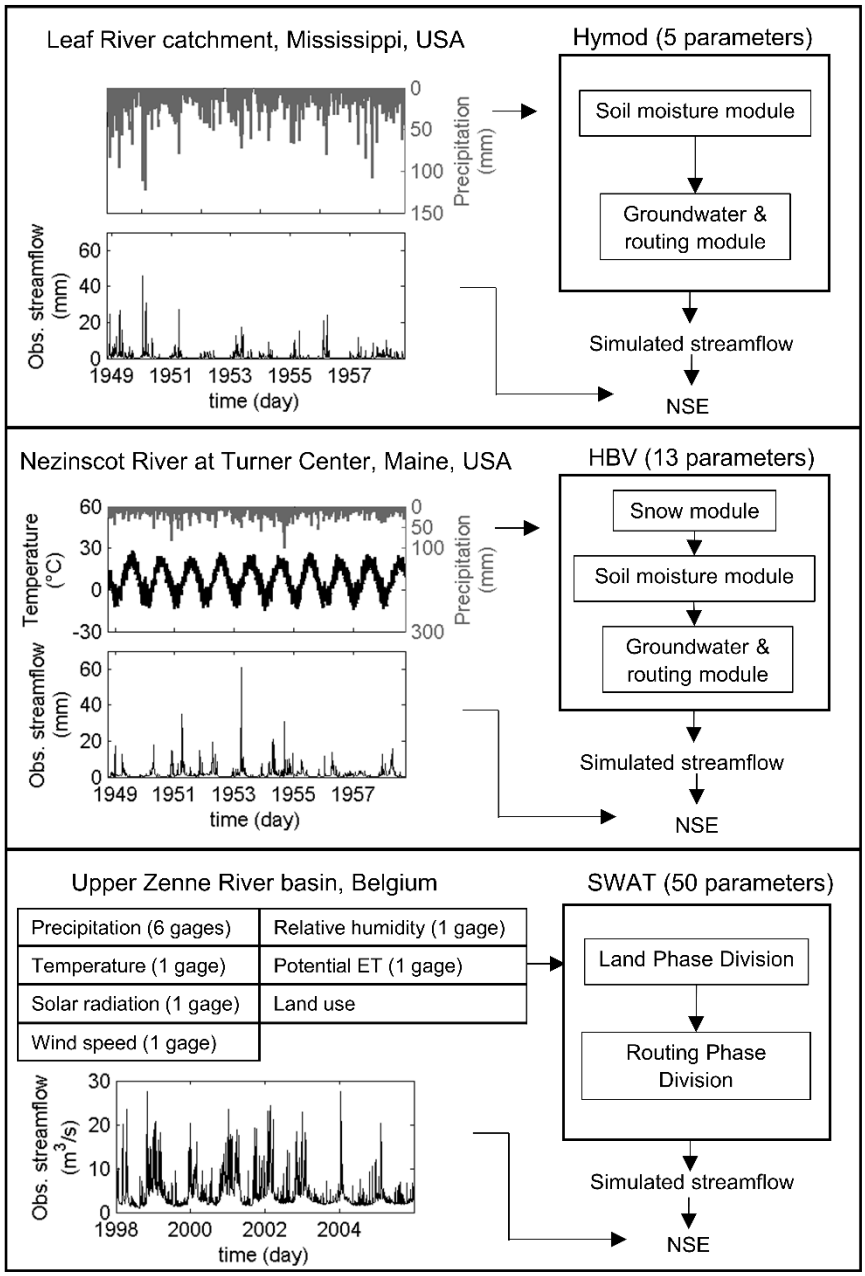


Figure 2.4 Data, routine and model output for the three case studies analysed.

HyMod model (5 parameters)

The HYdrological MODel (HyMod) model was first introduced by Boyle (2001) and is described in Wagener et al. (2001). It has been widely applied because of its simplicity (5 parameters), for instance in Vrugt et al. (2002), Kollat et al. (2012) and Gharari et al. (2013). The HyMod model produces a time series of stream flow predictions and is forced by precipitation and potential evapotranspiration. It is composed of a soil moisture routine (parameters BETA and SM) and a routing module (parameters

ALPHA, RS and RF). The latter module consists of two sets of parallel linear reservoirs, namely three linear reservoirs for the fast runoff component and a single linear reservoir for the slow runoff component. In this chapter, the model is evaluated with daily time step data over a simulation horizon of ten years, starting on 01/11/1948, including a one-year warm-up period. The application study site is the Leaf River catchment, a 1950 km2 catchment located north of Collins, Mississippi, USA. Sorooshian et al. (1983) provide a detailed description of the Leaf River catchment.

HBV model (13 parameters)

The Hydrologiska Byråns Vattenbalansavdelning (HBV) model was introduced by Bergström (1995) and is described in Seibert (1997). Although developed initially for applications in Scandinavia, the HBV model was used in many studies around the world (e.g. Kollat et al., 2012; Grillakis et al., 2010). The model produces a time series of stream flow predictions and is driven by precipitation, mean temperature, and potential evapotranspiration. We implement a version with 13 parameters. The model includes a snow module (parameters TS, CFMAX, CFR, CWH), a soil moisture module (parameters β, LP, FC) and a groundwater and routing module (parameters PERC, K0, K1, K2, UZL, MAXBAS). The latter is composed of an upper reservoir for the fast response, a lower reservoir for the slow response and a channel routing component with a triangular weighting function. Here, HBV is evaluated with daily time step data over a simulation horizon of ten years, starting on 01/10/1948, including a one-year warm-up period. We use hydrologic years to remove the carry-over impacts of snow storage. The application study site is the Nezinscot River at Turner Center, Maine, USA (USGS 01055500), a catchment of 438 km2 (Duan et al., 2006).

SWAT model (50 parameters)

The Soil and Water Assessment Tool, SWAT (Arnold et al., 1993, 1998), is a semi-distributed hydrological model developed by the USDA Agricultural Research Service. The model is used worldwide to study the impact of catchment management on water availability (e.g. Tram et al., 2014), sediments (e.g. Ali et al., 2014), nutrients (e.g. Bouraoui and Grizzetti, 2008) and pesticide load (e.g. Bannwarth et al., 2014), and the impact of land use (e.g. Vaché et al., 2002) and climate change (e.g. Bae et al., 2011). It is a complex model with more than 100 parameters (though not all are typically calibrated) that includes the major catchment processes. The simulation of the hydrology is separated in two divisions in the SWAT model. The first division is the land phase of the hydrologic cycle, which controls the amount of water, sediment, nutrient and pesticide loadings to the main channel in each sub-basin. It includes multiple modules: climate of the watershed (weather generator, soil temperature), snow pack, canopy interception, surface runoff, soil moisture, groundwater, surface storage (ponds), tributary channels, plant growth and erosion, sediment, nutrient and pesticide yield. The second division is the

water or routing phase of the hydrologic cycle, which is composed of four components: water, sediments, nutrients and organic chemicals. It includes the routing in the main channel or reach to the outlet and the routing in the reservoirs. A catchment is partitioned into multiple sub-basins, which are then divided into Hydrological Response Units (HRUs). Each HRU has unique land cover, soil characteristics, and management combination and therefore requires specific values for its parameters.

The flow at the outlet of the basin is evaluated with daily time-step data over a simulation horizon of eight years, starting on 01/01/1998, including a three-year-warm-up period. The application study site is the upper Zenne River basin, Belgium, a 642 km^2 catchment (Leta et al., 2015). We use a SWAT model version that includes 21 sub-basins and 155 HRUs. We study the sensitivity of 26 flow parameters typically considered for GSA (see for instance Nossent and Bauwens, 2012). To add more parameters to GSA, we analyse the sensitivity of 6 of these 26 parameters separately for the five land use types present in the basin - Agriculture (A), Urban (U), Forest (F), Pasture (P) and Range Brush (R). We therefore consider 50 parameters for SA. It is worth noting that these 6 parameters defined at the land use scale are controlling the properties of a part of the catchment only. Therefore, they are likely to be less sensitive than the corresponding parameters defined at the catchment scale. In this chapter, we use the 2009 version of the SWAT model (Neitsch et al., 2009).

The ranges of the parameters are taken from Wagener et al. (2001) for HyMod, from Kollat et al. (2012) for HBV and from personal communication for SWAT. An initial analysis was conducted to refine those ranges for the particular application sites (see Appendix A.2) since the chosen ranges influence the results of GSA as discussed in Table 2.1. Appendix A.1 reports the parameter ranges that are used for the analysis.

Output definition

We compute the Nash-Sutcliffe Efficiency (Nash and Sutcliffe, 1970) as scalar output for sensitivity analysis:

$$NSE = 1 - \frac{\sum_{t=1}^{H}\left(y_{o,t} - y_{s,t}\right)^2}{\sum_{t=1}^{H}\left(y_{o,t} - \overline{y_o}\right)^2}, \tag{2.19}$$

where H is the number of time steps, $y_{o,t}$ is the observed value of the stream flow at time t, $\overline{y_o}$ is the average value of the observations and $y_{s,t}$ is the simulated value of the stream flow at time t. NSE is widely used in hydrology. The better the simulations match the observations, the more NSE tends towards a value of 1. Instead, values of NSE below 0 indicate that the average of the observations provides a better estimate of the observed stream flow than the model simulations. It is worth noting

that the NSE has a tendency to focus on fitting high flows due to the use of squared residuals. Different results for the GSA could be obtained with different model outputs, such as a different performance measure (e.g. bias or absolute mean error, which focuses on the water balance), a prediction function (e.g. mean annual stream flow) or an output related to a state variable (e.g. soil moisture). For a rigorous diagnostic of the parameter sensitivity, different model outputs and environmental conditions should be taken into account as discussed in Sect. 2.2.1. This is, however, beyond the scope of this chapter, which only aims to provide indications on how to choose the sample size and validate screening in GSA through particular case studies.

Experimental set-up

The computational experiments were performed with the SAFE toolbox (Pianosi et al., 2015; available at http://www.bris.ac.uk/cabot/resources/safe-toolbox/). SAFE implements the three GSA methods tested (EET, RSA, VBSA), the bootstrap technique, tools for the convergence analysis and the HyMod and the HBV models as test examples. Table 2.2 summarises the experimental set-up adopted for the analyses.

We used the sensitivity indices of Eq. (2.3) and Eq. (2.4) for VBSA (S_i^{VBM} and S_i^{VBT}), the normalised index of Eq. (2.9) for EET (S_i^{EET}) and the index of Eq. (2.10) for RSA (S_i^{RSA}). We used a maximin Latin Hypercube sampling strategy to maximise the coverage of the input factor space as discusses in Sect. 2.2.2. First, we generated a Latin Hypercube Sample of maximum size and we derived smaller samples by randomly taking sub-samples of the original LHS. The advantage of this approach is that it limits the number of model evaluations. However, the sub-samples are dependent and therefore, the evolution of the sensitivity indices for increasing sample size is smoother than it would be when using independent samples. The computation of the bootstrap estimates partially overcomes this issue, since the bootstrap technique approximates the sample distribution. For a sub-sample of a given size N, the bootstrap estimates were obtained by resampling with replacement within this sub-sample. Moreover, the stratified structure of the Latin Hypercube is not maintained in the sub-samples. For RSA, we reduced the sample size by dropping parameterizations using the maximin criterion (i.e. we randomly generate ten subsamples and we take the one with the maximal value of the minimum Euclidian inter-point distance) in order to cover the search space as evenly as possible. For EET and VBSA instead, due to the particular structure of the samples, the parameterizations in the initial LHS were dropped randomly without using the maximin criterion. We note that strategies exist to avoid this loss of stratification and we refer for instance to Andres (1997) for further details.

Experimental set-up for sampling of the parameter space		
Parameter distribution	Uniform distribution (no a priori knowledge)	
Sampling strategy	Maximin Latin Hypercube (uniform 1D margins and maximisation of the minimum inter-point Euclidean distance)	
Experimental set-up for convergence analysis		
Threshold value for RSA		0.4
Number of bootstrap replicates ($\boldsymbol{N_{boot}}$)		1000
Maximum number of model evaluations for HyMod ($\boldsymbol{N_{max}}$)	EET	78,000
	RSA	10,000
	VBSA	420,000
Maximum number of model evaluations for HBV ($\boldsymbol{N_{max}}$)	EET	560,000
	RSA	20,000
	VBSA	600,000
Maximum number of model evaluations for SWAT ($\boldsymbol{N_{max}}$)	EET	102,000
	RSA	30,000
	VBSA	520,000
Experimental set-up for validation of screening		
Sample size for unconditional output ($\boldsymbol{N_u}$)		2000
Sample size for conditional output ($\boldsymbol{N_c}$)		1600
Number of conditioning values ($\boldsymbol{n_c}$)		20

Table 2.2 Experimental set-up for testing the methodology

2.5 Test results

In this section we present the results of the convergence study and of the screening threshold investigation. Figure 2.5 shows the evolution of the sensitivity indices for increasing numbers of model evaluations for the three GSA methods and the three case studies. It is worth noting that an apparent convergence of the bootstrap mean of the sensitivity index (flattening of the line) is a necessary but not sufficient condition for convergence. It can happen that the bootstrap mean takes similar values in two different samples 'by chance' while the actual statistical convergence is not reached yet (the confidence intervals are still very wide). Therefore, it is important to also include information about the confidence intervals. Fig. 2.6 shows the evolution of the convergence statistics defined in Sect. 2.1. In this figure, vertical lines indicate the number of model evaluations, N, suggested in Saltelli et al. (2008, Table 6.9) for the three GSA methods. These values correspond to a base sample size n equal to 10 for EET, 2000 for RSA and 1000 for VBSA. Table 2.3 reports the values of the sample sizes that ensure convergence of the indices ($Stat_{indices} < 0.05$), of ranking ($Stat_{ranking} < 1$) and of screening

($Stat_{screening}$ <0.05) when using our suggested convergence statistics. Note that for VBSA, the number of model evaluations necessary to reach convergence refers to the joint estimation of the two indices (VBM and VBT), because these two indices are obtained from the same sample of model evaluations.

2.5.1 Convergence of sensitivity indices

The top panels in Fig. 2.6 (a-c) show the values of the convergence statistic for the value of the sensitivity indices for increasing sample size. They show that the sensitivity indices converge first for RSA compared to the other methods (see also Table 2.3). RSA requires at most 15,000 model evaluations for the three case studies considered. EET and VBSA require a much larger number of model evaluations, generally of the order of magnitude of several hundreds of thousands, which is prohibitive when simulations are computationally expensive. In particular for the SWAT model (Fig. 2.6.c), sensitivity indices have not reached convergence even after 102,000 model evaluations for EET and 520,000 model evaluations for VBSA for both Main effect (VBM) and Total effect (VBT). Moreover, we observe that with the values suggested in the literature (vertical dashed lines in Fig. 2.6), the width of the confidence intervals of the sensitivity indices are quite wide, especially for EET. However, for RSA, $Stat_{indices}$ is already quite close to its threshold value after 2000 model evaluations since the width of the confidence intervals is equal to 0.09 for the HyMod and the SWAT models.

2.5.2 Convergence of parameter ranking

The middle panels in Fig. 2.6 (d-f) show the value of the convergence statistic for ranking for increasing sample size, for the three models. Fig. 2.5 reports the ordering of the most sensitive parameters.
We first observe from Fig. 2.5 that the three GSA methods provide different rankings of importance for the model parameters. This is reasonable since the three methods measure sensitivity according to different rationales and assumptions. The rankings given by EET and VBT indices are generally quite consistent with each other. In particular, the two methods identify the same group of most sensitive parameters (ALPHA and RF for HyMod in Fig 2.5a,d; FC and TS for HBV in Fig. 2.5.e,h; CN2_A, CH-K2 and CH-N2 for SWAT in Fig. 2.5.i,l) and of least sensitive parameters. For HBV and SWAT, the ranking given by RSA (Fig. 2.5.g,k) differs from the one produced by EET and VBSA, which might be explained by the fact that RSA does not detect many types of interactions (see p.190 in Saltelli et al., 2008).

We find that the ranking generally converges faster than the estimates of the sensitivity indices when comparing the middle panels of Fig. 2.6 (d-f) with the top panels (a-c) and the corresponding values of the sample sizes reported in Table 2.3. However, the results are different for RSA applied to SWAT.

The two more sensitive parameters (CN2_A and SLOPE_A) clearly separate out while the other parameters have very similar values of the sensitivity indices. Since this happens also for parameters that have a relatively high sensitivity index, minor fluctuations in these indices values can lead to large differences in ranking.

When comparing the convergence of the ranking across the three case studies, we observe that the number of model evaluations N required for convergence usually increases with the number of parameters, M, as expected. Interestingly, this does not seem to be the case for EET. We indeed observe that for the HBV model (Fig. 2.6.e), 7000 model evaluations are necessary for the convergence of the ranking provided by EET while for the SWAT model (Fig. 2.6.f) only 4590 model evaluations are necessary. The rate of convergence for the ranking appears to depend on the specific case study and on the relative value of the sensitivity indices among the different parameters. For EET applied to the SWAT model (Fig. 2.5.i), the sensitivity indices of the three most influential parameters are significantly higher than all the others, while for the HBV model (Fig. 2.5.e) they are more evenly spread. As a result, the ranking of the most influential parameters stabilises faster for SWAT than for HBV.

Analysing the rate of convergence across the three GSA methods, we observe that the convergence of the ranking is reached quickest for the RSA method compared to the other methods for the HyMod and the HBV models (Fig. 2.6, Table 2.3). EET appears to require fewer model evaluations than VBSA while providing a ranking consistent with VBT. We also note that the ranking obtained with the number of model evaluations suggested in the literature (vertical dashed lines in Fig. 2.6) is generally not robust for the two more complex models (HBV and SWAT) since $Stat_{ranking}$ takes values above 1 at these sample sizes. In particular for EET, with a base sample size of n =10 (N =140 for HBV and N =520 for SWAT), $Stat_{ranking}$ is higher than 4 for these two models. However, $Stat_{ranking}$ underestimates the rate of convergence for the VBSA method applied to the SWAT model (Fig. 2.6.f). We indeed note that the curve for VBM has large oscillations and only converges for a very high number of model evaluations. This is due to rank reversals happening between some low-sensitivity parameters, while the ranking of the most sensitive parameters stabilises already after a much lower number of model evaluations. This shows that rank reversals for low-sensitivity parameters can still have some influence on our proposed indicator $Stat_{ranking}$ (see Section Appendix A.3 for further analysis).

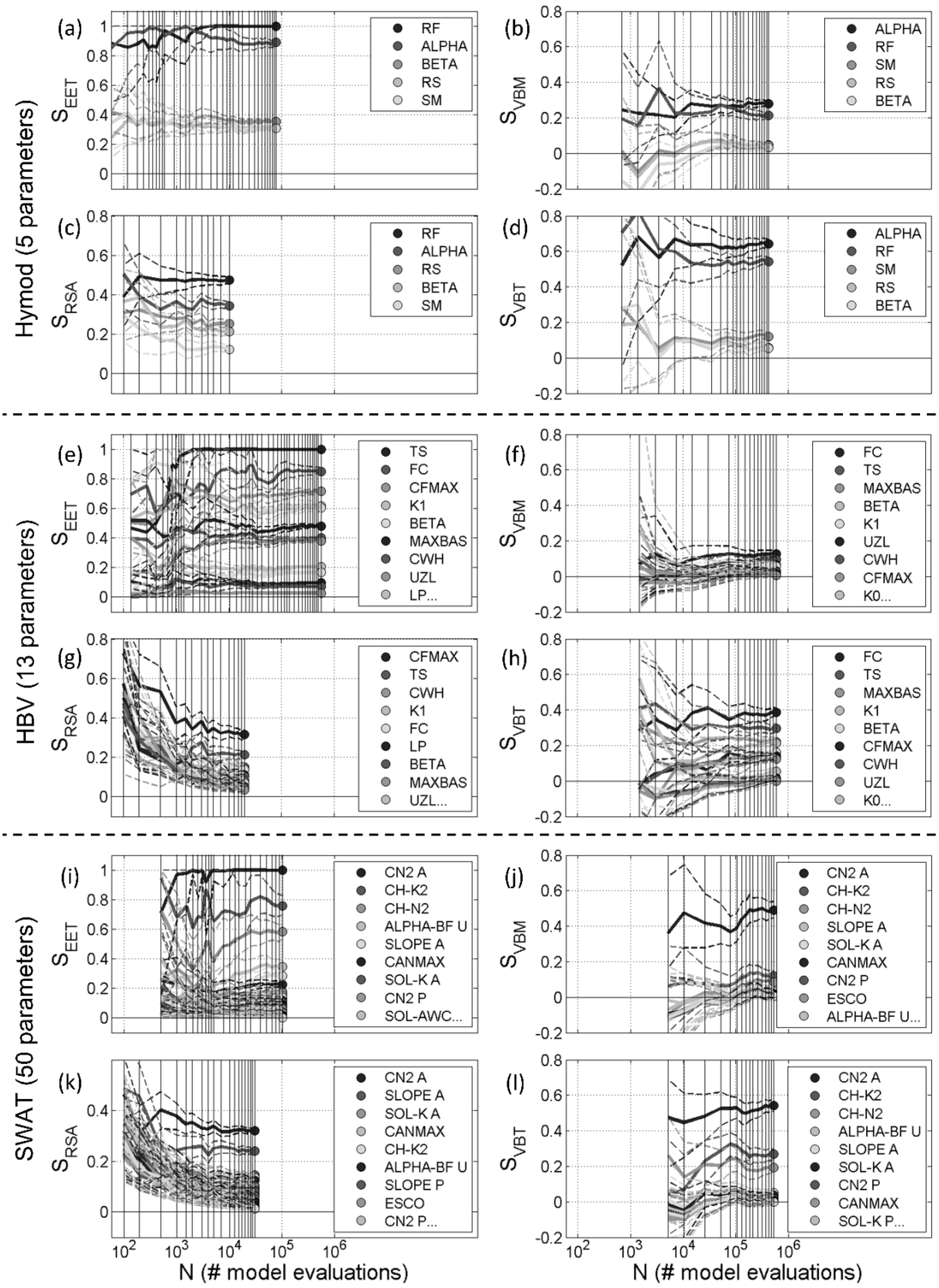


Figure 2.5 Convergence plots.

The figure represents the sensitivity indices of the model parameters for HyMod, HBV and SWAT, estimated using an increasing number of model evaluations *N*, computed for the different GSA methods, Elementary Effect Test (EET), Regional Sensitivity Analysis (RSA), Variance-based methods Main effect (VBM) and Total effect (VBT). The solid lines are the bootstrap means of the sensitivity indices and the dashed lines are the 95% bootstrap confidence intervals.

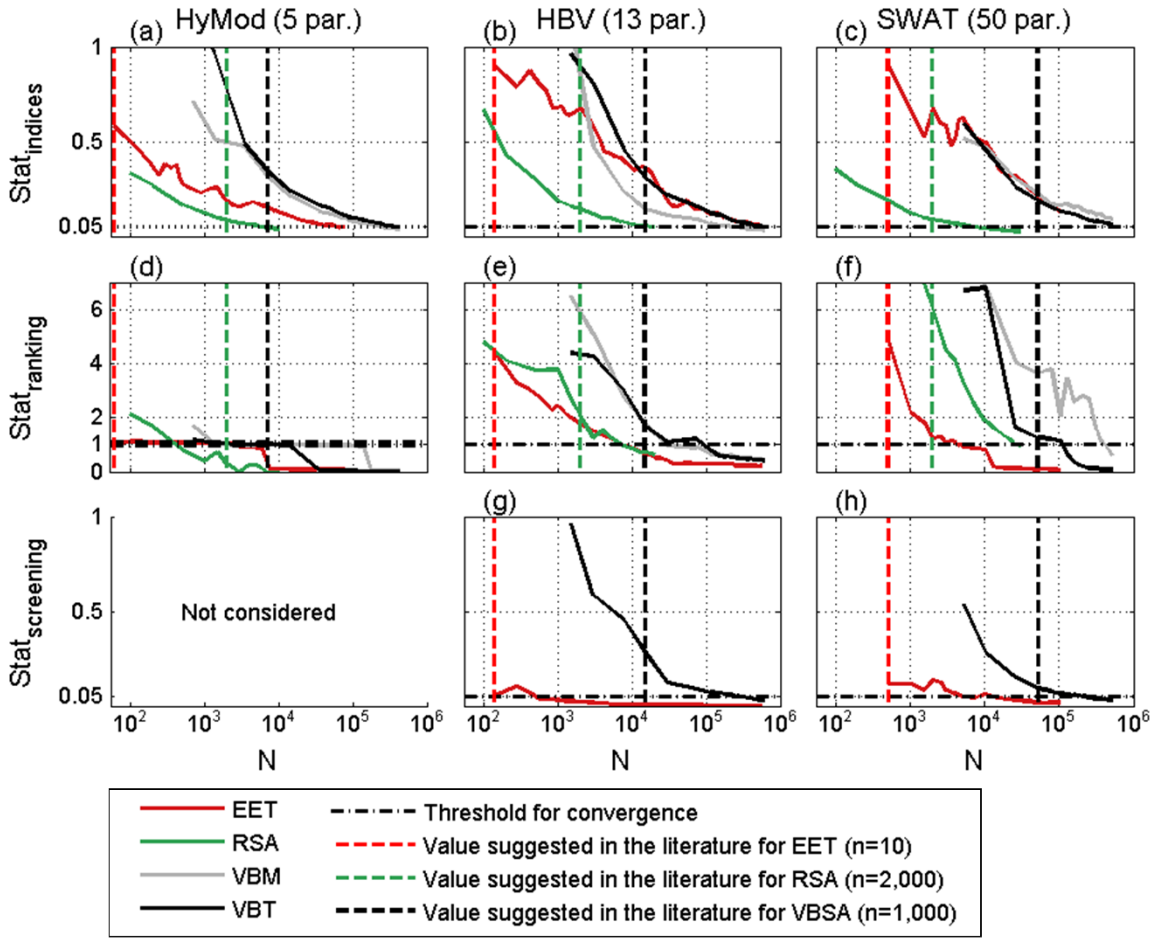


Figure 2.6 Value of the convergence statistics against number of model evaluations *N*. $Stat_{indices}$ (convergence of sensitivity indices) is the maximum width of the bootstrap confidence intervals, $Stat_{ranking}$ (convergence of ranking) is an adjusted rank correlation coefficient, and $Stat_{screening}$ (convergence of screening) is the maximum width of the bootstrap confidence intervals for the lower-sensitivity parameters. The results are reported for the three case studies. The dashed vertical lines report the number of model evaluations suggested in the literature (Saltelli et al., 2008, Table 6.9) and the dashed horizontal black lines represent the threshold values for the convergence statistics, below which we assume that convergence is reached.

GSA method	Objective	Number of model evaluations for convergence		
		HyMod (M = 5)	HBV (M = 13)	SWAT (M=50)
EET $\boldsymbol{N = n(M + 1)}$	Indices	N = 60,000 (n = 10,000)	N = 560,000 (n = 40,000)	N > 102,000 (n > 2000)
	Ranking	N = 3000 (n = 500)	N = 7000 (n = 500)	N = 4590 (n = 90)
	Screening		N = 560 (n = 40)	N = 5100 (n= 100)
RSA $\boldsymbol{N = n}$	Indices	N = 7000 (N/M = 1400)	N = 15,000 (N/M = 1154)	N = 7000 (N/M = 140)
	Ranking	N = 500 (N/M = 100)	N = 7000 (N/M = 538)	N = 25,000 (N/M = 500)
	Screening			
VBM $\boldsymbol{N = n(M + 2)}$	Indices	N = 210,000 (n = 30,000)	N = 225,000 (n = 15,000)	N > 520,000 (n > 10,000)
	Ranking	N = 1400 (n = 200)	N = 30,000 (n = 2000)	N = 416,000 (n = 8000)
	Screening			
VBT $\boldsymbol{N = n(M + 2)}$	Indices	N = 350,000 (n = 50,000)	N = 450,000 (n = 30,000)	N > 520,000 (n > 10,000)
	Ranking	N = 3500 (n = 500)	N = 112,500 (n = 7500)	N = 130,000 (n = 2500)
	Screening		N = 262,500 (n = 17,500)	N = 208,000 (n = 4000)

Table 2.3 Number of model evaluations N (and corresponding base sample size n) necessary to reach convergence of sensitivity indices ($Stat_{indices} < 0.05$), ranking ($\text{Stat}_{\text{ranking}} < 1$) and screening ($Stat_{screening} < 0.05$) based on analysis.

M is the number of parameters. RSA and VBM are not used for screening. We do not screen the HyMod parameters (they are all influential).

2.5.3 Convergence of parameter screening

The bottom panels in Fig. 2.6(g,h) show the value of the convergence statistic for screening for increasing sample size. Fig. 2.7 reports the results of the validation test.

Convergence of sensitivity indices for low-sensitivity parameters

For the purpose of screening, we consider only EET and VBT while we exclude RSA and VBM because they do not account for interactions. Furthermore, screening is not applied to the HyMod model since all its five parameters are found to be influential in our experimental set-up. We compute the convergence statistic for screening for the other two case studies (HBV and SWAT) (Fig. 2.6.g,h).

Comparing the bottom panels of Fig. 2.6 (g,h) with the top panels (b,c), we observe that the convergence of the sensitivity indices for the lower-influence parameters ($Stat_{screening}$) is quicker compared to the other parameters ($Stat_{indices}$). In particular, for EET, after 560 model evaluations (n =40) for HBV

and 5100 model evaluations (n =100) for SWAT, the lower-sensitivity parameters have already converged while it takes hundreds of thousands of model evaluations for all the sensitivity indices to converge (see also Table 2.3). The convergence of the indices for the lower-sensitivity parameters requires still more model evaluations than usually suggested in the literature.

Analysing the results obtained with EET across the two case studies, we observe that the convergence of the screening for HBV is reached for a smaller number of model evaluations than for SWAT, which is expected because HBV has a lower number of parameters. However, for VBT, the screening converges slightly earlier for the SWAT model than the HBV model. Therefore, as observed for the convergence of the ranking, the number of model evaluations required to stabilise the sensitivity indices of lower-sensitivity parameters does not necessarily increase with the number of parameters considered.

Moreover, we notice some oscillations in the value of $Stat_{screening}$ for EET when the number of model evaluations is small (Fig. 2.6.g,h). For small sample sizes, the bootstrap technique is not able to assess the 'true' variability of the sensitivity indices because the small samples may not contain enough information. We indeed note that the width of the bootstrap confidence interval for some low-influence parameters increases significantly with the sample size early on, before decreasing and reaching convergence when further increasing the sample size.

Validation of screening

We validated the screening and investigated the value of the screening threshold at the sample sizes for which convergence is reached (reported in Table 2.3). Figure 2.7 shows the estimated KS_{max} for increasing values of the assumed screening threshold for the two models and the two GSA methods. The Figure also shows the critical values of the KS-statistics at different significance levels. As explained in Sect. 2.2, we used a small value of the significance level (0.001) when applying the KS-test, so that the screening is not too conservative.

The screening results of EET and VBT are consistent. For the HBV model (Fig. 2.7.a,c), one insensitive parameter is identified (K2) by both EET and VBT. For the SWAT model (Fig. 2.7.b,d), 27 insensitive parameters are identified by EET and only 21 of those 27 insensitive parameters are identified with VBT. Therefore, for SWAT, EET identifies a higher number of non-influential parameters for a much smaller number of model evaluations than VBT. The reason is that for VBT, the 95% confidence intervals of the sensitivity index for the lower-sensitivity parameters are strongly overlapping while for EET we observe much less overlap. As a result, for SWAT, EET is able to differentiate the sensitivities among the lower-sensitivity parameters, allowing for a better separation between parameters that have a negligible impact on the output and parameters that have a low but non-negligible influence.

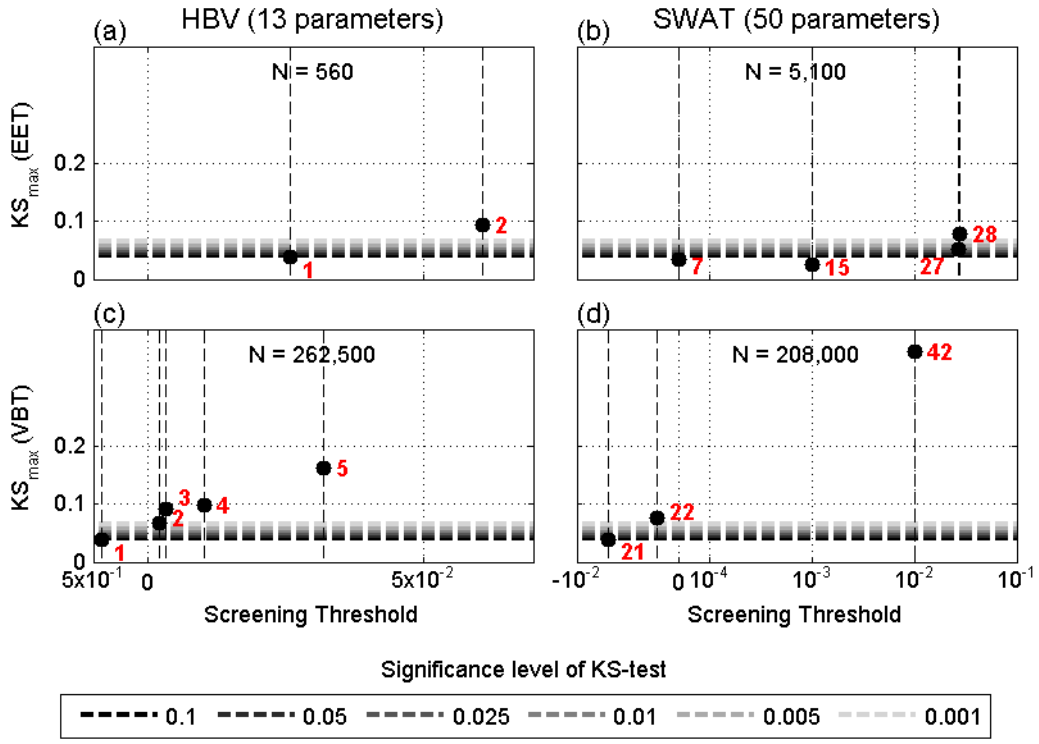


Figure 2.7 Validation of screening results.
KS_{max} against assumed Screening Threshold on the sensitivity indices, computed by EET and VBT methods for the HBV and SWAT models. N is the number of model evaluations used to compute the sensitivity indices (this was chosen to ensure screening convergence, see also Table 2.3). The critical values of the KS-statistic are reported for different significance levels. Next to each data point is the number of parameters that are fixed to compute the KS-statistic.

For EET (Fig. 2.7.a,b), the screening is validated for screening threshold values of 2.6×10^{-2} and 2.7×10^{-2} for HBV and SWAT respectively. For the VBT method (Fig. 2.7.c,d), it is validated for a screening threshold value of -8.4×10^{-3} and -4.0×10^{-3} for HBV and SWAT respectively. We note that the value of the screening threshold is slightly negative for VBT because the value of the sensitivity indices for the lower-sensitivity parameters has still not perfectly converged (the width of the confidence interval is equal to 0.05 and not 0). The values of the screening thresholds for VBT are lower than the value of 0.01 generally used in the literature (see for instance Tang et al., 2007; Sin et al., 2011; Cosenza et al., 2013; Vanrolleghem et al., 2015). For the SWAT model (Fig. 2.7.d), using a threshold of 0.01 would lead to a KS-statistic well above the critical values. On the other hand, the subset of parameters with a sensitivity index below a value of 0.01 seems to remain stable when the sample size is increased, since it does not change up to the maximum sample size analysed (N=520,000) (result not shown). Therefore, a screening threshold of 0.01 as used before appears to be too high to screen the parameters of the SWAT model because it would lead to a larger number of parameters being deemed insensitive.

2.6 Discussion of the results

2.6.1 Implications of the results for GSA implementation

The results of the present chapter can be used to derive guidelines for the implementation of GSA, while of course considering that we used specific case studies for testing, as well as a specific experimental set-up, e.g. the choice of parameter ranges, of the output function, of the simulation period, of the sampling strategy. Shin et al. (2013) for example demonstrated the influence of some of these choices on GSA results.

Firstly, we observe that different base sample size values are needed for different models. Interestingly, from our results, no clear relationship emerges between the number of parameters M and the number of model evaluations N needed to reach convergence (Table 2.3). The values of the base sample size n for EET and VBSA and of the ratio N/M for RSA vary between models. Moreover, while the number of model evaluations generally increases with model complexity in our case studies, for EET, the ranking of the 50 model parameters of the SWAT model converges before the ranking of the 13 model parameters of the HBV model. Besides model complexity, we believe that the rate of convergence depends on other factors, such as the relative sensitivity of the parameters (i.e. the closer the sensitivity of the influential parameters, the slower the convergence of the ranking). We expect to observe even more variations between convergence rate of GSA when considering a larger range of models and types of parameter variation. In particular, highly non-linear models (e.g. models that exhibit low frequency high consequence effects) are likely to show different convergence dynamics. Therefore, our study demonstrates that it is essential to check the convergence of the GSA results within the case under study and not only rely on numbers reported in previous studies. In particular, we find that sample sizes needed to reach the convergence are larger than suggested in the literature (e.g. Saltelli et al., 2008, Table 6.9) (Fig. 2.6, Table 2.3). This is consistent with results reported in the few other studies that specifically focused on GSA convergence (Fig. 2.1).

Secondly, we note that the convergence of ranking and screening can be reached at significantly lower number of model evaluations than the convergence of the sensitivity indices. This observation emphasises the importance of distinguishing between the three types of convergence (indices, ranking and screening) to use computational resources efficiently. Full convergence of the value of the sensitivity indices may not be needed if one is interested in ranking (Factor Prioritization) or screening (Factor Fixing). In this case, trying to reach the full convergence of the value of the sensitivity indices constitutes a Type III error or framing error (Saltelli et al., 2008) given that the right answer (sample size for convergence) is sought for the wrong questions (what is the exact value of the sensitivity indices?). If instead the objective would be Variance Cutting, a converged value of VBSA sensitivity

indices is needed. For RSA and EET, the values of the sensitivity indices do not provide much information by themselves and determining converged values seems to be of little interest.

Furthermore, we find that it is important to validate the screening results to avoid classifying influential parameters as insensitive. For EET, we obtain similar values of the screening threshold (between 0.025 and 0.03) for both the HBV and the SWAT models. A more exhaustive analysis would be needed to study the applicability of these values to other case studies. For VBT, we obtain negative values of the screening threshold very close to zero, even after a high number of model evaluations (more than 200,000). This indicates that the sensitivity indices for the insensitive parameters have not perfectly converged because their actual value should be zero or very small but positive. Consequently, the exact (converged) value of the screening threshold T cannot be determined in these case studies. Nonetheless, we showed that the typical value of the screening threshold used for the VBT method (0.01) is not suitable for screening the parameters of the SWAT model. Some influential parameters appear to have a sensitivity index below this threshold value and the actual value of the screening threshold is thus below 0.01.

Finally, we can draw a comparison between the different GSA methods. In the case studies analysed here, EET provides ranking and screening of the model parameters consistent with VBT using much fewer model evaluations. For the SWAT model, after applying the validation test, we even identified a higher number of insensitive parameters with EET compared to VBT. Therefore, for SWAT, a complex model, VBT may not be suitable, since it would require a very high number of model evaluations and EET may be a better choice. Moreover, the results provided by RSA converge quickly in the three case studies. More broadly, our results therefore support the recommendations of previous studies regarding the use of a multi-method approach (Borgonovo et al., 2017; Pianosi et al., 2017 as mentioned in Sect. 2.2.1). Given that the different GSA methods rely on different assumption, it is good practice to apply different GSA methods to the same case study within the limits of the available computational resources.

2.6.2 Consistency of our results with previous studies

Our study confirms the results found in the literature regarding the relative computational cost of the three GSA methods EET, RSA and VBSA (comparing our results in Table 2.3 and Fig. 2.6 with the results of other studies summarised in Fig. 2.1). We also find that for a given number of parameters, the ranking provided by RSA converges before the ranking provided by VBSA. Likewise, ranking and screening provided by EET converge before the ranking and the screening provided by VBSA. In particular, Campolongo et al. (2007) empirically demonstrated that EET allows for screening the model parameters with much fewer model evaluations than VBSA as reported in Sect. 2.2.2, which is corroborated by our study. However, we show that the result does not hold when the objective is the full

convergence of the indices: the order of magnitude of the number of model evaluations necessary to reach convergence of the indices is the same for EET and VBSA, while RSA is much less computationally expensive. Moreover, for HyMod and HBV, convergence of indices and ranking given by Variance-Based Main effect (VBM) requires fewer model evaluations than Variance-Based Total effect (VBT). Nossent et al. (2011) already noticed that the convergence of VBM tends to be faster than the convergence of VBT. No conclusion can be drawn from the results of the SWAT model on this point, because sensitivity indices did not converge for either VBM or VBT even after 520,000 model evaluations and the convergence statistic for ranking underestimates the rate of convergence in particular for VBM (see Sect. 2.5.2).

It is worth comparing our results with the study by Yang (2011) for the HyMod model applied to the Leaf River Catchment. Although the experimental set-up differs between Yang's and our study (simulation period, threshold for RSA, sampling strategy), we observe similarities in the results. The sample sizes for the convergence of the sensitivity indices (EET and RSA) and of ranking (VBSA and EET) have the same order of magnitude. Yet, we note a significant difference for the convergence of the sensitivity indices for VBSA (n = 30,000 in our study while n = 3,000 in Yang's study). In Yang's study, the convergence analysis was performed qualitatively and in ours, the quantitative criterion used may be conservative. No sample size was explicitly given for the convergence of the ranking for RSA in Yang's paper, while screening convergence was not considered.

Finally, as already briefly mentioned in Sect. 2.6.1, we observe a general coherence between EET and VBT results since both methods consistently separate out the most sensitive and the least sensitive parameters, in agreement with previous studies (e.g. Campolongo et al., 2007; Confalonieri et al., 2010; Herman et al., 2013). However, also in accordance with these previous studies, differences are observed in parameter ranking. EET may be a suitable alternative to VBT when model simulations are computationally expensive depending on the specific case study.

2.6.3 Limitations of the methodology proposed and direction for future improvements

This chapter introduces methods to formally assess the convergence of GSA results, relying on the definition of quantitative convergence statistics (Sect. 2.4.1). We set threshold values on these convergence statistics below which we assume that convergence has been reached. We believe that these threshold values ensure a sufficient degree of convergence of GSA results to obtain reliable results, although they could be conservative. The adjusted and weighted rank correlation coefficient here proposed, was shown to be more suitable than the Spearman rank correlation coefficient for comparison of parameter rankings. In fact, it emphasises the differences in ranking for the more influential parameters while it reduces the impact of the low-sensitivity parameters. However, low-sensitivity

parameters can still contribute to the value of the statistic when their sensitivity index is not negligible and when their rank is highly variable (see Appendix A.3 and discussion in Sect. 2.5.2). Therefore, the convergence statistic for ranking may underestimate the rate of convergence, which leads again to an estimate on the conservative side. In fact, when the statistic takes low values, GSA users can be quite confident that ranking convergence has been reached.

The methodology introduced to assess convergence may not be suitable when sample sizes are very small. In this case, the sample may not provide sufficient coverage of the parameter space so that bootstrapping may show wrongly low uncertainty of the sensitivity index estimates (see discussion in Sect. 2.5.3). This problem has for example been observed by Isaksson et al. (Isaksson et al., 2008). At very small sample sizes, our analysis could be misleading and incorrectly suggest that GSA has converged. Likewise, when low frequency high consequence events can occur in a model (i.e. a small number of input data points can produce a large effect on the output), bootstrapping might fail to assess the uncertainty of the sensitivity index estimates (if these highly influential values are not present in the sample).

The methodology for screening (introduced in Sect. 2.4.1) applies for models with a reasonable number of input factors so that it is computationally affordable to estimate the sensitivity for every single input factor. When in contrast the number of input factors is very high compared to available computational resources (like in supersaturated designs), it might not be possible to estimate the effect of every single input factor. Nevertheless, the methodology proposed in this chapter could be applied for such models if input factors are assigned to a given number of groups, and if GSA is performed by taking these groups of input factors as inputs. In this case, screening consists in identifying the influential groups (i.e. at least one input factor in the group is sensitive) and the non-influential ones (i.e. all input factors in the group are insensitive). We refer to Saltelli et al. (2008) for more details on screening for supersaturated designs and group sampling.

We propose in Sect. 2.4.2 a quantitative validation method for the screening results based on the computation of the KS-statistic between unconditional output (obtained by varying all parameters) and conditional output (only influential parameters are varied). One main drawback of this method is that it requires further model evaluations for the computation of the conditional outputs. Further investigation is needed regarding the robustness of the KS-test. One possibility would be to use the bootstrap technique to compute the KS-statistic. However, in this chapter, we found that bootstrapping tends to overestimate the KS-statistic (more details are given in Appendix A.4). An interesting complementary approach to our validation procedure for screening has been proposed by Khorashadi Zadeh et al. (2017). This approach is based on the use of a ‘dummy’ parameter, i.e. a parameter which is added to the set of

input factors, but which does not have any influence on the model predictions because it is not explicitly included into the model equations. The sensitivity index of the dummy parameter is an estimate of the approximation error of the Monte-Carlo procedure. Non-influential input factors are identified as the inputs that have a value of the sensitivity index below the one of the 'dummy' parameter.

Given all the above-mentioned limitations of the methodology introduced in this chapter, we believe that further investigation is needed for (1) convergence statistic for ranking in order to make it potentially less conservative, (2) the KS-test to formalise the assessment of its convergence, (3) the bootstrap technique to overcome its limitations in particular for the KS-statistic and (4) the LHS design to further examine and test strategies that would avoid a loss of stratification when increasing or decreasing the sample size as examined in Andres (1997). Regarding (4), this would enable increasing an already existing sample to reach convergence of the sensitivity analysis while maximising the coverage of the input factor space. In the context of a convergence analysis similar to the one performed in this chapter, this would provide more rigorous results because based on the analysis of a Latin Hypercube sample at each sample size, while our sampling strategies introduced loss of stratification (as discussed in Sect. 2.4.4). Additionally, future work should include a comparison of the convergence speed between different sampling strategies to help GSA users with this choice.

2.7 Conclusions

Global Sensitivity Analysis aims to apportion the uncertainty of the output of a model among the its different input factors. It regroups a large range of different methods based on different assumptions and rationales and that can therefore serve different purposes. The present chapter provides an overview of these different methods and in particular Variance-Based Sensitivity Analysis (VBSA, or Sobol' method), the Elementary Effect Test (EET, or method of Morris), Factor Mapping techniques such as Regional Sensitivity Analysis (RSA) and Classification and Regression Trees (CART) and finally it briefly introduces visual techniques such as parallel coordinate plots and scatter plots. GSA is widely used to support model parameter estimation, model verification, support the prioritization of efforts for uncertainty reduction and model control analysis or help with model-based decision-making. The technique can be applied to any model and any type of input factors, as long as the uncertainty of the input factors (distribution) can be quantified.

However, the application of GSA involves a number of critical choices that can lead to very different sensitivity outcomes and, importantly, that can affect the credibility and robustness of the results. More specifically, the implementation of GSA is generally sampling-based and therefore requires the specification of the sample size, and the identification of the non-influential input factors (screening) is

typically based on the definition of a threshold on the sensitivity indices value below which input factors are considered as non-influential. These choices have been subject to little investigation in previous studies. In this chapter, we aimed to fill this gap by defining and testing quantitative criteria to assess the convergence of sensitivity indices, of ranking (ordering among the influential parameters) and of screening (identification of the insensitive parameters). Second, we propose a quantitative and unconditional method to validate the screening results to avoid classifying influential parameters as non-influential.

This study demonstrates that it is indeed important to separately assess the convergence of sensitivity indices, ranking and screening, since these three objectives may require different numbers of model evaluations. It is not always necessary to reach the full convergence of the value of all the sensitivity indices. In fact, the parameter ranking and the sensitivity indices of the low-influential parameters (and therefore the screening) may converge first. We also observed that values of the sample size sometimes suggested in the literature (e.g. Saltelli et al., 2008, Table 6.9) can be insufficient to reach convergence of GSA results, as observed for the two more complex models analysed here (HBV and SWAT). Since, no clear relationship emerged between the number of model parameters subject to GSA and the number of model evaluations necessary to reach convergence, we recommend that GSA users always check the convergence of their results within their specific case study. The study also shows that EET can provide a good approximation of the ranking and screening given by VBSA for much fewer model evaluations, as has already been noted in previous studies. As far as RSA is concerned, it appears to converge quickly in the case studies considered, although, as discussed in previous studies, it cannot be used when the objective is to study parameter interactions.

The next chapters of the thesis will adopt some of these GSA methods. To inform the development of the vegetation-recharge model for karst areas (Chapter 3), we will use the EET, which has been confirmed to provide reliable ranking results in this chapter, and RSA, parallel coordinate plot and scatter plots to visualise the results. In Chapter 4, we will apply VBSA and CART to analyse the sensitivity of karst groundwater recharge to changes in climate and land cover. In all these applications, the convergence of GSA results will be analysed using the methodology proposed in the present chapter to ensure that the appropriate sample size is used.

Chapter 3. A parsimonious large-scale integrated vegetation-recharge model to simulate the impact of climate and land cover change in karst regions

This chapter is adapted from a work under review that has been partially modified to improve the consistency throughout the thesis:

> *Sarrazin, F., Hartmann, A., Pianosi, F. and Wagener, T.: V2Karst v1.0: A parsimonious large-scale integrated vegetation-recharge model to understand the impact of climate and land cover change in karst regions, Geoscientific Model Development Discussion, in review, doi:10.5194/gmd-2017-315, 2018.*

For this work, partial support to Andreas Hartmann was provided by the Emmy Noether-Programme of the German Research Foundation (DFG; grant number HA 8113/1-1; project 'Global Assessment of Water Stress in Karst Regions in a Changing World'). Francesca Pianosi was partially supported by a UK Engineering and Physical Sciences Research Council fellowship (EPSRC; grant number EP/R007330/1). This work has also benefitted from the advice of Rafael Rosolem on FLUXNET data processing and evapotranspiration modelling and from the advice of Timothy Foster on evapotranspiration modelling. Yoshihide Wada provided clarification on evapotranspiration representation in the PCR-GLOBWB model. Dimitri Rambourg is thanked for his analysis of the MODIS leaf area index and evapotranspiration products. The investigators of the four FLUXNET sites (DE-Hai, Es-Lju, FR-FBn and FR-Pue) are thanked for allowing us to use their data, and in particular Martina Mund and Manfred Fink (De-Hai) Penelope Serrano Ortiz, Francisco Domingo Poveda and Andrew Kowalski (ES-Lju), Richard Joffre and Serge Rambal (FR-Pue), and Guillaume Simioni (FR-FBn) shared information on the characteristics of the FLUXNET sites. Eleonora Canfora and Dario Papale from the FLUXNET European Fluxes Database Cluster provided information on the FLUXNET data. The work has also benefited from the comments of two anonymous referees and this chapter includes parts of the revisions that are discussed in the online interactive discussion (https://www.geosci-model-dev-discuss.net/gmd-2017-315/).

3.1 Introduction

Carbonate rocks, from which karst systems typically develop, are estimated to cover 10-15% of the world (Ford and Williams, 2007). Karst aquifers are an important source of drinking water for almost a quarter of the wold population (Ford and Williams, 2007) and have a critical role in sustaining food production because most karst areas present some form of agricultural activity (Coxon, 2011). In particular, in Europe, carbonate rock areas cover 14-29% of the land area, and some European countries such as Austria and Slovenia derive up to 50% of their total water supply from karst aquifers (COST, 1995; Chen et al., 2017).

Karst systems are characterised by a high spatial variability of bedrock and soil permeability due to the presence of preferential flow pathways (Hartmann et al., 2014). The soluble carbonate bedrock is structured by large dissolution fissures or conduits (Williams, 1983, 2008) and the typically clayey soil often contains cracks (Blume et al., 2010; Lu et al., 2016) where infiltrating water is concentrated. Therefore, a large part of the groundwater recharge occurs as concentrated and fast flow in large apertures and the other part as diffuse and slow flow in the matrix (Hartmann and Baker, 2017). Preferential flow pathways are particularly developed in karst, but they are also widely found in many other systems, due to root and organism activities, discontinuous subsurface layers, surface depressions, soil desiccation, tectonic processes and physical and chemical weathering (Hendrickx and Flury, 2001; Uhlenbrook, 2006; Beven and Germann, 2013)

Preferential infiltration is typically triggered when thresholds on the rain intensity and soil moisture are exceeded (Tritz et al., 2011; Rahman and Rosolem, 2017). When activated, preferential infiltration pathways may enhance groundwater recharge while limiting surface runoff (e.g. Bargués Tobella et al., 2014)). In karst, permeability is often so high that surface runoff is negligible, and virtually all precipitation infiltrates (Fleury et al., 2007; Contreras et al., 2008; Hartmann et al., 2014). Furthermore, preferential infiltration pathways can affect the temporal dynamics of recharge. For instance Cuthbert et al. (2013) showed that macro-pores in the soil can generate quick responses in the water table, and Arbel et al. (2010) observed that dripping rates in a karst cave can fluctuate following precipitation inter-seasonal and intra-seasonal variations.

Changes in weather patterns (e.g. due to climate change), and specifically in the precipitation intensity and frequency, may alter the activation of preferential flow pathways. From previous studies in non-karst areas we can learn that changes in the intensity and frequency of precipitation events have an impact on the water yield. For instance, using an analytical framework and synthetic experiments. Porporato et al. (2004) established a dependency between the soil water balance and both the frequency and intensity of precipitation events, while Joothityangkoon and Sivapalan (2009) determined different

theoretical hydrological regimes based on the intensity and frequency of the precipitation input over Australian catchments. Regarding groundwater recharge, observation records are scarce but data indicate a sensitivity to extreme rainfall in a semi-arid tropical region (Taylor et al., 2013a) and in a seasonally humid tropical region (Owor et al., 2009).

In karst areas, few modelling studies showed that groundwater recharge (Loáiciga et al., 2000; Hartmann et al., 2012a), spring discharge (Hao et al., 2006), and streamflow (Samuels et al., 2010) respond to changes in climate. However, to the authors' knowledge, only one study by Hartmann et al. (2017) analysed quantitatively the sensitivity of karst groundwater recharge to specific precipitation characteristics, namely the mean precipitation and the intensity of heavy precipitation events, and compared the results obtained over karst and non-karst areas. That study suggests that, due to the presence of preferential flow pathways, recharge in karst systems tends to show higher sensitivity to mean precipitation and to the intensity of heavy precipitation events in dry climates, and lower sensitivity in wet climates compared to non-karst systems.

Land cover/use change could also have a major impact on hydrological processes in the future (Vörösmarty, 2002; DeFries and Eshleman, 2004). Changes in land cover/use can impact the partitioning between green water (evapotranspiration losses) and blue water (water potentially available for human activities, namely groundwater recharge and runoff). Green water tends to be higher for forested areas than for shorter vegetation (e.g. Brown et al., 2005), which has also been found in few local studies in karst areas (Williams, 1993; Ford and Williams, 2007). Significant land cover/use changes are expected to occur in the future, including in karst areas in Europe, the Middle East and Northern Africa. These will partly be caused by modifications in socio-economic and technological factors, such as changes to food and wood demand or changes in agricultural management practices that could enhance agricultural yields (see e.g. Holman et al., 2017 for a European assessment; Hurtt et al., 2011 for a global assessment). Future changing environmental conditions such as modifications in atmospheric CO_2, nitrate deposition and climate, and natural disturbances such as wildfire, storm, bark beetle could also cause changes in land cover and in vegetation characteristics (e.g. leaf area index) (Seidl et al., 2014; Zhu et al., 2016).

The above review of the literature reveals that changes in climate characteristics (e.g. precipitation intensity and frequency) and in land cover properties are expected to have significant combined impacts on karst hydrology. Yet, the impact of preferential pathways on the partitioning between green and blue water and the effect of climate and land cover change has not been studied systematically. Determining how sensitive groundwater recharge is to climate and land cover may change in the presence of preferential pathways, and therefore to what extent findings obtained for non-karst areas may be

extrapolated to karst ones, is also essential to improve our understanding of future groundwater recharge at large-scales and ultimately to improve water resources management (Archfield et al., 2015). In this study we introduce a novel large-scale model that includes explicit representation of both karst and vegetation properties and systematically explore the sensitivity of its simulated recharge to climate and vegetation inputs. Our model builds on an existing karst hydrology model, called VarKarst, which was recently developed for large-scale applications and demonstrated over European and Mediterranean carbonate rock areas (Hartmann et al., 2015). However, VarKarst has a very simplistic representation of evapotranspiration and does not include land cover properties explicitly, which, up to now, prevented its application in land cover change impact studies.

The present study has two objectives that help us to overcome the previous limitation. First, we aim to add an explicit representation of land cover properties into VarKarst by improving the evapotranspiration (ET) estimation. While we seek to keep the model structure parsimonious, we want the new version of the model, called V2Karst (V1.0), to be appropriate for combined land cover and climate change impact studies. We test the plausibility of the V2Karst model behaviour by comparing its predictions against observations available at carbonate rock FLUXNET sites, and by analysing the dominant controls of simulated recharge. Second, we aim to understand the sensitivity of simulated groundwater recharge with V2Karst to changes in the vegetation characteristics and climate. We use a set of virtual experiments that allow us to control variations in climate and vegetation inputs, so that we can better explore their individual and combined effects on model outputs.

3.2 New version of VarKarst with explicit representation of land cover properties (V2Karst)

3.2.1 Challenges for modelling ET and representing land cover properties explicitly at large-scales

Representing explicitly land cover properties for ET estimation requires the specification of vegetation properties, such as leaf area index, vegetation height, stomatal resistance, canopy interception storage capacity, and the availability of time series of climate variables such as air temperature, net radiation, humidity and wind speed. Modelling ET at large-scales faces a range of challenges: (1) a lack of ET observations to compare with model simulations, (2) a lack of observations of vegetation properties, and (3) uncertainty in large-scale forcing weather variables.

Firstly, on the ground, measurements of actual ET (e.g. FLUXNET network, Baldocchi et al., 2001) are limited in number and are only representative of plot scale ET. Their footprint can extend to a few hundred metres or possibly to a few kilometres (Baldocchi and Ryu, 2011), which is much smaller than

the extent of typical large-scale model simulation units that are mostly between 9 km (5' grid) and 111 km (1° grid) (Bierkens, 2015). Moreover, ground measurements of the partitioning of ET among its main components (transpiration, evaporation from interception and soil evaporation) are lacking as reported in Miralles et al. (2016), and the ET partitioning assessed using isotope techniques has large uncertainties and limited spatial coverage (Coenders-Gerrits et al., 2014; Sutanto et al., 2014). Additionally, global gridded ET products are available. Yet, these products do not provide direct observations of actual ET, but they are estimates of actual ET assessed using models that assimilate remote-sensed variables and either solve the energy balance or use potential ET (PET) equations as discussed for instance in McCabe et al. (2016) and Miralles et al. (2016). Additionally, Jung et al. (2011) created a global gridded ET products based on model tree ensembles which are trained using observations from the FLUXNET network.

A second issue is that observations of large-scale vegetation properties are limited. Large-scale gridded land cover databases provide spatially distributed information about the type of vegetation present around the world. We refer to Smith (2016) for a review of land cover databases. Large-scale gridded measurements of vegetation characteristics are obtained using remote-sensing techniques, that permit to retrieve vegetation leaf area index (LAI) as discussed for instance in Fang et al. (2013) and other vegetation indices that can be only used as proxy for actual vegetation properties such as density or state of health, Vegetation Optical Depth (VOD), Normalized Difference Vegetation Index (NDVI) or Enhanced Vegetation Index (EVI) (see a review in Xue and Su, 2017). Moreover, such products suffer from a number of uncertainties, among which cloud contamination as reported for instance in Fang et al. (2013) regarding LAI, and do not allow to assess critical vegetation properties such as rooting depth, stomatal resistance or canopy interception capacity. Ground measurements of vegetation properties are sparse and only few studies report collected values for specific variables or regions, these include Breuer et al. (2003) for a range of vegetation properties in temperate climates, Körner (1995) for stomatal resistance and Schenk and Jackson (2002) for rooting depth. Since ground measurements are limited, they do not allow to capture the variability in vegetation characteristics, as discussed in Wang-Erlandsson et al. (2016) regarding rooting depth measurements. In particular, stomatal resistance presents a high temporal variability because it is determined by weather conditions and therefore its measurements are particularly difficult to interpret (2003) and to use in modelling applications. Ground measurements of canopy interception capacity are also sparse and uncertain (Van Dijk et al., 2015).

Thirdly, large-scale databases of historical weather data used to force model simulations are affected by large uncertainties because they have to rely on measurements with incomplete spatial coverage, in particular wind speed measurements (New et al., 2002). Moreover, the height from the ground at which these weather data are provided is not well characterised. Measurements are assumed to be provided at

standard heights, typically 10 m for wind speed and 2 m for temperature and humidity (see e.g. Rodell et al., 2004; Weedon et al., 2010), which may not be representative of the specific location.

3.2.2 Rationale to explicitly represent land cover properties into VarKarst

The new version of the VarKarst model should be appropriate to assess the impact of climate and land cover change on karst groundwater recharge, and should consider the range of challenges related to modelling ET at large-scales detailed in Sect. 3.2.1. According to that, we define the three following criteria to represent ET in the VarKarst model:

1. **The model should assess separately all three main ET components (bare soil evaporation in presence of sparse canopy, transpiration and evaporation from canopy interception).** In fact, these fluxes exhibit different dynamics and sensitivity to environmental conditions and therefore, they are likely to respond differently to climate and land cover changes (Savenije, 2004; Gerrits, 2010; Wang and Dickinson, 2012; Maxwell and Condon, 2016).
2. **The model should use Penman-Monteith formulation for potential evapotranspiration (PET)** (Monteith, 1965), to separate the effects of climate and land cover and assess specific rates for the different ET components. In fact, empirical PET formulations such as the Priestley-Taylor equation (Priestley and Taylor, 1972) do not represent explicitly land cover properties.
3. **All processes should be represented parsimoniously** in accordance with the modelling philosophy underpinning the first version of VarKarst (Hartmann et al., 2015). This criterion aims to avoid over-parameterisation given the limited amount of available information to constrain and test model simulations in particular at large-scales (Abramowitz et al., 2008; Beven and Cloke, 2012; Haughton et al., 2018; Hong et al., 2017; IPCC, 2013, Chapter 9, pp. 790-791; Young et al., 1996). In particular, parameters that account for physical properties of the system (e.g. soil and vegetation properties) are commonly taken from look-up tables but their physical meaning have been put into question and they may actually not be commensurate with field measurements as discussed in Hogue et al. (2006) and in Rosero et al. (2010). Therefore, it has been suggested that even these 'physical' parameters should be calibrated so to optimise the model performance (Chaney et al., 2016; Rosolem et al., 2013). Importantly, parsimony limits the computational time for model simulations and allows for assessing the impact of modelling choices and the uncertainty and sensitivity of model output using Monte-Carlo simulation (Hong et al., 2017; Young et al., 1996).

With respect to existing models of karst areas, to our knowledge, only four models that explicitly include ET and land cover processes were applied in karst studies, all of which were local studies with detailed on-site information. Three of these models (Canora et al., 2008; Doummar et al., 2012; Zhang et al.,

2011) were not specifically developed for karst areas, but they are distributed models that simply utilised the flexibility of their parameters to represent the variability in soil and bedrock properties. These models are heavily parameterised, which hampers their application at large-scales, and does not comply with criterion 3 (parsimony). The fourth model introduced in (Sauter, 1992) is lumped and is much more parsimonious than the three other models. However, the model does not represent soil evaporation, and uses empirical PET equations, which does not allow to separate the effect of climate and land cover (disagreement with criteria 1 and 2).

As for large-scale models, we can identify two main types: hydrological models, which focus on the assessment of hydrological fluxes, and land surface models, which also evaluate many other fluxes such as sensible heat, latent heat, ground heat flux, radiation and carbon fluxes (for a review, see Bierkens, 2015). Land surface models do not usually comply with criterion 3 because they have many parameters, including a number of empirical parameters that are difficult to constrain (Mendoza et al., 2015a). Moreover, it has been shown that land surface models could be simplified when the objective is to assess hydrological fluxes only. For instance, Cuntz et al. (2016) demonstrated that a large number of parameters of the Noah land surface model are non-influential or have very little influence on simulated runoff. In contrast, hydrological models focus on the representation of hydrological processes and include far fewer parameters. However, our literature review (summarised in Tables 3.1-3.4) showed that we cannot directly adopt any of their ET representation into VarKarst. In fact, as shown in Tables 3.1-3.4, the most parsimonious models (WBM, WaterGap and mHM) neglect some ET components and/or use empirical PET equations, which contradicts criteria 1 and 2, while models that comply with criteria 1 and 2 (PCR-GLOBWB, VIC and the model of Kergoat (1998)) use heavily parameterised schemes, such as a Jarvis type parameterisation of surface resistance (Jarvis, 1976; Stewart, 1988) and therefore do not satisfy criterion 3 (parsimony). Moreover, we found that large-scale models include empirical schemes with no clear origin, such as the reference crop formulation used in the PCR-GLOBWB model for PET calculation or the interception model used in LPJ and in the model of Kergoat (1998). Importantly, our review revealed the tremendous variability of approaches used in large-scale models to represent ET processes. A detailed list of all parameters involved in the representation of ET the models of Tables 3.1-3.4 can be found in Appendix B.1. Consequently, no clear indication emerged regarding a ‘best way’ to parameterise the different ET processes at large-scales, which leaves us with a large range of different formulations to choose from to implement an explicit representation of land cover processes into VarKarst.

Model		Energy balance	ET processes					Number of parameters for ET estimation [a]	Reference
							Carbon cycle		
WBM	daily	no	**yes**	no	no	no	no	3 (minimum)	(Federer et al., 2003; Vörösmarty et al., 1989; Vörösmarty et al., 1998)
WaterGap V2.2	daily	no	**yes**	no	no	**yes**	no	7	(Döll et al., 2003; Müller Schmied et al., 2014)
mHM	daily/ sub-daily	no	**yes**	no	no	**yes**	no	10 [b]	(Kumar et al., 2013; Samaniego et al., 2010, 2018)
LPJ	daily	no	**yes**	no	**yes**	**yes**	**yes**	14	(Gerten et al., 2004; Sitch et al., 2003)
Model of (Kergoat, 1998)	daily	no	**yes**	no	**yes**	**yes**	no	15	(Kergoat, 1998)
PCR-GLOBWB	daily	no	**yes**	no	**yes**	**yes**	no	15	(Van Beek and Bierkens, 2008; Van Beek, 2008; Sperna Weiland et al., 2015; Sutanudjaja et al., 2011)
Mac-PDM	daily	no	**yes**	**yes**	no	**yes**	no	16 [c]	(Arnell, 1999; Gosling and Arnell, 2011; Smith, 2016)
GLEAM V3	daily	no	**yes**	no	**yes**	**tall land cover**	no	18 [d]	(Martens et al., 2017; Miralles et al., 2010, 2011)
VIC V4.2	daily/ sub-daily	**optional**	**yes**	no	**yes**	**yes**	no	22	(Bohn and Vivoni, 2016; Liang et al., 1994)

Table 3.1 Characteristics of selected large-scale hydrological models: simulation time step (), solving of the energy balance, ET processes represented, number of parameters for ET estimation and references.

ET processes are: overstory transpiration, understory transpiration , soil evaporation , evaporation from canopy interception , and carbon cycle i.e. vegetation dynamic model.

[a] Number of parameters for a given land cover type, excluding parameters used in the representation of vegetation seasonality, carbon cycle (vegetation dynamic), sublimation from snowpack and snowmelt evaporation to make models more comparable.

[b] Number of parameters considering three soil layers.

[c] This number includes the parameters used for the computation of both understory and overstory (grass) transpiration.

[d] Number of parameters assuming tall vegetation (interception is considered for tall vegetation only).

Model	Potential evapotranspiration (PET)			Seasonality of vegetation
	Formulation [a]	Surface resistance	Number of parameters	
WBM	T, SW	constant when considered	0 (minimum)	not included
WaterGap V2.2	PT	not included	1	empirical growth model
mHM	HS	Not included	1 (aspect correction)	monthy values of canopy interception capacity calculated from monthly LAI
LPJ	empirical formula based on PT	function of CO_2 and photosynthesis	4	vegetation dynamic model
Model of (Kergoat, 1998)	PM	Jarvis type (Jarvis, 1976; Stewart, 1988)	10	set to zero during leaf-off season
PCR-GLOBWB	PM	empirical reference crop scheme (Allen et al., 1998) [b]	2	monthly values of crop factors and
Mac-PDM	PM	constant	8 [c]	not included
GLEAM V3	PT	not included	3 [d]	assimilation of VOD
VIC V4.2	PM	Jarvis type (Jarvis, 1976; Stewart, 1988)	12	monthly values of and assimilation of daily NDVI

Table 3.2 Representation of potential evapotranspiration (PET) and seasonality of vegetation in the large-scale hydrological models of Table 3.1.

[a] T: Thornthwaite (Thornthwaite, 1948); HS: Hargreaves and Samani (1985); PT: Priestley-Taylor (Priestley and Taylor, 1972); PM: Penman-Monteith (Monteith, 1965); SW: Shuttleworth-Wallace (Shuttleworth and Wallace, 1985).

[b] This approach consists of calculating a value of PET for a reference grass surface with known properties and to adjust this potential rate using land cover specific empirical crop factors. This formulation avoids the specification of the stomatal resistance whose value is largely uncertain (see Appendix B.1). Tabulated values of the crop factors for agricultural crops are provided in (Allen et al., 1998). However, the origin of the crop factor formulation for non-agricultural crops is not clear.

[c] This number includes the parameters used for the computation of PET for both understory and overstory (grass).

[d] Number of parameters assuming tall vegetation (interception is considered for tall vegetation only).

Model	Stress model for actual ET calculation from PET	Evaporation from canopy interception (	
		model	Number of parameters
WBM	function of soil moisture which multiplies PET	not included	0
WaterGap V2.2	demand-supply model (Federer, 1982)	overflow store	3
mHM	function of soil moisture which multiplies PET	overflow store	2
LPJ	demand-supply model for transpiration (Federer, 1982) and function of soil moisture which multiplies PET for soil evaporation	empirical: fraction of precipitation (Kergoat, 1998)	2
Model of (Kergoat, 1998)	function of soil moisture which multiplies	empirical: fraction of precipitation (Kergoat, 1998)	2
PCR-GLOBWB	function of soil moisture and soil hydraulic properties which multiplies PET	overflow store	2
Mac-PDM	function of soil moisture which multiplies PET	Calder (Calder, 1990)	3 [a]
GLEAM V3	function of soil moisture and vegetation optical depth which multiplies PET	Gash (Gash, 1979; Valente et al., 1997)	7 [b]
VIC V4.2	function of soil moisture which multiplies for transpiration and PET for soil evaporation	overflow store	3

Table 3.3 Representation of stress model for actual evapotranspiration (ET) calculation from Potential Evapotranspiration (PET) and evaporation from canopy interception for the large-scale models of Table 3.1.

[a] This number includes the parameters used for the computation of PET for both understory and overstory (grass).

[b] Number of parameters assuming tall vegetation (interception is considered for tall vegetation only).

