University of Potsdam
Institute of Environmental Science and Geography

Pursuing the understanding of uncertainties in hydrological modelling

Cumulative dissertation
for the degree of "doctor rerum naturalium" (Dr. rer. nat.)
in Hydrology

submitted to the
Faculty of Science
at the University of Potsdam, Germany

by
Tobias Pilz

Date of Disputation: 15 September 2020

Pursuing the understanding of uncertainties in hydrological modelling

Dissertation submitted to the Faculty of Science at the University of Potsdam, Germany, for the degree of "doctor rerum naturalium" (Dr. rer. nat.) in Hydrology

Potsdam, November 2019

Author: Tobias Pilz
University of Potsdam, Institute of Environmental Science and Geography, Potsdam, Germany
tpilz@uni-potsdam.de

1st Referee: (Supervisor) Prof. Dr. Axel Bronstert
University of Potsdam, Institute of Environmental Science and Geography, Potsdam, Germany

2nd Referee: Prof. Dr. Thorsten Wagener
University of Bristol, Department of Civil Engineering, Bristol, UK

3rd Referee: Prof. Dr. Niels Schütze
Technische Universität Dresden, Chair of Hydrology, Dresden, Germany

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-47664
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-476643

“The last one was a bitch. This one was three bitches and a bastard.” — George R. R. Martin in *A Dance with Dragons*

“I love deadlines. I love the whooshing noise they make as they go by.” — Douglas Adams in *The Salmon of Doubt*

Acknowledgements

This thesis is the result of a long journey through the vast universe of hydrological models. As this is a topic of enormous gravity, dilatation of time was a major issue along the journey. The final completion of my dissertation would not have been possible without the help and encouragement of a number of people, who deserve my special gratitude.

First I want to thank my supervisor Axel Bronstert for the opportunity to do my PhD in his working group, the freedom to realise my ideas, and for his continuous support in achieving my scientific goals. Special thanks are sent to Till Francke, who served as a mentor during my PhD period, providing me with inspiring ideas, technical and scientific support, always having time for my questions, and keeping me on track until the end.

I want to further thank all Hyklis, my colleagues from the working group of Hydrology and Climatology of the University of Potsdam, for the relaxed working atmosphere, inspiring discussions, and never boring chats during the lunch break.

Thank you to Gabriele Baroni for his invaluable comments and patiently answering all my questions about sensitivity analysis in general and, specifically, the great support for the last manuscript (Chapter 4). In this regard I also want to thank Thorsten Wagener for his inspiring suggestions that enabled, after several setbacks, to finally make some sense out of my model runs for Chapter 4.

During my PhD I spent some months in Brazil at the Research Institute for Meteorology and Water Resources of the state of Ceará (FUNCEME). In this regard I want to thank Eduardo Martins for the invitation and his support, José Carlos de Araújo, Alexandre Cunha Costa, and Pedro Medeiros for all the invaluable hints and suggestions for surviving and having a great time in Fortaleza and Brazil, and all the colleagues from FUNCEME for the friendly working atmosphere and fruitful discussions.

Thanks to my friends and family for their mental support and letting me forget all frustrations that working on a PhD sometimes bears.

Last but not least many thanks to Aline for all the scientific and mental (and financial) support, for exploring the world together, and for the whooshing noise *before* deadlines go by.

Contents

Summary

Hydrological models are important tools for the simulation and quantification of the water cycle. They therefore aid in the understanding of hydrological processes, prediction of river discharge, assessment of the impacts of land use and climate changes, or the management of water resources. However, uncertainties associated with hydrological modelling are still large. While significant research has been done on the quantification and reduction of uncertainties, there are still fields which have gained little attention so far, such as model structural uncertainties that are related to the process implementations in the models. This holds especially true for complex process-based models in contrast to simpler conceptual models. Consequently, the aim of this thesis is to improve the understanding of structural uncertainties with focus on process-based hydrological modelling, including methods for their quantification.

To identify common deficits of frequently used hydrological models and develop further strategies on how to reduce them, a survey among modellers was conducted. It was found that there is a certain degree of subjectivity in the perception of modellers, for instance with respect to the distinction of hydrological models into conceptual groups. It was further found that there are ambiguities on how to apply a certain hydrological model, for instance how many parameters should be calibrated, together with a large diversity of opinion regarding the deficits of models. Nevertheless, evapotranspiration processes are often represented in a more physically based manner, while processes of groundwater and soil water movement are often simplified, which many survey participants saw as a drawback. A large flexibility, for instance with respect to different alternative process implementations or a small number of parameters that needs to be calibrated, was generally seen as strength of a model.

Flexible and efficient software, which is straightforward to apply, has been increasingly acknowledged by the hydrological community. This work further elaborated on this topic in a twofold way. First, a software package for semi-automated landscape discretisation has been developed, which serves as a tool for model initialisation. This was complemented by a sensitivity analysis of important and commonly used discretisation parameters, of which the size of hydrological sub-catchments as well as the size and number of hydrologically uniform computational units appeared to be more influential than information considered for the characterisation of hillslope profiles. Second, a process-based hydrological model has been implemented into a flexible simulation environment with several alternative process representations and a number of numerical solvers. It turned out that, even though computation times were still long, enhanced computational capabilities nowadays in combination with

innovative methods for statistical analysis allow for the exploration of structural uncertainties of even complex process-based models, which up to now was often neglected by the modelling community.

In a further study it could be shown that process-based models may even be employed as tools for seasonal operational forecasting. In contrast to statistical models, which are faster to initialise and to apply, process-based models produce more information in addition to the target variable, even at finer spatial and temporal scales, and provide more insights into process behaviour and catchment functioning. However, the process-based model was much more dependent on reliable rainfall forecasts.

It seems unlikely that there exists a single best formulation for hydrological processes, even for a specific catchment. This supports the use of flexible model environments with alternative process representations instead of a single model structure. However, correlation and compensation effects between process formulations, their parametrisation, and other aspects such as numerical solver and model resolution, may lead to surprising results and potentially misleading conclusions. In future studies, such effects should be more explicitly addressed and quantified. Moreover, model functioning appeared to be highly dependent on the meteorological conditions and rainfall input generally was the most important source of uncertainty. It is still unclear, how this could be addressed, especially in the light of the aforementioned correlations. The use of innovative data products, e.g. remote sensing data in combination with station measurements, and efficient processing methods for the improvement of rainfall input and explicit consideration of associated uncertainties is advisable to bring more insights and make hydrological simulations and predictions more reliable.

Zusammenfassung

Hydrologische Modelle sind wichtige Werkzeuge zur Simulation und Quantifizierung des Wasserkreislaufs. Sie helfen bei der explorativen Analyse hydrologischer Prozesse, Abflussvorhersage, Abschätzung der Folgen von Klima- und Landnutzungswandel oder dem Management von Wasserressourcen. Allerdings sind die mit der hydrologischen Modellierung einhergehenden Unsicherheiten noch immer groß. Trotz der zahlreichen Forschungsarbeiten auf dem Gebiet der Quantifizierung und Reduktion der Unsicherheiten gibt es einige Bereiche, die bisher wenig erforscht wurden, wie beispielsweise strukturelle Unsicherheiten, welche sich unter anderem auf die Prozessimplementation in den Modellen beziehen. Dies betrifft vor allem komplexe prozessbasierte Modelle im Gegensatz zu einfacheren konzeptionellen Modellen. Gegenstand dieser Arbeit ist es daher, das Verständnis struktureller Unsicherheiten sowie Methoden für deren Quantifizierung innerhalb prozessbasierter hydrologischer Modellanwendungen zu erweitern.

Zur Identifikation typischer Defizite hydrologischer Modelle und Erarbeitung von Lösungsstrategien, um diese zu reduzieren, wurde eine Umfrage unter Modellanwendern durchgeführt. Dabei stellte sich heraus, dass ein hohes Maß an Subjektivität in der Wahrnehmung des Themas unter Modellieren herrscht, beispielsweise bei der Einordnung hydrologischer Modelle in konzeptionelle Klassen. Des Weiteren gibt es Unklarheiten in der Art und Weise, wie ein bestimmtes hydrologisches Modell angewendet werden sollte, wie etwa hinsichtlich der Kalibrierung bestimmter Parameter, sowie vielschichtige Auffassungen bezüglich der Modelldefizite. Letztlich stellte sich jedoch heraus, dass Verdunstungsprozesse vor allem physikalisch basiert abgebildet werden, während Prozesse im Bereich des Grundwassers und der Bodenwasserbewegung häufig vereinfacht abgebildet werden, was von vielen Umfrageteilnehmern als Nachteil empfunden wurde. Generell als Stärke wurde die Flexibilität einiger Modelle empfunden, zum Beispiel wenn diese verschiedene Implementationen eines Prozesses enthalten oder wenn nur eine geringe Zahl an Parametern kalibriert werden muss.

Flexible und effiziente Software, die darüber hinaus einfach zu bedienen ist, wird von der hydrologischen Gemeinschaft immer stärker in den Vordergrund gebracht. Daher greift diese Arbeit das Thema in zweifacher Hinsicht auf. Zum einen wurde ein Softwarepaket zur halbautomatischen Landschaftsdiskretisierung entwickelt, welches zudem als Werkzeug zur Modellinitialisierung gedacht ist. Damit einhergehend wurde eine Sensitivitätsanalyse wichtiger und häufig genutzter Diskretisierungsparameter durchgeführt, bei der die Größe hydrologischer Teileinzugsgebiete sowie die Anzahl und Größe hydrologischer Elementarflächen sich als maßgeblicher herausstellte als etwa raumbezogene Informationen zur Charakterisierung der

Hangprofile. Zum anderen wurde ein prozessbasiertes hydrologisches Modell in eine flexible Softwareumgebung integriert, der verschiedene alternative Prozessformulierungen sowie numerische Differentialgleichungslöser hinzugefügt wurden. Die Analyse struktureller Unsicherheiten komplexer prozessbasierter Modelle wurde in der Vergangenheit von der hydrologischen Gemeinschaft mit Verweis auf zu lange Rechenzeit oft vernachlässigt. Es zeigte sich jedoch, dass die mittlerweile zur Verfügung stehenden Computerressourcen, vor allem in Kombination mit innovativen statistischen Analyseverfahren, derartige Untersuchungen bereits ermöglichen.

In einer weiteren Studie konnte zudem gezeigt werden, dass auch prozessbasierte Modelle für den operationellen Einsatz in der saisonalen Vorhersage geeignet sind. Im Gegensatz zu statistischen Modellen, welche schneller initialisierbar und anwendbar sind, produzieren prozessbasierte Modelle neben der eigentlichen Zielgröße weitere potentiell relevante Informationen in höherer räumlicher und zeitlicher Auflösung und geben zudem tiefere Einblicke in die generelle Wirkungsweise der hydrologischen Prozesse in einem Einzugsgebiet. In der Studie stellte sich jedoch ebenso heraus, dass zuverlässige Niederschlagsvorhersagen für ein prozessbasiertes Modell umso wichtiger sind.

Allgemein erscheint es unwahrscheinlich, dass eine einzelne optimale Implementation für einen hydrologischen Prozess, selbst innerhalb eines bestimmten Einzugsgebietes, überhaupt existiert. Die Nutzung flexibler Modellumgebungen mit alternativen Prozessbeschreibungen anstelle eines einzelnen Modells scheint deshalb große Vorteile zu bringen. Mögliche Korrelationen zwischen Prozessbeschreibungen, deren Parametrisierung, sowie anderen Aspekte wie numerischen Lösern und Modellauflösung, können jedoch zu überraschenden Ergebnissen und letztlich falschen Schlussfolgerungen führen. In zukünftigen Studien sollten solche Effekte daher explizit berücksichtigt und quantifiziert werden. Darüber hinaus wird die Leistungsfähigkeit eines Modells maßgeblich von den meteorologischen Randbedingungen beeinflusst. Vor allem der Niederschlag erwies sich innerhalb dieser Arbeit als wichtigste Ursache für Unsicherheiten in der Modellierung. Allerdings ist nicht vollständig klar, wie dieser Umstand berücksichtigt werden kann und inwiefern die zuvor genannten Korrelationen hier einen Einfluss haben. Die Nutzung innovativer Datenprodukte, zum Beispiel Fernerkundungsdaten verbunden mit Stationsmessungen, in Kombination mit effizienten Prozessierungsalgorithmen zur Verbesserung des Niederschlagsinputs und expliziten Beachtung einhergehender Unsicherheiten wird angeraten. Dies verspricht bessere Einblicke in die Zusammenhänge verschiedener Unsicherheitsquellen zu gewinnen und letztlich hydrologische Simulationen und Vorhersagen zuverlässiger zu machen.

1. Introduction

1.1 Background

1.1.1 About hydrological modelling

Water is *a colourless, transparent, odourless liquid that forms the seas, lakes, rivers, and rain and is the basis of the fluids of living organisms* (Oxford dictionaries, 2018). The movement of water above and below the surface of Earth is commonly referred to as the *hydrological cycle*. This cycle involves different physical processes, such as evaporation, precipitation, runoff, and subsurface movement, describing the transport of water between different reservoirs, including oceans, rivers and lakes, soil, and the atmosphere. Humans as living organisms as well as their economy and infrastructure depend on water. However, water on Earth is unequally distributed and this distribution is not necessarily static in time. Therefore, ever since the beginning of civilisation, humans tried to control the supply of water. The construction of canals and reservoirs, for instance, serve the provision and storage of water during dry periods, when little or no rainfall occurs, whereas the alteration of natural rivers and the construction of dams and weirs provide controlled discharge of water and protection against floods during times of surplus. Nevertheless, hydrological extreme events such as droughts and floods still have the power to provoke damages on goods, threaten our economy, or even cause fatalities. Moreover, climate change is likely to aggravate such extreme events in many regions of the world (Jiménez Cisneros et al., 2014).

The rapid development of powerful computer technology over the last about 70 years enabled the evolution of *hydrological computer models*. Models are simplified and abstract mappings of reality into machine code. As such, hydrological models include process formulations aiming at the simulation of the hydrological cycle. Typically they serve the estimation of a discharge hydrograph in a river following a rainfall event, but often also provide information about the various components of the hydrological cycle. Today, they are invaluable tools for the exploration of hydrological processes, the assessment of climate and land use change impacts on the water cycle, the prediction of hydrological extremes, and the management of water resources.

Landscapes, however, are unique in their configuration of environmental attributes, such as soil types, vegetation cover, geology, or meteorological conditions. Even though the physical principles to describe the transport of water and energy at the microscale are known, at least under static and homogeneous boundary conditions, there is still a lack of understanding regarding the upscaling of such principles to the catchment scale. This includes, in particular, the mathematical formulation of interactions of system variables and boundary conditions within extremely heterogeneous landscape configurations, which are nonlinear, hysteretic, and scale-dependent (commonly referred to as *closure problem*) (Beven, 2006b).

To handle this upscaling problem, hydrological modelling approaches typically include empirically based relationships. However, a unified theory of catchment hydrology is still missing. Consequently, a large number of different hydrological models evolved varying in complexity and conceptualisation (Weiler and Beven, 2015). To decide on a specific model is therefore a difficult and often subjective task. Yet several formalised strategies exist, which are commonly based on the principle of parsimony seeking for the best compromise between model fit to observations and a minimum of complexity (Höge et al., 2018).

Physically based models incorporate explicit representations of physical processes. Thus, they ideally describe the flow of water within the hydrological cycle in a mechanistic manner following physical principles while the equations satisfy the laws of conservation of mass, energy, and momentum. This modelling philosophy has been pursued for a long time in hydrology (Abbott et al., 1986; Freeze and Harlan, 1969; Stephenson and Freeze, 1974; Woolhiser et al., 1990). It comes with a number of advantages: the equations are related to physical processes and as such the parameters are in principle directly measurable; the latter provides the benefit that, at least in theory, such models can be readily applied to different regions with a priori determined parameters and without calibration; quantities of interest, i.e. water fluxes such as runoff, can be inferred at any point in space and time along the model domain.

In practise, however, this approach became more and more debated due to a number of important drawbacks and limitations (Beven, 1989; Grayson et al., 1992b; Woolhiser, 1996): physically based models are often slow and computationally demanding; their parametrisation requires a large amount of data which is often lacking in availability and/or quality; the closure problem may lead to large errors in the scaling of water fluxes to the scale of interest. Consequently, also physically based models in practise rely on empirical simplifications and require the calibration of certain parameters, which is why they are often also referred to as *process-based* models instead of the rather imprecise and potentially misleading term "physically based".

In contrast to process-based approaches, *conceptual models* rely on empirical relationships to transform a precipitation signal into a discharge hydrograph. Probably the most straightforward and purely empirical method is the unit hydrograph (Dooge, 1959; Sherman, 1932). However, such straightforward empirical approaches are typically designed to describe a single rainfall–runoff event and are limited in the considered aspects; baseflow, for instance, is often neglected. Therefore, for the continuous simulation of discharge dynamics, more advanced approaches are needed, which contain some degree of process mapping.

Conceptual models typically comprise theoretical reservoirs (also known as *buckets*) reflecting, for instance, groundwater and surface water storages, which are filled by rainfall input and emptied following (non-)linear storage equations. The number of storages, included subprocesses such as evapotranspiration, and storage equations reflecting the concentration of flow in a river can vary greatly among models. Furthermore, the models can be applied in a spatially lumped manner, i.e. describing a watershed as a single entity, or spatially distributed, simulating the hydrological process dynamics independently in single sub-catchments and merging the outflows.

Conceptual models contain a number of parameters reflecting the environmental conditions of the landscape to be modelled. They need to be calibrated on observational data or related to observable quantities via statistical relationships. Such kind of models are fast and straightforward to apply and are therefore suitable for the simulation of (large) catchments with poor data availability. Furthermore, they are typically employed to generate operational predictions in real-time, such as discharge forecasts over time horizons of a few hours or days.