Model	Sub-grid variability of soil moisture [a]	Soil layers [b]	Sparse vegetation formulation [c]
WBM	no	1 layer	not included
WaterGap V2.2	**implicit**	1 layer	not included
mHM	**implicit**	3 layers (from all layers depending on their relative root fractions)	not included
LPJ	no	3 layers (from shallow layer and from all layers depending on their relative root fractions)	uncoupled (vegetated and bare soil tiles)
Model of (Kergoat, 1998)	no	1 layer	coupled moisture uptake
PCR-GLOBWB	**implicit**	2 layers (from shallow layer and from all layers depending on their relative root fractions)	coupled moisture uptake
Mac-PDM	**implicit**	1 layer for each tile	uncoupled (overstory and understory tiles)
GLEAM V3	no	3 layers (from shallower layer and in wettest layer)	uncoupled (vegetated and bare soil tiles)
VIC V4.2	**implicit**	2 layers (from shallower layer and from all layers depending on their relative root fractions)	coupled moisture uptake

Table 3.4 Representation of sub-grid variability of soil moisture, soil layers and sparse vegetation in the large-scale models of Table 3.1.

[a] None of these models account for karst processes as done by the VarKarst model (Hartmann et al., 2015).

[b] : actual soil evaporation; : actual vegetation transpiration.

[c] Uncoupled or 'tile' approaches consist of assessing separately the water balance for the vegetated and bare soil fractions (overstory and understory fractions for Mac-PDM). Therefore, this approach is based on the simplifying assumption that the vegetation roots do not extend beyond the surface area covered by the vegetation canopy. Instead, coupled approaches evaluate the overall water balance over both fraction, thus allowing for interactions for soil moisture uptake between vegetated and bare soil fractions. All models neglect aerodynamic interactions between vegetation and bare soil. This can be accounted for using for instance the Shuttleworth-Wallace PET equation (Shuttleworth and Wallace, 1985), which requires the specification of further resistance parameters compared to the Penman-Monteith equation. The Shuttleworth-Wallace equation was used anecdotally in the WBM model for a few applications.

We note that, in this research, we do not consider the carbon cycle and the effect of atmospheric CO_2 concentration on the vegetation. These processes are included for instance in land surface models (Bierkens, 2015) and in the LPJ model (Sitch et al., 2003; Gerten et al., 2004) but are neglected in other large-scale hydrological models of Tables 3.1-3.4. The impact of atmospheric CO_2 concentration on ET is not clear, since it is understood that a rising CO_2 concentration has two counteracting effect on transpiration losses (see e.g. Gerten et al., 2004; Döll et al., 2016; Ukkola et al., 2016; Lemordant et al., 2018). In fact, as discussed in the previously cited papers, a rising CO_2 concentration not only has a fertilisation effect (which enhances photosynthesis and plant growth and therefore enhances transpiration) but also leads to an increase in stomatal resistance (which in turn increases water use efficiency by the plants and therefore reduces transpiration). Moreover, the representation of the carbon cycle and of the effect of CO_2 on plants in models is difficult to constrain given the limited amount of

information available, especially at large-scales as explained in our criterion 3 for model development (parsimony). For instance, the value of the stomatal resistance and its dependence on CO_2 concentration is poorly characterised, because few ground measurements of stomatal resistance are available, as discussed in Sect. 3.2.1. Consequently, different representations are used in different models to account for these processes, and different models tend to produce different results. In particular, the model inter-comparison study by Davie et al. (2013) showed that models disagree on the effect of CO_2 concentration on runoff. Given the limited state of knowledge on the effect of atmospheric CO_2 concentration on hydrological fluxes and the large uncertainties related to the representation of the effect of CO_2 in models, we did not consider CO_2 in this study. Such in-depth investigation is beyond the scope of this research.

The next sections provide more details on the specific assumptions and choices made to develop the new ET component for the VarKarst model, which satisfies the three criteria defined in this section and utilises some of the schemes from other large-scale models.

3.2.3 Previous representation of ET processes in VarKarst

VarKarst (Hartmann et al., 2015) is currently the only karst recharge model developed for large-scale applications. It is a conceptual semi-distributed model that simulates daily karst potential recharge (Fig. 3.1.a). VarKarst includes two horizontal subsurface layers, a top layer called ‘soil’ and a deeper layer called ‘epikarst’. The soil layer corresponds to the layer from which ET can occur. The epikarst layer corresponds to the uppermost layer of weathered carbonate rocks where it is assumed that water cannot be lost through ET. Groundwater recharge predicted by VarKarst includes both the diffuse and concentration fractions, because for each model grid cell, the water balance is evaluated separately over a number of vertical compartments with varying soil and epikarst properties. The ET component of the VarKarst model is very simple and does not include explicit representation of land cover properties. ET is lumped in the soil layer and is estimated from PET and reduced by a water stress factor, which is estimated as a linear function of soil moisture. The PET rate is calculated with the empirical Priestley-Taylor equation (Priestley and Taylor, 1972) using a spatially and temporally uniform value of the empirical coefficient. This approach does not allow to separate the effect of climate and land cover, since the empirical coefficient reflects both climate and vegetation characteristics simultaneously. Therefore, the ET component of VarKarst needs to be modified if the model is to be used for large-scale land cover change impact assessment.

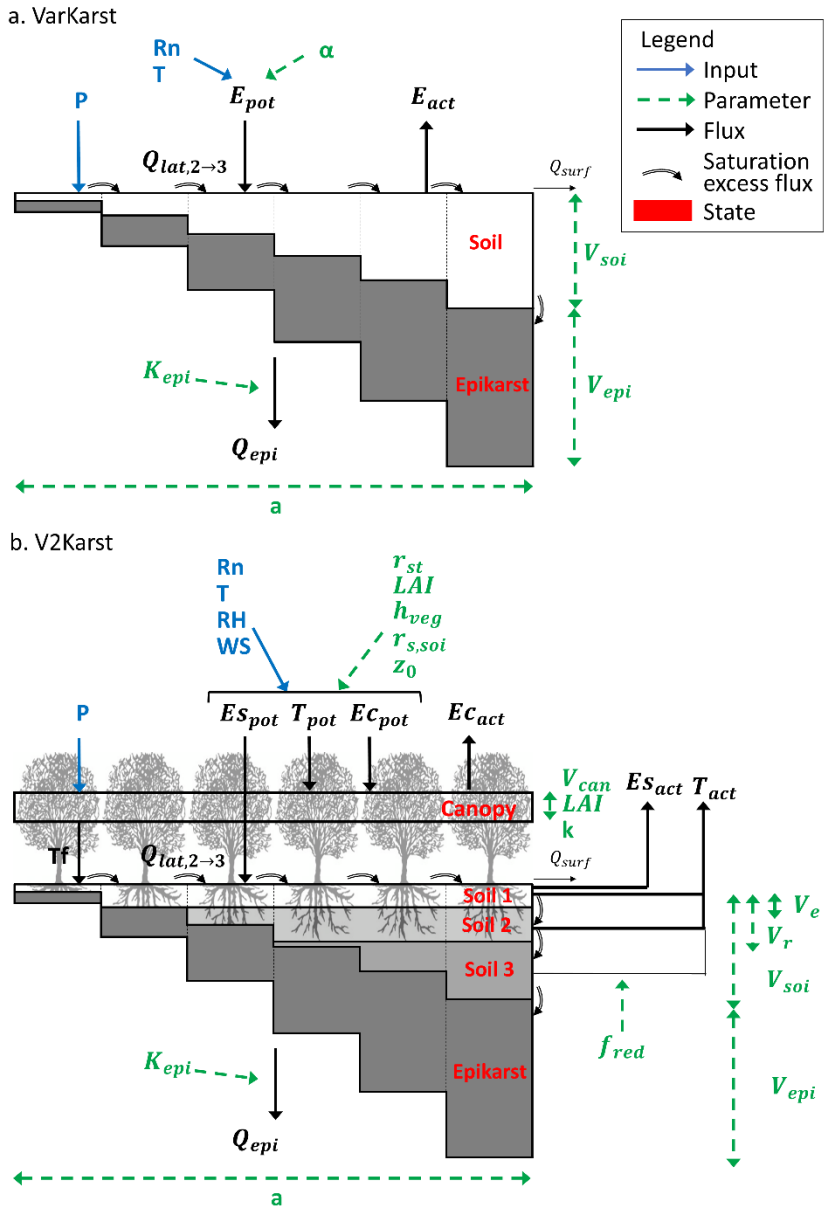


Figure 3.1 Schematic representation of (a) the VarKarst model (Hartmann et al., 2015) and (b) the new version of the model V2Karst using six vertical compartments.

Model parameters are in green (see their definition in Table 3.5**),** inputs are in blue (P precipitation, R_n net radiation, T temperature, RH relative humidity, WS wind speed), model fluxes are in black (E_{pot} potential total evapotranspiration, T_{pot} potential transpiration, Ec_{pot} potential evaporation from canopy interception, Es_{pot} potential soil evaporation, E_{act} total actual ET, T_{act} actual transpiration, Ec_{act} actual evaporation from canopy interception, Es_{act} actual bare soil evaporation, Tf throughfall, $Q_{lat,2\to3}$ lateral flow from the second to the third compartment, Q_{surf} surface runoff and Q_{epi} recharge) and state variables are in red.

Parameter	Description	unit	Lower limit	Upper limit	Category
h_{veg}	Vegetation height	[m]	0.2	Site specific	vegetation
r_{st}	Stomatal resistance	[s.m^{-1}]	20	600	vegetation
LAI_{min}	Reduction in leaf area index during the dormant season	[%]	5	100	vegetation
LAI_{max}	Annual maximum leaf area index	[m^2.m^{-2}]	0.5	8	vegetation
V_r	Maximum storage capacity of the root zone	[mm]	20	500	vegetation
V_{can}	Canopy storage capacity per unit of LAI	[mm LAI]	0.1	0.5	vegetation
k	Beer-Lambert's law extinction coefficient	[-]	0.4	0.7	vegetation
f_{red}	Reduction factor for transpiration below the root zone	[-]	0	0.15	soil
z_0	Soil roughness length	[m]	0.0003	0.013	soil
$r_{s,soi}$	Soil surface resistance	[s.m^{-1}]	0	100	soil
V_e	Maximum storage capacity of the first soil layer	[mm]	5	45	soil
a	Spatial variability coefficient	[-]	0	6	soil and epikarst
V_{soil}	Mean soil storage capacity	[mm]	20	800	soil
V_{epi}	Mean epikarst storage capacity	[mm]	200	700	epikarst
K_{epi}	Mean epikarst outflow coefficient	[d]	0	50	epikarst

Table 3.5 Description of V2Karst parameters, unconstrained ranges used in the application at the four FLUXNET sites to capture the variability across soil, epikarst and vegetation types, category of the parameters (which indicated whether the parameters depend on soil, epikarst or vegetation properties).

Parameters a, V_{soil}, V_{epi} and K_{epi} were already present in the previous version of the model (VarKarst). More information on how the ranges were determined is provided in Appendix B.2

3.2.4 V2Karst: the new version of VarKarst for integrated vegetation-recharge simulations over karst areas

In this section, we propose a new version of the VarKarst model, called V2Karst (Figure 3.1b). In accordance with the criteria 1 and 2 defined in Sect. 3.2.2, compared to VarKarst, the new V2Karst model (1) includes a physically based PET equation, (2) separates the evapotranspiration flux into three components (transpiration, bare soil evaporation and evaporation from canopy interception), (3) comprises three soil layers. Additionally, V2Karst represents parsimoniously the seasonal changes in the vegetation properties, which will allow us to analyse the importance of this process on simulated recharge. We assumed homogeneous above ground vegetation properties across model compartments.

We note that V2Karst has a total of 15 parameters (described in Table 3.5 and Figure 3.1.b), including the 4 parameters of VarKarst and 11 new parameters in the new ET component, that replaces the Priestley-Taylor empirical coefficient α used in VarKarst. In agreement with criteria 3 of Sect. 3.2.2 (parsimony), we sought to represent parsimoniously the different ET processes into VarKarst. In fact, V2Karst uses 12 parameters to represent ET and vegetation seasonality (including the 11 newly introduced parameters and the soil water capacity parameters V_{soi} already present in VarKarst). This is less than other existing large-scale models that use Penman-Monteith equation and separate the three ET components, since these models have over 15 parameters in their ET component (PCR-GLOBWB, VIC and the model of Kergoat, 1998, see Table 3.1) The new model is forced by time series of precipitation P, air temperature T and net radiation R_n as VarKarst. Additionally, time series of relative humidity RH and wind speed WS are now needed for PET calculation.

As VarKarst, V2Karst does not include a snow routine. Developing a snow routine for the V2Karst model is beyond the scope of this study, which aims to include an explicit representation of land cover processes into the VarKarst model. In this chapter, we neglect the effect of snow, which seems reasonable because snow is not a first order control at the four study sites that are examined (as discussed in Sect. 3.3.1). Future applications of V2Karst can account for the effect of snow using inputs of snow water equivalent as in previous done with the VarKarst model. In fact, Hartmann et al. (2015) used snow water equivalent from the GLDAS dataset (Rodell et al., 2004) to simulate VarKarst over the historical period and Hartmann et al. (2017) derived snow water equivalent from the ISI-MIP dataset (Warszawski et al., 2014) to simulate VarKarst over future periods. The effect of snow on hydrological fluxes and the representation of snow in future large-scale applications of the V2Karst are further discussed in Sect. 3.6.3.

Definition of soil and epikarst properties in V2Karst

The computation of water storage capacity of the entire soil column $V_{S,i}$ [mm] and of the epikarst $V_{E,i}$ [mm], and the epikarst outflow coefficient $K_{E,i}$ [d] for the ith model compartment is done as before in VarKarst:

$$V_{S,i} = V_{max,S}\left(\frac{i}{n_c}\right)^a,$$

$$V_{E,i} = V_{max,E}\left(\frac{i}{n_c}\right)^a, \tag{4.1}$$

$$K_{E,i} = K_{max,E}\left(\frac{n_c - i + 1}{n_c}\right)^a.$$

where $V_{max,S}$ [mm] is the maximum soil storage capacity over all model compartments, $V_{max,E}$ [mm] is the maximum epikarst storage capacity, $K_{max,E}$ [d] is the maximum outflow coefficient, n_c [−] is the number of model compartments, which is set to 15 following (Hartmann et al., 2013a, 2015) and a [−] is the spatial variability coefficient. A previous study showed that $V_{S,i}$, $V_{E,i}$ and $K_{E,i}$ can be determined using the same distribution coefficient a (Hartmann et al., 2013a). In V2Karst, $V_{max,S}$, $V_{max,E}$ and $K_{E,i}$ are computed as a function of the average properties of the cell using the following the formulas:

$$V_{max,S} = \frac{V_{soi} n_c}{\sum_{i=1}^{n} \left(\frac{i}{n_c}\right)^a},$$

$$V_{max,E} = \frac{V_{epi} n_c}{\sum_{i=1}^{n} \left(\frac{i}{n_c}\right)^a}, \tag{4.2}$$

$$K_{max,E} = \frac{K_{epi} n_c}{\sum_{i=1}^{n} \left(\frac{i}{n_c}\right)^a},$$

where V_{soi} [mm] is the mean soil storage capacity, V_{epi} [mm] is the mean epikarst storage capacity and K_{epi} [mm] is the mean epikarst outflow coefficient. We note that the definition of the three parameters V_{soi}, V_{epi} and K_{epi} is revised compared to VarKarst.

As in VarKarst, we neglect ET from the epikarst. Several studies showed that in presence of shallow soil and dry climate, plants can take up water in the weathered bedrock where soil pockets can sustain roots development (Schwinning, 2010). However, given the uncertainty in soil depth for large-scale applications, V2Kast does not allow ET from the epikarst to avoid over-parameterisation. Therefore, the V2Karst soil layer must be interpreted as a conceptual layer that does not exactly correspond to the physical soil layer (layer of loose material) but is defined as the portion of the subsurface where ET losses can occur.

In V2Karst, the soil layer is further divided into a shallow top layer from which water can be lost from both evaporation and transpiration, a second middle layer where only transpiration can occur and a third deeper layer below the root zone where transpiration can only take place when the first two layers are depleted. The maximum storage capacity of the first layer is noted as V_e [mm], and the maximum storage capacity of first and second layers combined is noted as V_r [mm],which corresponds to the maximum storage capacity of the root zone. The model assumes that V_e is smaller than V_r, which is in turn smaller than the storage capacity of the deeper model compartment $V_{S,n}$.

Soil water balance

The soil water storage $V_{soi,i}^{j}(t)$ [mm] in the ith compartment and the jth soil layer $j = 1,2,3$ is updated at the end of each time step t as follows:

$$\begin{aligned} V_{soi,i}^{1}(t) &= V_{soi,i}^{1}(t-1) + T_f(t) + Q_{lat,i-1\to i}(t) - Es_{act,i}(t) - T_{act,i}^{1}(t) - R_{12,i}(t), \\ V_{soi,i}^{2}(t) &= V_{soi,i}^{2}(t-1) + R_{12,i}(t) - T_{act,i}^{2}(t) - R_{23,i}(t), \\ V_{soi,i}^{3}(t) &= V_{soi,i}^{3}(t-1) + R_{23,i}(t) - T_{act,i}^{3}(t) - R_{epi,i}(t), \end{aligned} \tag{4.3}$$

where $T_f(t)$ [mm] is the throughfall i.e. the fraction of precipitation that is not evaporated from the interception store, $Q_{lat,i-1\to i}(t)$ [mm] is the lateral flow from the $(i-1)$th to the ith model compartment (Eq. (4.15)), $Es_{act,i}(t)$ [mm] is the actual soil evaporation (Eq. (4.7)), $T_{act,i}^{j}(t)$ [mm] is the actual transpiration in the jth soil layer (Eq. (4.9-10)), $R_{12,i}(t)$ [mm] is the downward flow from the first to the second soil layer, $R_{23,i}(t)$ [mm] is the downward flow from the second to the third soil layer and $R_{epi,i}(t)$ [mm] is the downward flow from the soil to the epikarst.

It is assumed that percolation from the unsaturated soil to the epikarst is negligible due to low permeability of the soil. This assumption seems reasonable since karst soils usually have a high clay content (Clapp and Hornberger, 1978; Blume et al., 2010). However, clayey soil typically present cracks (Lu et al., 2016), and therefore when the soil reaches saturation, preferential flow starts to occur in the soil cracks, which causes all saturation excess to quickly infiltrate to the epikarst. Just as in VarKarst, such preferential vertical flow is represented by the variable $R_{epi,i}(t)$ (used in Eq. (4.3)) and is set equal to the saturation excess in the (lowest) soil layer. In V2Karst, a similar approach is also used to assess the other vertical flows from one soil layer to another ($R_{12,i}(t)$ and $R_{23,i}(t)$) in Eq. (4.3)).

Vegetated and non-vegetated fractions

We adopt the representation of sparse vegetation proposed by (Bohn and Vivoni, 2016) for the VIC model and referred to as 'clumped' vegetation scheme. Each model compartment is divided into a vegetated and a non-vegetated fraction using a canopy cover fraction coefficient $f_c(t)$ $[-]$. The uptake of soil moisture for transpiration and soil evaporation is coupled in a way that, for each model compartment, we evaluate an overall water balance over the two fractions. Using such a coupled approach facilitates the representation of the seasonal variations in vegetated and non-vegetated fractions compared to an uncoupled 'tile' approach, in which a separate soil moisture state is represented for vegetated and bare soil fractions. Consistently with other existing large-scale models, aerodynamic interactions between both fractions are neglected to keep the number of parameters to a minimum (Table 3.2).

The canopy coefficient $f_c(t)$ is estimated in V2Karst using the Beer-Lambert's law as in (Van Dijk and Bruijnzeel, 2001; Ruiz et al., 2010). This law has been originally used to separate the fraction of incident radiation (and by extension of net radiation) absorbed by the canopy from the fraction penetrating the canopy (Ross, 1975; Shuttleworth and Wallace, 1985; Kergoat, 1998). The canopy cover fraction at time t is expressed as a function of the cell average leaf area index $LAI(t)$ [$\mathrm{m}^2.\mathrm{m}^{-2}$] and an extinction coefficient k [$-$], which is understood to vary across vegetation type since it accounts for leaf architecture (Ross, 1975):

$$f_c(t) = 1 - e^{-kLAI(t)}. \tag{4.4}$$

Notice that Eq. (4.4) allows to describe the seasonal variations in canopy cover fraction without introducing additional parameters in the model, given that they will simply follow the seasonal variations in LAI.

Canopy interception

It has been shown that a simple parameterization of daily interception can give reasonable simulation results (Savenije, 1997; De Groen, 2002; Gerrits, 2010). Following these studies, in V2Karst, interception is represented by a daily threshold model. Our formulation is as follows:

$$Ec_{act}(t) = f_c(t) min\left(Ec_{pot}(t), P(t), V_{can,max}(t)\right), \tag{4.5}$$

where $Ec_{pot}(t)$ [mm] is the potential evaporation from canopy interception (Eq. (4.12)), $P(t)$ [mm] is the precipitation and $V_{can,max}(t)$ [mm] is the interception storage capacity over the vegetated fraction of the cell (Eq. (4.6)). The factor $f_c(t)$ in Eq. (4.5) accounts for the fact that evaporation from canopy occurs over the vegetated fraction only. We note that the potential rate $Ec_{pot}(t)$ was not accounted for in the original formulation by (Savenije, 1997; De Groen, 2002; Gerrits, 2010).. The interception storage capacity over the vegetated fraction $V_{can,max}$ [mm] depends (1) on the leaf area index over the vegetated fraction, which is estimated by rescaling cell average leaf area index $LAI(t)$ using the vegetation cover fraction f_c(t) following Bohn and Vivoni (2016), and (2) on the canopy storage capacity per unit of leaf area index, denoted by V_{can}, which is understood to depend on the vegetation type since it accounts for leaf architecture (Gerrits, 2010). It is expressed as:

$$V_{can,max}(\mathrm{t}) = V_{can}\left(\frac{\mathrm{LAI(t)}}{f_c(\mathrm{t})}\right). \tag{4.6}$$

The model does not account for the carry-over of interception storage from one day to the next, which means that all precipitation which is not evaporated from the interception store reaches the ground as throughfall T_f [mm]. This assumption can be justified by the fact that the interception process is highly

dynamic at a sub-daily time scale, because the canopy can go through several wetting-drying cycles within a day (Gerrits, 2010). Therefore, when evaporation from canopy interception is estimated with a daily time step as in V2Karst, the canopy layer must be interpreted as a conceptual layer, whose storage capacity does not exactly correspond to the physical storage capacity of the canopy (i.e. the amount of water that can be hold at a given time), but to the cumulative amount of water that can be hold by the canopy over a day (Gerrits, 2010).

Bare soil evaporation

It is assumed that soil evaporation is a faster process than transpiration consistently with general knowledge on ET processes (Wang and Dickinson, 2012). Therefore, soil moisture can be first evaporated and then transpired if some available moisture remains for plant water uptake. Soil evaporation is withdrawn for the first soil layer as a function of the potential rate and soil moisture, similar to the previous version of VarKarst:

$$Es_{act,i}(t) = \min\left(\left(1 - f_c(t)\right) Es_{pot}(t) \frac{V^1_{soi,i}(t-1)}{V^1_{S,i}}, V^1_{soi,i}(t-1) + T_f(t) \right), \quad (4.7)$$

where $Es_{pot}(t)$ is the potential soil evaporation (Eq. (4.12)). The factor $(1 - f_c(t))$ in Eq. (4.7) accounts for the fact that soil evaporation occurs from the non-vegetated fraction only and therefore the potential rate has to be weighted by the bare soil cover fraction. The right term of the equation ($V^1_{soi,i}(t-1) + T_f(t)$) is not weighted, because we assume that the soil moisture is uniform over the fractions of each model compartment (we compute a unique water balance) and therefore the total moisture present in the first soil layer is available to soil evaporation because the vegetated fraction can supply moisture to the bare soil fraction.

Transpiration over the vegetated fraction

Transpiration mainly occurs in the first and second soil layers, and it switches to the third soil layer when the first two layers are depleted. The extraction of water by the roots below the root zone is documented in (Penman, 1950) and we account for this process by representing a soil layer below the root zone, which can provide water to the root zone through capillary rise as in the ISBA model (Boone et al., 1999). In V2Karst, the rate at which transpiration occurs in the two first soil layers $T^{12}_{rate,i}(t)$ [mm] and in the third soil layer $T^{3}_{rate,i}(t)$ [mm] are assessed as follows:

$$T_{rate,i}^{12}(t) = \big(1 - t_{wet}(t)\big) f_c(t) T_{pot}(t) \frac{V_{soi,i}^{1}(t-1) + V_{soi,i}^{2}(t-1)}{V_{S,i}^{1} + V_{S,i}^{2}},$$
$$T_{rate,i}^{3}(t) = \big(1 - t_{wet}(t)\big) f_c(t) T_{pot}(t) \frac{V_{soi,i}^{3}(t-1)}{V_{S,i}^{3}} f_{red}. \quad (4.8)$$

Where $T_{pot}(t)$ is the potential transpiration (Eq. (4.12)), $t_{wet}(t)$ $[-]$ is the fraction of the day with wet canopy (Eq. (4.11)) and f_{red} $[-]$ is a reduction factor which accounts for the fact that moisture below the root zone is less easily accessible to the roots than moisture in the root zone (Penman, 1950), and which is expected to vary across soil type since it is linked to the soil capability to supply water to the root zone . It is assumed that transpiration occurs in the two first soil layers when $T_{rate,i}^{12}(t)$ is higher than $T_{rate,i}^{3}(t)$, and that transpiration is drawn from the third soil layer otherwise. The actual transpiration in the two first soil layers $T_{act,i}^{12}(t)$ $[mm]$ and in the third soil layer $T_{act,i}^{3}(t)$ [mm] are therefore calculated as follows:

when $T_{rate,i}^{12}(t) \geq T_{rate,i}^{3}(t)$:

$$\begin{cases} T_{act,i}^{12}(t) = \min\left(T_{rate,i}^{12}(t), V_{soi,i}^{1}(t-1) + V_{soi,i}^{2}(t-1) + T_f(t) - Es_{act,i}(t)\right), \\ T_{act,i}^{3}(t) = 0, \end{cases}$$

when $T_{rate,i}^{12}(t) < T_{rate,i}^{3}(t)$: (4.9)

$$\begin{cases} T_{act,i}^{12}(t) = 0, \\ T_{act,i}^{3}(t) = \min\left(T_{rate,i}^{3}(t),\ V_{soi,i}^{3}(t-1) + R_{23,i}(t)\right). \end{cases}$$

Actual transpiration in the upper two layers $T_{act,i}^{12}(t)$ is partitioned between the two soil layers within the root zone as is used in the PCR-GLOBWB model (Van Beek, 2008). In V2Karst, the transpiration is attributed to the two first soil layers proportional to their storage content. This simple representation assumes that the roots can equally access the moisture stored in the first and second layer. Actual transpiration from the first layer $T_{act,i}^{1}(t)$ [mm] and the second layer $T_{act,i}^{2}(t)$ [mm] are computed as follows:

$$T_{act,i}^{1}(t) = \frac{V_{soi,i}^{1}(t-1) + T_f(t) - Es_{act,i}(t)}{V_{soi,i}^{1}(t-1) + T_f(t) - Es_{act,i}(t) + V_{soi,i}^{2}(t-1)} T_{act,i}^{12}(t),$$
$$T_{act,i}^{2}(t) = \frac{V_{soi,i}^{2}(t-1)}{V_{soi,i}^{1}(t-1) + T_f(t) - Es_{act,i}(t) + V_{soi,i}^{2}(t-1)} T_{act,i}^{12}(t). \quad (4.10)$$

The fraction of the day with wet canopy $t_{wet}(t)$ $[-]$ is estimated as the fraction of available energy that was used to evaporate water from the interception store similar to Kergoat (1998):

$$t_{wet}(t) = \frac{Ec_{act}(t)}{f_c(t)Ec_{pot}(t)} \tag{4.11}$$

Potential evapotranspiration

We replace the Priestley-Taylor potential evaporation equation used in the previous version of the model by the Penman-Monteith equation (Monteith, 1965). Potential transpiration rate over the vegetated fraction of the cell $T_{pot}(t)$ [mm] is estimated from the canopy aerodynamic resistance $r_{a,can}$(t) [s. m^{-1}] and surface resistance $r_{s,can}(t)$ [s. m^{-1}], potential evaporation from interception over the vegetated fraction of the cell $Ec_{pot}(t)$ [mm] is assessed assuming that the surface resistance is equal to 0 following e.g. (Shuttleworth, 1993), while potential bare soil evaporation rate over the bare soil fraction of the cell $Es_{pot}(t)$ [mm] is calculated from the soil aerodynamic resistance $r_{a,soi}(t)$ [s. m^{-1}] and surface resistance $r_{s,soi}$ [s. m^{-1}], using the following equations:

$$T_{pot}(t) = \frac{\Delta(t)R_n(t) + K_t\rho_a(t)c_p\dfrac{e_s(t) - e_a(t)}{r_{a,can}(t)}}{\lambda(t)\left(\Delta(t) + \gamma(t)\left(1 + \dfrac{r_{s,can}(t)}{r_{a,can}(t)}\right)\right)},$$

$$Ec_{pot}(t) = \frac{\Delta(t)R_n(t) + K_t\rho_a(t)c_p\dfrac{e_s(t) - e_a(t)}{r_{a,can}(t)}}{\lambda(t)\big(\Delta(t) + \gamma(t)\big)}, \tag{4.12}$$

$$Es_{pot}(t) = \frac{\Delta(t)R_n(t) + K_t\rho_a(t)c_p\dfrac{e_s(t) - e_a(t)}{r_{a,soi}(t)}}{\lambda(t)\left(\Delta(t) + \gamma(t)\left(1 + \dfrac{r_{s,soi}}{r_{a,soi}(t)}\right)\right)},$$

where $Rn(t)$ [$MJ.m^{-2}.d^{-1}$], $\lambda(t)$ [MJ. kg^{-1}] is the latent heat of vaporization of water, Δ(t) [kPa. °C^{-1}] is the gradient of the saturated vapour pressure-temperature function, $\gamma(t)$ [kPa. °C^{-1}]is the psychrometric constant $\rho_a(t)$ [kg. m^{-3}] is the air density, c_p [MJ. kg^{-1}. °C^{-1}] is the specific heat of the air and is equal $1.013.10^{-3}$MJ. kg^{-1}. °C^{-1}, $e_s(t)$ [kPa] is the saturation vapor pressure, $e_a(t)$ [kPa] is the actual vapor pressure, and K_t [s. d^{-1}] is a time conversion factor which corresponds to the number of seconds per simulation time step equal to 86,400 s. d^{-1}. We neglect ground heat flux, which seems to be reasonable for daily calculations (see e.g. Allen et al., 1998; Shuttleworth, 2012).

The aerodynamic resistances of canopy ($r_{a,can}$(t)) and of the soil ($r_{a,soi}(t)$), that depend on the properties of the land cover and the soil respectively, are computed using the formulation of Allen et al. (1998). To assess $r_{a,can}$(t)), roughness lengths and zero displacement plane for the canopy are estimated

from the vegetation height h_{veg} [m] (Allen et al., 1998). To calculate $r_{a,soi}(t)$), the zero plane displacement height is equal to zero ($d = 0$) and the roughness length for momentum and for heat and water vapor transfer are assumed to be equal, as in Šimůnek et al (2009), and denoted as z_0 [m].

Finally, the canopy surface resistance is computed by scaling the stomatal resistance r_{st} [$\mathrm{s.m^{-1}}$] to canopy level using the leaf area index over the vegetated fraction (as in Eq. (4.6) to assess canopy interception capacity), and therefore assuming a homogeneous response across all stomata in the canopy (Liang et al., 1994; Allen et al., 1998):

$$r_{s,can}(t) = \frac{r_{st}}{\left(\frac{LAI(t)}{f_c(t)}\right)}. \tag{4.13}$$

In other large-scale models, $r_{s,can}$ is also often expressed as a function of LAI, which allows to directly represents its seasonality following the variations in LAI.

Seasonality of vegetation

We represent the seasonality of vegetation by describing the seasonal variation of the cell average leaf area index LAI. We use two parameters, the maximum LAI_{max} [$\mathrm{m^2.m^{-2}}$], which is the annual maximum value of LAI during the growing season (assumed to be from June to August) and LAI_{min} [%], which is the percentage of reduction in LAI during the dormant season (assumed to be from December to February). The monthly value of leaf area index LAI_m [$\mathrm{m^2.m^{-2}}$] for the m^{th} month is computed using a continuous, piecewise linear function of LAI_{max} and LAI_{min}, which allows for a smooth transition between dormant and growing seasons and is similar to the function proposed by Allen et al. (1998) to assess the seasonality in crop factors:

$$\begin{aligned}
LAI_m &= \tfrac{LAI_{min}}{100} LAI_{max} && \text{when } m = 1, 2, 12 \\
LAI_m &= \tfrac{LAI_{min}}{100} \tfrac{LAI_{max}}{4} (6 - m) + \tfrac{LAI_{max}}{4} (m - 2) && \text{when } m = 3, 4, 5 \\
LAI_m &= LAI_{max} && \text{when } m = 6, 7, 8 \\
LAI_m &= \tfrac{LAI_{min}}{100} \tfrac{LAI_{max}}{4} (m - 8) + \tfrac{LAI_{max}}{4} (12 - m) && \text{when } m = 9, 10, 11.
\end{aligned} \tag{4.14}$$

The advantage of using this simple parameterisation is that it permits to easily analyse the effect of vegetation seasonality by studying the sensitivity of the model predictions to parameter LAI_{min}, which captures the strength of the seasonal variation in LAI. The timings of the four phases of the seasonality model are adapted for the application at the sites used in the present study that are located in Europe (Sect. 3.3.1).

Water storage in the epikarst

Epikarst water storage $V_{epi,i}(t)$ [mm] for the ith compartment is updated at the end of each time step t as follows:

$$\begin{aligned} V_{epi,i}(t) &= V_{epi,i}(t-1) + R_{epi,i}(t) - Q_{epi,i}(t) - Q_{lat,\mathrm{i}\to\mathrm{i+1}}(t) \quad \text{when } i < n_c, \\ V_{epi,n_c}(t) &= V_{epi,n_c}(t-1) + R_{epi,n_c}(t) - Q_{epi,n_c}(t) - Q_{surf,n_c}(t) \quad \text{when } i = n_c. \end{aligned} \tag{4.15}$$

where $Q_{epi,i}(t)$ [mm] is the potential recharge to the groundwater (Eq. (4.16)), $Q_{lat,\mathrm{i}\to i+1}(t)$ [mm] is the lateral flow from the ith to the $(i+1)$th model compartment and $Q_{surf,n_c}(t)$ [mm] is the surface runoff generated by the n_cth compartment. When soil and epikarst layers are saturated, the concentration flow component of the model is activated. The ith model compartment generates lateral flow towards the $(i+1)$th compartment $Q_{lat,\mathrm{i}\to i+1}(t)$ $[mm]$ equal to its saturation excess. Lateral flow from the n_cth compartment is lost from the cell as surface runoff while the other model compartments do not produce any surface runoff. The epikarst is simulated as a linear reservoir (Rimmer and Hartmann, 2012) with outflow coefficient $K_{E,i}$ [d]:

$$Q_{epi,i}(t) = \min\left(\frac{V_{epi,i}(t-1)}{K_{E,i}}, V_{epi,i}(t-1) + R_{epi,i}(t)\right). \tag{4.16}$$

3.3 Site and data for model testing

3.3.1 Site description

We test the model with plot scale measurements from sites of the FLUXNET network (Baldocchi et al., 2001). We identified four FLUXNET sites across European and Mediterranean carbonate rock areas for which sufficient data are available to force V2Karst and to test the model (see Sect. 3.3.2). A short summary of the sites' characteristics is provided in Figure 3.2 and more detailed information can be found in Table 3.6.

The sites have different climate and land cover properties. The first site (Hainich site, referred to as 'German site') is located in the protected Hainich National Park, Thuringia, central Germany, and is characterised by a suboceanic-submountain climate and a tall and dense deciduous broadleaf forest. The second site (Llano de los Juanes site referred to as 'Spanish site') is located on a plateau of the Sierra de Gádor mountains, south-eastern Spain, has a semi-arid mountain Mediterranean climate and is an open shrubland. The third site (Font-Blanche site, referred to as 'French 1 site') is located in south-eastern France, has a Mediterranean climate and its land cover is medium-height mixed evergreen forest. The

fourth site (Puéchabon site, referred to as 'French 2 site') is located in southern France and is characterised by a Mediterranean climate with a short evergreen broadleaf forest. Overground vegetation properties are well characterised at all sites, but subsurface properties are more uncertain. In particular, the rooting depth water capacity was only well investigated at the French 2 site. The four sites are appropriate for testing V2Karst since they satisfy the model assumptions, namely a karstified or fissured and fractured bedrock, overall high infiltration capacity with limited surface runoff and high clay content in the soil (Table 3.6). Additionally, we consider that snow is not a first order control at the sites. At the two French sites, that have a Mediterranean climate, snow events are rare. At the Spanish and German sites, snow events are typically ephemeral, i.e. the snowpack does not have a seasonal duration and comes and goes over the winter period (we refer to Cantón et al., 2010 for the Spanish site; Pinty et al., 2011 for the German site).

Site name		Hainich (German site)	Llano de los Juanes (Spanish site)
General information	Coordinates	51°04′45″N, 10°27′07″E	36°55'56''N, 2°44'55''W
	Elevation	430 m a.s.l.	1600 m a.s.l
Vegetation	Type	Deciduous broadleaf trees	Shrubs, herbs, bare soil, rock outcrops
	Maximum *LAI*	5 $m^2.m^{-2}$	2.71 $m^2.m^{-2}$
	Height	Around 33 m	0.5 m (average) - 1.2 m (maximum)
	Seasonality	Leaves from May to October	1.31 $m^2.m^{-2}$ (annual minimum)
	Rooting depth	Not available	Roots probably access water below the soil
Soil	Texture	Silty clay	Silt loam and clay loam
	Depth	0.5 - 0.7 m	0.1 – 0.3 m (occasionally up to 1.5 m)
	Available water capacity [a]	0.13 $m^3.m^{-3}$	0.25 $m^3.m^{-3}$
	Other properties	Permeable loess layer of 10 -50 cm between soil and bedrock	Rocky soil
Bedrock		Fissured and fractured limestone	Karstified dolomite and dolines
Hydrology	Surface runoff	Low	Low
	Recharge	Large part of the water balance	Diffuse and concentrated, high temporal variability
Measure-ments	Height for humidity and temperature	43.5 m	1.5m
	Height for wind speed	43.5 m	2.5 m
	Depth for soil moisture	0.05, 0.15, 0.3 m	0.15 m
References		(Knohl et al., 2003; Mund et al., 2010; Pinty et al., 2011), personal communication from Martina Mund and Manfred Fink	(Alcalá et al., 2011; Cantón et al., 2010; Contreras et al., 2008; Li et al., 2007, 2011; Pérez-Priego et al., 2013; Serrano-Ortiz et al., 2007)

Table 3.6 Description of the carbonate rock FLUXNET sites.

[a] between wilting point and field capacity.

Site name		Font-Blanche (French 1 site)	Puéchabon (French 2 site)
General information	Coordinates	43°14′27″N, 5°40′45″E	43°44’29’’N, 3°35’45’’E
	Elevation	420 m a.s.l	270 m a.s.l
Vegetation	Type	Evergreen trees (30% broadleaf and 70% needleleaf)	Evergreen broadleaf trees
	Maximum *LAI*	2.2 $m^2.m^{-2}$	2.9 ± 0.4 $m^2.m^{-2}$
	Height	6 m (broadleaf) and 12 m (needleleaf)	5.5 m
	Seasonality	Not available	Not available
	Rooting depth	Roots probably access water below the soil	4.5 m (150 mm available water capacity)
Soil	Texture	Sandy clay loam	Silty clay loam and clay loam
	Depth	0.6 m (maximum)	No clear limit between soil and epikarst
	Available water capacity [a]	49 mm	No clear limit between soil and epikarst
	Other properties	Rocky soil	Rocky soil
Bedrock		Karstified limestone	Karstified limestone
Hydrology	Surface runoff	Low	Inexistent
	Recharge	Not available	Not available
Measure-ments	Height for humidity and temperature	16 m	12.2 m
	Height for wind speed	16 m	12.2 m
	Depth for soil moisture	Not measured	Not measured
References		(Ecofor, n.d.; Gea-Izquierdo et al., 2015; Simioni et al., 2013), personal communication from Guillaume Simioni,	(Rambal, 1992, 2011; Rambal et al., 2003; Reichstein et al., 2002)

Table 3.6 Continued.

[a] between wilting point and field capacity.

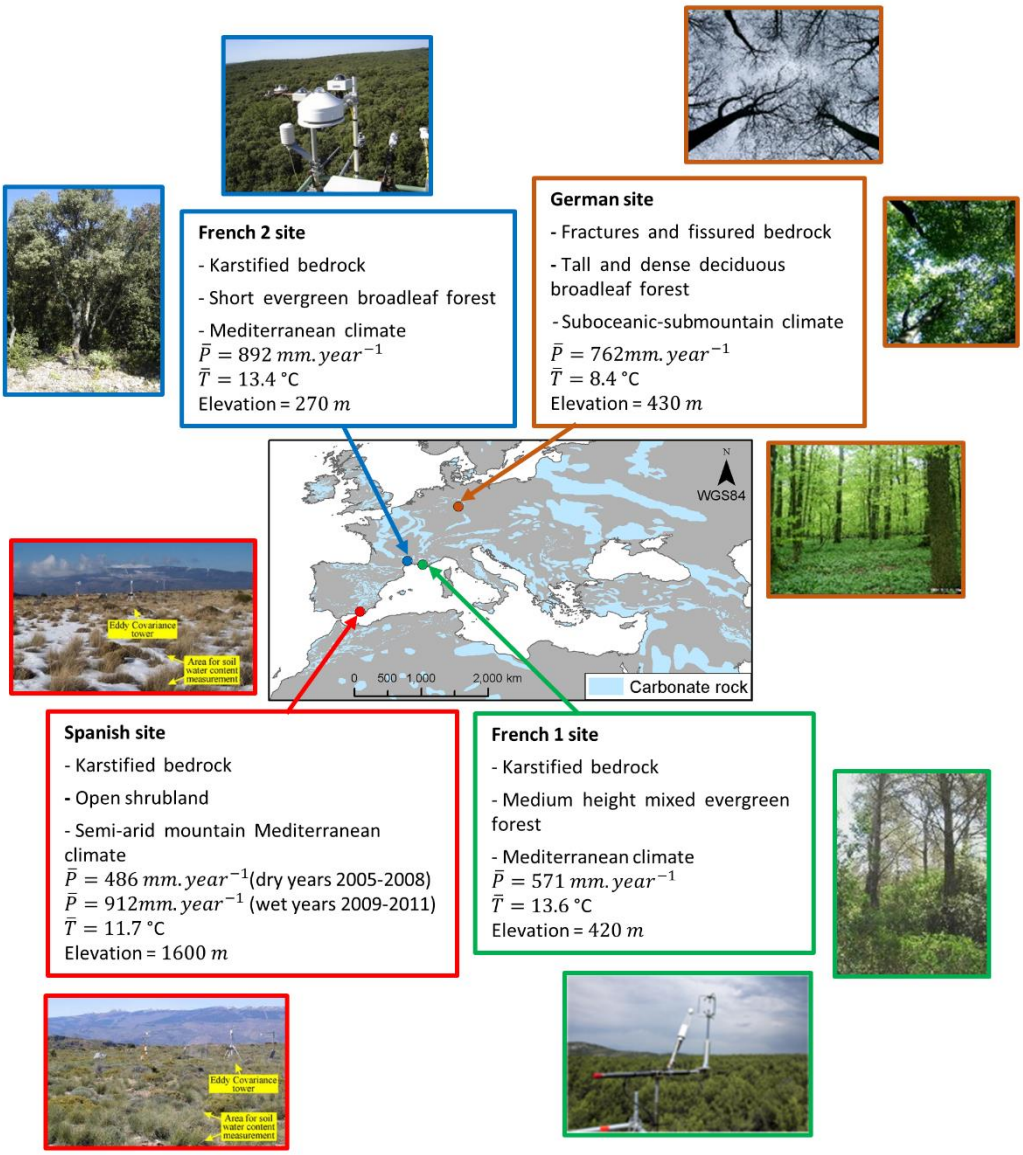


Figure 3.2 Four carbonate rock FLUXNET sites selected for the analyses.

Mean annual precipitation $\bar{P}$ and mean annual temperature $\bar{T}$ were estimated over the period 1 January 2001-17 December 2009 for the German site, 1 January 2006-31 December 2008 for the Spanish site (dry years), 1 January 2009-30 December 2011 for the Spanish site (wet years), 1 January 2010-30 December 2011 for the French 1 site and 1 April 2003-31 March 2009 for the French 2 site.

Sources of the photos: (Pinty et al., 2011) for the German site, (Alcalá et al., 2011) for the Spanish site, http://www.gip-ecofor.org/f-ore-t/fontBlanche.php for the French 1 site, http://puechabon.cefe.cnrs.fr/ for the French 2 site. Source of the carbonate rock and country map: (Williams and Ford, 2006) (country map obtained from Terraspace, Russian space agency).

3.3.2 Data description and preparation

Data available at the four FLUXNET sites include measurements of precipitation, temperature, net radiation, relative humidity and wind speed to force the model, and eddy-covariance measurements of latent heat and at the German and Spanish sites measurements of soil moisture to estimate the model parameters (Sect. 3.4.1). Specifically, at the German site, soil moisture was measured in one vertical soil profile at three different depths (5, 15 and 30 cm) with Theta-probes (Knohl et al., 2003). We selected the measurement at 30 cm depth, which we deem to be most representative of the entire soil column which has a depth between 50 and 60 cm. At the Spanish site, soil moisture was assessed at a depth of 15 cm using a water content reflectometer (Pérez-Priego et al., 2013).

Regarding the data processing, data to force the model were gap-filled and aggregated from 30 min to daily time scale. V2Karst output observations, namely latent heat and soil moisture measurements, were aggregated from 30 min to monthly time scale and we discarded the months when more than 20 % of 30 min data were missing. We also removed monthly aggregated latent heat measurements when the mismatch in the energy balance closure was higher than 50% similar to Miralles et al. (2011). Additionally, we discarded the monthly observations of latent heat and soil moisture for months in which the forcing data contain many gaps, and therefore the impact of the gap-filling of the data on the simulation results is likely to be too significant to sensibly compare simulated and observed soil moisture and latent heat. Further details on the data processing and is reported in Appendix B.3.

Table 3.7 reports the simulation period and the number of monthly latent heat and soil moisture observations that were used to estimate the model parameters at the four FLUXNET sites. We extracted a continuous time series of forcing data covering about 10 years at the German site, 7 years at the Spanish site, 3 years at the French 1 site and 8 years at the French 2 site, while latent heat and soil moisture measurements are not available over the entire simulation time series. All model simulations were performed using a one-year warmup period, which we found to be sufficient to remove the impact of the initial conditions on the simulation results (see Appendix B.4).

Moreover, we corrected latent heat measurements and analysed their uncertainty. We derived two corrected estimates of actual ET, obtained by forcing the closure in the energy balance following Twine et al. (2000) Foken et al. (2012), namely:

1. a corrected value $E_{act,bow}$ $[\mathrm{mm.month^{-1}}]$ that assumes that latent heat (LE $[\mathrm{MJ.m^{-2}.month^{-1}}]$) and sensible heat ($H$ $[\mathrm{MJ.m^{-2}.month^{-1}}]$) have similar errors (referred to as Bowen ratio estimate):

$$E_{act,bow} = \frac{R_n}{\lambda.(1+\frac{H}{LE})}; \quad (4.17)$$

2. a corrected value $E_{act,res}$ [mm. month^{-1}] that assumes errors in latent heat only (referred to as residual estimate):

$$E_{act,res} = \lambda.(R_n - H). \quad (4.18)$$

An additional analysis showed that the two corrected estimates of Eq. (4.17-18) and the uncorrected measure of actual ET are well correlated at the FLUXNET sites, which gives us some confidence regarding the temporal variations in actual ET measurements, while relative errors between corrected and uncorrected estimates can be quite high (see Appendix B.3). We chose to use the Bowen ratio estimate (Eq. (4.17)) to calibrate the model. In fact, it is not clear whether one of the two turbulent fluxes may be more uncertain than the other (Foken et al., 2012).

Site	Simulation period (including a one-year warm-up period)		Number of months with latent heat measurement for calibration	Number of months with soil moisture measurement for calibration
	Start	End		
German site	1 Jan. 2000	17 Dec. 2009	62	74
Spanish site	1 Jan. 2005	30 Dec. 2011	12	12
French 1 site	2 Jan. 2009	30 Dec. 2011	13	Not measured
French 2 site	18 Apr. 2002	29 Jun. 2009	37	Not measured

Table 3.7 Simulation period at the four FLUXNET sites, and number of months where latent heat measurements and soil moisture measurements are available to calibrate the model.
Soil moisture measurements are not provided at the two French sites.

3.4 Methods

To test the plausibility of V2Karst realisations at FLUXNET sites, we estimate the model parameters by constraining model simulations with actual ET and soil moisture observations (Sect. 3.4.1), and we perform two Sensitivity Analyses using measured data to force the model (Sect. 3.4.2), and synthetic forcing data and land cover change scenarios (Sect. 3.4.3). All the analyses were performed using the SAFE toolbox for Global Sensitivity Analysis (Pianosi et al., 2015).

3.4.1 Parameter estimation at the FLUXNET sites using soft rules

We investigate whether it is possible to estimate parameter values that produce plausible simulations based on information available at each FLUXNET site. To this end, and similarly to (Hartmann et al., 2015), we use 'soft rules' to accept or reject parameter combinations based on the consistency between monthly model simulations on one side, and monthly observations and a priori information on model

fluxes on the other side. Using soft rules instead of 'hard rules' (i.e. minimisation of the mismatch between observations and simulations) allows to identify a set of plausible model simulations and accounts for the fact that (1) the observed soil moisture is not strictly commensurate with simulated soil moisture, (2) observations are affected by uncertainties (see Sect. 4.3.2) and (3) it is not expected that V2Karst simulations closely match site-specific data, since the model structure is based on general understanding of karst systems for large-scale applications and may not account for some site specificities. We define five soft rules to identify acceptable ('behavioural') parameter combinations:

1. **The bias between observed and simulated actual ET is below 20**%:

$$Bias = \left| \frac{\sum_{t \in M_{ET}} (E_{act,sim}(t) - E_{act,bow}(t))}{\sum_{t \in M_{ET}} E_{act,bow}(t)} \right| < 20\%, \quad (4.19)$$

   where $E_{act,sim}(t)$[mm] is the simulated actual ET for month t (sum of transpiration, soil evaporation and evaporation from canopy interception), $E_{act,bow}(t)$[mm] is the Bowen ratio correction of observed actual ET (Eq. (4.17)), and M_{ET} is the set of months for which latent heat measurements are available. This rule allows to constrain the simulated water balance.

2. **The correlation coefficient (ρ_{ET}) between observed monthly actual ET ($E_{act,bw}$) and simulated total actual ET ($E_{act,sim}$) is above 0.6.** This rule ensures that the temporal pattern of simulated ET follows the observed pattern.
3. **The correlation coefficient (ρ_{SM}) between observed monthly soil moisture (SM_{obs} [% soil saturation]) and simulated monthly soil moisture (SM_{sim} [$\mathbf{m^3 . m^{-3}}$ soil volume]) is above 0.6.** Simulated soil moisture SM_{sim} for month t is calculated as the average soil moisture within the root zone over all model compartments. This rule guarantees that soil moisture variations are consistent with observations.
4. **Total simulated surface runoff (Q_{surf}) is less than 10% of precipitation**, in accordance with a priori information on the carbonate rock sites, which attests that runoff is negligible (see section 3.3.1).
5. **Soil and vegetation parameter values are consistent with a priori information**, i.e. they fall within constrained (site-specific) ranges. This rule applies to the parameters for which a priori information is available at the FLUXNET sites, namely h_{veg}, r_{st}, LAI_{min}, LAI_{max} V_r and V_{soi} and the constrained ranges are reported in Table 3.8. This rule ensures that acceptable model outputs are produced using plausible parameter values.

We note that the different soft rules focus on different aspect of the model's behaviour. Therefore, we expect that each rule will play a role in the constraining of the parameter space and will allow to identify additional non-behavioural parameter combinations compared to application of the other rules. For each

site, we derived a sample of size 100,000 for the 15 parameters of V2Karst using Latin Hypercube Sampling (LHS) and unconstrained (wide) ranges for the model parameters to explore a large range of soil and vegetation types (reported in Table 3.5), and we applied the above rules in sequence to either reject or accept the sampled parameter combinations. We sampled more densely the constrained parameter ranges used in rule 5 so that a sufficiently large number of parameterisations remain after applying rule 5. Similarly to Hartmann et al. (2015), a priori information on parameter ranges (rule 5) is applied last so that we can first assess the constraining of the parameter space based on information on model output only (rules 1 to 4), and then the consistency of this constraining with a priori information (rule 5).

We also note that the thresholds used in rules 1 to 3 are stricter compared to the study by Hartmann et al. (2015), in which the threshold for the bias rule (1) was set to 75% and for the correlation rules (2 and 3) was set to 0. The reason is that in Hartmann et al. (2015) behavioural parameter sets had to be consistent with observations at all sites within each climate zone defined in the study, while here we perform the parameter estimation for each site separately and therefore we expect better model performances.

Parameter	Unit	German site (deciduous forest)		Spanish site (shrubland)		French 1 site (evergreen forest)		French 2 site (evergreen forest)	
		Lower limit	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit
h_{veg}	[m]	23.1	42.9	0.35	0.85	7.1	13.3	3.9	7.2
r_{st}	[s.m^{-1}]	275	400	195	350	320	455	320	455
LAI_{min}	[%]	5	20	34	63	80	100	80	100
LAI_{max}	[m^2.m^{-2}]	3.5	6.5	1.9	3.5	1.5	2.9	2.0	3.8
V_r	[mm]	60	300	30	200	30	200	30	200
V_{soi}	[mm]	60	400	30	300	30	300	30	300

Table 3.8 Site-specific constrained parameter ranges at the four FLUXNET sites for the vegetation parameters (h_{veg}, r_{st}, LAI_{min}, LAI_{max}, V_r) and for the soil storage capacity (V_{soi}).
Information on how the ranges were determined is provided in Appendix B.2. Parameters are defined in Table 3.5.

3.4.2 Parameter Global Sensitivity Analysis

We use the Elementary Effect Test (Saltelli et al., 2008), or method of Morris (Morris, 1991) described in detail in Sect. 2.2.2. This is a global sensitivity analysis method, and therefore it permits to analyse sensitivity across the entire parameter variability space, it is well suited for identifying uninfluential parameters (Campolongo et al., 2007; Saltelli et al., 2008) and it can be applied to dependent parameters

(in V2Karst it is assumed that $V_e \leq V_r \leq V_{S,n}$ as explained in Sect. 3.2.4). The method requires the computation of the Elementary Effects (EEs) of each parameter in n different baseline points in the parameter space (Eq. (2.6)). For a given model parameter i, the sensitivity indices analysed in the present chapter are the mean of the absolute values of the EEs (denoted by μ_i^*, Eq. (2.7)) introduced in Campolong et al. (2007), which is a measure of the total effect of the ith parameter, and the standard deviation of the EEs (σ_i) proposed in Morris (1991), which is an aggregate measure of the intensity of the interactions of the ith parameter with the other parameters and of the degree of non-linearity in the model response to changes in the ith parameter.

The total number of model evaluations required to compute these two sensitivity indices is $n.(M+1)$, where n is the number of baseline points chosen by the user. We implement the Elementary Effect Test following the radial design proposed by Campolongo et al. (2011) as explained in Sect. 2.2.2. The baseline points were randomly selected using LHS (as in Chapter 2) for the 15 parameters of V2Karst, and dropping the parameter sets that did not meet the condition $V_e \leq V_r \leq V_{S,n}$. In our application, we used $n = 500$ points, which means that we needed 8000 model evaluations for each sensitivity analysis for each of the four FLUXNET sites. We derived confidence intervals on the sensitivity indices via bootstrapping using 1000 bootstrap resamples and checked the convergence of the results at the chosen sample size, as in Chapter 2 (Sarrazin et al., 2016).

3.4.3 Virtual experiments to analyse sensitivity to climate and land cover change

Our last analysis consists of a set virtual experiments to investigate the sensitivity of recharge and actual ET simulated by V2Karst to changes in (1) the precipitation properties (specifically precipitation average amount and temporal distribution) and (2) land cover (specifically from forest to shrub and vice versa).

Virtual experiments using numerical models permit full control on experimental conditions, and thus to unequivocally attribute changes in model outputs to changes in model inputs (see e.g. Weiler and McDonnell, 2004; Pechlivanidis et al., 2016). Several studies have used virtual experiments to analyse the impact of precipitation spatial and temporal variability on hydrologic model outputs. In fact, using historical precipitation time series or future projections only allow to explore a limited range of possible realisations, which makes it difficult to disentangle the effects of different precipitation properties on model outputs. Instead, synthetic precipitation time series can be tailored to analyse the impact of specific precipitation characteristics, for instance precipitation spatial distribution (Van Werkhoven et al., 2008b; Pechlivanidis et al., 2016) and precipitation temporal distribution, namely frequency and intensity (Porporato et al., 2004; Jothityangkoon and Sivapalan, 2009), storminess (Jothityangkoon and

Sivapalan, 2009) and seasonality (Laio et al., 2002; Botter et al., 2009; Jothityangkoon and Sivapalan, 2009; Yin et al., 2014). In this chapter, we create a synthetic precipitation time series where the same precipitation event is periodically repeated. The precipitation time series is characterised by the intensity of precipitation events I_p [$\mathrm{mm.\,d^{-1}}$] and the interval between two wet days H_p [d]. The duration of each precipitation event here is set to one day. The average monthly precipitation P_m [$\mathrm{mm.\,month^{-1}}$] for an average month with 30 days is therefore equal to:

$$P_m = 30.\frac{I_p}{1 + H_p} \tag{4.20}$$

To determine the possible range of variation of the three variables, P_m, I_p and H_p, we analysed their distributions at the four FLUXNET sites and over all European and Mediterranean carbonate rock areas using GLDAS data (Rodell et al., 2004) (distributions are reported in Appendix B.5). We found that wide but plausible ranges are: P_m varies between 0 and 500 mm.month^{-1}, I_p varies between 0 and 200 m.d^{-1} and H_p varies between 0 and 89 d (note that $H_p = 0$ means that it rains every day). We then derived a set of 2266 precipitation time series by deterministically sampling P_m, and H_p within those ranges (and consequently deriving a sampled value of I_p from Eq. (4.20)). We sampled more densely closer to the lower bound of the ranges since lower values of P_m and H_p are more likely to occur.

For each of the precipitation time series so obtained, we ran the V2Karst model until the simulated fluxes reached a steady-state (i.e. periodic oscillations of all state and flux variables) and we analysed the steady-state monthly average of recharge, transpiration, soil evaporation and evaporation from interception.

The experiments are conducted at two virtual sites that are designed based on the characteristics of the FLUXNET sites. Specifically, we use a virtual 'forest site' that has the characteristics of the German site (i.e. its behavioural parameterisations for the soil, epikarst and vegetation parameters) and a virtual 'shrub site' that has the characteristics of the Spanish site. The forest site also inherits the suboceanic-submountain climate characteristics of the German site (i.e. we force the model by the average values of air temperature, net radiation, humidity and wind speed measured at that site), while the shrub site inherits the semi-arid climate of the Spanish site. To investigate the impact of a change in land cover at these virtual sites, we swapped the vegetation parameters (indicated in Table 3.5) between the two virtual sites.

We do not investigate the effects of varying temperature, net radiation, relative humidity and wind speed characteristics as we did for precipitation, because these weather variables are correlated (see e.g. Ivanov et al., 2007) and therefore they cannot be varied independently. Instead, we account for their overall

combined effect in a simple way by analysing the changes in sensitivity when these variables are set to winter (low energy for ET) and summer (high energy for ET) conditions. Table 3.9 reports the values of the parameters and weather variables used at the two virtual sites.

V2Karst input		Unit	Virtual forest site	Virtual shrub site
Vegetation parameter	h_{veg}	[m]	32.1	0.4
	r_{st}	[s.m^{-1}]	390	291
	LAI_{min}	[%]	16	38
	LAI_{max}	[m^2.m^{-2}]	5.0	2.0
	V_r	[mm]	289	151
	V_{can}	[mm LAI]	0.29	0.35
	k	[-]	0.53	0.45
Soil and epikarst parameter	f_{red}	[-]	0.010	0.080
	z_0	[m]	0.0110	0.0045
	$r_{s,soi}$	[s.m^{-1}]	56	61
	V_e	[mm]	11	8
	a	[-]	1.8	1.9
	V_{soil}	[mm]	373	174
	V_{epi}	[mm]	396	519
	K_{epi}	[d]	33	15
Weather input (winter)	R_n	[MJ.m^{-2}.d^{-1}]	-0.0	2.2
	T	[°C]	0.1	4.9
	RH	[%]	89	61
	WS [a]	[m.s^{-1}]	3.5	4.0
Weather input (summer)	R_n	[MJ.m^{-2}.d^{-1}]	10.5	12.1
	T	[°C]	16.6	20.4
	RH	[%]	72	43
	WS [a]	[m.s^{-1}]	2.6	3.4

Table 3.9 Values of V2Karst parameters and weather variables used in the virtual experiment. Values for the virtual forest site and the virtual shrub site are based on the characteristics of the German FLUXNET site and Spanish FLUXNET site respectively. Values of the model parameters (parameters are defined in Table 3.5) correspond to behavioural parameterisations obtained when calibrating the model and values of the weather variables (R_n net radiation, T temperature, RH relative humidity, WS wind speed) correspond to the average values calculated at FLUXNET sites.

[a] At the virtual shrub site, WS was recalculated at a height of 43.5 m because the original measurement provided at a height of 2.5 m at the Spanish site was too low to simulate a change of land cover to tall vegetation (forest). More details on this are reported in Appendix B.3.

3.5 Results

3.5.1 Parameter estimation

Constraining of the parameter space

Figure 3.3 shows that behavioural parameterisations consistent with all rules can be identified at all sites, but their number is very different from one site to another. Specifically, out of the initial 100,000 randomly generated parameter samples, we found 36,838 behavioural parameterisations at the German site, 147 at the Spanish site, 6354 at the French 1 site and 4077 at the French 2 site. From Fig. 3.3, we also see that the application of each rule reduces the number of behavioural parameterisations, except for rule 4 (value of total surface runoff < 10% of precipitation), since all model simulations produce less than 7% of surface runoff at all sites. This can be explained by the fact that V2Karst gives priority to recharge production over surface runoff. Therefore, the latter only occurs under extremely wet conditions when all model compartments are saturated. Our results confirm that each rule, apart from the rule 4 (rule on surface runoff), brings additional information on the model's behaviour and plays a role in the constraining of the parameter space.

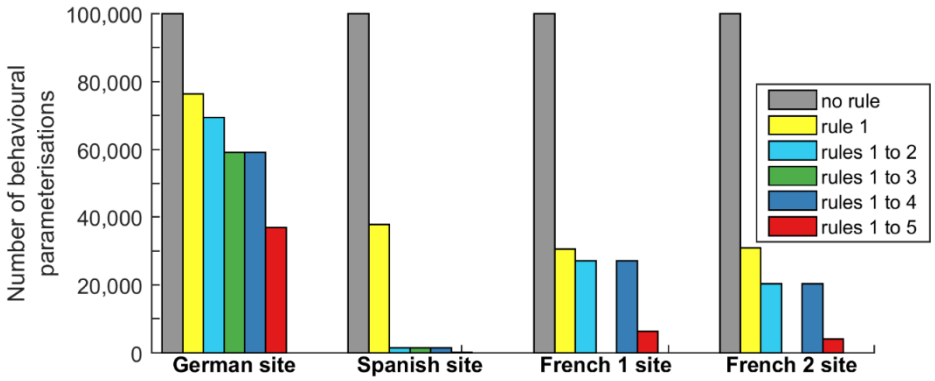


Figure 3.3 Reduction in the number of behavioural parameterisations of the V2Karst model at the four FLUXNET sites, when applying sequentially the five soft rules.
The soft rules are defined in Sect. 3.4.1. no rule: initial sample; rule 1: ET bias; rule 2: ET correlation; rule 3: soil moisture correlation; rule 4: runoff; rule 5: a priori information. Rule 3 could not be applied to the French sites where soil moisture observations are not available.

Figure 3.4 reports a parallel coordinate plot of the behavioural parameter sets and associated values of the output metrics after sequential application of the soft rules. The application of rules 1 to 4 does not significantly reduce the parameter ranges, but it only allows to discard low values of parameters V_r and V_{soi} at all sites (dark blue lines in Fig. 3.4). Instead, the application of rule 5 (a priori parameter ranges, red lines in Fig. 3.4) permits a significant reduction in parameter ranges, not only for the parameters that are directly constrained by this rule (h_{veg}, r_{st}, LAI_{min}, LAI_{max}, V_r and V_{soi}) but also for the spatial

variability coefficient a. Specifically, behavioural values of parameter a are found to be between 0 and 3.2 at the French 1 site, between 0 and 2.8 at the French 2 site. At the Spanish site, we also observe that the behavioural simulations (red lines) cover more densely some portions of the ranges, specifically higher values of parameters $r_{s,soi}$ and a, and lower values of z_0 and V_e. This means that the value for these parameters is more likely to be within these sub-ranges.

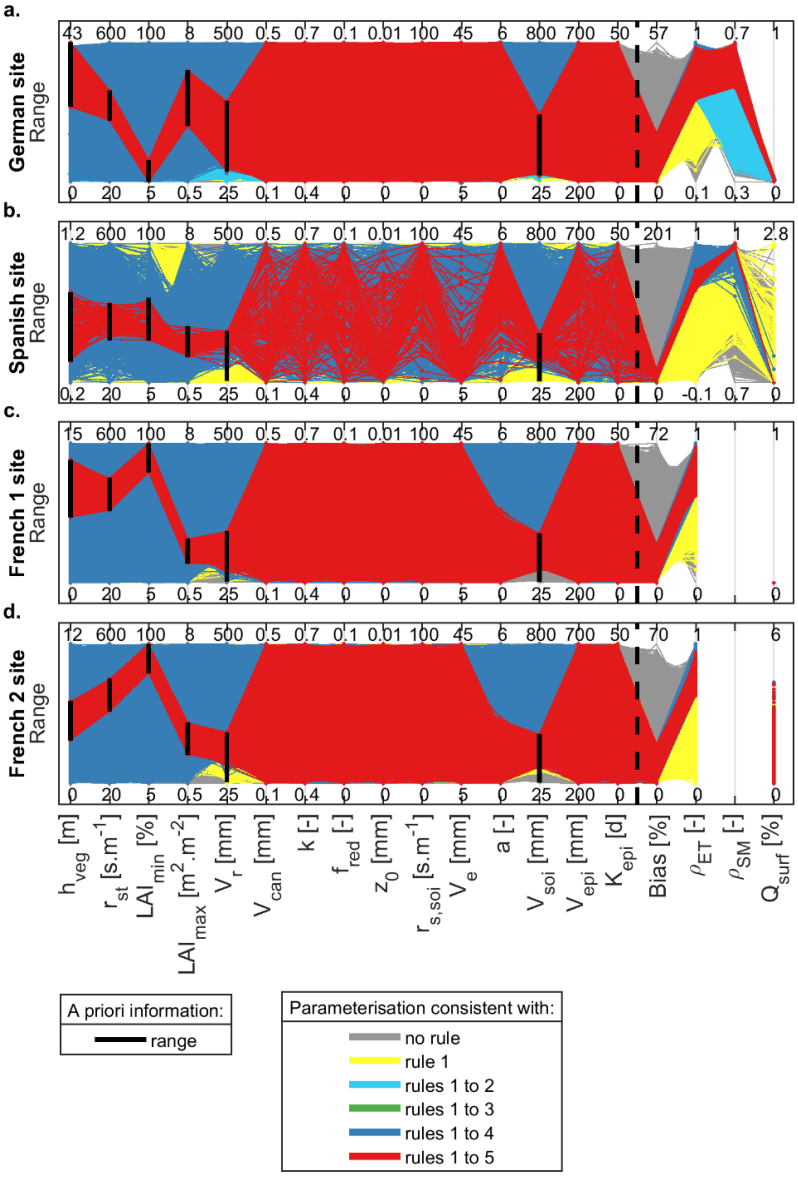


Figure 3.4 Parallel coordinate plots representing V2Karst behavioural parameterisations, and their corresponding simulated output values, identified when sequentially applying the five soft rules at (a) the German site, (b) the Spanish site, (c) the French 1 site and (d) the French 2 site.

The soft rules are defined in Sect. 3.4.1 Parameters are defined in Table 3.5. $BIAS$ absolute mean error between observed and simulated total actual ET (rule 1), ρ_{ET} correlation coefficient between observed and simulated total actual ET (rule 2), ρ_{SM} correlation coefficient between observed and simulated soil moisture (rule 3), Q_{surf} surface runoff (rule 4). Rule 5 corresponds to application of a priori information on parameter ranges (black vertical bars).

Analysis of model simulations

In this paragraph, we analyse the partitioning of the water fluxes simulated by the V2Karst using the behavioural parameterisations. Figure 3.5a compares the total simulated recharge and the total actual ET, expressed in percentage of total precipitation at the four FLUXNET sites (mean and 95% confidence interval across the behavioural parameterisations). At the Spanish site, we present the results over two different time periods that have very different precipitation amounts, namely a drier period from 1 January 2006 to 31 December 2008 and a wetter period from 1 January 2009 to 30 December 2011 (see Fig. 3.2). Figure 3.5 shows that, apart from extremely wet periods at Spanish site, in all other cases the fraction of recharge (Q_{epi}) is significantly lower than the fraction actual ET (ET_{act}). Figure 3.5b shows the partitioning of ET among its different components (transpiration, soil evaporation and interception). We observe that transpiration (T_{act}) is the largest component at all sites, while the relative importance of evaporation from canopy interception (Ec_{act}) and soil evaporation (Es_{act}) varies across sites. In particular, at the German site, Ec_{act} is on average particularly high compared to the other sites, which may be partly explained by the fact that summer LAI (parameter LAI_{max}) is higher at this densely forested site compared to the other sites, and therefore the summer canopy storage capacity is higher as well.

Finally, Fig. 3.6 presents the time series of monthly precipitation input (P), simulated monthly recharge (Q_{epi}), total actual ET (E_{act}) and soil moisture in the root zone (SM_{sim}) at the four FLUXNET sites. Observation of soil moisture and actual ET are also reported and the blue lines correspond to the Bowen ratio corrected estimate used in rules 1-2 for parameter estimation (see Sect. 3.4.1). We see that the soft rules allow to significantly reduce the uncertainty in model outputs at all sites. In fact, the width of the behavioural ensemble, i.e. the ensemble of simulations obtained by application of the rules (black lines), is much narrower than the non-behavioural ensemble (grey lines). Simulated actual ET (E_{act}) is also closer to the observations (blue line) in the behavioural ensemble compared to the non-behavioural one. This means that the application of the soft rules and a priori information on parameter ranges allows not only to improve the precision of the simulated states and fluxes (reduced uncertainty ranges of the simulations), but also the accuracy of simulated actual ET (simulations close to observations). Moreover, the model structure is flexible enough to capture most corrected and uncorrected ET observations, since the non-behavioural model ensemble (grey) includes most corrected and uncorrected ET values.

From Fig. 3.6, we also observe that the seasonal variations in model predictions are consistent with our understanding of the sites over the entire simulation horizon and not only over the months for which ET and soil moisture observations are used to estimate the parameters (blue and red areas in the plot).

Specifically, at the German site we find a marked seasonality of simulated E_{act} and SM_{sim}, with low E_{act} and high SM_{sim} in winter, and high E_{act} and low SM_{sim} in spring and summer. In fact, in winter, the energy available for ET is low and the deciduous vegetation is not able to transpire or intercept large amounts of precipitation, while in spring and summer more energy is available for ET and the vegetation has a higher value of LAI, and therefore ET losses can occur and deplete the soil moisture. At the other sites we observe a similar pattern for SM_{sim}, while E_{act} tends to peak in spring and to be lower in summer when the ET fluxes are more water-limited than at the German site.

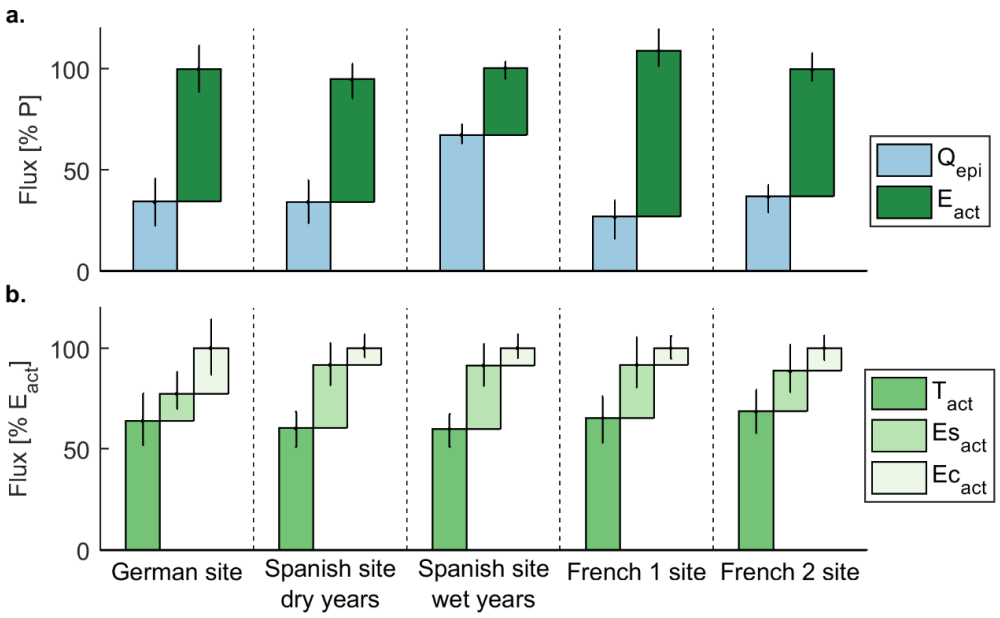


Figure 3.5 (a) Simulated recharge (Q_{epi}) and actual ET (E_{act}) expressed as a percentage of total precipitation and (b) simulated actual transpiration (T_{act}), actual soil evaporation (Es_{act}) and actual evaporation from interception (Ec_{act}) expressed as a percentage E_{act}.

figure reports the ensemble mean and 95% confidence intervals calculated over the behavioural simulation ensemble of the V2Karst model at the four FLUXNET sites. Simulated fluxes were evaluated over the period 1 January 2001-17 December 2009 for the German site, 1 January 2006-31 December 2008 for the Spanish site (dry years), 1 January 2009-30 December 2011 for the Spanish site (wet years), 1 January 2010-30 December 2011 for the French 1 site and 1 April 2003-31 March 2009 for the French 2 site.

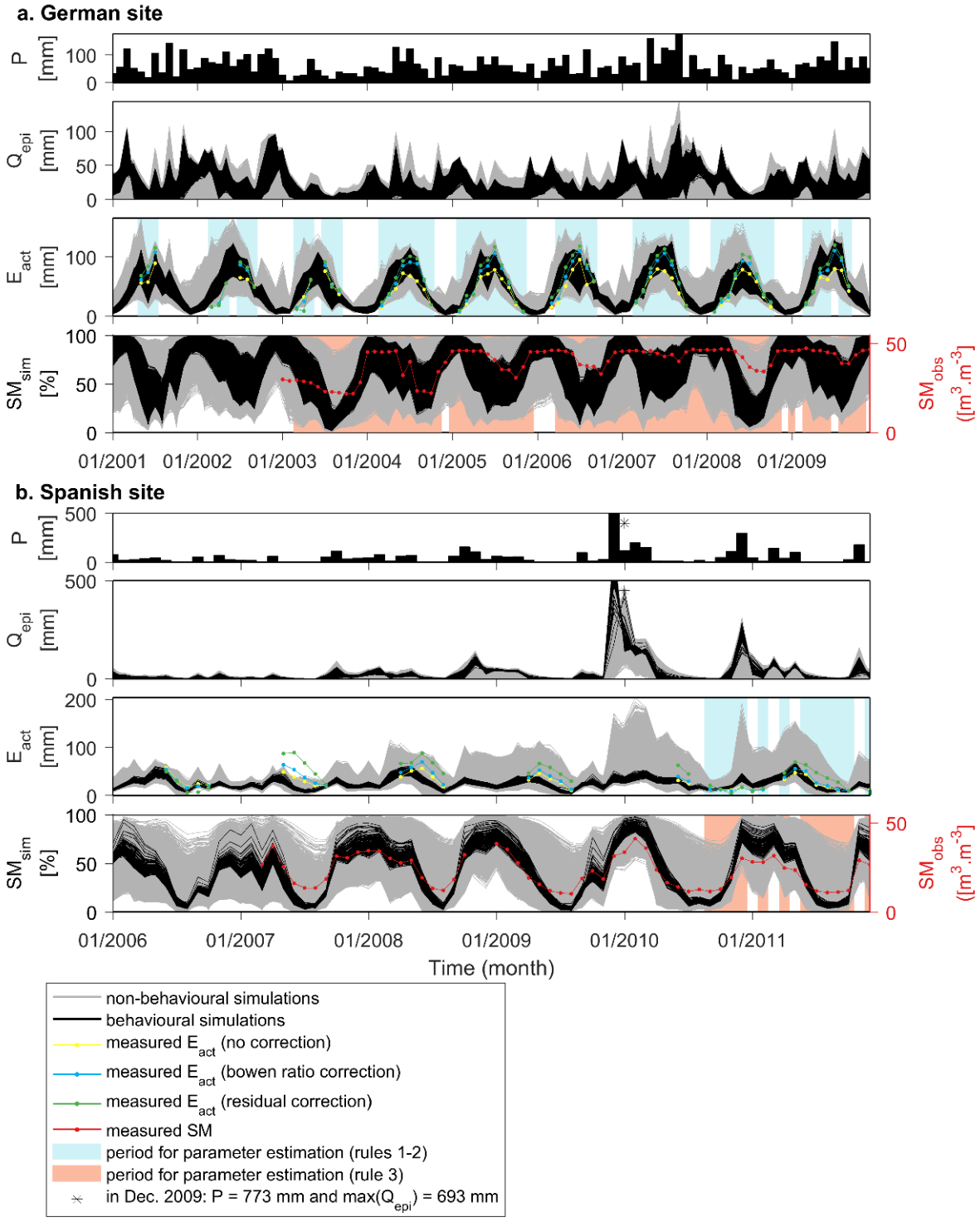


Figure 3.6 Monthly time series of precipitation input (P), simulated recharge (Q_{epi}), simulated actual ET (E_{act}, which is the sum of evaporation from canopy interception, transpiration and soil evaporation), simulated soil moisture within the root zone (SM_{sim}), and monthly observations of actual ET and soil moisture at the FLUXNET sites.

Blu and red shaded areas correspond to the periods in which observation of ET and soil moisture respectively were selected to apply the soft rules of Sect. 3.4.1 (further details on data processing in Sect. 3.3.2).

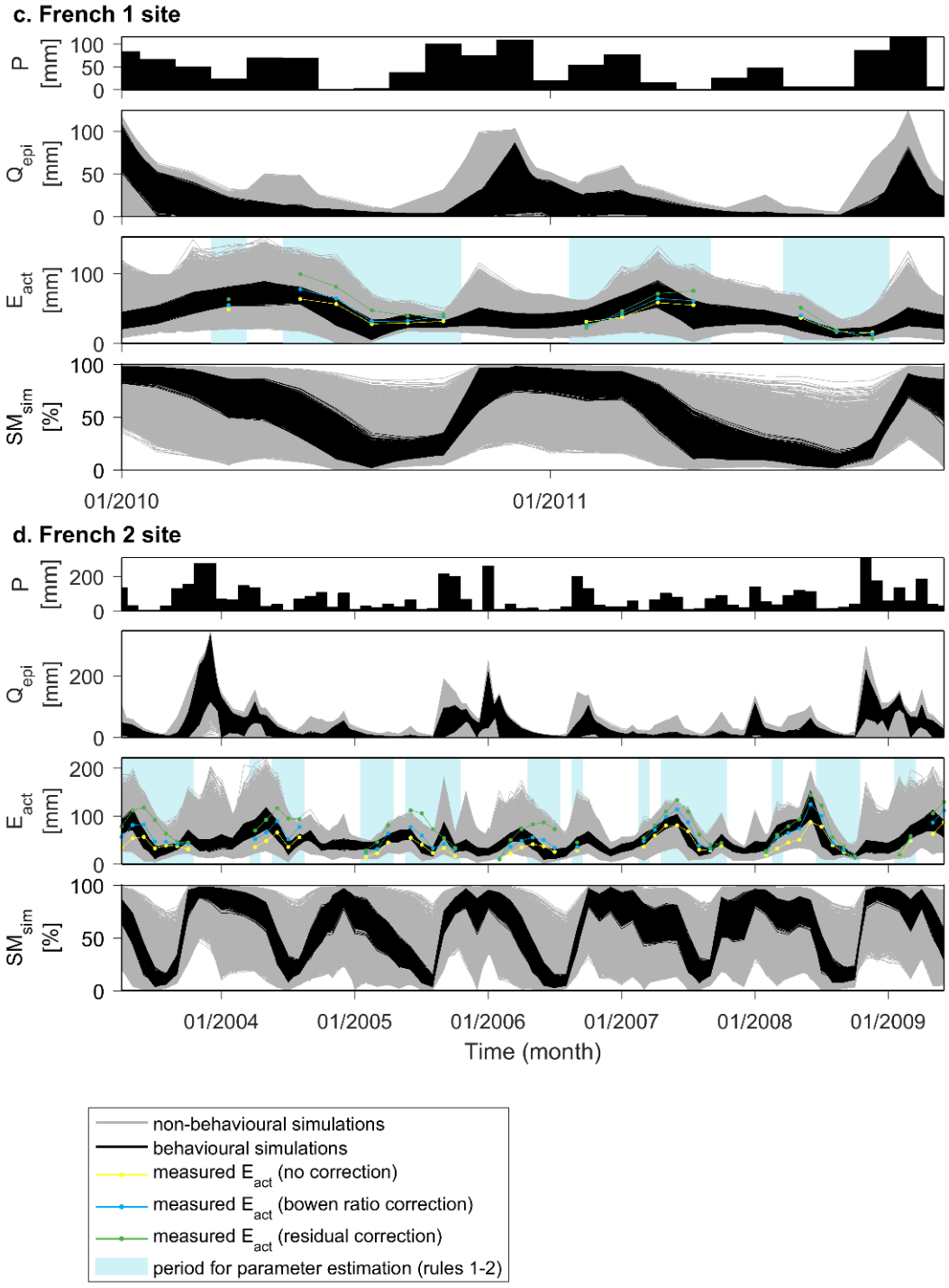


Figure 3.6 Continued.

3.5.2 Parameter Global Sensitivity Analysis

The sensitivity analysis results refer to the sensitivity of total simulated recharge (expressed as a percentage of total precipitation) to the 15 parameters of the V2Karst model. For each parameter, the plots in Fig. 3.7 report on the horizontal axis the absolute mean (μ^*) of the Elementary Effects and on the vertical axis their standard deviation (σ). In all plots, we observe that the bootstrap confidence intervals of the sensitivity indices are narrow and show little overlap, which gives confidence that the sensitivity results are robust. Similar to the analysis of the simulated fluxes in Sect. 3.5.1 (Fig. 3.5), at the Spanish site we present the results for two different time periods with different precipitation amounts.

Global Sensitivity Analysis with constrained parameter ranges

We first examine the left panels in Fig. 3.7, which show the sensitivity results when (h_{veg}, r_{st}, LAI_{min}, LAI_{max} and V_r) and the soil storage capacity V_{soi} are sampled within constrained ranges to inform model calibration in future model applications, since such parameter ranges capture the uncertainty in parameter values left after considering site-specific information. We first note that μ^* and σ take a non-zero value for all parameters at all sites, which means that all parameters are influential and have a non-linear effect on recharge, possibly through interactions with other parameters. The existence of parameter interactions can explain the limited reduction in some parameter ranges during our parameter estimation (Sect. 3.4.1).

We observe that the spatial variability coefficient a has by far the largest influence, followed by parameters V_{soi} and V_r. In fact, their value of μ^* is significantly higher than the other parameters at all sites. The implication for model calibration in future applications of V2Karst is that efforts should primarily seek to reduce the uncertainty in parameters a, V_{soi} and V_r. These three parameters also have a significantly large value of σ, which indicates non-linearities in the model response to variations in these parameters and which is coherent with the nature of Eq. (4.1-2). Interestingly, parameter V_{can}, that controls evaporation from interception, and $r_{s,soi}$, that controls soil evaporation, have an impact on recharge at most sites and at the Spanish site during wet years respectively. This shows that the processes of evaporation from interception and soil evaporation can be important for recharge simulations.

Moreover, we observe that parameters LAI_{min}, z_0 k and V_e have a very small impact on total recharge at all sites ($\mu^* < 3$ %). However, Appendix B.6 reports additional sensitivity analysis results for other model outputs and shows that the most influential parameters that should be the focus of the calibration strategy vary depending on the output of interest. In particular, parameter V_e has a significant impact on the fraction of actual transpiration in total ET, and therefore on the partitioning of ET among its different components.

Global Sensitivity Analysis with unconstrained parameter ranges

The right panels of Fig. 3.7 show the sensitivity indices when sampling parameters within unconstrained ranges. This analysis allows to test the plausibility of the model structure through the assessment of the model sensitivity across a large spectrum of soil and vegetation conditions.

The most apparent difference with respect to the previous SA results is that vegetation parameters (h_{veg}, r_{st}, LAI_{min} LAI_{max} and V_r) now have a much higher value of the sensitivity indices (both μ^* and σ). More specifically, LAI_{max} has a very high sensitivity index at all sites ($\mu^* > 10.5\%$), which can be attributed to the fact that this parameter is used to calculate different model components. Interestingly, the seasonality of leaf area index appears to play an important role in V2Karst since μ^* for LAI_{min}, although always lower than μ^* for LAI_{max}, stands out at all sites.

When comparing parameter sensitivities across sites, we see some significant differences, that we can interpret by considering their climatic differences. In fact, we would expect transpiration to be mainly energy-limited at the German site, given that it has a suboceanic-submountain climate and mainly water-limited at the French sites, which have a Mediterranean climate, and at the Spanish site, which has a semi-arid Mediterranean climate. Specifically, the most influential parameter at the Spanish site is by far parameter a (high μ^*), which has an impact on the water storage in the soil and therefore on the amount of water available to sustain ET between rain events, while at the German site parameter LAI_{max}, which is used to calculate PET, has the largest effect on recharge (high μ^*).We also notice that parameters r_{st} and h_{veg}, that control PET, are more influential at the German site compared to the other sites.

Finally, we observe that, the parameters that specifically control the volume of transpiration (r_{st} and V_r) have a significantly higher value of μ^* than the parameters that specifically control soil evaporation (z_0, $r_{s,soi}$ and V_e) and evaporation from interception (V_{can}). Moreover, z_0, $r_{s,soi}$ and V_e have a very small impact ($\mu^* < 3$ %), while parameter V_{can} can have an important effect at the German site ($\mu^* = 5.7$ %). This suggests that transpiration is overall dominating the ET fluxes at these sites when exploring a wide range of soil and land cover properties and that interception is an important process under the climate of the German site. Additionally, we see that parameter f_{red}, that controls transpiration from the third soil layer, has an impact on recharge simulated at the Spanish site.

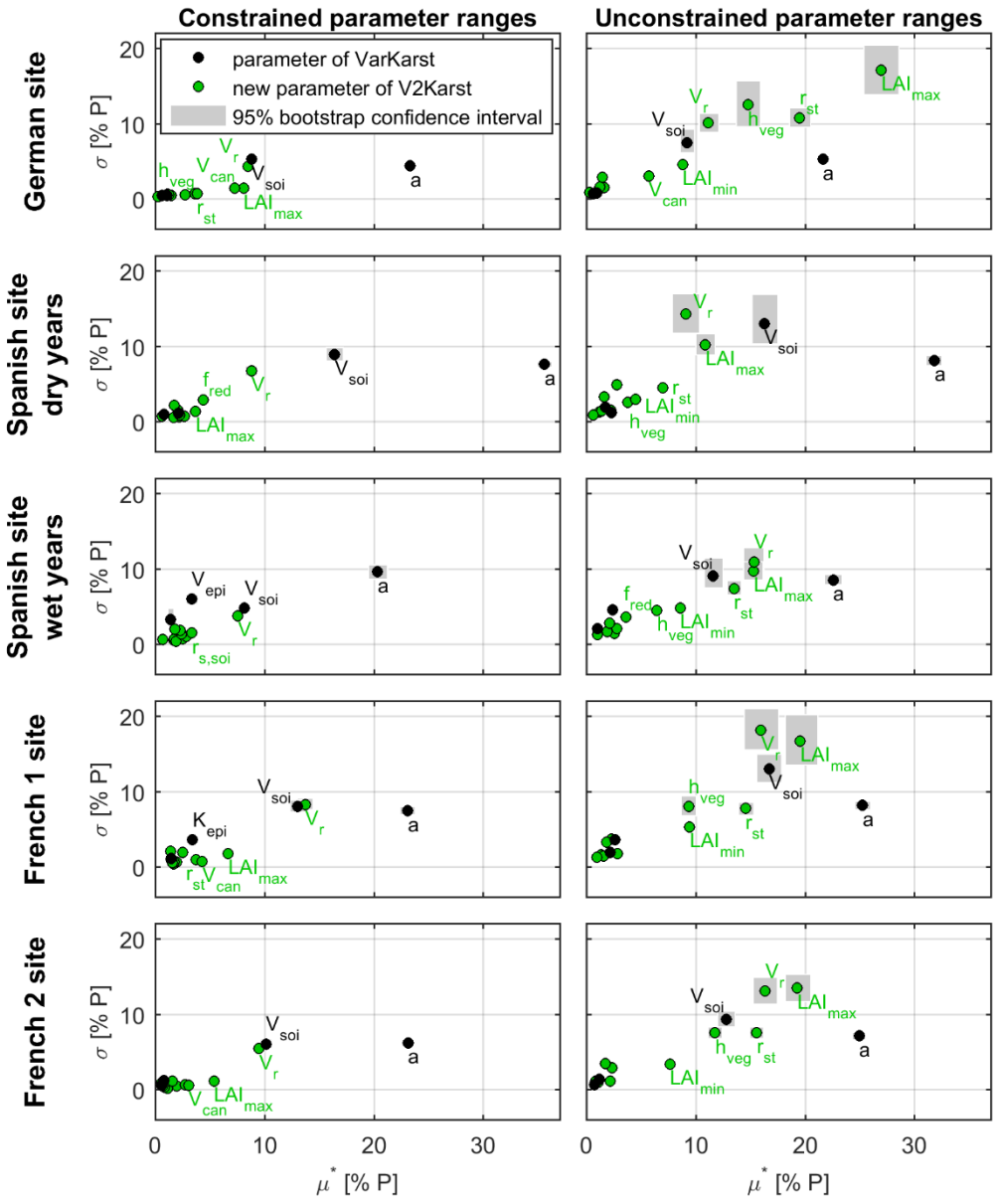


Figure 3.7 Sensitivity indices of the V2Karst parameters (μ^* is the mean of the absolute Elementary Effects and σ is the standard deviation of the Elementary Effects) for total simulated recharge (expressed as a percentage of total precipitation) at the four FLUXNET sites, when constrained (site-specific) parameter ranges are used (ranges of Table 3.5) and when unconstrained ranges are used (ranges of Table 3.8).

Sensitivity indices were computed over the period 1 January 2001-17 December 2009 for the German site, 1 January 2006-31 December 2008 for the Spanish site (dry years), 1 January 2009-30 December 2011 for the Spanish site (wet years), 1 January 2010-30 December 2011 for the French 1 site and 1 April 2003-31 March 2009 for the French 2 site.

3.5.3 Virtual experiments

Sensitivity of simulated fluxes to precipitation characteristics

Figure 3.8 shows the monthly average value of simulated recharge Q_{epi}, for different values of the precipitation monthly amount P_m (x-axis) and the interval between rainy days H_p (y-axis) at the virtual forest and shrub sites. We do not report Q_{epi} values in the top right of the plots because this region corresponds to very intense precipitation events (higher than 200 mm.d^{-1}) that have a very low probability of occurrence (see Sect. 3.4.3).

From the top left panel of Fig.3.8, we see that winter Q_{epi} is mostly sensitive to P_m, in fact simulated recharge increases when moving along the horizontal direction from left to right, but shows little variations along the vertical direction (when H_p is varied). This result is due to the fact that actual ET is very limited in winter because of the low energy available. We indeed estimated that the maximum value of total ET across the different precipitation inputs is 13 mm.month^{-1} at the forest site and 35 mm.month^{-1} at the shrub site. Therefore, a large part of precipitation becomes recharge rather independently of its temporal distribution.

From the right panel of Fig. 3.8, we observe a systematic reduction in summer Q_{epi} compared to winter at both virtual sites. Moreover, summer recharge is overall highly sensitive not only to P_m but also to H_p, since it increases when moving along the vertical direction from bottom to top, i.e. when the same amount of monthly precipitation falls in less frequent but more intense events. This result can be explained by the fact that in summer potential ET is larger and therefore, if events are less intense, a larger part of the precipitation is lost via ET, while if instead events are more intense, the canopy and soil stores reach saturation and precipitation generates a saturation excess flow to the epikarst and hence more recharge and less ET. Moreover, in summer, Q_{epi} shows a limited sensitivity to P_m and H_p when these quantities take low values (brown and red dots on the left of the plots), because only few soil compartments reach saturation under drier conditions and therefore little recharge can be generated. We also see that at the shrub site, Q_{epi} is a significant flux ($Q_{epi} > 5mm$) for smaller values of P_m and H_p compared to the forest site, which may be due to the fact that at the shrub site, the soil water capacity (V_{soi}) is much smaller and therefore the soil compartments can reach saturation under drier conditions.

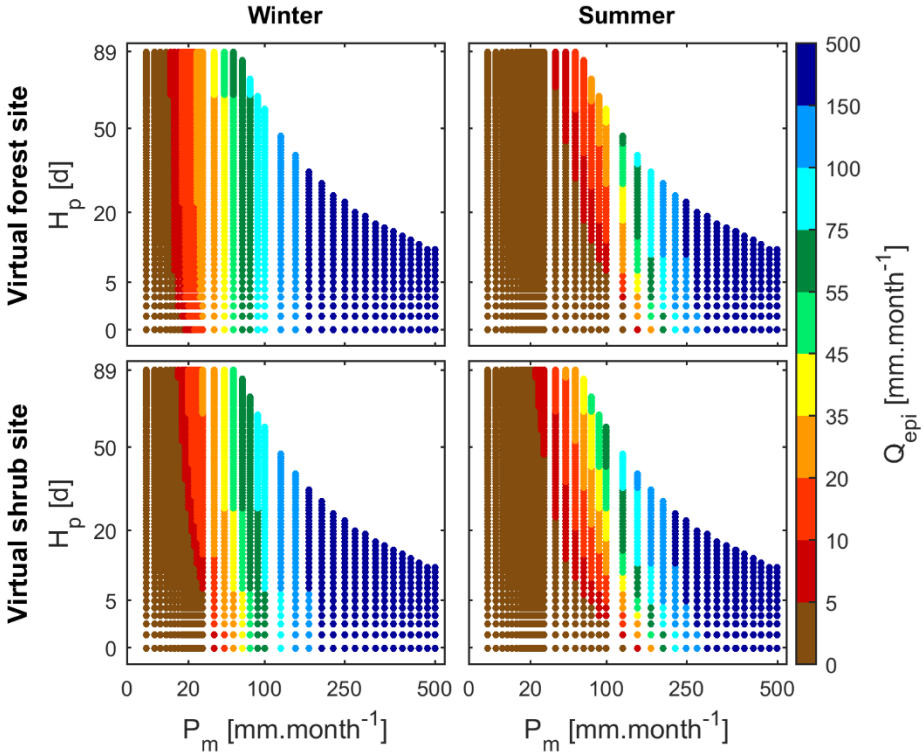


Figure 3.8 Average monthly recharge (Q_{epi}) simulated with V2Karst for different values of the average monthly precipitation amount P_m [mm. month^{-1}] and the interval between wet days H_p [d] of the synthetic periodic precipitation input used to force the model at the virtual forest and shurb sites and under winter and summer conditions.

Sensitivity of simulated fluxes to land cover change

Figure 3.9 reports the results of another virtual experiment similar to Fig. 3.8 but focusing on the impact of land cover change. Specifically, the panels in Fig. 3.9 show the variation in simulated recharge when land cover is changed from forest to shrub at the virtual forest site (and vice versa at the virtual shrub site), and more specifically, Fig.3.9 reports $\Delta Q_{epi} = Q_{epi}^{shrub} - Q_{epi}^{forest}$. We see that in all plots ΔQ_{epi} is positive, which means that recharge is larger and therefore actual ET is lower under shrub compared to forest land cover for both sites. From the left panels of Fig. 3.9, we observe that ΔQ_{epi} is very limited in winter, which is expected since ET fluxes are small in winter (Fig. 3.8).

Instead, the right panels of Fig. 3.9 show that summer ΔQ_{epi} is much higher compared to winter conditions. The value of summer ΔQ_{epi} is largest when the monthly precipitation P_m is high and the interval between wet days H_p is low (green dots at the virtual forest site and dark blue dots at the virtual shrub site), because under these precipitation conditions the amount of moisture available for ET is maximum. Interestingly, for both virtual sites, summer ΔQ_{epi} is sensitive to both P_m and H_p, but its

sensitivity is highly variable across the different precipitation inputs, and more specifically an increase in H_p can have a different effect on ΔQ_{epi} depending on the value of P_m (no variation, increase or decrease in ΔQ_{epi}). In fact, when P_m is low, ΔQ_{epi} is always low and does not vary sensibly when P_m and H_p are varied (brown area in the left end of the plot), since recharge is always low under these precipitation conditions as shown in Fig. 3.8. For intermediate values of P_m, ΔQ_{epi} has a similar pattern at both sites and increases when either H_p or P_m increases. Instead, for high values of P_m, we see that for both sites ΔQ_{epi} decreases when H_p increases and that at the virtual forest site, ΔQ_{epi} increases when P_m increases.

Importantly, our results also show that the impact of a change in land cover can vary greatly across sites, since at the virtual shrub site summer ΔQ_{epi} reaches much higher values and is sensitive to P_m and H_p over a larger range of values of P_m and H_p compared to the virtual forest site.

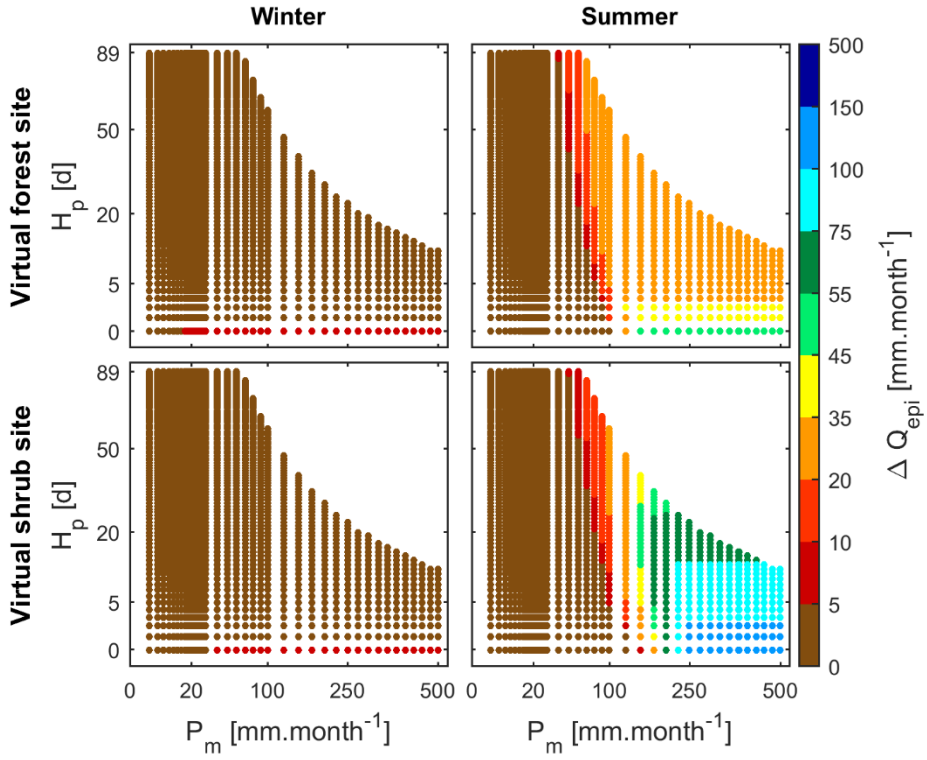


Figure 3.9 Change in monthly recharge ($\Delta Q_{epi} = Q_{epi}^{shrub} - Q_{epi}^{forest}$) simulated with V2Karst when the land cover is set to shrub compared to forest for different values of the average monthly precipitation amount P_m [mm. month^{-1}] and the interval between wet days H_p [d] of the synthetic periodic precipitation input used to force the model at the virtual forest and shurb sites and under winter and summer conditions.

3.6 Discussion

3.6.1 Plausibility of V2Karst simulations

We tested the model by evaluating its ability to reproduce observations at four carbonate rock FLUXNET sites, which is a standard approach to model testing, used for instance to test the previous version of the model Varkarst (Hartmann et al., 2015) and large-scale ET products (Miralles et al., 2011; McCabe et al., 2016; Martens et al., 2017). We demonstrated that V2Karst is able to produce behavioural simulations consistent with observations and a priori information at FLUXNET sites, and additionally the time series of the model outputs are coherent with our understanding of the sites. A different number of behavioural parameterisations was identified at the different sites, because we used the same constraints across sites. The fact that the highest number of behavioural parameterisations was found at the more humid German site and the lowest at the semi-arid Spanish site is coherent with previous findings that higher fit-to-observation can be obtained at wetter locations (Atkinson et al., 2002; Bai et al., 2015).

Interestingly, for the French 1 site, the results of the parameter estimation allow to corroborate the hypothesis that root water uptake is likely to extend below the physical soil layer as communicated by Guillaume Simioni (investigator of the site). In fact, we found here that behavioural values of parameter V_r are higher than 59 mm, while site-specific information indicates that the physical soil layer has a storage capacity of 49 mm (Table 3.6). This result further attests to the realism of V2Karst structure.

Moreover, the GSA using constrained parameter ranges, that are representative of a wide range of different land cover and soil types, showed a set of sensitivities that are interpretable in light of the different climatic conditions at the four FLUXNET sites. This suggests that the model behaves sensibly and consistently with our understanding of the key vegetation-recharge processes we aim at reproducing.

3.6.2 Sensitivity of simulated groundwater recharge to changes in climate and vegetation characteristics in karst areas

In this chapter, we investigated the sensitivity of simulated recharge to both climate and land cover change, through a global sensitivity analysis of the model parameters at the FLUXNET sites, and through virtual experiments using a simple synthetic periodic precipitation input.

Firstly, the results of Elementary Effect Test using unconstrained (wide) ranges showed that the vegetation parameters (h_{veg}, r_{st}, LAI_{min}, LAI_{max} and V_r and additionally V_{can} at the German site) have a significant impact on simulated recharge at the FLUXNET sites, which means that simulated recharge is sensitive to changes in land cover properties. More specifically, the maximum leaf area index

(LAI_{max}) was highly influential at all sites, and to a lesser extent the parameter controlling the seasonality in LAI (LAI_{min}). This is consistent with the findings of previous studies, since (Tesemma et al., 2015) found that assimilating year-to-year monthly LAI in the VIC model can significantly improve runoff simulations compared to using long-term average LAI and Rosero et al. (2010) determined that LAI has a large influence on simulated latent heat in the Noah land surface model. Therefore, the future potential increasing trend in global LAI documented by Zhu et al. (2016) could have a significant impact on the partitioning between green and blue water, including in karst areas.

Our results are also comparable to the sensitivity analysis results obtained for the WaterGap model in Güntner et al (2007) and Werth et al. (2009) with respect to continental water storage and additionally runoff for the latter study. These two studies are the only ones to the author knowledge that performed a parameter global sensitivity analysis including land cover parameters for the large-scale models of Table 3.1. Similar to our results, both studies found that highly influential parameters are parameters that control PET (Priestley-Taylor empirical coefficient in WaterGap, which is replaced by parameters r_{st} and h_{veg} in V2Karst), the water storage capacity in the root zone (denoted as V_r in V2Karst) and at a few sites the interception capacity per unit of LAI (denoted as V_{can} in V2Karst). We note that the impact of parameter LAI was not reported and vegetation seasonality was not considered in these two studies.

Secondly, the results of our virtual experiment showed that simulated recharge is sensitive not only to changes in the precipitation amount but also in the precipitation temporal distribution (interval between wet days) and in land cover, and that its sensitivity is highly dependent on the precipitation properties and on the value of the other weather variables that are used to calculate PET (temperature, net radiation, relative humidity and wind speed). This result is in line with a previous study by Hartmann et al. (2017) also found that recharge simulated with VarKarst is sensitive to the precipitation amount and temporal distribution (specifically intensity of heavy precipitation events), using historical weather time series. Here we complemented the study of Hartmann et al. (2017) by unequivocally attributing the changes in recharge to changes in precipitation properties using virtual experiments. Our results are also consistent with past studies for non-karst areas that established dependencies between hydrological fluxes on one side and precipitation properties on the other side, using synthetic precipitation inputs (Jothityangkoon and Sivapalan, 2009; Porporato et al., 2004) and observations of recharge in a semi-arid tropical region (Taylor et al., 2013a) and in a seasonally humid tropical region (Owor et al., 2009). This further confirm the plausibility of V2Karst sensitivities.

Although, precipitation patterns are more complex than simple periodic variations and the steady state conditions may never be reached in practice, we believe that performing virtual experiments similar to

the ones proposed in this study is a complementary approach to application of climate projections provided by Global Circulation Models (GCMs) and future land cover change scenarios (Holman et al., 2017; e.g. Hurtt et al., 2011), to understand the sensitivity of a model to changes in input characteristics and to determine which aspects of a model input would be worth further investigating.

3.6.3 Applying V2Karst over larger domains

In this section, we first discuss the importance of the new processes that we introduced in V2Karst for large-scale applications and second the strategy to estimate the model parameters over large domains.

The results of our global sensitivity analyses suggest that all newly introduced processes into V2Karst (transpiration, soil evaporation, evaporation from canopy interception, vegetation seasonality and contribution of the water stored below the root zone to transpiration) are relevant for applications over large domains because all of them can affect simulated recharge, depending on the climatic, soil and land cover conditions. Specifically, the results of our Sensitivity Analyses across a large range of soil and land cover conditions (wide unconstrained ranges) showed that overall transpiration and vegetation seasonality are important processes under the climate of the four FLUXNET sites, and additionally evaporation from canopy interception and the contribution of water stored below the root zone are also important model components under the climate of the German site and Spanish site respectively. Moreover, the Sensitivity Analysis using site-specific constrained ranges revealed that the process of evaporation from canopy interception has an effect on simulated recharge at all forested sites (German site and two French sites) and that the process of soil evaporation has an impact on simulated recharge at the semi-arid site with sparse and short vegetation (Spanish site). The importance of representing canopy interception, in particular for forested land covers, was already mentioned in previous studies (; Savenije, 2004; Gerrits, 2010) and the significance of separating transpiration and soil evaporation was reported in Wang and Dickinson (2012) and in Maxwell and Condon (2016).

Regarding the estimation of V2Karst parameters, in this study, we showed that the application of the soft rules based on the comparison between observed and simulated variables and on a priori information on parameter ranges (Sect. 3.4.1) allowed to estimate V2Karst parameters and constrain the model predictions at the four FLUXNET sites. Therefore, to confine V2Karst parameter ranges over a large modelling domain, future studies will investigate the application of an approach similar to the strategy presented in this chapter and in Hartmann et al. (2015) for the VarKarst model, based on soft rules and on the grouping of the model grid cells across the application domain into typical karst-vegetation landscapes. In addition to a priori information on the value of the soil water capacity V_{soi} used in Hartmann et al. (2015), a priori information on the vegetation parameters will also need to be derived from large-scale databases of vegetation properties (see Sect. 3.2.1). We can anticipate that the

estimation of the parameters that characterise sub-surface properties (a, V_{soi}, V_r) may be particularly critical, since our Sensitivity Analyses using site-specific constrained parameter ranges showed that these parameters have the largest impact on simulated recharge. In addition, unlike above-ground vegetation properties that can be more easily observed (e.g. LAI, vegetation height), sub-surface properties are not often well investigated.

The model can account for sub-grid heterogeneity in vegetation type using a 'tile' approach, which consists of subdividing each model grid cell in a number of independent units (tiles), each of which has a specific land cover (e.g. short or tall vegetation). The model can then be evaluated separately over each tile. The overall simulated fluxes for a given grid cell are computed as the area weighted average of the fluxes calculated over the tiles. The same approach is also used in other large-scale hydrological models, such as Mac-PDM (Gosling and Arnell, 2011) and VIC (Bohn and Vivoni, 2016).

We note that the effect of snow was neglected in this chapter because we deemed that snow was not a first order control at the four FLUXNET sites, as explained in Sect. 3.3.1. However, we expect that snow will have a significant effect on karst groundwater recharge in snow-dominated areas, such as mountain regions. The effect of snow on hydrological fluxes in karst areas has been investigated in a few studies. These studies showed that snow dynamics affect streamflow timing in mountainous upper catchments of the Jordan river (Samuels et al., 2010) and affect the dynamics of spring discharge in an Alpine karst catchment (Chen et al., 2018). Additionally, previous studies in non-karst areas demonstrated that snow dynamics have an effect not only on the dynamic of the hydrological fluxes but also on the long-term partitioning of precipitation between available water (streamflow) and evapotranspiration (e.g. Berghuijs et al., 2014; Barnhart et al., 2016). In particular, representing snow seems to be highly relevant for climate change impact studies, since future projections indicate a decrease in snow cover in the Northern Hemisphere (IPCC, 2013), including in the Alpine region in Europe (e.g. Marty et al., 2017) and in the Western United States (e.g. Fyfe et al., 2017; Gergel et al., 2017). This is expected to impact future water availability.

Given all the above, we recommend that future applications of V2Karst over snow-dominated areas account for the effect of snow. This could be done by using inputs of snow water equivalent as in previous applications of the VarKarst model (Hartmann et al., 2015, 2017) as discussed in Sect. 3.2.4. This approach will be implemented in Chapter 4 to estimate recharge with V2Karst over carbonate rock areas in Europe, the Middle East and Northern Africa. Alternatively, an explicit snow routine could be coupled to the V2Karst model, using for instance a simple degree day method similar to the snow routine of the HBV model (Bergström, 1995, the HBV model was introduced in Sect. 2.4.4). A snow routine based on the degree day method was used in a few local studies in mountainous karst areas (Butscher

and Huggenberger, 2008; Samuels et al., 2010; Chen et al., 2018). However, these studies have highlighted the challenges related to the development of a snow routine and the estimation of the snow parameters given the lack of snow observations. Additionally, the effect of future decreasing snow cover on snowmelt and therefore on hydrological fluxes is not well understood and involves complex mechanisms. In fact, Musselman et al. (2017) found that snowmelt rates depend not only on temperature as widely accepted, but also on the depth of the snow pack, which is not accounted for in current modelling approaches. The results of Musselman et al. (2017) suggest that snowmelt rates could decrease in a warmer world. Therefore, the representation of snow processes in models, and in particular in karst areas, will require in-depth investigation in future studies and we recommend that future studies examine the representation of explicit snow processes into the V2Karst model.

One question that we think is still insufficiently addressed in large-scale hydrological modelling is the issue of which parameters should be varied during parameter estimation and uncertainty analysis, and instead which parameters can be reasonably fixed to a constant value across the modelling domain to simplify the analyses. Other studies have reported on the issue, and in particular a study by Cuntz (2016) showed that some constant parameters of the Noah-MP land surface model can be highly influential for some model outputs. Likewise, in this chapter, we found that parameters V_{can} and $r_{s,soi}$, that are typically fixed in the other large-scale models of Table 3.1, do have an impact on total recharge at least one FLUXNET site. Moreover, as mentioned in Sect. 3.2.4, V_{can} and $r_{s,soi}$ are understood to vary across land cover type and soil type respectively, even if no clear ranges of these parameters have been established across land cover and soil types respectively. Therefore, fixing these two parameters could potentially introduce large uncertainties in V2Karst simulations.

The reason for the modellers' decision to fix a given parameter could for example have been based on the finding that the parameter might not have been influential for a particular site at which sensitivity was analysed. However, it might be that the same parameter is influential for other systems with different characteristics since parameter sensitivity can show a high variability across places as suggested by this chapter and as further demonstrated in Güntner et al. (2007) and in Van Werkhoven et al. (2008a). It is therefore particularly important to assess the sensitivity of model parameters across the modelling domain to test the suitability of fixing model parameters, as done in this study at FLUXNET sites.

3.7 Conclusions

The objectives of this chapter were (1) to develop and test an ET component with explicit representation of land cover processes for the large-scale karst recharge model VarKarst, so that the model can be used for climate and land cover change impact assessment, (2) to evaluate the mechanisms of recharge

production in the model as well as the model's sensitivity to temporal precipitation patterns and land cover using virtual experiment.

Many different approaches are used to represent ET in large-scale hydrologic models, and the lack of in-situ ET observations makes it difficult to assess and compare the performance of these different formulations. Moreover, some models use a large number of parameters that can be only poorly constrained by the few available observations. High model complexity also makes Monte Carlo simulation computationally expensive and hampers uncertainty and sensitivity analysis. The new version of the VarKarst model developed here, V2Karst (V1.0), is the first large-scale model to include explicit representation of both karst and land cover processes. We sought to include parsimoniously processes that are understood to be relevant for climate and land cover impact assessment, namely, (1) a representation of the three ET components (transpiration, soil evaporation in presence of sparse canopy and evaporation from canopy interception) and (2) a physically-based PET equation (Penman-Monteith). The model also comprises a parsimonious representation of vegetation seasonality.

We showed that V2Karst was able to produce plausible simulations at four carbonate rock FLUXNET sites, since its simulations were consistent with observations of latent heat and soil moisture and a priori information at the sites, and the parameters that dominate the model sensitivity were in accordance with our perception of expected controls on recharge. Additionally, it was also shown that all newly introduced processes in V2Karst can have an impact on simulated recharge depending on the climate, the soil properties and the land cover.

Virtual experiments, using synthetic periodic precipitation inputs to force the model, allowed to characterise the sensitivity of simulated recharge to the precipitation temporal distribution, the precipitation amount, the seasonal conditions of the other climate variables and the land cover. This had been little examined in previous studies in karst areas. Our results call for a large-scale assessment of the combined impact of future changes in climate (and more specifically the precipitation amount and temporal distribution) and in land cover on groundwater recharge in karst areas.

Importantly, our study demonstrate that global sensitivity analysis can provide valuable insights for model development, since it can help to determine which processes should be included in models and which parameters can be fixed to constant values with little impact on the simulations. Moreover, GSA, allows to characterise a model sensitivity to changes in climate and land cover. We therefore believe that large-scale hydrology would benefit from a more exhaustive evaluation of the models' sensitivities over their application domain, since so far sensitivity analyses of large-scale models are very few and many of them explore a limited ranges of possible parameter combinations only.

The code of the V2Karst model is open source and freely available under the terms of the GNU General Public License version 3.0. The model code, written in Matlab, is provided through a Github repository: https://github.com/fannysarrazin/V2Karst_model (https://doi.org/10.5281/zenodo.1205463).

Chapter 4. Sensitivity of karst groundwater recharge to climate and land cover changes across Europe, the Middle East and Northern Africa

This chapter is part of a work in preparation:

Sarrazin F, Hartmann A, Pianosi F, Rosolem R, Wagener, T., How do land cover and subsurface heterogeneity modulate climate change impacts on future groundwater recharge in karst areas?, in preparation.

For this work, partial support to Andreas Hartmann was provided by the Emmy Noether-Programme of the German Research Foundation (DFG; grant number HA 8113/1-1; project 'Global Assessment of Water Stress in Karst Regions in a Changing World'). Francesca Pianosi was partially supported by a UK Engineering and Physical Sciences Research Council fellowship (EPSRC; grant number EP/R007330/1). Numerical experiments have been performed using the computational facilities of the advanced Computing Research Centre, University of Bristol (http://www.bris.ac.uk/acrc).

4.1 Introduction

Groundwater widely sustains domestic water consumption, agricultural irrigation and ecosystem health (Giordano, 2009; Kløve et al., 2011; Siebert et al., 2010). Aquifer recharge is therefore a key component of the water balance because it characterises the amount of renewable groundwater available to meet human and environmental/ecosystem demand (Scanlon et al., 2006; Döll and Fiedler, 2008; Wada et al., 2012). Future changes in climate and land cover are expected to modify groundwater recharge patterns, while the future global population growth will increase human water needs (Vörösmarty, 2002; Taylor et al., 2013b; Kim and Jackson, 2012; Mohan et al., 2018). The design of sustainable groundwater management strategies is therefore required to guarantee future water and food security, to preserve biodiversity and to stop the trend of decreasing groundwater storage that has already been noted in some regions (Aeschbach-Hertig and Gleeson, 2012; Rodell et al., 2018). The design of land cover/use management strategies, which has been discussed in recent studies (e.g. Foley et al., 2011; Holman et al., 2017) could also play a significant role for the availability of groundwater resources.

Critically, the combined impact of changes in climate and land cover on karst groundwater recharge has not been investigated so far at large-scales. Yet such studies are essential to guide future water management in karst areas, where major springs have already dried up because of groundwater overexploitation (Hartmann et al., 2012; Li et al., 2012; Zakhem and Kattaa, 2017). Karst aquifers are of high relevance because they supply drinking water to 10-25% of the world population (Ford and Williams, 2007; Stevanović, 2018). Karst groundwater recharge is also likely to be particularly sensitive to future changes in climate and land cover (Hartmann et al., 2017). In fact, karst systems typically develop from the weathering of carbonate rocks, which results in extremely heterogeneous subsurface properties and very high infiltration capacities (Ford and Williams, 2007; Hartmann et al., 2014).

To simulate the impact of climate and land cover/use change, previous studies have widely adopted a ‘top-down’ or ‘scenario-led’ approach, such as the study by Gosling et al. (2017) to assess the impact of climate change on river runoff and the study by Rajib and Merwade (2017) to evaluate the hydrological impact of land use change. The ‘top-down’ approach consists of the propagation of the uncertainty arising from emission and socio-economic scenarios, climate and land cover/use projection model, downscaling technique (to derive high resolution climate and land cover/use projections) and impact model (e.g. hydrological models). This approach has the advantage of being based on a physically constrained set of climate and land cover/use model projections. However, it tends to produce a large range of possible futures which is not easily interpretable by decision makers, while climate projections only cover a limited number of scenarios (partly due to computational limitations) and therefore it is likely to explore only partially the range of future uncertainties (Stainforth et al., 2007;

Wilby and Dessai, 2010; Kundzewicz et al., 2018). An alternative approach, referred to as 'bottom-up', 'scenario discovery' or 'scenario neutral' strategy, has been proposed to investigate the vulnerability of the variable of interest (e.g. groundwater recharge), i.e. its responsiveness to changes in climate and land cover/use beyond projections by climate and land cover/use models (Groves and Lempert, 2007; Wilby and Dessai, 2010; Prudhomme et al., 2013; Singh et al., 2014; Ray and Brown, 2015). The approach aims to characterise the impact model (e.g. hydrological model) response surface and in particular potential climate and/or land use conditions that would lead to unwanted outcomes typically defined by one or more stakeholders. One critical part of such bottom-up approaches is the definition of credible climate and land cover/use inputs to feed into the impact model as discussed for instance in Prudhomme et al. (2013) and Guo et al. (2017) with respect to climate scenarios.

The present study combines both the 'top-down' and 'bottom-up' approaches. We use virtual land cover change scenario and we analyse climate model projections by 'trading space for time' to explore more exhaustively the space of climate variability and the space of land cover management options. 'Trading space for time' consists of analysing spatial differences across places as a proxy for temporal variations at a given place (e.g. Keese et al., 2005; Singh et al., 2011; Maxwell and Condon, 2016). Keese et al. (2005) assessed the controls of recharge simulated by an unsaturated flow model along a precipitation gradient. Singh et al. (2011) traded space for time to support model calibration. More specifically, they estimated the parameters of a hydrological model considering climate conditions beyond historical observations, so that the model may produce more realistic projections when forced by future climate scenarios. Maxwell and Condon (2016) applied a distributed hydrological model over a large domain in the USA. They traded space for time by jointly analysing the differences between simulated transpiration and soil evaporation across the domain, so to investigate the controls of the partitioning of evapotranspiration into bare soil evaporation and plant transpiration.

The present study aims to assess the sensitivity of karst groundwater recharge to climate and land cover change in karst regions. We focus our study on carbonate rock areas in Europe, the Middle East and Northern Africa because aquifers are an importance source of drinking water in these regions (COST, 1995; Stevanović, 2018). More specifically we attempt to answer the following research questions:

1. What is the relative importance of land cover and subsurface heterogeneity in modulating groundwater recharge under climate variability?
2. How do climate, land cover and subsurface heterogeneity combine to define the propensity of a location to produce more or less recharge?
3. In which European/Mediterranean regions recharge is most sensitive to changes in land cover under climate change?

To address the above mentioned research questions, we simulated groundwater recharge using the V2Karst model introduced in Chapter 3 and forcing climate data from the Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP, Warszawski et al., 2014). We utilised Global Sensitivity Analysis techniques to analyse the controls of simulated groundwater recharge.

4.2 Data and methods

4.2.1 Hydrological model for assessing groundwater recharge in karst regions

Groundwater recharge was simulated using the V2karst model, which is a daily large-scale hydrological model that includes an explicit representation of vegetation properties and karst processes (Chapter 3; Sarrazin et al., 2018). V2Karst is particularly suitable for applications in karst areas, because it explicitly represents sub-grid subsurface heterogeneity by assessing the daily water balance over a number of model vertical compartments with different storage capacity and permeability, and because it accounts for the process of lateral flow that concentrates the infiltrating water from low permeability to high permeability areas (Hartmann et al., 2015, 2017). The subsurface is divided in three soil layers overlying an epikarst layer, which represents the layer of highly weathered bedrock typical of karst systems. Moisture in the soil can be lost back to atmosphere through the processes of transpiration and soil evaporation, while the soil saturation excess infiltrates to the epikarst. Additional evaporation losses can occur from a canopy interception store above ground. Potential ET (PET) in V2Karst is assessed using the Penman-Monteith equation (Monteith, 1965). Groundwater (potential) recharge is calculated as the flux leaving the bottom of the epikarst.

V2Karst can simulate groundwater recharge in karst systems with different land cover properties, by varying the value of the vegetation parameters, and with different degree of subsurface heterogeneity by varying the spatial variability parameter a that controls the variability in storage capacity and permeability across the model vertical compartments (V2Karst parameters are described in Table 4.1). V2karst can also represents the sub-grid variability in land cover type following a 'tile' approach. This approach consists of dividing each grid cell into simulation units with homogeneous land cover properties ('tiles'), of evaluating the model separately over each tile and of calculating the overall simulated fluxes over the grid cell as the area weighted average over the tiles, as explained in Sect. 3.6.3.

Figure 4.1 shows a conceptualisation of the V2Karst representations for two land cover types (grass/crop and forest) and two degrees of subsurface heterogeneity (low and high). Qualitatively, a higher degree of subsurface heterogeneity (Fig. 4.1.b,d) tends to enhance infiltration and recharge, while reducing ET and/or surface runoff (Hartmann et al., 2017). Forest land cover (Fig. 4.1.c,d) tends to produce more ET compared to grass/crop land cover (Fig. 4.1.a,b) and therefore less recharge and/or surface runoff

(Chapter 3). In Chapter 3, it was also shown that V2Karst tends to produce very low amounts of surface runoff, as typically observed in karst systems (Fleury et al., 2007; Contreras et al., 2008; Hartmann et al., 2014).

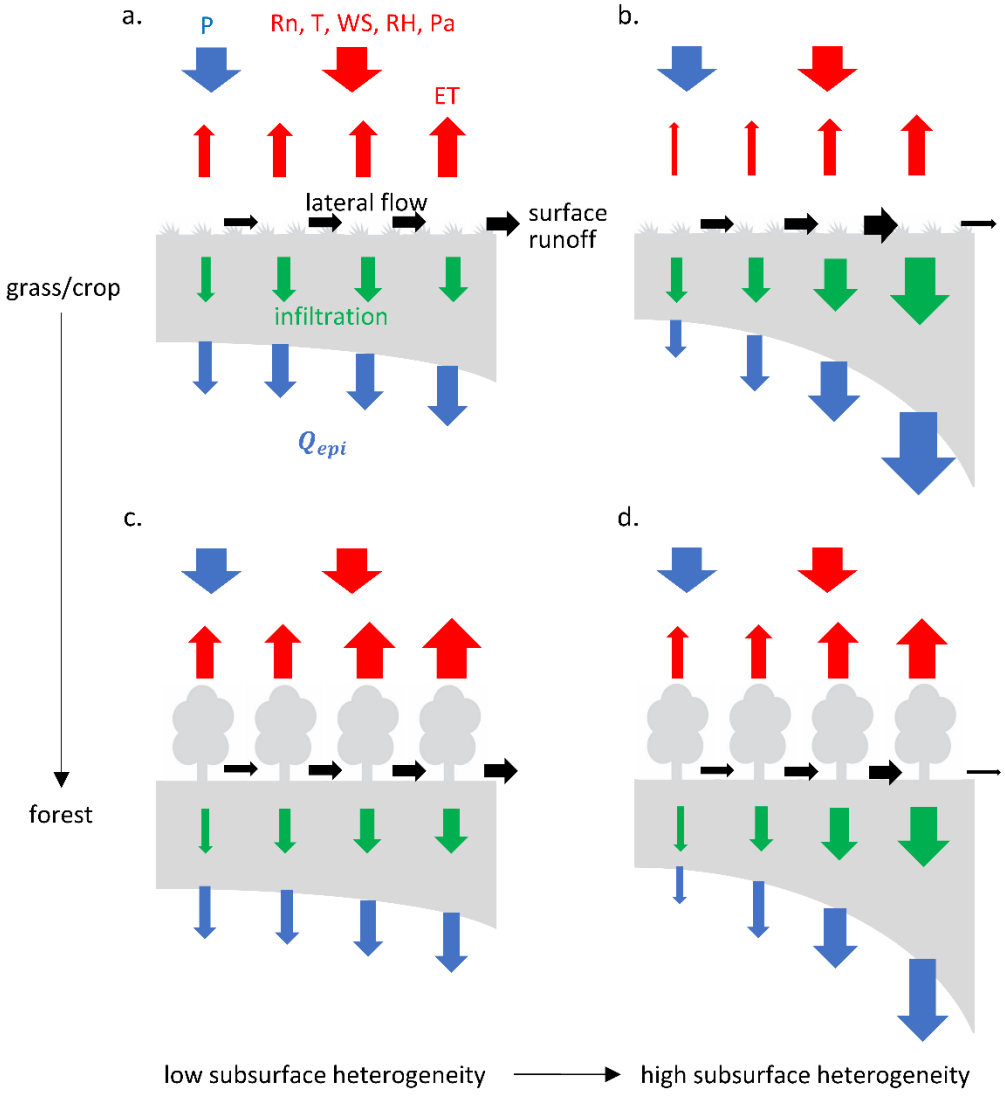


Figure 4.1 : Conceptualisation of the V2Karst model processes within one simulation grid cell for two land cover types and two degrees of subsurface heterogeneity.
a. Grass/crop land cover and low degree of subsurface heterogeneity. b. Grass/crop land cover and high degree of subsurface heterogeneity. c. Forest land cover and low degree of subsurface heterogeneity. d. Forest land cover and high degree of subsurface heterogeneity. P precipitation, R_n net radiation, T temperature, WS wind speed, RH relative humidity, P_a atmospheric pressure, ET evapotranspiration, Q_{epi} groundwater recharge.

Parameter	Description	unit	Category	Lower limit	Upper limit	Reference for values
h_{veg}	Vegetation height	[m]	vegetation	G/C: 0.4 F: 10	G/C: 1.5 F: 35	Appendix C.1
r_{st}	Stomatal resistance	[s.m^{-1}]	vegetation	G/C: 40 F: 100	G/C: 250 F: 500	Appendix C.1
LAI_{min}	Reduction in leaf area index during the dormant season (vegetation seasonality)	[%]	vegetation	5	100	Appendix C.1
LAI_{max}	Annual maximum leaf area index	[m^2.m^{-2}]	vegetation	G/C: 1.5 F: 5	G/C: 5 F: 10	Appendix C.1
V_r	Maximum storage capacity of the root zone	[mm]	vegetation	G/C: 100 F: 200	G/C: 200 F: 800	Appendix C.1
α_{veg}	Vegetation albedo	[-]	vegetation	G/C: 0.16 F: 0.1	G/C: 0.25 F: 0.16	Appendix C.1
V_{can}	Canopy storage capacity per unit of *LAI*	[mm LAI]	vegetation	0.2	0.4	(Dickinson, 1984; Döll et al., 2003)
k	Beer-Lambert's law extinction coefficient	[-]	vegetation	0.4	0.7	(Chapter 3 ; Sarrazin et al., 2018)
f_{red}	Reduction factor for transpiration below the root zone	[-]	soil	0	0.15	(Chapter 3; Sarrazin et al., 2018)
z_0	Soil roughness length	[m]	soil	0.0003	0.013	(Chapter 3; Sarrazin et al., 2018)
$r_{s,soi}$	Soil surface resistance	[s.m^{-1}]	soil	0	100	(Chapter 3; Sarrazin et al., 2018)
V_e	Maximum storage capacity of the first soil layer	[mm]	soil	5	45	(Chapter 3 Sarrazin et al., 2018)
α_{soi}	Soil albedo	[-]	soil	0.3	0.4	(Hagemann, 2002; Kumar et al., 2011; Müller Schmied et al., 2014)
a	Spatial variability coefficient	[-]	soil and epikarst	HUM: 1 MTN: 0 MED: 1 DES: 0	HUM: 3.5 MTN: 3 MED: 6 DES: 6	(Hartmann et al., 2015)
V_{soil}	Mean soil storage capacity	[mm]	soil	HUM: 450 MTN: 250 MED: 50 DES: 0	HUM: 650 MTN: 450 MED: 250 DES: 50	Relative values among landscapes from (Hartmann et al., 2015)
V_{epi}	Mean epikarst storage capacity	[mm]	epikarst	200	700	(Hartmann et al., 2015)
K_{epi}	Mean epikarst outflow coefficient	[d]	epikarst	0	50	(Hartmann et al., 2015)

Table 4.1 The 17 parameters of the V2Karst V1.1 model, and the ranges used in the present study. Compared to V2Karst V1.0 (Chapter 3) two new parameters have been added (α_{veg} and α_{soi}) in V2Karst V1.1. Different ranges for the vegetation parameters h_{veg}, r_{st}, LAI_{max}, V_r and α_{veg} were used for the different land cover types, i.e. for grass/crop (G/C) and for forest (F). Different ranges of the parameters a and V_{soil} were used for four different karst landscapes, i.e. humid (HUM), mountain (MTN), Mediterranean (MED) and desert (DES).

The model version used in the present study (V2Karst V1.1) includes 17 model parameters (see Table 4.1). Compared to V2Karst V1.0 (presented in Chapter 3), the new version presents two additional parameters to compute the net radiation $R_n(t)$ $[MJ.m^{-2}\ d^{-1}]$ at each simulation time step t, namely the vegetation albedo α_{veg} $[-]$ and the bare soil albedo α_{soi} $[-]$, while in V2Karst V1.0 net radiation was simply a model input. In V2Karst V1.1, it is computed as follows:

$$R_n(t) = (1-\alpha)SW_{down}(t) + LW_{down}(t) - \sigma(T(t)+273.15)^4 \quad (4.1)$$

where $SW_{down}(t)$ $[MJ.m^{-2}\ d^{-1}]$ is the downwelling shortwave radiation, $LW_{down}(t)$ $[MJ.m^{-2}\ d^{-1}]$ is the upwelling longwave radiation, T $[°C]$ is the air temperature, α $[-]$ is the surface albedo and σ $[MJ.\ K^{-4}\ m^{-2}\ d^{-1}]$ is the Stefan-Boltzmann constant equal to 4.90 $MJ.\ K^{-4}\ m^{-2}\ d^{-1}$. We note that Eq. (4.1) assumes a value of the surface emissivity equal to 1 and uses air temperature, which differs from the theoretical equation which uses surface temperature (e.g. Bonan, 2015b). The formulation of Eq. (4.1) was shown to reasonably reproduce radiation observations at a range of sites from the FLUXNET network (the network is described in Baldocchi et al., 2001), and we report the results in Appendix C.2. The surface albedo α $[-]$ for each grid cell is assessed as the area weighted average of the vegetation albedo α_{veg} $[-]$ and bare soil albedo α_{soi} $[-]$, since each grid cell includes a vegetated and a bare soil fraction:

$$\alpha = f_c(t)\alpha_{veg} + (1-f_c(t))\alpha_{soi} \quad (4.2)$$

where $f_c(t)$ $[-]$ is the canopy cover fraction.

The modelling approach used in this chapter should also account for the effect of snow on the hydrological fluxes. In fact, a large part of the study domain falls within mountainous areas (see Fig. 2.a) where snow is expected to have a large effect on recharge (as discussed in Sect. 3.6.3). V2Karst V1.1 uses inputs of snow water equivalent to account for the impact of snow on hydrological fluxes, similar to Hartmann et al. (2015, 2017). V2Karst also assumes that no soil evaporation and transpiration occur in presence of snow cover. As discussed in Sect. 3.6.3, we recommend that future studies examine the representation of explicit snow processes into the V2Karst model.

4.2.2 Climate and land cover scenarios for the study domain

We forced the V2Karst model using climate projections of precipitation, air temperature, relative humidity, wind speed, downwelling shortwave radiation, upwelling longwave radiation and atmospheric pressure from the ISI-MIP database (Warszawski et al., 2014). These projections were bias corrected and downscaled to 0.5x0.5° resolution (Hempel et al., 2013). We used all scenarios provided in the ISI-MIP ensemble for the carbonate rock areas in Europe, the Middle East and Northern Africa, i.e. for four

future Representative Concentration Pathways (RCPs, Van Vuuren et al., 2011) and the historical period, and for five General Circulation Models (GCMs), which gives a total of 25 climate scenarios for each grid cell. Relative humidity is provided at a height of 2m and wind speed at a height of 10m. To simulate the effect of forest land cover (tall vegetation), both variables were transformed at a height of 40m following Lhomme et al. (2014) similar to Chapter 3 (Sect. 3.4.3), since these variables need to be above canopy level to evaluate the Penman-Monteith equation. Daily snow water equivalent was derived from monthly values, that were evaluated using the Joint UK Land Environment Simulator (JULES, Best et al., 2011) and that are provided in the ISI-MIP database, using linear interpolation, similar to Hartmann et al. (2017).

Our analyses consider two land cover types: crop/grass (i.e. short vegetation) and forest (i.e. tall vegetation). To assess historical recharge over the study domain, we used the forest cover fraction map for the historical period (year 2004) from the Harmonized Global Land Use database V1 (Hurtt et al., 2011; Chini et al., 2014) available at 0.5x0.5° resolution. This map is reported in Fig. 1.4 in Chapter 1. In the Harmonized Global Land use database, the non-forested fraction of the cells is further classified as crop, pasture or natural non-forested vegetation. Our crop/grass land cover type includes all three classes, since our study focuses on assessing the impact of changes in land cover and not in land use. The Harmonized Global Land use database V1 also provides land use projections for the future period for four RCPs. However, these projections capture a limited range of uncertainties in future land cover because only one projection is provided for a given grid cell for each of the four RCPs. Additionally, projections were constructed using strong modelling assumptions, which resulted in classifying about 80% of the cells in our study domain as not-forested in the historical period and not subject to any afforestation in the future (Hurtt et al., 2011). Consequently, we did not use these land cover projections for our analyses, but we developed virtual land cover scenarios to further explore the possible effects of variations in land cover type.

4.2.3 Experimental setup

We ran the V2Karst model over the study domain to identify controls on simulated recharge. A previous study identified four karst landscapes within the simulation domain based on climatic and topographic characteristics, namely humid, mountain, Mediterranean and desert landscapes (Hartmann et al., 2015; see Fig. 2.a). We used different model configurations to estimate recharge for the different landscapes. Each landscape is understood to have specific soil water capacity, with higher values for wetter landscapes (humid and mountain) compared to drier landscapes (Mediterranean and desert). The different landscapes also tend to have different degrees of subsurface heterogeneity. We therefore used different ranges for the spatial variability parameter a and the mean soil water capacity for the different

landscapes following Hartmann et al. (2015) and as reported in Table 4.1. We note from Table 4.1 that the ranges of the a parameters tend to be quite wide, especially for the Mediterranean and the desert regions. This is due to the lack of data at large-scales to constrain these parameter values.

We created a dataset of 17,000,000 simulated recharge values by performing 5000 simulations for each of the 3400 grid cells of the domain. Each of the 5000 simulations correspond to a different set of conditions, obtained by combining 25 climate scenarios from the ISI-MIP database (4 RCPs and the historical period and 5 GCMs), 2 virtual climate scenarios (forest or grass/crop) and 100 parameter sets derived by Latin Hypercube Sampling of the V2Karst parameters (using the ranges in Table 4.1). Each model simulation was performed over a 30-year period, consistently with previous studies that analysed the impact of climate change on hydrological fluxes (e.g. Davie et al., 2013; Mishra et al., 2017; Marx et al., 2018). We selected the period 1st October 1974 - 30th September 2004 to run historical climate scenarios and the period 1st October 2069 - 30th September 2099 to run climate projections corresponding to one of the four RCPs. We used a 3-year warmup period previous to each 30-year period to remove the effect of the initial condition on simulated recharge, similar to Hartmann et al. (2017). To analyse the effect of specific climate properties, we calculated a range of climate descriptors to characterise each of the climate scenario used to force the model. These climate descriptors and their pairwise linear correlations are reported in Appendix C.3, while in this chapter we only analyse the climate descriptors that were identified by our analyses as mostly influential on recharge.

4.2.4 Global Sensitivity Analysis methods

We adopted two very different Global Sensitivity Analysis methods to analyse our input-output dataset with the aim of understanding dominant recharge controls, namely Variance-Based Sensitivity Analysis (VBSA) and Classification and Regression Tree (CART). The sensitivity analyses were performed using the Matlab SAFE toolbox (Pianosi et al., 2015) and the Matlab Statistics and Machine Learning Toolbox (Mathworks, 2016).

Variance-Based Sensitivity Analysis (VBSA)

Variance-Based Sensitivity Analysis (Sobol', 1990; Saltelli et al., 2008; Sect. 2.2.2) allows to assess the contribution of the different input factors (here climate variability, land cover type, subsurface heterogeneity described by the a parameter, and the uncertainty arising from the remaining 16 parameters of the V2karst model) to the model output variance (here recharge). This method is particularly appropriate to quantitatively determine the relative impact of the input factors on a model output and was used for instance in Guo et al. (2017) to examine the effect of climate properties on runoff. In the present study, we focused on the main effects indices (described in Eq. (2.1)), which

measure the direct contribution of individual input factors (i.e. excluding interactions) to the output variance. The higher the main effect for a given input factor, the higher the impact of that input factor on the output. We also assessed the contribution of the interactions among factors, which accounts for the fraction of variance unexplained by the main effects. In other words, the sum of the main effects and of the contribution of input factor interactions is equal to 1. In this study, we applied the computational technique proposed by Saltelli (2002) and presented in Sect. 2.2.2 to calculate the main effects. This computational technique requires a tailored sampling strategy to derive the input sample. The number of model evaluation required to perform VBSA is $n(M + 2)$ where M is the number of input factors (here 4) and n is a base sample size chosen by the user (here set equal to 40,000). We performed VBSA for each of the four karst landscapes, resulting in an additional input-output dataset of size 960,000. The input samples were derived using the ranges reported in Table 4.1 based on Latin Hypercube sampling as described in Sect. 2.2.2. We calculated confidence intervals on the variance fractions using 1000 bootstrap resamples and checked the convergence of the variance fractions following the methodology proposed in Chapter 2. We note that VBSA could be applied using a generic input-output sample as proposed for instance in Anderson et al. (2014) or in Stanfill et al. (2015), which would reduce the computational cost of the analyses.

Classification and regression tree (CART)

Classification and Regression Trees (CART) is a machine learning technique that has been proposed to build non-linear prediction models from a dataset (Breiman et al., 1984). The method was introduced in Sect. 2.2.2 as a possible technique to perform Factor Mapping. Unlike VBSA, CART does not allow to quantitatively measure the relative contribution of the different input factors, but it permits the identification of those specific subranges and combinations of input factors that produce a specific outcome (e.g. low or high recharge values). The results of CART are typically visualised as binary trees composed of nodes and branches. Each internal node of the tree corresponds to a logical expression based on the value of a specific input (e.g. "parameter a is below 1") and splits into two branches that correspond to the two possible outcomes (i.e. logical expression is "true" or "false"). A given leaf node of the tree indicates the outcome which is obtained (e.g. low recharge values) under the conditions identified by the branches leading to that node. The input factors that appear in the internal nodes of the tree can be interpreted as being influential, and the higher they appear in the tree the larger their influence on the output.

Classification trees were applied for instance in Singh et al. (2014) to identify climate and land cover properties that produce lower/higher changes in runoff and in Almeida et al. (2017) to reveal the precipitation conditions and slope properties that can trigger a landslide. In this study, we used

Classification Trees to determine how the different input factors combine to produce more or less recharge. Classification Trees are constructed following a binary recursive algorithm that partitions the input factor space based on the output classes and on a selection criterion that aims to maximise the homogeneity in output classes at each node. We adopted the widely used Gini impurity index as selection criterion, which is implemented in the Matlab Statistics and Machine Learning Toolbox (Mathworks, 2016). The classification error of the tree can then be calculated to assess the quality of the classification and the predictive power of the tree.