As a result, many such models have emerged over time (e.g. Bergström, 1992; Fenicia et al., 2011; Jakeman and Hornberger, 1993; Perrin et al., 2003). Even though they are much simpler, they are often able to achieve performances similar to or even better than process-based models, especially in case of limited data availability or restricted computational resources (Kokkonen and Jakeman, 2001; Refsgaard and Knudsen, 1996). They may even serve for an advanced process understanding in a study area, e.g. by testing alternative landscape discretisations, incorporating different kind of information, or using different model structures, which reflect hydrological processes in varying degree of detail (Fenicia et al., 2016).

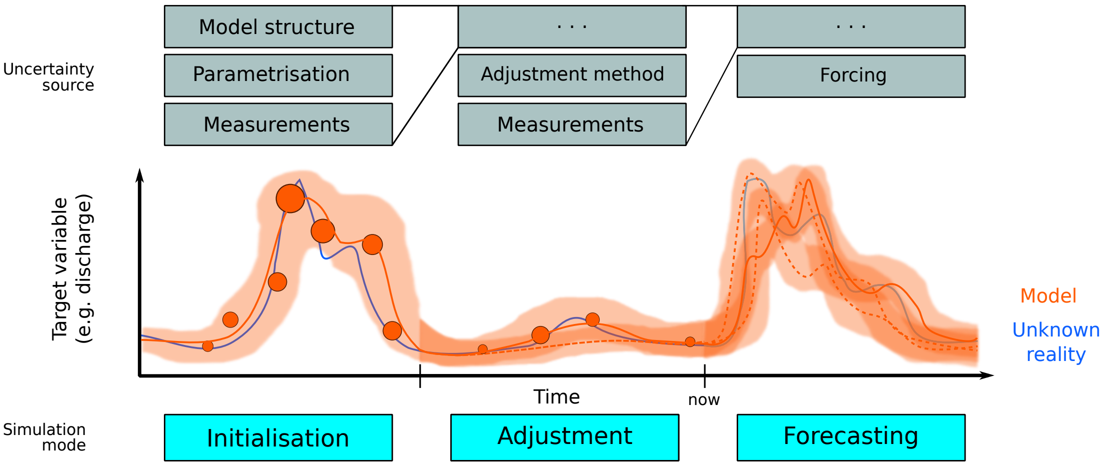

Figure 1.1: Illustration of the influence and propagation of different sources of uncertainty in a typical hydrological model application.

Eventually it has to be noted that there is no strict separation between the model conceptions. Many models are process-based in a way that they explicitly represent physical processes, whereas the mathematical formulations rely on empirical relationships. Moreover, in the last years, growing efforts on the syntheses of existing modelling approaches could be noted (Clark et al., 2017; Fatichi et al., 2016; Hrachowitz and Clark, 2017) and there are even discussions regarding the development of a community model (Weiler and Beven, 2015).

1.1.2 Uncertainties in hydrological simulations

Models as abstract simplifications of the real world are necessarily imperfect and, consequently, model-based simulations are accompanied by a number of uncertainties. The large variety of different model concepts and individual models in hydrology induces a large diversity of uncertainties, which can be specific to individual models, model conceptions, or even individual applications. Nevertheless, the problem can be simplified into three basic types of uncertainties in hydrological modelling. They are related to model structure, model parametrisation, and imperfect data (Liu and Gupta, 2007; Wagener and Gupta, 2005). The influence and propagation of such uncertainties along a typical application of hydrological simulation and prediction are illustrated in Fig. 1.1. In the following, these fundamental types of uncertainty will be briefly described while in Sect. 1.1.3 some approaches for dealing with these uncertainties will be introduced.

Structural uncertainties are related to the conception of a model, the incorporated mathematical equations, and the translation of the latter into computer code. This starts already with the perceptual model as a basis of every hydrological computer model, reflecting the perception of the real-world system and the understanding of incorporated processes (Beven, 2009). Such a perceptual model is typically based on evidence from experimental data and experience of the model developer (Wrede et al., 2015). Therefore, it is subject to a certain degree of subjectivity and uncertainties may arise from imperfect process understanding or misconception. What follows the perceptual model is the mathematical representation. This involves the definition of state variables, boundary conditions, input and output variables, flux equations, and parameters.

Apart from a possible lack of understanding, modellers are confronted with the problem that for a certain process often various alternative equations exist, varying in the degree of detail and physical realism. The selection of an equation is therefore associated with a high degree of subjectivity and is usually problem-specific. For instance, to represent evapotranspiration processes in an agricultural region in Germany, the empirical Haude formula (Haude, 1958) might be sufficient, whereas for a model to be applied in different regions of the world, a more general or even physically based approach should be used, but also requires much more data.

The mathematical equations then need to be translated into computer code. This step furthermore involves decisions regarding the spatial and temporal resolution of the model and how to address the closure problem, i.e. how to integrate the process equations both in space and time (Gupta et al., 2012). Besides, the skeletal structure of the model needs to be designed and programmed, including the reading and possibly transformation of input and writing of output information, definition of data structures and formats, construction of a user interface, etc. Structural uncertainty may therefore arise not just from a lack of understanding, misconception, or the degree of abstraction of the real world, but also from a certain degree of subjectivity and ambiguity, numerical imprecision (e.g. when solving differential equations), and random errors (e.g. during programming).

Measurement uncertainty refers to imperfect observations. Errors in the measurements can arise randomly, e.g. by failure of a device or errors in the recording or transcription of data, and can be systematic, which is, for example, associated with the imprecision or limited resolution of a measurement device. In the context of hydrological modelling, especially uncertain discharge data have a large impact on model results as discharge typically is the target variable of hydrological simulations and is furthermore frequently used for model calibration (e.g. McMillan and Westerberg, 2015; Ocio et al., 2017).

Rainfall observations as main forcing of hydrological models pose another important source of measurement uncertainty. In this case, a large portion of uncertainty in particular arises from the interpolation of point measurements to the (sub-)catchment scale (McMillan et al., 2012). Quantitative precipitation estimates based on remote sensing (such as RADAR data) try to tackle that problem but come along with further uncertainties, such as artefacts, signal attenuation with increasing distance to the station, etc., and are not yet necessarily superior to gauge-based interpolations (Abon et al., 2016). Measurement uncertainties may have far-reaching impacts on hydrological simulations as they can influence our system understanding, deteriorate model initialisation and adjustment, or interfere with model evaluation and, consequently, may lead to wrong conclusions (Westerberg and Birkel, 2015). As such, rainfall uncertainty impacts all simulation modes outlined in Fig. 1.1, including the forecast mode, where it is the principle driving force of hydrological models (together with other meteorological variables) and needs to be adequately predicted before starting the hydrological forecasting.

Parametric uncertainty generally refers to the estimation of realisations of model parameters for a specific application. The estimation procedure typically focusses on *effective* parameters which are aggregated *"representations of spatially and temporally heterogeneous properties of the real system"* (Liu and Gupta, 2007). Parametrisations can essentially be determined in two ways: measurement and calibration.

Measurement in this case refers to parameter estimation by direct derivation from field work or indirect determination such as from remote sensing data. Such measurements are important to characterise the physical properties of a landscape and therefore includes the effective description of the spatial distribution of landscape attributes, such as soil types, vegetation, geology, or the position of rivers and lakes. This further involves the parametrisation of such landscape attributes, e.g. bulk density and hydraulic conductivity of a certain soil type, stomatal resistance and albedo of the vegetation types, or the width and shape of a river section.

For an effective landscape representation, a model needs to be spatially discretised into entities, which can comprise either the whole catchment (in lumped discretisation approaches), smaller sub-catchments, hydrological uniform areas (referred to as semi-distributed modelling), or even regular raster cells (fully distributed approach). Therefore, in a broader sense, the discretisation of a landscape into model entities, which typically involves not only the acquisition but also the transformation and analysis of data, can be attributed to model parametrisation and belongs to the initialisation mode of a model application (Fig. 1.1). Uncertainties may arise from imperfect measurements and the applied algorithms of data manipulation and transfer from the scale of measurement to the model scale (Blöschl and Sivapalan, 1995).

Parametrisation by *calibration* follows model initialisation and is referred to as *adjustment* in Fig. 1.1. It is done by adjusting the parameters in a way that the model output matches observations as best as possible. This step includes a large degree of ambiguity and subjectivity, and suffers from the lack of a generally applicable methodology. It typically requires a number of decisions including the parameters of the model that shall be calibrated (for instance, because they cannot be measured or measurements are very uncertain), the definition of feasible value ranges for each parameter, the target variable(s) to be used for optimisation (in hydrology this is often but not exclusively discharge, also limited by data availability), the objective function as quantification of the goodness of model fit to observations, and the algorithm to be employed.

Over the last decades, a tremendous number of different algorithms and strategies has been developed, ranging from manual trial and error to automated methods, from deterministic to probabilistic approaches, from (non-)linear regression of traditional statistics to Bayesian philosophy, from single to multi-objective, from local to global calibration. In this context, it is impossible to provide an overview, instead the interested reader is referred to, for example, Beven (2009), Efstratiadis and Koutsoyiannis (2010), Kavetski et al. (2018), and Moriasi et al. (2007).

A common problem of calibration is equifinality, i.e. that a calibration procedure may result in not one superior but in a number of different parameter realisations achieving the same goodness of fit (Beven, 1993, 2006a). Furthermore, overfitting or overparametrisation of a model may result in a good fit of model simulations to measurements but not necessarily in an adequate predictive power when applied to new data. The relationships between equifinality, overfitting, goodness of fit, and parametric uncertainty of model results are, however, not well understood and can be model and catchment specific (e.g. Fenicia et al., 2016; Her and Chaubey, 2015; Schoups et al., 2008; Whittaker et al., 2010).

1.1.3 Dealing with uncertainties

The understanding and quantification of uncertainties along the modelling process are vital steps towards, on the one hand, the assessment of reliability of simulations and predictions and, on the other hand, the improvement of models and the reduction of uncertainties. Furthermore, also an appropriate communication of uncertainties, which is often neglected, can be of value for water managers and decision makers (McMillan et al., 2017). The quantification of uncertainties in hydrology is a difficult task and, again, no common procedure exists. As the equations for hydrological models are complex and nonlinear, the application of analytical methods to directly infer the impact of uncertain variables and parameters on the output is not possible. In addition, already the uncertainty quantification of just a single parameter or variable, such as discharge, is often challenging and hardly achievable (Kiang et al., 2018). Therefore, existing methods and frameworks for the quantification of hydrological model uncertainties typically consider multiple sources of uncertainty in an aggregated manner.

Uncertainty analysis aims at the quantification of uncertainty of the target variable within a model application. It is closely related to *sensitivity analysis*, which attributes this uncertainty to different sources of uncertainty (also referred to as input factors) (Pianosi et al., 2016). *Identifiability analysis* can then be employed by taking observations into account in order to constrain sensitive input factors, i.e. identify input factor realisations, which lead to optimal model performance and hence reduce model uncertainty (Ghasemizade et al., 2017). The rather general term of identifiability analysis includes, for instance, parameter calibration of a model, but extends to the questions of whether an optimal parametrisation with a given model structure and dataset of observations can be determined at all and why this is the case (Guillaume et al., 2019).

The most straightforward procedure for a combined uncertainty analysis is to apply a Monte Carlo (MC) simulation. Therein, multiple realisations (also termed an *ensemble*) of the uncertain input factors are generated. This, however, already requires some knowledge about the uncertainty structure of the input factors (e.g. their underlying distribution or their upper and lower bounds). The model is then run for each ensemble member individually and the resulting spread of the output ensemble can be regarded as a measure of uncertainty.

Another common strategy for dealing with uncertainty is to acknowledge inherent uncertainty in natural processes instead of seeking for the one optimum model and therefore rely on probability theory and Bayesian statistics. The application of Bayesian theory on hydrological modelling requires probabilistic in contrast to deterministic model output. When employing a deterministic model, this can be achieved by the application of MC simulation. This first requires to sample realisations of sensitive model input factors, in Bayesian theory referred to as *prior distribution*, because a priori knowledge of the uncertainty structure is required. In a second step, additional information, such as observations, are taken into account and compared with runs of the model, expressed as the *likelihood* that observations can be reproduced by the model. This eventually leads to the *posterior distribution*, i.e. a refined space of input factors instead of a single best model realisation.

In order to take model uncertainty explicitly into account, Bayesian strategies, e.g. for calibration or uncertainty analysis, require assumptions about the error structure of the model output in the form of an error model. In the simplest form, that error model is assumed to be a Gaussian distribution of residuals (deviations of model output from observations). The parameters of that error model (e.g. mean and variance) are estimated as *latent* (also termed *nuisance*) variables during model calibration (in Bayesian statistics termed *Likelihood maximisation*) in a way that the error model optimally describes the observed residuals.

A large number of algorithms has been developed pursuing that methodology and varying in how they define the prior distributions, how the prior distribution is sampled (typically some MC variant), the definition of the likelihood function and the maximisation method, and assumptions about the error model and which sources of uncertainty are therein explicitly considered (e.g. Beven and Binley, 1992; Kavetski et al., 2006; Krzysztofowicz, 1999; Todini, 2008; Vrugt et al., 2003). However, such approaches are in turn generally associated with a large degree of uncertainty due to, for example, the requirement to define prior distributions of uncertainty inputs, which are usually not exactly known, the underlying assumptions regarding the structure of the residuals, or because sources of uncertainty are not incorporated (often structural uncertainty is neglected).

Especially the quantification of model *structural uncertainty* is challenging and labour intensive and has therefore often been disregarded, especially for complex process-based models. A prominent strategy to tackle that issue is to employ multiple instead of just a single model. This can be done by the direct comparison of model outputs and performance metrics (e.g. Breuer et al., 2009) or within mathematical more rigorous frameworks, such as in *Bayesian model averaging* (BMA) (Duan et al., 2007; Hoeting et al., 1999). BMA can also be used to estimate the structural uncertainty along with other sources of uncertainty (Ajami et al., 2007). This, however, involves the determination of the so-called *Bayesian model evidence* (BME) for which several methods exist (Schöniger et al., 2014; Volpi et al., 2017), but which is also subject to a large degree of uncertainty (Schöniger et al., 2015).

To support the idea of employing multi-model ensembles and reduce the expenses for model initialisation, several flexible model environments have been developed to enable the exchange of model structures and/or single process representations in one application (Clark et al., 2011a, 2008b; Fenicia et al., 2011). In the past, such frameworks were primarily focussing on conceptual modelling approaches. Recently, more flexible environments have become available (Kneis, 2015; Walker et al., 2018) as well as frameworks explicitly focussing on process-based modelling approaches (Clark et al., 2015b,c).

Uncertainty quantification also imposes the question of how such information can be used to improve hydrological models and/or how to reduce the uncertainty of model predictions (i.e. model adjustment apart from mere parameter calibration). A common approach in short-term flood forecasting is to assess an error correction model from previously observed prediction errors (e.g. by means of time series analysis, provided there is a temporal correlation of errors) and apply that correction model on forecasts of the target variable (e.g. discharge) to obtain improved predictions (see Pinzinger et al., 2014, for an overview). The advantage of such approaches is that they are fast and easily applicable. Such uncertainty *(post-)processors* can as well be implemented in a Bayesian context (Herr and Krzysztofowicz, 2015).

Data assimilation (DA) is another strategy for the improvement of model simulations and predictions. The term *data assimilation* can be understood in different ways, but herein shall be defined in the sense of the assimilation of additional measurements during model runtime to update the internal model states and provide an improved foundation for the generation of model predictions. A prominent algorithm for DA is the *Kalman filter* (KF) proposed by Kalman (1960). This procedure aims at the filtering of observations into the model to update the state variables while explicitly considering uncertainties in both the model states and the observations. Further developments such as the extended KF (EKF) and the ensemble KF (EnKF) enable the use of the KF approach also for non-linear and computationally expensive models (Evensen, 1994; Reichle, 2008). The EnKF and related derivatives gained lots of attention in hydrological modelling (e.g. Chen et al., 2013; Clark et al., 2008a; Komma et al., 2008; McMillan et al., 2013).

DA is especially attractive along with the use of remote sensing data, which are typically available over larger scales and avoid labour intensive and costly field work (e.g. Blöschl et al., 2014; Tian et al., 2017). However, some important drawbacks exist, which involves the need for an adequate quantification of model and observation uncertainties (which is hardly achievable and often done in a simplified manner), assumptions about the error structures that are usually violated in hydrological application (i.e. Gaussian distributed and homoscedastic residuals), or the decision on algorithmic parameters and which model state(s) should be updated (Sun et al., 2016; Thiboult and Anctil, 2015).

The particle filter (PF) approach is a different method for DA that relaxes the assumption of a Gaussian error model. It gained some interest for hydrological application and produces more robust results and uncertainty quantifications than the EnKF, but poses a higher computational burden, still includes the challenge of an adequate uncertainty quantification, and furthermore depends on the selected sampling strategy (DeChant and Moradkhani, 2012; Moradkhani et al., 2005; Vrugt et al., 2013). Another DA approach is variational DA (VDA), a batch method assimilating multiple observations at once. This is often used in meteorology but only gained little attention in hydrological forecasting. While it is less computationally demanding than EnKF, it requires the complicated derivation of an adjoint model (Ercolani and Castelli, 2017; Seo et al., 2003).

In hydrological *forecasting*, the most important source of uncertainty poses the uncertain evolution of precipitation, which is the dominant forcing of hydrological models. To account for uncertain rainfall forecasts, a whole ensemble of predictions is usually considered (as illustrated for the forecast mode in Fig. 1.1). Such ensembles of numerical weather predictions are typically generated by small perturbations of model states, which can have a large impact on the forecasts. These forecasts are in turn model-based products and therefore incorporate uncertainties. The resulting ensemble is then propagated into a hydrological model (Cloke and Pappenberger, 2009; Kneis et al., 2012). Thereby, possible mismatches of spatial scales between the meteorological and hydrological model need to be accounted for (also referred to as *downscaling*). In addition, before application of the hydrological model, rainfall forecasts can be analysed and corrected, e.g. via *bias correction* and other *pre-processing* methods (Bürger et al., 2009; Gudmundsson et al., 2012; Kelly and Krzysztofowicz, 2000; Reggiani and Weerts, 2008; Verkade et al., 2013).