4.3 Results

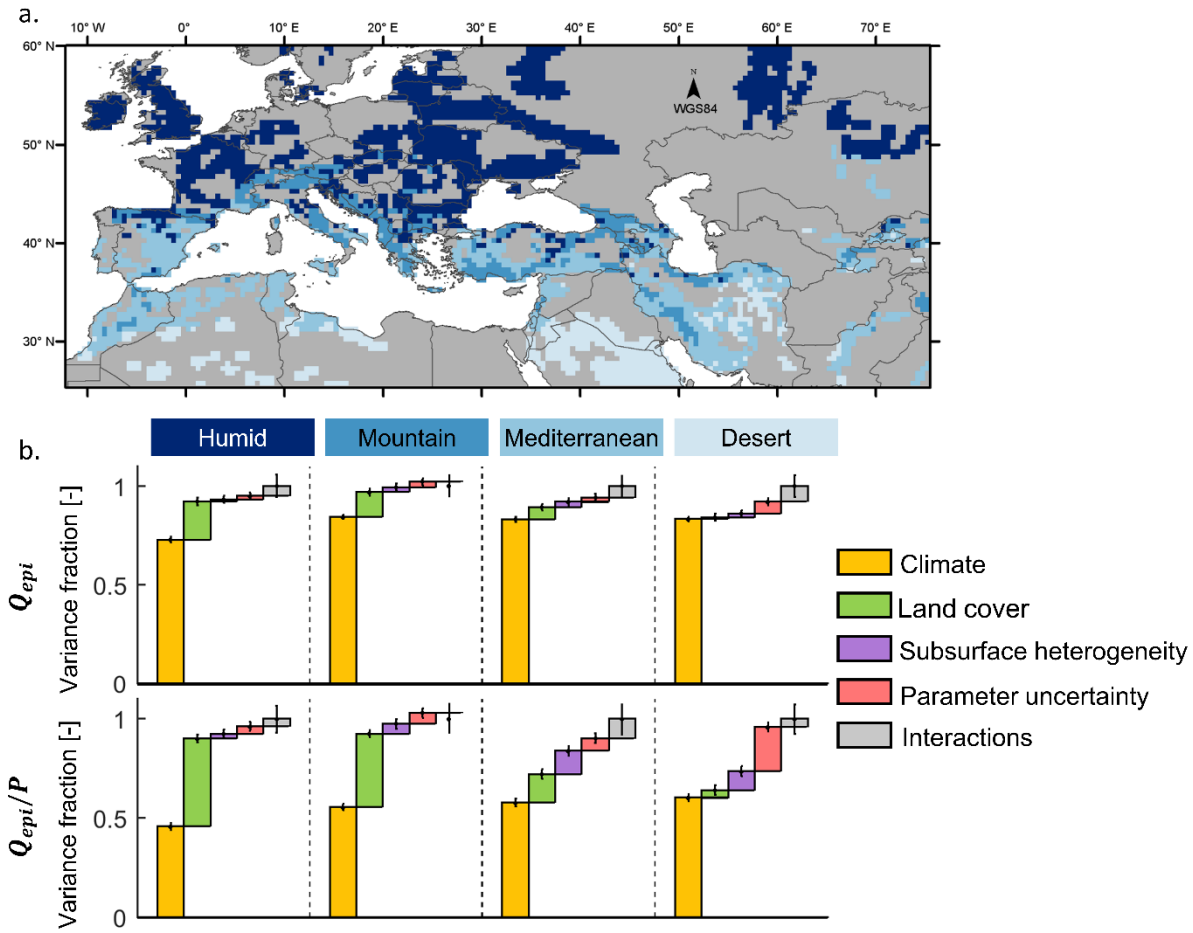


Figure 4.2 Variance decomposition of simulated mean annual recharge and recharge ratio for the four karst landscapes.

(a) Spatial extent of the four karst landscapes. (b) Variance fractions of mean annual recharge Q_{epi} $[mm.y^{-1}]$ and recharge ratio Q_{epi}/P $[-]$ explained by climate variability, land cover type, degree of subsurface heterogeneity and model parameter uncertainty as well as interactions among factors. The graph reports the average values and the 95% confidence intervals (vertical black lines) of the variance fractions (i.e. the VBSA Main effect) calculated over 1000 bootstrap resamples.

Source of the data: Carbonate rock and country map from Williams and Ford (2006) and in particular the country map was obtained from Terraspace, Russian space agency. Karst landscape map from Hartmann et al. (2015, 2017).

In this section, we present the results in terms of simulated recharge only, even though these results could also be interpreted in terms of simulated ET. In fact, we find that simulated runoff with V2Karst is negligible in most simulations in that, for all four karst landscapes, mean annual runoff is less than 1% of mean annual precipitation in 99% of the simulations. Therefore, virtually all precipitation that does not become recharge is evaporated. This can be explained by the fact that runoff occurs in V2Karst only when all model compartments (both for the soil and epikarst layers) are saturated. It is indeed assumed that the soil is highly permeable and that lateral flow can redistribute all saturation excess from the saturated compartments to the unsaturated compartments.

We analyse two aspects of recharge, namely the mean annual recharge denoted as Q_{epi} $[mm.y^{-1}]$ and the recharge ratio denoted as Q_{epi}/P $[-]$, which is the ratio of mean annual recharge over mean annual precipitation (e.g. Kim and Jackson, 2012; Jasechko et al., 2014). Recharge ratio describes the partitioning of precipitation between ET and recharge.

4.3.1 Relative impact of climate variability, land cover type and degree of subsurface heterogeneity on recharge

Figure 4.2.b shows the variance decomposition among the four input factors examined i.e. climate variability, land cover type, degree of subsurface heterogeneity and parameter uncertainty, and the interactions among these factors, obtained from application of the VBSA method for the four karst landscapes. We note that the variance fractions were derived using all ISI-MIP climate scenarios (historical period, i.e. 1974-2004, and all RCPs, i.e. 2069-2099). The figure reveals that the sensitivity of groundwater recharge and recharge ratio to the four factors varies significantly across the karst landscapes. Given that the bootstrap confidence intervals around the sensitivity indices are very narrow, the estimated order of importance among the input factors can be considered very robust.

From the top panel of Fig. 4.2.b, we see that groundwater recharge (Q_{epi}) is largely controlled by climate variability, which accounts for 72% to 85% of the total variance, but that additional factors play a significant role depending on the landscape. The impact of land cover is important for wetter landscapes (humid and mountain) and decreases when moving towards drier landscapes (Mediterranean and desert). The relative effect of parameter uncertainty appears to be more pronounced for desert areas, while it is very small for the other landscapes. The degree of subsurface heterogeneity only explains a small fraction of the variance for all landscapes.

The recharge ratio exhibits different sensitivities compared to Q_{epi}, as shown in the bottom panel of Fig. 4.2.b. The relative impact of climate on Q_{epi}/P reduces greatly compared to Q_{epi}, while the effects of

the other factors intensify. This is reasonable given that the effect of climate is partially removed in the very definition of the recharge ratio due to the normalisation by precipitation. Climate variability explains between 44% and 62% of the variance of Q_{epi}/P. The effect of land cover decreases moving from wetter to drier landscapes and, interestingly, it is as large as the effect of climate for the humid landscape. Parameter uncertainty in desert areas gains importance for Q_{epi}/P (20-24% of total variance) compared to Q_{epi} (4-8% of total variance). Unlike for Q_{epi}, subsurface heterogeneity is found to have significant effect on Q_{epi}/P, and particularly in the drier landscapes (Mediterranean and desert) where its influence is similar to that of land cover.

The fact that the impact of the different input factors varies across karst landscapes suggests that these factors interact across the domain. However, the variance decomposition in Fig. 4.2.b does not allow us to draw robust conclusions about the importance of factor interaction within landscapes because the confidence intervals of the interactions' contributions are very large.

An additional correlation analysis allows identification of the specific climate properties and model parameters that have the largest effect on recharge. Figure 4.3 reports the values of the Spearman rank correlation coefficient for selected model parameters and climate properties, namely precipitation intensity I_p $[mm.d^{-1}]$, mean annual precipitation P_m $[mm.y^{-1}]$ and aridity index $AI[-]$ (see definitions and further technical details in caption of Fig. 4.3). The correlation coefficients for the other model parameters, that have a negligible effect on recharge, and for a range of additional climate descriptors are reported in Appendix C.4.

The correlation analysis corroborates the results of the variance decomposition, in that it shows similar relative impact among the different groups of input factors. Once again, climate properties are the most influential factors, followed by land cover, subsurface heterogeneity and parameter uncertainty, depending on the landscape under study. From the left panels of Fig. 4.3 we further observe that, among climate properties, P_m and AI have a larger effect on Q_{epi} compared to I_p but also that the relative importance of I_p for Q_{epi} is larger for drier landscapes (Mediterranean and desert) compared to wetter landscapes (humid and mountain). By comparing the left and right panels of Fig. 4.3, we also see that the relative importance of I_p compared to P_m and AI is larger for Q_{epi}/P compared to Q_{epi} and that in particular I_p appears to be the most influential factor on Q_{epi}/P for the desert landscape. The correlation coefficient of I_p is always positive, which means that higher values of I_p tend to produce higher values of Q_{epi} and Q_{epi}/P.

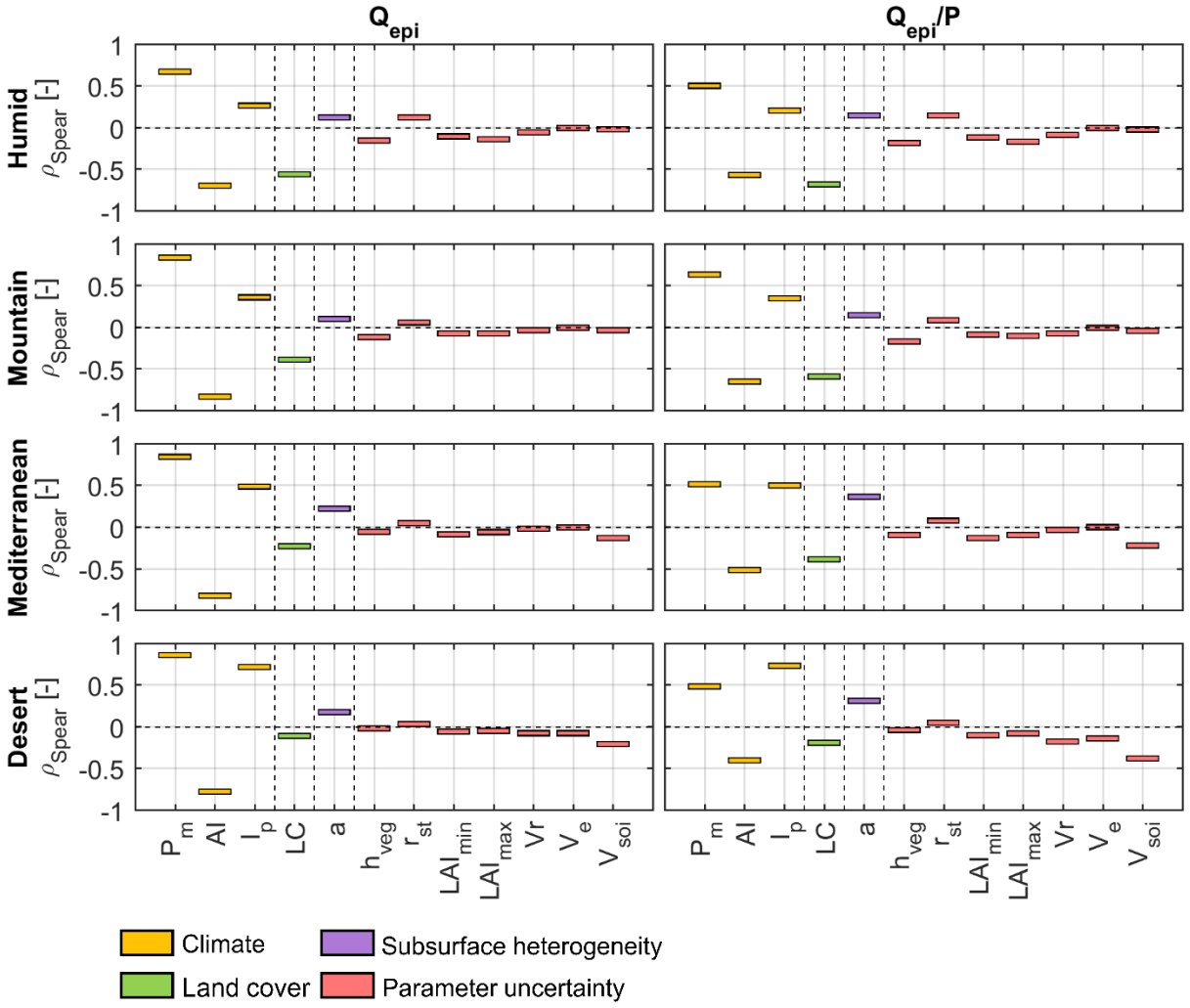


Figure 4.3 Spearman correlation coefficient ρ_{spear} $[-]$ between selected model input factors and simulated mean annual recharge and simulated recharge ratio for the four karst landscapes.
Q_{epi} $[mm.y^{-1}]$mean annual recharge, Q_{epi}/P $[-]$, recharge ratio, $P_m[mm.y^{-1}]$ mean annual precipitation, AI $[-]$ aridity index ($AI = PET_m/P_m, PET_m$[mm.y^{-1}] mean annual PET), I_p $[mm.d^{-1}]$ precipitation intensity (calculated over days with precipitation totals ≥ 1mm), LC land cover (which takes a value of 0 for grass/crop and of 1 for forest to compute the correlation coefficient), a spatial variability coefficient, h_{veg} vegetation height, r_{st} stomatal resistance, LAI_{min} reduction in leaf area index during the dormant season, LAI_{max} annual maximum leaf area index, V_r maximum storage capacity of the rooting zone, V_e maximum storage capacity of the first soil layer, V_{soi} mean soil storage capacity (model parameters are defined in Table 4.1). We note that since the ranges of the vegetation parameters depend on the land cover type, to assess the effect of the uncertainty in vegetation parameters (i.e. residual uncertainty in parameter values for a given land cover), we calculated the correlation coefficient for the vegetation parameters for the two land cover types (forest and grass/crop) separately and the figure reports the maximum values over the two land cover types. Correlation results for a range of additional climate descriptors and for the remaining V2Karst parameters are reported in Appendix C.4.

Regarding the effect of parameter uncertainty, we find again from Fig. 4.3 that it mainly has an impact in the desert landscape. In Fig. 4.3 we further find that, among the various soil parameters, it is the mean soil water capacity V_{soi} that has the strongest correlation with Q_{epi} and Q_{epi}/P. Moreover, the correlation coefficient of V_{soi} is negative, which means that higher values of V_{soi} reduces recharge and enhance ET (more moisture can be stored in the soil and made available to plant transpiration and soil evaporation).

4.3.2 Combinations of climate properties, land cover type and degree of subsurface heterogeneity for producing more or less recharge

The CART technique allows for further investigating the effect of the different parameters and climate properties and to identify combinations of climate properties, land cover type, degree of subsurface heterogeneity and parameter values that produce more a less recharge. Figures 4.4 and 4.5 visualise the classification tree for Q_{epi} and Q_{epi}/P respectively. We report climate descriptors that were selected by the CART algorithm only (Fig. 4.4.e and 4.5.e), while additional descriptors included in the analyses are presented in Appendix C.3. As for Sect. 4.3.1, the trees were derived using all ISI-MIP climate scenarios (historical period, i.e. 1974-2004, and all RCPs, i.e. 2069-2099).

Definition of recharge classes

Both classification trees were built with respect to four classes of recharge values. The definition of the classes for Q_{epi} (Fig. 4.4.c) is consistent with the WHYMAP dataset (Groundwater resources of the world, WHYMAP GWR © BGR & UNESCO 2015, BGR and UNESCO 2008) that provides a global map of mean annual recharge. We did not consider the lower class of values in the WHYMAP ($Q_{epi} < 2mm.y^{-1}$), since very few of our simulations fall in this category. We verified that the chosen classes reasonably capture the variability across the simulation domain (Fig. 4.4.d). We note that Fig. 4.4.d allows to qualitatively visualise the results obtained with VBSA and the correlation analysis, since it indicates differences in the classes of values of Q_{epi} for grass/crop compared to forest land cover (with higher values for grass/crop) and that these differences are more pronounced for the wetter landscapes (humid and mountain). Additionally, Fig. 4.4.d shows that, as expected, Q_{epi} is higher for wetter landscapes (humid and mountain,) where more moisture is available to produce recharge.

Regarding recharge ratio (Fig. 4.5), classes of values were chosen to describe the variability in Q_{epi}/P across the simulations (Fig. 4.5.c,d). Interestingly, Fig. 4.5.d shows no marked differences in the distribution in classes of Q_{epi}/P across landscapes as observed for Q_{epi} (Fig. 4.4). Our class definition differs significantly from the one used in Jasechko et al. (2014, Figure 1) to visualise global estimates

of Q_{epi}/P. Jasechko et al. (2014) considered seven classes for Q_{epi}/P values below 0.1, which all correspond to the class C1 defined here (Fig. 4.5.c); on the other hand, their highest class comprises all Q_{epi}/P values above 0.25, while here we have three classes for values up to 0.3 (C2), between 0.3 and 0.5 (C3), and above 0.5 (C4). The differences in these definitions are presumably due to the fact that our estimates of Q_{epi}/P are much higher, and that the classification chosen by Jasechko et al. (2014) aims at capturing the global variability in Q_{epi}/P without a specific focus on carbonate rock regions, while here we constrained our simulation domain to carbonate regions of Europe, the Middle East and Northern Africa.

Classification performance

The trees were pruned i.e. they were reduced in size by removing nodes further down the trees, using the Matlab Statistics and Machine Learning Toolbox (Mathworks, 2016). Since the objective here is not to build a prediction model for recharge but to identify the controls of recharge, we chose a level of pruning that lead to a reasonable number of leaf nodes, so that the trees are easily interpretable. The classification trees selected for Q_{epi} and Q_{epi}/P include 22 and 23 leaf nodes respectively and are characterised by a classification performance of 74.6% and 64.3% respectively. We note that the classification performance for Q_{epi}/P is lower than for Q_{epi}. Our previous analyses in Sect. 4.3.1 showed that Q_{epi} is largely controlled by climate properties (and in particular by P_m and AI), while the controls of Q_{epi}/P are more equally spread over the different input factors. Therefore, it is not surprising that Q_{epi}/P is more difficult to predict than Q_{epi}. In the following paragraphs, we analyse the dominant classes that emerge in the different leaf nodes.

Analysis of the controls of recharge and their combinations for producing more or less recharge

The climate descriptors selected by CART algorithm are: the aridity index AI $[-]$ and the mean precipitation intensity I_p $[mm.d^{-1}]$ for both trees; the mean annual precipitation P_m $[mm.y^{-1}]$, the precipitation seasonality index P_{SI} $[-]$ and the correlation coefficient between monthly precipitation and PET ρ_{P-PET} $[-]$ for the tree of Q_{epi}; and the mean precipitation duration D_p $[d]$ for the tree of Q_{epi}/P. It is important to note that all these climate descriptors generally exhibit low pairwise linear correlations (reported in Table C.6), apart from D_p and P_{SI} that are correlated with P_m (linear correlation coefficient equal to 0.84 for D_p and to -0.62 for P_{SI}). This result gives us confidence that, apart from D_p and P_{SI} , the climate descriptors are likely to appear in the tree because of their own effect on recharge, and not as a side-effect of their correlations.

The classification tree for Q_{epi} (Fig. 4.4) suggests that P_m dominates the model response as it appears in the first split of the tree. The fact that P_m is also present in lower levels of the tree suggest that it also interacts with the other input factors. Three main precipitation regimes can be identified from Fig. 4.4: a first regime for low P_m values ($P_m < 303\ mm.y^{-1}$) in which Q_{epi} also takes low values (C1 and C2 classes); a second regime for intermediate values of ($303 \leq P_m < 723\ mm.y^{-1}$) in which Q_{epi} can be in any class, depending on a range of other factors including other climate descriptors; and a third regime for high P_m values ($P_m \geq 723\ mm.y^{-1}$) in which Q_{epi} tends to be high (mostly C3 and C4 classes but also C2 class for few simulations).

The land cover type appears in the second split of the tree for the intermediate and the high precipitation regimes, which suggests that it is the second most important control on Q_{epi} after P_m. Generally, higher values of Q_{epi} occur for grass/crop compared to forest land cover. In particular, when land cover is forest, very high Q_{epi} values (class C4) only occur for a very small number of simulations (< 5%) corresponding to high P_m ($\geq 723\ mm.y^{-1}$) and very low aridity index ($AI < 0.4$), which means that PET is very limited compared to P_m, i.e. the conditions are energy limited and ET losses are reduced. On the contrary, when the land cover is grass/crop, Q_{epi} never takes very small values (C1 class) and only takes small values (C2 class) in less than 5% of the simulations, corresponding to low-intermediate P_m ($303 \leq P_m < 499\ mm.y^{-1}$), and high AI (≥ 1.46), i.e. the conditions are water limited and favourable to ET losses.

In the lower precipitation regime (first branches from the left in Fig. 4.4), I_p is the second control after P_m. Here, high values of I_p ($\geq 5.9\ mm.d^{-1}$) ensure that Q_{epi} never falls into the C1 class, while lower values of I_p tend to produce lower recharge values. This positive effect of I_p and its larger importance under drier conditions are in line with the results of the previous correlation analysis (Fig. 4.3).

Figure 4.4 also shows that the degree of subsurface heterogeneity (a) and the soil water capacity (V_{soi}) play a role in moderating the effect of climate on Q_{epi} in both the lower and in the intermediate precipitation regimes for forest land cover. These two factors are particularly important in separating C1 (very small) from C2 (small) recharge values. Higher values of a and lower values of V_{soi} enhance recharge and reduce ET. The tree also illuminates critical thresholds on the value of a and V_{soi} and shows that Q_{epi} can be in the C2 class in the lower precipitation regime and for low values of I_p ($I_p < 5.9\ mm.d^{-1}$) under rather strict conditions: either when V_{soi} is very small ($< 16\ mm$), which can only occur in desert landscape (see Table 4.1), or when a is high (> 2.9), which only occurs in the Mediterranean and the desert landscapes.

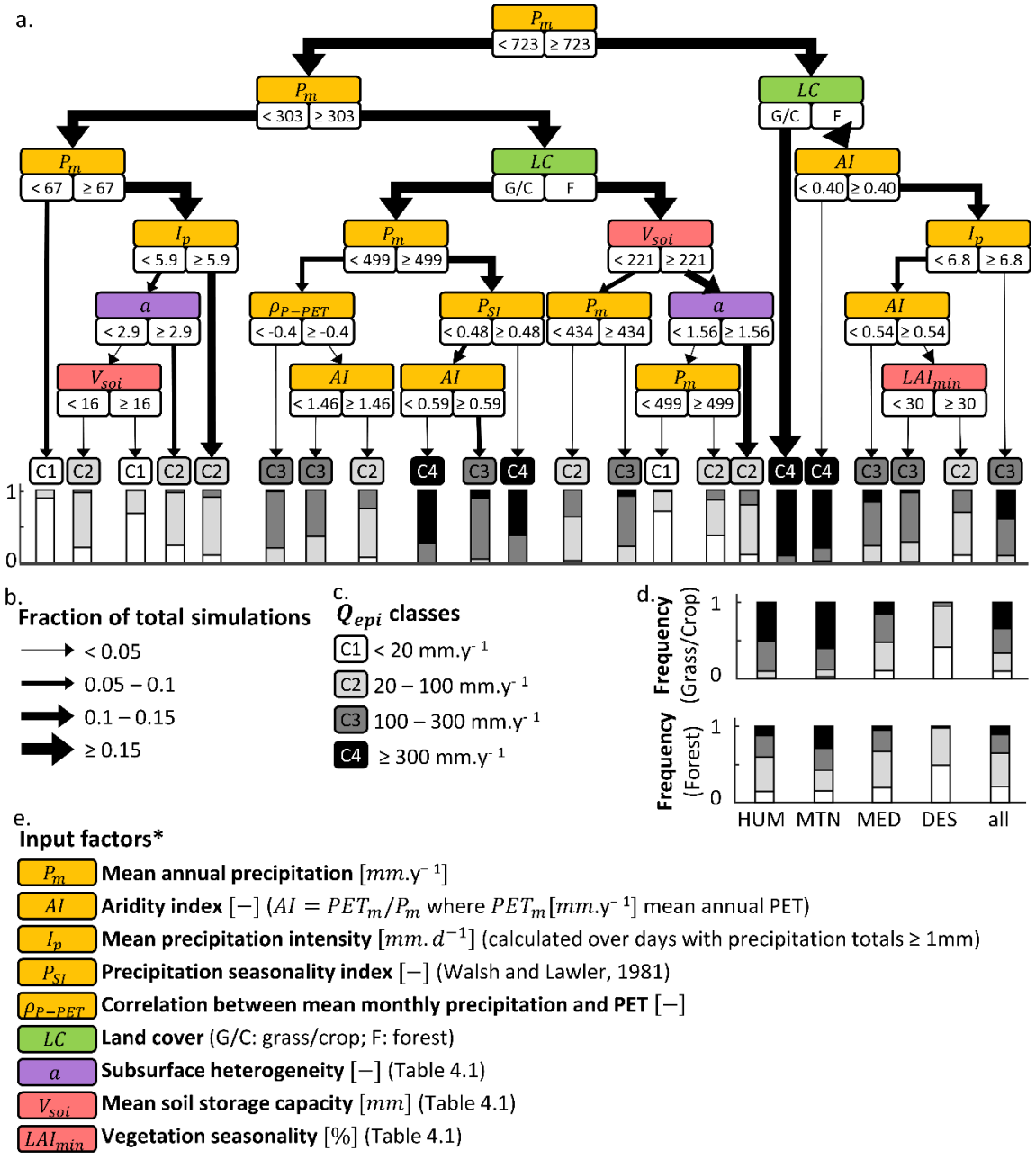


Figure 4.4 Classification Tree of simulated mean annual recharge.

Each internal node of the tree corresponds to one of the input factors (the inputs are coloured as in with Fig. 4.2 and 4.3). The two branches leaving each internal node indicate a binary split based on the input factor value. The thickness of the arrow represents the fraction of total simulations that fall into a given branch (b). The tree was built based on four classes of mean annual recharge (Q_{epi}) values (c). Each terminal node shows the predominant recharge class and the frequencies of the four recharge classes at that node. We report the frequencies of the four recharge classes for the four karst landscapes (HUM humid, MTN mountain, MED Mediterranean, DES desert) and all the domain, and for grass/crop and forest land cover (d). (*) Potential evapotranspiration (PET) required to assess the aridity index *AI* $[-]$ was computed using the Penman-Monteith equation (Monteith, 1965) using the properties of a reference crop reported in Allen et al. (1998). In this way, the variability in *AI* captures the variability in evaporative demand due to climate conditions only. Further details on the definition of the climate descriptors and the full list of climate descriptors included in the analyses is reported in Appendix C.4.

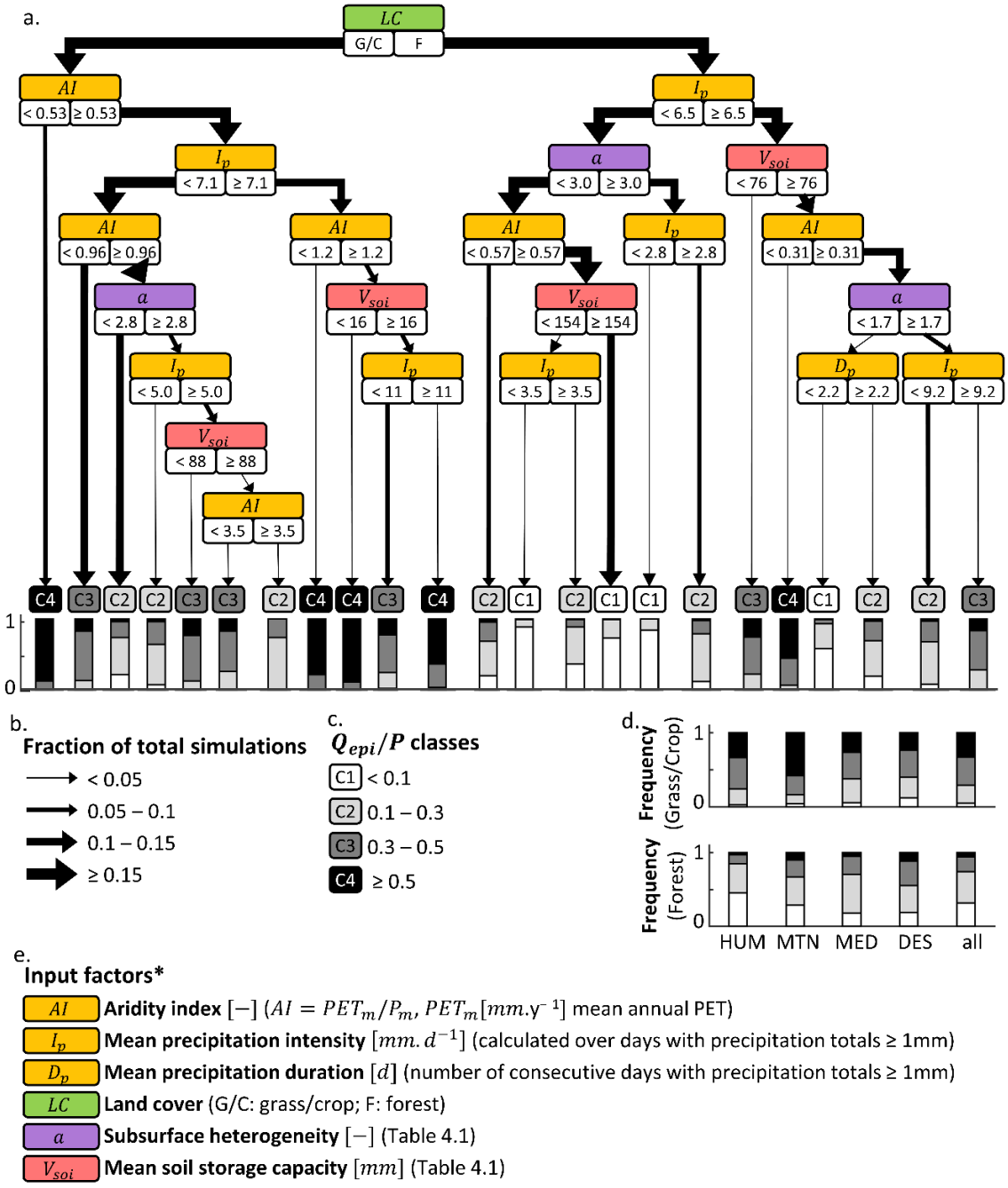


Figure 4.5 Classification Tree of simulated recharge ratio.

The tree was built based on four classes of recharge ratio (Q_{epi}/P) values (c). We refer to the caption of Fig. 4.4 (Classification Tree of simulated mean annual recharge) for additional information.

(*) Potential evapotranspiration (PET) required to assess the aridity index *AI* [−] was computed using the Penman-Monteith equation (Monteith, 1965) using the properties of a reference crop reported in Allen et al. (1998). In this way, the variability in *AI* captures the variability in evaporative demand due to climate conditions only. Further details on the definition of the climate descriptors and the full list of climate descriptors included in the analyses is reported in Appendix C.4.

The tree of Q_{epi}/P (Fig. 4.5) is very different from the tree of Q_{epi} in that P_m is not selected. However, since D_p is included on the right end side of the tree, and given that D_p is highly correlated with P_m (as previously discussed), the presence of D_p in the tree indicates that P_m may be a relevant descriptor to explain Q_{epi}/P as well, which is consistent with our Spearman correlation analysis (Fig. 4.3 and Appendix C.4). The tree of Fig. 4.5 reveals that the main control of Q_{epi}/P is the land cover (first split) and highlights the importance of AI, I_p, a and V_{soi} in moderating the effect of the land cover and in determining the class of Q_{epi}/P.

In Sect. 4.3.3, we further disaggregate the sensitivity of groundwater recharge to land cover change for the different cells of the study domain separately. We expect such sensitivity to vary across the domain given that our analyses so far suggest that the effect of land cover depends on the interactions between climate properties, degree of subsurface heterogeneity and soil water capacity.

4.3.3 Sensitivity of recharge across the study domain

Historical mean annual recharge (1974-2004**)**

Figure 4.6 reports the simulated mean annual recharge Q_{epi} and its projected changes at the end of the 21st century for the study domain (average values over the simulation ensembles obtained by combining 5 GCMs, 4 RCPs and 100 parameter samples). Historical Q_{epi} (Fig. 4.6.a) was assessed using the historical climate scenarios from the 5 GCMs and the historical land cover map of forest and grass/crop fraction (data are described in Sect. 4.2.2). Total recharge for each cell over both forest and grass/crop fraction was assessed following the 'tile' approach described in Sect. 4.2.1. We compared the recharge map of Fig. 4.6.a to the recharge values of the WHYMAP that was briefly introduced in Sect. 4.3.2. The WHYMAP was established by evaluating the WaterGap model (Döll et al., 2003), which is a large-scale hydrological model that does not represent karst processes, over the period 1961-1990. We found an agreement in the recharge class between the two datasets for about 30% of the cells, while for almost all the remaining 70% of the cells, V2Karst produces higher recharge classes compared to the WHYMAP. In particular, the level of agreement is higher over the desert landscape (43.7%), where recharge values tend to be lower compared to the other landscapes in both datasets, because of limited moisture availability. The lower bound of the 95% confidence interval of recharge simulated with V2Karst is more consistent with the WHYMAP (44% of agreement across the domain) but is still largely higher. Further details on this comparison, and the lower and upper of bound of recharge simulated with V2Karst and the WHYMAP are reported in Appendix C.5.

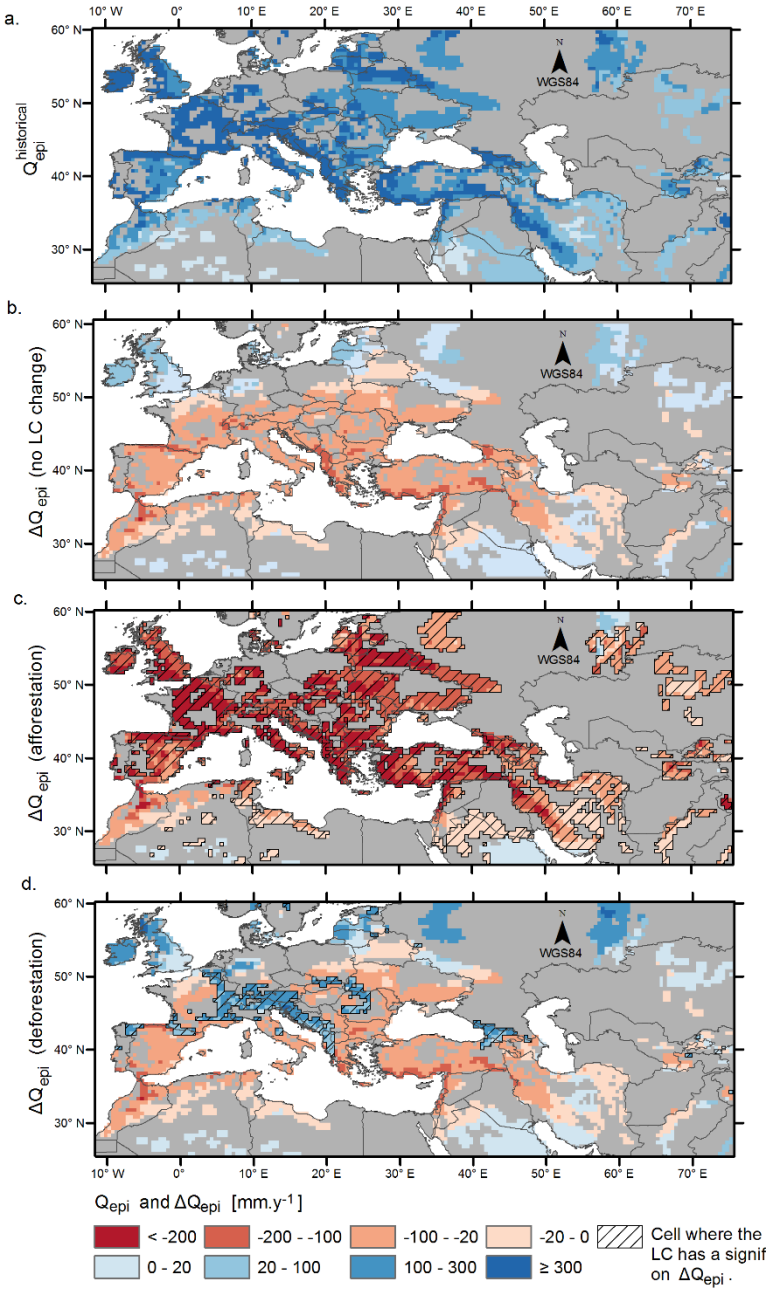


Figure 4.6 Historical mean annual recharge for the period 1974-2004 (top) and grid cell sensitivity of recharge to combined climate change and land cover change for three different land cover change scenarios (no change, afforestation, deforestation) from the period 1974-2004 to 2069-2099 in carbonate rock areas in Europe, the Middle East and Northern Africa.

(a) Historical mean annual recharge $Q_{epi}^{historical}[mm.y^{-1}]$, (b) projected change $\Delta Q_{epi} = Q_{epi}^{future} - Q_{epi}^{historical}$ $[mm.y^{-1}]$ assuming no change in land cover (LC), (c) assuming complete afforestation and (d) assuming complete deforestation. The figure reports the mean values over the simulation ensemble (lower and upper bounds are reported in Appendix C.5). Hatched cells show a significant difference in ΔQ_{epi} between the 'no land cover change' scenario and the relevant land cover change scenario (afforestation or deforestation). Source of the data: Carbonate rock and country map from Williams and Ford (2006) and in particular the country map was obtained from Terraspace, Russian space agency.

Changes in mean annual recharge in response to change in climate and land cover

We examine changes in mean annual recharge ΔQ_{epi} under changes in climate (from 1st October 1974 - 30th September 2004 to 1st October 2069 - 30th September 2099) and under three different land cover change scenarios, namely: (1) no change in land cover, (2) complete afforestation, i.e. the current grass/crop fraction of all cells becomes forest, and (3) complete deforestation, i.e. the forest fraction of all cells becomes grass/crop. By considering these three extreme scenarios, we aim at exploring the envelope of change in land cover. We note that a complete afforestation or deforestation of all grid cells of the domain simultaneously is highly unlikely. The results of this analysis, which are reported in Fig. 4.6, show the sensitivity of the different grid cells to changes in climate and land cover and are illustrative of what could happen at the scale of the grid cell.

The 'no land cover change' scenario (Fig. 4.6.b) shows a decrease in Q_{epi} ($\Delta Q_{epi} < 0$) for about two thirds of the cells. Increases in Q_{epi} mainly occur over the northern part of the domain, the Middle East and Northern Africa and are generally no higher than 100mm. These values are the mean over the simulation ensemble (i.e. over all combinations of the RCPs, GCMs and parameter sets), however an uncertainty analysis reveals that for as many as 92% of the cells, Q_{epi} can either increase or decrease in the simulation ensemble (uncertainty bounds are reported in Appendix C.6). In particular, under the 'pessimistic' case (lower uncertainty bound), Q_{epi} decreases in virtually all cells.

The afforestation scenario (Fig. 4.6.c) results in values of ΔQ_{epi} that are mostly negative (for 94% of the cells) and that are always lower than the values obtained for the 'no land cover change' scenario for all cells. This is because higher forest fractions enhance ET losses. We estimate that for 84% of the cells (hatched areas in Fig. 4.6.c), the effect of afforestation is significant relative to the effect of climate change. In fact, for 27% of the cells, afforestation modifies the direction of change: ΔQ_{epi} is positive when land cover does not change but is negative when afforestation occurs. For 57 % of the cells, ΔQ_{epi} is negative under both land cover scenarios, but under the afforestation it more than twice the value obtained in the 'no land cover change' scenario. This means that afforestation produces an additional effect on Q_{epi} which is higher than the effect of climate change alone. Therefore, the cells that are hatched in Fig. 4.6.c can be considered highly sensitive to afforestation under future 'average conditions' (i.e. simulation ensemble mean). These cells include most of the domain, apart from some southern areas that mostly belong to the desert or Mediterranean landscape and that show little response to changes in land cover (consistently with our previous analyses in Sect. 4.3.1 and 4.3.2), and apart from areas located in the north-east part of the domain where the forest fraction is currently already very high (Fig. 1.4 in Chapter 1). With respect to the lower uncertainty bound of ΔQ_{epi} ('pessimistic' case, Fig. C.6), the impact of afforestation is still significant relative to climate for about half of the cells hatched in Fig

4.6.c. In fact, for these cells the value of ΔQ_{epi} under the afforestation scenario is more than twice the values of ΔQ_{epi} under the climate change only scenario. Regarding the upper uncertainty bound of ΔQ_{epi} ('optimistic' case, Fig. C.7) for over half of the cells hatched in Fig. 4.6.c the impact of afforestation is still critical because afforestation produces a change of sign for ΔQ_{epi}.

Finally, the deforestation scenario (Fig. 4.6.d) affect 20% of the cells only, since the remaining 80% of the cells are currently covered by grass/crop vegetation only (Fig. 1.4). For about half of the currently forested cells (hatched cells in Fig. 4.6.d), deforestation reverses the trend in ΔQ_{epi}, i.e ΔQ_{epi} is negative under the 'no land cover change' scenario while it becomes positive under the deforestation scenario. Regarding the lower uncertainty bound of ΔQ_{epi} ('pessimistic' case, Fig. C.6), we further found that deforestation produces a change in the sign of ΔQ_{epi} for over half of the forested cells, but that these cells are mostly different to the ones hatched in Fig. 4.6.d.

This analysis shows that, for most cells of the domain, afforestation and deforestation (for cells that are partially or fully forested) could have a significant effect on the direction of change and magnitude of mean annual recharge relative to the effect of climate change.

4.4 Discussion

To identify the controls of groundwater recharge in carbonate rock areas in Europe, the Middle East and Northern Africa, we simulated groundwater recharge using the integrated vegetation-recharge V2Karst model for karst areas. We jointly analysed the model inputs (climate projection, virtual land cover scenarios, degree of subsurface heterogeneity and parameter uncertainty) and output (mean annual recharge and recharge ratio) dataset using Global Sensitivity Analysis (GSA). We found consistent sensitivities across the different GSA methods even though these rely on different assumptions, which attests to the robustness of our results. The different GSA methods also proved to be complementary to derive information from the input-output dataset in terms of the quantitative relative impact of the different input factors and their percentage contribution to the variability of recharge, in terms of the direction of change in recharge associated with the variations in the input factors, and in terms of specific ranges of values of the input factors and their combination in producing more or less recharge.

Sensitivity of recharge to the degree of subsurface heterogeneity

Our sensitivity analyses demonstrated that the higher the degree of subsurface heterogeneity, the higher the mean annual recharge and the higher the recharge ratio. This result is in line with an earlier study by Hartmann et al. (2017) using a previous version of the V2Karst model for the same simulation domain,

that showed that mean annual recharge values produced using a heterogeneous representation of the subsurface tend to be much higher compared to a homogeneous representation. We also found that subsurface heterogeneity can alter the impact of climate and land cover on recharge and that the effect of subsurface heterogeneity is more pronounced under drier conditions and for forest land cover. Hartmann et al. (2017) also found that the representation of subsurface heterogeneity modifies recharge sensitivity to changes in climate, and in particular they demonstrated that climate change leads to much higher absolute decreases in recharge under subsurface heterogenous conditions compared to homogeneous conditions for the most extreme future climate scenarios (RCP8.5, Van Vuuren et al., 2011). Here, we complemented the results of Hartmann et al. (2017) by assessing the relative impact of subsurface heterogeneity with regards to climate properties and land cover and by establishing how these three input factors combine to determine recharge values.

The comparison of current mean annual recharge simulated with V2Karst and with the WaterGap (WHYMAP), which is a hydrological model that does not include an explicit representation of subsurface heterogeneity, showed that V2Karst tends to produce higher recharge values. This results further support the results of our sensitivity analyses and the conclusion of the study by Hartmann et al.(2017) that subsurface heterogeneity has a significant effect on recharge. However, part of the deviations observed between V2Karst and WaterGap predictions could be also explained by differences in the set-up of the two models to perform the simulations (e.g. the forcing data used).

Sensitivity of recharge to precipitation intensity

The CART and Spearman correlation analysis suggest, apart from mean annual precipitation and aridity index, that mean precipitation intensity is an important controlling factor and that the higher the mean precipitation intensity, the higher the recharge and the recharge ratio. We note that this study used climate projections, which provide consistent and physically constrained climate scenarios, but this also makes it more difficult to disentangle the impacts of the different climate properties because these properties may be involved in complex interactions among each other. A previous idealised study using a synthetic periodic precipitation input and constant value of the other climatic inputs to force the V2Karst model (Chapter 3; Sarrazin et al., 2018) unequivocally demonstrated that precipitation intensity has a significant positive effect on recharge. This previous idealised study provides further confidence regarding the validity of the results obtained in this study. An important and positive effect of precipitation intensity on recharge was also found in observational studies in non-karst areas, namely in a seasonally humid tropical region in the upper Nile basin in Owor (2009) and in a semi-arid tropical region in East Africa in Taylor et al. (2013a). Surprisingly, Fig. 3 in Hartmann et al. (2017) reveals that precipitation intensity can have not only a positive but also a negative effect on recharge. This

inconsistency between the results of Hartmann et al. (2017) and ours may be due to the fact that the former study did not trade space for time and assessed the sensitivity of recharge for each grid cell separately. The previous approach explores a limited number of climate conditions for each grid cells (i.e. the climate projections for that particular cell), which may not permit to reliably isolate the effect of precipitation intensity from the effect of the other climate properties.

Sensitivity of recharge to land cover type

Our sensitivity analyses based on the 'trade space for time' strategy showed that land cover appears as the secondary control of mean annual recharge after mean annual precipitation, apart from drier regions where the effect of land cover is limited. The importance of land cover is also more pronounced for recharge ratio. The differences in recharge sensitivity to land cover observed across the karst landscapes can be explained by the fact that wetter landscapes have deeper soils and tend to have energy-limited conditions, hence the factors that control the energy available for evapotranspiration (and therefore vegetation properties) may play a larger role compared to drier landscapes. The latter have shallower soils and tend to have water-limited conditions, and hence factors that control the soil storage capacity (ability of the soil to hold water to be evapotranspired during dry periods) may play a larger role. The virtual experiments of Chapter 3 (Sarrazin et al., 2018) used synthetic climatic and land cover change scenarios (a change between grass and forest) to force the V2Karst model. These virtual experiments also found the land cover has an important impact on recharge and that this impact is more pronounced under wetter conditions. The results of the present study complement these previous virtual experiments and confirms the importance of vegetation using a more realistic climate forcing.

The sensitivity of groundwater recharge to land cover revealed by this research is in line with previous studies that analysed the effect of vegetation on groundwater recharge in non-karst areas. Mohan et al. (2018) examined a global database of recharge observations and found that mean annual precipitation is the first control of mean annual recharge followed by PET and land cover/use. However, they did not analyse the differences in recharge sensitivity to land cover/use across different climate types. Keese et al. (2005) conducted a modelling study along a precipitation gradient in the south western USA and found large decreases in recharge when including vegetation in the simulations compared to a non-vegetated model configuration. The global and observation-based study of Kim and Jackson (2012) established that vegetation type is overall the second control of recharge after water input (sum of precipitation of irrigation). The latter study found a significant effect of vegetation along the entire climate gradient. It also revealed interactions among vegetation, soil types and climate in producing recharge.

Moreover, when analysing the sensitivity of recharge to land cover change for the different cells of domain separately, and therefore when constraining the climate variability, we showed that for most cells a change in land cover could have a significant effect on recharge compared to the effect of climate change, including some arid cells (desert landscape). The studies of Keese et al. (2005) and Kim and Jackson (2012) also found that a change in land cover could have a significant effect on recharge in arid and semi-arid regions.

Additionally, previous studies that analysed the control of ET using results from paired catchment experiments also found an effect of vegetation on ET consistently with the present study. Our results qualitatively compare with the empirical relationship derived by Zhang et al. (2001) between mean annual precipitation, mean annual ET and land cover (forest and grass). These 'Zhang curves' show that ET both grass and forest is very similar when precipitation is low and that the higher the precipitation, the higher the discrepancy in ET between the two vegetation types. These curves were tested against observations in Zhang et al. (2001) and in the review of paired catchment studies of Brown et al. (2005). The study by Peel et al. (2010) also highlighted that the importance of vegetation on ET strongly varies across climate types.

4.5 Conclusions

In conclusion, we find that the degree of subsurface heterogeneity, the precipitation intensity and the land cover type are important controls of recharge beyond mean annual precipitation and aridity index. Our study has three main implications that are described below.

First, karst processes should be included in model representation to produce plausible sensitivities of groundwater recharge to climate and land cover change. More research is also needed to further constrain the degree of subsurface heterogeneity in the different landscapes of the domain, in particular for the desert landscape in which subsurface heterogeneity was shown to be particularly influential, while the range of variation of subsurface heterogeneity is poorly constrained in this landscape.

Second, changes in precipitation intensity should be included in future climate and land cover impact studies on karst groundwater. This is critical because significant trends in mean daily precipitation intensity, either positive or negative, have already been observed over parts of the domain (Fig. 2.33 in IPCC, 2013) and precipitation intensity was shown here to be an important control of groundwater recharge. Efforts should be directed towards reducing the uncertainty in precipitation intensity in climate projections to produce more realistic recharge projections, since precipitation projection largely suffers for biases (Chan et al., 2018).

And third, to ensure future water security, future land cover management strategies should careful consider the sensitivity of karst systems to changes in land cover. Although the present study analysed extreme changes in land cover (total afforestation or deforestation), it suggests that changes in vegetation type could potentially have a critical effect on groundwater recharge in carbonate rock regions in Europe, the Middle East and northern Africa.

Chapter 5. Summary and outlook

Section 5.1 of this chapter is adapted from the abstracts of three works that are published, in review or in preparation:

1. *Sarrazin, F., Pianosi, F. and Wagener, T.: Global Sensitivity Analysis of environmental models: Convergence and validation, Environmental Modelling and Software, 79, 135–152, doi:10.1016/j.envsoft.2016.02.005, 2016.*
2. *Sarrazin, F., Hartmann, A., Pianosi, F. and Wagener, T.: V2Karst v1.0: A parsimonious large-scale integrated vegetation-recharge model to understand the impact of climate and land cover change in karst regions, Geoscientific Model Development Discussion, doi:10.5194/gmd-2017-315, in review.*
3. *Sarrazin F, Hartmann A, Pianosi F, Rosolem R, Wagener, T.: How do land cover and subsurface heterogeneity modulate climate change impacts on future groundwater recharge in karst areas?, in preparation.*

5.1 Summary of this research

The overall aim of thesis was to improve our understanding of the impact of changes in climate and land cover on karst groundwater recharge at large-scales and over large domains, which is critically needed to inform water management and ensure future water security in karst regions. To address this issue, this study investigated and developed methodologies and a hydrological model (Chapter 2 and 3), that were then applied over carbonate rock areas in Europe, the Middle East and Northern Africa to assess the sensitivity of recharge to changes in climate and land cover in this region (Chapter 4). More specifically, this thesis makes three main contributions: (1) it proposes novel criteria to assess the robustness of Global Sensitivity Analysis results, (2) it introduces a new model to study the impact of climate and land cover change on karst groundwater recharge and (3) it assesses the sensitivity of karst groundwater recharge to climate and land cover change in Europe, the Middle East and Northern Africa.

Novel criteria to assess the robustness of Global Sensitivity Analysis results

Global Sensitivity Analysis (GSA) is a crucial tool for guiding model development and understanding the controls of the output of a model (e.g. groundwater recharge). GSA comprehends a range of different methods that are based on different assumptions and rationales. These methods can therefore serve different purposes, such as identifying the insensitive input factors of a model (screening) ordering the input factors according to their sensitivity (ranking) and identifying specific combinations of input factor values that produce a model outcome of interest to the GSA user (mapping).

However, GSA implementation involves a number of choices that need to be careful examined to ensure the robustness and credibility of the results. In this research, we addressed two critical choices in GSA: the choice of the sample size and the choice of the threshold for screening the input factors. Guidance to assist users with those two choices is still insufficient and we aimed at filling this gap. Firstly, we defined criteria to quantify the convergence of the value of the sensitivity indices, of the ranking and of the screening, based on a bootstrap approach. Secondly, we investigated the screening threshold with a quantitative validation procedure for screening results. We applied the proposed methodologies to three hydrological models with varying complexity utilising three widely-used GSA methods, namely Variance-Base Sensitivity Analysis (VBSA), Elementary Effect Test (EET, or method of Morris) and Regional Sensitivity Analysis (RSA).

We demonstrated that convergence of screening and ranking can be reached before sensitivity estimates stabilise. Convergence dynamics appear to be case-dependent, which suggests that “fit-for-all” rules for sample sizes should not be used. Other modellers can easily adopt our criteria and procedures for a wide range of GSA methods and cases. The convergence criteria have been used in the sensitivity analyses

conducted in this research to inform the development of a new hydrological model and to assess the sensitivity of simulated groundwater recharge to changes in climate and land cover in karst areas.

A new model to study the impact of climate and land cover change on karst groundwater recharge

Karst aquifers are an important source of fresh water for agricultural activities and domestic use in many regions of the world. Karst areas are highly permeable and produce large amounts of groundwater recharge, while surface runoff is typically negligible. As a result, recharge in these systems may have a different sensitivity to climate and land cover changes compared to other less permeable systems.

Yet, little effort has been directed toward assessing the impact of climate and land cover change in karst areas at large-scales. In this research, we addressed this gap by (1) introducing the first large-scale hydrological model including an explicit representation of both karst and land cover properties, and by (2) analysing the model's recharge production behaviour. To achieve these points, we first improved the evapotranspiration estimation of a previous large-scale karst recharge model (VarKarst). Given the lack of data to constrain model simulations at a large-scale, the new model (V2Karst V1.0) includes a parsimonious representation of relevant ET processes for climate and land cover change impact studies. In this way, all model components can be subject to sensitivity and uncertainty analysis.

We demonstrated the plausibility of V2Karst predictions at carbonate rock FLUXNET sites using soft rules and GSA (RSA and EET methods). Then, virtual experiments with synthetic data confirmed that the model has sensitivities to precipitation (overall amount and temporal distribution) and land cover that are in line with previous studies. In addition, results allowed to quantify the relationship between changes in simulated recharge and changes in precipitation and land cover characteristics. Large-scale assessment of future karst groundwater recharge should therefore consider the combined impact of changes in land cover and precipitation properties if it is to produce realistic projections of changes in recharge.

Sensitivity of karst groundwater recharge to climate and land cover change in Europe, the Middle East and Northern Africa

A significant part of the population in Europe, the Middle East and Northern Africa relies on karst aquifers as a source of drinking water. Yet, poor water management has already led to the drying up of major karst springs in this region. Future changes in climate and land cover are expected to affect groundwater recharge, while human water demand is likely to increase due to population growth, thus potentially aggravating the issue of groundwater depletion in this region.

In this research, we aimed to assess the sensitivity of karst groundwater recharge to climate and land cover in Europe, the Middle East and Northern Africa. To this end, we simulated recharge using the V2Karst model, climate projection from the Inter-Sectoral Impact Model Inter-comparison Project (ISI-MIP) dataset and selected land cover scenarios (forest and grass/crop), while accounting for the variability in the degree of subsurface heterogeneity and the uncertainty in model parameters. The model input-output dataset so created was analysed using a 'trade for time' approach and GSA techniques (VBSA, Spearman correlation and Classification Tree) to uncover the controls of simulated recharge.

First, our results revealed that higher degrees of subsurface heterogeneity enhance recharge and can alter the impact of climate variability and land cover type, especially under drier conditions. Consequently, an appropriate representation of karst processes should be included in model representation to produce plausible sensitivities of recharge to climate and land cover change.

Second, beyond mean annual precipitation and aridity index, precipitation intensity was found to have a strong positive effect on recharge. Therefore, efforts should be directed towards reducing the uncertainty in precipitation intensity in climate projections to produce more realistic recharge projections.

Third, using the 'trade space for time' approach, land cover type was shown to have a significant impact on recharge relative to climate variability, in particular in wetter climates. An additional analysis, conducted for each cell of the domain separately, suggested that changes in land cover could potentially have an important effect on the sign or the magnitude of the projected change in recharge relative to the effect of the projected change in climate for many cells. To ensure water security in Europe, the Middle East and Northern Africa, we recommend that future land cover management strategies consider the sensitivity of karst groundwater recharge to changes in land cover.

This research provides crucial information on the sensitivity of karst groundwater recharge to changes in climate and land cover and on potential future changes in karst recharge in Europe, the Middle East and Northern Africa. Since groundwater recharge indicates the amount of renewable groundwater, this research characterises future water availability in karst areas. In addition, information on future water demand would be needed to identify critical recharge thresholds below which water demand cannot be met, and to ultimately guide decision and policy making to ensure that water demand can be satisfied.

5.2 Directions for future research

The present thesis focused on the assessment of the impact of climate and land cover change on the amount of groundwater recharge, i.e. on water quantity in karst systems, while neglecting the effect of future changes in atmospheric CO_2 concentration. Changes in atmospheric CO_2 concentration are known to affect both the state of the vegetation and the rate of weathering of the carbonate rocks, while changes in climate and land cover/use are also known to have a significant impact on water quality and the geomorphology of karst systems. All these mechanisms could, in turn, further affect karst water resources. Finally, as highlighted in Sect. 5.1, additional information on water demand in karst areas would be needed to guide decision and policy making. In this section, we further elaborate on these issues and we delineate directions for future research.

Effect of atmospheric CO_2 concentration on karst groundwater recharge

As discussed in Sect. 3.2.2, we did not consider the effect of CO_2 on the vegetation (specifically on plant fertilisation and stomatal resistance) in this research, given the limited state of knowledge on the effect of CO_2 and the large uncertainties related to the representation of the effect of CO_2 in hydrological models. However, previous studies suggest that future changes in CO_2 concentration could have a significant impact on vegetation water use and on the hydrological cycle (e.g. Ukkola et al., 2016; Lemordant et al., 2018) and therefore an in-depth investigation of the effect of CO_2 on plants and its uncertainty would be needed. This issue regards not only karst but also non-karst areas.

In addition, in karst areas, the projected increase in atmospheric CO_2 concentration is expected to cause an intensification in the weathering of carbonate rocks, with possible detectable feedback on the carbon cycle, i.e. a decrease in atmospheric CO_2 (Beaulieu et al., 2012; Cao et al., 2016). This process of weathering could affect preferential flow in karst systems and the dynamics and amount of groundwater recharge. However, it is not well understood how potentially higher dissolution rates of carbonate rocks will affect karst hydrology over the next centuries.

Water demand in karst areas

Estimating human water demand and water use involves large uncertainties, as discussed e.g. in Wada et al. (2016), where different methods and datasets for assessing water withdrawal from different sectors (agricultural, industrial and domestic) are reviewed and are shown to produce different results. Estimating future water demand is all the more challenging given the uncertainty in future demographic and socio-economic scenarios (Shared Socio-economic Pathways, SSP, Van Vuuren et al., 2014; United Nations, 2017), as discussed in Wada et al. (2016).

More specifically, in karst areas, an estimation of current groundwater withdrawal for human consumption is lacking. An assessment of the percentage of karst water in total water supply for the different European countries was performed at country scale more than twenty years ago (COST, 1995) and an up-to-date assessment would therefore be needed.

Environmental flow requirements are also important components of the water demand. These refer to the flow amount, its timing and duration necessary to maintain ecosystem health and ecosystem services (Pastor et al., 2014; Wada et al., 2016). Environmental flow requirements are also difficult to estimate and empirical methods are often used, as discussed in Pastor et al. (2014) and in Wada et al. (2016).

Water quality in karst systems

Karst aquifers are particularly vulnerable to contamination due to the presence of large subsurface apertures in karst systems (preferential flow). Contaminants can rapidly infiltrate to the groundwater and be transported in the aquifer to the springs, with little degradation through chemical or biological reactions, filtration, dilution and evaporation of volatile compounds (Kačaroğlu, 1999; Zwahlen, 2003; Coxon, 2011). As a result, pollution of karst aquifers has been a long standing issue as discussed for instance in Coxon et al. (2011) who reports numerous examples of karst groundwater contamination. In particular, as highlighted in Coxon et al. (2011), the fact that agricultural lands are widespread in karst areas increases the risk of pollution due to the use of fertilisers, manure and pesticides. Investigating contaminant transport is complex because of the large range of pollutants that can be found in agricultural systems with very different solubility, mobility and adsorption capacities (Kačaroğlu, 1999; Coxon, 2011). It is all the more complex in karst due to the high subsurface heterogeneity of these systems and the multiplicity of their flow mechanisms. Importantly, in karst, all types of contaminants seem to present a threat to water resources. Soluble contaminants (such as nitrate which is found in fertilisers) can be dissolved and can quickly infiltrate to the groundwater through preferential flow pathways (Coxon, 2011). The presence of larger apertures in karst underground enhances the transport of soil particles to the aquifer. This not only causes problem of turbidity in water, but also leads to the presence in water of those contaminants that are adsorbed to the soil particles (such as phosphorous, which is found in manure and fertiliser, and some pesticides) (Coxon, 2011).

Current approaches to the issue of groundwater contamination in karst systems typically consist of the assessment of the systems vulnerability through qualitative or semi-quantitative indicators that account for the properties of the karst systems and possibly the properties of the climate and the contaminants. Such approaches are reviewed for instance in Iván and Mádl-Szőnyi (2017). They have an empirical basis, they rely on strong simplifications of the geology and the hydrology of the karst systems and the vulnerability maps derived from these approaches tend to lack a sufficient validation (Foster et al., 2013;

Iván and Mádl-Szőnyi, 2017). Another limitation of these approaches is that very few of them consider precipitation intensity, such as the framework develop by the European Commission (COST action 620, Zwahlen, 2003), while observational evidence suggests a large impact of precipitation intensity on pollutant transport (Huebsch et al., 2014). Very few studies used simulation models to determine the fate of contaminants in karst systems (e.g. Jeannin et al., 2001; Butscher and Huggenberger, 2008, 2009; Hartmann et al., 2016). Yet, a simulation model approach allows for a more physically-based and quantitative assessment of aquifer contamination and the analysis of the dynamic relationship between pollutant transport, climatic conditions and land cover/use.

Erosion in karst systems

Karst soils are prone to erosion because they are often relatively shallow and the large subsurface apertures facilitate the transport of soil particles by the infiltrating water to deeper underground horizons (Coxon, 2011). Severe soil erosion can lead to rocky desertification, i.e. bedrock exposure, which is a widespread issue for instance in southwest China (Wang et al., 2004; Ford and Williams, 2007; Xiong et al., 2009; Jianhua et al., 2015). Karst rocky desertification is understood to be linked to changes in climate and in particular to increases in temperature and heavy rainfall (Xiong et al., 2009) and to changes in land cover such as deforestation (Wang et al., 2004). It has disastrous consequences for water security, because it reduces water retention at the surface, which leads to water shortages (Wang et al., 2004). Furthermore, deforestation in karst can lead to the infilling of the weathered bedrock by soil particles, which reduces the infiltration capacities of karst systems, decreases groundwater recharge and enhances surface runoff (Chandler and Bisogni, 1999). Therefore, soil erosion can have an important impact on the flow mechanisms in karst systems. However, few studies have used modelling approaches to quantitatively assess the effect of climate and land cover/use change on soil erosion in karst systems. An exception is the study by Febles-González et al. (2012) that analysed the effect of heavy precipitation on soil erosion in karst regions in Cuba.

Recommendations for future research

This research highlighted that the effects (and their representation in models) of future rising CO_2 atmospheric concentration on plant water use (as discussed here above and in Sect. 3.2.2) and of future changes in climate on snow dynamics (as discusses in Sect. 3.6.3) would require in-depth investigation in future studies, for both karst and non-karst areas.

Importantly, we also highlighted a number of open and poorly explored scientific issues specific to karst areas related to water quality and geomorphology (karsitification and soil erosion). We recommend that future studies investigate the sensitivity of both karst water quantity and quality to combined climate

change and land cover/use change, and their interactions with human activities, through modelling experiments at a large-scale. The effect of afforestation and deforestation on karst water resources needs to be examined in a broader modelling framework that considers not only the potential of forest to reduce water quantity through an increase in evapotranspiration, but also its beneficial effect on water quality and on soil protection against erosion that are highlighted for instance in Hamilton et al. (2008). The modelling framework would also benefit from a consideration of water demand and more broadly of human-water interactions (Wada et al., 2017), which are currently poorly characterised in karst areas. Such integrated studies would further inform future land cover/use management strategies, whose careful design will be crucial to meet future food and water demand (Foley et al., 2011). A major hindrance to such studies will be the paucity of data to constrain model simulations, since such data, as we have highlighted throughout this thesis, are particularly lacking for large-scale applications.

Appendix A. Supplements to Chapter 2

Appendix A is adapted from the supplemental material of a published work that has been partially modified to improve the consistency throughout the thesis:

> *Sarrazin, F., Pianosi, F. and Wagener, T.: Global Sensitivity Analysis of environmental models: Convergence and validation, Environmental Modelling and Software, 79, 135–152, doi:10.1016/j.envsoft.2016.02.005, 2016.*

The main article for this work supports Chapter 2.

A.1 Description of the model parameters for testing the methodology

Parameter	Definition	Module	Unit	Range
SM	Maximum soil moisture	Soil moisture	[mm]	70 - 400
BETA	Exponential coefficient in the soil routine	Soil moisture	[-]	0 - 0.95
ALPHA	Partitioning factor	Groundwater and routing	[-]	0 - 0.9
RS	Slow reservoir outflow coefficient	Groundwater and routing	[d^{-1}]	0 - 0.1
RF	Fast reservoir outflow coefficient	Groundwater and routing	[d^{-1}]	0.1 - 1

Table A.1 Parameters of the HyMod model considered for GSA.
We refined the ranges reported in Wagener et al. (2001) (details in Appendix A.2).

Parameter	Definition	Module	Unit	Range
TS	Temperature threshold for snow/melt	Snow	[°C]	-3 - 3
CFMAX	Degree day factor	Snow	[$mm.°C^{-1}\,d^{-1}$]	0 - 20
CFR	Refreezing coefficient	Snow	[-]	0 - 1
CWH	Water holding capacity of snow	Snow	[-]	0 - 0.8
BETA	Exponential coefficient	Soil moisture	[-]	0 - 7
LP	Limit Evapotranspiration	Soil moisture	[-]	0.3 - 1
FC	Field capacity	Soil moisture	[mm]	0 - 2000
PERC	Maximum percolation rate	Groundwater and routing	[$mm.d^{-1}$]	0 - 100
K0	Near surface flow coefficient	Groundwater and routing	[d^{-1}]	0.05 - 2
K1	Upper zone outflow coefficient	Groundwater and routing	[d^{-1}]	0.01 - 1
K2	Lower zone outflow coefficient	Groundwater and routing	[d^{-1}]	0.05 - 0.1
UZL	Near surface flow threshold	Groundwater and routing	[mm]	0 - 100
MAXBAS	Flow routing coefficient	Groundwater and routing	[d]	1 - 5

Table A.2 Parameters of the HBV model considered for GSA.
The ranges are taken from Kollat et al. (2012).

Parameter	Definition	Module	Unit	Range	Level
CN2[a,b]	SCS runoff curve number for moisture condition II	Surface Runoff	[-]	-50 - 25	HRU
SLOPE[b]	Average slope steepness	Lateral flow/Surface Runoff	[$m.m^{-1}$]	0 - 1	HRU
SLSUBBSN	Average slope length	Surface Runoff	[m]	10 - 150	HRU
SURLAG	Surface runoff lag coefficient	Surface Runoff	[d]	0.5 - 10	BSN
SOL_K[b]	Soil conductivity	Soil	[$mm.h^{-1}$]	0 - 2000	HRU
SOL_AWC[a]	Available water capacity of the soil layer	Soil	[mm]	-25 - 60	HRU
SMFMX	Maximum melt rate for snow	Snow	[$mm.°C^{-1} d^{-1}$]	0 - 10	BSN
SMFMN	Minimum melt rate for snow	Snow	[$mm.°C^{-1} d^{-1}$]	0 - 10	BSN
SFTMP	Snowfall temperature	Snow	[°C]	-5 - 5	BSN
SMTMP	Snow melt base temperature	Snow	[°C]	-5 - 5	BSN
TIMP	Snow pack temperature lag factor	Snow	[-]	0 - 1	BSN
CANMAX	Maximum canopy storage	Evapotranspiration	[mm]	0 - 10	HRU
ESCO	Soil evaporation compensation factor	Evapotranspiration	[-]	0 - 1	HRU
EPCO	Plant uptake compensation factor	Evapotranspiration	[-]	0.1 - 1	HRU
SOL_ALB	Soil albedo	Evapotranspiration	[-]	0 - 0.25	HRU
ALPHA_BF[b]	Baseflow recession factor	Groundwater	[d^{-1}]	0 - 1	HRU
GW_DELAY	Groundwater delay	Groundwater	[d]	1 - 60	HRU
GW_REVAP	Groundwater 'revap' coefficient	Groundwater	[-]	0.02 - 0.2	HRU
RCHRG_DP[b]	Deep aquifer percolation fraction	Groundwater	[-]	0 - 1	HRU
REVAP_MN	Threshold water depth in shallow aquifer for 'revap'	Groundwater	[mm]	1 - 500	HRU
GWQMN[b]	Threshold water depth in shallow aquifer for return flow	Groundwater	[mm]	10 - 500	HRU
CH_N2	Manning coefficient for channel	Routing	[-]	0 - 0.3	SUBBSN
CH_K2	Hydraulic conductivity in channel	Routing	[$mm.h^{-1}$]	0.5 - 150	SUBBSN
TLAPS	Temperature laps rate	Temperature	[$°C.km^{-1}$]	-10 - 10	SUBBSN
BLAI	Maximum potential leaf area index for crops	Plant growth	[-]	0.5 - 10	BSN
BIOMIX	Biological mixing efficiency	Nutrient cycle	[-]	0 - 1	HRU

Table A.3 Parameters of the SWAT model considered for GSA.

The ranges come from personal communication. HRU Hydrological Response Unit (set of unique land cover, soil characteristics, and management combination), SUBBSN sub-basin, BSN basin.

[a] The value of the parameter is changed by applying a relative change to a given baseline value for the different HRUs (different values of the parameter are applied for the different HRUs).

[b] The value of the parameter is changed separately for the five land use types (Agriculture (A), Pasture (P), Urban (U), Forest (F) and Range Brush (R))

A.2 Investigation of the parameter ranges

We investigated the parameter ranges by identifying the parameterisations that produce poor performance of the model output. We denote the number of time steps as H, the observed value of the stream flow at time t as $y_{o,t}$, the average value of the observations as $\overline{y_o}$, the simulated value of the stream flow at time t as $y_{s,t}$ and the average value of the simulations as $\overline{y_s}$. To assess the model performance, we considered three objective functions:

1. Mean Absolute Error of the simulated stream flow normalised by the standard deviation of the observed stream flow:

$$MAE = \frac{\frac{1}{H}\sum_{t=1}^{H}\left|y_{o,t} - y_{s,t}\right|}{\sqrt{\frac{1}{H}\sum_{t=1}^{H}\left(y_{o,t} - \overline{y_o}\right)^2}}; \tag{A.1}$$

2. Nash Sutcliff Efficiency of the simulated stream flow:

$$\mathrm{NSE} = 1 - \frac{\sum_{t=1}^{H}\left(y_{o,t} - y_{s,t}\right)^2}{\sum_{t=1}^{H}\left(y_{o,t} - \overline{y_o}\right)^2}; \tag{A.2}$$

3. BIAS of the simulated stream flow normalised by the mean of the observed streamflow:

$$BIAS = \frac{\left|\overline{y_o} - \overline{y_s}\right|}{\overline{y_o}}. \tag{A.3}$$

To identify non-behavioural parameterisations, we set threshold values on the three objective functions defined above. We drew a sample of the parameter space within the initial ranges of the parameters, taken from Wagener et al. (2001) for HyMod, from Kollat et al. (2012) for HBV and from personal communication from SWAT. Parallel coordinate plots (Fig. A.1) allowed us to identify possible regions of the parameter ranges that systematically produce non-behavioural values of the model output.

For the HyMod model, we observe non-behavioural regions in the range of parameters SM, BETA and ALPHA. Therefore, we could reduce the ranges for these parameters by taking out these non-behavioural values. We note that we considered as non-behavioural the parameterisations highlighted in red in the plot even if it gives acceptable values of the objective functions. This parameterisation is indeed an outlier since its value of BETA is much higher compared to the other behavioural parameterisations.

Instead, for the HBV and the SWAT model, the behavioural parameterisations are spread across all the parameter space. Therefore, it is not possible to refine the parameter ranges. For the SWAT model, very

high values of parameter CN2 for Agriculture land use produce non-behavioural parameterisations. However, we kept the original range since the possible range reduction was not significant.

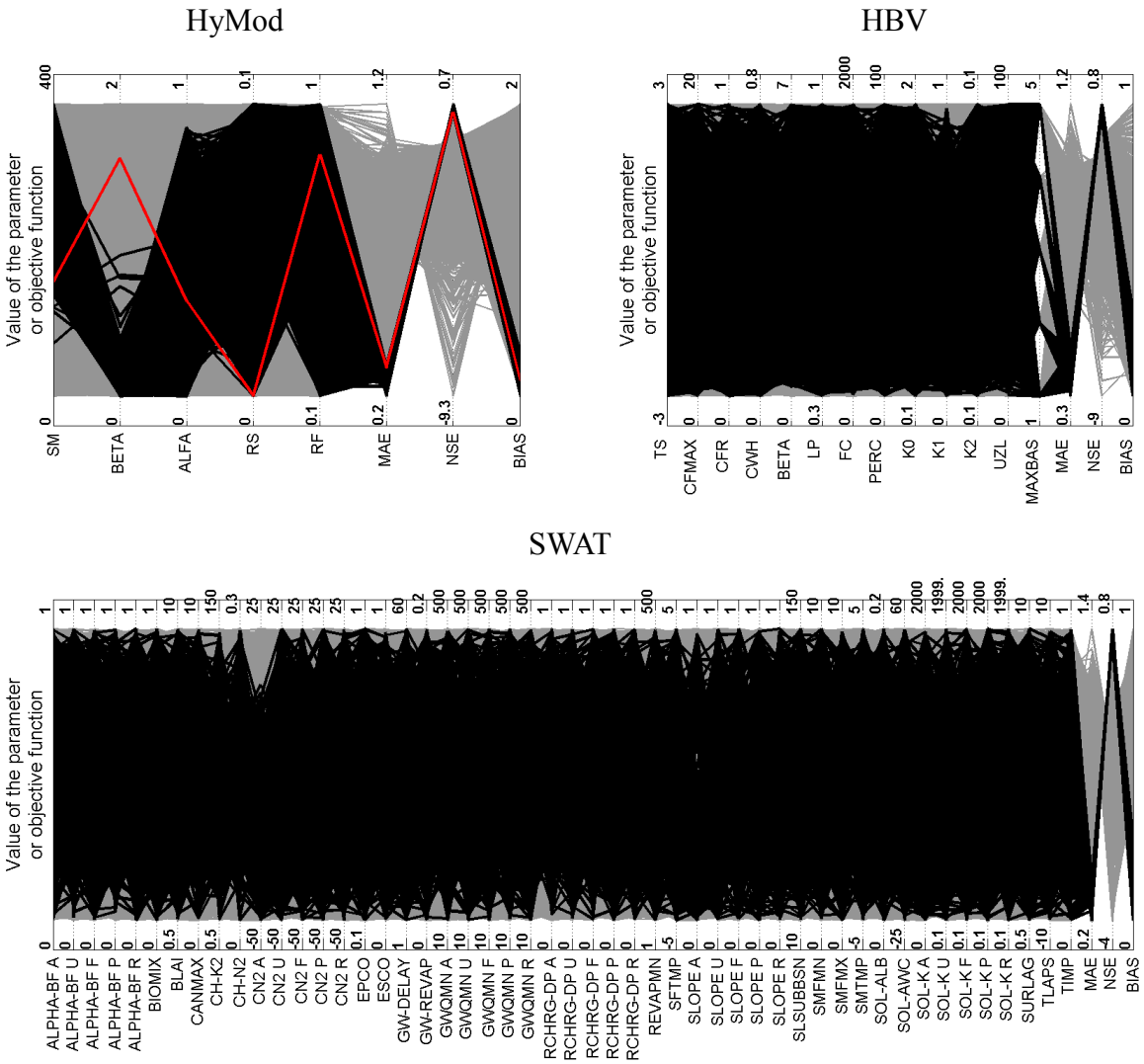


Figure A.1 Parallel coordinate plots for the parameters and objective functions of the HyMod, HBV and SWAT models.

On the x-axis are reported the different parameters and objective functions and on the y-axis are the values of the parameters and the corresponding values of the objective functions. Each line corresponds to a parameterisation (the ranges are normalised so that they can be reported on the same axis). The grey lines represent the non-behavioural parameterisations and the black lines plotted on top represent the behavioural ones. The red line in the plot for the HyMod model identifies the behavioural parameterisation we discarded to refine the parameter ranges, since it appears to be an outlier (high value of BETA).

A.3 Analysis of the convergence statistic for ranking

To select an appropriate statistic for the assessment of the convergence of the ranking, we analysed five potential statistics. We were looking for an indicator that would emphasise the disagreements in the ranking for the most sensitive parameters while neglecting the disagreements for the least sensitive parameters. We compared three categories of indices:

1) The un-weighted Spearman rank correlation coefficient (Spearman, 1904):

$$\rho_1 = 1 - \frac{6\sum_{i=1}^{M}\left(R_i^j - R_i^k\right)^2}{M(M^2-1)}, \tag{A.4}$$

Where R_i^j and R_i^k are the rank of the i-th parameter in the ranking generated using sample k and j respectively, and M is the number of parameters.

2) Three weighted Spearman rank correlation coefficients proposed by Dancelli et al.(2013):

$$\rho_2 = 1 - 2\frac{\sum_{i=1}^{M}\left(R_i^j - R_i^k\right)^2\left(\frac{1}{R_i^j} + \frac{1}{R_i^k}\right)}{\max\left[\sum_{i=1}^{M}\left(R_i^j - R_i^k\right)^2\left(\frac{1}{R_i^j} + \frac{1}{R_i^k}\right)\right]}, \tag{A.5}$$

$$\rho_3 = 1 - 2\frac{\sum_{i=1}^{M}\left(R_i^j - R_i^k\right)^2\left(\frac{1}{R_i^j.R_i^k}\right)}{\max\left[\sum_{i=1}^{M}\left(R_i^j - R_i^k\right)^2\left(\frac{1}{R_i^j.R_i^k}\right)\right]} \tag{A.6}$$

$$\rho_4 = 1 - 2\frac{\sum_{i=1}^{M}\left(R_i^j - R_i^k\right)^2\left(\frac{1}{R_i^j + R_i^k}\right)}{\max\left[\sum_{i=1}^{M}\left(R_i^j - R_i^k\right)^2\left(\frac{1}{R_i^j + R_i^k}\right)\right]} \tag{A.7}$$

3) A correlation coefficient computed on Savage scores (Savage, 1956) proposed by Iman and Conover (1987):

$$\rho_5 = \frac{\sum_{i=1}^{M} SS_i^j SS_i^k - M}{M - \sum_{h=1}^{M}\frac{1}{h}}, \tag{A.8}$$

where SS_i^j and SS_i^k are the value of the Savage score for the i-th parameter in sample j and k respectively. The Savage score for the i-th parameter is defined as follows, assuming that this parameter is ranked in the m-th position:

$$SS_i = \sum_{h=m}^{M} \frac{1}{h}. \tag{A.9}$$

We can demonstrate that this coefficient can be re-written as follows:

$$\rho_5 = 1 - \frac{\sum_{i=1}^{M}(SS_i^j - SS_i^k)^2}{2(M - \sum_{h=1}^{M}\frac{1}{h})}. \tag{A.10}$$

4) An adjusted and weighted rank correlation coefficient introduced in this study:

$$\rho_6 = \sum_{i=1}^{M} \left|R_i^j - R_i^k\right| \frac{\max_{j,k}\left(S_i^j, S_i^k\right)^2}{\sum_{i=1}^{M}\max_{j,k}\left(S_i^j, S_i^k\right)^2}, \tag{A.11}$$

where S_i^j and S_i^k are the value of the sensitivity index of the i-th parameter in sample j and k respectively.

The values of the un-weighted Spearman (Eq. (A.4)), of the weighted Spearman rank correlation coefficients (Eq. (A.5) to (A.7)) and of the coefficient computed on Savage scores (Eq (A.10)) vary between -1 and 1, and the higher the coefficient, the higher the agreement between the two rankings. The value of the adjusted and weighted rank correlation coefficient (Eq. (A.11)) is always positive and the lower the coefficient, the higher the agreement between the two rankings.

We tested these six coefficients with two theoretical examples with 30 model parameters. The values of the sensitivity indices and the ranking for the 30 model parameters in the two samples and for the two examples are reported in Fig. A.2 and A.3. In Example 1, in the two samples, 24 parameters have a low sensitivity (sensitivity index below 0.024) and 6 have a significant value of the sensitivity index (above 0.15). Between sample 1 and sample 2 there is a ranking reversal among the 24 least sensitive parameters while the ranking for the 6 most sensitive parameters is identical in the two samples. In Example 2, as in the Example 1, in the two samples 24 parameters have a low sensitivity (sensitivity index below 0.024) and 6 have a significant value of the sensitivity index (above 0.15). Between sample 1 and sample 2, there is a ranking reversal among the 24 least sensitive parameters and among the 6 most sensitive parameters. In the first example, an appropriate rank coefficient should reveal that the rankings provided

by the two samples are consistent. In the second example, it should indicate that the two rankings are different.

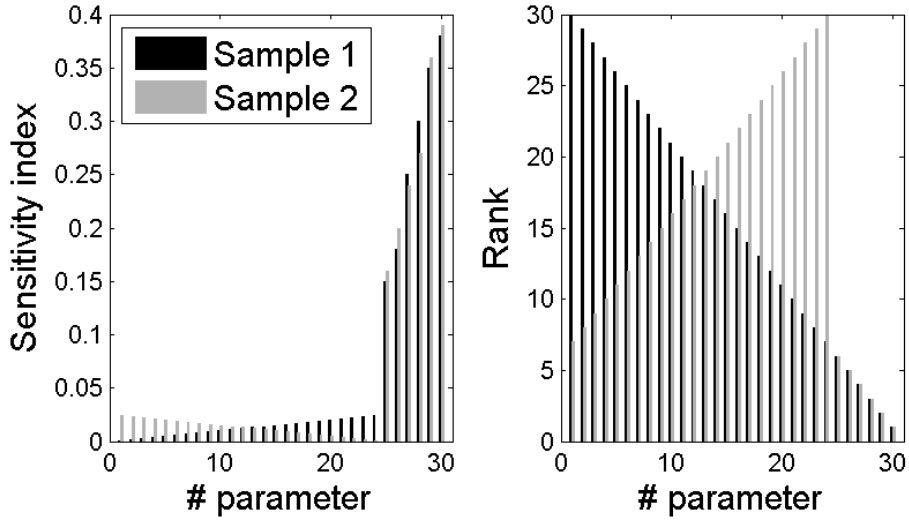


Figure A.2 Value of sensitivity index and rank of the 30 parameters in Example 1.

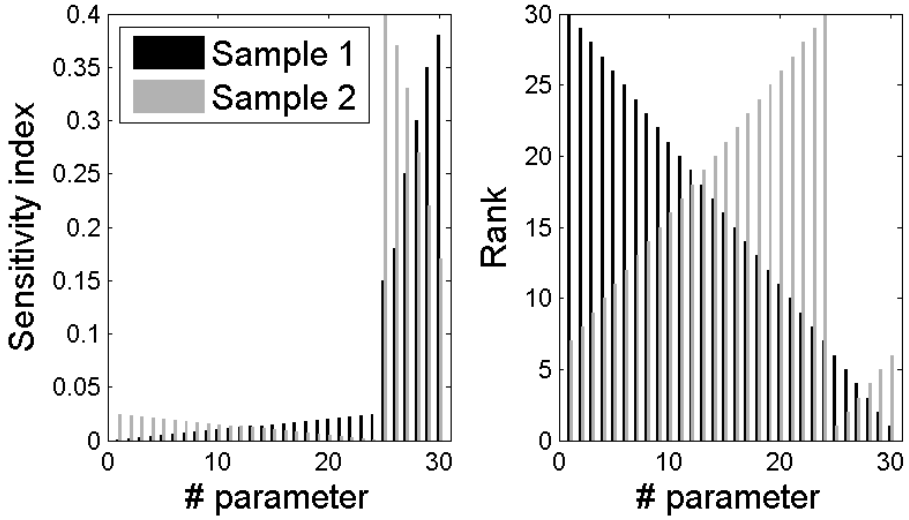


Figure A.3 Value of sensitivity index and rank of the 30 parameters in Example 2.

For a given rank correlation coefficient ρ_m $m = 1 \dots 6$, we note F_i the factor that expresses the contribution of the i-th parameter to the total coefficient:

$$\rho_m = 1 - \sum_{i=1}^{M} F_i \;\; m = 1 \dots 5, \tag{A.12}$$

$$\rho_6 = \sum_{i=1}^{M} F_i. \tag{A.13}$$

Figures A.4 and A.5 represents the value of F_i for the 30 model parameters, for the six rank correlation coefficients in the two examples. In Example 1 (Fig. A.4), we observe that for the six coefficients, the factors F_i for the 24 least sensitive parameters only are non-zero since the ranking does not vary for the other 6 most sensitive parameters. Instead, in Example 2 (Fig. A.5), all the parameters contribute to the rank correlation coefficient. The values of F_i are still the same as in Example 1 for the 24 least sensitive parameters since their sensitivity index does not change. We also note that the relative values of the factors F_i among the parameters vary across the six statistics. For the un-weighted and weighted Spearman coefficient ($\rho_m\ m = 1 \ldots 4$) and for the coefficient based on Savage scores (ρ_5) , the total contribution of parameters 1 to 24 is larger than the total contribution of parameters 25 to 30 while for the modified rank correlation coefficient (ρ_6), the contribution of parameters 1 to 24 is negligible compared to the contribution of parameter 25 to 30. We note however, that the coefficient based on Savage score (ρ_5) seems to perform better compared to the weighted and unweighted Spearman coefficients ($\rho_m\ m = 1 \ldots 4$).

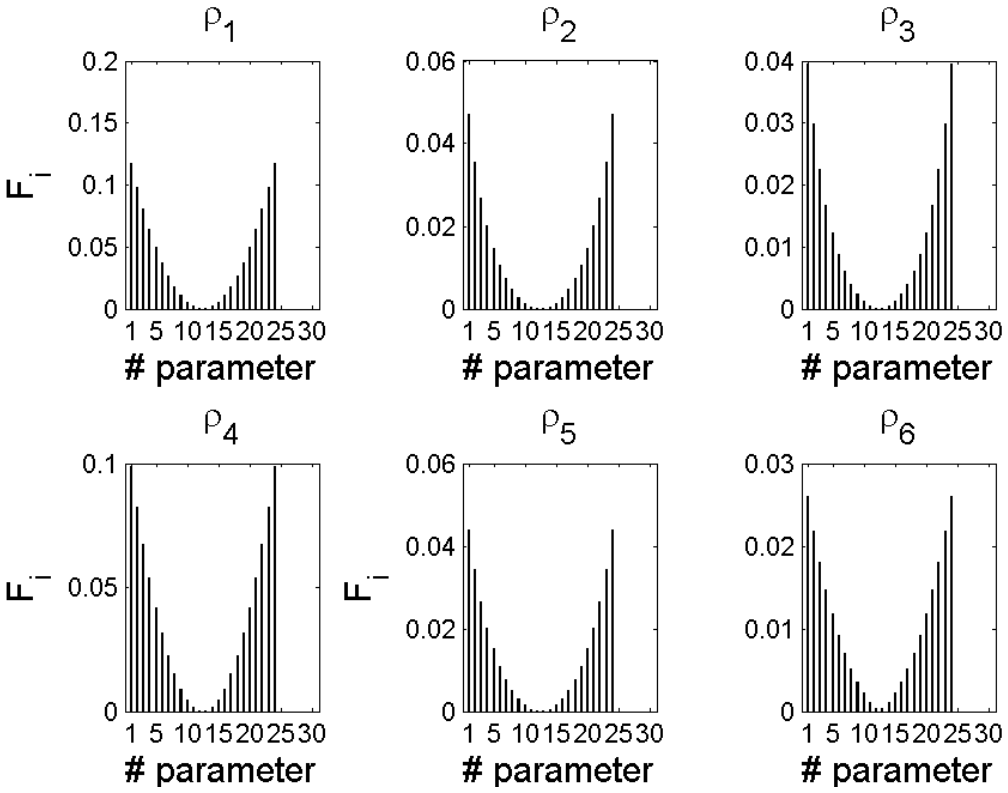


Figure A.4 Parameter contribution to the six rank correlation coefficients for the 30 parameters in Example 1.

F_i Parameter contribution (Eq.(A.12)).

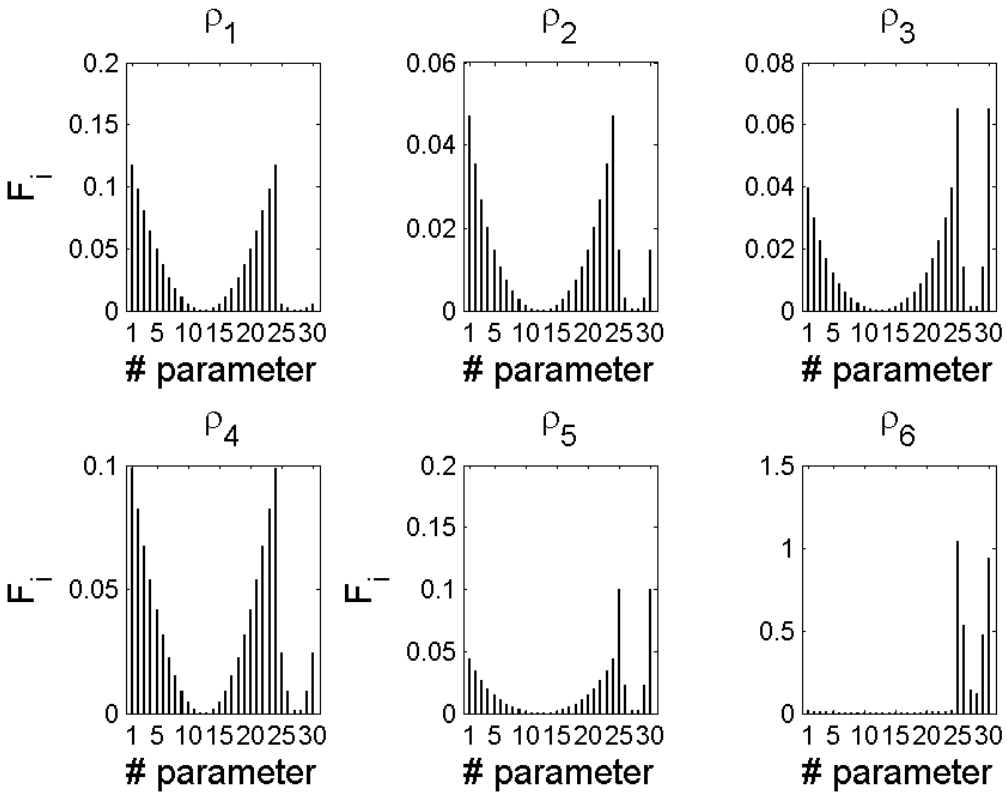


Figure A.5 Parameter contribution to the six rank correlation coefficients for the 30 parameters in Example 2.
F_i Parameter contribution (Eq.(A.12)).

The four Spearman rank correlation coefficients ($\rho_m\ m = 1 \dots 4$) and the coefficient computed on Savage scores (ρ_5) take similar values in the two examples (Table A.4).

The un-weighted Spearman coefficient (ρ_1) gives the same importance to differences in ranking among the highest and the lowest parameters. This can be explained considering that the quantity $R_i^j - R_i^k$ is sometimes higher among the low-sensitivity parameters than among the high-sensitivity parameters and the number of low-sensitivity parameters is higher than the number of high-sensitivity parameters. As a result, the low-sensitivity parameters give the major contribution to the rank correlation coefficient and produce low values of ρ_1.

Likewise, the contribution of the low-sensitivity parameters is still dominant for the weighted Spearman coefficients (ρ_2, ρ_3 and ρ_4) and the coefficient computed on Savage scores (ρ_5) . The weights do not succeed in reducing the contribution of the low-sensitivity parameters compared to the high-sensitivity ones.

The adjusted and weighted rank correlation coefficient (ρ_6), emphasises the differences in the rankings for the high-sensitivity parameters and it is able to discriminate Examples 1 and 2. We see from Table A.4 that in Example 1, ρ_6 takes a very low value of 0.24, which means that the two rankings are consistent. In Example 2, ρ_6 takes a high value of 3.4. Therefore, the adjusted and weighted rank correlation coefficient of Eq. (A.11) was adopted for the convergence study.

	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
Example 1	-0.023	0.66	0.71	0.14	0.66	0.24
Example 2	-0.04	0.62	0.55	0.07	0.41	3.4

Table A.4 Value of the six rank correlation coefficients in the two test examples.

However, we observed that the factors F_i in ρ_6 for low-sensitivity parameters can still take significant values when the value of the sensitivity index for these parameters is not negligible and when their position in the ranking varies significantly between the two samples. Figures A.6 and A.7 report results we obtained when computing the sensitivity indices with Variance-Based method for the SWAT model. We note as w_i the weights of the adjusted and weighted rank correlation coefficient (Fig. A.7):

$$w_i = \frac{\max_{j,k}\left(S_i^j, S_i^k\right)^2}{\sum_{i=1}^{M} \max_{j,k}\left(S_i^j, S_i^k\right)^2}, \tag{A.14}$$

In Fig. A.6 we see that the rank of the first five parameters does not vary between the two samples while some low-sensitivity parameters are changing rank. As a result, we see from Fig A.7 that some low-sensitivity parameters produce a high value of ρ_6. The adjusted rank correlation coefficient in the reported example is high (3.92) while the ranking for the high-sensitivity parameters is similar between the two samples. This shows some limitation of the adjusted and weighted rank correlation coefficient in measuring the agreement between two rankings of parameters.

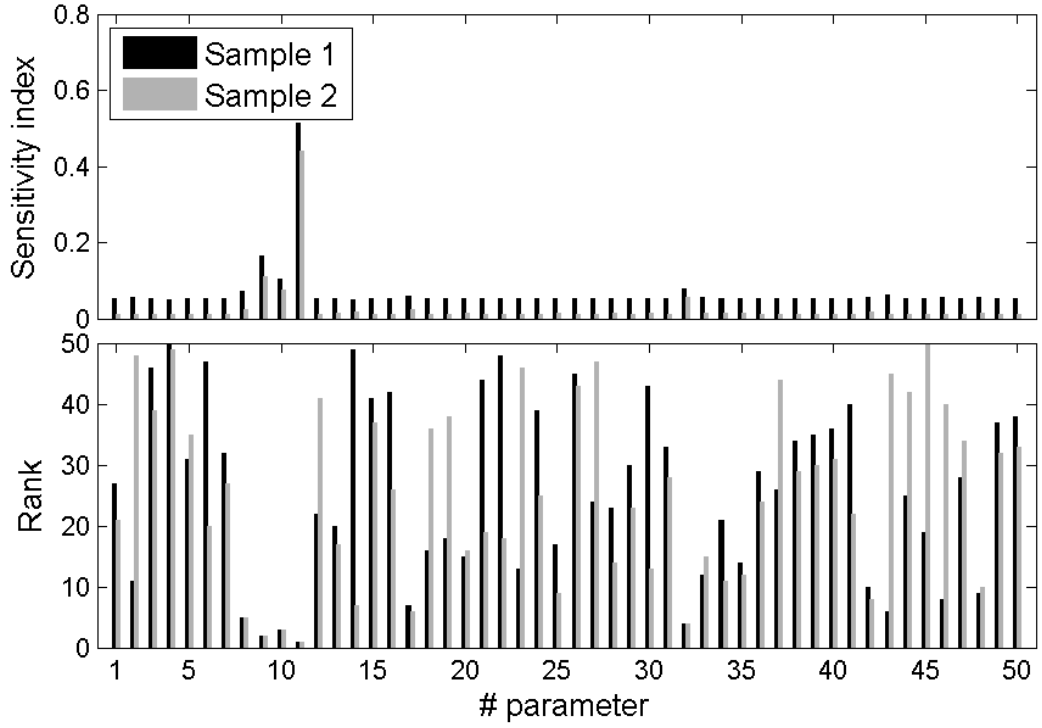


Figure A.6 Value of the VB Main Effect sensitivity index and corresponding rank for two different bootstrap resamples of the 50 parameters of SWAT.

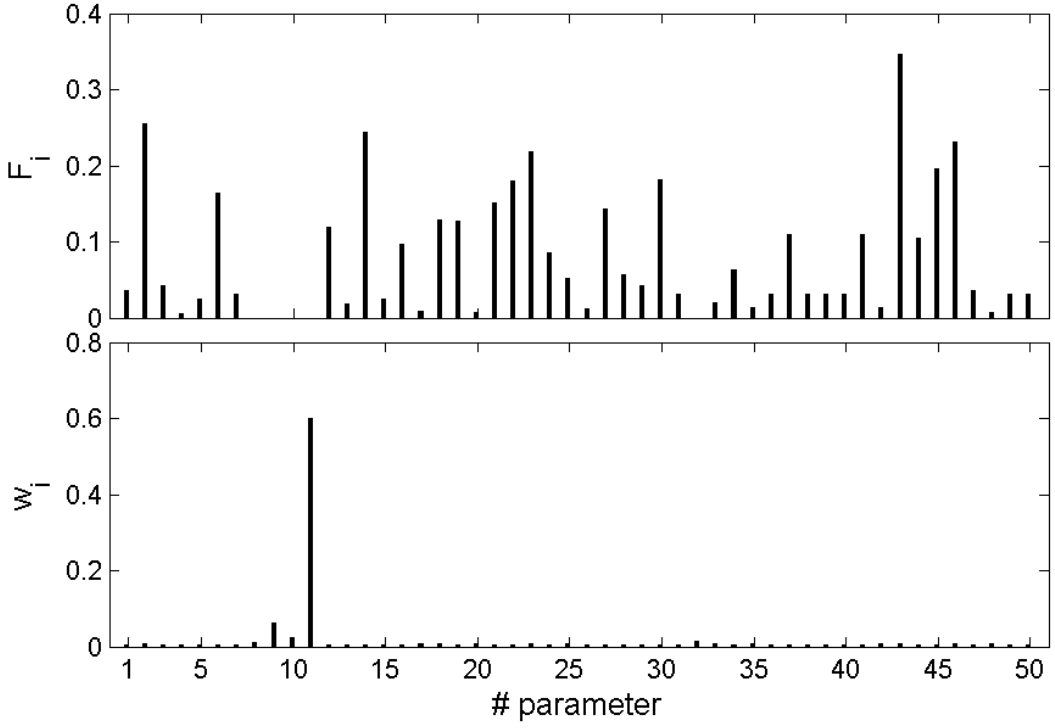


Figure A.7 Parameter contribution to the value of the adjusted rank correlation coefficient (F_6) and corresponding weights of the model parameters w_6 for the two bootstrap resamples of Fig. A.6.

A.4 Choice of the sample sizes for the validation test

To implement the validation test described in Sect. 2.4.2 of Chapter 2 and to set the value of the sample size for conditional (N_c) and unconditional (N_u) outputs, we assessed the convergence of the results of the KS-test. We set a higher value for N_u than for N_c since the input space to be explored is larger for the unconditional output. We computed bootstrap estimates of the KS-statistic using resampling with replacement for increasing sample size. We used 20 conditioning values ($n_c = 20$). Figure A.8 reports the results obtained with bootstrapping (resampling with replacement) and without bootstrapping (sample value) for the HBV model, by fixing different sets of parameters. We observe that the value of the KS-statistic computed without bootstrapping is always below the bootstrap mean in particular when the value of the KS-statistic takes low values. Therefore, the bootstrap estimates tend to be an overestimation of the KS-statistic, especially for low-sensitivity parameters. When using resampling with replacement, the same value of the input/output can appear several times in the bootstrap resamples: shifts are induced in the resamples' CDF and the value of the KS-statistic is increased.

We also derived bootstrap estimates by using resampling without replacement (each bootstrap resample is constructed by taking randomly 80% of the parameter sets in the total sample) which are reported in Fig. A.9. However, we found that the bootstrap estimates are still overestimating the KS-statistic for low-sensitivity parameters. Therefore, we did not use the bootstrap technique to assess the convergence of the KS-test.

From the results obtained without bootstrapping, we conducted a visual analysis to determine suitable values of the sample sizes. We used N_u =2,000 and N_c =1,600 to ensure convergence.

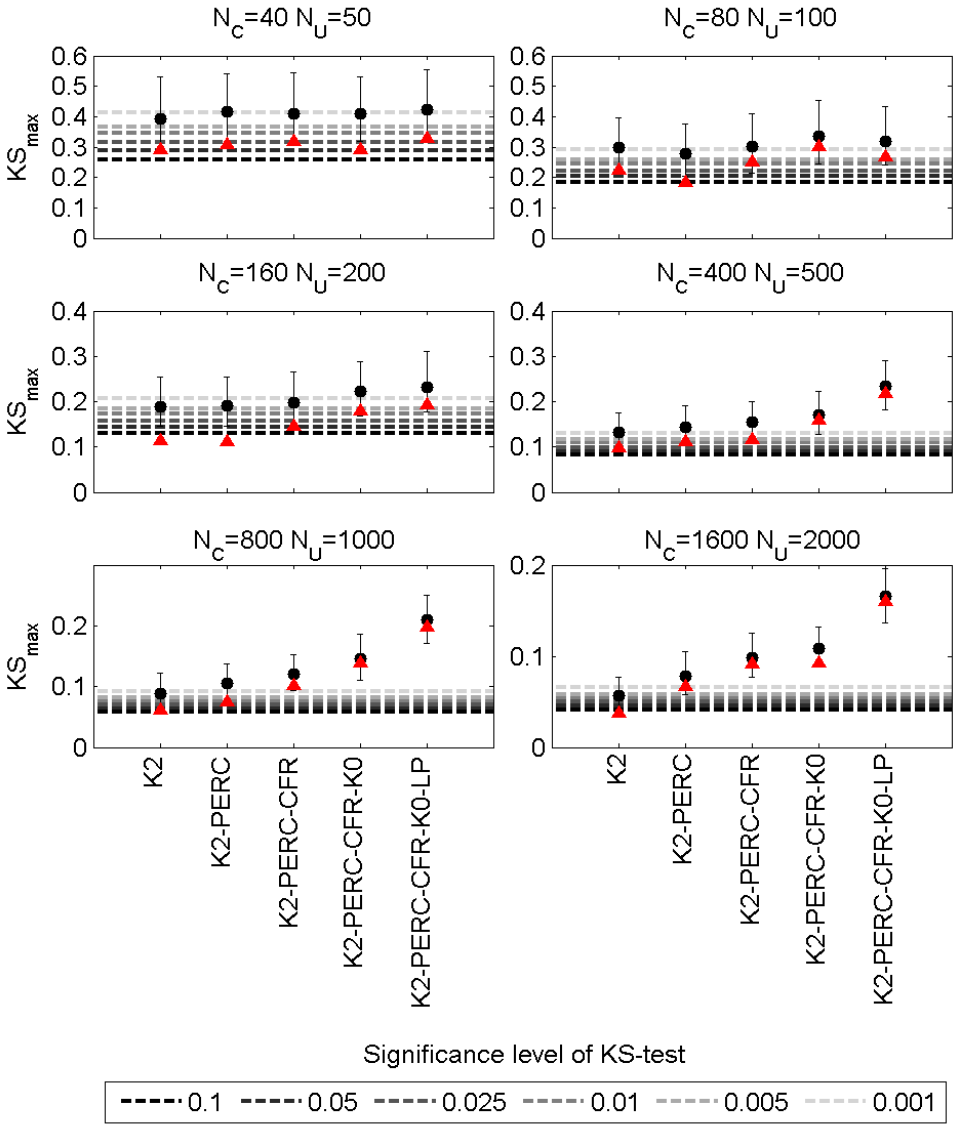


Figure A.8 Value of KS_{max} computed by fixing different sets of parameters of the HBV model for increasing value of the sample sizes for conditional and unconditional outputs.

N_c and N_u sample sizes for conditional and unconditional outputs respectively. The black dots are the bootstrap means and the black error bars are the bootstrap confidence intervals (bootstrap estimates were computed using resampling **with replacement**). The red triangles are the sample values (computed without bootstrapping). On the x-axis are indicated the set of parameters that were fixed to assess the KS statistics.

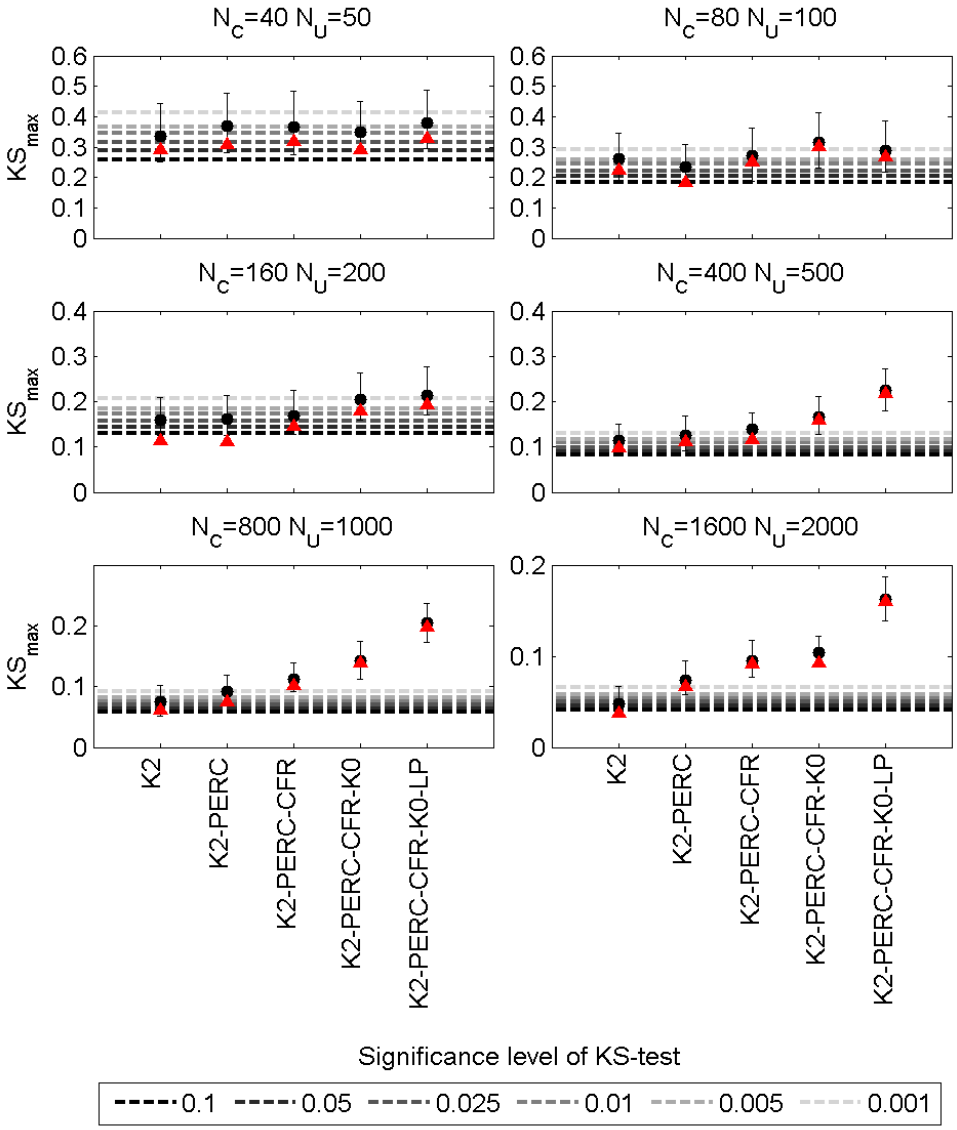


Figure A.9 Value of KS_{max} computed by fixing different sets of parameters of the HBV model for increasing value of the sample sizes for conditional and unconditional outputs.
N_c and N_u sample sizes for conditional and unconditional outputs respectively. The black dots are the bootstrap means and the black error bars are the bootstrap confidence intervals (bootstrap estimates were computed using resampling **without replacement**). The red triangles are the sample values (computed without bootstrapping). On the x-axis are indicated the set of parameters that were fixed to assess the KS statistics.

Appendix B. Supplements to Chapter 3

Appendix B is adapted from the supplemental material of a work under review that has been partially modified to improve the consistency throughout the thesis:

> *Sarrazin, F., Hartmann, A., Pianosi, F. and Wagener, T.: V2Karst v1.0: A parsimonious large-scale integrated vegetation-recharge model to understand the impact of climate and land cover change in karst regions, Geoscientific Model Development Discussion, in review, doi:10.5194/gmd-2017-315, 2018.*

The main article for this work supports Chapter 3.

B.1 Parameters used for ET estimation in large-scale hydrological models

Parameter	Description	Module [a]	Category	Unit	Reference
Z_r	Rooting depth	Stress	Vegetation	[m]	(Vörösmarty et al., 1989)
AWC	Soil available water capacity	Stress	Soil	[$m^3.m^{-3}$]	(Vörösmarty et al., 1989)
α	Empirical coefficient of the drying curve (set to 5)	Stress	Constant	[-]	(Vörösmarty et al., 1998)

Table B.1 Parameters used for ET estimation in the WBM model.

The model includes a minimum of 3 parameters (reported in the table), and additional parameters depending on the PET formulation which is used (namely the Thornthwaite equation (Thornthwaite, 1948) in (Vörösmarty et al., 1996), the Shuttleworth-Wallace (Shuttleworth and Wallace, 1985) equation in (Federer et al., 2003), and a range of different PET equations in (Vörösmarty et al., 1998)).

[a] Stress: Stress model for actual ET calculation

Parameter	Description	Module [a]	Category	Unit	Reference
β_{28}	Aspect correction factor of PET	PET	Terrain	[-]	(Kumar et al., 2013; Samaniego et al., 2010)
β_1	Effective maximum capacity storage	Interception/ Seasonality	Vegetation	[mm]	(Kumar et al., 2013; Samaniego et al., 2010)
Exp_{can}	Exponent to assess the wet canopy fraction (set to 2/3)	Interception	Constant	[-]	(Samaniego et al., 2010)
β_{15}	Permanent wilting point	Stress	Vegetation and soil	[-]	(Samaniego et al., 2010)
β_{16}	Soil moisture limit above which the actual transpiration is equal to PET	Stress	Vegetation and soil	[-]	(Samaniego et al., 2010)
β_{17}^1	Fraction of roots in soil layer 1	Stress	Vegetation		(Rakovec et al., 2016; Samaniego et al., 2010)
β_{17}^2	Fraction of roots in soil layer 2	Stress	Vegetation		(Rakovec et al., 2016; Samaniego et al., 2010)
d_1	depth soil layer 1 (set to 0.05 m)	Soil layers	Constant	[m]	(Rakovec et al., 2016)
d_2	depth soil layer 2 (set to 0.25 m)	Soil layers	Constant	[m]	(Rakovec et al., 2016)
d_3	depth soil layer 3 (set to 1 m)	Soil layers	Constant	[m]	(Rakovec et al., 2016)

Table B.2 Parameters used for ET estimation in the mHM model.

[a] PET: potential evapotranspiration equation; Stress: Stress model for actual ET calculation from PET.

Parameter	Description	Module [a]	Category	Unit	Reference
α_{PT}	Priestley-Taylor empirical coefficient (1.26 in semiarid and arid areas and 1.74 in humid areas)	PET	Climate	[-]	(Döll et al., 2003)
$E_{pot,max}$	Maximum potential evapotranspiration (20 mmd^{-1} in semiarid and arid areas and 10 mmd^{-1} in humid areas)	Stress	Climate	[$mm.d^{-1}$]	(Müller Schmied et al., 2014)
Z_r	Rooting depth	Stress	Vegetation	[m]	(Müller Schmied et al., 2014)
AWC	Soil available water capacity	Stress	Soil	[$m^3.m^{-3}$]	(Döll et al., 2003)
V_{can}	Interception storage capacity per unit of LAI (set to 0.3 mm LAI)	Interception	Constant	[mm LAI]	(Döll et al., 2003)
Exp_{can}	Exponent to assess the wet canopy fraction (set to 2/3)	Interception	Constant	[-]	(Deardorff, 1978; Döll et al., 2003)
LAI_{max}	Maximum leaf area index	Interception	Vegetation	[$m^2.m^{-2}$]	(Müller Schmied et al., 2014)
$f_{d,lc}$	Fraction of deciduous plants in LAI growth model	Seasonality	Vegetation	[-]	(Müller Schmied et al., 2014)
$c_{e,lc}$	Reduction factor for evergreen plants in LAI growth model	Seasonality	Vegetation	[-]	(Müller Schmied et al., 2014)
t_{min}	Initial days to start/end with growing season in LAI growth model	Seasonality	Vegetation	[d]	(Müller Schmied et al., 2014)
LAI_{min}	Minimum leaf area index for deciduous plants in LAI growth model (set to 0.1 $m^2.m^{-2}$)	Seasonality	Constant	[$m^2.m^{-2}$]	(Müller Schmied et al., 2014)
T_{min}	Daily temperature threshold to initiate the growing season in LAI growth model (set to 8°C)	Seasonality	Constant	[°C]	(Müller Schmied et al., 2014)
$P_{min,cum}$	Cumulative precipitation threshold to initiate the growing season in LAI growth model (set to 40mm)	Seasonality	Constant	[mm]	(Müller Schmied et al., 2014)
$P_{min,daily}$	Minimum daily precipitation to keep growing season growing in semi-arid and arid regions in LAI growth model (set to 0.5mm)	Seasonality	Constant	[$mm.d^{-1}$]	(Müller Schmied et al., 2014)
t_{growth}	Number of days for LAI to increase from its minimum to its maximum value or to decrease from its maximum to its minimum value in LAI growth model (set to 30 d)	Seasonality	Constant	[d]	(Müller Schmied et al., 2014)

Table B.3 Parameters used for ET estimation in the WaterGap V2.2 model.

[a] PET: potential evapotranspiration equation; Stress: Stress model for actual ET calculation from PET.

Parameter	Description	Module [a]	Category	Unit	Reference
g_{min}	Minimum canopy conductance	PET	Vegetation	[mm.s^{-1}]	(Gerten et al., 2004; Sitch et al., 2003)
g_m	Scaling conductance in the evaporative demand function (set to 3.26 mm.s^{-1})	PET	Constant	[mm.s^{-1}]	(Gerten et al., 2004)
α_m	Priestley-Taylor empirical coefficient (set to 1.391)	PET	Constant	[-]	(Gerten et al., 2004)
α_{PT}	Priestley-Taylor empirical coefficient (set to 1.32)	PET	Constant	[-]	(Gerten et al., 2004)
i	Empirical coefficient for calculation of interception (same formulation as (Kergoat, 1998))	Interception	Vegetation	[-]	(Gerten et al., 2004)
LAI	Leaf area index (determined as a function of daily phenomenology)	Interception	Vegetation	[m^2.m^{-2}]	(Gerten et al., 2004)
$E_{pot,max}$	Maximum potential evapotranspiration (5-7 mm.d^{-1})	Stress	Vegetation	[mm.d^{-1}]	(Gerten et al., 2004)
AWC	Soil available water capacity	Stress	Soil	[m^3.m^{-3}]	(Gerten et al., 2004)
$f_{root,0}$	Weighting constant to determine fraction of roots in evaporation layer (set to 1.3)	Stress	Constant	[-]	(Gerten et al., 2004)
$f_{root,1}$	fraction of roots in soil layer 1	Stress	Vegetation	[-]	(Gerten et al., 2004; Sitch et al., 2003)
d_1	depth soil layer 1 (set to 0.5 m)	Soil layers	Constant	[m]	(Gerten et al., 2004)
d_2	depth soil layer 2 (set to 1 m)	Soil layers	Constant	[m]	(Gerten et al., 2004)
d_0	depth evaporation layer (set to 0.2 m)	Soil layers	Constant	[m]	(Gerten et al., 2004)
f_c	Vegetation cover fraction (determined as a function of daily phenomenology)	Sparse vegetation	Vegetation	[-]	(Gerten et al., 2004)

Table B.4 Parameters used for ET estimation in the LPJ model.

[a] PET: potential evapotranspiration equation; Stress: Stress model for actual ET calculation from PET.

Parameter	Description	Module [a]	Category	Unit	Reference
$r_{a,veg}$	Vegetation aerodynamic resistance	PET	Vegetation	[s.m^{-1}]	(Kergoat, 1998)
r_{st}	Minimum stomatal resistance	PET	Vegetation	[s.m^{-1}]	(Kergoat, 1998)
$r_{a,soi}$	Soil aerodynamic resistance (set to 100 s.m^{-1})	PET	Constant	[s.m^{-1}]	(Kergoat, 1998)
$r_{s,soi}$	Soil surface resistance (set to 50 s.m^{-1})	PET	Constant	[s.m^{-1}]	(Kergoat, 1998)
LAI	Leaf area index	PET and interception	Vegetation	[m^2.m^{-2}]	(Kergoat, 1998)
β	Empirical coefficient for calculation of interception	Interception	Vegetation	[-]	(Kergoat, 1998)
S_1	Constant in radiation term in stomatal resistance parameterization (set to 10 W PAR.m$^{-2)}$	PET (surface resistance)	Constant	[W PAR .m^{-2}]	(Kergoat, 1998)
f_s	Fraction of photosynthetically active solar radiation (set to 0.48)	PET (surface resistance)	Constant	[-]	(Kergoat, 1998)
D_1	First coefficient of the vapour pressure deficit term in stomatal resistance parameterization (set to 3000 Pa)	PET (surface resistance)	Constant	[Pa]	(Kergoat, 1998)
D_2	Second coefficient of the vapour pressure deficit term in stomatal resistance parameterization (set to 3500 Pa)	PET (surface resistance)	Constant	[Pa]	(Kergoat, 1998)
k	Beer- Lambert extinction coefficient (set to 0.5)	PET (surface resistance) and Sparse vegetation	Constant	[-]	(Kergoat, 1998)
Z_r	Rooting depth	Stress	Vegetation	[m]	(Kergoat, 1998)
AWC	Soil available water capacity	Stress	Soil	[m^3.m^{-3}]	(Kergoat, 1998)
W_1	Soil water constant for stomatal closure as a fraction of soil water storage (set to 0.4)	Stress	Constant	[-]	(Kergoat, 1998)
W_2	Soil water constant for soil evaporation reduction (set to 0.6)	Stress	Constant	[-]	(Kergoat, 1998)

Table B.5 Parameters used for ET estimation in the model proposed by (Kergoat, 1998)**.**
We did not review the light limitation sub-model of the model, which is used to calculate an equilibrium value of LAI.

[a] PET: potential evapotranspiration equation; Stress: Stress model for actual ET calculation from PET.

Parameter	Description	Module [a]	Category	Unit	Reference
K_c	Crop factor (monthly values estimated as a function of land cover and climatology)	PET (and seasonality)	Vegetation	[-]	(Van Beek, 2008)
$K_{c,min}$	Minimum crop factor for bare soil (set to 0.2)	PET	Constant	[-]	(Van Beek, 2008; Sperna Weiland et al., 2015)
LAI	Leaf area index (monthly values estimated as a function of land cover and climatology)	Interception (and seasonality)	Vegetation	[$m^2.m^{-2}$]	(Van Beek, 2008; Sutanudjaja et al., 2011)
V_{can}	Interception storage capacity (set to 0.3 mm LAI)	Interception	Constant	[mm LAI]	(Sutanudjaja et al., 2011)
$f_{root,1}$	Root fraction in soil layer 1	Stress	Vegetation	[-]	(Van Beek, 2008; Sperna Weiland et al., 2015; Sutanudjaja et al., 2011)
β_1	Coefficient of the soil water retention curve in soil layer 1	Stress	Soil	[-]	(Van Beek, 2008; Sutanudjaja et al., 2011)
β_2	Coefficient of the soil water retention curve in soil layer 2	Stress	Soil	[-]	(Van Beek, 2008; Sutanudjaja et al., 2011)
$W_{sat,1}$	Saturated volumetric moisture content in soil layer 1	Stress	Soil	[$m^3.m^{-3}$]	(Van Beek and Bierkens, 2008; Sperna Weiland et al., 2015)
$W_{sat,2}$	Saturated volumetric moisture content in soil layer 2	Stress	Soil	[$m^3.m^{-3}$]	(Van Beek and Bierkens, 2008; Sperna Weiland et al., 2015)
$k_{sat,1}$	Saturated hydraulic conductivity in soil layer 1	Stress (soil evaporation)	Soil	[$m.d^{-1}$]	(Van Beek, 2008; Sutanudjaja et al., 2011)
$\Psi_{sat,1}$	Matric soil suction at saturation in soil layer 1	Stress (transpiration)	Soil	[m]	(Sutanudjaja et al., 2011)
$\Psi_{sat,2}$	Matric soil suction at saturation in soil layer 2	Stress (transpiration)	Soil	[m]	(Sutanudjaja et al., 2011)
$\Psi_{50\%}$	Matric soil suction at which transpiration is halved (set for instance equal to 3.33m)	Stress (transpiration)	Constant	[m]	(Sutanudjaja et al., 2011)
d_1	Depth of soil layer 1 (set to 0.3 m)	Stress	Constant	[m]	(Van Beek and Bierkens, 2008)
d_2	Depth of soil layer 2 (set to 1.2 m)	Stress	Constant	[m]	(Van Beek and Bierkens, 2008)

Table B.6 Parameters used for ET estimation in the PCR-GLOBWB model.

[a] PET: potential evapotranspiration equation; Stress: Stress model for actual ET calculation from PET.

Parameter	Description	Module [a]	Category	Unit	Reference
$h_{veg,over}$	Overstory vegetation height	PET	Overstory vegetation	[m]	(Gosling and Arnell, 2011; Smith, 2016)
$r_{st,over}$	Overstory vegetation stomatal resistance	PET	Overstory vegetation	[s.m^{-1}]	(Gosling and Arnell, 2011; Smith, 2016)
LAI_{over}	Overstory leaf area index	PET	Overstory vegetation	[m^2.m^{-2}]	(Gosling and Arnell, 2011; Smith, 2016)
$h_{veg,over}$	Understory vegetation height (set to value for grass)	PET	Understory vegetation	[m]	(Gosling and Arnell, 2011; Smith, 2016)
$r_{st,under}$	Understory vegetation stomatal resistance (set to value for grass)	PET	Understory vegetation	[s.m^{-1}]	(Gosling and Arnell, 2011; Smith, 2016)
LAI_{under}	Understory leaf area index (set to value for grass)	PET	Understory vegetation	[m^2.m^{-2}]	(Gosling and Arnell, 2011; Smith, 2016)
K	Radiation coefficient to calculate canopy surface resistance (set to 0.7)	PET	Constant	[-]	(Smith, 2016)
$r_{s,soi}$	(Soil) resistance to calculate canopy surface resistance (set to 100 s.m^{-1})	PET	Constant	[s.m^{-1}]	(Smith, 2016)
$Z_{r,over}$	Overstory rooting depth	Stress	Overstory vegetation	[m]	(Gosling and Arnell, 2011; Smith, 2016)
$Z_{r,under}$	Understory rooting depth (set to value for grass)	Stress	Understory vegetation	[m]	(Gosling and Arnell, 2011; Smith, 2016)
FC	Soil field capacity	Stress	Soil	[m^3.m^{-3}]	(Gosling and Arnell, 2011; Smith, 2016)
S_{max}	Soil saturation capacity	Stress	Soil	[m^3.m^{-3}]	(Gosling and Arnell, 2011; Smith, 2016)
γ_{over}	Overstory interception capacity	Interception	Overstory vegetation	[mm]	(Gosling and Arnell, 2011; Smith, 2016)
γ_{under}	Understory interception capacity (set to value for grass)	Interception	Understory vegetation	[mm]	(Gosling and Arnell, 2011; Smith, 2016)
δ	Empirical parameter of interception model (set to 0.75)	Interception	Constant	[-]	(Arnell, 1999; Smith, 2016)
$Percov$	Percent overstory cover	Sparse vegetation	Overstory vegetation	[%]	(Gosling and Arnell, 2011; Smith, 2016)

Table B.7 Parameters used for ET estimation in the Mac-PDM model.

[a] PET: potential evapotranspiration equation; Stress: Stress model for actual ET calculation from PET.

Parameter	Description	Module [a]	Category	Unit	Reference
α_{PT}	Priestley-Taylor empirical coefficient	PET	Vegetation	[-]	(Miralles et al., 2011)
f_G	Ground heat as a fraction of net radiation	PET	Vegetation	[-]	(Miralles et al., 2011)
β	Correction factor for transpiration to account for hours with wet canopy (set to 0.07)	PET (tall vegetation)	Constant	[-]	(Miralles et al., 2011)
VOD	Vegetation optical depth (remotely sensed)	Stress and seasonality	Vegetation	[-]	(Martens et al., 2017; Miralles et al., 2011)
VOD_{max}	Maximum vegetation optical depth	Stress	Vegetation	[-]	(Martens et al., 2017)
Z_r	Rooting depth	Stress	Vegetation	[m]	(Miralles et al., 2011)
WP	Wilting point	Stress	Soil	[$m^3.m^{-3}$]	(Martens et al., 2017)
FC	Field capacity	Stress	Soil	[$m^3.m^{-3}$]	(Martens et al., 2017)
S_c	Canopy storage for tall vegetation (set to 1.2 mm)	Interception (tall vegetation)	Constant	[mm]	(Miralles et al., 2010)
$\overline{E_c}$	Mean evaporation rate for interception for tall vegetation (set to 0.3 $mm.h^{-1}$)	Interception (tall vegetation)	Constant	[$mm.h^{-1}$]	(Miralles et al., 2010)
$\overline{R_s}$	Mean (synoptic) rainfall rate for tall vegetation (set to 1.5 $mm.h^{-1}$)	Interception (tall vegetation)	Constant	[$mm.h^{-1}$]	(Miralles et al., 2010)
$\overline{R_c}$	Mean (convective) rainfall rate for tall vegetation (set to 5.6 $mm.h^{-1}$)	Interception (tall vegetation)	Constant**)**	[$mm.h^{-1}$]	(Miralles et al., 2010)
p_d	Fraction of rain to trunks for tall vegetation (set to 0.02)	Interception (tall vegetation)	Constant	[-]	(Miralles et al., 2010)
e	Fraction of trunk evaporation for tall vegetation (set to 0.02)	Interception (tall vegetation)	Constant	[-]	(Miralles et al., 2010)
S_t	Trunk capacity for tall vegetation (set to 0.02 mm)	Interception (tall vegetation)	Constant	[mm]	(Miralles et al., 2010)
d_1	Depth at the bottom of the first soil layer (set to 0.05m)	Soil layers	Constant	[m]	(Miralles et al., 2011)
d_2	Depth at the bottom of the second soil layer (set to 1 m)	Soil layers	Constant	[m]	(Miralles et al., 2011)
d_3	Depth at the bottom of the third soil layer (set to 2.5 m)	Soil layers	Constant	[m]	(Miralles et al., 2011)

Table B.8 Parameters used for ET estimation in the GLEAM V3 model.

[a] PET: potential evapotranspiration equation; Stress model for actual ET calculation from PET.

Parameter	Description	Module [a]	Category	Unit	Reference
z_0	Surface roughness length	PET	Vegetation	[m]	(Liang et al., 1994)
r_{st}	Minimum stomatal resistance	PET	Vegetation	[s.m^{-1}]	(Bohn and Vivoni, 2016; Liang et al., 1994)
r_{arc}	Vegetation architectural resistance (boundary layer resistance)	PET	Vegetation	[s.m^{-1}]	(Bohn and Vivoni, 2016; Liang et al., 1994)
d_0	Vegetation zero plane displacement height	PET	Vegetation	[m]	(Liang et al., 1994)
$r_{s,soi}$	Soil surface resistance (set to 0 s.m^{-1})	PET	Constant	[s.m^{-1}]	(Bohn and Vivoni, 2016)
$r_{arc,soi}$	Soil architectural resistance (set to 0 s.m^{-1})	PET	Constant	[s.m^{-1}]	(Bohn and Vivoni, 2016)
LAI	Leaf area index (average monthly values)	PET and interception	Vegetation	[m^2.m^{-2}]	(Bohn and Vivoni, 2016; Liang et al., 1994)
V_{can}	Interception storage capacity per unit of LAI (set to 0.2 mm LAI)	Interception	Constant	[mm LAI]	(Liang et al., 1994)
Exp_{can}	Exponent to assess the wet canopy fraction (set to 2/3)	Interception	Constant	[-]	(Deardorff, 1978; Liang et al., 1994)
R_{GL}	Limit value of incoming solar radiation	PET (surface resistance)	Vegetation	[W m^{-2}]	(Bohn and Vivoni, 2016)
$r_{st,max}$	Maximum surface resistance	PET (surface resistance)	Constant	[s.m^{-1}]	(Bohn and Vivoni, 2016)
f_s	Fraction of photosynthetically active solar radiation	PET (surface resistance)	Constant	[-]	(Bohn and Vivoni, 2016)
g	Coefficient of the vapour pressure deficit term	PET (surface resistance)	Constant	[hPa^{-1}]	(Bohn and Vivoni, 2016)
k_T	Coefficient of the temperature term	PET (surface resistance)	Constant	[K^{-2}]	(Bohn and Vivoni, 2016)
$f_{root,1}$	Root fraction in first soil layer	Stress	Vegetation	[-]	(Liang et al., 1994)
W_{crit}	Critical soil moisture in stomatal resistance parameterization as a fraction of soil saturation	Stress	Soil	[m^3.m^{-3}]	(Bohn and Vivoni, 2016; Liang et al., 1994)
WP	Wilting point	Stress	Soil	[m^3.m^{-3}]	(Bohn and Vivoni, 2016; Liang et al., 1994)
d_1	Depth of soil layer 1 (e.g. set to 0.3 m)	Stress	Constant	[m]	(Liang et al., 1994)
d_2	Depth of soil layer 2 (e.g. set to 0.7 m)	Stress	Constant	[m]	(Liang et al., 1994)
$NDVI$	Normalized Difference Vegetation Index (remotely sensed daily values)	Sparse vegetation and seasonality	Vegetation	[-]	(Bohn and Vivoni, 2016)
$NDVI_{min}$	Minimum Normalized Difference Vegetation Index (set to 0.1)	Sparse vegetation	Constant	[-]	(Bohn and Vivoni, 2016)
$NDVI_{max}$	Maximum Normalized Difference Vegetation Index (set to 0.8)	Sparse vegetation	Constant	[-]	(Bohn and Vivoni, 2016)

Table B.9 Parameters used for ET estimation in the VIC V4.2 model.
Additional information on model parameters was found in the GLDAS project (https://ldas.gsfc.nasa.gov/gldas/GLDASmapveg.php).

[a] PET: potential evapotranspiration equation; Stress model for actual ET calculation from PET

B.2 References for V2Karst parameter ranges

Parameter	unit	Lower limit	Upper limit	Note and references for parameter range
h_{veg}	[m]	0.2	Site specific	The upper bound is set for each site specifically so that it is lower than the measurement heights reported in Table B1.
r_{st}	[s.m^{-1}]	20	600	The range includes the 70th percentiles of the values for the different vegetation types in temperate climate (Breuer et al., 2003).
LAI_{min}	[%]	5	100	Best guess estimate.
LAI_{max}	[m^2.m^{-2}]	0.5	8	The range includes the 70th percentiles calculated for the different vegetation types in temperate climate (Breuer et al., 2003).
V_r	[mm]	20	500	The range includes the 70th percentiles of the values of rooting depth (provided in [m]) for the different vegetation types in temperate climate (Breuer et al., 2003) multiplied by an average value of soil available water capacity of 0.2 m^{3}m^{-3} (Bonan, 2015; Miralles et al., 2011; Salter and Williams, 1965).
V_{can}	[mm LAI]	0.1	0.5	The range includes the value used in WaterGap (Döll et al., 2003) for daily application (0.3 mm LAI); in VIC (Liang et al., 1994) and ISBA (Noilhan and Planton, 1989) for subdaily applications as proposed in (Dickinson, 1984) (0.2 mm LAI); in the Distributed Hydrology-Soil-Vegetation model (Wigmosta et al., 1994) for subdaily applications (0.1 mm LAI); the maximum value used in Mac-PDM [Gosling and Arnell, 2011] (0.5 mm LAI for open shrublands).
k	[-]	0.4	0.7	The range includes the value reported in (Van Dijk and Bruijnzeel, 2001; Granier et al., 1999; Kergoat, 1998; Ruiz et al., 2010) (0.5); in (Shuttleworth and Wallace, 1985) (0.7).
f_{red}	[-]	0	0.15	The range includes the value reported in (Penman, 1950; Wagener et al., 2003) (1/12).
z_0	[m]	0.0003	0.013	The range includes the value used in MOSES (Essery et al., 2001) (0.0003m); in Hydrus (Šimůnek et al., 2009) (0.001 m); in NOAH (Yang et al., 2011) and the Community Land model (Oleson et al., 2010) (0.01 m); in (Masson et al., 2003) (0.013 m).
$r_{s,soi}$	[s.m^{-1}]	0	100	The range includes value used in VIC (Bohn and Vivoni, 2016) and SWAP (Kroes et al., 2008) (0 m.s^{-1}); in (Kergoat, 1998) (50 m.s^{-1}); in MacPDM (Smith, 2016) (100 m.s^{-1}); in (Van de Griend and Owe, 1994) (10 m.s^{-1}).
V_e	[mm]	5	45	Range includes the average depth of 0.1-0.15 m recommended in (Allen et al., 1998) multiplied by a large value of the soil water capacity of 0.3 m^{3}m^{-3} ((Bonan, 2015; Salter and Williams, 1965)).
a	[-]	0	6	(Hartmann et al., 2015)
V_{soil}	[mm]	20	800	Best guess estimate.
V_{epi}	[mm]	200	700	(Hartmann et al., 2015)
K_{epi}	[d]	0	50	(Hartmann et al., 2015)

Table B.10 References for the determination of the unconstrained parameter ranges of V2Karst for the application at the four FLUXNET sites.

The ranges were selected to capture the variability across soil, epikarst and vegetation types. Parameters a, V_{soil}, V_{epi} and K_{epi} were already present in the previous version of the model (VarKarst).

Parameter	Unit	German site (deciduous forest)		Spanish site (shrubland)		French 1 site (evergreen forest)		French 2 site (evergreen forest)		Note and reference for parameter ranges
		Lower limit	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit	
h_{veg}	[m]	23.1	42.9	0.35	0.85	7.1	13.3	3.9	7.2	The range corresponds to the average value reported in Table B1 for the site ±30%. At the Spanish site, the upper bound is set higher due to the presence of a few plants taller than average.
r_{st}	[s.m^{-1}]	275	400	195	350	320	455	320	455	40th and 60th percentile values reported in (Breuer et al., 2003) for the specific land cover at the site.
LAI_{min}	[%]	5	20	34	63	80	100	80	100	At the Spanish site, the range corresponds to the value reported in Table B1 for the site ±30%, and it is a best guess estimates for the other sites.
LAI_{max}	[m^2.m^{-2}]	3.5	6.5	1.9	3.5	1.5	2.9	2.0	3.8	The range corresponds to the value reported in Table B1 for the site ±30%.
V_r	[mm]	60	300	30	200	30	200	30	200	The range includes the average value of the soil available water capacity for the German, Spanish and French 2 sites, and the value of the available water capacity of the root zone for the French 2 site. The upper bound is set to a high value to include uncertainty and to account for the fact that at the German, Spanish and French 1 sites, roots could extend below the soil because the soil is quite shallow.
V_{soi}	[mm]	60	400	30	300	30	300	30	300	Best guess estimates.

Table B.11 References for the determination of the constrained (site-specific) parameter ranges of V2Karst for the application at the four FLUXNET sites.

h_{veg}, r_{st}, LAI_{min}, LAI_{max}, V_r are vegetation parameters and V_{soi} is the soil storage capacity

B.3 Analysis and processing of FLUXNET data

Processing of forcing data

Measurements of precipitation, air temperature, net radiation, relative humidity and wind speed were gap-filled and then aggregated from 30 min to daily time scale. Missing precipitation data were filled with zero values for short gaps only (less or equal to 3 h). For all other variables, we used the following procedure for gap-filling:

- short gaps (less of equal to 3h) were filled using linear interpolation;
- medium gaps (from 3.5 h to 15 days) were filled using moving window averaging, i.e. the values for same time of the day for the previous and following days were averaged. For each gap we expanded progressively the width of the moving window until a minimum of four values to calculate the average were found. The maximum width of the moving window was 30 days.
- long gaps (from 15 to 80 days) were filled using long term averaging, i.e. for each month, we derived an average value for each time of the day by calculating the average over the entire time series.
- After gap-filling, we could extract for each site a simulation period for which no gap remained.

We then identified the 'poor' months for which the data contained many gaps, and therefore for which the impact of the gap-filling on the simulation results is likely to be significant. 'Poor' months had more than 20 % of the days that contained gap-filled data or were following months in which more than 20 % of the days contained gap-filled data. In fact, after each period of months that contained many gaps, we assumed that the impact of the gap-filling is still significant over a subsequent period containing the same number of months. During such 'poor' months we did not compare model simulations with latent heat and soil moisture observations when applying the soft rules for parameter estimation (Sect. 3.4.1 of the main paper).

Analysis of the uncertainty in observed ET

We analysed the uncertainty in observed ET by calculating the relative difference and the monthly correlation coefficient between the uncorrected actual ET and the Bowen ratio corrected estimates (Eq.(3.17)) and the residual corrected estimates (Eq.(3.18)) at the four FLUXNET sites. Results are reported in Table B.12. We observe that the relative difference can be quite large, especially between the uncorrected and the residual corrected estimates, since the relative difference can be as high as 77 %. We see that the Bowen ratio corrected estimate provides an intermediate value, between the uncorrected and the residual corrected estimate. However, the monthly correlation coefficient was

always high at all sites (above 0.86), which means that all three estimated have similar temporal dynamics.

Therefore, the magnitude of observed actual ET has large uncertainties at the FLUXNET sites, while the temporal dynamic of observed actual ET seems to be well captured by the measurements.

Site	Relative difference [%]		Monthly correlation coefficient [-]	
	$E_{act,bow}$	$E_{act,res}$	$E_{act,bow}$	$E_{act,res}$
German	16	23	0.97	0.94
Spanish	17	76	0.99	0.87
French 1	10	30	0.97	0.91
French 2	34	77	0.97	0.86

Table B.12 Relative difference and correlation coefficient between monthly measured actual ET and monthly corrected actual ET using the Bowen method and the energy residual method at the four FLUXNET sites.

$E_{act,obs}$ monthly measured actual ET, $E_{act,bow}$ corrected actual ET using the Bowen method, $E_{act,res}$ corrected actual ET using the energy residual method.

Estimation of wind speed at the FLUXNET Spanish for the virtual experiment

To setup the virtual experiment, we transformed the wind speed measurements at the Spanish site to estimate their value at the same height as measured at the German site (43.5 m). In fact, at the Spanish site wind speed is measured at a low height (2.5 m), since the vegetation is short. Therefore, to simulate the impact of a change to tall vegetation (forest) at the shrub virtual site, wind speed should be estimated at a height which is above canopy level, as required by the Penman Monteith equation. We assumed a logarithmic wind profile as e.g. in in Lhomme et al. (2014). We note that we modified Eq. (6) in Lhomme et al. (2014), which is valid when the vegetation is fully covering the ground, to account for sparse vegetation. We calculated the value of wind speed at 43.5 m over vegetated and non-vegetated fraction separately using Eq. (6) in Lhomme et al. (2014) and estimated the overall wind speed at 43.5 m for the site as the area weighted value over both fractions. The other climate variables (air temperature and humidity) are assumed to be the same at 43.5 m compared to 2.5 m. We deemed that these assumptions were reasonable, since the objective of the virtual experiment is to understand recharge sensitivity and not to predict future recharge.

B.4 Analysis of the warm-up period

The analyses reported in this section aim to identify an appropriate value of the warm-up period (denoted as H_w), to evaluate V2Karst at the four FLUXNET sites. The warm-up period corresponds to the initial time period which is discarded to reduce the impact of the choice of the value of the model initial states on the simulations. We assessed the sensitivity of the fluxes simulated with V2Karst to H_w by evaluating the model over a range of values of H_w. For a given FLUXNET site, the date of the first day following the warm-up period is kept constant across the simulations (1 January 2001 at the German site, 1 January 2006 at the Spanish site, 1 January 2010 at the French 1 site and 1 April 2003 at the French 2 site). Instead, the date of the first day of the warm-up period is varied according to the value of H_w. In this way, simulated fluxes are assessed over the same time horizon for all values of H_w and therefore simulations using different values of H_w can be compared among each other. We varied H_w between 2 and 12 months and we assessed the sensitivity of the total simulated recharge (Q_{epi}) and actual ET (E_{act}) to H_w by estimating the metrics ΔQ_{epi} [mm] and ΔE_{act} [mm] defined as follows:

$$\Delta Q_{epi}(H_w = h_w) = Q_{epi}(H_w = h_w) - Q_{epi}(H_w = 12)$$

$$\Delta E_{act}(H_w = h_w) = E_{act}(H_w = h_w) - E_{act}(H_w = 12) \quad \text{(B.1)}$$

$$\text{where } h_w = 2, \dots, 11 \ months$$

The two metrics of Eq. (B.1) measure the difference in Q_{epi} and E_{act} when H_w is set to 12 months compared to when H_w is set to lower values. A large value of ΔQ_{epi} or ΔE_{act} means that the choice of H_w has an impact on simulated recharge and actual ET, while a small value of ΔQ_{epi} or ΔE_{act} means that H_w has little effect on the simulation results. Initially, we assumed that the soil and epikarst stores of V2Karst are saturated. For each of the 11 values of H_w that were tested, we repeated the simulations over 1,000 parameter sets sampled using latin hypercube sampling and the ranges reported in Table B.10, and therefore for each site we performed a total number of 11,000 model evaluations.

Figure B.1 reports ΔQ_{epi} (left panels) and ΔE_{act} (right panels) against H_w for the 1,000 parameter sets for each FLUXNET site. We see that when H_w increases, the width of the simulation ensemble decreases, which means that the impact of H_w on the simulations decreases. In general, the value of ΔQ_{epi} and ΔE_{act} becomes very small ($-5\ mm < \Delta Q_{epi} < 5mm$ and $-5\ mm < \Delta E_{act} < 5mm$) when H_w is equal to or larger than 10 months, apart from one parameterisation at the Spanish site for which ΔQ_{epi} and ΔE_{act} becomes very small when H_w is equal to 11 months. Therefore, the simulated fluxes show generally little changes in response to changes in H_w when H_w is higher than 10 months.

Consequently, we deemed reasonable to set the warm-up period equal to 12 months at all FLUXNET sites to perform the parameter estimation and the sensitivity analysis in Chapter 3.