1.2 Objectives

The preceding literature study illustrates that there are numerous strategies focussing on the quantification, attribution, and reduction of uncertainties. However, there still is a large gap of knowledge and some sources of uncertainty have hardly been investigated. Therefore, the goal of this thesis is to identify and close such gaps of research on model deficiencies, uncertainty analysis, and their quantification. To achieve this, the following research questions shall be investigated in more detail.

What are the strengths and deficits of frequently applied hydrological models? What are the reasons for deficits and how can they be alleviated?

In hydrology a large variety of models of different conceptualisations exist. These are accompanied by different strengths and deficits. Therefore, common strengths and deficits of prominent hydrological models shall be identified, including the sources for such deficits and which actions could be started for their elimination.

How does the methodology of discretisation of landscapes into spatial model units influence simulation results?

There are many studies focussing on parametric model uncertainty and the impact of calibration on model results. However, even though it is an integral part of every hydrological model application, there are hardly any studies focussing on the impact of decisions during landscape discretisation on simulation results. Consequently, existing approaches of landscape discretisation and software for their implementation shall be reviewed and common parameters analysed in more detail.

What is the adequate structure for a process-based hydrological model?

The determination of adequate representations of hydrological processes for a computer model is difficult and labour intensive. Moreover, while simplified numerical time step integration enables rapid model application, it may deteriorate model results and conclusions. Many model developers neglect this issue. This holds especially true for computationally demanding process-based models. However, rapidly growing computational resources and technologies offer new possibilities. Therefore, the identifiability of complex process-based model structures will be investigated.

Are process-based models suitable tools for operational forecasting? What are the deficits and how can forecasts be improved?

Operational forecasting usually relies on the application of statistical approaches or conceptual models, which are fast and straightforward in their implementation. However, in contrast to process-based models, they usually cannot provide as much detail on hydrological process behaviour or the spatial distribution of hydrological variables. In addition, increasing computer power nowadays allows for the application of more complex models in an operational context. Consequently, the suitability of a process-based model for seasonal drought prediction shall be investigated for a dryland catchment in northeast Brazil. Besides, the main sources of uncertainty shall be identified and guidelines for future model improvement be given.

1.3 Thesis outline and author contribution

This cumulative thesis combines three published articles and one manuscript submitted to a scientific journal. All the manuscripts were prepared during my doctoral research time at the University of Potsdam. They comprise the combined efforts of different teams of authors. In the following, an overview of the four papers along with my individual contributions shall be given.

Chapter 2: "Charakterisierung und Analyse hydrologischer Modelle im deutschsprachigen Raum" ***("Analysis and characterisation of hydrological models in German-speaking countries")***

The paper gives an overview of prominent hydrological models employed in German-speaking countries based on an online survey. Furthermore, general model conception and process representation along with strengths and deficits are analysed in more detail. The paper is published in German language. However, in order to fit into the context of this thesis, the article has been translated into English. It has been published as:

Guse, B., Pilz, T., Stoelzle, M., and Bormann, H. (2019). "Charakterisierung und Analyse hydrologischer Modelle im deutschsprachigen Raum". In: *Wasser und Abfall* 5.2019, pp. 43–52.

Own contribution: support to idea and methodology; compilation of the online survey with support of all co-authors; data analysis with support of all co-authors; contributions to figures and tables; contribution to the German manuscript; translation into English.

Chapter 3: "lumpR 2.0.0: an R package facilitating landscape discretisation for hillslope-based hydrological models"

The paper introduces lumpR, a software for semi-automated landscape discretisation and model initialisation. The most important discretisation parameters were investigated within a case study in a dryland catchment. The article has been published as:

Pilz, T., Francke, T., and Bronstert, A. (2017). "lumpR 2.0.0: an R package facilitating landscape discretisation for hillslope-based hydrological models". In: *Geosci. Model Dev.* 10.8, pp. 3001–3023. DOI: `10.5194/gmd-10-3001-2017`.

Own contribution: idea and methodology in consultation with the co-authors; development of the lumpR package with significant contributions by T. Francke; realisation of experiments, analysis of results, and compilation of figures and tables with support of T. Francke; writing of the manuscript with contributions by the co-authors.

Chapter 4: "How to Tailor my Process-based Model? Dynamic Identifiability Analysis of Flexible Model Structures"

The manuscript introduces a framework for identifiability analysis of complex process-based model structures. It is seeking for the most adequate model structure in terms of process representation and numerical solver while explicitly accounting for parameter interactions. Moreover, the variability of identifiability in space and time is investigated. The manuscript has been submitted as:

Pilz, T., Francke, T., Baroni, G., and Bronstert, A. (2019b). "How to Tailor my Process-based Model? Dynamic Identifiability Analysis of Flexible Model Structures". Submitted to Water Resources Research.

Own contribution: experiment design with contributions by all co-authors; programming of process formulations; model initialisation and realisation of experiments; analysis of results, compilation of figures and tables, and writing of the manuscript with support of all co-authors.

Chapter 5: "Seasonal drought prediction for semiarid northeast Brazil: what is the added value of a process-based hydrological model?"

This study compares the abilities of a statistical regression approach and a process-based hydrological model to generate drought forecasts in a dryland catchment. Moreover, uncertainties and limitations of the approaches are analysed and discussed. The paper has been published as:

Pilz, T., Delgado, J. M., Voss, S., Vormoor, K., Francke, T., Costa, A. C., Martins, E., and Bronstert, A. (2019a). "Seasonal drought prediction for semiarid northeast Brazil: what is the added value of a process-based hydrological model?" In: *Hydrol. Earth Syst. Sc.* 23.4, pp. 1951–1971. DOI: 10.5194/hess-23-1951-2019.

Own contribution: experiment design with contributions by all co-authors; realisation of experiments regarding the WASA-SED model; analysis of results, compilation of figures and tables, and writing of the manuscript with support of all co-authors.

2. Analysis and characterisation of hydrological models in German-speaking countries

Abstract

In order to investigate different aspects of hydrology and water management, a large number of models of different conceptions with various process representations is employed. This study presents the results of an online survey regarding the classification of models and the characterisation of associated strengths and deficits.

Condensed summary:

- For the analysis of a survey among modellers, 47 completed questionnaires comprising 26 different models of various conceptions were available.
- There are small differences among groups of model conception with respect to temporal resolution and spatial domain of application.
- Processes of evapotranspiration are often implemented in a process-based manner while for processes related to groundwater and soil water movement conceptual approaches dominate.

Published as:
Guse, B., Pilz, T., Stoelzle, M., and Bormann, H. (2019). “Charakterisierung und Analyse hydrologischer Modelle im deutschsprachigen Raum”. In: *Wasser und Abfall* 5.2019, pp. 43–52
The original article has been translated from German into English.

2.1 Introduction

Hydrological models differ in various aspects, such as conception, field of application, spatial and temporal resolution, computational demand, and degree of detail in the representation of hydrological processes (Bormann et al., 2009a). The decision for a specific model may vary depending on the goal of a study. What follows are different demands regarding the degree of detail and physical realism in the representations of hydrological processes, as their importance may vary depending on the type of application and the catchment area (Wagener et al., 2001). Following Clark et al. (2015b), there are at least four different aspects that should be considered when deciding for a specific model: the processes that shall be considered; degree of detail in the representation of these processes; applicability, e.g. with respect to computational demand; data availability. However, the experience of a modeller may play a further important role.

A simplified process formulation may exhibit a less realistic simulation in space and time (Clark et al., 2011c; Wagener et al., 2001). On the other hand, in this way computational demand and efforts for data preparation can be reduced. If a simplified formulation is not sufficient to adequately represent the hydrological processes, a more physically-based representation should be used (e.g. an approach based on the energy balance instead of the degree-day method for to calculate snowmelt). To simulate a specific hydrological process, there typically various alternative processes representations available (Clark et al., 2015b), such as different equations to determine evaporation (e.g. Hargreaves, Priestley-Taylor, Penman-Monteith).

In order to structure the different process representations for the hydrological community, hydrological models can be classified (Dyck and Peschke, 1995). This enables the rapid evaluation of the functionality of a yet unknown model a modeller is confronted with. Usually, model can be grouped into at least three categories of conceptualisation.

- Conceptual models represent processes with simplified and abstract concepts, such as cascades of storages. These models needs to be calibrated and the transferability of the parametrisation to other catchments is limited (Merz and Blöschl, 2004).
- Conceptualised process-based models explicitly represent the (sub-)processes of rainfall-runoff transformation. Yet they also contain simplifying conceptual components.
- Physically based models aim at an exact representation of processes based on the laws of physics (Zehe et al., 2001). Their parameters have a physical meaning and can be derived by measurements, which should result in less efforts regarding calibration.

To sort a model into one of these groups, typically the representation of key processes is used (Bormann et al., 2009a; Wagener et al., 2001), such as evapotranspiration (often Penman-Monteith vs. conceptual approaches) or soil water movement (e.g. Richards' equation vs. storage-based models).

The simplified representation of the hydrological system, also with respect to a specific application, necessarily leads to deficits (Beven, 2007). Such deficits may arise from the conceptualisation (Fenicia et al., 2011) or can be attributed to specific characteristics of a study area. For instance, a simplified formulation of processes related to snow might be sufficient for study areas in northern Germany, while it can cause large errors in alpine regions. In that way, the conception of a model is directly connected with the possibilities of meaningful model applications. The knowledge about strengths and deficits of a model is a condition for its consistent and productive application (Elfert and Bormann, 2010) and may further aid in the interpretation of results (Guse et al., 2014).

Up to now, deficits in process formulations usually have been investigated in combination with a specific model. In this way it is possible to determine the deficits of a process formulation in a specific model (Guse et al., 2014). However, such model-specific insights are not necessarily transferable to other models. More systematic analyses are typically based on the comparison of multiple models (e.g. Clark et al., 2015a). Therein, different models are applied with unified datasets and their strengths and deficits are subsequently analysed respecting the specific objectives of the study (Breuer et al., 2009; Holländer et al., 2009). Such analyses showed that models of different conception represent specific parts of the hydrograph with varying quality (Breuer et al., 2009) and react differently on issues with data availability (Bormann et al., 2009b).

In the hydrological community there is a debate whether individual models should be optimised or if it is more productive to discuss model structures and process representations in a joint effort in order to develop a broadly accepted community hydrological model (Weiler and Beven, 2015). Pursuing the approach of systematic model comparison, Clark et al. (2016) call for a community modelling process to compare process implementations of different models.

The goal of this work is to provide an overview over frequently used hydrological models in German-speaking countries, focussing on model characteristics, strengths, and deficits. To achieve this, an online survey among model developers and users was conducted. The following research questions are addressed:

- How large is the *spectrum* of models and is there a consistent classification?
- How are specific *processes* represented in models with different conceptions?
- What *strengths and deficits* with respect to process formulations are attributed to the models mentioned along the survey and can systematic causes for deficits be inferred?

2.2 Methods

2.2.1 Survey design

Between November 2017 and January 2018 an online survey was conducted among users of hydrological models. The survey was distributed over the network of the Deutsche Hydrologische Gesellschaft e.V. (DHG) and further directed towards research groups in Switzerland and Austria. The completion of the questionnaire was done anonymously. The complete questionnaire can be retrieved from the DHG web page: `https://www.dhydrog.de/dhg/arbeitskreise/` (German only; last access 5 June 2019).

The questionnaire consisted of four topics: model properties, implemented processes and parameters, strengths and deficits, and application. For each topic, different questions were formulated, which could mostly be answered in a quantitative manner (Table 2.1). Thereby, three categories of answers were be distinguished. Free text enabled to fill in arbitrary text. Predefined fields gave a limited number of choices, of which one had to be selected. For some predefined fields it was allowed to select multiple choices. A further category was a drop-down menu, from which a specific answer had to be chosen.

2.2.2 Survey analysis

In order to improve the comparability of the answers of the questionnaire, some of the information were grouped into categories. Only completed questionnaires were used for further analysis. However, it was not mandatory to provide answers to all questions, which resulted in gaps in some of the datasets. Small deviations in model names (e.g. due to varying versions or typos) have been manually adapted. Specifications from free-text fields were categorised into predefined classes. In case of multiple answers of a participant within one free-text field, all the answers were counted individually.

2.3 Results and discussion

The survey was started by 81 persons and completed by 47 participants. Only the latter were therefore used for further analysis. In the 47 datasets 26 different models were mentioned.

2.3.1 Model conception

Most of the participants classified their model as conceptualised process-based (22 datasets, 48 %), while 12 models (26 %) were grouped as conceptual and physically based, respectively (Table 2.2). For one model (CMF) no classification was given. Eight of the ten models, which were mentioned several times, have been classified into different groups of conception by different participants. The models SWAT, J2000, and TRAIN, for instance, have been classified as conceptualised process-based and physically based, respectively, models NASIM, COSERO, and LARSIM as conceptual and conceptualised process-based, respectively. The exceptions are HBV, which consistently has been classified as conceptual model, and WaSiM-ETH, which

Table 2.1: Overview over the topics of the questionnaire.

Topic	Request	Category of answer
Model properties	Name + version	Free text
	Conception	Predefined
	Spatial discretisation	Predefined
	Temporal discretisation	Predefined (multiple choices)
	Scale of application	Predefined (multiple choices)
	Determinism	Drop-down menu
	Dimensions	Drop-down menu
Model application	Operating system	Predefined (multiple choices)
	Handling	Predefined (multiple choices)
	Field of application	Free text
	Catchments / regions	Free text
	Key publications	Free text
Processes and parameters	Implemented processes	Predefined (multiple choices)
	Further processes	Free text
	Number of calibration parameters	Predefined
Strengths and deficits	Strengths	Free text
	Deficits	Free text
	Causes for deficits	Free text
	Ideas for elimination of deficits	Free text

was always seen as physically based. This illustrates the large degree of subjectivity in the perception of such a widely accepted classification scheme. Consequently, their is a rather diffuse transition from conceptual to physically based models.

2.3.2 Model discretisation

With respect to discretisation, three categories were distinguished in the survey: spatial extent, spatial resolution or model entities, and temporal resolution (see Fig. 2.1). All categories were analysed separately with respect to the different conceptual model groups. One assumption along the survey was, that models are applied differently with respect to their conceptualisation as they usually have been developed for specific scales, objectives, and data availability. Consequently, a characteristic pattern regarding model conception was expected. Depending on the degree of detail in the underlying process representations, models are only applicable with specific temporal and spatial resolution, and spatial units. Another expectation was that some models would appear as all-rounders and others more specialised with respect to their discretisation.

Spatial scale

With respect to spatial scale, all conceptual model groups are applied over broad ranges (Fig. 2.1a). Nevertheless, it can be seen that physically based model are rarely applied over large scales and conceptual models are less apparent for small scales. Physically based model usually have been developed for field scales, whereas conceptual models are commonly designed for catchment scales. A more dominant pattern supporting this fact was expected.

A potential cause for this mismatch might be the higher data availability nowadays. Detailed soil and land use maps with large spatial coverage and the increasing availability of remotely sensed data may enable the applicability of conceptual models over small areas and the use of physically based models over larger catchments. This tendency is further enhanced by increasing computational resources and less limitations for numerically demanding complex models. Another reason might be that modellers rather stick to models they are experienced

Table 2.2: Hydrological models mentioned along the survey, including their conceptions as specified by the participants. The full list of references can be accessed via the DHG web pages: `https://www.dhydrog.de/dhg/arbeitskreise/` (last access 5 June 2019).

Model	Conceptual	Conceptualised process-based	Physically based	Reference
BlueM	0	1	0	Bach et al., 2009
CMF	0	0	0	Kraft et al., 2011
COSERO	1	4	0	Kling & Nachtnebel, 2009
GSFLOW	0	1	0	Carroll et al., 2016
HBV	5	0	0	Bergström, 1976
HEC-HMS	0	1	0	HEC, 2013
HILLFLOW	0	0	1	Bronstert & Plate, 1997
HydroGeoSphere	0	0	1	Jones et al., 2008
J2000	0	1	1	Nepal et al., 2014
LARSIM	1	1	0	Ludwig & Bremicker, 2006
LWF-Brook90	0	1	0	Hammel & Kennel, 2001
mHM	0	1	0	Samaniego et al., 2010
MIKE-SHE	0	0	1	Abbott et al., 1986
PANTA RHEI	0	1	0	Meon et al., 2015
RoGeR	0	1	0	Steinbrich et al., 2016
SES	0	0	1	Asztalos, 2004
Simulat	0	0	1	Diekkrüger & Arning, 1995
SWAT	0	2	2	Arnold et al., 1998
SWIM	0	1	0	Krysanova et al., 1998
TOPMODEL	1	0	0	Beven & Kirkby, 1979
TRAIN	0	1	1	Menzel, 1997
UHP	1	0	0	Bormann & Diekkrüger, 2003
WASA-SED	0	3	1	Güntner und Bronstert, 2004
WaSiM-ETH	0	0	2	Schulla, 1997
WaterGAP	1	1	0	Döll et al., 2003
Sum	**12**	**22**	**12**	

with and avoid the efforts to become acquainted with a new model, even though the application might exceed the limits of applicability of their model (Holländer et al., 2014).

Spatial resolution

With respect to spatial resolution, the pattern appears more as expected (Fig. 2.1b, Wood (1995)). Physically based model either use a raster-based discretisation or irregularly shaped elementary units, e.g. hydrological response units (HRUs). The same counts for conceptualised process-based models. In contrast, for conceptual models use all types of spatial resolution occur, i.e. lumped approaches, discretisations based on sub-catchments (also termed subbasins), as well as HRUs, and raster cells. Therefore, a clear distinction between conceptual model on the one hand and conceptualised process-based and physically based model discretisations on the other hand can be made. Models without spatial discretisation (lumped approaches) were exclusively classified into the conceptual group.

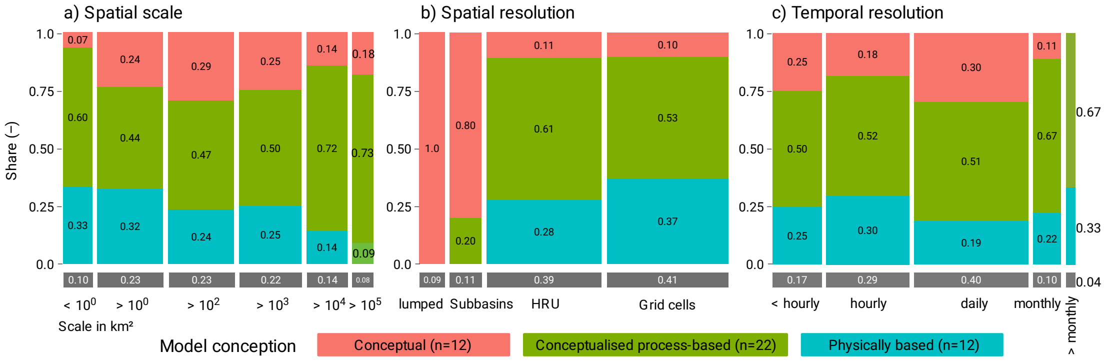

Figure 2.1: Share of answers regarding a) catchment size, b) size of spatial entities, and c) temporal resolution with respect to different model conceptions.

Temporal resolution

Temporal resolution could be specified between less than hourly and more than monthly (Fig. 2.1c). In the survey, models with hourly and daily resolution dominated (70 %). With respect to conceptual classification, no clear pattern can be distinguished, as all conceptual groups are applied with all kinds of temporal resolution. This was surprising as physically based models are usually designed for application with small time steps. As the included processes are typically represented with a high degree of detail, a low temporal resolution increases the risk of numerical instabilities during the integration of the underlying differential equations. In addition, it was expected that conceptual models are predominantly applied with larger time steps as their simplified process representations are designed for lower resolutions. This leads to the question, whether process representation does not anymore depend on structural model conception. It seems that, probably due to increasing data availability and computer resources, models are more frequently applied outside of their original scope of application.

When focussing on spatial scale and temporal resolution of a specific selection of eight models (Fig. 2.2), some models, such as SWAT and COSERO, are applied over almost all combinations of spatial scale and temporal resolution. Other models are more flexible with respect to spatial scale but are constrained regarding the temporal resolution (e.g. TRAIN, LARSIM) and vice versa (e.g. WaSiM-ETH). Subsequent analyses may help in identifying overlapping patterns of models with different conceptualisation.

2.3.3 Model application

The analysis of fields of model application with respect to their conception (Fig. 2.3) shows that flood forecasting and engineering tasks are primarily conducted with conceptual models. In contrast, for process analyses and assessments related to matter transport and water budget more complex models are preferred (process-based and physically based). Presumably, such models seem more appropriate to modellers when extrapolation under changing boundary conditions is necessary.

2.3.4 Processes and parameters

In order to calibrate models to the specific conditions of a study area, they contain parameters, which implicitly represent relevant properties that are not explicitly represented in the underlying conceptual design or the data used to initialise the model. The number of parameters that is used for calibration varies substantially among models and conceptions (Fig. 2.4). Thereby, the full range of predefined parameter numbers can be found. Even the specified numbers for a specific model can differ substantially according to the rating of survey participants. For model COSERO, for instance, the given minimum number of necessary calibration parameters varies between 1 and 10, while as maximum number values between 10 and more than 50 were given. For SWAT the minimum number ranges from 0 to 20 and the maximum number from 10 to more than 50. Both examples illustrate the huge variety, which cannot just be explained by catchment areas of different complexity. As these ratings vary between individual modellers, likely reasons

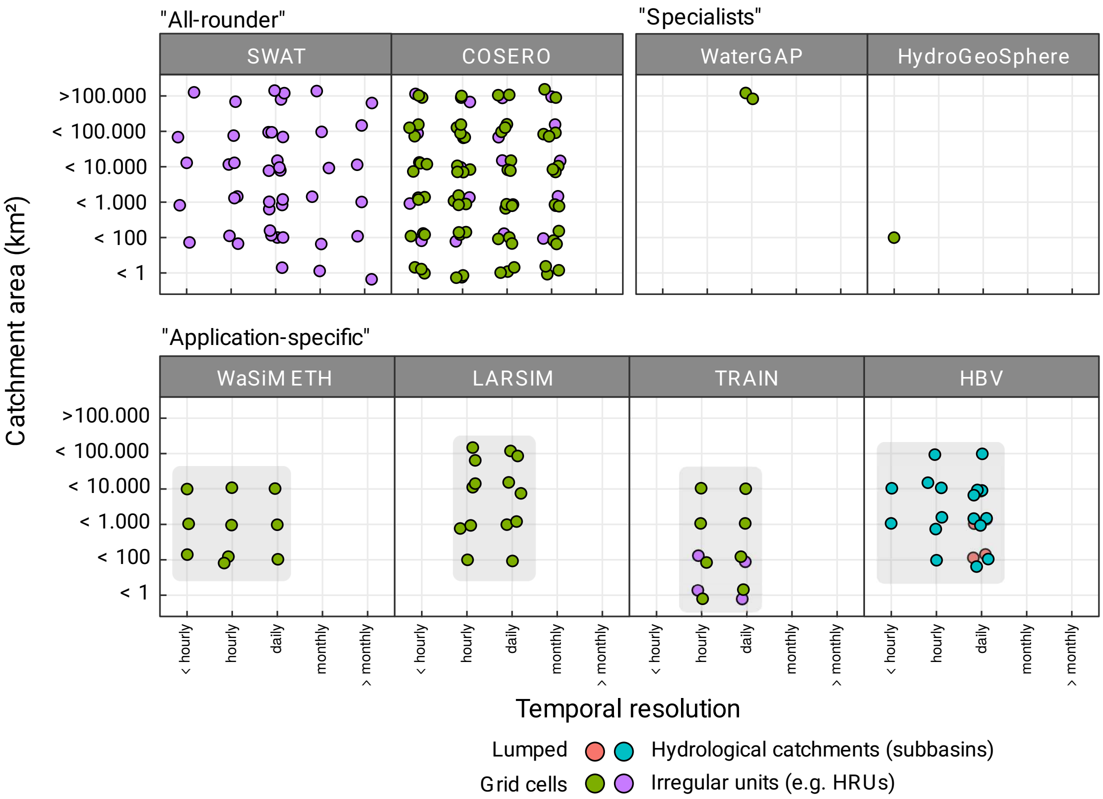

Figure 2.2: Relation between spatial scale and temporal resolution for some selected models.

for the large variety are different objectives of their model application or the use of standard values by some modellers, while other prefer to calibrate more parameters (Holländer et al., 2014).

Despite the variety in specifications it can be seen that the number of calibration parameters is lower for physically based models. Also the differences between minimum and maximum parameter numbers are lower. For physically based models it seems that it is more obvious, which parameters need to be calibrated and which need to be directly inferred from observations.

For the conceptual and process-based models the number of parameters differs to a larger degree. On the one hand this can be explained by the higher number of potential calibration parameters, on the other hand subjective decisions might play a role, which parameters should be calibrated and which should not. With increasing experience with a specific model it is more likely that the parameters, that need to be calibrated, can be narrowed down. On the other hand, the number of calibration parameters may as well vary in dependence of the calibration method.

In the questionnaire, some selected processes were evaluated regarding their conceptual implementation (Fig. 2.5). For all processes it was assessed whether they are implemented empirically, conceptually or physically based. In the analysis the overall model conception was further considered. Overall a consistent picture appears as the conceptual implementation of processes in most cases is in accordance with model conception.

Processes of evaporation and transpiration hold the largest share of physically based approaches across all model conceptual groups (45 %). In that way, model are more independent from the application in a specific region. On the other hand it can also be seen that physically based models may as well contain processes with conceptual or empirical implementation. For instance, subsurface processes (baseflow, interflow, soil water movement) are often implemented in a conceptual manner (67 %). This suggests that for some processes, independent of the overall model conception, a certain implementation is commonly employed. In that way a physically based implementation of evapotranspiration can by implemented in models of different overall complexity.

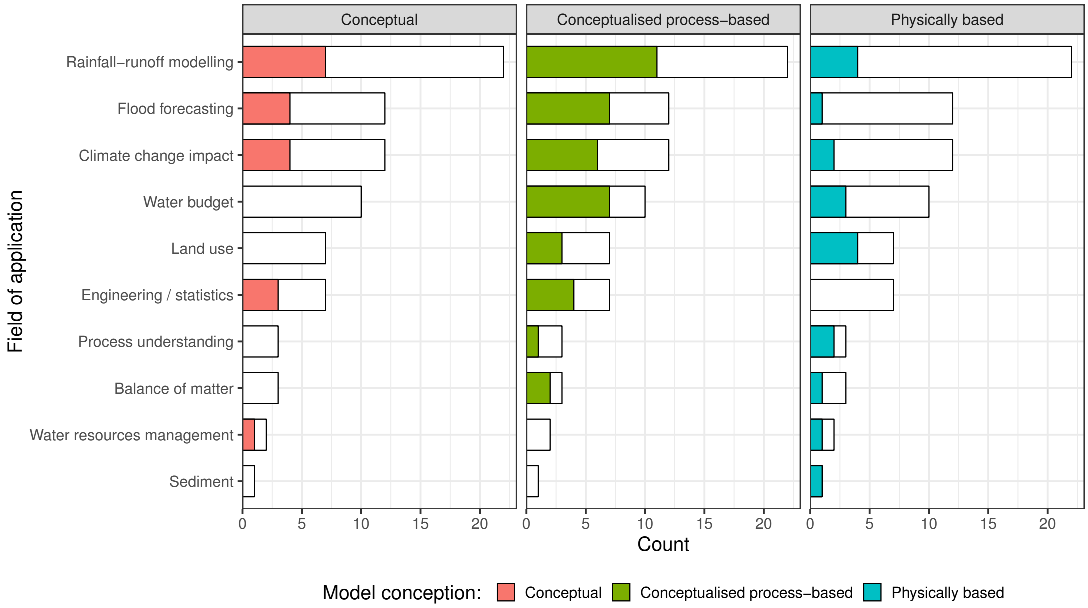

Figure 2.3: Fields of application of hydrological models with respect to their conception.

Due to the varying conceptualisations of certain process implementations in a single model there is necessarily a certain degree of subjectivity in the rating of the overall model conception. The classification into a conceptual group presumably depends on the implementation of certain key processes for the study area. Consequently, certain parts of the model are seen as characteristic for the whole model and changes of these parts may lead to a different classification of the model. Therefore, the empirical snow module of a model applied in an alpine catchment may lead to a classification as conceptual model, while modellers applying the same model in northern Germany might come to a different conclusion.

2.3.5 Strengths and deficits of implemented processes

In Fig. 2.6, strengths and deficits of the implementations of important hydrological processes are shown together with the further model properties of flexibility and efficiency. Irrespective of model conceptual class, processes implementations of soil water / interflow, groundwater / baseflow, and snow are most often seen both as strengths and deficits. Efficiency is commonly regarded as a strength of conceptual and conceptualised process-based models. Thereby, efficiency is related to an efficient handling, straightforward model initialisation or low computational demand. Flexibility, i.e. the lack of it, was only referred to as deficit of many models. Flexibility here is defined as a modularised design of a model or a broad field of application. It can be seen that conceptual models were primarily associated with deficits in process representations and strengths in terms of flexibility and efficiency.

The sources for model deficits as seen by the survey participants can basically be related to general model structure and conception, and missing implementations for certain processes (Fig. 2.7). Limited processes understanding, computational demand, and numerical deficits are almost exclusively related to physically based models. However, these findings should be interpreted with some caution as the absolute number of specifications for physically based models is low. On the other hand, it seems to be a more robust finding that conceptualised process-based models are limited by missing process implementations, inadequate model structure, and problematic model calibration. Furthermore, issues of data availability and quality are less often seen as problematic for conceptualised process-based models than for conceptual and physically based models. Yet it seems unclear if an improved data basis alone would already reduce the model deficits or if it is in addition necessary to improve the process representations (Francke et al., 2018b).

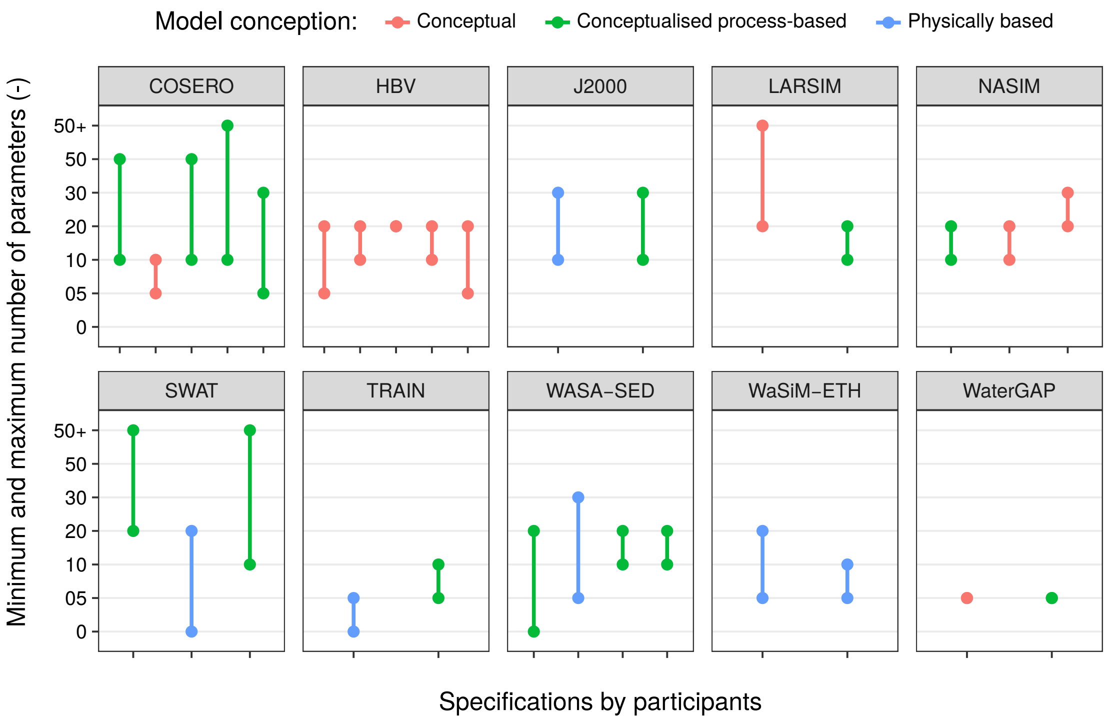

Figure 2.4: Minimum and maximum number of parameters needed for the calibration of a specific model (only models are shown that were mentioned more than once in the survey).

In order to improve hydrological models and eliminate their deficits, improved process implementations are seen as the major strategy by survey participants across all conceptual groups (Fig. 2.8). In addition, improvements with respect to spatial resolution and general model conception as well as more resources (time, money, employees) were mentioned. It remains to be debated, whether more complex process representations are the main key to reduce deficits or if this would rather cause more problems related to data availability, parametrisation, and calibration. Moreover, it is to be expected that increasing model complexity will also lead to less straightforward applicability.

2.4 Conclusions

How large is the spectrum of models and is there a consistent classification?

The variety of different models that have been mentioned in the survey (n = 16) illustrates the large number of models that is used in German-speaking countries. With respect to model conception, most models are conceptualised process-based (22), followed by conceptual and physically based models (12 each). This study is not statistically representative and does therefore not necessarily reflect the true conditions of the hydrological landscape in the German-speaking countries. However, it shows the large diversity of models that can be applied by the hydrological community and indicates that the improvement or combination of existing models should be preferred over the development of new ones.

It was found that different modellers may group a specific model into different conceptual groups. Even though the classification scheme with respect to the structural conception, which was used in this work, is frequently employed among hydrologists, there still seems to be a large degree of subjectivity. This holds also true for the limits of applicability of certain model groups. Nevertheless, there is a certain consensus about the objectives for the different model conceptions. Eventually, it seems to be practical to further use the distinction of conceptual groups for the characterisation of hydrological models.

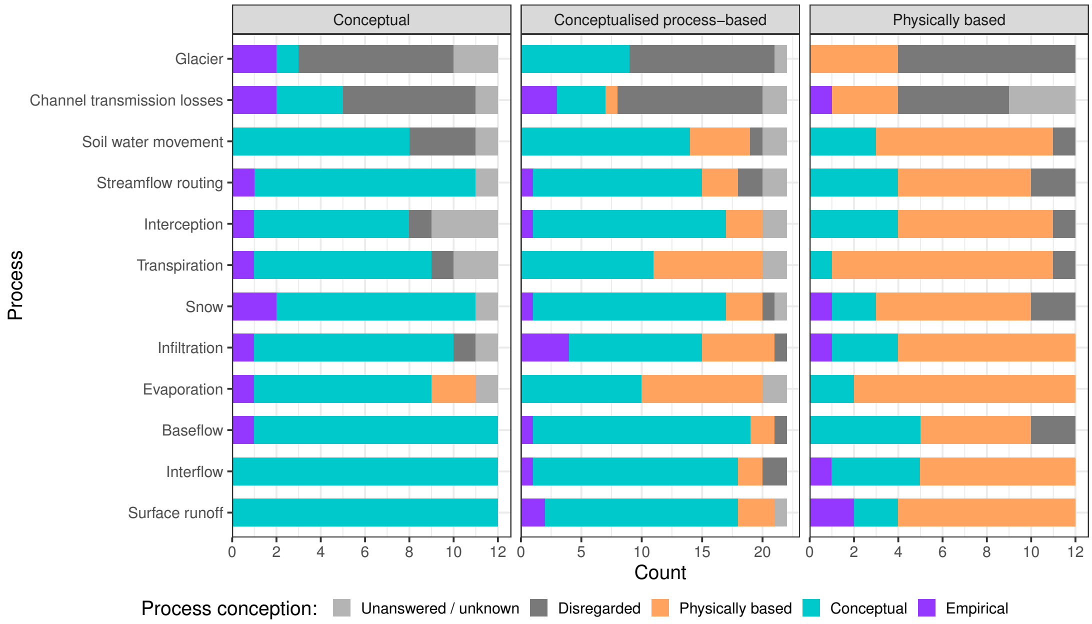

Figure 2.5: Conception of model implementation of selected hydrological processes.