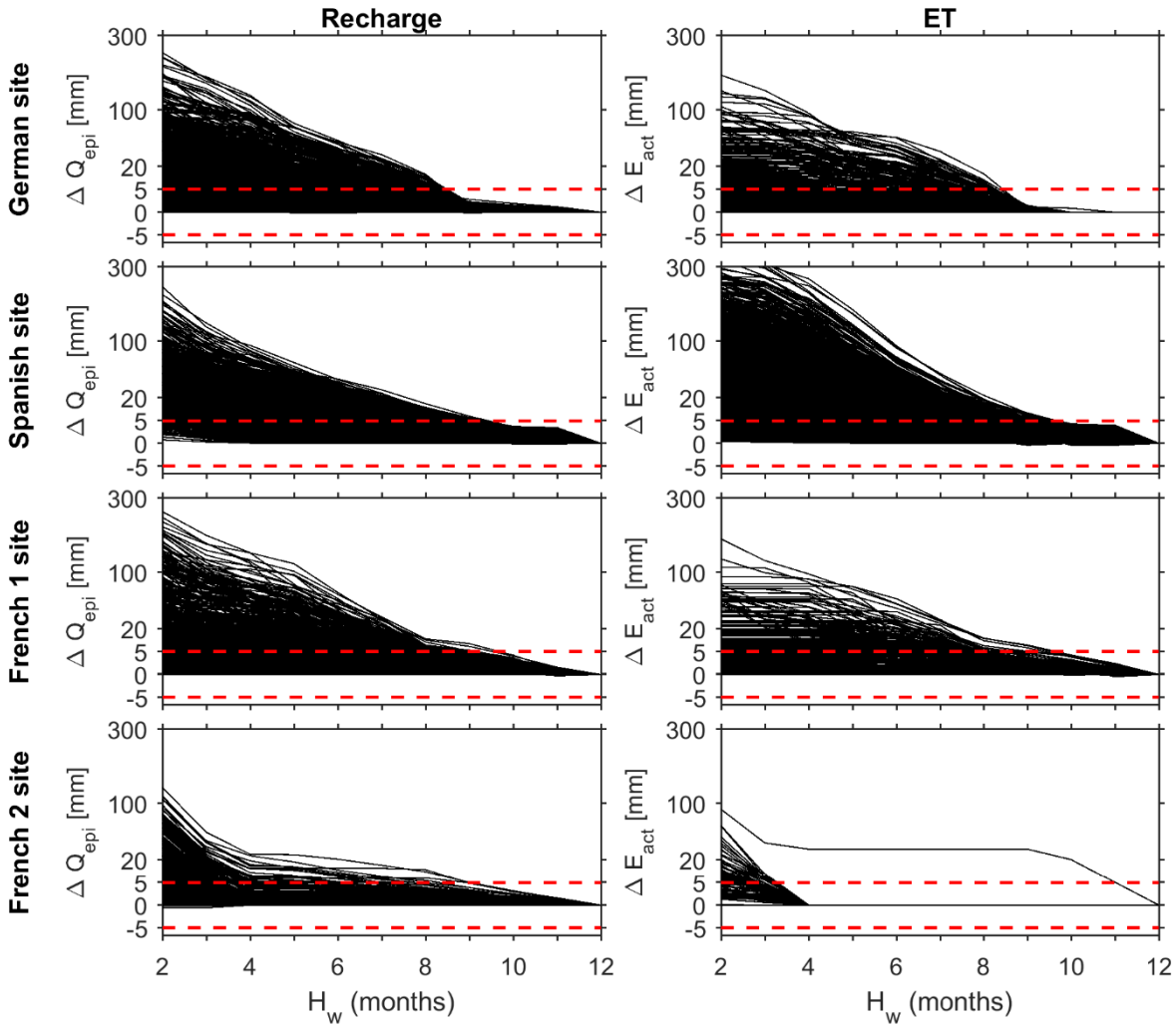


Figure B.1 Difference in simulated recharge and actual ET estimated for varying values of the warm-up period.

ΔQ_{epi} differences in simulated recharge, ΔE_{act} differences in simulate actual ET, H_w length of the warm-up period. The figure reports the difference between recharge (resp. actual ET) simulated when using the value of H_w reported on the x-axis of the plots compared to a value of H_w of 12 months (see Eq. (B.1)) at the four FLUXNET sites.

B.5 Range of variation of precipitation characteristics

This section reports the cumulative distribution function of monthly precipitation P_m [$\mathrm{mm.month^{-1}}$] (Fig. B.2), precipitation intensity I_p [$\mathrm{mm.d^{-1}}$] (Fig. B.3) and interval between wet days H_p [d] (Fig. B.4) for:

- the whole domain, which is all European and Mediterranean carbonate rock areas reported in the carbonate rock map of Williams and Ford (2006) presented in Fig. 31. For this, precipitation from the GLDAS database is used (Rodell et al., 2004);
- the four carbonate rock sites of the FLUXNET network (Baldocchi et al., 2001) analysed in this study presented in Fig. 3.1 and Table 3.4.

These three figures allowed to inform the choice of the ranges of P_m, I_p and H_p to derive the synthetic precipitation inputs used in the virtual experiment presented in Sect. 3.4.3.

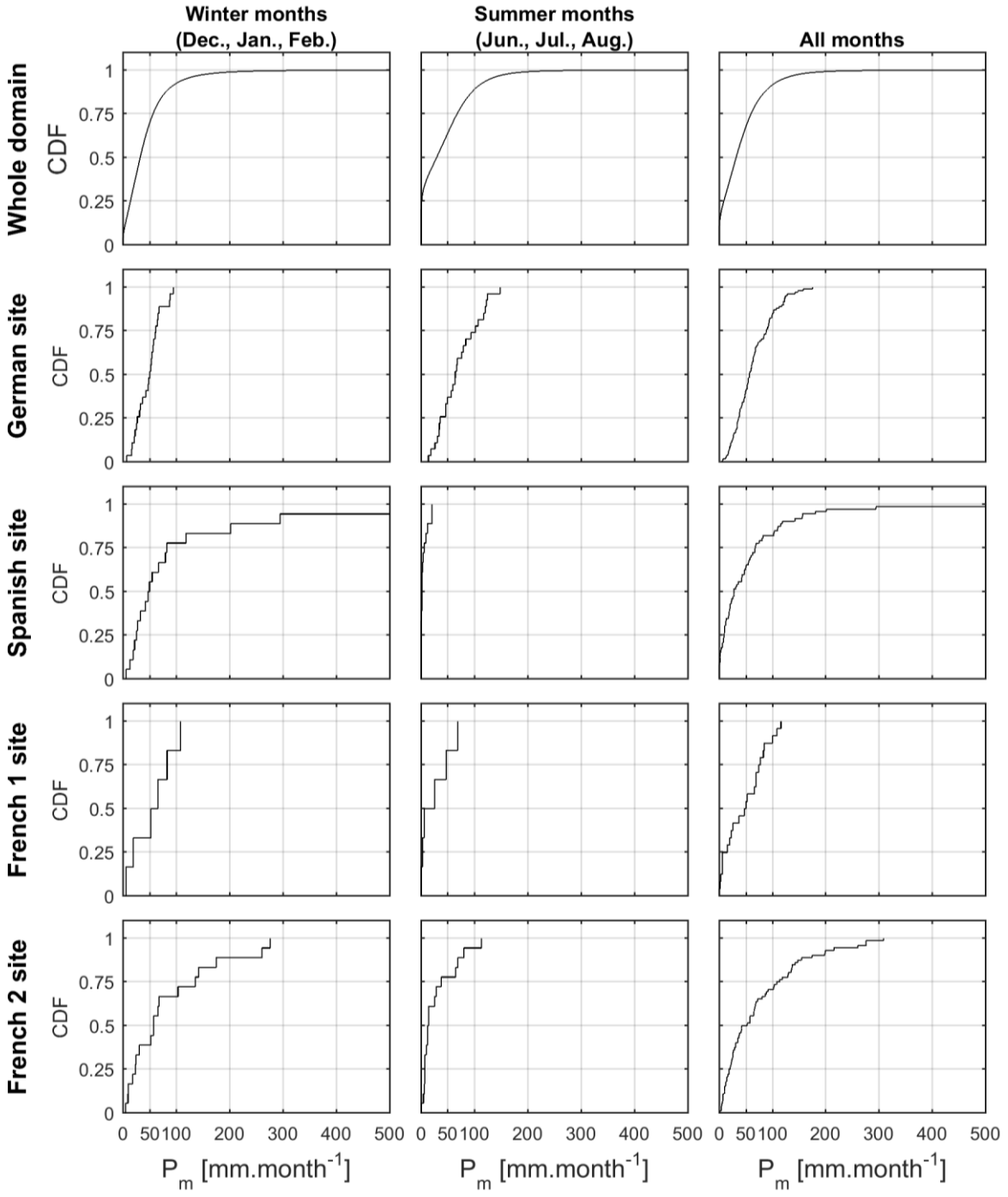


Figure B.2 Cumulative distribution functions of monthly precipitation.
P_m [mm. month^{-1}] monthly precipitation. Distributions were estimated over winter months (Dec., Jan. Feb.), summer months (Jun., Jul., Aug.) and all months of the year, for the whole domain (all European and Mediterranean carbonate rock areas) over the period 1 October 2002–30 September 2012, at the German FLUXNET site over the period 1 January 2001–17 December 2009, at the Spanish FLUXNET site over the period 1 January 2006-30 December 2011, at the French 1 FLUXNET site over the period 1 January 2010–30 December 2011 and at the French 2 FLUXNET site over the period 1 April 2003–31 March 2009.

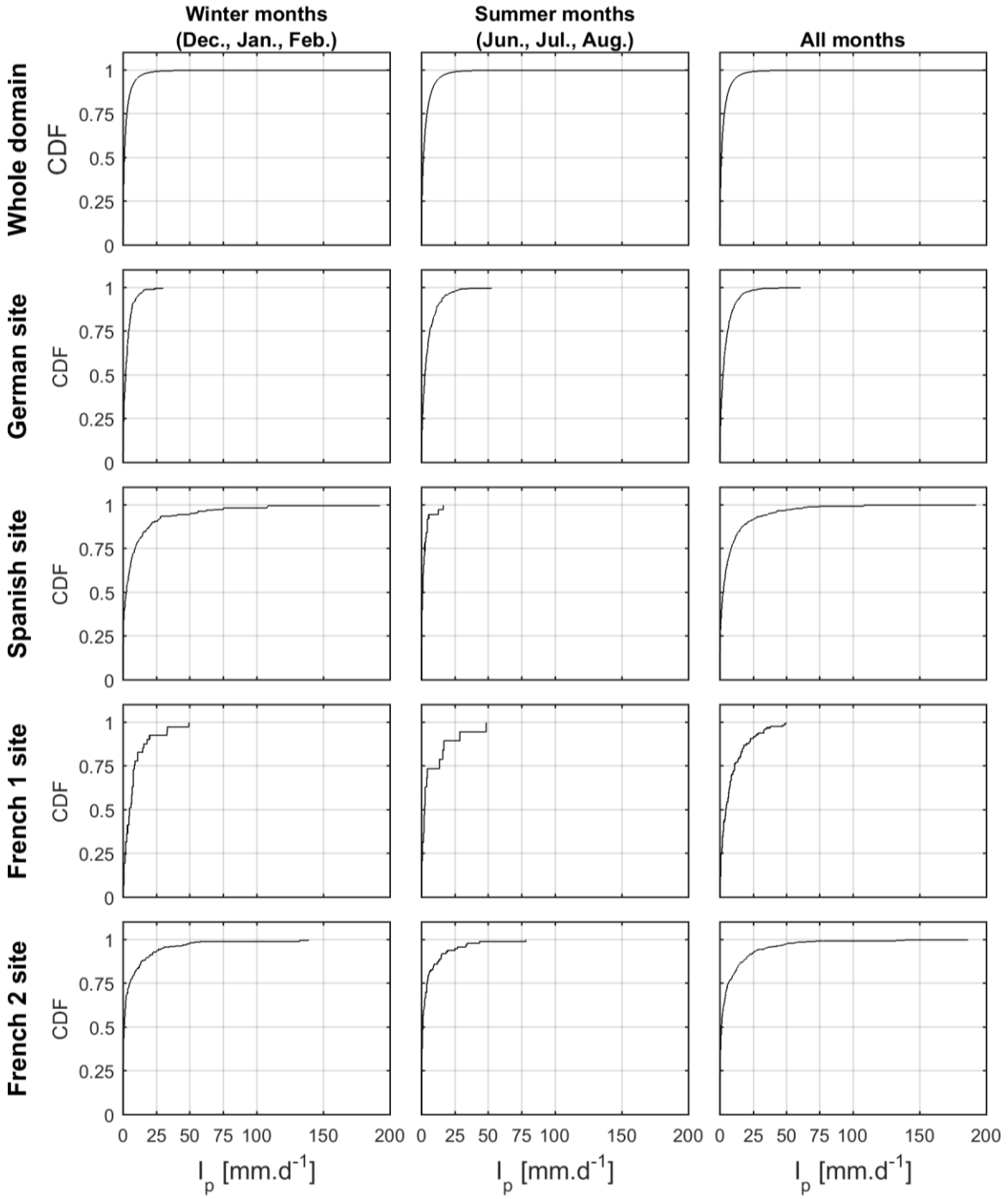


Figure B.3 Cumulative distribution functions of the intensity of precipitation.
I_p [mm. d^{-1}] intensity of precipitation. Distributions were estimated over winter months (Dec., Jan. Feb.), summer months (Jun., Jul., Aug.) and all months of the year, for the whole domain (all European and Mediterranean carbonate rock areas) over the period 1 October 2002–30 September 2012, at the German FLUXNET site over the period 1 January 2001–17 December 2009, at the Spanish FLUXNET site over the period 1 January 2006-30 December 2011, at the French 1 FLUXNET site over the period 1 January 2010–30 December 2011 and at the French 2 FLUXNET site over the period 1 April 2003–31 March 2009. Only days that had a precipitation amount above 0.1 mm were included in the calculation.

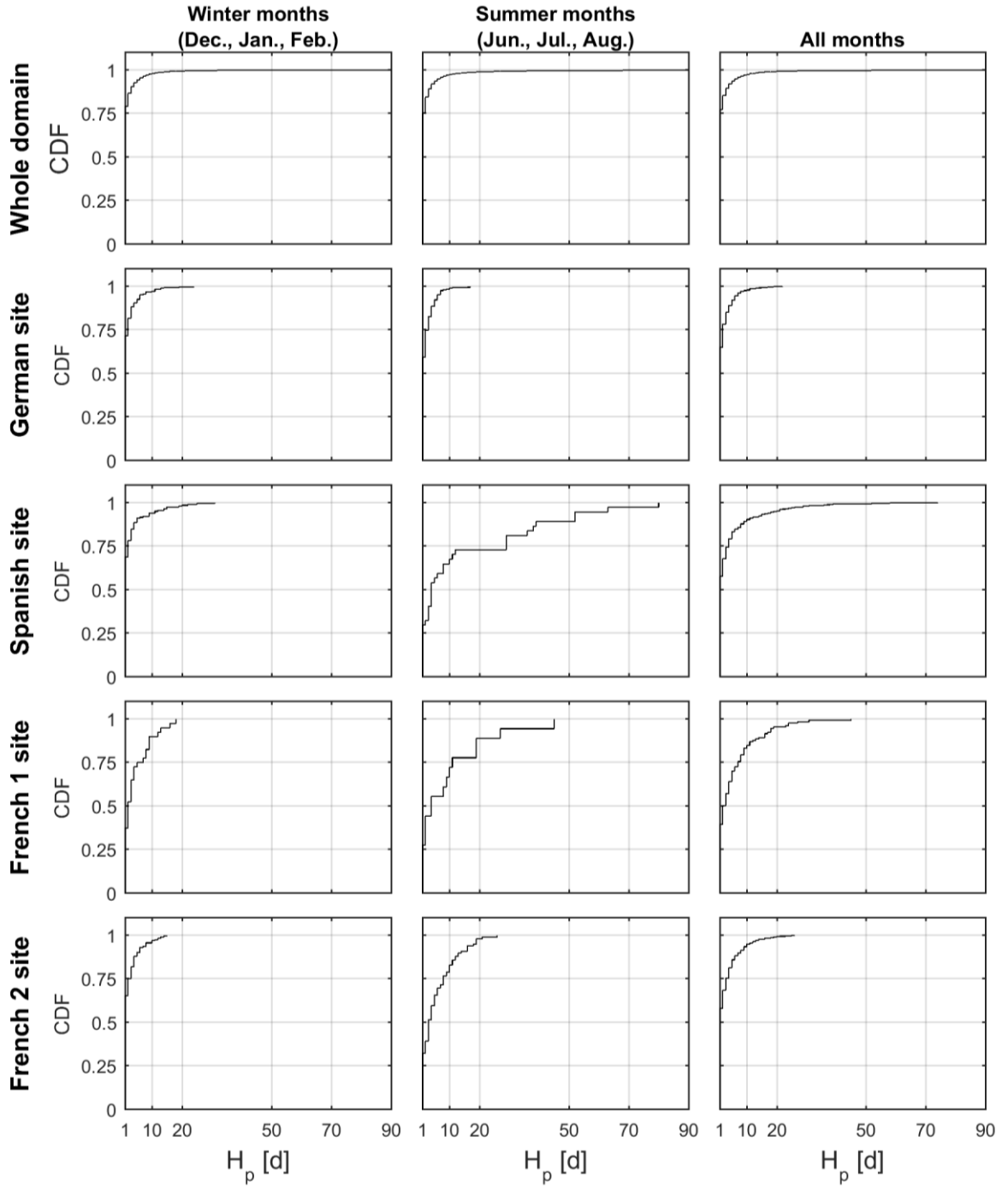


Figure B.4 Cumulative distribution functions of the interval between wet days.
H_p $[d]$ interval between wet days. Distributions were estimated over winter months (Dec., Jan. Feb.), summer months (Jun., Jul., Aug.) and all months of the year, for the whole domain (all European and Mediterranean carbonate rock areas) over the period 1 October 2002–30 September 2012, at the German FLUXNET site over the period 1 January 2001–17 December 2009, at the Spanish FLUXNET site over the period 1 January 2006-30 December 2011, at the French 1 FLUXNET site over the period 1 January 2010–30 December 2011 and at the French 2 FLUXNET site over the period 1 April 2003–31 March 2009. A wet day is defined as a day with more than 0.1 mm of precipitation.

B.6 Sensitivity of the standard deviation of monthly simulated recharge and simulated actual transpiration

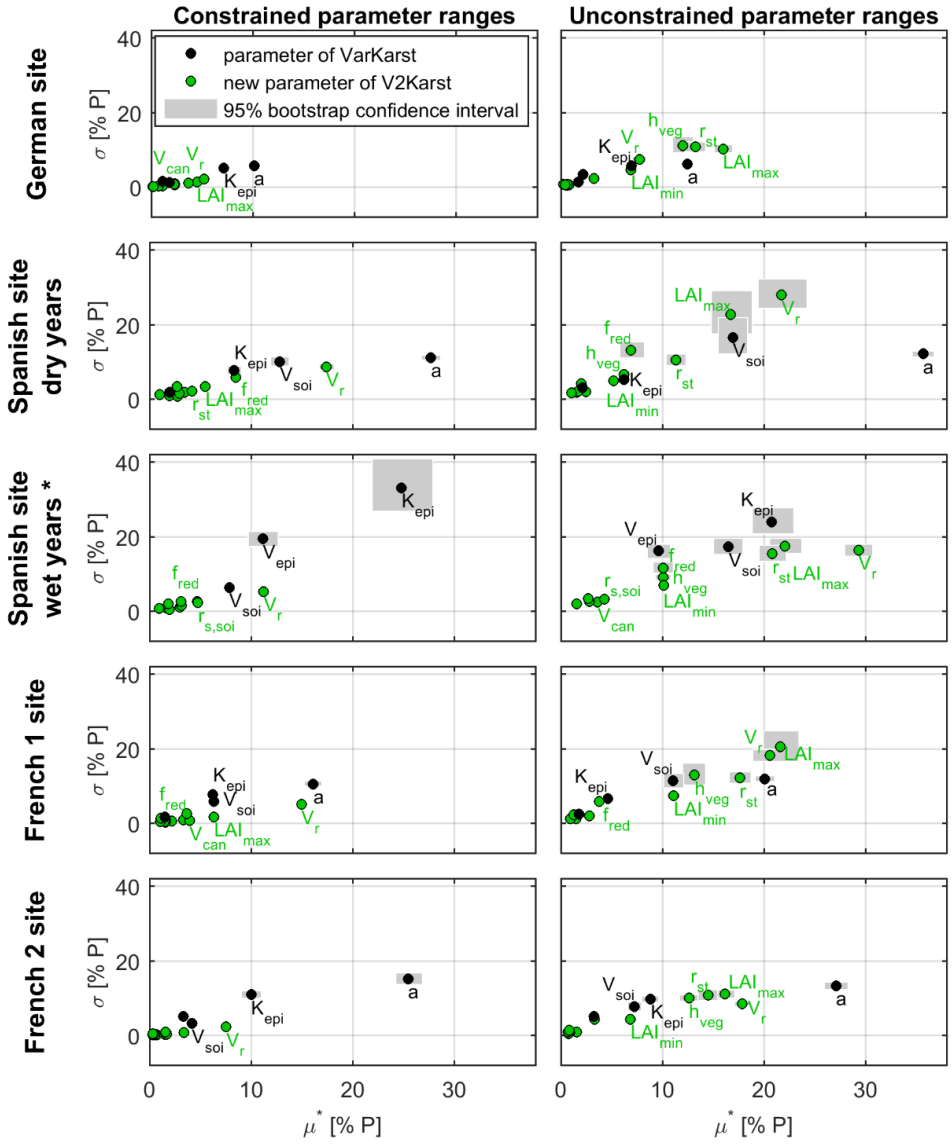


Figure B.5 Sensitivity indices of V2Karst parameters for the standard deviation of simulated monthly recharge at the four FLUXNET sites.

Recharge is expressed as a percentage of mean monthly precipitation. μ^* is the mean of the absolute Elementary Effects and σ is the standard deviation of the Elementary Effects. Results are presents for the constrained (site-specific) and the unconstrained ranges are used). Sensitivity indices were computed over the period 1 January 2001-17 December 2009 for the German site, 1 January 2006-31 December 2008 for the Spanish site (dry years), 1 January 2009-30 December 2011 for the Spanish site (wet years), 1 January 2010-30 December 2011 for the French 1 site and 1 April 2003-31 March 2009 for the French 2 site.

* Sensitivity indices for parameter a are not reported in the plots for the Spanish site wet years because they are significantly higher than the other parameters ($\mu_a^* = 68$ % and $\sigma_a = 51$ % for constrained ranges and $\mu_a^* = 68$ % and $\sigma_a = 38$ % for unconstrained ranges).

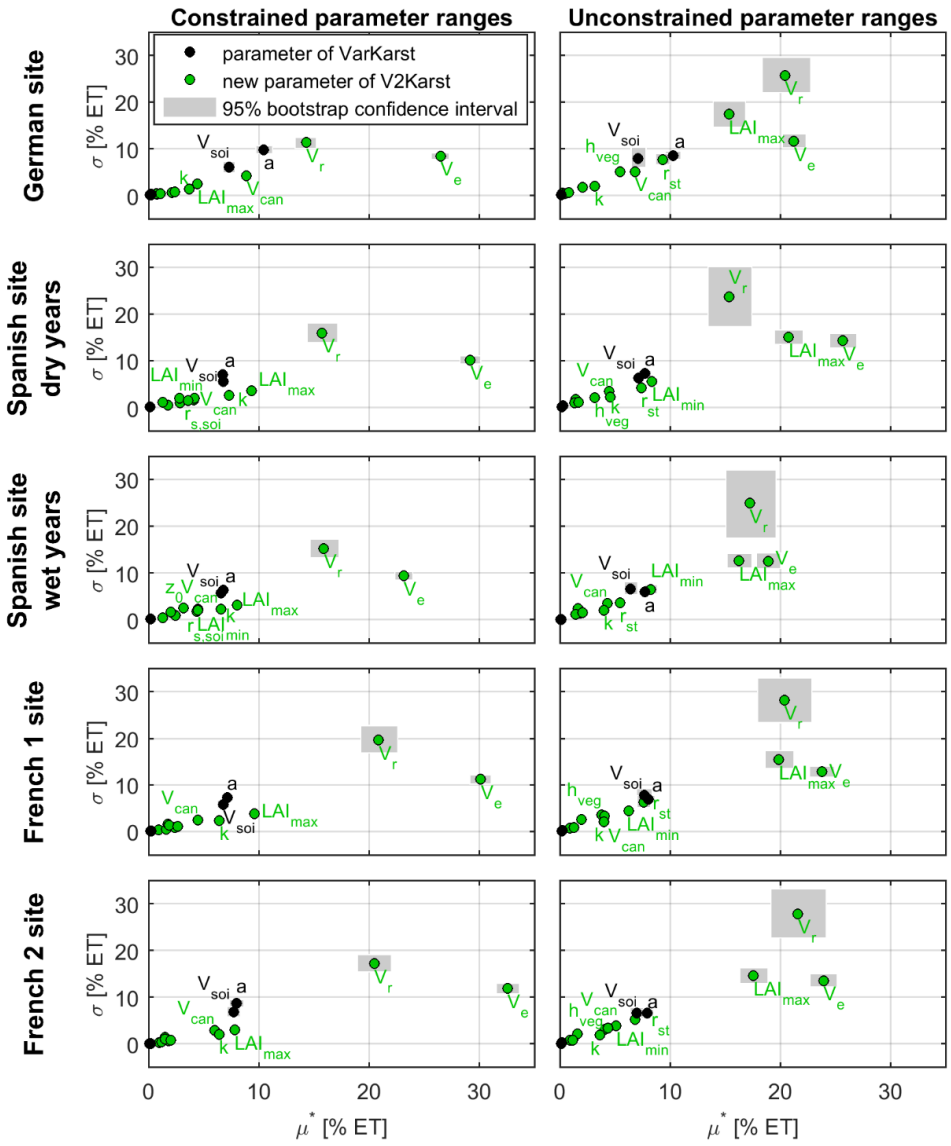


Figure B.6 Sensitivity indices of V2Karst parameters for the standard deviation of simulated actual transpiration at the four FLUXNET sites.

Actual transpiration is expressed as a percentage of total ET. μ^* is the mean of the absolute Elementary Effects and σ is the standard deviation of the Elementary Effects. Results are presents for the constrained (site-specific) and the unconstrained ranges are used). Sensitivity indices were computed over the period 1 January 2001-17 December 2009 for the German site, 1 January 2006-31 December 2008 for the Spanish site (dry years), 1 January 2009-30 December 2011 for the Spanish site (wet years), 1 January 2010-30 December 2011 for the French 1 site and 1 April 2003-31 March 2009 for the French 2 site.

* Sensitivity indices for parameter a are not reported in the plots for the Spanish site wet years because they are significantly higher than the other parameters ($\mu_a^* = 68$ % and $\sigma_a = 51$ % for constrained ranges and $\mu_a^* = 68$ % and $\sigma_a = 38$ % for unconstrained ranges).

Appendix C. Supplements to Chapter 4

Appendix C is part of a work in preparation:

> *Sarrazin F, Hartmann A, Pianosi F, Rosolem R, Wagener, T., How do land cover and subsurface heterogeneity modulate climate change impacts on future groundwater recharge in karst areas?, in preparation.*

The main research for this work supports Chapter 4.

C.1 References for V2Karst vegetation parameter ranges

Reference	Parameter values											
	h_{veg} $[m]$		r_{st} $[s.m^{-1}]$		LAI_{min} [%]		LAI_{max} $[m^2.m^{-2}]$		V_r $[mm]$ [b]		α_{veg} $[-]$	
	Lower limit	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit
LAD model (Milly and Shmakin, 2002)	n. s.	n. s..	n. s.	n. s.	n. s.	n. s.	n. s.	n. s.	120	220	0.13	0.11
(Cox et al., 1999)	10.0	29.4	n. s.	n. s.	n. s.	n. s.	4.0	9.0	180	300	0.12	0.14
Mac-PDM.09 model (Gosling and Arnell, 2011)	10.0	29.4	n. s.	n. s.	n. s.	n. s.	4.0	9.0	180	300	n. c.	n. c.
WaterGap 2.2 model (Müller Schmied et al., 2014)	n. s.	n. s.	n. s.	n. s.	10	100	4.02	4.78	400	800	0.07	0.13
(Kergoat, 1998)	n. c.	n. c.	167	217	n. s.	n. s.	Equilibrium model		240	300	0.19 [c]	0.19 [c]
Noah model (Wei et al., 2013)	n. s.	n. s.	n. s.	n. s.	17	83	5.99	6.0	n. s.	n. s.	n. s.	n. s.
Noah model (Kumar et al., 2011)	n. s.	n. s.	100	150	48	100	3.31	6.48	n. s.	n. s.	0.12 [d]	0.25 [d]
(Masson et al., 2003)	n. s.	n. s.	150	250	n. s.	n. s.	n. s.	n. s.	n. s.	n. s.	0.10	0.15
HTESSEL model (Boussetta et al., 2013)	n. s.	n. s.	175	500	n. s.	n. s.	5.0	6.0	n. s.	n. s.	n. s.	n. s.
CLM v4.5 (Oleson et al., 2013)	14.0	35.0	n. s.	n. s.	Cell-specific monthly value based on MODIS product				n. s.	n. s.	n. s.	n. s.
(Hagemann, 2002) (used iPCR-GLOBWB model (Van Beek et al., 2011)	n. s.	n. s.	n. s.	n. s.	2	100	3.4	9.9	n. s.	n. s.	0.12	0.16
(Oke, 1987) as reported in (Bonan, 2015b)	n. s.	n. s.	n. s.	n. s.	n. s.	n. s.	n. s.	n. s.	n. s.	n. s.	0.05	0.20
VIC model in the GLDAS project (NASA, 2018)	n. s.	n. s.	80	120	34	77	4.4	5.0	200 [c]	200 [c]	0.12	0.18
Mosaic model in the GLDAS project (NASA, 2018)	n. s.	n. s.	n. s.	n. s.	7	81	5.12	10.76	n. s.	n. s.	n. s.	n. s.
(Breuer et al., 2003) [a]	n. s.	n. s.	278	532	n. s.	n. s.	4.3	6.3	260	400	0.10 [d]	0.29 [d]
Ranges used in Chapter 4	**10**	**35**	**100**	**500**	**5**	**100**	**5**	**10**	**200**	**800**	**0.1**	**0.16**

Table C.1 Ranges of six vegetation parameters for <u>forest</u> land cover in previous studies and in Chapter 4.

Ranges used in Chapter 4 were set to include the ranges found in previous studies. Records highlighted in black were considered as outliers and were discarded. h_{veg} vegetation height, r_{st} stomatal resistance, LAI_{min} reduction in leaf area index during the dormant season, LAI_{max} annual maximum leaf area index, V_r maximum storage capacity of the root zone, V_{can} canopy storage capacity per unit of leaf area index, α_{veg} vegetation albedo. n.s. value not specified.

[a] Reported values include the 40th and 60th percentile reported in Breuer et al. (2003) for coniferous and deciduous forests.

[b] V_r in previous studies was calculated as the product between the rooting depth for forest land cover and an average value of soil available water capacity of 0.2 m^3m^{-3} (Bonan, 2015a; Miralles et al., 2011; Salter and Williams, 1965).

[c] Constant parameter across vegetation types.

[d] The range of values covers the seasonal variations in albedo, while in V2karst the seasonality of albedo is not considered.

Reference	Parameter values											
	h_{veg} $[m]$		r_{st} $[s.m^{-1}]$		LAI_{min} [%]		LAI_{max} $[m^2.m^{-2}]$		V_r $[mm]$ [b]		α_{veg} $[-]$	
	Lower limit	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit
LAD model (Milly and Shmakin, 2002)	n. s.	n. s.	n. s.	n. s.	n. s.	n. s.	n. s.	n. s.	8	120	0.16	0.20
(Cox et al., 1999)	0.4	1.0	n. s.	n. s.	n. s.	n. s.	2.0	5.0	100	160	0.12	0.25
Mac-PDM.09 model (Gosling and Arnell, 2011)	0.6	14.9	n. s.	n. s.	n. s.	n. s.	3.0	5.0	120	240	n. s.	n. s.
WaterGap 2.2 model (Müller Schmied et al., 2014)	n. s.	n. s.	n. s.	n. s..	10	50	1.71	3.62	200	200	0.18	0.25
(Kergoat, 1998)	n. s.	n. s.	128	128	n. s.	n. s..	Equilibrium model		120	120	0.19 [c]	0.19 [c]
Noah model (Wei et al., 2013)	n. s.	n. s.	n. s.	n. s.	25	26	2.64	3.0	n. s.	n. s.	n. s.	n. s.
Noah model (Kumar et al., 2011)	n. s.	n. s.	40	40	18	100	2.9	5.68	n. s.	n. s.	0.17 [d]	0.25 [d]
(Masson et al., 2003)	n. s.	n. s.	40	120	n. s.	n. s.	n. s.	n. s.	n. s.	n. s.	0.20	0.20
HTESSEL model (Boussetta et al., 2013)	n. s.	n. s.	100	180	n. s.	n. s.	2.0	3.0	n. s.	n. s.	n. s.	n. s.
CLM v4.5 (Oleson et al., 2013)	0.5	0.5	n. s.	n. s.	Cell-specific monthly value based on MODIS product				n. s.	n. s.	n. s.	n. s.
(Hagemann, 2002) (used in the PCR-GLOBWB model (Van Beek et al., 2011)	n. s.	n. s.	n. s.	n. s.	0	46	1.5	4.6	n. s.	n. s.	0.15	0.20
(Oke, 1987) as reported in (Bonan, 2015b)	n. s.	n. s.	n. s.	n. s.	n. s.	n. s.	n. s.	n. s.	n. s.	n. s.	0.16	0.26
VIC model in the GLDAS project (NASA, 2018)	n. s.	n. s.	80	80	0.4	52	3.85	5.0	200 [c]	200 [c]	0.10 [d]	0.20 [d]
Mosaic model in the GLDAS project (NASA, 2018)	n. s.	n. s.	n. s.	n. s.	16	16	4.78	4.78	n. s.	n. s.	n. s.	n. s.
(Breuer et al., 2003) [a]	1.0	1.5	163	275	n. s.	n. s.	3.4	6.5	140	300	0.2 [d]	0.28 [d]
Ranges used in Chapter 4	**0.4**	**1.5**	**40**	**250**	**5**	**100**	**1.5**	**5**	**100**	**200**	**0.16**	**0.25**

Table C.2 Ranges of six vegetation parameters for <u>grass/crop</u> land cover in previous studies and in Chapter 4.

Ranges used in Chapter 4 were set to include the ranges found in previous studies. Records highlighted in black were considered as outliers and were discarded. h_{veg} vegetation height, r_{st} stomatal resistance, LAI_{min} reduction in leaf area index during the dormant season, LAI_{max} annual maximum leaf area index, V_r maximum storage capacity of the root zone, V_{can} canopy storage capacity per unit of leaf area index, α_{veg} vegetation albedo. n.s. value not specified.

[a] Reported values include the 40th and 60th percentile reported in Breuer et al. (2003) for herbs, forbs, grasses and crops.

[b] V_r in previous studies was calculated as the product between the rooting depth for forest land cover and an average value of soil available water capacity of 0.2 m^3m^{-3} (Bonan, 2015a; Miralles et al., 2011; Salter and Williams, 1965).

[c] Constant parameter across vegetation types.

[d] The range of values covers the seasonal variations in albedo, while in V2karst the seasonality of albedo is not considered.

C.2 Testing of the longwave upwelling radiation model proposed for V2Karst

Daily longwave upwelling radiation simulated with V2Karst V1.1 at time $t\ [d]$, $LW_{up,sim}(t)\ [MJ.m^{-2}\ d^{-1}]$, is assessed as follows:

$$LW_{up,sim}(t) = \sigma(T(t) + 273.15)^4 \quad \text{(C.1)}$$

where $T(t)\ [°C]$ is the air temperature, $\alpha\ [-]$ is the surface albedo, $\varepsilon\ [-]$ is the surface emissivity and $\sigma\ [MJ.\ K^{-4}\ m^{-2}\ d^{-1}]$ is the Stefan-Boltzmann constant equal to $4.90\ 10^{-9}\ MJ.\ K^{-4}\ m^{-2}\ d^{-1}$. We note that Eq. (C.1) assumes a value of the surface emissivity equal to 1 and uses air temperature, which differs from the theoretical equation which uses surface temperature (e.g. Bonan, 2015b). To assess the applicability of Eq. (C.1), we compared observed and calculated daily longwave upwelling radiation ($LW_{up,sim}$ and $LW_{up,obs}$) at sites from the FLUXNET network (Baldocchi et al., 2001) that are presented in Table C3. The sites cover a range of climates and land cover types. We aggregated the observations of temperature and longwave upwelling radiation from half-hourly to daily time scale. To test the model, we considered only days for which no half-hourly observations were missing. Daily time series of observed and simulated LW_{up} are reported in Fig. C.1.

Site name	Short name	Country	Coordinates	Land cover	Reference	Number of days used for testing
Hainich	DE-Hai	Germany	51°04′45″N, 10°27′07″E	deciduous forest	(Pinty et al., 2011)	3579
Soroe	DK-Sor	Denmark	55°29’13’’N, 11°38’45’’E	deciduous forest	(Pilegaard et al., 2001)	1278
Font Blanche	FR-FBn	France	43°14′27″N, 5°40′45″E	evergreen forest	(Simioni et al., 2013)	885
Puéchabon	FR-Pue	France	43°44’29’’N, 3°35’45’’E	evergreen forest	(Reichstein et al., 2002)	3277
Yatir	IL-Yat	Israel	31°20’49’’N, 35°03’07’’E	evergreen forest	(Grünzweig et al., 2003)	2038
Collelongo	IT-Col	Italy	41°50’58’’N, 13°35’17’’E	deciduous forest	(Papale et al., 2015)	2535
Lavarone	IT-Lav	Italy	45°57’22’’N, 11°16’53’’E	mixed forest	(Papale et al., 2015)	3920
Rotholz	AT-Rtz	Austria	47°23’N, 11°48’E	grassland	(Wohlfahrt et al., 2010)	1812
Früebüel	CH-Fru	Switzerland	47°06'57.0" N, 8°32'16.0" E	grassland	(ETH Zürich, 2018)	2350
Lusignan	FR-Lus	France	46°24’56’’N, 0°7’9’’E	grassland	(Carrer et al., 2012)	336
Amplero	IT-Amp	Italy	41°54’15’’N, 13°36’19’’E	grassland	(Papale et al., 2015)	441

Table C.3 Description of the FLUXNET sites used for testing the longwave upwelling radiation model.
The number of days used for testing the longwave upwelling radiation model of V2Karst is the number of days for which no half-hourly observations of air temperature and longwave upwelling radiation were missing.

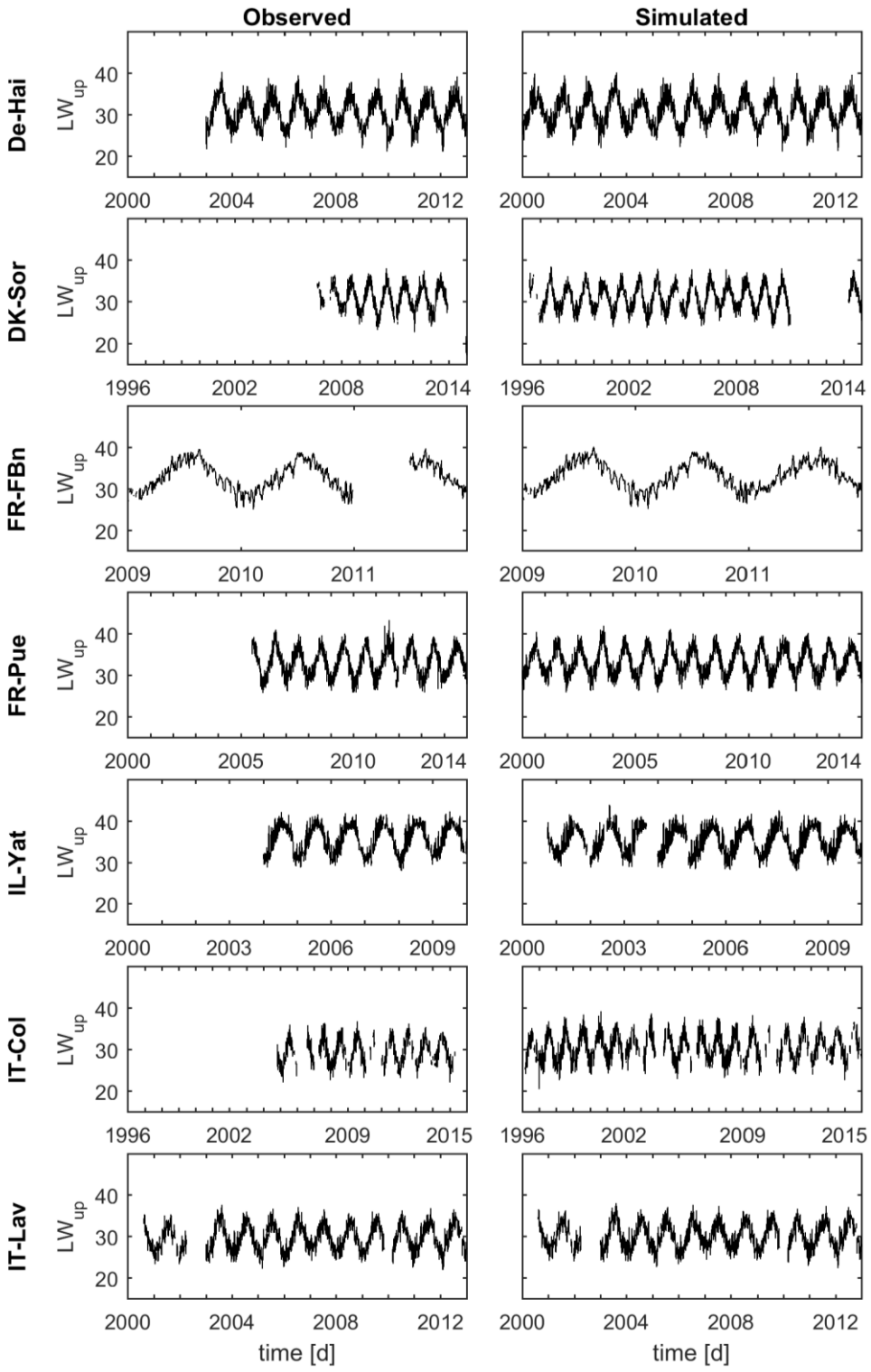


Figure C.1 Daily time series of observed and simulated longwave upwelling radiation at the FLUXNET site.

LW_{up} $[MJ.m^{-2}\ d^{-1}]$ longwave upwelling radiation. Ticks on the x-axis indicate January 1st for each year. The time series are discontinuous because some observations of temperature (to assess net longwave radiation from Eq.(C.1)) and some observations of upwelling longwave radiation are missing. FLUXNET sites are presented in Table C.3.

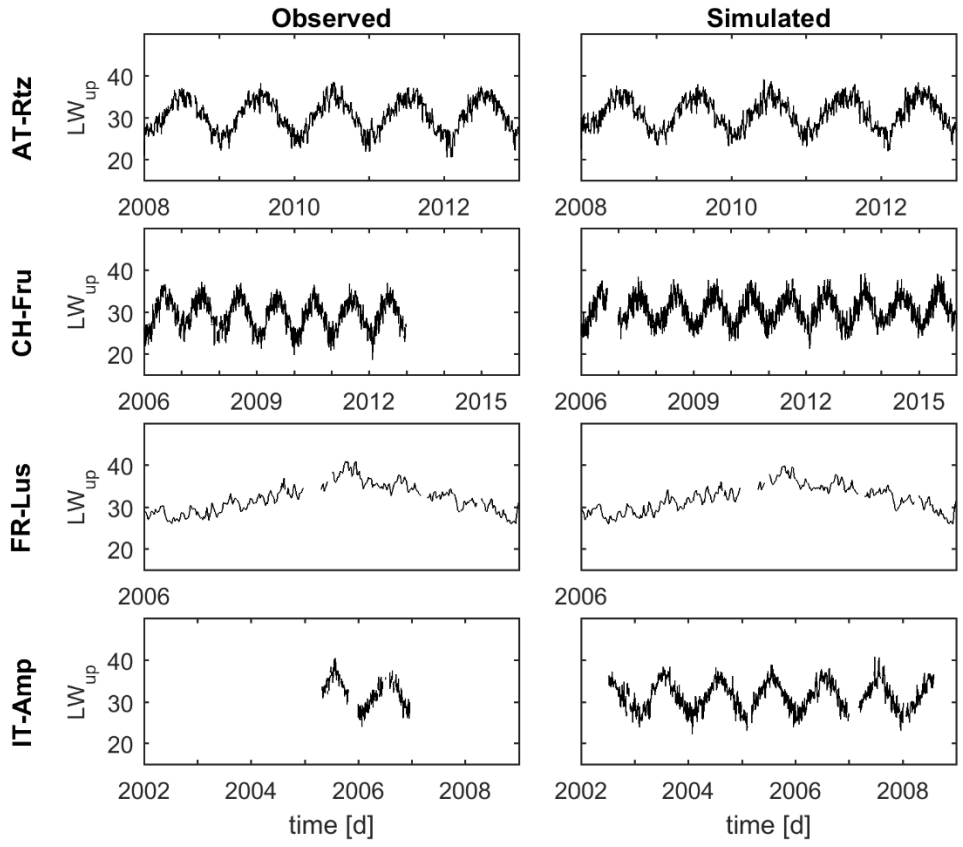


Figure C.1 continued.

A visual comparison of the scatter plots of observed against simulated LW_{up} (Fig. C.2) shows a general agreement between the two variables, since they usually plot along the 1:1 line apart from the CH-Fru site where we observe more scatter. From Fig. C.2 we also see that for the DK-Sor and FR-Pue sites, a few points are distant from the 1:1 line. This is likely to be due to measurement errors. Additionally, we defined three performance metrics to quantify the level of agreement between $LW_{up,sim}$ and $LW_{up,obs}$:

1. The daily Pearson correlation coefficient $\rho_{pear}[-]$;
2. the $Bias$ [%]:

$$Bias = 100\frac{\overline{LW_{up,sim}} - \overline{LW_{up,obs}}}{\left|\overline{LW_{up,obs}}\right|}, \quad \text{(C.2)}$$

   where $\overline{LW_{up,sim}}$ $[MJ.m^{-2}\,d^{-1}]$ and $\overline{LW_{up,obs}}$ $[MJ.m^{-2}\,d^{-1}]$ are the average daily value of the simulated and observed upwelling longwave radiation respectively;
3. the average of the daily Relative Error RE_m[%]:

$$RE_m = \frac{1}{H}\sum_{t=1}^{H}\left(100\left|\frac{LW_{up,sim}(t) - LW_{up,obs}(t)}{LW_{up,obs}(t)}\right|\right), \quad \text{(C.3)}$$

where H is the number of days used for testing;

4. the 95% quantile of the daily Relative Error $RE_{95}[\%]$:

$$RE_{95} = Q_{\underset{t}{0.95}}\left(100\left|\frac{LW_{up,sim}(t) - LW_{up,obs}(t)}{LW_{up,obs}(t)}\right|\right). \quad \text{(C.4)}$$

Results are reported in Table C.4 and indicate that ρ_{pear} is always very high (higher than 0.96) and therefore the dynamic of simulated and observed LW_{up} are consistent. $Bias$ is generally positive, which indicates that the model tends to overestimate observed LW_{up} at the FLUXNET sites analysed. $Bias$ always takes small values below 4%. Finally, RE_m is always below 4.2% and for most sites RE_{95} is below 5.1% apart from one site (FR-Lus) site for which it reaches 11.5%.

Therefore, our analyses shows that Eq. (C.1) generally produce reasonable results at the FLUXNET sites analysed.

Site	ρ_{pear} $[-]$	$Bias$ [%]	RE_m [%]	RE_{95} [%]
DE-Hai	1.00	0.6	0.8	1.8
DK-Sor	0.96	0.5	0.9	1.4
FR-FBn	1.00	0.9	1.0	2.5
FR-Pue	0.99	0.9	1.1	2.3
IL-Yat	0.99	- 1.2	1.5	3.9
IT-Col	1.00	3.3	3.3	5.1
IT-Lav	1.00	1.6	1.6	3.4
AT-Rtz	0.99	1.1	1.7	5.0
CH-Fru	0.96	3.8	4.2	11.5
FR-Lus	0.99	0.3	1.4	3.1
IT-Amp	0.97	- 0.9	1.8	4.7

Table C.4 Performance of the longwave upwelling radiation model at the FLUXNET sites. $\boldsymbol{\rho_{pear}}$ Spearman correlation coefficient, $BIAS$ (Eq. (C.2)), RE_m average of the daily Relative Error (Eq.(C.3)), RE_{95} 95% quantile of the daily Relative Error (Eq.(C.4)). FLUXNET sites are presented in Table C.3.

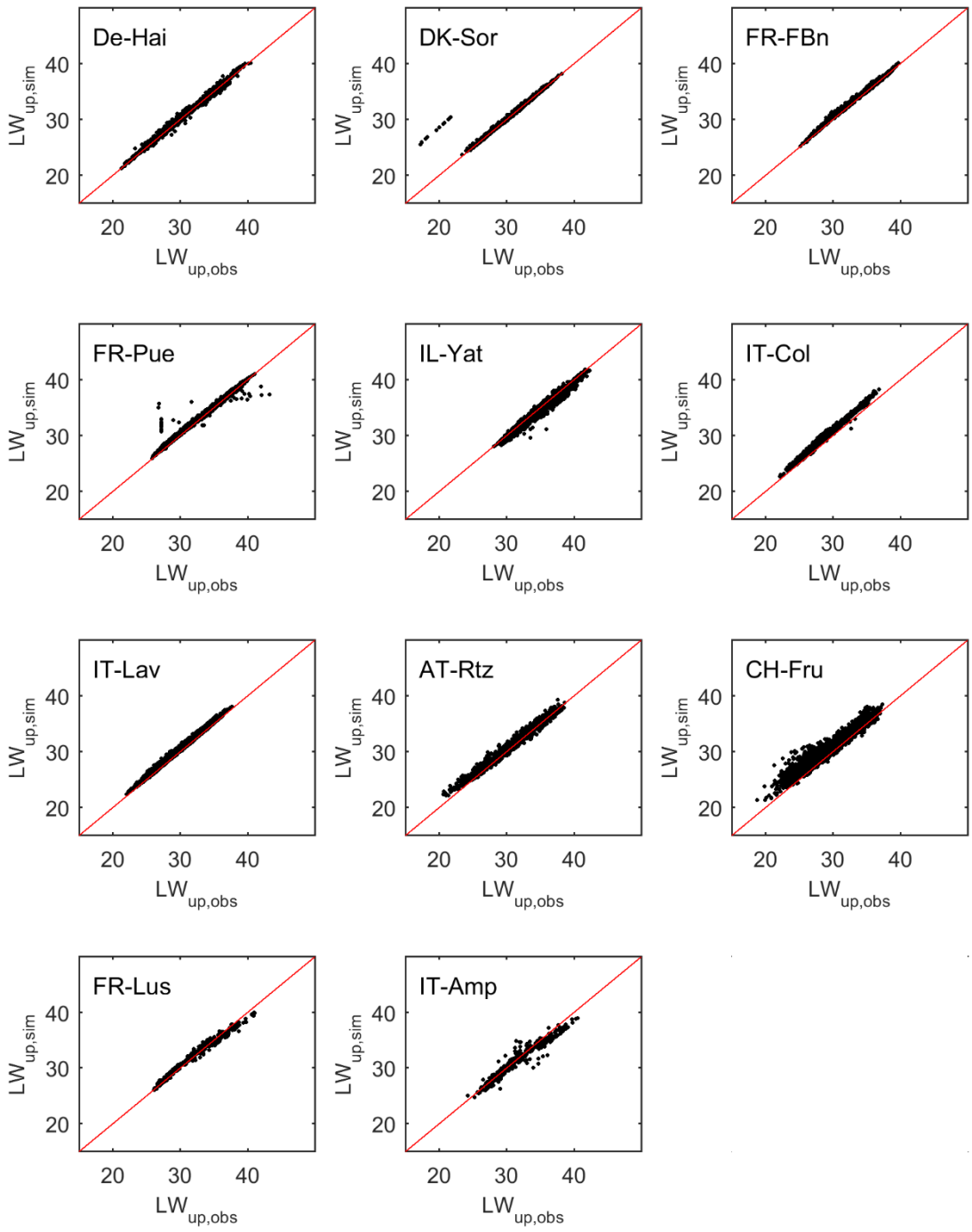


Figure C.2 Simulated against observed daily longwave upwelling radiation at the FLUXNET sites. $LW_{upsim}[MJ.m^{-2}\ d^{-1}]$ simulated longwave upwelling radiation, $LW_{up,obs}\ [MJ.m^{-2}\ d^{-1}]$ observed longwave upwelling radiation. The 1:1 line is plotted in red. FLUXNET sites are presented in Table C.3.

C.3 Correlation analysis and selection of climate descriptors

Climate descriptor	Unit	Description and reference
$\boldsymbol{P_m}$	**[mm.y^{-1}]**	**Mean annual precipitation.**
[b]PET_m	[mm**.y^{-1}**]	Mean annual potential evapotranspiration. PET_m using the Penman Monteith equation (Monteith, 1965) for a reference crop surface with complete canopy cover, a height of 0.12 m, a surface resistance of 70 s.m^{-1} and an albedo of 0.23 following Allen et al. (1998). In this way, PET_m captures the variability in evaporative demand due to climatic conditions only and does not depend on the land cover for the specific cell.
[b]T_m	[°C]	Mean temperature.
[b]DTR_m	[°C]	Mean diurnal temperature range, defined as the difference between daily maximum and minimum temperature (Davy et al., 2017; Lewis and Karoly, 2013).
[b]RH_m	[%]	Mean relative humidity.
[b]WS_m	[m.s^{-1}]	Mean wind speed.
[b]$R_{down,m}$		Mean downwelling radiation. $R_{down,m} = SW_{down,m} + LW_{down,m}$ where $SW_{down,m}$ and $LW_{down,m}$[MJ.m^{-2}.d^{-1}] are the mean shortwave and longwave downwelling radiation respectively.
$\boldsymbol{AI}$	**[-]**	**Aridity index** (e.g. Zhang et al., 2001)**:** $\boldsymbol{AI = \frac{PET_m}{P_m}}$.
$\boldsymbol{\rho_{P-PET}}$	**[-]**	**Correlation coefficient between mean monthly precipitation and mean monthly potential evapotranspiration (a similar descriptor was used in Troch et al., 2013 but temperature was used instead of PET).**
$\boldsymbol{P_{SI}}$	**[-]**	**Precipitation seasonality index (Beck et al., 2015; Troch et al., 2013; Walsh and Lawler, 1981):** $\boldsymbol{P_{SI} = \frac{1}{P_m}\sum_{i=1}^{12}\left\lvert P_i - \frac{P_m}{12}\right\rvert}$, **where $\boldsymbol{P_i[mm.month^{-1}]}$ is the mean monthly precipitation for the $\boldsymbol{i}$-th month of the year.**
[a]P_{CI}	[-]	Precipitation concentration index (Oliver, 1980; Raziei, 2017): $P_{CI} = 100\frac{\sum_{i=1}^{12} P_i^2}{(\sum_{i=1}^{12} P_i)^2}$, where $P_i[mm.month^{-1}]$ is the mean monthly precipitation for the ith month of the year.
[a]P_{RE}	[-]	Precipitation relative entropy (Feng et al., 2013; Raziei, 2017): $P_{RE} = \sum_{i=1}^{12}\frac{P_i}{P_m} log_2\left(12\frac{P_i}{P_m}\right)$, where $P_i[mm.month^{-1}]$ is the mean monthly precipitation for the i-th month of the year. The terms of the sum are set to 0 when $P_i = 0$.
[a]P_{CV}	[-]	Monthly precipitation coefficient of variation.
[b]PET_{SI}	[-]	Potential evapotranspiration seasonality index (same definition as P_{SI})
[a]PET_{CI}	[-]	Potential evapotranspiration concentration index (same definition as P_{CI})
[a]PET_{RE}	[-]	Potential evapotranspiration relative entropy (same definition as P_{RE})
[a]PET_{CV}	[-]	Monthly potential evapotranspiration coefficient of variation.

Table C.5 Definition of the climate descriptors analysed.

Descriptors highlighted in bold were selected by the CART algorithm in Chapter 4 (Fig. 4.4 and 4.5).

[a]Descriptor not included in the sensitivity analyses of Chapter 4 (CART) because a correlation analyses showed that it is redundant (correlation analysis results are reported in Table C.5).

[b] Descriptor included in the sensitivity analyses of Chapter 4, but that was not selected as by the CART algorithm.

Climate descriptor	Unit	Description and reference
$\boldsymbol{I_p}$	**[mm.d^{-1}]**	**Mean precipitation intensity defined as the mean daily precipitation over wet days, i.e. days with precipitation totals of 1mm or higher** (Giorgi et al., 2011)**.**
[a]I_{95}	[mm.d^{-1}]	Intensity of heavy precipitation defined as the 95% quantile of daily precipitation over wet days, i.e. days with precipitation totals of 1mm or higher (IPCC, 2014; Salinger and Griffiths, 2001).
[a]I_{75}	[mm.d^{-1}]	Intensity of heavy precipitation defined as the mean daily precipitation of the upper quartile of daily precipitation as used in Hartmann et al. (2017). As for I_p and I_{95}, I_{75} was calculated here over wet days, i.e. days with precipitation totals of 1mm or higher.
[b]H_p	[d]	Mean interval between wet days that are days with precipitation totals of 1mm or higher.
[b]H_{95}	[d]	95% quantile of the length of dry spells. Dry spells are defined as periods of at least five consecutive days with daily precipitation below 1 mm (IPCC, 2014).
$\boldsymbol{D_p}$	**[d]**	**Mean precipitation duration, defined as the number of consecutive wet days that are days with precipitation totals of 1mm or higher.**
[a]D_{95}	[d]	95% quantile of rainfall duration. Rainfall duration is defined as the number of consecutive wet days that are days precipitation totals of 1mm or higher.
[b]DS	[d]	Mean annual number of day with snow cover (Hartmann et al., 2015)

Table C.5 continued.

	P_m	[b]PET_m	[b]T_m	[b]DTR_m	[b]RH_m	[b]WS_m	[b]$R_{down,m}$	AI	ρ_{P-PET}	P_{SI}	[a]P_{CI}	[a]P_{RE}	[a]P_{CV}	[b]PET_{SI}	[a]PET_{CI}	[a]PET_{RE}	[a]PET_{ETCV}	I_p	[a]I_{95}	[a]I_{75}	[b]H_p	[b]H_{95}	D_p	[a]D_{95}	[b]DS
P_m	**1.00**	**-0.68**	**-0.54**	**-0.69**	**0.72**	**-0.03**	**-0.64**	**-0.13**	**0.15**	**-0.62**	**-0.55**	**-0.61**	**-0.63**	**0.61**	**0.56**	**0.59**	**0.60**	**0.08**	**-0.03**	**-0.01**	**-0.18**	**-0.36**	**0.84**	**0.82**	**0.24**
[b]PET_m	**-0.68**	1.00	0.83	0.66	-0.83	0.28	0.88	**0.23**	**-0.39**	**0.77**	0.75	0.80	0.78	-0.86	-0.80	-0.82	-0.86	**0.23**	0.31	0.30	0.30	0.54	**-0.55**	-0.51	-0.48
[b]T_m	**-0.54**	0.83	1.00	0.65	-0.79	0.02	0.97	**0.15**	**-0.52**	**0.74**	0.67	0.73	0.74	-0.86	-0.86	-0.84	-0.87	**0.41**	0.43	0.43	0.21	0.40	**-0.46**	-0.42	-0.75
[b]DTR_m	**-0.69**	0.66	0.65	1.00	-0.86	-0.31	0.73	**0.12**	**-0.34**	**0.74**	0.63	0.71	0.73	-0.65	-0.65	-0.63	-0.66	**0.26**	0.32	0.31	0.17	0.36	**-0.58**	-0.56	-0.41
[b]RH_m	**0.72**	-0.83	-0.79	-0.86	1.00	0.11	-0.84	**-0.15**	**0.41**	**-0.84**	-0.74	-0.82	-0.83	0.81	0.79	0.79	0.81	**-0.38**	-0.45	-0.44	-0.20	-0.43	**0.60**	0.57	0.50
[b]WS_m	**-0.03**	0.28	0.02	-0.31	0.11	1.00	0.01	**0.13**	**-0.09**	**-0.02**	0.06	0.03	-0.00	-0.14	-0.08	-0.15	-0.13	**-0.18**	-0.14	-0.15	0.15	0.20	**0.04**	0.07	-0.03
[b]$R_{down,m}$	**-0.64**	0.88	0.97	0.73	-0.84	0.01	1.00	**0.15**	**-0.51**	**0.79**	0.70	0.77	0.78	-0.91	-0.90	-0.89	-0.91	**0.38**	0.41	0.41	0.20	0.41	**-0.53**	-0.49	-0.71
AI	**-0.13**	**0.23**	**0.15**	**0.12**	**-0.15**	**0.13**	**0.15**	**1.00**	**-0.04**	**0.19**	**0.30**	**0.24**	**0.23**	**-0.16**	**-0.13**	**-0.14**	**-0.16**	**-0.11**	**-0.06**	**-0.08**	**0.85**	**0.73**	**-0.14**	**-0.14**	**-0.06**
ρ_{P-PET}	**0.15**	**-0.39**	**-0.52**	**-0.34**	**0.41**	**-0.09**	**-0.51**	**-0.04**	**1.00**	**-0.45**	**-0.33**	**-0.41**	**-0.42**	**0.42**	**0.40**	**0.42**	**0.41**	**-0.37**	**-0.33**	**-0.34**	**-0.05**	**-0.15**	**0.03**	**0.00**	**0.50**
P_{SI}	**-0.62**	**0.77**	**0.74**	**0.74**	**-0.84**	**-0.02**	**0.79**	**0.19**	**-0.45**	**1.00**	**0.90**	**0.98**	**0.99**	**-0.74**	**-0.68**	**-0.70**	**-0.73**	**0.44**	**0.51**	**0.50**	**0.25**	**0.49**	**-0.46**	**-0.44**	**-0.40**
[a]P_{CI}	**-0.55**	0.75	0.67	0.63	-0.74	0.06	0.70	**0.30**	**-0.33**	**0.90**	1.00	0.96	0.95	-0.68	-0.61	-0.64	-0.68	**0.36**	0.46	0.44	0.37	0.62	**-0.40**	-0.38	-0.34
[a]P_{RE}	**-0.61**	0.80	0.73	0.71	-0.82	0.03	0.77	**0.24**	**-0.41**	**0.98**	0.96	1.00	0.98	-0.74	-0.67	-0.70	-0.73	**0.41**	0.50	0.49	0.31	0.56	**-0.45**	-0.43	-0.38
[a]P_{CV}	**-0.63**	0.78	0.74	0.73	-0.83	-0.00	0.78	**0.23**	**-0.42**	**0.99**	0.95	0.98	1.00	-0.74	-0.68	-0.70	-0.73	**0.42**	0.50	0.48	0.29	0.54	**-0.47**	-0.45	-0.39
[b]PET_{SI}	**0.61**	-0.86	-0.86	-0.65	0.81	-0.14	-0.91	**-0.16**	**0.42**	**-0.74**	-0.68	-0.74	-0.74	1.00	0.98	0.99	1.00	**-0.32**	-0.37	-0.37	-0.22	-0.43	**0.47**	0.43	0.66
[a]PET_{CI}	**0.56**	-0.80	-0.86	-0.65	0.79	-0.08	-0.90	**-0.13**	**0.40**	**-0.68**	-0.61	-0.67	-0.68	0.98	1.00	0.99	0.99	**-0.32**	-0.35	-0.35	-0.18	-0.37	**0.46**	0.42	0.72
[a]PET_{RE}	**0.59**	-0.82	-0.84	-0.63	0.79	-0.15	-0.89	**-0.14**	**0.42**	**-0.70**	-0.64	-0.70	-0.70	0.99	0.99	1.00	0.99	**-0.31**	-0.35	-0.35	-0.19	-0.39	**0.45**	0.42	0.67
[a]PET_{ETCV}	**0.60**	-0.86	-0.87	-0.66	0.81	-0.13	-0.91	**-0.16**	**0.41**	**-0.73**	-0.68	-0.73	-0.73	1.00	0.99	0.99	1.00	**-0.32**	-0.37	-0.37	-0.22	-0.42	**0.47**	0.43	0.67
I_p	**0.08**	**0.23**	**0.41**	**0.26**	**-0.38**	**-0.18**	**0.38**	**-0.11**	**-0.37**	**0.44**	**0.36**	**0.41**	**0.42**	**-0.32**	**-0.32**	**-0.31**	**-0.32**	**1.00**	**0.96**	**0.98**	**-0.09**	**-0.03**	**-0.01**	**-0.01**	**-0.31**
[a]I_{95}	**-0.03**	0.31	0.43	0.32	-0.45	-0.14	0.41	**-0.06**	**-0.33**	**0.51**	0.46	0.50	0.50	-0.37	-0.35	-0.35	-0.37	**0.96**	1.00	0.99	-0.04	0.07	**-0.07**	-0.07	-0.28
[a]I_{75}	**-0.01**	0.30	0.43	0.31	-0.44	-0.15	0.41	**-0.08**	**-0.34**	**0.50**	0.44	0.49	0.48	-0.37	-0.35	-0.35	-0.37	**0.98**	0.99	1.00	-0.06	0.04	**-0.05**	-0.06	-0.29
[b]H_p	**-0.18**	0.30	0.21	0.17	-0.20	0.15	0.20	**0.85**	**-0.05**	**0.25**	0.37	0.31	0.29	-0.22	-0.18	-0.19	-0.22	**-0.09**	-0.04	-0.06	1.00	0.79	**-0.16**	-0.15	-0.08
[b]H_{95}	**-0.36**	0.54	0.40	0.36	-0.43	0.20	0.41	**0.73**	**-0.15**	**0.49**	0.62	0.56	0.54	-0.43	-0.37	-0.39	-0.42	**-0.03**	0.07	0.04	0.79	1.00	**-0.28**	-0.27	-0.18
D_p	**0.84**	**-0.55**	**-0.46**	**-0.58**	**0.60**	**0.04**	**-0.53**	**-0.14**	**0.03**	**-0.46**	**-0.40**	**-0.45**	**-0.47**	**0.47**	**0.46**	**0.45**	**0.47**	**-0.01**	**-0.07**	**-0.05**	**-0.16**	**-0.28**	**1.00**	**0.98**	**0.21**
[a]D_{95}	**0.82**	-0.51	-0.42	-0.56	0.57	0.07	-0.49	**-0.14**	**0.00**	**-0.44**	-0.38	-0.43	-0.45	0.43	0.42	0.42	0.43	**-0.01**	-0.07	-0.06	-0.15	-0.27	**0.98**	1.00	0.17
[b]DS	**0.24**	-0.48	-0.75	-0.41	0.50	-0.03	-0.71	**-0.06**	**0.50**	**-0.40**	-0.34	-0.38	-0.39	0.66	0.72	0.67	0.67	**-0.31**	-0.28	-0.29	-0.08	-0.18	**0.21**	0.17	1.00

Table C.6 Pearson correlation matrix among climate descriptors.

Climate descriptors are defined in Table C.5. To facilitate the reading of the table, values of the correlation coefficient higher than 0.8 or lower than -0.8 are shaded in red, and values of the correlation coefficient between 0.5 and 0.8 or between -0.8 and -0.5 are shaded in yellow. Descriptors highlighted in bold were selected by the CART algorithm in Chapter 4. The descriptor I_{75} was used in Hartmann et al. (2017) to describe precipitation intensity.

[a] Descriptor not included in the sensitivity analyses of Chapter 4 because it shows very high correlation with other descriptors as shows in this table, and has a similar effect on recharge as shown in Fig. C.3 and C.4 (and therefore it is considered to be redundant).

[b] Descriptor included in the sensitivity analyses of Chapter 4, but that was not selected as by the CART algorithm.

C.4 Correlation analysis between simulated recharge and climate, land cover and V2Karst parameters

In this section, we report the results of the correlation analysis between all climate descriptors analysed, all model parameters and the land cover types. We note that since the ranges of the vegetation parameters depend on the land cover type, to assess the effect of the uncertainty in vegetation parameters (i.e. residual uncertainty in parameter values for a given land cover), we calculated the correlation coefficient for the vegetation parameters for the two land cover types (forest and grass/crop) separately and Figures C.3 and C4 and Tables C.7 and C.8 report the maximum values over the two land cover types.

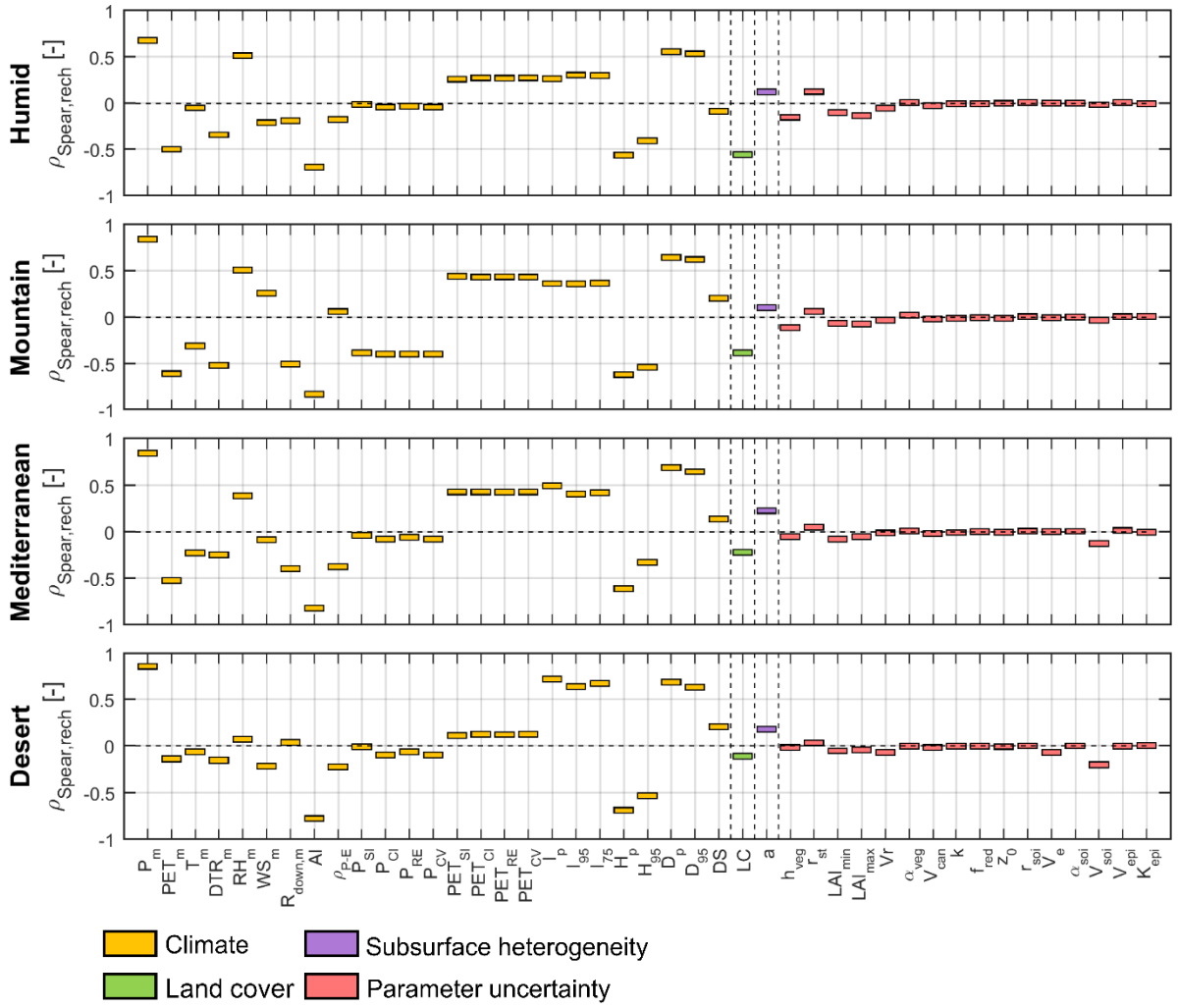


Figure C.3 Spearman correlation coefficient between the model input factors and simulated <u>mean annual recharge</u> for the four karst landscapes.

$\rho_{spear,rech}$ [−] Spearman correlation coefficient for mean annual recharge, LC land cover type (which takes a value of 0 for grass/crop and of 1 for forest to compute the correlation coefficient). Climate descriptors are defined in Table C.5. Model parameters are defined in Table 4.1. Records in bold indicate the inputs that were selected by the CART algorithm.

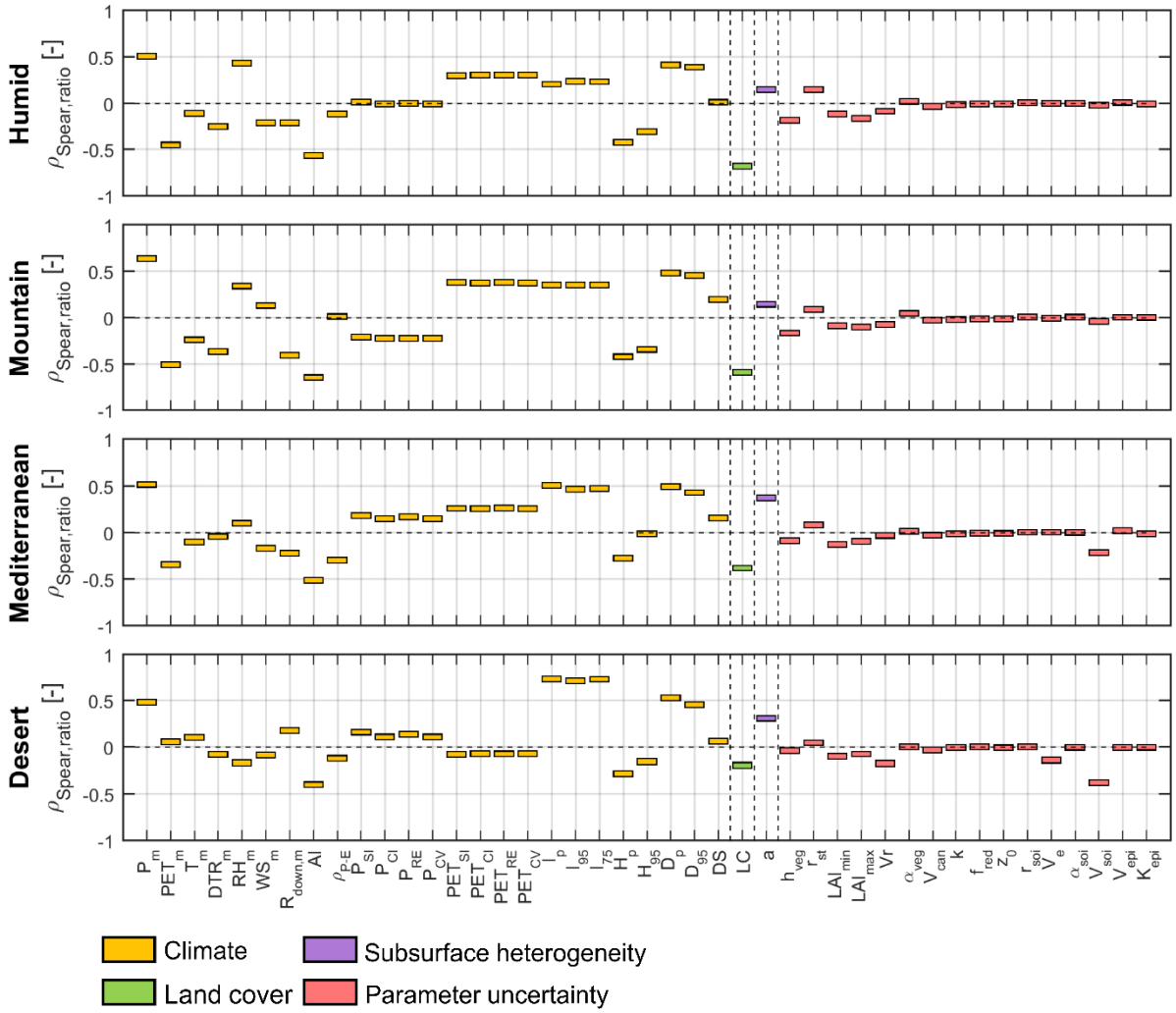


Figure C.4 Spearman correlation coefficient between the model input factors and simulated recharge ratio for the four karst landscapes.

$\rho_{spear,ratio}$ [−] Spearman correlation coefficient for recharge ratio, LC land cover type (which takes a value of 0 for grass/crop and of 1 for forest to compute the correlation coefficient). Climate descriptors are defined in Table C.5. Model parameters are defined in Table 4.1. Records in bold indicate the inputs that were selected by the CART algorithm.

	Humid		Mountain		Mediterranean		Desert	
	$\rho_{spear,rech}$	p-value	$\rho_{spear,rech}$	p-value	$\rho_{spear,rech}$	p-value	$\rho_{spear,rech}$	p-value
$\boldsymbol{P_m}$	**0.68**	**0.00**	**0.84**	**0.00**	**0.85**	**0.00**	**0.86**	**0.00**
[b]PET_m	-0.50	0.00	-0.61	0.00	-0.53	0.00	-0.14	0.00
[b]T_m	-0.05	0.00	-0.31	0.00	-0.23	0.00	-0.06	0.00
[b]DTR_m	-0.34	0.00	-0.52	0.00	-0.25	0.00	-0.15	0.00
[b]RH_m	0.51	0.00	0.51	0.00	0.39	0.00	0.07	0.00
[b]WS_m	-0.21	0.00	0.26	0.00	-0.09	0.00	-0.21	0.00
[b]$R_{down,m}$	-0.19	0.00	-0.50	0.00	-0.40	0.00	0.04	0.00
$\boldsymbol{AI}$	**-0.69**	**0.00**	**-0.83**	**0.00**	**-0.82**	**0.00**	**-0.78**	**0.00**
$\boldsymbol{\rho_{P-PET}}$	**-0.18**	**0.00**	**0.06**	**0.00**	**-0.38**	**0.00**	**-0.22**	**0.00**
$\boldsymbol{P_{SI}}$	**-0.01**	**0.00**	**-0.39**	**0.00**	**-0.04**	**0.00**	**-0.01**	**0.06**
[a]P_{CI}	-0.04	0.00	-0.40	0.00	-0.08	0.00	-0.10	0.00
[a]P_{RE}	-0.03	0.00	-0.40	0.00	-0.06	0.00	-0.06	0.00
[a]P_{CV}	-0.04	0.00	-0.40	0.00	-0.08	0.00	-0.10	0.00
[b]PET_{SI}	0.26	0.00	0.44	0.00	0.43	0.00	0.12	0.00
[a]PET_{CI}	0.27	0.00	0.43	0.00	0.43	0.00	0.13	0.00
[a]PET_{RE}	0.27	0.00	0.44	0.00	0.43	0.00	0.12	0.00
[a]PET_{ETCV}	0.27	0.00	0.43	0.00	0.43	0.00	0.13	0.00
$\boldsymbol{I_p}$	**0.27**	**0.00**	**0.36**	**0.00**	**0.49**	**0.00**	**0.72**	**0.00**
[a]I_{95}	0.30	0.00	0.36	0.00	0.41	0.00	0.64	0.00
[a]I_{75}	0.30	0.00	0.37	0.00	0.42	0.00	0.68	0.00
[b]H_p	-0.56	0.00	-0.62	0.00	-0.62	0.00	-0.69	0.00
[b]H_{95}	-0.40	0.00	-0.54	0.00	-0.33	0.00	-0.53	0.00
[b]D_p	0.56	0.00	0.64	0.00	0.69	0.00	0.69	0.00
[a]D_{95}	0.53	0.00	0.62	0.00	0.65	0.00	0.64	0.00
[b]DS	-0.09	0.00	0.21	0.00	0.14	0.00	0.21	0.00
LC	**-0.55**	**0.00**	**-0.39**	**0.00**	**-0.22**	**0.00**	**-0.11**	**0.00**
$\boldsymbol{a}$	**0.12**	**0.00**	**0.10**	**0.00**	**0.23**	**0.00**	**0.18**	**0.00**
h_{veg}	-0.15	0.00	-0.11	0.00	-0.05	0.00	-0.02	0.00
r_{st}	0.13	0.00	0.06	0.00	0.05	0.00	0.04	0.00
$\boldsymbol{LAI_{min}}$	**-0.10**	**0.00**	**-0.06**	**0.00**	**-0.08**	**0.00**	**-0.05**	**0.00**
LAI_{max}	-0.14	0.00	-0.07	0.00	-0.06	0.00	-0.04	0.00
V_r	-0.06	0.00	-0.03	0.00	-0.02	0.00	-0.07	0.00
α_{veg}	0.01	0.06	0.02	0.00	0.01	0.27	-0.00	0.79
V_{can}	-0.03	0.00	-0.02	0.00	-0.02	0.00	-0.02	0.00
k	-0.01	0.02	-0.01	0.00	-0.01	0.01	-0.00	0.34
f_{red}	-0.01	0.15	-0.01	0.13	-0.00	0.80	0.00	0.78
z_0	-0.00	0.60	-0.01	0.00	- 0.00	0.25	-0.01	0.13
$r_{s,soi}$	0.01	0.16	0.01	0.09	0.00	0.25	0.00	0.59
V_e	0.00	0.87	-0.00	0.56	0.00	0.44	- 0.07	0.00
α_{soi}	-0.00	0.64	0.00	0.45	0.00	0.29	0.00	0.62
$\boldsymbol{V_{soil}}$	**-0.02**	**0.00**	**-0.03**	**0.00**	**-0.13**	**0.00**	**-0.20**	**0.00**
V_{epi}	0.01	0.01	0.01	0.09	0.01	0.00	-0.00	0.61
K_{epi}	-0.01	0.08	0.01	0.05	-0.01	0.09	0.00	0.35

Table C.7 Spearman correlation coefficient and p-value between the model input factors and simulated <u>mean annual recharge</u> for the four karst landscapes.

$\rho_{spear,rech}$ [−] Spearman correlation coefficient for mean annual recharge, LC land cover type (forest or grass/crop). Climate descriptors are defined in Table C.5. Model parameters are defined in Table 4.1. Records in bold indicate the inputs that were selected by the CART algorithm. Records are shaded in grey when the p-value is higher than 0.05 (5% significance level).

[a] Climate descriptor not included in the sensitivity analyses of Chapter 4 (CART) because a correlation analyses showed that it is redundant (Table C.5). [b] Climate descriptor included in the sensitivity analyses of Chapter 4, but that was not selected as by the CART algorithm for mean annual recharge.

	Humid		Mountain		Mediterranean		Desert	
	$\rho_{spear,ratio}$	p-value	$\rho_{spear,ratio}$	p-value	$\rho_{spear,ratio}$	p-value	$\rho_{spear,ratio}$	p-value
[b]P_m	0.51	0.00	0.64	0.00	0.52	0.00	0.49	0.00
[b]PET_m	-0.45	0.00	-0.50	0.00	-0.34	0.00	0.06	0.00
[b]T_m	-0.11	0.00	-0.24	0.00	-0.10	0.00	0.11	0.00
[b]DTR_m	-0.25	0.00	-0.36	0.00	-0.04	0.00	-0.07	0.00
[b]RH_m	0.44	0.00	0.34	0.00	0.10	0.00	-0.16	0.00
[b]WS_m	-0.21	0.00	0.13	0.00	-0.17	0.00	-0.08	0.00
[b]$R_{down,m}$	-0.21	0.00	-0.41	0.00	-0.22	0.00	0.18	0.00
AI	**-0.56**	**0.00**	**-0.64**	**0.00**	**-0.51**	**0.00**	**-0.40**	**0.00**
[b]ρ_{P-PET}	-0.12	0.00	0.01	0.00	-0.30	0.00	-0.12	0.00
[b]P_{SI}	0.01	0.00	-0.21	0.00	0.18	0.00	0.16	0.00
[a]P_{CI}	-0.01	0.03	-0.22	0.00	0.15	0.00	0.11	0.00
[a]P_{RE}	-0.00	0.76	-0.22	0.00	0.17	0.00	0.14	0.00
[a]P_{CV}	-0.01	0.03	-0.22	0.00	0.15	0.00	0.11	0.00
[b]PET_{SI}	0.30	0.00	0.38	0.00	0.26	0.00	-0.08	0.00
[a]PET_{CI}	0.31	0.00	0.37	0.00	0.26	0.00	-0.07	0.00
[a]PET_{RE}	0.30	0.00	0.38	0.00	0.26	0.00	-0.07	0.00
[a]PET_{ETCV}	0.31	0.00	0.37	0.00	0.26	0.00	-0.07	0.00
I_p	**0.21**	**0.00**	**0.35**	**0.00**	**0.51**	**0.00**	**0.74**	**0.00**
[a]I_{95}	0.24	0.00	0.35	0.00	0.47	0.00	0.72	0.00
[a]I_{75}	0.23	0.00	0.36	0.00	0.47	0.00	0.73	0.00
[b]H_p	-0.42	0.00	-0.42	0.00	-0.27	0.00	-0.28	0.00
[b]H_{95}	-0.30	0.00	-0.34	0.00	-0.01	0.00	-0.15	0.00
D_p	**0.41**	**0.00**	**0.48**	**0.00**	**0.49**	**0.00**	**0.53**	**0.00**
[a]D_{95}	0.39	0.00	0.45	0.00	0.43	0.00	0.46	0.00
[b]DS	0.01	0.00	0.20	0.00	0.16	0.00	0.07	0.00
LC	**-0.68**	**0.00**	**-0.59**	**0.00**	**-0.38**	**0.00**	**-0.19**	**0.00**
α	**0.15**	**0.00**	**0.15**	**0.00**	**0.37**	**0.00**	**0.31**	**0.00**
h_{veg}	-0.18	0.00	-0.16	0.00	-0.09	0.00	-0.03	0.00
r_{st}	0.15	0.00	0.09	0.00	0.08	0.00	0.05	0.00
LAI_{min}	-0.12	0.00	-0.08	0.00	-0.13	0.00	-0.10	0.00
LAI_{max}	-0.16	0.00	-0.10	0.00	-0.09	0.00	-0.07	0.00
V_r	-0.08	0.00	-0.07	0.00	-0.03	0.00	-0.17	0.00
α_{veg}	0.02	0.00	0.05	0.00	0.01	0.01	0.00	0.56
V_{can}	-0.03	0.00	-0.02	0.00	-0.03	0.00	-0.03	0.00
k	-0.01	0.00	-0.02	0.00	-0.02	0.00	-0.00	0.30
f_{red}	-0.01	0.09	-0.01	0.01	-0.01	0.04	0.00	0.40
z_0	-0.01	0.13	-0.01	0.00	-0.01	0.01	-0.00	0.25
$r_{s,soi}$	0.01	0.02	0.01	0.02	0.00	0.28	0.00	0.40
V_e	0.00	0.71	-0.00	0.34	0.00	0.29	-0.14	0.00
α_{soi}	0.00	0.75	0.01	0.08	0.00	0.31	0.00	0.70
V_{soil}	**-0.02**	**0.00**	**-0.04**	**0.00**	**-0.22**	**0.00**	**-0.37**	**0.00**
V_{epi}	0.01	0.00	0.01	0.10	0.02	0.00	-0.00	0.57
K_{epi}	-0.01	0.03	0.01	0.15	-0.01	0.00	0.00	0.63

Table C.8 Spearman correlation coefficient and p-value between the model input factors and simulated recharge ratio for the four karst landscapes.

$\rho_{spear,ratio}$ $[-]$ Spearman correlation coefficient for recharge ratio, LC land cover type (forest or grass/crop). Climate descriptors are defined in Table C.5. Model parameters are defined in Table 4.1. Records in bold indicate the inputs that were selected by the CART algorithm. Records are shaded in grey when the p-value is higher than 0.05 (5% significance level).

[a] Climate descriptor not included in the sensitivity analyses of Chapter 4 (CART) because a correlation analyses showed that it is redundant (Table C.5). [b] Climate descriptor included in the sensitivity analyses of Chapter 4, but that was not selected as by the CART algorithm for recharge ratio.

C.5 Comparison between present recharge simulated with V2Karst and the WHYMAP

Estimate of recharge compared with the WHYMAP	Percentage of cells with recharge classes lower than WHYMAP	Percentage of cells with recharge classes higher than WHYMAP	Percentage of cells with recharge classes in agreement with WHYMAP
Mean of V2Karst ensemble	**Domain: 0.5%** Humid: 1.1% Mountain: 0.2% Mediterranean: 0% Desert: 0%	**Domain; 69.4%** Humid: 65.5% Mountain: 76.7% Mediterranean: 79.4% Desert: 56.3%	**Domain: 30.1%** Humid: 33.4% Mountain: 23.1% Mediterranean: 20.6% Desert: 43.7%
Lower bound of V2Karst ensemble	**Domain: 3.0%** Humid: 5.0% Mountain: 1.1% Mediterranean: 2.2% Desert: 0.6%	**Domain: 43.7%** Humid: 46.5% Mountain: 66.5% Mediterranean: 47.2% Desert: 8.6%	**Domain: 53.3%** Humid: 48.5% Mountain: 32.4% Mediterranean: 50.6% Desert: 90.8%
Upper bound of V2Karst ensemble	**Domain: 0.1%** Humid: 0.2% Mountain: 0.2% Mediterranean: 0% Desert: 0%	**Domain: 80.5%** Humid: 78.3% Mountain: 78.8% Mediterranean: 91.1% Desert: 69.2%	**Domain: 19.4%** Humid: 21.5% Mountain: 21.0% Mediterranean: 8.9% Desert: 30.8%

Table C.9 Comparison between present mean annual recharge simulated by V2Karst (1964 - 2004) and the WHYMAP (2069 – 2099) over carbonate rock areas in Europe, the Middle East and Northern Africa.

For each 0.5x0.5° cell, three values of recharge simulated by V2Karst for the present period are considered, i.e. the mean, the lower bound of the 95% confidence interval and the upper bound of the 95% confidence interval. Values are reported for the entire domain and the four karst landscapes (humid, mountain, Mediterranean and desert). The agreement between V2Karst simulations and the WHYMAP was assessed in terms of recharge classes. Four classes of recharge were considered for most cells, apart for areas indicated as shallow and local aquifers in the WHYMAP, for which two classes of recharge were considered as in the WHYMAP (classes are reported in Fig. C.5).

Source of the data: The WHYMAP comes from (WHYMAP GWR © BGR & UNESCO 2015, BGR and UNESCO 2008) and was gridded in this study (0.5x0.5° grid) to allow comparison with recharge simulated with V2Karst (map is provided in Fig, C.5.d).

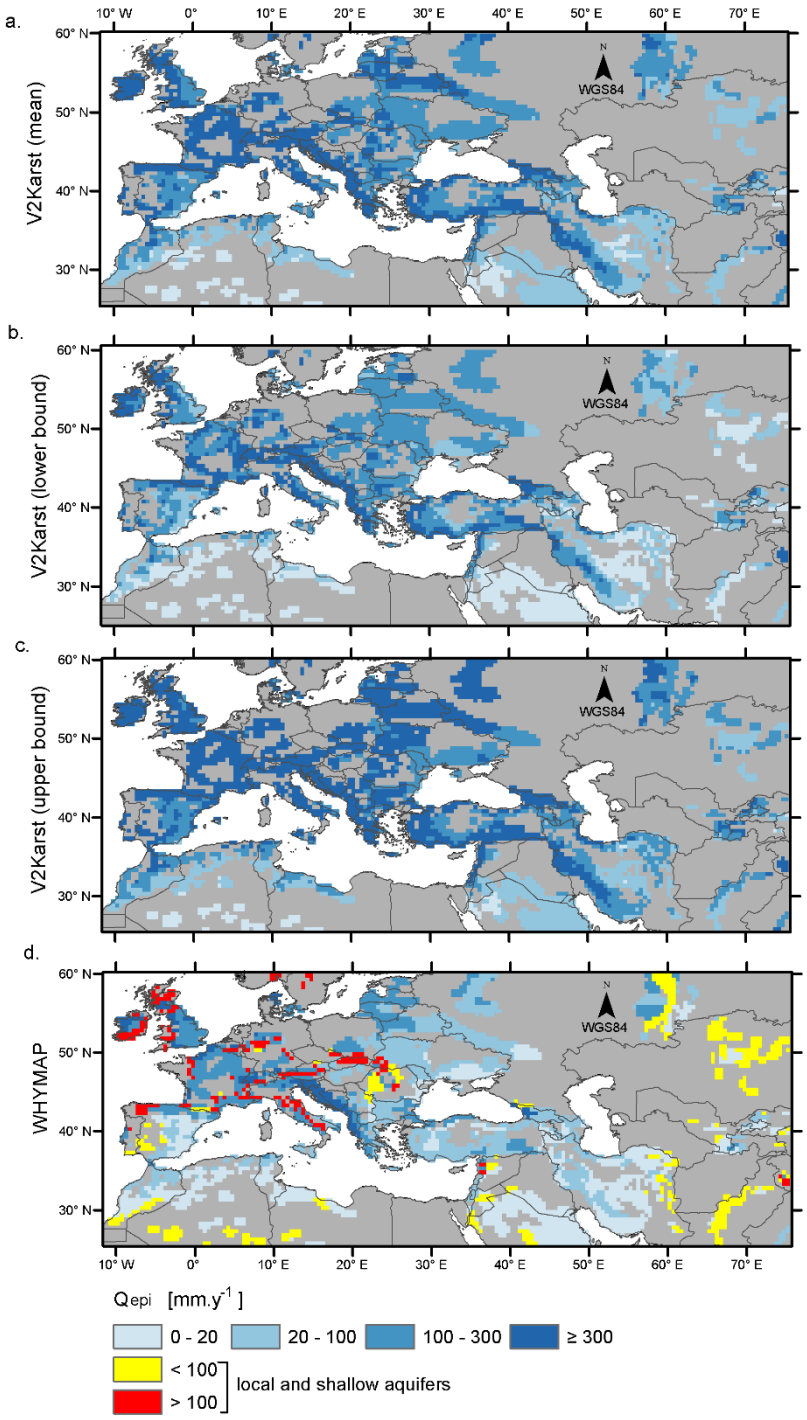


Figure C.5 Historical mean annual recharge simulated by V2Karst (1964 – 2004) and WHYMAP (1961-1990) in carbonate rock areas in Europe, the Middle East and Northern Africa.

The figure reports (a) the mean, (b) the lower bound of the 95% confidence interval and (c) the upper of the 95% confidence interval of the V2karst simulation ensemble, and (d) the values reported in the WHYMAP. The WHYMAP identifies only two recharge classes for local and shallow and aquifers (< 100 or >100 $mm.y^{-1}$). Q_{epi} $[mm.y^{-1}]$ mean annual recharge. Source of the data: Carbonate rock and country map from Williams and Ford (2006) and in particular the country map was obtained from Terraspace, Russian space agency. For (c): mean annual recharge from the WHYMAP (WHYMAP GWR © BGR & UNESCO 2015, BGR and UNESCO 2008), that was gridded in this study (0.5x0.5° grid) to allow comparison with recharge simulated with V2Karst.

C.6 Uncertainty bounds of the projected change in recharge

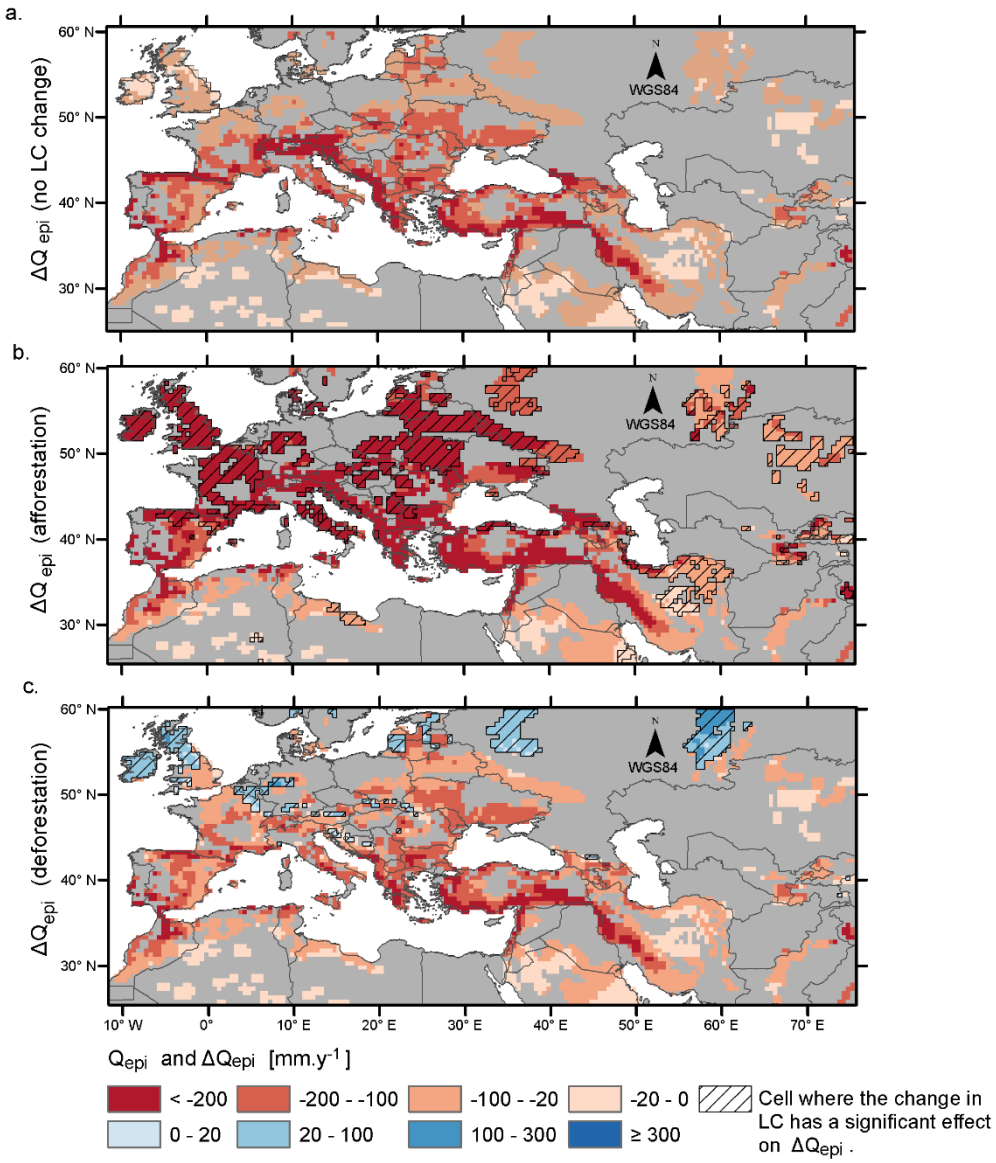


Figure C.6 Lower bound of the grid cell sensitivity of recharge to combined climate change and land cover change for three different land cover change scenarios from the period 1974-2004 to 2069-2099 in carbonate rock areas in Europe, the Middle East and Northern Africa.

(a) Projected change in mean annual recharge $\Delta Q_{epi} = \mathrm{Q}_{\mathrm{epi}}^{future} - \mathrm{Q}_{\mathrm{epi}}^{historical}$ $[mm.y^{-1}]$ assuming no change in land cover (LC), (b) assuming complete afforestation and (c) assuming complete deforestation. The figure reports for each cell the lower bound of the 95% confidence interval of the simulation ensemble ('pessimistic case'). Hatched cells show a significant difference in ΔQ_{epi} between the 'no land cover change' scenario and the relevant land cover change scenario (afforestation or deforestation) as discussed in Sect. 4.3.3.

Source of the data: Carbonate rock and country map from Williams and Ford (2006) and in particular the country map was obtained from Terraspace, Russian space agency.

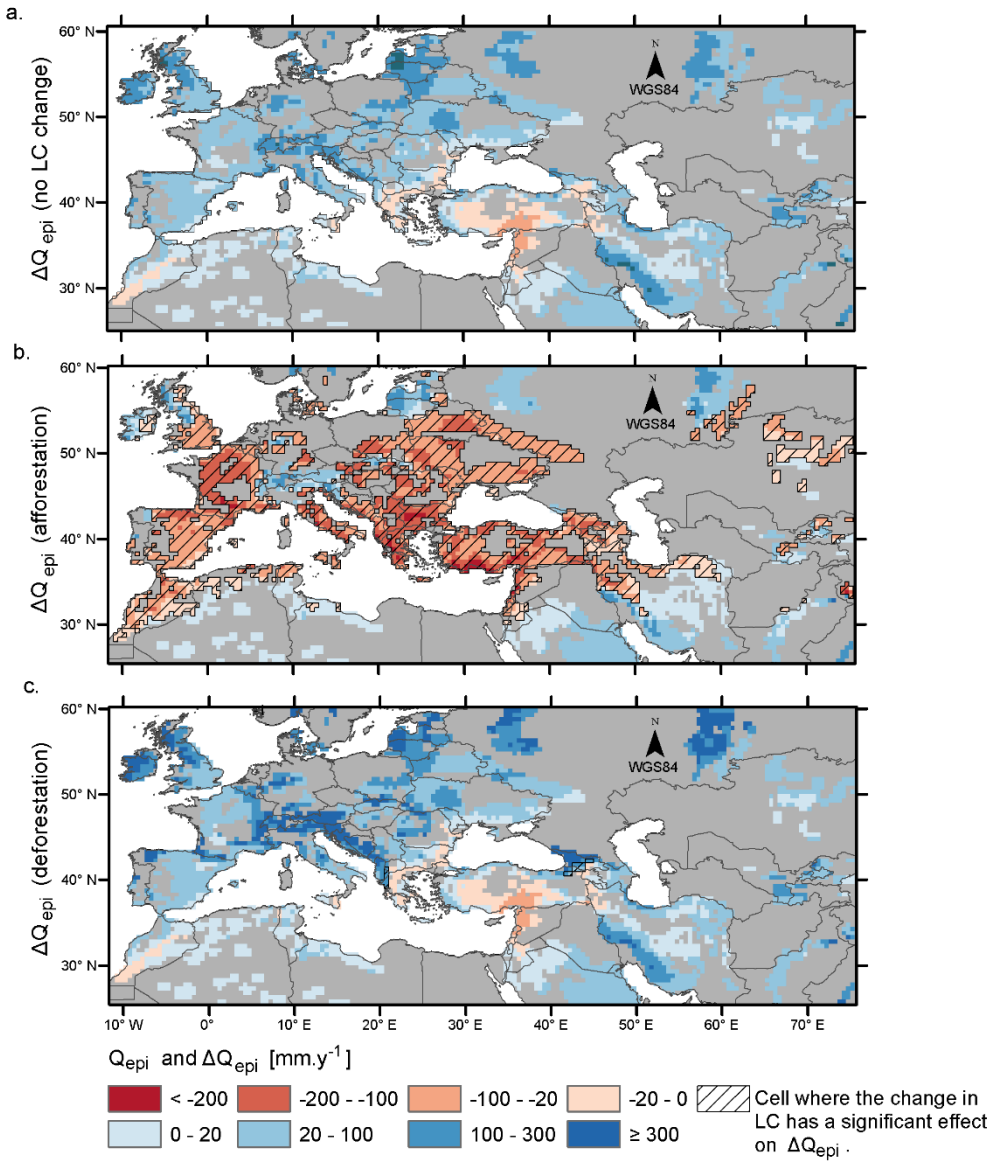


Figure C.7 Upper bound of the grid cell sensitivity of recharge to combined climate change and land cover change for three different land cover change scenarios from the period 1974-2004 to 2069-2099 in carbonate rock areas in Europe, the Middle East and Northern Africa.

(a) Projected change in mean annual recharge $\Delta Q_{epi} = \mathrm{Q}_{\mathrm{epi}}^{future} - \mathrm{Q}_{\mathrm{epi}}^{historical}$ $[mm.y^{-1}]$ assuming no change in land cover (LC), (b) assuming complete afforestation and (c) assuming complete deforestation. The figure reports for each cell the upper bound of the 95% confidence interval of the simulation ensemble ('optimistic case'). Hatched cells show a significant difference in ΔQ_{epi} between the 'no land cover change' scenario and the relevant land cover change scenario (afforestation or deforestation) as discussed in Sect. 4.3.3.

Source of the data: Carbonate rock and country map from Williams and Ford (2006) and in particular the country map was obtained from Terraspace, Russian space agency.

Appendix D Curriculum Vitae

Fanny Sarrazin

(fanny.sarrazin@bristol.ac.uk)

EDUCATION

04/2014 - 06/2018	**PhD** in Water and Environmental Engineering University of Bristol, United Kingdom *Thesis: Understanding the sensitivity of karst groundwater recharge to climate and land cover changes at a large-scale* *Advisors: Prof. Thorsten Wagener, Dr Francesca Pianosi*
09/2011 - 12/2013	**MSc** in Environmental and Land Planning Engineering (Laurea Magistrale) Politecnico di Milano, Italy (Double degree programme with Ecole Centrale de Lyon) Programme in Italian Graduated summa cum laude (110/110) *Master Thesis: Land use changes in river floodplains - Analysis of Rivers Po and Oglio*
09/2009 - 12/2013	**MSc** in General Engineering **(**Diplôme d'ingénieur généraliste) Ecole Centrale de Lyon, France *4-month internship: Contribution to the development of an action plan to improve the yield of a drinking water distribution system. Rhône Saône centre of Lyonnaise des Eaux, Caluire, France*
09/2007 - 06/2009	2-year programme to prepare for national competitive examinations to enter a « Grande Ecole » of Engineering (Classe préparatoire MPSI-MP*) Lycée Champollion, Grenoble, France *Main subjects: Mathematics and Physics*

AWARDS

03/2014	4-year University of Bristol Alumni Postgraduate Scholarship
05/2011	Recognition for being in the top 12% of the cohort of 304 students at Ecole Centrale de Lyon (Encouragements)

SERVICE

05/2016	Organiser of the British Hydrological Society Peter Wolf Symposium 2016 - Early Career Hydrologists' Event in Bristol, United Kingdom *2-day Symposium with 36 oral/poster presentations from delegates and 2 keynotes*
02/2016 - present	Reviewer for: Hydrology and Earth System Sciences, Environmental Modelling and Software, Journal of Hydrology, Hydrology Research, Journal of Hydrologic Engineering, Cogent Engineering, Ecological Modelling (certificate of outstanding contribution in reviewing manuscripts for Ecological Modelling awarded in October 2017)

TEACHING

09/2014 - 06/2018	Teaching Assistant for Postgraduate modules at University of Bristol, United Kingdom: support to Matlab laboratory and marking, MSc in Water and Environmental Management.
07/2016	Teaching Assistant for sensitivity analysis workshop, Cabot Institute Summer School on risk and uncertainty in natural hazards in Bristol, United Kingdom
07/2016 - 08/2016	Co-advisor for 2-month internship of Dimitri Rambourg (MSc) at University of Bristol, United Kingdom: Analysis of large-scale databases for a better understanding of evapotranspiration and groundwater recharge in karst areas.

SKILLS

Languages	French (Mother tongue), English (Fluent), Italian (Fluent)
Software	Matlab - Arcgis - Quantum Gis

SOFTWARE DEVELOPMENT

SAFE, Matlab toolbox for global sensitivity analysis

(https://www.safetoolbox.info/)

Co-developer of the SAFE toolbox, which implements several global sensitivity analysis methods, visualisation tools and functions to assess the robustness and convergence of the results. SAFE is designed for both non-specialist users and users with advanced skills in sensitivity analysis. The software is open source and has been freely available for academic and non-commercial purpose since 2015. SAFE has been adopted by over 1300 academic users in over 50 countries for both research and teaching purposes.

V2Karst, integrated vegetation-recharge model for karst areas

(https://zenodo.org/record/1205463)

Co-developer of the V2karst model, which is the first large-scale integrated vegetation-recharge model for karst areas. The model is implemented in Matlab and builds on its previous version called VarKarst (Harmann et al., 2015, Geoscientific Model Development). V2Karst is parsimonious and was designed to assess the impact of future changes in climate and land cover on groundwater recharge in karst regions.

PUBLICATIONS

Journal (peer reviewed):

Sarrazin F, Hartmann A, Pianosi F, Rosolem R, Wagener, T (in preparation) How do land cover and subsurface heterogeneity modulate climate change impacts on future groundwater recharge in karst areas?.

Sarrazin F, Hartmann A, Pianosi F, Wagener T (in review) V2Karst v1.0: A parsimonious large-scale integrated vegetation-recharge model to understand the impact of climate and land cover change in karst regions. Geosci Model Dev Discuss. doi: 10.5194/gmd-2017-315

Khorashadi Zadeh F, 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 Softw 91:210–222. doi: 10.1016/j.envsoft.2017.02.001

Sarrazin F, Pianosi F, Wagener T (2016) Global Sensitivity Analysis of environmental models: Convergence and validation. Environ Model Softw 79:135–152. doi: 10.1016/j.envsoft.2016.02.005

Pianosi F, **Sarrazin F**, Wagener T (2015) A Matlab toolbox for Global Sensitivity Analysis. Environ Model Softw 70:80–85. doi: 10.1016/j.envsoft.2015.04.009

Book chapter:

Sarrazin F, Pianosi F, Wagener T (2017) An introduction to the SAFE Matlab Toolbox with practical examples and guidelines. In: Petropoulos G, Srivastava P (eds) Sensitivity Analysis in Earth Observation Modelling. Elsevier Inc., pp 363–378.

CONFERENCE PRESENTATIONS

Oral presentations:

A large-scale integrated karst-vegetation recharge model to understand the impact of climate and land cover change, European Geosciences Union General Assembly 2017, Vienna, Austria.

What is the relative impact of climate and land use changes on groundwater recharge given subsurface heterogeneity?, Eurokarst 2016, Neuchâtel, Switzerland.

Global Sensitivity Analysis of environmental models: Robustness, convergence and validation British Hydrological Society Peter Wolf Symposium - Early Career Hydrologists' Event 2015, Liverpool, United Kingdom.

Poster presentations:

How does future groundwater recharge depend on climate, land cover and subsurface heterogeneities in the European and Mediterranean region?, European Geosciences Union General Assembly 2018, Vienna, Austria.

To what extent do preferential flow pathways alter the impact of climate and land cover change on groundwater recharge?, American Geophysical Union Fall meeting 2016, San Francisco, USA.

How much does subsurface heterogeneity alter the impact of climate and land use changes on groundwater recharge?, European Geosciences Union General Assembly 2016, Vienna, Austria.

Does the relative impact of climate and land use changes on groundwater recharge vary depending on the degree of subsurface heterogeneity?, American Geophysical Union Fall meeting 2015, San Francisco, USA.

Global Sensitivity Analysis of Environmental Models: Convergence, Robustness and Validation, European Geosciences Union General Assembly 2015, Vienna, Austria.

Global Sensitivity Analysis of Environmental Models: Convergence, Robustness and Accuracy Analysis, American Geophysical Union Fall meeting 2014, San Francisco, USA.

Bibliography

Abramowitz, G., Leuning, R., Clark, M. and Pitman, A.: Evaluating the performance of land surface Models, J. Clim., 21(21), 5468–5481, doi:10.1175/2008JCLI2378.1, 2008.

Aeschbach-Hertig, W. and Gleeson, T.: Regional strategies for the accelerating global problem of groundwater depletion, Nat. Geosci., 5(12), 853–861, doi:10.1038/ngeo1617, 2012.

Alcalá, F. J., Cantón, Y., Contreras, S., Were, A., Serrano-Ortiz, P., Puigdefábregas, J., Solé-Benet, A., Custodio, E. and Domingo, F.: Diffuse and concentrated recharge evaluation using physical and tracer techniques: Alcalá, Francisco J., Yolanda Cantón, Sergio Contreras, Ana Were, Penélope Serrano-Ortiz, Juan Puigdefábregas, Albert Solé-Benet, Emilio Custodio, and Francisco Domingo. 20, Environ. Earth Sci., 62(3), 541–557, doi:10.1007/s12665-010-0546-y, 2011.

Ali, Y. S. A., Crosato, A., Mohamed, Y. A., Abdalla, S. H. and Wright, N. G.: Sediment balances in the Blue Nile River Basin, Int. J. Sediment Res., 29(3), 316–328, doi:10.1016/S1001-6279(14)60047-0, 2014.

Allen, R. G., Pereira, L. S., Raes, D. and Smith, M.: Crop evapotranspiration: Guidelines for computing crop requirements, FAO Irrigation and Drainage Paper 56, Food and Agriculture Organization (FAO), Rome, Italy, 1998.

Almeida, S., Ann Holcombe, E., Pianosi, F. and Wagener, T.: Dealing with deep uncertainties in landslide modelling for disaster risk reduction under climate change, Nat. Hazards Earth Syst. Sci., 17(2), 225–241, doi:10.5194/nhess-17-225-2017, 2017.

Anderson, B., Borgonovo, E., Galeotti, M. and Roson, R.: Uncertainty in climate change modeling: Can global sensitivity analysis be of help?, Risk Anal., 34(2), 271–293, doi:10.1111/risa.12117, 2014.

Andres, T. H.: Sampling methods and sensitivity analysis for large parameter sets, J. Stat. Comput. Simul., 57(1–4), 77–110, doi:10.1080/00949659708811804, 1997.

Arbel, Y., Greenbaum, N., Lange, J. and Inbar, M.: Infiltration processes and flow rates in developed karst vadose zone using tracers in cave drips, Earth Surf. Process. Landforms, 35(14), 1682–1693, doi:10.1002/esp.2010, 2010.

Archer, G. E. B., Saltelli, A. and Sobol, I. M.: Sensitivity measures, ANOVA-like techniques and the use of boostrap, J. Stat. Comput. Simul., 58(2), 99–120, doi:10.1080/00949659708811825, 1997.

Archfield, S. A., Clark, M., Arheimer, B., Hay, L. E., McMillan, H., Kiang, J. E., Seibert, J., Hakala, K., Bock, A., Wagener, T., Farmer, W. H., Andréassian, V., Attinger, S., Viglione, A., Knight, R., Markstrom, S. and Over, T.: Accelerating advances in continental domain hydrologic modeling, Water Resour. Res., 51(12), 10078–10091, doi:10.1002/2015WR017498, 2015.

Arnell, N. W.: A simple water balance model for the simulation of streamflow over a large geographic domain, J. Hydrol., 217(3–4), 314–335, doi:10.1016/S0022-1694(99)00023-2, 1999.

Arnold, J. G., Allen, P. M. and Bernhardt, G.: A comprehensive surface-groundwater flow model, J. Hydrol., 142(1–4), 47–69, doi:10.1016/0022-1694(93)90004-S, 1993.

Arnold, J. G., Srinivasan, R., Muttiah, R. S. and Williams, J. R.: Large area hydrologic modeling and assesment Part I: Model development, JAWRA J. Am. Water Resour. Assoc., 34(1), 73–89, doi:10.1111/j.1752-1688.1998.tb05961.x, 1998.

Atkinson, S. E., Woods, R. A. and Sivapalan, M.: Climate and landscape controls on water balance model complexity over changing timescales, Water Resour. Res., 38(12), 50-1-50–17, doi:10.1029/2002WR001487, 2002.

Bae, D. H., Jung, I. W. and Lettenmaier, D. P.: Hydrologic uncertainties in climate change from IPCC AR4 GCM simulations of the Chungju Basin, Korea, J. Hydrol., 401(1–2), 90–105, doi:10.1016/j.jhydrol.2011.02.012, 2011.

Bai, P., Liu, X., Liang, K. and Liu, C.: Comparison of performance of twelve monthly water balance models in different climatic catchments of China, J. Hydrol., 529, 1030–1040, doi:10.1016/j.jhydrol.2015.09.015, 2015.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw, K. T., Pilegaard, K., Schmid, H. P., Valentini, R., Verma, S., Vesala, T., Wilson, K. and Wofsy, S.: FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities, Bull. Am. Meteorol. Soc., 82(11), 2415–2434, doi:10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2, 2001.

Baldocchi, D. D. and Ryu, Y.: A Synthesis of Forest Evaporation Fluxes – from Days to Years – as Measured with Eddy Covariance, in Forest Hydrology and Biogeochemistry, Synthesis of Past Research and Future Directions, edited by D. F. Levia, D. Carlyle-Moses, and T. Tanaka, pp. 101–116, Springer Netherlands, Dordrecht., 2011.

Bannwarth, M. A., Sangchan, W., Hugenschmidt, C., Lamers, M., Ingwersen, J., Ziegler, A. D. and Streck, T.: Pesticide transport simulation in a tropical catchment by SWAT, Environ. Pollut., 191, 70–79, doi:10.1016/j.envpol.2014.04.011, 2014.

Bargués Tobella, A., Reese, H., Almaw, A., Bayala, J., Malmer, A., Laudon, H. and Ilstedt, U.: The effect of trees on preferential flow and soil infiltrability in an agroforestry parkland in semiarid Burkina Faso, Water Resour. Res., 50(4), 3342–3354, doi:10.1002/2013WR015197, 2014.

Barnhart, T. B., Molotch, N. P., Livneh, B., Harpold, A. A., Knowles, J. F. and Schneider, D.: Snowmelt rate dictates streamflow, Geophys. Res. Lett., 43(15), 8006–8016, doi:10.1002/2016GL069690, 2016.

Baroni, G. and Tarantola, S.: A General Probabilistic Framework for uncertainty and global sensitivity analysis of deterministic models: A hydrological case study, Environ. Model. Softw., 51, 26–34, doi:10.1016/j.envsoft.2013.09.022, 2014.

Bartolino, J. R. and Cunningham, W. L.: Ground-Water Depletion Across the Nation, U.S Geol. Surv. Fact Sheet, 103–3 [online] Available from: https://pubs.usgs.gov/fs/fs-103-03/#pdf, 2003.

Beaulieu, E., Goddéris, Y., Donnadieu, Y., Labat, D. and Roelandt, C.: High sensitivity of the continental-weathering carbon dioxide sink to future climate change, Nat. Clim. Chang., 2(5), 346–349, doi:10.1038/nclimate1419, 2012.

Beck, H. E., de Roo, A. and van Dijk, A. I. J. M.: Global Maps of Streamflow Characteristics Based on Observations from Several Thousand Catchments, J. Hydrometeorol., 16(4), 1478–1501, doi:10.1175/JHM-D-14-0155.1, 2015.

Berghuijs, W. R., Woods, R. A. and Hrachowitz, M.: A precipitation shift from snow towards rain leads to a decrease in streamflow, Nat. Clim. Chang., 4(7), 583–586, doi:10.1038/nclimate2246, 2014.

Bergström, S.: The HBV model (Chapter 13), in Computer Models of Watershed Hydrology, edited by V. P. Singh, pp. 443–476, Water Resources Publications, Highlands Ranch, Colorado, USA., 1995.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. . L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B. and Harding, R. J.: The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes, Geosci. Model Dev., 4(3), 677–699, doi:10.5194/gmd-4-677-2011, 2011.

Beven, K.: How far can we go in distributed hydrological modelling?, Hydrol. Earth Syst. Sci., 5(1), 1–12, doi:10.5194/hess-5-1-2001, 2001.

Beven, K. J. and Cloke, H. L.: Comment on "hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water" by Eric F. Wood et al., Water Resour. Res., 48(1), W01801, doi:10.1029/2010WR010090, 2012.

Beven, K. and Germann, P.: Macropores and water flow in soils revisited, Water Resour. Res., 49(6), 3071–3092, doi:10.1002/wrcr.20156, 2013.

Bierkens, M. F. P.: Global hydrology 2015: State, trends, and directions, Water Resour. Res., 51(7), 4923–4947, doi:10.1002/2015WR017173, 2016.

Blume, H.-P., Brümmer, G. W., Horn, R., Kandeler, E., Kögel-Knabner, I., Kretzschmar, R., Stahr, K. and Wilke, B.-M.: Lehrbuch der Bodenkunde, Springer-Verlag, Berlin Heidelberg, doi: 10.1007/978-3-662-49960-3., 2010.

Bohn, T. J. and Vivoni, E. R.: Process-based characterization of evapotranspiration sources over the North American monsoon region, Water Resour. Res., 52(1), 358–384, doi:10.1002/2015WR017934, 2016.

Bonacci, O., Pipan, Ta. and Culver, D. C.: A framework for karst ecohydrology, Environ. Geol., 56(5), 891–900, doi:0.1007/s00254-008-1189-0, 2009.

Bonan, G.: Soil Physics, in Ecological Climatology: Concepts and Applications, pp. 143–154, Cambridge University Press, Cambridge., 2015a.

Bonan, G.: Surface energy fluxes, in Ecological climatology: Concepts and Applications, pp. 193–208, Cambridge University Press, Cambridge, United Kingdom., 2015b.

Boone, A., Calvet, J.-C. and Noilhan, J.: Inclusion of a third soil layer in a land surface scheme using the force–restore method, J. Appl. Meteorol., 38(11), 1611–1630, doi:10.1175/1520-0450(1999)038<1611:IOATSL>2.0.CO;2, 1999.

Borgonovo, E., Lu, X., Plischke, E., Rakovec, O. and Hill, M. C.: Making the most out of a hydrological model data set: Sensitivity analyses to open the model black-box, Water Resour. Res., 53(9), 7933–7950, doi:10.1002/2017WR020767, 2017.

Botter, G., Porporato, A., Rodriguez-Iturbe, I. and Rinaldo, A.: Nonlinear storage-discharge relations and catchment streamflow regimes, Water Resour. Res., 45(10), W10427, doi:10.1029/2008WR007658, 2009.

Bouraoui, F. and Grizzetti, B.: An integrated modelling framework to estimate the fate of nutrients: Application to the Loire (France), Ecol. Modell., 212(3–4), 450–459, doi:10.1016/j.ecolmodel.2007.10.037, 2008.

Boussetta, S., Balsamo, G., Beljaars, A., Kral, T. and Jarlan, L.: Impact of a satellite-derived leaf area index monthly climatology in a global numerical weather prediction model, Int. J. Remote Sens., 34(9–10), 3520–3542, doi:10.1080/01431161.2012.716543, 2013.

Boyle, D.: Multicriteria calibration of hydrological models, PhD thesis, University of Arizona, Tucson, 2001.

Breiman, L., Friedman, J., Stone, C. J. and Olshen, R. A.: Classification and regression trees, Chapman & Hall, London., 1984.

Breuer, L., Eckhardt, K. and Frede, H. G.: Plant parameter values for models in temperate climates, Ecol. Modell., 169(2–3), 237–293, doi:10.1016/S0304-3800(03)00274-6, 2003.

Brown, A. E., Zhang, L., Mcmahon, T. A., Western, A. W. and Vertessy, R. A.: A review of paired catchment studies for determining changes in water yield resulting from alterations in vegetation, J. Hydrol., 310(1–4), 28–61, doi:10.1016/j.jhydrol.2004.12.010, 2005.

Bundesanstalt für Geowissenschaften und Rohstoffe (BGR) & UNESCO: Groundwater Resources of the World 1 : 25 000 000 – Hannover, Paris, https://www.whymap.org [last accessed on 22/06/2018], 2008.

Butscher, C. and Huggenberger, P.: Intrinsic vulnerability assessment in karst areas: A numerical modeling approach, Water Resour. Res., 44(3), W03408, doi:10.1029/2007WR006277, 2008.

Butscher, C. and Huggenberger, P.: Enhanced vulnerability assessment in karst areas by combining mapping with modeling approaches, Sci. Total Environ., 407(3), 1153–1163, doi:10.1016/j.scitotenv.2008.09.033, 2009.

Calder, I. R.: Evaporation in the Uplands, John Wiley & Sons, Ltd, Chichester, UK., 1990.

Campolongo, F. and Saltelli, A.: Sensitivity analysis of an environmental model: an application of different analysis methods, Reliab. Eng. Syst. Saf., 57(1), 49–69, doi:10.1016/S0951-8320(97)00021-5, 1997.

Campolongo, F., Cariboni, J. and Saltelli, A.: An effective screening design for sensitivity analysis of large models, Environ. Model. Softw., 22(10), 1509–1518, doi:10.1016/j.envsoft.2006.10.004, 2007.

Campolongo, F., Saltelli, A. and Cariboni, J.: From screening to quantitative sensitivity analysis. A unified approach, Comput. Phys. Commun., 182(4), 978–988, doi:10.1016/j.cpc.2010.12.039, 2011.

Canora, F., Fidelibus, M. D., Sciortino, A. and Spilotro, G.: Variation of infiltration rate through karstic surfaces due to land use changes: A case study in Murgia (SE-Italy), Eng. Geol., 99(3–4), 210–227, doi:10.1016/j.enggeo.2007.11.018, 2008.

Cantón, Y., Villagarcía, L., José Moro, M., Serrano-Ortíz, P., Were, A., Javier Alcalá, F., Kowalski, A. S., Solé-Benet, A., Lázaro, R. and Domingo, F.: Temporal dynamics of soil water balance components in a karst range in southeastern Spain: estimation of potential recharge, Hydrol. Sci. J., 55(5), 737–753, doi:10.1080/02626667.2010.490530, 2010.

Cao, J., Hu, B., Groves, C., Huang, F., Yang, H. and Zhang, C.: Karst dynamic system and the carbon cycle, Zeitschrift für Geomorphol., 60(Suppl. 2), 35–55, doi:10.1127/zfg_suppl/2016/00304, 2016.

Carrer, D., Lafont, S., Roujean, J.-L., Calvet, J.-C., Meurey, C., Le Moigne, P. and Trigo, I. F.: Incoming Solar and Infrared Radiation Derived from METEOSAT: Impact on the Modeled Land Water and Energy Budget over France, J. Hydrometeorol., 13(2), 504–520, doi:10.1175/JHM-D-11-059.1, 2012.

Chan, S. C., Kahana, R., Kendon, E. J. and Fowler, H. J.: Projected changes in extreme precipitation over Scotland and Northern England using a high-resolution regional climate model, Clim. Dyn., 1–19, doi:10.1007/s00382-018-4096-4, 2018.

Chandler, D. G. and Bisogni, J. J.: The use of alkalinity as a conservative tracer in a study of near-surface hydrologic change in tropical karst, J. Hydrol., 216(3–4), 172–182, doi:10.1016/S0022-1694(99)00006-2, 1999.

Chaney, N. W., Herman, J. D., Ek, M. B. and Wood, E. F.: Deriving global parameter estimates for the Noah land surface model using FLUXNET and machine learning, J. Geophys. Res., 121(22), 13,218-13,235, doi:10.1002/2016JD024821, 2016.

Chen, Z., Hartmann, A., Wagener, T. and Goldscheider, N.: Dynamics of water fluxes and storages in an Alpine karst catchment under current and potential future climate conditions, Hydrol. Earth Syst. Sci. Discuss., 22(7), 3807–3823, doi:10.5194/hess-22-3807-2018, 2018.

Chen, Z., Auler, A. S., Bakalowicz, M., Drew, D., Griger, F., Hartmann, J., Jiang, G., Moosdorf, N., Richts, A., Stevanovic, Z., Veni, G. and Goldscheider, N.: The World Karst Aquifer Mapping project: concept, mapping procedure and map of Europe, Hydrogeol. J., 25(3), 771–785, doi:10.1007/s10040-016-1519-3, 2017.

Chini, L. P., Hurtt, G. C. and Frolking, S.: Harmonized Global Land Use for Years 1500 – 2100, V1, Data set, Available on-line [http://daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA, doi:10.3334/ORNLDAAC/1248, 2014.

CIESIN - Center for International Earth Science Information Network- Columbia University, Documentation for the Gridded Population of the World, Version 4 (GPWv4), Revision 10, Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC), doi:10.7927/H4B56GPT, 2017a.

CIESIN - Center for International Earth Science Information Network- Columbia University, Gridded Population of the World, Version 4 (GPWv4): Population Density Adjusted to Match 2015 Revision UN WPP Country Totals, Revision 10, Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC), doi:10.7927/H49884ZR., 2017b.

Clapp, R. B. and Hornberger, G. M.: Empirical equations for some soil hydraulic properties, Water Resour. Res., 14(4), 601–604, doi:10.1029/WR014i004p00601, 1978.

Coenders-Gerrits, A. M. J., Van der Ent, R. J., Bogaard, T. A., Wang-Erlandsson, L., Hrachowitz, M. and Savenije, H. H. G.: Uncertainties in transpiration estimates, Nature, 506, E1–E2, doi:10.1038/nature12925, 2014.

Confalonieri, R., Bellocchi, G., Tarantola, S., Acutis, M., Donatelli, M. and Genovese, G.: Sensitivity analysis of the rice model WARM in Europe: Exploring the effects of different locations, climates and methods of analysis on model sensitivity to crop parameters, Environ. Model. Softw., 25(4), 479–488, doi:10.1016/j.envsoft.2009.10.005, 2010.

Contreras, S., Boer, M. M., Alcala, F. J., Domingo, F., Garcia, M., Pulido-Bosch, A. and Puigdefabregas, J.: An ecohydrological modelling approach for assessing long-term recharge rates in semiarid karstic landscapes, J. Hydrol., 351(1–2), 42–57, doi:10.1016/j.jhydrol.2007.11.039, 2008.

Cosenza, A., Mannina, G., Vanrolleghem, P. A. and Neumann, M. B.: Global sensitivity analysis in wastewater applications: A comprehensive comparison of different methods, Environ. Model. Softw., 49, 40–52, doi:10.1016/j.envsoft.2013.07.009, 2013.

COST: Cost action 65 - Hydrogeological aspects of groundwater protection in karstic areas, Report EUR 16547, European Commission, Directorate-General XII Science, Research Development, Luxembourg., 1995.

Cox, P. M., Betts, R. A., Bunton, C. B., Essery, R. L. H., Rowntree, P. R. and Smith, J.: The impact of new land surface physics on the GCM simulation of climate and climate sensitivity, Clim. Dyn., 15(3), 183–203, doi:10.1007/s003820050276, 1999.

Coxon, C.: Agriculture and Karst, in Karst Management, edited by P. E. van Beynen, pp. 103–138, Springer Netherlands, Dordrecht., 2011.

Cuntz, M., Mai, J., Samaniego, L., Clark, M., Wulfmeyer, V., Branch, O., Attinger, S. and Thober, S.: The impact of standard and hard-coded parameters on the hydrologic fluxes in the Noah-MP land surface model, J. Geophys. Res., 121(18), 10,676-10,700, doi:10.1002/2016JD025097, 2016.

Custodio, E.: Aquifer overexploitation: What does it mean?, Hydrogeol. J., 10(2), 254–277, doi:10.1007/s10040-002-0188-6, 2002.

Cuthbert, M. O., Mackay, R. and Nimmo, J. R.: Linking soil moisture balance and source-responsive models to estimate diffuse and preferential components of groundwater recharge, Hydrol. Earth Syst. Sci., 17, 1003–1019, doi:10.5194/hess-17-1003-2013, 2013.

Dai, H. and Ye, M.: Variance-based global sensitivity analysis for multiple scenarios and models with implementation using sparse grid collocation, J. Hydrol., 528, 286–300, doi:10.1016/j.jhydrol.2015.06.034, 2015.

Dai, H., Ye, M., Walker, A. P. and Chen, X.: A new process sensitivity index to identify important system processes under process model and parametric uncertainty, Water Resour. Res., 53(4), 3476–3490, doi:10.1002/2016WR019715, 2017.

Dancelli, L., Manisera, M. and Vezzoli, M.: On two classes of weighted rank correlation measures deriving from the spearman's ρ, in Statistical Models for Data Analysis. Studies in Classification, Data Analysis, and Knowledge Organization, edited by P. Giudici, S. Ingrassia, and M. Vichi, pp. 107–114, The Springer, Heidelberg, 2013.

Davie, J. C. S., Falloon, P. D., Kahana, R., Dankers, R., Betts, R., Portmann, F. T., Wisser, D., Clark, D. B., Ito, A., Masaki, Y., Nishina, K., Fekete, B., Tessler, Z., Wada, Y., Liu, X., Tang, Q., Hagemann, S., Stacke, T., Pavlick, R., Schaphoff, S., Gosling, S. N., Franssen, W. and Arnell, N.: Comparing projections of future changes in runoff from hydrological and biome models in ISI-MIP, Earth Syst. Dyn., 4(2), 359–374, doi:10.5194/esd-4-359-2013, 2013.

Davy, R., Esau, I., Chernokulsky, A., Stephen, O. and Zilitinkevich, S.: Diurnal asymmetry to the observed global warming, Int. J. Climatol., 31(1), doi:10.1002/joc.4688, 2017.

Deardorff, J. W.: Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation, J. Geophys. Res., 83(C4), 1889–1903, doi:10.1029/JC083iC04p01889, 1978.

DeFries, R. and Eshleman, K. N.: Land-use change and hydrologic processes: a major focus for the future, Hydrol. Process., 18(11), 2183–2186, doi:10.1002/hyp.5584, 2004.

De Groen, M. M.: Modelling interception and transpiration at monthly time steps : introducing daily variability through Markov chains, Ph.D. thesis, Delft University of Technology, Delft, The Netherlands. [online] Available from: https://books.google.com/books?id=-birGVG44nkC&pgis=1, 2002.

Déqué, M., Rowell, D. P., Lüthi, D., Giorgi, F., Christensen, J. H., Rockel, B., Jacob, D., Kjellström, E., De Castro, M. and Van Den Hurk, B.: An intercomparison of regional climate simulations for Europe: Assessing uncertainties in model projections, Clim. Change, 81(SUPPL. 1), 53–70, doi:10.1007/s10584-006-9228-x, 2007.

Dickinson, R. E.: Modeling evapotranspiration for three dimensional global climate models, in Climate Processes and Climate Sensitivity, edited by J. E. Hansen and T. Takahashi, American Geophysical Union, Washington, D. C., 1984.

Döll, P. and Fiedler, K.: Global-scale modeling of groundwater recharge, Hydrol. Earth Syst. Sci., 12(3), 863–885, doi:10.5194/hess-12-863-2008, 2008.

Döll, P., Kaspar, F. and Lehner, B.: A global hydrological model for deriving water availability indicators: Model tuning and validation, J. Hydrol., 270(1–2), 105–134, doi:10.1016/S0022-1694(02)00283-4, 2003.

Döll, P., Douville, H., Güntner, A., Müller Schmied, H. and Wada, Y.: Modelling Freshwater Resources at the Global Scale: Challenges and Prospects, Surv. Geophys., 37(2), 195–221, doi:10.1007/s10712-015-9343-1, 2016.

Doummar, J., Sauter, M. and Geyer, T.: Simulation of flow processes in a large scale karst system with an integrated catchment model (Mike She) – Identification of relevant parameters influencing spring discharge, J. Hydrol., 426–427, 112–123, doi:10.1016/j.jhydrol.2012.01.021, 2012.

Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H. V., Gusev, Y. M., Habets, F., Hall, A., Hay, L., Hogue, T., Huang, M., Leavesley, G., Liang, X., Nasonova, O. N., Noilhan, J., Oudin, L., Sorooshian, S., Wagener, T. and Wood, E. F.: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, J. Hydrol., 320(1–2), 3–17, doi:10.1016/j.jhydrol.2005.07.031, 2006.

Eckhardt, K., Breuer, L. and Frede, H.-G.: Parameter uncertainty and the significance of simulated land use change effects, J. Hydrol., 273(1–4), 164–176, doi:10.1016/S0022-1694(02)00395-5, 2003.

Ecofor: Site atelier de Font Blanche, [online] Available from: http://www.gip-ecofor.org/f-ore-t/fontBlanche.php (Accessed 13 December 2016), n.d.

Efron, B. and Tibshirani, R. J.: An Introduction to the bootstrap, Chapman and Hall/CRC, New York, USA., 1994.

Ekström, M., Grose, M. R. and Whetton, P. H.: An appraisal of downscaling methods used in climate change research, WIREs Clim. Chang., 6(3), 301–319, doi:10.1002/wcc.339, 2015.

Essery, R., Best, M. and Cox, P.: MOSES 2.2 Technical Documentation, Berks, UK. [online] Available from: http://www.metoffice.gov.uk/media/pdf/9/j/HCTN_30.pdf, 2001.

ETH Zürich - Department of Environmental Systems Science - Institute of Agricultural Sciences: Grassland Sciences - Früebüel (ZG), [online] Available from: http://www.gl.ethz.ch/infrastructure/sites/frueebueel.html (Accessed 30 April 2018), 2018.

Falkenmark, M. and Rockström, J.: The new blue and green water paradigm : breaking new ground for water resources planning and management, J. Water Resour. Plan. Manag., 132(3), 129–132, doi:10.1061/(ASCE)0733-9496(2006)132:3(129), 2006.

Fang, H., Jiang, C., Li, W., Wei, S., Baret, F., Chen, J. M., Garcia-Haro, J., Liang, S., Liu, R., Myneni, R. B., Pinty, B., Xiao, Z. and Zhu, Z.: Characterization and intercomparison of global moderate resolution leaf area index (LAI) products: Analysis of climatologies and theoretical uncertainties, J. Geophys. Res. Biogeosciences, 118(2), 529–548, doi:10.1002/jgrg.20051, 2013.

Febles-González, J. M., Vega-Carreño, M. B., Tolón-Becerra, A. and Lastra-Bravo, X.: Assessment of soil erosion in karst regions of Havana, Cuba, L. Degrad. Dev., 23(5), 465–474, doi:10.1002/ldr.1089, 2012.

Federer, C. A.: Transpirational Supply and Demand: plant, soil, and atmospheric effects evaluated by simulation, Water Resour. Res., 18(2), 355–362, doi:10.1029/WR018i002p00355, 1982.

Federer, C. A., Vörösmarty, C. and Fekete, B.: Sensitivity of Annual Evaporation to Soil and Root Properties in Two Models of Contrasting Complexity, J. Hydrometeorol., 4(6), 1276–1290, doi:10.1175/1525-7541(2003)004<1276:SOAETS>2.0.CO;2, 2003.

Feng, X., Porporato, A. and Rodriguez-Iturbe, I.: Changes in rainfall seasonality in the tropics, Nat. Clim. Chang., 3(9), 811–815, doi:10.1038/nclimate1907, 2013.

Ferguson, G. and Gleeson, T.: Vulnerability of coastal aquifers to groundwater use and climate change, Nat. Clim. Chang., 2(5), 342–345, doi:10.1038/nclimate1413, 2012.

Fleury, P., Plagnes, V. and Bakalowicz, M.: Modelling of the functioning of karst aquifers with a reservoir model : Application to Fontaine de Vaucluse (South of France), J. Hydrol., 345(1–2), 38–49, doi:doi.org/10.1016/j.jhydrol.2007.07.014, 2007.

Foken, T., Leuning, R., Oncley, S. R., Mauder, M. and Aubinet, M.: Corrections and Data Quality Control, in Eddy Covariance: A Practical Guide to Measurement and Data Analysis, edited by M. Aubinet, T. Vesala, and D. Papale, Springer Netherlands, Dordrecht., 85–131, 2012.

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O'Connell, C., Ray, D. K., West, P. C., Balzer, C., Bennett, E. M., Carpenter, S. R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., Tilman, D. and Zaks, D. P. M.: Solutions for a cultivated planet, Nature, 478(7369), 337–342, doi:10.1038/nature10452, 2011.

Ford, D. and Williams, P.: Karst Hydrogeology and Geomorphology, John Wiley & Sons Ltd, Chichester, UK, 2007.

Forrester, A., Sobester, A. and Keane, A.: Engineering design via surrogate modelling: a practical guide, John Wiley & Sons., 2008.

Foster, S., Hirata, R. and Andreo, B.: The aquifer pollution vulnerability concept: aid or impediment in promoting groundwater protection?, Hydrogeol. J., 21(7), 1389–1392, doi:10.1007/s10040-013-1019-7, 2013.

Freer, J., Beven, K. and Ambroise, B.: Bayesian estimation of uncertainty in runoff production and the value of data: An application of the GLUE approach, Water Resour. Res., 32(7), 2161–2173, doi:10.1029/95WR03723, 1996.

Fyfe, J. C., Derksen, C., Mudryk, L., Flato, G. M., Santer, B. D., Swart, N. C., Molotch, N. P., Zhang, X., Wan, H., Arora, V. K., Scinocca, J. and Jiao, Y.: Large near-Term projected snowpack loss over the western United States, Nat. Commun., 8, 14996, doi:10.1038/ncomms14996, 2017.

Gash, J. H. C.: An analytical model of rainfall interception by forests, Quaterly J. R. Meteorol. Soc., 105(443), 43–55, doi:10.1002/qj.49710544304, 1979.

Gea-Izquierdo, G., Guibal, F., Joffre, R., Ourcival, J. M., Simioni, G. and Guiot, J.: Modelling the climatic drivers determining photosynthesis and carbon allocation in evergreen Mediterranean forests using multiproxy long time series, Biogeosciences, 12(12), 3695–3712, doi:10.5194/bg-12-3695-2015, 2015.

Gergel, D. R., Nijssen, B., Abatzoglou, J. T., Lettenmaier, D. P. and Stumbaugh, M. R.: Effects of climate change on snowpack and fire potential in the western USA, Clim. Change, 141(2), 287–299, doi:10.1007/s10584-017-1899-y, 2017.

Gerrits, M.: The role of interception in the hydrological cycle, Ph.D. thesis, Delft University of Technology, Delft, The Netherlands. [online] Available from: https://repository.tudelft.nl/islandora/object/uuid:7dd2523b-2169-4e7e-992c-365d2294d02e?collection=research, 2010.

Gerten, D., Schaphoff, S., Haberlandt, U., Lucht, W. and Sitch, S.: Terrestrial vegetation and water balance - Hydrological evaluation of a dynamic global vegetation model, J. Hydrol., 286(1–4), 249–270, doi:10.1016/j.jhydrol.2003.09.029, 2004.

Gharari, S., Hrachowitz, M., Fenicia, F. and Savenije, H. H. G.: An approach to identify time consistent model parameters: Sub-period calibration, Hydrol. Earth Syst. Sci., 17(1), 149–161, doi:10.5194/hess-17-149-2013, 2013.

Giordano, M.: Global Groundwater? Issues and Solutions, Annu. Rev. Environ. Resour., 34(1), 153–178, doi:10.1146/annurev.environ.030308.100251, 2009.

Giorgi, F., Im, E. S., Coppola, E., Diffenbaugh, N. S., Gao, X. J., Mariotti, L. and Shi, Y.: Higher hydroclimatic intensity with global warming, J. Clim., 24(20), 5309–5324, doi:10.1175/2011JCLI3979.1, 2011.

Giuntoli, I., Vidal, J.-P., Prudhomme, C. and Hannah, D. M.: Future hydrological extremes: the uncertainty from multiple global climate and global hydrological models, Earth Syst. Dyn., 6(1), 267–285, doi:10.5194/esd-6-267-2015, 2015.

Gleeson, T., VanderSteen, J., Sophocleous, M. A., Taniguchi, M., Alley, W. M., Allen, D. M. and Zhou, Y.: Groundwater sustainability strategies, Nat. Geosci., 3(6), 378–379, doi:10.1038/ngeo881, 2010.

Gleeson, T., Befus, K. M., Jasechko, S., Luijendijk, E. and Cardenas, M. B.: The global volume and distribution of modern groundwater, Nat. Geosci., 9(2), 161–167, doi:10.1038/ngeo2590, 2016.

Gordon, L. J., Steffen, W., Jönsson, B. F., Folke, C., Falkenmark, M. and Johannessen, A.: Human modification of global water vapor flows from the land surface, Proc. Natl. Acad. Sci., 102(21), 7612–7617, doi:10.1073/pnas.0500208102, 2005.

Gosling, S. N. and Arnell, N. W.: Simulating current global river runoff with a global hydrological model: Model revisions, validation, and sensitivity analysis, Hydrol. Process., 25(7), 1129–1145, doi:10.1002/hyp.7727, 2011.

Gosling, S. N., Zaherpour, J., Mount, N. J., Hattermann, F. F., Dankers, R., Arheimer, B., Breuer, L., Ding, J., Haddeland, I., Kumar, R., Kundu, D., Liu, J., van Griensven, A., Veldkamp, T. I. E., Vetter, T., Wang, X. and Zhang, X.: A comparison of changes in river runoff from multiple global and catchment-scale hydrological models under global warming scenarios of 1 C, 2 C and 3 C, Clim. Change, 141(3), 577–595, doi:10.1007/s10584-016-1773-3, 2017.