How are specific processes represented in models with different conceptions?

Within models of a specific conceptual class, different processes can be represented with varying degree of detail. Physically based model may as well contain conceptual process implementations and vice versa. Overall it was found that evapotranspiration processes are often implemented in a physically based manner in all model conceptual groups. In contrast, processes related to groundwater and soil water movement are usually empirically or conceptually represented. Only for some models a clear and objective classification into a conceptual group is possible. Therefore, it is always advisable to check the implementation of the most relevant processes before applying a model in a specific catchment.

What strengths and deficits with respect to process formulations are attributed to the models mentioned along the survey and can systematic causes for deficits be inferred?

There is a large diversity of specifications made by survey participants with respect to strengths and deficits of models. Frequently, certain processes were seen as inadequately represented in one model and as strength in another model. Nevertheless, implementations of soil and groundwater processes were most often rated as a deficit. As strengths of more conceptual models were usually seen aspects such as flexibility and efficiency in their application, while for more physically based models certain process implementations and their degree of detail (often implementations related to snow and soil water processes) were mentioned. In general, flexible process implementations with multiple alternatives rather than static representations were seen as strength of models. The opinions about deficits vary greatly among models, which favours a model-specific investigation rather than a universal prioritisation of deficits.

As a common reason for deficits in models, limited or missing process implementations were mentioned. Consequently, suggestions for model improvement were mostly directed towards improved process representations. Furthermore, new methods of remote sensing and further improvement of data availability should be exploited to check alternative process implementations for their capabilities of improved process representation.

2.4.1 Outlook

Future studies could focus on frequently used models of different conception in a more detailed analysis or questionnaire among model practitioners. Based on the conclusions of this work, model could be selected that are applied on similar spatial and temporal scales. Model

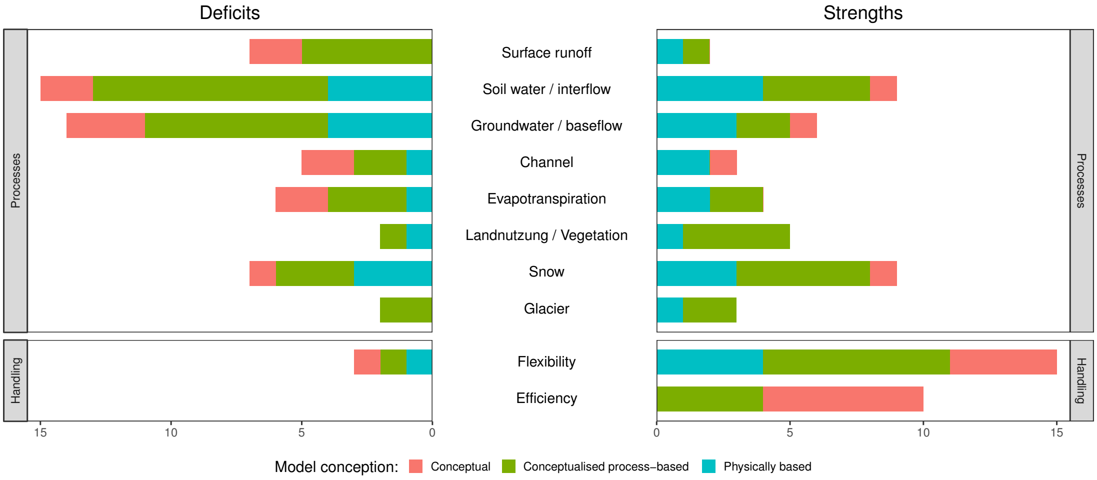

Figure 2.6: Mentioned strengths and deficits of hydrological models.

developers could then elaborate solutions for model improvement, e.g. in a workshop. Moreover, processes that are frequently rated as a deficit could be further investigated in studies of model intercomparison in order to identify more suitable representations and test their applicability.

Acknowledgements

Authors thank the Deutsche Hydrologische Gesellschaft e.V. (DHG) for their support, all participants of the questionnaire for their specifications, and SoSci Survey GmbH for their survey tool. Furthermore, the DHG working group *Prozessabbildung in hydrologischen Modellen* is acknowledged for their support and discussion on the survey design as well as all colleagues from CAU Kiel, University of Potsdam, and University of Freiburg who tested the initial questionnaire and gave feedback for improvements.

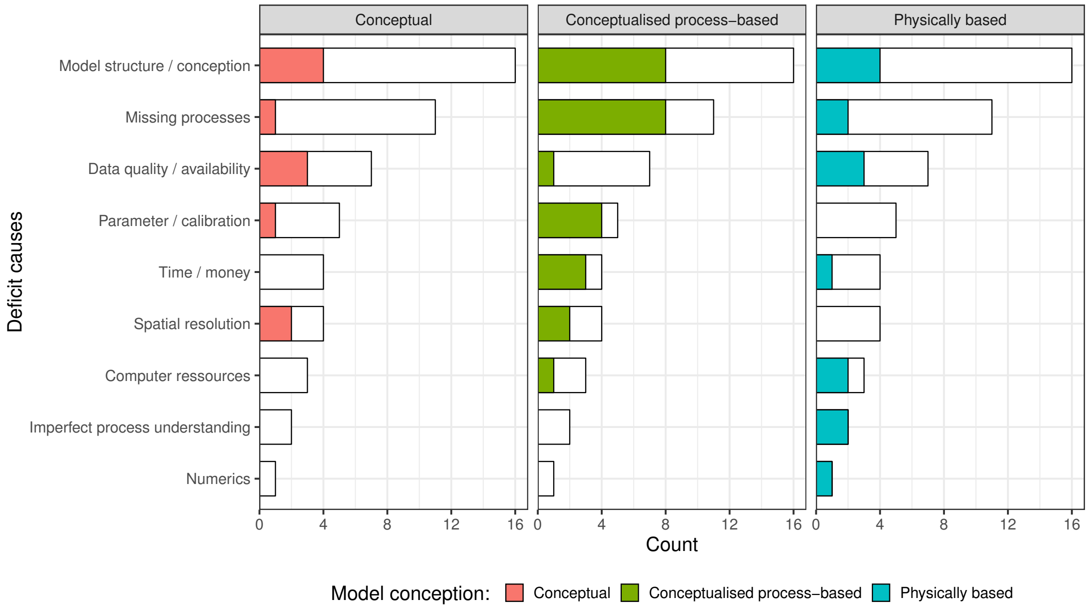

Figure 2.7: Potential sources for model deficits mentioned by survey participants in relation to model conception.

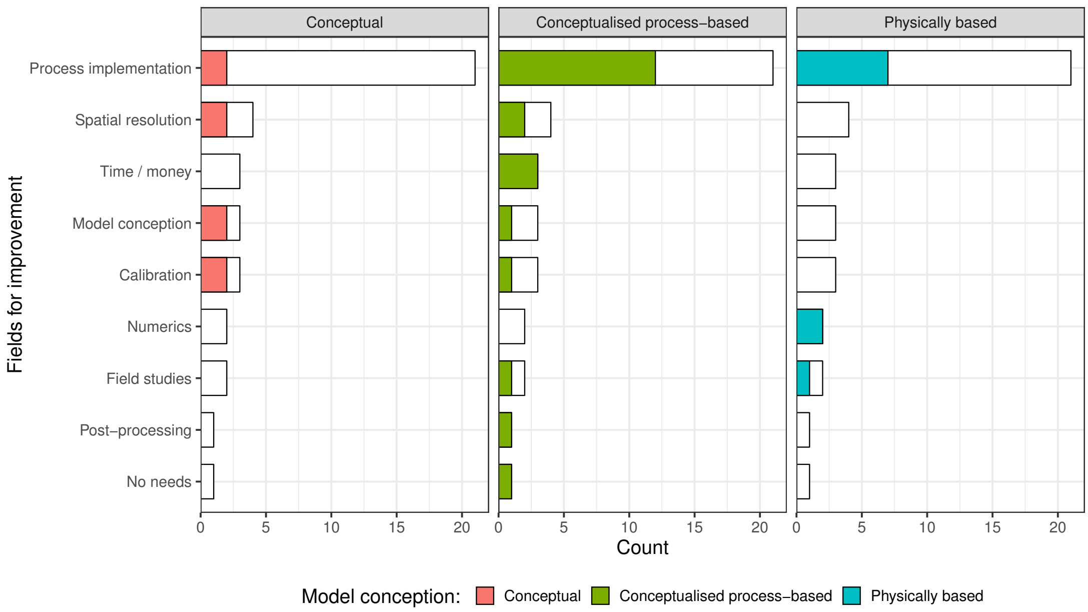

Figure 2.8: Ideas for model improvement mentioned by survey participants.

3. lumpR 2.0.0: an R package facilitating landscape discretisation for hillslope-based hydrological models

Abstract

The characteristics of a landscape pose essential factors for hydrological processes. Therefore, an adequate representation of the landscape of a catchment in hydrological models is vital. However, many of such models exist differing, amongst others, in spatial concept and discretisation. The latter constitutes an essential pre-processing step, for which many different algorithms along with numerous software implementations exist. In that context, existing solutions are often model specific, commercial or depend on commercial back-end software, and allow only a limited or no workflow automation at all.

Consequently, a new package for the scientific software and scripting environment R, called *lumpR*, was developed. lumpR employs an algorithm for hillslope-based landscape discretisation directed to large-scale application via a hierarchical multi-scale approach. The package addresses existing limitations as it is free and open source, easily extendible to other hydrological models, and the workflow can be fully automated. Moreover, it is user-friendly as the direct coupling to a GIS allows for immediate visual inspection and manual adjustment. Sufficient control is furthermore retained via parameter specification and the option to include expert knowledge. Conversely, completely automatic operation also allows for extensive analysis of aspects related to landscape discretisation.

In a case study, the application of the package is presented. A sensitivity analysis of the most important discretisation parameters demonstrates its efficient workflow automation. Considering multiple streamflow metrics, the employed model proved reasonably robust to the discretisation parameters. However, parameters determining the sizes of subbasins and hillslopes proved to be more important than the others, including the number of representative hillslopes, the number of attributes employed for the lumping algorithm, and the number of sub-discretisations of the representative hillslopes.

Published as:

Pilz, T., Francke, T., and Bronstert, A. (2017). “lumpR 2.0.0: an R package facilitating landscape discretisation for hillslope-based hydrological models”. In: *Geosci. Model Dev.* 10.8, pp. 3001–3023. DOI: `10.5194/gmd-10-3001-2017`

3.1 Introduction

Hydrological simulation models are important tools for gaining process understanding, forecasting streamflow, supporting water managers, and climate and/or land use change impact studies. However, the lack of a unified theory in catchment hydrology led to a large number of competing models (e.g. Weiler and Beven, 2015). These differ in terms of aim and scope, process representation and parameterisation, temporal resolution, and spatial discretisation. The latter includes the size and shape of units where the model's underlying mathematical equations are solved.

The most straightforward procedure in terms of data handling and computation is the discretisation of the landscape into quadratic grid cells of equal size, commonly referred to as *fully distributed* approach. This, however, suffers from several drawbacks: high computational burden and large memory demands for small grid cells, while for large grid cells limitations in representing landscape variability and processes at scales smaller than given by the grid's resolution may occur. Another strategy is to treat a hydrological catchment as a single unit, known as *lumped* approach. While being easy to implement along with simple and fast computable conceptual process representations, this approach fails to adequately represent complex landscapes and process interactions often leading to a deteriorated simulation of hydrological catchment behaviour (e.g. Yang et al., 2000). Several approaches exist addressing this scaling problem for which no generic solution has been found so far (Beven, 2006b). Subgrid parameterisation is a common procedure, describing the variability of a response function lying below a model's spatial resolution by the estimation of spatial distribution functions (e.g. Beven, 1995; Bronstert and Bárdossy, 1999; Nijzink et al., 2016). A different approach is to introduce different spatial levels in a model as compromise between distributed and lumped discretisation schemes. This is sometimes referred to as *semi-distributed* approach, often being the preferred choice in practical applications (e.g. Euser et al., 2015; Kumar et al., 2010). Although no clear definition exists, this family of discretisation typically consists of a hierarchical multi-scale scheme dividing a hydrological basin into several subbasins which in turn contain irregularly shaped computational elements being hydrologically uniform entities (e.g. Krysanova et al., 1998). Thus, the properties of both distributed and lumped modelling can be found, often extended by integrating subgrid parameterisation schemes. On the one hand, hydrological processes are simulated at different locations in the study area taking into account distributed model input (such as meteorological forcing and landscape parameters) and producing spatially variable output (such as lateral and vertical water flows). On the other hand, naturally heterogeneous but hydrologically similar areas are aggregated and parameterised in the same manner. The spatial heterogeneity of parameters or state variables of the model, such as hydraulic conductivity and soil moisture, respectively, may thereby be described by constitutive relationships. In favour of more computational efficiency, the topology of individual elements is often neglected. This, however, imposes the threat that significant hydrological connectivity between the elements might not be correctly represented.

In the last decades, a large number of landscape discretisation procedures has been developed for the delineation of spatial units for hydrological models. The number of accompanying software solutions is even larger. This makes it a difficult task to choose a specific model, the corresponding discretisation approach, and potential tools for landscape pre-processing. Therefore, the *first objective* of this paper is to provide an overview of existing landscape discretisation algorithms and software implementations, thereby solely focussing on semi-distributed hydrological model application (Sect. 3.2).

Among the semi-distributed approaches, hillslope-based modelling is an efficient option for representing heterogeneous runoff generation processes while accounting for phenomena of hydrological connectivity (see Sect. 3.2.2). However, so far only few computer programmes exist that aid in the pre-processing of hillslope-based models. Furthermore, these are often model specific, have a limited applicability, and are not freely available or can only be used along with commercial software (see Sect. 3.2.3). The *second objective* therefore is to present a new software package for the pre-processing of hillslope-based multi-scale hydrological models addressing these limitations. It is introduced and described in Sect. 3.3.

The role of detail of discretisation (i.e. the spatial resolution) of hydrological models, partly also as a user decision during the pre-processing process, has long been acknowledged in numerous studies (e.g. Euser et al., 2015; González et al., 2016; Haghnegahdar et al., 2015; Han et al., 2014; Kumar et al., 2010; Wood et al., 1988). For grid-based models, this influence has been thoroughly assessed (e.g. Molnár and Julien, 2000; Refsgaard, 1997; Sulis et al., 2011; Zhang and Montgomery, 1994). Such analyses are relatively straightforward, as changing grid resolution is easily attained. For semi-distributed models, however, systematic and objective analyses covering multiple scale are less common as they require an efficient and automated workflow for the creation of many realisations. The proposed software package is able to provide this. Thus, the *third and last objective* of this paper is to present an example case study in a semi-arid catchment using the WASA-SED model, thereby conducting a sensitivity analysis of crucial discretisation parameters (Sect. 3.4).

Eventually, the findings of this study are discussed and conclusions are inferred (Sects. 3.5 and 3.6, respectively).

3.2 Review of landscape representation in hydrological modelling

As being the starting point of any representation of landscapes in computer models, this section starts with a short overview of basic approaches for the representation of topography in computer models (Sect. 3.2.1). What follows is a review of methods for the delineation of spatial entities in semi-distributed hydrological modelling (Sect. 3.2.2) and, finally, an appraisal if and how these are supported by existing software tools (Sect. 3.2.3).

As a basis for watershed delineation and further landscape discretisation, a number of pre-processing steps have to be performed. These shall be briefly mentioned but not further discussed herein. Typically included is *(i)* pit filling of the digital elevation model (DEM) to remove sink areas and ensure a proper drainage of water from the catchment; *(ii)* computation of flow directions; *(iii)* computing upslope contributing area (i.e. the accumulated flow for each DEM unit); *(iv)* derivation of the river network, typically based on flow accumulation from the previous step; *(v)* delineation of the hydrological basin and subbasins. For each step different algorithms have been proposed whose selection depends on the type of DEM, model to be employed, aim and scope of the study, or just personal preferences, e.g. in terms of software to be used (e.g. Costa-Cabral and Burges, 1994; Lacroix et al., 2002; Moretti and Orlandini, 2008; O'Callaghan and Mark, 1984; Tarboton et al., 1991; Vivoni et al., 2004).

3.2.1 Topography representation in computer models

Topography is typically mapped using DEMs. *Contour-based* DEMs store terrain information as contour lines (or x, y coordinate pairs) of specific elevation (Moore et al., 1988). Moore and Grayson (1991) and Maurer (1997) showed example applications of contour-based terrain analysis for hydrological model application where this type of DEM proved to be powerful as its structure is based on how water flows on (albeit not necessarily below) the land surface. However, although they have been further investigated for hydrological application (e.g. Dawes and Short, 1994; Maunder, 1999; Moretti and Orlandini, 2008; Zhang et al., 1999) contour-based DEMs come along with some limitations. They have a relatively high data-storage demand, topographic attributes are complicated to derive, and they provide no computational advantages (Moore et al., 1991).

Triangulated irregular networks (TINs) form a type of DEM sampling elevation points at specific landscape features, such as peaks or ridges, and form an irregular network of x, y, and z coordinates. They are very flexible as, due to their irregular structure, they are able to map regions of high heterogeneity with more data points than smooth terrain and thus avoid redundancy and increase data-storage efficiency (DeVantier and Feldman, 1993; Moore et al., 1991). TINs also proved to be useful in a number of hydrological applications (e.g. Freitas et al., 2016; Ivanov et al., 2004; Tucker et al., 2001; Vivoni et al., 2004). Their irregularity, however, makes the computation of topographic attributes more difficult and there can be problems when determining upslope connections for watershed derivation (Moore et al., 1991).

Table 3.1: Classification of prominent landscape discretisation approaches for semi-distributed hydrological modelling. Approaches are ordered as they appear in the text. For the meaning of abbreviations of the approaches see text. Listed model and software solutions are non-exhaustive examples. Key references refer to the introductory or an illustrative example publication.

Approach	Class	Model	Software solution	Key reference
REA	Areal unit	–	–	Wood et al. (1988)
GRU	Areal unit	SIMPLE, CHARM	WATFLOOD	Kouwen et al. (1993)
ASA	Areal unit	SLURP	SLURPAZ	Kite (1995)
Hydrologic landscapes	Areal unit	–	–	Winter (2001)
HRU	Areal unit	PRMS, SWAT, SWIM, PREVAH, GSFLOW, MHYDAS	IOSWAT, AGWA, AVSWAT, WINHRU, Geo-MHYDAS	Flügel (1995) and Leavesley et al. (1983)
HSB	Hillslope	h3D	–	Troch et al. (2003)
Single Hillslope	Hillslope	KINEROS, IHDM, CATFLOW, HILLSLOPE	–	Bronstert (1999)
Representative Hillslope	Hillslope	WEPP	GeoWEPP, LUMP	Flanagan and Nearing (1995)
ECS	Hillslope	–	SMART	Khan et al. (2014)
flow-interval scheme	Hillslope	–	–	Yang et al. (2002)
REW	Functional unit	–	–	Reggiani et al. (1998)
Functional response units	Functional unit	TAC	–	Uhlenbrook and Leibundgut (2002) and Zehe et al. (2014)
Multiple scales	Mixed	RHESSys, WASA-SED	LUMP	Güntner (2002)

Grid-based DEMs store elevation information as a regularly spaced mesh. There are a number of drawbacks as the regular structure might impose artefacts and discontinuities while sub-resolution landscape features cannot be captured limiting the applicability for hydrological purposes. Furthermore, when increasing the grid's resolution to reduce these problems, computational burden and memory requirements are increased reducing their suitability for large-scale applications. Nevertheless, grid-based DEMs are the most widely used data structures due to their straightforward generation from remote sensing data, direct applicability for further investigations, and efficient calculation of geomorphological characteristics (DeVantier and Feldman, 1993; Moore et al., 1991).

Some models explicitly use TINs, such as tRIBS (Ivanov et al., 2004), or modelling units derived from contour-based DEMs, e.g. THALES (Grayson et al., 1992a). The majority, however, delineate irregularly shaped polygons as computational model units derived from grid-based DEMs which will be further discussed in the following subsection.

3.2.2 Discretisation approaches in semi-distributed hydrological modelling

There exists a large number of landscape discretisation schemes for semi-distributed hydrological model application. In this context, with discretisation we understand the process of deriving computational units for a hydrological computer model from spatial input data. We do not consider temporal resolution here. The spatial discretisation in a model determines size, hierarchy, and topology of model elements. For semi-distributed hydrological models, these comprise landscape elements such as (sub-)catchments, river segments, hillslopes, hillslope segments (e.g. different slope sectors along a hillslope), hydrologically homogeneous areas, soil units, and horizons. In correspondence to the dominating hydrological processes, the objects of higher hierarchy are commonly determined by topography, while for the lower hierarchy soil and vegetation are the distinguishing factor.

For our review we identified four general classes of discretisation approaches which are presented along with specific examples in Tab. 3.1. These include *areal unit* schemes, which delineate spatial polygons as fundamental modelling units; approaches taking *hillslopes* as spatial elements; *functional units* with a focus on a homogeneous process description rather than mere spatial entities; and a *mixed* class, typically comprising a hierarchical scheme of different spatial scales. With the latter we mean conceptions exceeding the common *watershed*–

subbasin–element scheme. In the following, the presented approaches shall be briefly described whereas more information on the mentioned software solutions can be found in Sect. 3.2.3.

Wood et al. (1988) focussed on how to define a hydrological (sub-)catchment and studied its dominant controls. They defined the smallest discernible averaging watershed where statistics of runoff generation did not further change as *representative elementary area* (REA) and applied simple conceptual equations for the simulation of runoff generation. They found the size of their REA to be primarily influenced by topography. Similar, though conceptually different, is the study of Kouwen et al. (1993) who were looking for an approach to represent watershed heterogeneity over large basins. They defined *grouped response units* (GRUs), small watersheds as computational units with uniform meteorological forcing that are considered to be hydrologically heterogeneous by consisting of a range of land cover (or some other attribute's) characteristics, where only percent cover is used for characterisation instead of an explicit spatial reference. Topography and the other relevant attributes (e.g. land cover) are assumed to be the major factors influencing runoff generation. A similar conception is utilised by the *aggregated simulation area* (ASA) approach developed for the SLURP model (Kite, 1995). Winter (2001) aimed for an integrated inspection of the complete hydrological system in different terrain types by introducing *hydrologic landscapes*. These consist of variations and multiples of fundamental hydrological landscape units as building blocks characterised by land surface form, geologic framework, and climatic setting to describe movements of surface water, groundwater, and atmospheric water, respectively.

The concept of *hydrological response units* (HRUs) is directly related to a smaller spatial level. Introduced by Leavesley et al. (1983) for their Precipitation–Runoff Modeling System (PRMS) and further elaborated by Flügel (1995), it evolved into a prominent landscape discretisation scheme utilised by many models. An HRU is assumed to be a homogeneous set of hydrological process dynamics formed by a pedo–topo–geological association with specific land use and as such controlled by land use management and physical landscape properties. The conception has been adopted for models such as SWAT (Manguerra and Engel, 1998), SWIM (therein termed *hydrotopes*) (Krysanova et al., 1998), PREVAH (Viviroli et al., 2009), or GSFLOW (Markstrom et al., 2008). MHYDAS is a process-based hydrological model for which the HRU concept has been further pursued for application in agricultural management contexts by including man-made hydrological discontinuities such as ditches and field boundaries (Moussa et al., 2002). However, the HRU approach commonly does not preserve topological information for the spatial units. Instead of a direct representation of water flow pathways, generated runoff is typically summed over all HRUs of a watershed and routed along a representative channel element.

Other approaches divide the watershed into *representative hillslopes* as one- or two-dimensional (1-D or 2-D) approximation of a three-dimensional soil catena separated by drainage network and ridges. For instance, Troch et al. (2003) developed the hillslope-storage Boussinesq (HSB) equation to simulate drainage and soil moisture storage dynamics along a hillslope described by a polynomial function. In their hybrid 3-D hillslope hydrological model (h3D), Hazenberg et al. (2015) employed the latter along with the Richards' equation for vertical flow and a diffusive wave approximation of the shallow water equations for overland flow as an efficient physically based modelling approach aimed for use in continental and global-scale Earth system models. Flanagan and Nearing (1995) introduced WEPP, a complex process-based soil erosion prediction model applicable over hillslopes or small watersheds comprised of multiple hillslopes, channels, and impoundments. They lump individual hillslopes by calculating and averaging quantitative hillslope characteristics. Examples for models treating single hillslopes over smaller scales include KINEROS (Woolhiser et al., 1990), IHDM (Beven et al., 1987), HILLSLOPE (Bronstert, 1994), or CATFLOW (Maurer, 1997). Several studies focussed on how to delineate and describe hillslopes from a DEM (e.g. Cochrane and Flanagan, 2003; Noël et al., 2014), discussing morphometric controls on hillslope parameters (Bogaart and Troch, 2006), or investigating the role of hydrologic connectivity (Smith et al., 2013). Khan et al. (2014), for instance, formulated *equivalent cross sections* (ECSs) as representatives of a part or an entire subbasin and proposed different averaging algorithms based on topographical and physiographical properties. Yang et al. (2002) introduced the *flow-interval hillslope scheme* where a catchment is subdivided into a number of connected flow intervals which are defined by the width (i.e. the number of streams) and the geomorphological area (i.e. the drainage area) as functions of distance from the watershed outlet. Overall, hillslope hydrological modelling has been proven to be useful in regions with steeply sloping landscapes and heterogeneous runoff generation mechanisms where lateral water fluxes are relevant (Bronstert, 1999). On the other hand, especially when iterating through individual hillslopes of a watershed, applicability is clearly limited to smaller scales due to the large computational demand.

In a different concept, a stronger focus is put on *functional units* rather than mere spatial units which always came along with the assumption of homogeneous process dynamics in a certain area. Herein, each unit is characterised by a specific dominant process and an accompanying model conceptualisation, such as for the TAC model of Uhlenbrook and Leibundgut (2002). Reggiani et al. (1998) delineated the basin into autonomous functional subbasins, the *representative elementary watersheds* (REWs), using the drainage network as basic organising structure. The REW is then divided into five functional sub-regions whereas micro-scale physical conservation equations for each sub-region are simplified and averaged, further accounting for thermodynamic exchange between sub-regions and REWs. Zehe et al. (2014) proposed three *functional response units* separating radiation-driven vertical flow from rainfall-driven lateral flow processes on similar functional entities within a hydro-geomorphic homogeneous subbasin. However, it still is a challenge to define the size of averaging volumes and the closure relationships of boundary fluxes (Beven, 2006b).

Regardless of the chosen approach, for the application at large scales it is necessary to find a compromise between sufficient detail in landscape representation and computational feasibility. Thus, several studies examined the impact of discretisation complexity on model performance. They showed that (semi-)distributed models are usually more suitable for the representation of landscape heterogeneity and the exploration of hydrological processes, and are also more able to reproduce observed discharge dynamics than lumped models (Euser et al., 2015; Kumar et al., 2010). However, there is a threshold of subdivision level above which no more improvements can be achieved (Haghnegahdar et al., 2015; Han et al., 2014; Wood et al., 1988). On the other hand, natural variability outside the models' spatial discretisation level can still exhibit an important limitation in the representation of natural processes. This can be accounted for by different spatial scales in a model application (e.g. the scales of meteorological forcing, model application, and basin characteristics) and combining them using parameter regionalisation (e.g. Samaniego et al., 2010).

The concepts described above are often extended by a hierarchical multi-scale discretisation scheme. It commonly includes the basin of investigation being discretised into subbasins (i.e. hydrological sub-catchments) and hydrologically homogeneous modelling elements (e.g. HRUs). Thus, elementary units are grouped into a structure of higher order, summarising (and potentially defining a topology of) their in- and outputs. For the modelling system RHESSys, for instance, the landscape is partitioned into a hierarchy of progressively finer units modelling different processes associated with a particular scale. A given spatial level is represented as object type with a set of states, fluxes, process representations, and corresponding model parameters (Band et al., 2000; Tague and Band, 2004).

For their WASA model, Güntner and Bronstert (2004) developed an even more complex scheme of six spatial levels. Starting at the watershed, subbasins are delineated, sub-divided into representative hillslopes termed landscape units which are further separated into specific parts of the hillslope, the terrain components to account for lateral redistribution of water flows, whereas vertical processes and runoff generation are simulated over individual homogeneous soil–vegetation components for which a representative soil profile with respective soil layers has to be given. As for the GRU approach, at the smaller scales an explicit spatial representation is omitted in favour of a percentage cover representation in order to better capture landscape heterogeneity while keeping storage and computational demands at a minimum. This concept has been proven to be efficient and successful for the simulation of heterogeneous semi-arid landscapes with complex hillslopes and patchy vegetation over large scales dominated by Hortonian overland flow and runoff redistribution mechanisms. As hydrologic connectivity of the landscape can be represented in a realistic manner, the model has been used for a number of studies investigating runoff redistribution and erosion processes (e.g. Bronstert et al., 2014; Güntner and Bronstert, 2004; Medeiros et al., 2010; Mueller et al., 2010).

3.2.3 Software for model pre-processing and landscape discretisation

The previous section described conceptual approaches of landscape discretisation and gave examples for models where these concepts are utilised. Their implementation into a model, however, requires a number of more or less complex pre-processing steps. Together with improving computer facilities and increasingly available DEMs and processing algorithms, software for terrain analysis and discretisation started to evolve. Already in the early 1990s, DeVantier and Feldman (1993) published a review of applications of *geographical information systems* (GIS) in hydrological modelling. Especially for grid-based models, many tasks during spatial data pre-processing can be performed with standard GIS functionality. However, other steps require more specific operations. Thus, many researchers started writing their own scripts tailored to their needs and sometimes later on published or distributed their solutions both commercially and in non-profit manners.

TAPES-G is an early terrain analysis programme written in FORTRAN-77 and C for use on Unix machines already including several algorithms for specific tasks, e.g. five methods alone to calculate flow accumulation (Gallant and Wilson, 1996). A prominent and widely used example is the TOPAZ software package for automated analysis of digital landscape topography addressed to guide farmers, engineers, and scientists in both research and practical application. The programme is free of charge and available on request but the development stopped in 1999 (Garbrecht and Martz, 1999). Tarboton (2003) introduced TauDEM, a freely usable terrain analysis programme for Windows that can be applied independently from the command line or comes along with a *graphical user interface* (GUI) as extension for the commercial ArcMap software and is still being further developed. More information on hydrologically relevant software of the commercial ArcGIS family, such as ArcHydro, can be found at the software's community web pages: `http://resources.arcgis.com/en/communities/hydro/` (accessed 23 June 2016). The *free and open-source software* (FOSS) GIS GRASS provides a number of elaborated and still evolving tools for hydrological model pre-processing, such as *r.watershed*, that have been successfully applied in a number of studies (e.g. Kinner et al., 2005; Metz et al., 2011; Neteler et al., 2012). Another recent example of stand-alone software is GeoNet, a tool for automatic channel head, channel network, and channel morphological parameter extraction from high resolution topography data that can be employed within MATLAB or, in a more recent version, as Python programme (Passalacqua et al., 2010; Sangireddy et al., 2016).

As being a frequently employed discretisation scheme used in well-known hydrological models such as SWAT, many software solutions exist implementing the HRU concept. Typically, the delineation process is based on the intersection of spatial raster data including land cover, soil, and/or geology. The programmes basically differ in terms of additional processing steps (such as terrain analysis), used algorithms, supported data formats and whether they are tailored to a specific model or are stand-alone, GIS back-end (mostly ArcGIS or GRASS), supported operating systems, and whether they provide a GUI or have to be run from command line. Sanzana et al. (2013) developed a number of stand-alone and model-independent Python scripts for terrain analysis and HRU mesh generation making use of GRASS functionalities. Also relying on GRASS and Python, Schwartze (2008) created an extension for QGIS (a FOSS GIS with user-friendly GUI) as a HRU delineation tool. Other software is model specific, such as IOSWAT (Haverkamp et al., 2005), AGWA (Miller et al., 2007, 2002), or AVSWAT (Di Luzio et al., 2004), being addressed to SWAT, WINHRU (Viviroli et al., 2009) written for PREVAH, or Geo-MHYDAS (Lagacherie et al., 2010), which is a collection of SHELL and PERL scripts using GRASS to help users of MHYDAS with the model pre-processing. Some of these are mere wraps around individual programmes to guide through the whole modelling process, including terrain analysis, HRU delineation, preparation of input files, model execution and parameter calibration, and graphical and/or statistical analysis of simulation results.

The pool of software packages for other landscape discretisation schemes is less rich. Lacroix et al. (2002) presented SLURPAZ, an interface between the TOPAZ terrain analysis tool and the SLURP model for the delineation of ASAs. The WATFLOOD flow forecasting system is a framework consisting of a hydrological model (CHARM) including pre- and post-processors, incorporating the GRU approach (Kouwen, 2016) . Around the hillslope-based WEPP model, the geo-spatial assessment software GeoWEPP has been developed integrating TOPAZ, WEPP, and other tools for detailed analysis of spatially and temporally variable environmental and management scenarios (Renschler, 2003). It is presently integrated into the ArcGIS project with a lightweight web-based interface for less advanced users and ad hoc model application (Flanagan et al., 2013). RHESSys comes with interfaces for both GRASS and ArcGIS to assist in landscape pre-processing (Band et al., 2000). Ajami et al. (2016) published SMART, a MATLAB toolbox integrating TauDEM as terrain analysis tool, performing rainfall–runoff simulations over hillslopes in the sense of ECSs with several options for their derivation, and providing functions for the post-processing of modelling results.

Francke et al. (2008) published an algorithm for the semi-automated delineation of representative hillslopes. This *Landscape Unit Mapping Program* (LUMP) first discretises the landscape into various hillslopes, the elementary hillslope areas, and computes a representative catena for each of them. As a next step, similar catenae are grouped into landscape units whereas for the classification several variables can be taken into account such as horizontal and vertical catena length, shape of the profile, and sets of supplemental attributes further characterising the hillslope, e.g. qualitative data such as soil type and land cover class, or quantitative attributes such as leaf area index or further terrain characteristics. These representative catenae are eventually sub-divided into terrain components, e.g. into upslope, middle, and downslope parts. The approach will be further discussed in Sect. 3.3.2 and is illustrated in Fig. 3.2. LUMP is semi-automated in way that the hillslope-based landscape parameterisation is largely automated generating reproducible results and reducing required user decisions to a minimum whereas, on the other hand, expert knowledge can be easily incorporated to improve the discretisation outcome. Contrary to other hillslope-based algorithms, due to the discretisation at multiple scales, it is applicable over large areas with relatively little effort. Even though LUMP was directed to the pre-processing of the WASA-SED model (Güntner and Bronstert, 2004), it a is stand-alone model and the output can as well be used for other hillslope-based models. However, the programme is basically a collection of freely available scripts written in MATLAB and SHELL using GRASS functionalities. Additionally to the dependence on commercial software, the workflow still requires a considerable number of pre-processing steps and user interaction.