Granier, A., Bréda, N., Biron, P. and Villette, S.: A lumped water balance model to evaluate duration and intensity of drought constraints in forest stands, Ecol. Modell., 116(2–3), 269–283, doi:10.1016/S0304-3800(98)00205-1, 1999.

Grillakis, M. G., Tsanis, I. K. and Koutroulis, A. G.: Application of the HBV hydrological model in a flash flood case in Slovenia, Nat. Hazards Earth Syst. Sci., 10(12), 2713–2725, doi:10.5194/nhess-10-2713-2010, 2010.

Groves, D. G. and Lempert, R. J.: A new analytic method for finding policy-relevant scenarios, Glob. Environ. Chang., 17(1), 73–85, doi:10.1016/j.gloenvcha.2006.11.006, 2007.

Grünzweig, J. M., Lin, T., Rotenberg, E., Schwartz, A. and Yakir, D.: Carbon sequestration in arid-land forest, Glob. Chang. Biol., 9(5), 791–799, doi:10.1046/j.1365-2486.2003.00612.x, 2003.

Güntner, A., Stuck, J., Werth, S., Döll, P., Verzano, K. and Merz, B.: A global analysis of temporal and spatial variations in continental water storage, Water Resour. Res., 43(5), W05416, doi:10.1029/2006WR005247, 2007.

Guo, D., Westra, S. and Maier, H. R.: Use of a scenario-neutral approach to identify the key hydro-meteorological attributes that impact runoff from a natural catchment, J. Hydrol., 554, 317–330, doi:10.1016/j.jhydrol.2017.09.021, 2017.

Hagemann, S.: An Improved Land Surface Parameter Dataset for Global and Regional Climate Models, Hamburg, Germany. [online] Available from: http://epub.sub.uni-hamburg.de/epub/volltexte/2010/5248/pdf/max_scirep_336.pdf, 2002.

Hamdia, K. M., Msekh, M. A., Silani, M., Vu-Bac, N., Zhuang, X., Nguyen-Thoi, T. and Rabczuk, T.: Uncertainty quantification of the fracture properties of polymeric nanocomposites based on phase field modeling, Compos. Struct., 133, 1177–1190, doi:10.1016/j.compstruct.2015.08.051, 2015.

Hamilton, L. S., Dudley, N., Greminger, G., Hassan, N., Lamb, D., S., S. and Tognetti, S.: Forest and Water, FAO forestry paper 155, Food and Agriculture Organization of the United Nations, Rome, Forestry, 2008.

Hao, Y., Yeh, T. C. J., Gao, Z., Wang, Y. and Zhao, Y.: A gray system model for studying the response to climatic change: The Liulin karst springs, China, J. Hydrol., 328(3–4), 668–676, doi:10.1016/j.jhydrol.2006.01.022, 2006.

Hargreaves, G. H. and Samani, Z. A.: Reference crop evapotranspiration from temperature, Appl. Eng. Agric., 1(2), 96–99, doi:10.13031/2013.26773, 1985.

Hartmann, A. and Baker, A.: Modelling karst vadose zone hydrology and its relevance for paleoclimate reconstruction, Earth-Sciences Rev., 172, 178–192, doi:10.1016/j.earscirev.2017.08.001, 2017.

Hartmann, A., Lange, J., Vivó Aguado, À., Mizyed, N., Smiatek, G. and Kunstmann, H.: A multi-model approach for improved simulations of future water availability at a large Eastern Mediterranean karst spring, J. Hydrol., 468–469, 130–138, doi:10.1016/j.jhydrol.2012.08.024, 2012a.

Hartmann, A., Lange, J., Weiler, M., Arbel, Y. and Greenbaum, N.: A new approach to model the spatial and temporal variability of recharge to karst aquifers, Hydrol. Earth Syst. Sci., 16(7), 2219–2231, doi:10.5194/hess-16-2219-2012, 2012b.

Hartmann, A., Barberá, J. A., Lange, J., Andreo, B. and Weiler, M.: Progress in the hydrologic simulation of time variant recharge areas of karst systems - Exemplified at a karst spring in Southern Spain, Adv. Water Resour., 54, 149–160, doi:10.1016/j.advwatres.2013.01.010, 2013a.

Hartmann, A., Wagener, T., Rimmer, A., Lange, J., Brielmann, H. and Weiler, M.: Testing the realism of model structures to identify karst system processes using water quality and quantity signatures, Water Resour. Res., 49(6), 3345–3358, doi:10.1002/wrcr.20229, 2013b.

Hartmann, A., Goldscheider, N., Wagener, T., Lange, J. and Weiler, M.: Karst water resources in a changing world: Review of hydrological modeling approaches, Rev. Geophys., 52(3), 218–242, doi:10.1002/2013RG000443, 2014.

Hartmann, A., Gleeson, T., Rosolem, R., Pianosi, F., Wada, Y. and Wagener, T.: A large-scale simulation model to assess karstic groundwater recharge over Europe and the Mediterranean, Geosci. Model Dev., 8(6), 1729–1746, doi:10.5194/gmd-8-1729-2015, 2015.

Hartmann, A., Kobler, J., Kralik, M., Dirnböck, T., Humer, F. and Weiler, M.: Model-aided quantification of dissolved carbon and nitrogen release after windthrow disturbance in an Austrian karst system, Biogeosciences, 13(1), 159–174, doi:10.5194/bg-13-159-2016, 2016.

Hartmann, A., Gleeson, T., Wada, Y. and Wagener, T.: Enhanced groundwater recharge rates and altered recharge sensitivity to climate variability through subsurface heterogeneity, Proc. Natl. Acad. Sci., 114(11), 2842–2847, doi:10.1073/pnas.1614941114, 2017.

Hattermann, F. F., Vetter, T., Breuer, L., Su, B., Daggupati, P., Donnelly, C., Fekete, B., Flörke, F., Gosling, S. N. and Hoffmann, P.: Sources of uncertainty in hydrological climate impact assessment: a cross-scale study, Environ. Res. Lett., 13(1), 15006, doi:10.1088/1748-9326/aa9938, 2018.

Haughton, N., Abramowitz, G. and Pitman, A. J.: On the predictability of land surface fluxes from meteorological variables, Geosci. Model Dev., 11(1), 195–212, doi:10.5194/gmd-11-195-2018, 2018.

Hempel, S., Frieler, K., Warszawski, L., Schewe, J. and Piontek, F.: A trend-preserving bias correction – the ISI-MIP approach, Earth Syst. Dyn., 4(2), 219–236, doi:10.5194/esd-4-219-2013, 2013.

Hendrickx, J. M. H. and Flury, M.: Uniform and Preferential Flow Mechanisms in the Vadose Zone, in Conceptual Models of Flow and Transport in the Fractured Vadose Zone, edited by National Research Council, The National Academies Press, Washington, D.C., 149–188, 2001.

Hendriks, D. M. D., Kuijper, M. J. M. and van Ek, R.: Groundwater impact on environmental flow needs of streams in sandy catchments in the Netherlands, Hydrol. Sci. J., 59(3–4), 562–577, doi:10.1080/02626667.2014.892601, 2014.

Herman, J. D., Kollat, J. B., Reed, P. M. and Wagener, T.: 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, doi:10.5194/hess-17-2893-2013, 2013.

Hill, M. C. and Tiedeman, C. R.: Effective Groundwater Model Calibration: With Analysis of Data, Sensitivities, Predictions, and Uncertainty, John Wiley & Sons, Inc., 2007.

Hogue, T. S., Bastidas, L. A., Gupta, H. V. and Sorooshian, S.: Evaluating model performance and parameter behavior for varying levels of land surface model complexity, Water Resour. Res., 42(8), W08430, doi:10.1029/2005WR004440, 2006.

Holman, I. P., Brown, C., Janes, V. and Sandars, D.: Can we be certain about future land use change in Europe? A multi-scenario, integrated-assessment analysis, Agric. Syst., 151, 126–135, doi:10.1016/j.agsy.2016.12.001, 2017.

Hong, E.-M., Pachepsky, Y. A., Whelan, G. and Nicholson, T.: Simpler models in environmental studies and predictions, Crit. Rev. Environ. Sci. Technol., 47(18), 1669–1712, doi:10.1080/10643389.2017.1393264, 2017.

Huebsch, M., Fenton, O., Horan, B., Hennessy, D., Richards, K. G., Jordan, P., Goldscheider, N., Butscher, C. and Blum, P.: Mobilisation or dilution? Nitrate response of karst springs to high rainfall events, Hydrol. Earth Syst. Sci., 18(11), 4423–4435, doi:10.5194/hess-18-4423-2014, 2014.

Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R, A., Feddema, J., Fischer, G., Fisk, J. P., Hibbard, K., Houghton, R. A., Janetos, A., Jones, C. D., Kindermann, G., Kinoshita, T., Klein Goldewijk, K., Riahi, K., Shevliakova, E., Smith, S., Stehfest, E., Thomson, A., Thornton, P., van Vuuren, D. P. and Wang, Y. P.: Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands, Clim. Change, 109, 117–161, doi:10.1007/s10584-011-0153-2, 2011.

Iman, R. L. and Conover, W. J.: A measure of top-down correlation, Technometrics, 29(3), 351–357, doi:10.2307/1269344, 1987.

Inselberg, A.: Parallel coordinates: Visual multidimensional geometry and its applications, Springer, New York., 2009.

IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by V. B. and P. M. M. Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, Cambridge University Press., 2013.

IPCC: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Part B: Regional Aspects, edited by V. R. Barros, C. B. Field, D. J. Dokken, M. D. Mastrandrea, K. J. Mach, T. E. Bilir, E. S. Kissel, M. Chatterjee, K. L. Ebi, Y. O. Estrada, R. C. Genova, B. Girma, A. N. Levy, S. MacCracken, P. R. Mastrandrea, and L. L. White, Cambridge University Press., 2014.

Isaksson, A., Wallman, M., Göransson, H. and Gustafsson, M. G.: Cross-validation and bootstrapping are unreliable in small sample classification, Pattern Recognit. Lett., 29(14), 1960–1965, doi:10.1016/j.patrec.2008.06.018, 2008.

Ishigami, T. and Homma, T.: An importance quantification technique in uncertainty analysis for computer models, in Proceedings of the ISUMA 90', First International Symposium on Uncertainty Modelling and Analysis, pp. 398–403, University of Maryland, 1990.

Iván, V. and Mádl-Szőnyi, J.: State of the art of karst vulnerability assessment: overview, evaluation and outlook, Environ. Earth Sci., 76, 112, doi:10.1007/s12665-017-6422-2, 2017.

Ivanov, V. Y., Bras, R. L. and Curtis, D. C.: A weather generator for hydrological, ecological, and agricultural applications, Water Resour. Res., 43(10), doi:10.1029/2006WR005364, 2007.

Jarvis, P. G.: The Interpretation of the Variations in Leaf Water Potential and Stomatal Conductance Found in Canopies in the Field, Phil. Trans. R. Soc. Lond. B., 273(927), 593–610, doi:10.1098/rstb.1976.0035, 1976.

Jasechko, S., Birks, S. J., Gleeson, T., Wada, Y., Fawcett, P. J., Sharp, Z. D., McDonnell, J. J. and Welker, J. M.: The pronounced seasonlity of global groundwater recharge, Water Resour. Res., 50(11), 8845–8867, doi:10.1002/2014WR015809, 2014.

Jeannin, P.-Y. and Grasso, D. A.: Permeability and hydrodynamic behavior of karstic environment, in Karst Waters Environmental Impact, edited by G. Gunay and A. I. Johnson, pp. 335–342, A.A. Balkema, Rotterdam, Netherlands., 1997.

Jeannin, P.-Y., Cornaton, F., Zwahlen, F. and Perrochet, P.: VULK: a tool for intrinsic vulnerability assessment and validation. Sciences et techniques de l'environnement, Sci. Tech. l'environnement Mémoire hors-série, 185–190, 2001.

Jia, Z., Zang, H., Zheng, X. and Xu, Y.: Climate Change and Its Influence on the Karst Groundwater Recharge in the Jinci Spring Region, Northern China, Water, 9(4), 267, doi:10.3390/w9040267, 2017.

Jianhua, C., Daoxian, Y., Liqiang, T., Mallik, A., Hui, Y. and Fen, H.: An Overview of Karst Ecosystem in Southwest China: Current State and Future Management, J. Resour. Ecol., 6(4), 247–256, doi:10.5814/j.issn.1674-764x.2015.04.008, 2015.

Jothityangkoon, C. and Sivapalan, M.: Framework for exploration of climatic and landscape controls on catchment water balance, with emphasis on inter-annual variability, J. Hydrol., 371(1–4), 154–168, doi:10.1016/j.jhydrol.2009.03.030, 2009.

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B. E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E. J., Papale, D., Sottocornola, M., Vaccari, F. and Williams, C.: Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, J. Geophys. Res. Biogeosciences, 116(G3), G00J07, doi:10.1029/2010JG001566, 2011.

Kačaroğlu, F.: Review of groundwater pollution and protection in karst areas, Water. Air. Soil Pollut., 113(1–4), 337–356, doi:10.1023/A:1005014532330, 1999.

Kannan, N., White, S. M., Worrall, F. and Whelan, M. J.: Sensitivity analysis and identification of the best evapotranspiration and runoff options for hydrological modelling in SWAT-2000, J. Hydrol., 332(3–4), 456–466, doi:10.1016/j.jhydrol.2006.08.001, 2007.

Kasprzyk, J. R., Nataraj, S., Reed, P. M. and Lempert, R. J.: Many objective robust decision making for complex environmental systems undergoing change, Environ. Model. Softw., 42, 55–71, doi:10.1016/j.envsoft.2012.12.007, 2013.

Keese, K. E., Scanlon, B. R. and Reedy, R. C.: Assessing controls on diffuse groundwater recharge using unsaturated flow modeling, Water Resour. Res., 41(6), W06010, doi:10.1029/2004WR003841, 2005.

Kelleher, C., Wagener, T., McGlynn, B., Ward, A. S., Gooseff, M. N. and Payn, R. A.: Identifiability of transient storage model parameters along a mountain stream, Water Resour. Res., 49(9), 5290–5306, doi:10.1002/wrcr.20413, 2013.

Kergoat, L.: A model for hydrological equilibrium of leaf area index on a global scale, J. Hydrol., 212–213, 268–286, doi:10.1016/S0022-1694(98)00211-X, 1998.

Khorashadi Zadeh, F., Nossent, J., Sarrazin, F., Pianosi, F., van Griensven, A., Wagener, T. and Bauwens, W.: Comparison of variance-based and moment-independent global sensitivity analysis approaches by application to the SWAT model, Environ. Model. Softw., 91, 210–222, doi:10.1016/j.envsoft.2017.02.001, 2017.

Kim, J. H. and Jackson, R. B.: A Global Analysis of Groundwater Recharge for Vegetation, Climate, and Soils, Vadose Zo. J., 11(1), doi:10.2136/vzj2011.0021RA, 2012.

Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water Resour. Res., 42(3), W03S04, doi:10.1029/2005WR004362, 2006.

Kløve, B., Ala-aho, P., Bertrand, G., Boukalova, Z., Ertürk, A., Goldscheider, N., Ilmonen, J., Karakaya, N., Kupfersberger, H., Kvœrner, J., Lundberg, A., Mileusnić, M., Moszczynska, A., Muotka, T., Preda, E., Rossi, P., Siergieiev, D., Šimek, J., Wachniew, P., Angheluta, V. and Widerlund, A.: Groundwater dependent ecosystems. Part I: Hydroecological status and trends, Environ. Sci. Policy, 14(7), 770–781, doi:10.1016/j.envsci.2011.04.002, 2011.

Knohl, A., Schulze, E. D., Kolle, O. and Buchmann, N.: Large carbon uptake by an unmanaged 250-year-old deciduous forest in Central Germany, Agric. For. Meteorol., 118(3–4), 151–167, doi:10.1016/S0168-1923(03)00115-1, 2003.

Kollat, J. B., Reed, P. M. and Wagener, T.: When are multiobjective calibration trade-offs in hydrologic models meaningful?, Water Resour. Res., 48(3), W03520, doi:10.1029/2011WR011534, 2012.

Kolmogorov, A.: Sulla determinazione empirica di una legge di distribuzione, G. dell'Istituto Ital. degli Attuari, 4, 83–91, 1933.

Körner, C.: Leaf diffusive conductances in the major vegetation types of the globe, in Ecophysiology of photosynthesis, edited by E. D. Schulze and M. M. Caldwell, pp. 463–490, Springer Study Edition, vol 100, Springer, Berlin, Heidelberg., 1995.

Kroes, J. G., van Dam, J. C., Groenendijk, P., Hendriks, A. R. F. and Jacobs, C. M. J.: SWAP version 3.2: Theory description and user manual, Report 1649, Alterra, Wageningen, The Netherlands., 2008.

Kumar, R., Samaniego, L. and Attinger, S.: Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations, Water Resour. Res., 49(1), 360–379, doi:10.1029/2012WR012195, 2013.

Kumar, A., Chen, F., Niyogi, D., Alfieri, J. G., Ek, M. and Mitchell, K.: Evaluation of a Photosynthesis-Based Canopy Resistance Formulation in the Noah Land-Surface Model, Boundary-Layer Meteorol., 138(2), 263–284, doi:10.1007/s10546-010-9559-z, 2011.

Kundzewicz, Z. W., Krysanova, V., Benestad, R. E., Hov, Ø., Piniewski, M. and Otto, I. M.: Uncertainty in climate change impacts on water resources, Environ. Sci. Policy, 79, 1–8, doi:10.1016/j.envsci.2017.10.008, 2018.

Laio, F., Porporato, A., Ridolfi, L. and Rodriguez-Iturbe, I.: On the seasonal dynamics of mean soil moisture, J. Geophys. Res., 107(D15), ACL 8-1–ACL 8-9, doi:10.1029/2001JD001252, 2002.

Lawrence, D. M., Hurtt, G. C., Arneth, A., Brovkin, V., Calvin, K. V., Jones, A. D., Jones, C. D., Lawrence, P. J., Noblet-Ducoudré, N. De, Pongratz, J., Seneviratne, S. I. and Shevliakova, E.: The Land Use Model Intercomparison Project (LUMIP) contribution to CMIP6: Rationale and

experimental design, Geosci. Model Dev., 9(9), 2973–2998, doi:10.5194/gmd-9-2973-2016, 2016.

Lemordant, L., Gentine, P., Swann, A. S., Cook, B. I. and Scheff, J.: Critical impact of vegetation physiology on the continental hydrologic cycle in response to increasing CO2, Proc. Natl. Acad. Sci., 115(16), 4093–4098, doi:10.1073/pnas.1720712115, 2018.

Leta, O. T., Nossent, J., Velez, C., Shrestha, N. K., van Griensven, A. and Bauwens, W.: Assessment of the different sources of uncertainty in a SWAT model ofthe River Senne (Belgium), Environ. Model. Softw., 68, 129–146, doi:10.1016/j.envsoft.2015.02.010, 2015.

Lewis, S. C. and Karoly, D. J.: Evaluation of historical diurnal temperature range trends in CMIP5 models, J. Clim., 26(22), 9077–9089, doi:10.1175/JCLI-D-13-00032.1, 2013.

Lhomme, J. P., Boudhina, N. and Masmoudi, M. M.: Technical Note: On the Matt–Shuttleworth approach to estimate crop water requirements, Hydrol. Earth Syst. Sci., 18(11), 4341–4348, doi:10.5194/hess-18-4341-2014, 2014.

Li, X., Shu, L., Liu, L., Yin, D. and Wen, J.: Sensitivity analysis of groundwater level in Jinci Spring Basin (China) based on artificial neural network modeling, Hydrogeol. J., 20(4), 727–738, doi:10.1007/s10040-012-0843-5, 2012.

Li, X. Y., Contreras, S. and Solé-Benet, A.: Spatial distribution of rock fragments in dolines: A case study in a semiarid Mediterranean mountain-range (Sierra de Gádor, SE Spain), Catena, 70(3), 366–374, doi:10.1016/j.catena.2006.11.003, 2007.

Li, X. Y., Contreras, S., Solé-Benet, A., Cantón, Y., Domingo, F., Lázaro, R., Lin, H., Van Wesemael, B. and Puigdefábregas, J.: Controls of infiltration-runoff processes in Mediterranean karst rangelands in SE Spain, Catena, 86(2), 98–109, doi:10.1016/j.catena.2011.03.003, 2011.

Liang, X., Lettenmaier, D. P., Wood, E. F. and Burges, S. J.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, J. Geophys. Res., 99(D7), 14415–14428, doi:10.1029/94JD00483, 1994.

Ljung: System Identification: Theory for the User, Second., Prentice Hall, Upper Saddle River, NJ., 1999.

Loáiciga, H. A., Maidment, D. R. and Valdes, J. B.: Climate-change impacts in a regional karst aquifer, Texas, USA, J. Hydrol., 227(1–4), 173–194, doi:10.1016/S0022-1694(99)00179-1, 2000.

Lu, Y., Liu, S., Weng, L., Wang, L., Li, Z. and Xu, L.: Fractal analysis of cracking in a clayey soil under freeze-thaw cycles, Eng. Geol., 208, 93–99, doi:10.1016/j.enggeo.2016.04.023, 2016.

Maréchal, J.-C., Ladouche, B., Dörfliger, N. and Lachassagne, P.: Interpretation of pumping tests in a mixed flow karst system, Water Resour. Res., 44(5), W05401, doi:10.1029/2007WR006288, 2008.

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Férnandez-Prieto, D., Beck, H. E., Dorigo, W. A. and Verhoest, N. E. C.: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, Geosci. Model Dev., 10(5), 1903–1925, doi:10.5194/gmd-10-1903-2017, 2017.

Marty, C., Schlögl, S., Bavay, M. and Lehning, M.: How much can we save? Impact of different emission scenarios on future snow cover in the Alps, Cryosphere, 11(1), 517–529, doi:10.5194/tc-11-517-2017, 2017.

Marx, A., Kumar, R., Thober, S., Rakovec, O., Wanders, N., Zink, M., Wood, E. F., Ming, P., Sheffield, J. and Samaniego, L.: Climate change alters low flows in Europe under a 1.5, 2, and 3 degree global warming, Hydrol. Earth Syst. Sci., 22(2), 1017–1032, doi:hess-22-1017-2018, 2018.

Masson, V., Champeaux, J. L., Chauvin, F., Meriguet, C. and Lacaze, R.: A global database of land surface parameters at 1-km resolution in meteorological and climate models, J. Clim., 16(9), 1261–1282, doi:10.1175/1520-0442-16.9.1261, 2003.

Mathworks: Matlab Statistics and Machine Learning Toolbox Release 2016a, Mathworks Inc., Natwick, MA, USA, 2016.

Maxwell, R. M. and Condon, L. E.: Connections between groundwater flow and transpiration partitioning, Science, 353(6297), 377–380, doi:10.1126/science.aaf7891, 2016.

McCabe, M. F., Ershadi, A., Jimenez, C., Miralles, D. G., Michel, D. and Wood, E. F.: The GEWEX LandFlux project: Evaluation of model evaporation using tower-based and globally gridded forcing data, Geosci. Model Dev., 9(1), 283–305, doi:10.5194/gmd-9-283-2016, 2016.

Melsen, L. A., Addor, N., Mizukami, N., Newman, A. J., Torfs, P. J. J. F., Clark, M. P., Uijlenhoet, R. and Teuling, A. J.: Mapping (dis)agreement in hydrologic projections, Hydrol. Earth Syst. Sci., 22(3), 1775–1791, doi:10.5194/hess-22-1775-2018, 2018.

Mendoza, P. A., Clark, M. P., Barlage, M., Rajagopalan, B., Samaniego, L., Abramowitz, G. and Gupta, H.: Are we unnecessarily constraining the agility of complex process-based models?, Water Resour. Res., 51(1), 716–728, doi:10.1002/2014WR015820, 2015a.

Mendoza, P. A., Clark, M. P., Mizukami, N., Newman, A. J., Barlage, M., Gutmann, E. D., Rasmussen, R. M., Rajagopalan, B., Brekke, L. D. and Arnold, J. R.: Effects of Hydrologic Model Choice and Calibration on the Portrayal of Climate Change Impacts, J. Hydrometeorol., 16(2), 762–780, doi:10.1175/JHM-D-14-0104.1, 2015b.

Mendoza, P. A., Clark, M. P., Mizukami, N., Gutmann, E. D., Arnold, J. R., Brekke, L. D. and Rajagopalan, B.: How do hydrologic modeling decisions affect the portrayal of climate change impacts?, Hydrol. Process., 30(7), 1071–1095, doi:10.1002/hyp.10684, 2016.

Milly, P. C. D. and Shmakin, A. B.: Global Modeling of Land Water and Energy Balances. Part I: The Land Dynamics (LaD) Model, J. Hydrometeorol., 3(3), 283–299, doi:10.1175/1525-754110.1175/1525-7541(2002)003<0283:GMOLWA>2.0.CO;2, 2002.

Miralles, D. G., Gash, J. H., Holmes, T. R. H., De Jeu, R. A. M. and Dolman, A. J.: Global canopy interception from satellite observations, J. Geophys. Res., 115(D16), D16122, doi:10.1029/2009JD013530, 2010.

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A. and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, Hydrol. Earth Syst. Sci., 15(2), 453–469, doi:10.5194/hess-15-453-2011, 2011.

Miralles, D. G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M. F., Hirschi, M., Martens, B., Dolman, A. J., Fisher, J. B., Mu, Q., Seneviratne, S. I., Wood, E. F. and Fernaìndez-Prieto, D.: The WACMOS-ET project - Part 2: Evaluation of global terrestrial evaporation data sets, Hydrol. Earth Syst. Sci., 20(2), 823–842, doi:doi.org/10.5194/hess-20-823-2016, 2016.

Mishra, V., Kumar, R., Shah, H. L., Samaniego, L., Eisner, S. and Yang, T.: Multimodel assessment of sensitivity and uncertainty of evapotranspiration and a proxy for available water resources under climate change, Clim. Change, 141(3), 451–465, doi:10.1007/s10584-016-1886-8, 2017.

Mockler, E. M., O'Loughlin, F. E. and Bruen, M.: Understanding hydrological flow paths in conceptual catchment models using uncertainty and sensitivity analysis, Comput. Geosci., 90, 66–77, doi:10.1016/j.cageo.2015.08.015, 2015.

Mohan, C., Western, A. W., Wei, Y. and Saft, M.: Predicting groundwater recharge for varying land cover and climate conditions – a global meta-study, Hydrol. Earth Syst. Sci., 22(5), 2689–2703, doi:10.5194/hess-22-2689-2018, 2018.

Monteith, J. L.: Evaporation and environment, Symp. Soc. Exp. Biol., 19, 205–234, 1965.

Monteith, J. L.: Accommodation between transpiring vegetation and the convective boundary layer, J. Hydrol., 166(3–4), 251–263, doi:10.1016/0022-1694(94)05086-D, 1995.

Morris, M. D.: Factorial sampling plans for preliminary computational experiments, Technometrics, 33(2), 161–174, doi:10.2307/1269043, 1991.

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M. and Döll, P.: Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration, Hydrol. Earth Syst. Sci., 18(9), 3511–3538, doi:10.5194/hess-18-3511-2014, 2014.

Mund, M., Kutsch, W. L., Wirth, C., Kahl, T., Knohl, A., Skomarkova, M. V. and Schulze, E. D.: The influence of climate and fructification on the inter-annual variability of stem growth and net primary productivity in an old-growth, mixed beech forest, Tree Physiol., 30(6), 689–704, doi:10.1093/treephys/tpq027, 2010.

Musselman, K. N., Clark, M. P., Liu, C., Ikeda, K. and Rasmussen, R.: Slower snowmelt in a warmer world, Nat. Clim. Chang., 7, 214–219, doi:10.1038/NCLIMATE3225, 2017.

NASA: LDAS Land Data Assimilation Systems - Vegetation Parameters, [online] Available from: https://ldas.gsfc.nasa.gov/gldas/GLDASmapveg.php (Accessed 31 March 2018), 2018.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I - A discussion of principles, J. Hydrol., 10(3), 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.

Neitsch, S., Arnold, J.., Kiniry, J., and Williams, J.: Soil and Water Assessment Tool Theoretical Documentation - Version 2009, Technical Report no 406, College Station, Texas, 2009.

New, M., Lister, D., Hulme, M. and Makin, I.: A high-resolution data set of surface climate over global land areas, Clim. Res., 21(1), 1–25, doi:10.3354/cr021001, 2002.

Noilhan, J. and Planton, S.: A Simple Parameterization of Land Surface Processes for Meteorological Models, Mon. Weather Rev., 117(3), 536–549, doi:10.1175/1520-0493(1989)117<0536:ASPOLS>2.0.CO;2, 1989.

Norton, J.: An introduction to sensitivity assessment of simulation models, Environ. Model. Softw., 69, 166–174, doi:10.1016/j.envsoft.2015.03.020, 2015.

Nossent, J. and Bauwens, W.: 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(3), 539–549, doi:10.2166/wst.2012.884, 2012.

Nossent, J., Elsen, P. and Bauwens, W.: Sobol' sensitivity analysis of a complex environmental model, Environ. Model. Softw., 26(12), 1515–1525, doi:10.1016/j.envsoft.2011.08.010, 2011.

Oke, T. R.: Boundary Layer Climates, 2nd ed., Routledge, London, United Kingdom, 1987.

Oleson, K. W., Lawrence, D. M., Bonan, G. B., Flanner, M. G., Kluzek, E., Lawrence, P. J., Levis, S., Swenson, S. C., Peter, Thornton, E., Dai, A., Decker, M., Dickinson, R., Feddema, J., Heald, C. L., Hoffman, F., Lamarque, J.-F., Mahowald, N., Niu, G.-Y., Qian, T., Randerson, J., Running, S., Sakaguchi, K., Slater, A., Stöckli, R., Wang, A., Yang, Z.-L., Zeng, X. and Zeng, X.: Technical Description of version 4.0 of the Community Land Model (CLM), NCAR/TN-478+STR, NCAR technical note, National Center for Atmospheric Research (NCAR), Boulder, Colorado., 2010.

Oleson, K. W., Lawrence, D. M., Bonan, G. B., Drewniak, B., Huang, M., Koven, C. D., Levis, S., Li, F., Riley, W. J., Subin, Z. M., Swenson, S. C., Thornton, P. E., Bozbiyik, A., Fisher, R., Heald, C. L., Kluzek, E., Jean-Francois Lamarque, P. J. L., Leung, L. R., Lipscomb, W., Muszala, S., Ricciuto, D. M., Sacks, W., Sun, Y., Tang, J. and Yang, Z.-L.: Technical description of version 4.5 of the Community Land Model (CLM), National Center for Atmospheric Research (NCAR), Boulder, Colorado., 2013.

Oliver, J. E.: Monthly precipitation distribution: a comparative index, Mon. Precip. Distrib. a Comp. index, 32(3), 300–309, doi:10.1111/j.0033-0124.1980.00300.x, 1980.

Owor, M., Taylor, R. G., Tindimugaya, C. and Mwesigwa, D.: Rainfall intensity and groundwater recharge: empirical evidence from the Upper Nile Basin, Environ. Res. Lett., 4(3), 35009, doi:10.1088/1748-9326/4/3/035009, 2009.

Papale, D., Migliavacca, M., Cremonese, E., Cescatti, A., Alberti, G., Balzarolo, M., Marchesini, L. B., Canfora, E., Casa, R., Duce, P., Facini, O., Galvagno, M., Genesio, L., Gianelle, D., Magliulo, V., Matteucci, G., Montagnani, L., Petrella, F., Pitacco, A., Seufert, G., Spano, D., Stefani, P., Vaccari, F. P. and Valentini, R.: Carbon, Water and Energy Fluxes of Terrestrial Ecosystems in Italy, in The Greenhouse Gas Balance of Italy, edited by R. Valentini and F. Miglietta, pp. 11–45, Springer, Berlin, Heidelberg., 2015.

Pastor, A. V, Ludwig, F., Biemans, H., Hoff, H. and Kabat, P.: Accounting for environmental flow requirements in global water assessments, Hydrol. Earth Syst. Sci., 18(12), 5041–5059, doi:10.5194/hess-18-5041-2014, 2014.

Pechlivanidis, I. G., McIntyre, N. and Wheater, H. S.: The significance of spatial variability of rainfall on simulated runoff: an evaluation based on the Upper Lee catchment, UK, Hydrol. Res., 48(6), nh2016038, doi:10.2166/nh.2016.038, 2016.

Peel, M. C., McMahon, T. A. and Finlayson, B. L.: Vegetation impact on mean annual evapotranspiration at a global catchment scale, Water Resour. Res., 46(9), W09508, doi:10.1029/2009WR008233, 2010.

Penman, H. L.: The dependance of transpiration on weather and soil conditions, J. Soil Sci., 1(1), 74–89, doi:10.1111/j.1365-2389.1950.tb00720.x, 1950.

Pérez-Priego, O., Serrano-Ortiz, P., Sánchez-Cañete, E. P., Domingo, F. and Kowalski, A. S.: Isolating the effect of subterranean ventilation on CO2 emissions from drylands to the atmosphere, Agric. For. Meteorol., 180, 194–202, doi:10.1016/j.agrformet.2013.06.014, 2013.

Pianosi, F. and Wagener, T.: A simple and efficient method for global sensitivity analysis based on cumulative distribution functions, Environ. Model. Softw., 67, 1–11, doi:10.1016/j.envsoft.2015.01.004, 2015.

Pianosi, F. and Wagener, T.: Understanding the time-varying importance of different uncertainty sources in hydrological modelling using global sensitivity analysis, Hydrol. Process., 30(22), 3991–4003, doi:10.1002/hyp.10968, 2016.

Pianosi, F., Sarrazin, F. and Wagener, T.: A Matlab toolbox for Global Sensitivity Analysis, Environ. Model. Softw., 70, 80–85, doi:10.1016/j.envsoft.2015.04.009, 2015.

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B. and Wagener, T.: Sensitivity analysis of environmental models: A systematic review with practical workflow, Environ. Model. Softw., 79, 214–232, doi:10.1016/j.envsoft.2016.02.008, 2016.

Pianosi, F., Iwema, J., Rosolem, R. and Wagener, T.: A multimethod Global Sensitivity Analysis approach to support the calibration and evaluation of Land Surface Models, in Sensitivity Analysis in Earth Observation Modelling, edited by G. Petropoulos and P. Srivastava, pp. 125–144, Elsevier Inc., 2017.

Pilegaard, K., Hummelshøj, P., Jensen, N. O. and Chen, Z.: Two years of continuous CO2 eddy-flux measurements over a Danish beech forest, Agric. For. Meteorol., 107(1), 29–41, doi:10.1016/S0168-1923(00)00227-6, 2001.

Pinty, B., Jung, M., Kaminski, T., Lavergne, T., Mund, M., Plummer, S., Thomas, E. and Widlowski, J. L.: Evaluation of the JRC-TIP 0.01° products over a mid-latitude deciduous forest site, Remote Sens. Environ., 115(12), 3567–3581, doi:10.1016/j.rse.2011.08.018, 2011.

Porporato, A., Daly, E. and Rodríguez-Iturbe, I.: Soil water balance and ecosystem response to climate change, Am. Nat., 164(5), 625–632, doi:10.1086/424970, 2004.

Press, W., Teukolsky, S., Vetterling, W. and Flannery, B.: Numerical recipes in C. The art of scientific computing, second ed., Cambridge University Press., 1992.

Priestley, C. H. B. and Taylor, R. J.: On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters, Mon. Weather Rev., 100(2), 81–92, doi:10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2, 1972.

Prudhomme, C., Crooks, S., Kay, A. L. and Reynard, N.: Climate change and river flooding: Part 1 classifying the sensitivity of British catchments, Clim. Change, 119(3–4), 933–948, doi:10.1007/s10584-013-0748-x, 2013.

Rahman, M. and Rosolem, R.: Towards a simple representation of chalk hydrology in land surface modelling, Hydrol. Earth Syst. Sci., 21, 459–471, doi:10.5194/hess-21-459-2017, 2017.

Rajib, A. and Merwade, V.: Hydrologic response to future land use change in the Upper Mississippi River Basin by the end of 21st century, Hydrol. Process., 31(21), 3645–3661, doi:10.1002/hyp.11282, 2017.

Rambal, S.: Quercus ilex facing water stress: a functional equilibrium hypothesis, in Quercus ilex L. ecosystems: function, dynamics and management, Advances in vegetation science (AIVS, volume 13), edited by F. Romane and J. Terradas, pp. 147–153, Springer, Dordrecht, The Netherlands., 1992.

Rambal, S.: Le Paradoxe hydrologique des écosystèmes méditerranéens sur des sols karstiques, in Numéro spécial des Annales de la Société d'Horticulture et d'Histoire Naturelle de l'Hérault, pp. 61–67., 2011.

Rambal, S., Ourcival, J. M., Joffre, R., Mouillot, F., Nouvellon, Y., Reichstein, M. and Rocheteau, A.: Drought controls over conductance and assimilation of a Mediterranean evergreen ecosystem: Scaling from leaf to canopy, Glob. Chang. Biol., 9(12), 1813–1824, doi:10.1111/j.1365-2486.2003.00687.x, 2003.

Ray, P. A. and Brown, C. M.: Confronting climate uncertainty in water resources planning and project design: the decision tree framework, World Bank, Washington, DC. [online] Available from: http://hdl.handle.net/10986/22544, 2015.

Raziei, T.: An analysis of daily and monthly precipitation seasonality and regimes in Iran and the associated changes in 1951 – 2014, Theor. Appl. Climatol., 1–22, doi:10.1007/s00704-017-2317-0, 2017.

Reichstein, M., Tenhunen, J. D., Roupsard, O., Ourcival, J. M., Rambal, S., Miglietta, F., Peressotti, A., Pecchiari, M., Tirone, G. and Valentini, R.: Severe drought effects on ecosystem CO2 and H2O fluxes at three Mediterranean evergreen sites: Revision of current hypotheses?, Glob. Chang. Biol., 8(10), 999–1017, doi:10.1046/j.1365-2486.2002.00530.x, 2002.

Reusser, D. E. and Zehe, E.: Inferring model structural deficits by analyzing temporal dynamics of model performance and parameter sensitivity, Water Resour. Res., 47(7), W07550, doi:10.1029/2010WR009946, 2011.

Rimmer, A. and Hartmann, A.: Simplified conceptual structures and analytical solutions for groundwater discharge using reservoir equations, in Water resources management and modeling, edited by P. Nayak, pp. 217–238, InTech, Kakinada, India., 2012.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. . K., Walker, J. P., Lohmann, D. and Toll, D.: The Global Land Data Assimilation System, Bull. Am. Meteorol. Soc., 85(3), 381–394, doi:10.1175/BAMS-85-3-381, 2004.

Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J. T., Beaudoing, H. K., Landerer, F. W. and Lo, M.-H.: Emerging trends in global freshwater availability, Nature, 557, 651–659, doi:10.1038/s41586-018-0123-1, 2018.

Romano, J. P. and Shaikh, A. M.: On the uniform asymptotic validity of subsampling and the bootstrap, Ann. Stat., 40(6), 2798–2822, doi:10.1214/12-AOS1051, 2012.

Rosero, E., Yang, Z. L., Wagener, T., Gulden, L. E., Yatheendradas, S. and Niu, G.-Y.: Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the Noah land surface model over transition zones during the warm season, J. Geophys. Res. Atmos., 115(D3), D03106, doi:10.1029/2009JD012035, 2010.

Rosolem, R., Gupta, H. V., Shuttleworth, W. J., Zeng, X. and De Gonçalves, L. G. G.: A fully multiple-criteria implementation of the Sobol' method for parameter sensitivity analysis, J. Geophys. Res., 117(D7), D07103, doi:10.1029/2011JD016355, 2012.

Rosolem, R., Gupta, H. V., Shuttleworth, W. J., Gonçalves de Gonçalves, L. G. and Zeng, X.: Towards a comprehensive approach to parameter estimation in land surface parameterization schemes, Hydrol. Process., 27(14), 2075–2097, doi:10.1002/hyp.9362, 2013.

Ross, J.: Radiative transfer in plant communities, in Vegetation and the Atmosphere, volume I Principles, edited by J. Monteith, pp. 13–55, Academic Press, London., 1975.

Ruiz, L., Varma, M. R. R., Kumar, M. M. S., Sekhar, M., Maréchal, J.-C., Descloitres, M., Riotte, J., Kumar, S., Kumar, C. and Braun, J.-J.: Water balance modelling in a tropical watershed under deciduous forest (Mule Hole, India): Regolith matric storage buffers the groundwater recharge process, J. Hydrol., 380(3–4), 460–472, doi:10.1016/j.jhydrol.2009.11.020, 2010.

Salinger, M. J. and Griffiths, G. A.: Trends in New Zealand Daily Temperature and Rainfall Extremes, Int. J. Climatol., 21(12), 1437–1452, doi:doi.org/10.1002/joc.694, 2001.

Saltelli, A.: Making best use of model evaluations to compute sensitivity indices, Comput. Phys. Commun., 145(2), 280–297, doi:10.1016/S0010-4655(02)00280-1, 2002.

Saltelli, A. and Tarantola, S.: On the relative importance of input factors in mathematical models: Safety assessment for nuclear waste disposal, J. Am. Stat. Assoc., 97(459), 702–709, doi:10.1198/016214502388618447, 2002.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S.: Global Sensitivity Analysis, The Primer, John Wiley & Sons, Ltd, Chichester, UK, 2008.

Salter, P. J. and Williams, J. B.: The influence of texture on the moisture characteristics of soils- II. Available water capacity and moisture release characteristics, J. Soil Sci., 16(2), 310–317, doi:10.1111/j.1365-2389.1965.tb01442.x, 1965.

Samaniego, L., Kumar, R. and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resour. Res., 46(5), W05523, doi:10.1029/2008WR007327, 2010.

Samaniego, L., Brenner, J., Demirel, C. M., Jing, M., Kaluza, M., Kumar, R., Langenberg, B., Rakovec, O., Schäfer, D., Schrön, M., Schweppe, R. and Thober, S.: The mesoscale Hydrologic Model mHM, Documentation for version 5.9, Helmoltz Centre for Environmental Research (UFZ), Leipzig, Germany., 2018.

Samuels, R., Rimmer, A., Hartmann, A., Krichak, S. and Alpert, P.: Climate Change Impacts on Jordan River Flow: Downscaling Application from a Regional Climate Model, J. Hydrometeorol., 11(4), 860–879, doi:10.1175/2010JHM1177.1, 2010.

Sarrazin, F., Pianosi, F. and Wagener, T.: Global Sensitivity Analysis of environmental models: Convergence and validation, Environ. Model. Softw., 79, 135–152, doi:10.1016/j.envsoft.2016.02.005, 2016.

Sarrazin, F., Pianosi, F. and Wagener, T.: An introduction to the SAFE Matlab Toolbox with practical examples and guidelines, in Sensitivity Analysis in Earth Observation Modelling, edited by G. Petropoulos and P. Srivastava, pp. 363–378, Elsevier Inc., 2017.

Sarrazin, F., Hartmann, A., Pianosi, F. and Wagener, T.: V2Karst v1.0: A parsimonious large-scale integrated vegetation-recharge model to understand the impact of climate and land cover change in karst regions, Geosci. Model Dev. Discuss., in review, doi:10.5194/gmd-2017-315, 2018.

Sauter, M.: Quantification and Forecasting of Regional Groundwater Flow and Transport in a Karst Aquifer (Gallusquelle, Malm, SW. Germany). Ph.D. thesis, Tübinger Universität, Tübinger, Germany., 1992.

Savage, J., Pianosi, F., Bates, P., Freer, J. and Wagener, T.: Quantifying the importance of spatial resolution and other factors through global sensitivity analysis of a flood inundation model, Water Resour. Res., 52(11), 9146–9163, doi:10.1002/2015WR018198, 2016.

Savage, R.: Contributions to the theory of rank order statistics - the two-sample case, Ann. Math. Stat., 26(3), 590–615, doi:10.1214/aoms/1177728170, 1956.

Savenije, H. H. G.: Determination of evaporation from a catchment water balance at a monthly time scale, Hydrol. Earth Syst. Sci., 1(1), 93–100, doi:10.5194/hess-1-93-1997, 1997.

Savenije, H. H. G.: The importance of interception and why we should delete the term evapotranspiration from our vocabulary, Hydrol. Process., 18(8), 1507–1511, doi:10.1002/hyp.5563, 2004.

Scanlon, B. R., Keese, K. E., Flint, A. L., Flint, L. E., Gaye, C. B., Edmunds, W. M. and Simmers, I.: Global synthesis of groundwater recharge in semiarid and arid regions, Hydrol. Prpcesses, 20(15), 3335–3370, doi:doi.org/10.1002/hyp.6335, 2006.

Schenk, H. J. and Jackson, R. B.: The Global Biogeography of Roots, Ecol. Monogr., 72(3), 311–328, doi:10.1890/0012-9615(2002)072[0311:TGBOR]2.0.CO;2, 2002.

Schwinning, S.: The ecohydrology of roots in rocks, Ecohydrology Bearings - Invited Commentary, Ecohydrology, 3(2), 238–245, doi:10.1002/eco.134, 2010.

Seibert, J.: Estimation of Parameter Uncertainty in the HBV Model, Nord. Hydrol., 28(4–5), 247–262, 1997.

Seidl, R., Schelhaas, M.-J., Rammer, W. and Verkerk, P. J.: Increasing forest disturbances in Europe and their impact on carbon storage, Nat. Clim. Chang., 4, 806–810, doi:10.1038/nclimate2393, 2014.

Serrano-Ortiz, P., Kowalski, A. S., Domingo, F., Rey, A., Pegoraro, E., Villagarcía, L. and Alados-Arboledas, L.: Variations in daytime net carbon and water exchange in a montane shrubland ecosystem in southeast Spain, Photosynthetica, 45(1), 30–35, doi:10.1007/s11099-007-0005-5, 2007.

Sheffield, J., Wood, E. F. and Roderick, M. L.: Little change in global drought over the past 60 years., Nature, 491(7424), 435–438, doi:10.1038/nature11575, 2012.

Shin, M. J., Guillaume, J. H. A., Croke, B. F. W. and Jakeman, A. J.: Addressing ten questions about conceptual rainfall-runoff models with global sensitivity analyses in R, J. Hydrol., 503, 135–152, doi:10.1016/j.jhydrol.2013.08.047, 2013.

Shuttleworth, W. J.: Evapotranspiration, in Handbook of Hydrology, edited by D. R. Maidment, p. 4.1-4.53, McGraw-Hill inc., New York., 1993.

Shuttleworth, W. J.: Terrestrial Hydrometeorology, John Wiley & Sons, Ltd, Chichester, UK., 2012.

Shuttleworth, W. J. and Wallace, J. S.: Evaporation From Spare Crops - An Energy Combination Theory, Q. J. R. Meteorol. Soc., 111(469), 839–855, doi:10.1002/qj.49711146910, 1985.

Sieber, A. and Uhlenbrook, S.: Sensitivity analyses of a distributed catchment model to verify the model structure, J. Hydrol., 310(1–4), 216–235, doi:10.1016/j.jhydrol.2005.01.004, 2005.

Siebert, S., Burke, J., Faures, J. M., Frenken, K., Hoogeveen, J., Döll, P. and Portmann, F. T.: Groundwater use for irrigation - A global inventory, Hydrol. Earth Syst. Sci., 14(10), 1863–1880, doi:10.5194/hess-14-1863-2010, 2010.

Simioni, G., Durand-Gillmann, M. and Huc, R.: Asymmetric competition increases leaf inclination effect on light absorption in mixed canopies, Ann. For. Sci., 70(2), 123–131, doi:10.1007/s13595-012-0246-8, 2013.

Šimůnek, J., Šejna, M., Saito, H., Sakai, M. and van Genuchten, M. T.: The HYDRUS-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media, Version 4.08, University of California Riverside, Riverside, USA, 2009.

Sin, G., Gernaey, K. V., Neumann, M. B., van Loosdrecht, M. C. M. and Gujer, W.: Global sensitivity analysis in wastewater treatment plant model applications: Prioritizing sources of uncertainty, Water Res., 45(2), 639–651, doi:10.1016/j.watres.2010.08.025, 2011.

Singh, R., Wagener, T., Van Werkhoven, K., Mann, M. E. and Crane, R.: A trading-space-for-time approach to probabilistic continuous streamflow predictions in a changing climate-accounting for changing watershed behavior, Hydrol. Earth Syst. Sci., 15(11), 3591–3603, doi:10.5194/hess-15-3591-2011, 2011.

Singh, R., Wagener, T., Crane, R., Mann, M. E. and Ning, L.: 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, doi:10.1002/2013WR014988, 2014.

Sitch, S., Smith, B., Prentice, I. C., Arneth, a., Bondeau, a., Cramer, W., Kaplan, J. O., Levis, S., Lucht, W., Sykes, M. T., Thonicke, K. and Venevsky, S.: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model , Glob. Chang. Biol., 9(2), 161–185, doi:10.1046/j.1365-2486.2003.00569.x, 2003.

Smirnov, N.: On the estimation of the discrepancy between empirical curves of distribution for two independent samples, Bull. Mathématique l'Université Moscou, 2, 3–14, 1939.

Smith, K. A.: Investigating Uncertainty in Global Hydrology Modelling, Ph.D. thesis, University of Nottingham, Nottingham, UK, 2016.

Sobol', I. M.: Sensitivity estimates for nonlinear mathematical models, Mat. Model. 2, 112-118 (in Russ. Transl. English (1993). Math. Model. Comput. Exp., 1, 407–414, 1990.

Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M. and Xu, C.: Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications, J. Hydrol., 523, 739–757, doi:10.1016/j.jhydrol.2015.02.013, 2015.

Sorooshian, S., Gupta, V. K. and Fulton, J. L.: Evaluation of Maximum Likelihood Parameter Estimation Techniques for and Length on Model Credibility, Water Resour. Res., 19(1), 251–259, doi:10.1029/WR019i001p00251, 1983.

Spear, R. C. and Hornberger, G. M.: Eutrophication in peel inlet - II. Identification of critical uncertainties via generalized sensitivity analysis, Water Res., 14, 43–49, doi:10.1016/0043-1354(80)90040-8, 1980.

Spearman, C.: The proof and measurement of association between two things, Am. J. Psychol., 15(1), 72–101, doi:10.2307/1412159, 1904.

Sperna Weiland, F. C., Vrugt, J. A., Van Beek, R. L. P. H., Weerts, A. H. and Bierkens, M. F. P.: Significant uncertainty in global scale hydrological modeling from precipitation data errors, J. Hydrol., 529(3), 1095–1115, doi:10.1016/j.jhydrol.2015.08.061, 2015.

Stainforth, D. A., Downing, T. E., Washington, R., Lopez, A. and New, M.: Issues in the interpretation of climate model ensembles to inform decisions, Philos. Trans. R. Soc. A, 365(1857), 2163–2177, doi:10.1098/rsta.2007.2073, 2007.

Stanfill, B., Mielenz, H., Clifford, D. and Thorburn, P.: Simple approach to emulating complex computer models for global sensitivity analysis, Environ. Model. Softw., 74, 140–155, doi:10.1016/j.envsoft.2015.09.011, 2015.

Stevanović, Z.: Global distribution and use of water from karst aquifers, Geol. Soc. London, Spec. Publ., 466, doi:doi.org/10.1144/SP466.17, 2018.

Stewart, J. B.: Modelling surface conductance of pine forest, Agric. For. Meteorol., 43(1), 19–35, doi:10.1016/0168-1923(88)90003-2, 1988.

Sutanto, S. J., Van Den Hurk, B., Dirmeyer, P. A., Seneviratne, S. I., Röckmann, T., Trenberth, K. E., Blyth, E. M., Wenninger, J. and Hoffmann, G.: HESS Opinions "a perspective on isotope versus non-isotope approaches to determine the contribution of transpiration to total evaporation," Hydrol. Earth Syst. Sci., 18(8), 2815–2827, doi:10.5194/hess-18-2815-2014, 2014.

Sutanudjaja, E. H., Van Beek, L. P. H., De Jong, S. M., Van Geer, F. C. and Bierkens, M. F. P.: Large-scale groundwater modeling using global datasets: A test case for the Rhine-Meuse basin, Hydrol. Earth Syst. Sci., 15(9), 2913–2935, doi:10.5194/hess-15-2913-2011, 2011.

Tang, Y., Reed, P., Wagener, T. and Van Werkhoven, K.: Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation, Hydrol. Earth Syst. Sci., 11, 793–817, doi:10.5194/hess-11-793-2007, 2007.

Taylor, R. G., Todd, M. C., Kongola, L., Maurice, L., Nahozya, E., Sanga, H. and MacDonald, A. M.: Evidence of the dependence of groundwater resources on extreme rainfall in East Africa, Nat. Clim. Chang., 3, 374–378, doi:10.1038/nclimate1731, 2013a.

Taylor, R. G., Scanlon, B., Döll, P., Rodell, M., Van Beek, R., Wada, Y., Longuevergne, L., Leblanc, M., Famiglietti, J. S., Edmunds, M., Konikow, L., Green, T. R., Chen, J., Taniguchi, M., Bierkens, M. F. P., Macdonald, A., Fan, Y., Maxwell, R. M., Yechieli, Y., Gurdak, J. J., Allen, D. M., Shamsudduha, M., Hiscock, K., Yeh, P. J. F., Holman, I. and Treidel, H.: Ground water and climate change, Nat. Clim. Chang., 3(4), 322–329, doi:10.1038/nclimate1744, 2013b.

Tesemma, Z. K., Wei, Y., Peel, M. C. and Western, A. W.: The effect of year-to-year variability of leaf area index on Variable Infiltration Capacity model performance and simulation of runoff, Adv. Water Resour., 83, 310–322, doi:10.1016/j.advwatres.2015.07.002, 2015.

Thober, S., Kumar, R., Wanders, N., Marx, A., Pan, M., Rakovec, O., Samaniego, L., Sheffield, J., Wood, E. F. and Zink, M.: Multi-model ensemble projections of European river floods and high flows at 1.5, 2, and 3 degrees global warming, Environ. Res. Lett., 13, 14003, doi:10.1088/1748-9326/aa9e35, 2018.

Thornthwaite, C. W.: An Approach toward a Rational Classification of Climate, Geogr. Rev., 38(1), 55–94, doi:10.2307/210739, 1948.

Tram, V. N. Q., Liem, N. D. and Loi, N. K.: Assessing Water Availability in PoKo Catchment using SWAT model, Khon Kaen Agric., 42(Suppl. 2), 73–84, 2014.

Tritz, S., Guinot, V. and Jourde, H.: Modelling the behaviour of a karst system catchment using non-linear hysteretic conceptual model, J. Hydrol., 397(3–4), 250–262, doi:10.1016/j.jhydrol.2010.12.001, 2011.

Troch, P. A., Carrillo, G., Sivapalan, M., Wagener, T. and Sawicz, K.: Climate-vegetation-soil interactions and long-term hydrologic partitioning: Signatures of catchment co-evolution, Hydrol. Earth Syst. Sci., 17(6), 2209–2217, doi:10.5194/hess-17-2209-2013, 2013.

Twine, T. E., Kustas, W. P., Norman, J. M., Cook, D. R., Houser, P. R., Meyers, T. P., Prueger, J. H., Starks, P. J. and Wesley, M. L.: Correcting eddy covariance flux underestimates over grassland, Agric. For. Meteorol., 103(3), 279–300, doi:10.1016/S0168-1923(00)00123-4, 2000.

Uhlenbrook, S.: Catchment hydrology—a science in which all processes are preferential, Hydrol. Process., 20(16), 3581–3585, doi:10.1002/hyp.6564, 2006.

Ukkola, A. M., Prentice, I. C., Keenan, T. F., Van Dijk, A. I. J. M., Viney, N. R., Myneni, R. B. and Bi, J.: Reduced streamflow in water-stressed climates consistent with CO2 effects on vegetation, Nat. Clim. Chang., 6, 75–78, doi:10.1038/NCLIMATE2831, 2016.

United Nations: World Population Prospects, 2017 Revision, [online] Available from: https://esa.un.org/unpd/wpp/ (Accessed 29 April 2018), 2017.

Vaché, K. B., Eilers, J. M. and Santelmann, M. V.: Water quality modeling of alternative agricultural scenarios in the U.S. Corn Belt, J. Am. Water Resour. Assoc., 38(3), 773–787, doi:10.1111/j.1752-1688.2002.tb00996.x, 2002.

Valente, F., David, J. S. and Gash, J. H. C.: Modelling interception loss for two sparse eucalypt and pine forests in central Portugal using reformulated Rutter and Gash analytical models, J. Hydrol., 190(1–2), 141–162, doi:10.1016/S0022-1694(96)03066-1, 1997.

Van Beek, L. P. H. and Bierkens, M. F. P.: The Global Hydrological Model PCR-GLOBWB: Conceptualization, Parameterization and Verification, Report Department of Physical Geography, Utrecht University, Netherlands. [online] Available from: http://vanbeek.geo.uu.nl/suppinfo/vanbeekbierkens2009.pdf, 2008.

Van Beek, L. P. H., Wada, Y. and Bierkens, M. F. P.: Global monthly water stress: 1. Water balance and water availability, Water Resour. Res., 47(7), W07517, doi:10.1029/2010WR009791, 2011.

Van Beek, R.: Forcing PCR-GLOBWB with CRU data, Utrecht University, the Netherlands. [online] Available from: http://vanbeek.geo.uu.nl/suppinfo/vanbeek2008.pdf, 2008.

Van de Griend, A. A. and Owe, M.: Bare soil surface resistance to evaporation by vapor diffusion under semiarid conditions, Water Resour. Res., 30(2), 181–188, doi:10.1029/93WR02747, 1994.

Van Dijk, A. I. J. M. and Bruijnzeel, L. A.: Modelling rainfall interception by vegetation of variable density using an adapted analytical model. Part 1. Model description, J. Hydrol., 247(3–4), 230–238, doi:10.1016/S0022-1694(01)00392-4, 2001.

Van Dijk, A. I. J. M., Gash, J. H., Van Gorsel, E., Blanken, P. D., Cescatti, A., Emmel, C., Gielen, B., Harman, I. N., Kiely, G., Merbold, L., Montagnani, L., Moors, E., Sottocornola, M., Varlagin, A., Williams, C. A. and Wohlfahrt, G.: Rainfall interception and the coupled surface water and energy balance, Agric. For. Meteorol., 214–215, 402–415, doi:10.1016/j.agrformet.2015.09.006, 2015.

Vanrolleghem, P. A., Mannina, G., Cosenza, A. and Neumann, M. B.: Global sensitivity analysis for urban water quality modelling: Terminology, convergence and comparison of different methods, J. Hydrol., 522, 339–352, doi:10.1016/j.jhydrol.2014.12.056, 2015.

Vanuytrecht, E., Raes, D. and Willems, P.: Global sensitivity analysis of yield output from the water productivity model, Environ. Model. Softw., 51, 323–332, doi:10.1016/j.envsoft.2013.10.017, 2014.

Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J. F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J. and Rose, S. K.: The representative concentration pathways: An overview, Clim. Change, 109, 5–31, doi:10.1007/s10584-011-0148-z, 2011.

Van Vuuren, D. P., Kriegler, E., O'Neill, B. C., Ebi, K. L., Riahi, K., Carter, T. R., Edmonds, J., Hallegatte, S., Kram, T., Mathur, R. and Winkler, H.: A new scenario framework for Climate Change Research: scenario matrix architecture, Clim. Change, 122(3), 373–386, doi:10.1007/s10584-013-0906-1, 2014.

Van Werkhoven, K., Wagener, T., Reed, P. and Tang, Y.: Characterization of watershed model behavior across a hydroclimatic gradient, Water Resour. Res., 44(1), W01429, doi:10.1029/2007WR006271, 2008a.

Van Werkhoven, K., Wagener, T., Reed, P. and Tang, Y.: Rainfall characteristics define the value of streamflow observations for distributed watershed model identification, Geophys. Res. Lett., 35(11), L11403, doi:10.1029/2008GL034162, 2008b.

Vetter, T., Huang, S., Aich, V., Yang, T., Wang, X., Krysanova, V. and Hattermann, F.: Multi-model climate impact assessment and intercomparison for three large-scale river basins on three continents, Earth Syst. Dyn., 6(1), 17–43, doi:10.5194/esd-6-17-2015, 2015.

Vidal, J.-P., Hingray, B., Magand, C., Sauquet, E. and Ducharne, A.: Hierarchy of climate and hydrological uncertainties in transient low-flow projections, Hydrol. Earth Syst. Sci., 20(9), 3651–3672, doi:10.5194/hess-20-3651-2016, 2016.

Vörösmarty, C. J.: Global change , the water cycle, and our search for Mauna Loa, Hydrol. Process., 16(1), 135–139, doi:10.1002/hyp.527, 2002.

Vörösmarty, C. J., Moore, B., Grace, A. L., Gildea, M. P., Melillo, J. M., Peterson, B. J., Rastetter, E. B. and Steudler, P. A.: Continental scale models of water balance and fluvial transport: An application to South America, Global Biogeochem. Cycles, 3(3), 241–265, doi:10.1029/GB003i003p00241, 1989.

Vörösmarty, C. J., Willmott, C. J., Choudhury, B. J., Schloss, A. L., Stearns, T. K., Robeson, S. M. and Dorman, T. J.: Analyzing the discharge regime of a large tropical river through remote sensing , ground-based climatic data , and modeling, , 32(10), 3137–3150, doi:10.1029/96WR01333, 1996.

Vörösmarty, C. J., Federer, C. A. and Schloss, A. L.: Potential evapotranspiration functions compared on US watersheds: implications for global-scale water balance and terrestrial ecosystem modeling, J. Hydrol., 207(3–4), 147–169, doi:10.1016/S0022-1694(98)00109-7, 1998.

Vrugt, J. A., Bouten, W., Gupta, H. V. and Sorooshian, S.: Toward improved identifiability of hydrologic model parameters: The information content of experimental data, Water Resour. Res., 38(12), 48-1-48–13, doi:10.1029/2001WR001118, 2002.

Wada, Y.: Impacts of Groundwater Pumping on Regional and Global Water Resources, in Terrestrial Water Cycle and Climate Change: Natural and Human-Induced Impacts, Geophysical Monograph 221, edited by Q. Tang and T. Oki, pp. 71–101, John Wiley & Sons, Inc., 2016.

Wada, Y., Van Beek, L. P. H. and Bierkens, M. F. P.: Nonsustainable groundwater sustaining irrigation: A global assessment, Water Resour. Res., 48(6), W00L06, doi:10.1029/2011WR010562, 2012.

Wada, Y., Flörke, M., Hanasaki, N., Eisner, S., Fischer, G., Tramberend, S., Satoh, Y., Van Vliet, M. T. H., Yillia, P., Ringler, C., Burek, P. and Wiberg, D.: Modeling global water use for the 21st century : the Water Futures and Solutions (WFaS) initiative and its approaches, Geosci. Model Dev., 9(1), 175–222, doi:10.5194/gmd-9-175-2016, 2016.

Wada, Y., Bierkens, M. F. P., De Roo, A., Dirmeyer, P. A., Famiglietti, J. S., Hanasaki, N., Konar, M., Liu, J., Müller Schmied, H., Oki, T., Pokhre, Y., Sivapalan, M., Troy, T. J., Van Dijk, A. I. J. M., Van Emmerik, T., Van Huijgevoort, Marjolein H. J. Van Lanen, H. A. J., Vörösmarty, C. J., Wanders, N. and Wheater, H.: Human – water interface in hydrological modelling : current status and future directions, Hydrol. Earth Syst. Sci., 21(8), 4169–4193, doi:10.5194/hess-21-4169-2017, 2017.

Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V. and Sorooshian, S.: A framework for development and application of hydrological models, Hydrol. Earth Syst. Sci., 5, 13–26, doi:10.5194/hess-5-13-2001, 2001.

Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S. and Gupta, H. V.: Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis, Hydrol. Process., 17(2), 455–476, doi:10.1002/hyp.1135, 2003.

Wall, J. V.: Practical statistics for astronomers - II. Correlation, data-modelling and sample comparison, Q. J. R. Astron. Soc., 37, 519–563, 1996.

Walsh, R. P. D. and Lawler, D. M.: Rainfall seasonality: description, spatial patterns and change through time, Weather, 36(7), 201–208, doi:10.1002/j.1477-8696.1981.tb05400.x, 1981.

Wang, J., Li, X., Lu, L. and Fang, F.: Parameter sensitivity analysis of crop growth models based on the extended Fourier Amplitude Sensitivity Test method, Environ. Model. Softw., 48, 171–182, doi:10.1016/j.envsoft.2013.06.007, 2013.

Wang, K. and Dickinson, R. E.: A review of global terrestrial evapotranspiration: observation, modelling, climatology, and climatic variability, Rev. Geophys., 50(2), 1–54, doi:10.1029/2011RG000373, 2012.

Wang, S. J., Liu, Q. M. and Zhang, D. F.: Karst rocky desertification in southwestern China: Geomorphology, landuse, impact and rehabilitation, L. Degrad. Dev., 15(2), 115–121, doi:10.1002/ldr.592, 2004.

Wang-Erlandsson, L., Bastiaanssen, W. G. M., Gao, H., Jägermeyr, J., Senay, G. B., Van Dijk, A. I. J. M., Guerschman, J. P., Keys, P. W., Gordon, L. J. and Savenije, H. H. G.: Global root zone storage capacity from satellite-based evaporation, Hydrol. Earth Syst. Sci., 20(4), 1459–1481, doi:10.5194/hess-20-1459-2016, 2016.

Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O. and Schewe, J.: The Inter-Sectoral Impact Model Intercomparison Project (ISI–MIP): Project framework, Proc. Natl. Acad. Sci., 111(9), 3228–3232, doi:10.1073/pnas.1312330110, 2014.

Weedon, G. P., Gomes, S., Viterbo, P., Österle, H., Adam, J. C., Bellouin, N., Boucher, O. and Best, M.: The Watch Forcing Data 1958-2001: a Meteorological Forcing Dataset for Land Surface- and Hydrological-Models. Technical Report 22., 2010.

Wei, H., Xia, Y., Mitchell, K. E. and Ek, M. B.: Improvement of the Noah land surface model for warm season processes: Evaluation of water and energy flux simulation, Hydrol. Process., 27(2), 297–303, doi:10.1002/hyp.9214, 2013.

Weiler, M. and McDonnell, J.: Virtual experiments: A new approach for improving process conceptualization in hillslope hydrology, J. Hydrol., 285(1–4), 3–18, doi:10.1016/S0022-1694(03)00271-3, 2004.

Werth, S., Güntner, A., Petrovic, S. and Schmidt, R.: Integration of GRACE mass variations into a global hydrological model, Earth Planet. Sci. Lett., 277(1–2), 166–173, doi:10.1016/j.epsl.2008.10.021, 2009.

White, J., Stengel, V., Rendon, S. and Banta, J.: The importance of parameterization when simulating the hydrologic response of vegetative land-cover change, Hydrol. Earth Syst. Sci., 21(8), 3975–3989, doi:10.5194/hess-21-3975-2017, 2017.

Wigmosta, M. S., Vail, L. W. and Lettenmaier, D. P.: A distributed hydrology-vegetation model for complex terrain, Water Resour. Res., 30(6), 1665–1679, doi:10.1029/94WR00436, 1994.

Wilby, R. L. and Dessai, S.: Robust adaptation to climate change, Weather, 65(7), 180–185, doi:0.1002/wea.543, 2010.

Williams, P. W.: The role of the subcutaneous zone in karst hydrology, J. Hydrol., 61(1–3), 45–67, doi:10.1016/0022-1694(83)90234-2, 1983.

Williams, P. W.: Environmental change and human impact on karst terrains: an introduction, in Karst Terrains - Environmental changes and human impact, edited by P. W. Williams, pp. 1–19, Catena Supplement 25, Catena Verlag, Cremlingen-Destedt, Germany., 1993.

Williams, P. W.: The role of the epikarst in karst and cave hydrogeology : a review, Int. J. Speleol., 37, 1–10, doi:10.5038/1827-806X.37.1.1, 2008.

Williams, P. W. and Ford, D. C.: Global distribution of carbonate rocks, Zeitschrift für Geomorphol. Suppl., 147, 1–2, 2006.

Wohlfahrt, G., Pilloni, S., Hörtnagl, L. and Hammerle, a.: Estimating carbon dioxide fluxes from temperate mountain grasslands using broad-band vegetation indices, Biogeosciences, 7(2), 683–694, doi:10.5194/bg-7-683-2010, 2010.

Xiong, Y. J., Qiu, G. Y., Mo, D. K., Lin, H., Sun, H., Wang, Q. X., Zhao, S. H. and Yin, J.: Rocky desertification and its causes in karst areas: A case study in Yongshun County, Hunan Province, China, Environ. Geol., 57(7), 1481–1488, doi:10.1007/s00254-008-1425-7, 2009.

Xue, J. and Su, B.: Significant remote sensing vegetation indices: A review of developments and applications, J. Sensors, 2017, Article ID 1353691, doi:10.1155/2017/1353691, 2017.

Yang, J.: Convergence and uncertainty analyses in Monte-Carlo based sensitivity analysis, Environ. Model. Softw., 26(4), 444–457, doi:10.1016/j.envsoft.2010.10.007, 2011.

Yang, Z.-L., Cai, X., Zhang, G., Tavakoly, A. A., Jin, Q., Meyer, L. H. and Guan, X.: The Community Noah Land Surface Model with Multi-Parameterization Options (Noah-MP) - Technical description, University of Texas at Austin, Austin, USA. [online] Available from: http://www.jsg.utexas.edu/noah-mp/files/noah-mp_technote_v0.2.pdf, 2011.

Yin, J., Porporato, A. and Albertson, J.: Interplay of climate seasonality and soil moisture-rainfall feedback, Water Resour. Res., 50(7), 6053–6066, doi:10.1002/2013WR014772, 2014.

Young, P. C., Spear, R. C. and Hornberger, G. M.: Modelling badly defined systems: some further thoughts, in Proceedings of SIMSIG Simulation Conference, pp. 24–32, Australian National University, Canberra., 1978.

Young, P., Parkinson, S. and Lees, M.: Simplicity out of complexity in environmental modelling: Occam's razor revisited, J. Appl. Stat., 23(2–3), 165–210, doi:10.1080/02664769624206, 1996.

Zakhem, A. B. and Kattaa, B.: Overexploitation and cumulative drought trend effect on Ras El Ain karstic spring discharge (Khabour Sub-basin, Syria), J. Earth Syst. Sci., 126(7), 93, doi:10.1007/s12040-017-0882-3, 2017.

Zhang, L., Dawes, W. R. and Walker, G. R.: Response of mean annual evapotranspiration to vegetation changes at catchment scale, Water Resour. Res., 37(3), 701–708, doi:10.1029/2000WR900325, 2001.

Zhang, Z., Chen, X., Ghadouani, A. and Shi, P.: Modelling hydrological processes influenced by soil, rock and vegetation in a small karst basin of southwest China, Hydrol. Process., 25(15), 2456–2470, doi:10.1002/hyp.8022, 2011.

Zhu, Z., Piao, S., Myneni, R. B., Huang, M., Zeng, Z., Canadell, J. G., Ciais, P., Sitch, S., Friedlingstein, P., Arneth, A., Cao, C., Cheng, L., Kato, E., Koven, C., Li, Y., Lian, X., Liu, Y., Liu, R., Mao, J., Pan, Y., Peng, S., Peñuelas, J., Poulter, B., Pugh, T. A. M., Stocker, B. D., Viovy, N., Wang, X., Wang, Y., Xiao, Z., Yang, H., Zaehle, S. and Zeng, N.: Greening of the Earth and its drivers, Nat. Clim. Chang., 6, 791–795, doi:10.1038/nclimate3004, 2016.

Zwahlen, F.: Vulnerability and risk mapping for the protection of carbonate (karst) aquifers, final report (COST action 620). European Commission for Science, Directorate-General XII, vol 297, Luxembourg, 2003.