Considering the above-mentioned merits of hillslope-based landscape discretisation, the number of tools for automating this tasks is low. On the other hand, manual derivation of such an discretisation is labour intensive, prone to error and rarely fully reproducible, which generally precludes its application on the larger scale. Thus, to meet the second objective of this study to develop a user-friendly and efficient tool for hillslope-based landscape discretisation, it was decided to build upon the LUMP algorithm which already remedies some of the above-mentioned shortcomings.

3.3 lumpR: R package description

The ***l**andscape **u**nit **m**apping **p**rogram for **R*** (lumpR) was developed with the aim to obtain a lightweight user-friendly and efficient tool for hillslope-based landscape discretisation and serve as a pre-processing tool for the WASA-SED model (Güntner and Bronstert, 2004, see also Sect. 3.4.2). In a more general sense, however, it should meet the requirements of being *(i)* platform independent, *(ii)* applicable for other hillslope-based models, too, *(iii)* free and open source, *(iv)* automated as far as possible reducing subjectivity but *(v)* allowing to include expert knowledge. In order to produce an easily applicable software and to meet objectives (i) and (iii) in particular, it was decided to use the scripting language R (R Core Team, 2015) and assemble *lumpR* as a software package for this environment, licensed under the GNU General Public Licence (GPL) version 3 or later.

Figure 3.1 gives an overview of the structure and functionalities of the package. In the following these shall be explained in more detail. For more information on how to install and use the package and to inform about updates the reader is referred to the package's documentation (see Sect. 3.6).

3.3.1 Prerequisites and general workflow

As it is a package of the scripting language R, lumpR requires the statistical software R together with various packages it depends on. It employs a number of external calls to GRASS GIS and

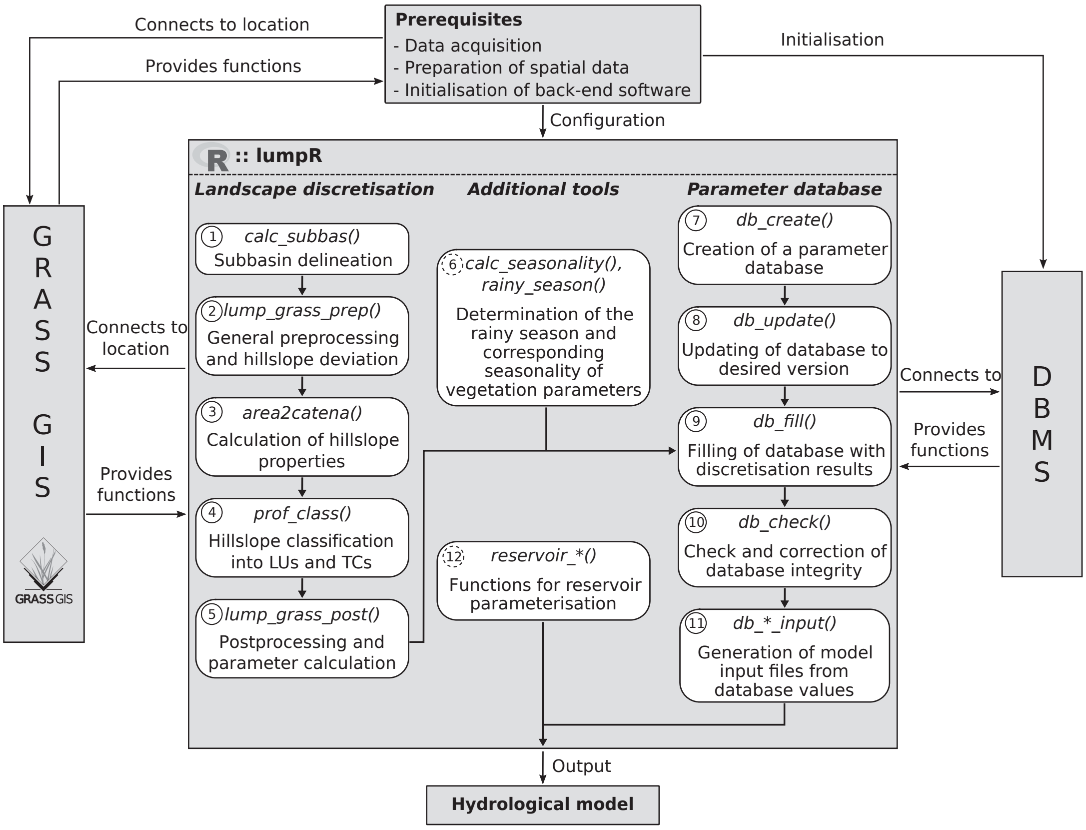

Figure 3.1: Schematic overview of the functionalities of the lumpR package. Shown are all available functions (*in italics*) with a short explanation, interactions with external components, and the typical flow of information during the application. Processing order is indicated by numbers, whereas optional steps are indicated by dashed circles. Note that the acronym DBMS refers to *DataBase Management System*.

thus requires having GRASS to be installed. A third requirement is a *database management system* (DBMS) which will be accessed via the *open database connectivity* (ODBC) which has to be set up as well. So far the DBMSs MySQL/MariaDB, SQLite, and MS Access are supported. When all settings and algorithm parameters can be determined, all processing steps can be run completely automatically. However, it is generally recommendable to process the steps successively to check the intermediate results, if necessary.

After installing the package and all additionally needed software, the user has to acquire and prepare all needed spatial data in a location in GRASS GIS. Internally the package's functions connect to that location, use the given data for processing while partly employing GRASS functions, and finally stores the spatial output in the GRASS location and/or text files in a specified directory for immediate inspection after each function call. As a first step in a new R session, before executing any of the package's function, R has to be connected to the GRASS location by the user. A template script guiding through the processing steps of landscape discretisation and the parameter database management has been prepared and is provided along with the package (see Sect. 3.6).

3.3.2 Landscape discretisation

As is sketched in Fig. 3.1, the process of landscape discretisation involves five functions that should be applied in the following order. This can be ensured by customising the provided template script. Figure 3.2 gives an illustrative example for the outcomes of the following steps (i) to (iv). *(i) calc_subbas()* sub-divides the hydrological basin into subbasins using a given grid-based DEM (black outlined polygons S1–3 in Fig. 3.2). Subbasin size can be influenced by the user by either giving a set of coordinates of drainage locations inferred beforehand or by specifying the parameter `thresh_sub` being the minimum size of a subbasin in number of grid cells internally used by GRASS function r.watershed. Furthermore, the river network is inferred from calculated flow accumulation (the number of upstream raster cells draining through a specific raster cell) via a user-defined threshold (blue lines in Fig. 3.2). *(ii) lump_grass_prep()* does several pre-processing steps needed for later use such as computing *soil–vegetation components* (SVCs) as simple overlay of soil and vegetation raster maps, inferring Horton stream order, and DEM-processing steps including the calculation of flow direction, flow accumulation, relative elevation (i.e. elevations above next downstream river grid cell), and distance to next river grid cell. The results are stored in the specified GRASS location. Furthermore, the function infers *elementary hillslope areas* (EHAs) based on the size parameter `eha_thres`. These are the basic units for calculation of representative catenae and, thus, one can think of them as single hillslopes (denoted as small polygons in Fig. 3.2).

(iii) area2catena() takes data from step (ii) and supplemental raster maps of quantitative and/or qualitative attributes to calculate a representative catena for every EHA (grey boxes in Fig. 3.2). It is characterised by horizontal and vertical length, shape (in terms of cumulated elevation along the hillslope), slope width (approximated by taking the number of grid cells at a profile point divided by the total number of grid cells representing the whole hillslope), and all supplemental data. This reduction is based on the work of Cochrane and Flanagan (2003) for the WEPP model. *(iv) prof_class()* classifies the representative catenae into *landscape units* (LUs) (coloured areas in Fig. 3.2). In this step similar catenae are identified and lumped together based on the calculated properties employing an unsupervised K-means clustering method. The user has to specify the number of classes to generate from each attribute during the clustering, which is done separately for each attribute. The final class assignment for each catena results from the combination of these attribute-wise classifications. Each LU is then further sub-divided into *terrain components* (TCs), i.e. planar elements representing, e.g. upper, middle, and downslope parts of the lumped catena (coloured diagrams in Fig. 3.2). The number of TCs to be generated for each LU can be specified and the partition is done by evaluating the derived LU properties and employing a minimisation of variances approach. Topological relations between SVCs, TCs, and LUs are established, expressed as percentages of covered area and along-slope location of TCs within a LU rather than spatial coordinates. For visual inspection, the user has the option to let the functions generate plots during steps (iii) and (iv). The employed algorithms for steps (iii) and (iv) are explained in more detail by Francke et al. (2008).

Finally, *(v) lump_grass_post()* establishes the topology between subbasins and LUs, again expressed as the percentage of covered area. Subbasins are the only spatial units with explicit reference in terms of geographic coordinates. In addition, subbasin and LU-specific parameters such as representative channel geometry and routing parameters, groundwater, and landscape coefficients, are approximated. Hereby, rather simple relationships or typical standard values are employed. The output of this function is mainly designed to provide a complete plausible parameterisation. Where alternative information is available, it should be used. See the function's documentation for more details.

A summary of the most important parameters for the landscape discretisation process is given in Tab. 3.2. Their meaning along with a sensitivity analysis will be further discussed in Sect. 3.4.4.

3.3.3 Additional tools

In order to meet further capabilities of the WASA-SED model, functions *reservoir_*()* have been introduced. They facilitate the pre-processing of reservoir-specific input files for the model using spatial reservoir data and pre-compiled parameterisations. The additional function *rainy_season()* calculates start and end dates of the rainy season for every year based on a time series of daily precipitation values using a statistical approach described by Gerstengarbe and Werner (1999). *calc_seasonality()* then uses the output of the former function and information about seasonal variation of a vegetation parameter to calculate a daily time series of that parameter by linear interpolation of the parameter's node points depending on the current start and end dates of the rainy season for a specific year. In hydrological models such as

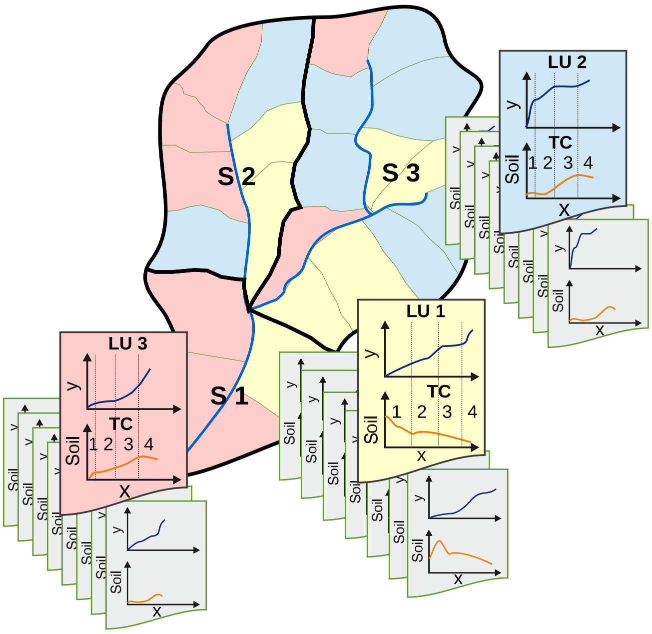

Figure 3.2: Example of the outcome of steps (i) to (iv) of landscape discretisation. Shown is a hydrological basin sub-divided into subbasins (labelled S 1–3), EHAs (small polygons), and LUs (red, blue, yellow) with TCs. Along-slope catena properties for the EHAs and LUs are summarised by the diagrams.

WASA-SED or WaSiM-ETH (Schulla and Jasper, 2007), such information can be used to describe intra-annual variations of vegetation parameters.

3.3.4 Parameter database

To store the results of landscape discretisation and allow easy maintenance, manipulation, and the generation of necessary model input files, all processing results are stored in a database. For flexibility, lumpR currently supports multiple DBMS, including MS Access, MySQL / MariaDB, and SQLite. Details for database configuration are provided in the wiki of lumpR's web page (see Sect. 3.6). For interacting with the database, the package provides a set of functions. Again, these functions should be applied in the recommended order as indicated in Fig. 3.1 and will be further explained in the following.

(i) db_create() creates an empty parameter database. As the package underlies continuous further development, *(ii) db_update()* ensures backward compatibility to previous database versions. *(iii) db_fill()* assimilates the output of the landscape discretisation steps and other pre-processing not included in the package (e.g. readily prepared soil and vegetation parameters) and imports the data into the respective tables of the parameter database. *(iv) db_check()* performs a number of checks to identify and (if possible and desired) automatically resolve inconsistencies and/or missing information in the database. These include filtering of tiny and spurious areas (e.g. spatial entities smaller than a specified threshold) to reduce computational overhead, checking that all TCs have a slope larger than zero, defining special areas for separate treatment (this currently includes SVCs marked as water or impervious), removing special areas marked as water, computing fractions of impervious surfaces at TC level, removing impervious surfaces, estimating a storage coefficient for groundwater delay at LU level, deleting obsolete datasets (i.e. unused spatial entities), checking for completeness (all IDs in the `*_contains_*` tables exist within the respective referenced tables), and computing the subbasin order from upstream to downstream. The user can decide, which checks to perform, how to deal with inconsistencies, and define thresholds for certain checks. For reproducibility, any changes to the database will be logged in table `meta_info`.

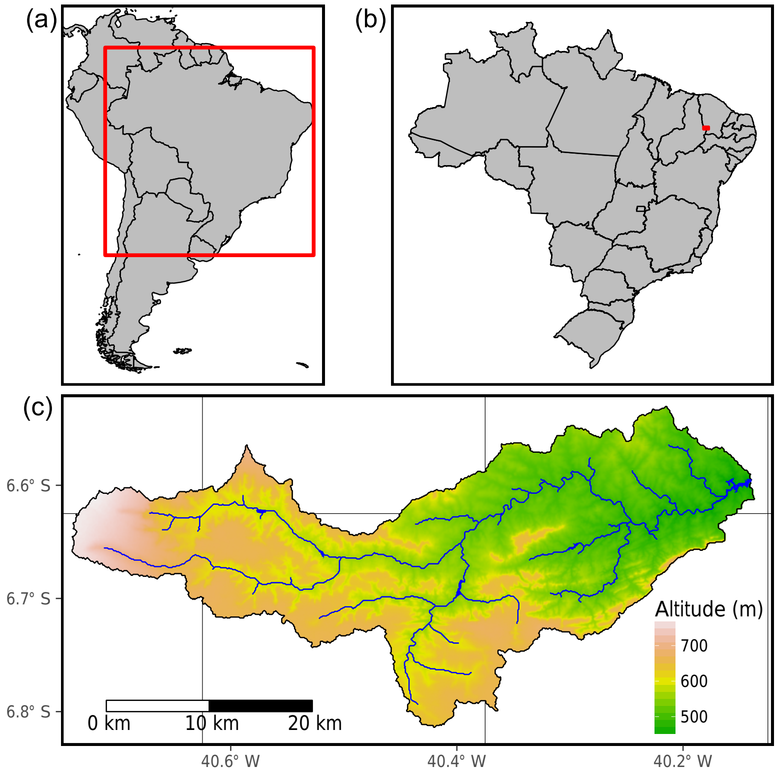

Figure 3.3: Overview over the Benguê catchment (c) and location within Brazil (b) and South America (a). Black lines in the background of the lower panel correspond to the grid of the meteorological dataset of Xavier et al. (2016) that was used in this analysis (see Sect. 3.4.3).

At any point during the processing the user can freely inspect and adjust the parameter database by means other than the functions provided by lumpR. In the current version, the package provides the function *db_wasa_input()* to convert the values of the parameter database into input files for WASA-SED. However, the user may as well export the values needed and compile the input files for any other model. Furthermore, upon request, export function for other models can easily be added.

3.4 Example application and sensitivity analysis

Within an example application, multiple realisations of hillslope discretisations of the same model domain are generated with lumpR version 2.0.0 by varying five parameters that control the creation of model entities on different scales. These realisations are then used with the hydrological model WASA-SED and their effect on the model output is analysed. In the following subsections, the study site, the WASA-SED model and data used for model initialisation, the sensitivity analysis of the discretisation parameters, and, finally, the results of this study are described.

3.4.1 Study site

To demonstrate the functionalities of lumpR the Benguê catchment was selected (Fig. 3.3). It is part of the upper Jaguaribe River catchment in the northeast of Brazil within the federal state of Ceará. The area has been investigated in a number of studies, in many cases employing the WASA-SED model, and was thus selected to ensure the suitability of the model for the catchment (de Araújo and Medeiros, 2013; Bronstert et al., 2014; de Figueiredo et al., 2016; Krol et al., 2011; Medeiros and de Araújo, 2014; Medeiros et al., 2014, 2010).

The Benguê catchment drains an area of about 926 km^2. At its outlet, the ephemeral Umbuzeiro River disembogues a reservoir built in 2000 with a storage capacity of 19.6 million

m^3 to enhance water supply and reliability in the region. As the area is located within the "drought polygon" of Brazil, annual average precipitation is low with about 600 mm in comparison to a potential evapotranspiration of more than 2000 mm. The mean annual temperature is 25 °C with little variation. Climate is further characterised by a strong intra-annual variation of precipitation leading to distinct rainy (January to May with more than 80 % of annual rainfall) and dry seasons. Rainfall is mostly convective and concentrated in only a few high-intensity events per year.. Inter-annual variation of precipitation, however, is also high causing recurrent droughts which in cases may last over several consecutive years. The dominant natural vegetation is Caatinga consisting of deciduous bushland with xerophytic species ranging from dense dry forests to almost desert-like sites. The environment is further characterised by mainly sedimentary plateaus in the southern and western parts of the study site with steep terrain and deep (> 1 m) permeable soils, predominantly Latosols. In the north and east crystalline bedrock is prevailing with shallow Luvisols (< 1 m) causing high runoff coefficients. In alluvial zones, Planosols are dominant. The population density is low (6.4 inhabitants per km^2) with rural lifestyle. Parts of the area are used for small-scale farming including cattle breeding and growing of maize and beans in particular.

3.4.2 The WASA-SED model

The WASA-SED model, revision 247 from 29 September 2016, is used to study the effects of different landscape discretisations realised with lumpR on simulated streamflow dynamics. WASA-SED is a deterministic, process-based, semi-distributed, time-continuous hydrological model. The model was first introduced by and is described in detail within Güntner (2002), with special focus on its application in semi-arid environments. It has been frequently employed in semi-arid areas such as northeastern Brazil (including the Benguê catchment; for references see Sect. 3.4.1), India (Jackisch et al., 2014) and Spain (Bronstert et al., 2014; Mueller et al., 2009, 2010).

The model incorporates the Shuttleworth–Wallace approach for evapotranspiration calculation over sparsely vegetated surfaces and an infiltration approach based on Green–Ampt accounting for Horton-type infiltration. Via the complex hierarchical spatial disaggregation scheme (see Sect. 3.2.2), lateral redistribution processes as well as re-infiltration along a hillslope are considered while the model can still be applied over large scales (up to the order of magnitude of 100,000 km^2). Large strategic reservoirs can be represented in an explicit manner while smaller ones are treated as lumped water bodies of different size classes to efficiently account for water retention of many small reservoirs in a study region. The model has been subsequently expanded, e.g. to account for sediment dynamics and was renamed from WASA to WASA-SED (Mueller et al., 2010). It should, however, be noted that in this study erosion is not modelled.

3.4.3 Data and model set-up

Meteorological data in daily resolution to drive WASA-SED, in particular precipitation, air temperature (obtained by minimum and maximum temperatures by simple averaging), relative humidity, and incoming shortwave radiation, have been derived from the dataset described by Xavier et al. (2016). This gridded dataset is based on station information that has been checked, corrected, and interpolated to a grid with 0.25° x 0.25° resolution. For the analysis, data from 20 grid cells were considered, whereas the catchment itself directly intersects with six cells (see Fig. 3.3). Extraterrestrial radiation as a further driver that has been calculated from astronomical relationships employing the R package *sirad*. Meteorological data have been interpolated to the locations of subbasin centroids using an inverse-distance weighting approach by employing the R package *geostat*.[1]

The basis of terrain analyses was the 90 m x 90 m SRTM DEM. As pre-processing step, the raw raster was sink-filled employing the GRASS function *r.terraflow*. Soil information has been derived from the database of Jacomine et al. (1973). The parameters needed by WASA-SED have been inferred by pedo-transfer functions based on soil texture information using the R

[1] This package is not on CRAN but available via the ECHSE tools library from `https://github.com/echse/echse_tools`.

Table 3.2: User decisions affecting landscape discretisation complexity and their realisations used for sensitivity analysis.

Identifier	Meaning	Realisations
SUB_thresh	Minimum size of subbasins in number of grid cells	1000, 2000, 5000, 10000, 30000
EHA_thresh	Minimum size of EHAs in number of grid cells	25, 50, 100, 200, 500, 750, 1000
LU_no	Maximum number of LUs to be classified	5, 10, 20, 50, 75, 100, 150, 200, 250, 300
LU_atts	Number of attributes to be considered during LU classification	1...7
TC_no	Number of TCs to be deviated for every LU	1...5

packages *soilwaterfun* and *soilwaterptf* available via `http://soilwater.r-forge.r-project.org/`. Vegetation parameters for the types occurring within the study area have been elaborated during the development of WASA (Güntner, 2002) and have been adopted for this study. An updated shapefile of landcover distribution was obtained from the Brazilian Ministry for the Environment and the land cover classes have been reclassified to those used in Güntner (2002). Data on reservoirs and the geology of the area had been collected and processed within the SESAM project; see `http://www.uni-potsdam.de/sesam`.

No calibration of any model parameters has been done as a comparison with observational data shall not be part of this study. As has been mentioned, however, the model has been used in and proved its ability for the catchment (see references given in Sect. 3.4.1).

3.4.4 Sensitivity analysis of landscape discretisation parameters

During the process of landscape discretisation, a user commonly has to make a number of – often subjective – decisions. These are directly influencing the complexity of the discretisation and thereby affecting computational efforts and, possibly, model results. Although this issue has been acknowledged (e.g. Ajami et al., 2016; Fenicia et al., 2016), we are not aware of any study systematically analysing this effect for hillslope-based approaches and at multiple spatial scales at the same time. This may be mainly because of the associated manual effort and computational burden, which has become accessible using lumpR. Its fully automatic integration allows for conducting a comprehensive numerical experiment, reflecting the complexity and multi-dimensionality in the discretisation process. Henceforward, we consider these as parameters within a model sensitivity analysis.

Experimental set-up

For landscape discretisation using lumpR, five parameters reflecting the most important user decisions have been identified and are summarised in Tab. 3.2. Their presented realisations are based on expert knowledge. They result from a reasonable range of values while striving for maximum possible variation in the generated spatial units. Therefore, for some parameters such as `SUB_thresh`, also non-uniform distributions of the values have been taken into account. What follows is a reasoning on selected parameter realisations for the experiments that can as well be used as guidelines for lumpR applications.

`SUB_thresh` and `EHA_thresh` are size thresholds affecting the size and thereby the number of delineated subbasins and EHAs, respectively. As their realisations are given in number of grid cells, their choice depends on the resolution of the GRASS location which should be oriented on the DEM (here, the SRTM DEM resolution of 90 m x 90 m), and catchment size. `LU_no` and `LU_atts` control the process of clustering EHAs into LUs. The maximum possible value for `LU_atts` depends on the number of attributes that can be used for classification. These, by defaults, include the shape of EHAs, their horizontal and vertical extension, and a proxy for hillslope width which are all inferred from a DEM. Further supplemental attributes can be added which in this study included maps of soil types, land cover, geology, and SVCs that resulted in a total of seven attributes. In this study, different realisations of `LU_atts` thus simulate a

Table 3.3: Streamflow indices used as scalar response functions for sensitivity analysis.

Symbol	Index	Calculation	Unit
RR	Runoff ratio	Sum of daily streamflow values divided by sum of daily precipitation over the whole period of analysis multiplied by 100	%
P_{flow}	Probability for significant streamflow	Number of days with significant[a] streamflow divided by total number of values multiplied by 100	%
Q_{avmax}	Average annual maximum flow	Average over all annual maximum streamflow values	$m^3\,s^{-1}$
SFDC	Slope of flow duration curve	Average slope of the flow duration curve for significant[a] medium ranged[b] streamflow values; high values stand for a more variable whereas low values represent a more damped flow regime (Sawicz et al., 2011)	dimensionless
f_{low}	Frequency of low flows	Average number of insignificant[a] flow events[c] per year	$year^{-1}$
f_{high}	Frequency of high flows	Average number of high flow[d] events[c] per year	$year^{-1}$
RC_{rise}	Rate of change during rise	Average rate of change of the rising limbs of high flow[d] events[c]	$m^3\,s^{-1}\,day^{-1}$
RC_{fall}	Rate of change during fall	Average rate of change of the falling limbs of high flow[d] events[c]	$m^3\,s^{-1}\,day^{-1}$

[a] (In-) significant streamflow defined as those values (less than or equal to) larger than 0.01 $m^3\,s^{-1}$.
[b] Values between the 33 % and 66 % percentiles.
[c] An event is defined as a period of consecutive days a certain condition is fulfilled.
[d] High flows are those values being larger than a flow threshold which is herein defined as the 90 % percentile of all significant[1] flow values from all 12,250 model realisations during the analysis period.

differing degree of information available for the deviation of LUs. For `LU_atts` less than seven, the aforementioned attributes were sampled randomly. `LU_no` defines the maximum number of LUs to be generated. As the LU classification is done successively for each attribute, this number results from the product of the number of classes N_i specified for each of the `LU_atts` considered attributes i:

$$\mathrm{LU\_no} = \prod_{i}^{\mathrm{LU\_atts}} N_i \tag{3.1}$$

Thus, conversely, when `LU_no` is pre-specified, the values of N_i need to be determined under the above-mentioned constraint as follows: one of the considered attributes is randomly selected and its N_i increased by one. This procedure is repeated until Eq. 3.1 is satisfied, i.e. the actual number of LUs, is greater than or equal to `LU_no`. Finally, `TC_no` is the number of TCs that will be delineated for every LU.

For the sensitivity analysis all possible combinations of parameter realisations were employed which resulted in a total of 12,250 realisations of discretisations. These comprise varying complexities within all spatial levels and different degrees of data availability. Finally, WASA-SED was run with each realisation over a 13-year period, where the first 5 years were considered as warm-up and thus have been excluded from the analysis.

Scalar model output

The target variable of the analysis is the time series of simulated daily river contributions to the Benguê reservoir located at the catchment outlet. However, for conducting the desired sensitivity analysis, a scalar target function is needed. As it is impossible to summarise all important characteristics of a streamflow time series in a single scalar value, we employed multiple indices and performed the sensitivity analysis for each index separately. The indices are presented and described in Tab. 3.3. The indices were chosen to describe a wide range of aspects of streamflow behaviour ranging from the magnitude of flow (RR, P_{flow}, Q_{avmax}) over flow regime (SFDC) to frequency (f_{low}, f_{high}) and runoff concentration time (RC_{rise}, RC_{fall}).

Analysis method

Numerous approaches for sensitivity analysis exist (Pianosi et al., 2016). Their choice depends on the objective of the study, the nature and complexity of the model, its parameters and outputs, and available computing resources.

The goals of this analysis were, first, a *ranking* of the described descritisation parameters in terms of their influence (sometimes also referred to as *priorisation*) and, second, the identification of those parameters with negligible influence on the respective streamflow index (also referred to as *screening* or *fixing*). The above-mentioned sampling procedure for the parameters corresponds to a *global* sensitivity analysis with *all-at-a-time* sampling. This allows a variance- or density-based approach Pianosi et al. (2016). The former is based on the calculation of sensitivity indices based on the variance of the response function. This, however, requires the assumption that the variance is a good proxy for describing the variation of the value range. For multi-modal or highly skewed distributions this cannot be guaranteed. In such a case Pianosi et al. (2016) recommend density-based methods. Rather than the variance alone this family of sensitivity analyses considers the probability density function of the response surface.

For the above-mentioned reasons, we chose the recently introduced PAWN method by Pianosi and Wagener (2015) as this density-based method can cope with skewed distributions and is relatively easy and straightforward to implement. PAWN uses empirical approximations of the unconditional cumulative distribution function $F_{y_i}(y_i)$, with y_i being one of the eight streamflow indices selected as scalar response functions over all 12,250 realisations, and the conditional cumulative distribution functions $F_{y_i|p_j}(y_i)$ where a certain parameter p_j is fixed at a specific value. The PAWN index T_j as a sensitivity measure can then be calculated for each parameter employing a numerical approximation of the Kolmogorov–Smirnov statistic KS following

$$\text{KS}(p_j) = \max_{y_i} \left| F_{y_i}(y_i) - F_{y_i|p_j}(y_i) \right| \tag{3.2}$$

and

$$T_j = \underset{p_j}{\text{median}} \left[\text{KS}(p_j) \right] \tag{3.3}$$

T_j varies between 0 and 1, where low values of T_j identify the less influential parameters. For parameter screening the two-sample Kolmogorov–Smirnov test was employed. It calculates a critical value KS_{crit} above which a parameter is significant as its conditional cumulative distribution function differs significantly from the unconditional one at a certain confidence level α:

$$\text{KS}_{\text{crit}} = c(\alpha) \sqrt{\frac{n+m}{nm}} \tag{3.4}$$

where n and m are the number of samples to estimate $F_{y_i}(y_i)$ and $F_{y_i|p_j}(y_i)$, respectively, and taking the tabulated value of $c(\alpha) = 1.36$ for an $\alpha = 0.05$.

3.4.5 Results

Instead of the simulated river discharge (i.e. the model output used to calculate the streamflow indices, see Sect. 3.4.4), Fig. 3.4 provides an overview over the simulated reservoir storages in comparison to observations for the Benguê reservoir at the catchment outlet. We chose the latter because of the very episodic characteristic of the river discharge while the volumes, for visual comparison, constitute a more informative representation and are more directly related to available measurements. Figure 3.4 furthermore shows the catchment's areal precipitation used to drive the model. The absolute deviations of these uncalibrated model runs from

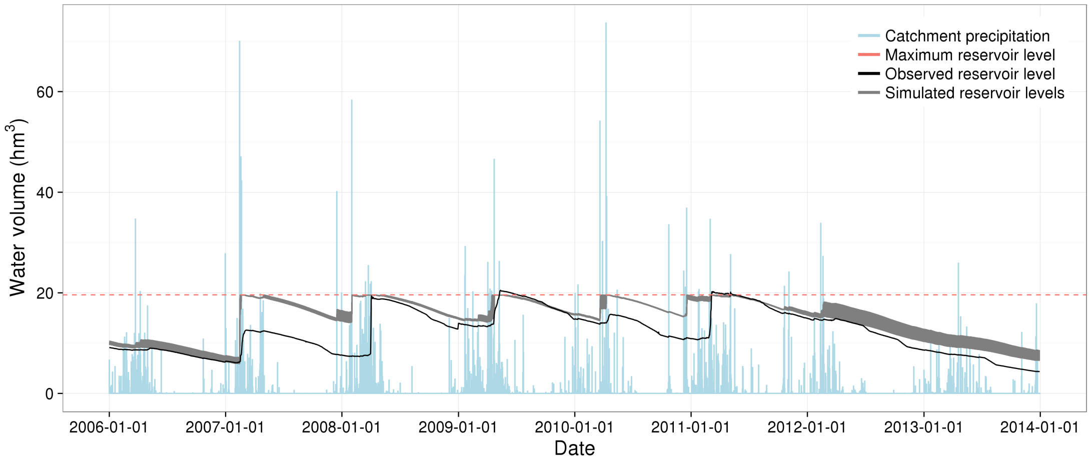

Figure 3.4: Time series plot of daily resolution of simulated (all 12,250 parameter realisations) and observed Benguê reservoir volume and areal precipitation. Note that the maximum reservoir level corresponds to the official value which has been used for model parameterisation. Observations sometime show slight deviations due to, e.g. random measurement errors or changed rating curves.

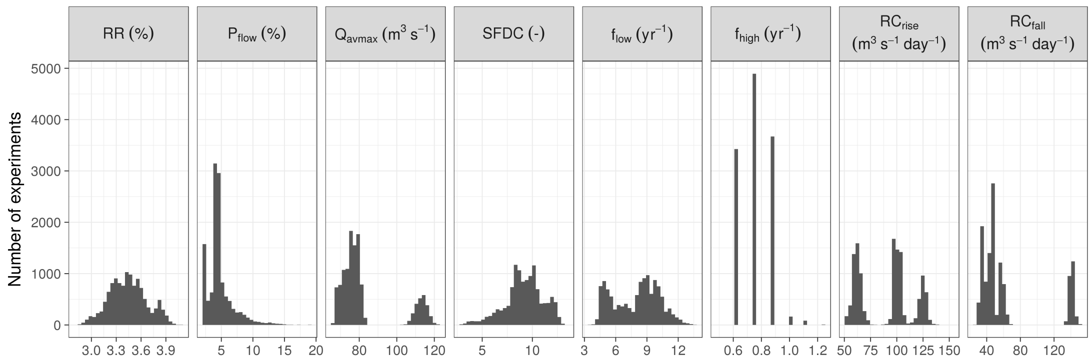

Figure 3.5: Histograms showing the value distributions of streamflow indices over the 12,250 realisations.

the observations are, in parts, considerable, whereas qualitative behaviour is matched well. However, in some years (2008 and 2011) the model simulates reservoir filling within the rainy season much earlier than observed. Simulated reservoir depletion is often slower than it can be observed which might be the result from an imperfect parameterisation of reservoir abstractions for water consumption. Overall it should be noted that variability caused by different parameter realisations is rather small. Deviations mainly appear in a way that different discretisations result in slightly different amounts of generated runoff and, as a consequence, different reservoir level changes. Naturally, this cannot be observed during runoff events causing the maximum reservoir level to be exceeded in which additional runoff is lost as reservoir overspill. Furthermore, it can be observed that a rainfall volume of at least about 35 hm^3 (which is almost equal to 35 mm) seems to be necessary to produce noticeable reservoir inflow.

Figure 3.5 gives an overview of the streamflow index value distributions from the 12,250 realisations. Some indices, namely Q_{avmax}, $\mathrm{RC}_{\mathrm{rise}}$, and $\mathrm{RC}_{\mathrm{fall}}$, show distinct multi-modal value distributions. Distributions for the other indices are slightly bi-modal (f_{low}) or skewed (P_{flow} and SFDC) with RR as the only index being more or less normally distributed. In general, the runoff coefficient RR is consistently low ranging between 3 % and 4 % and streamflow can be characterised as ephemeral due to a low probability of days showing significant streamflow (P_{flow} about 5 % for most of the experiments). Opposed to low flow periods, high flow events do rarely occur. The variability of f_{low} is relatively high whereas f_{high} shows relatively low variance.

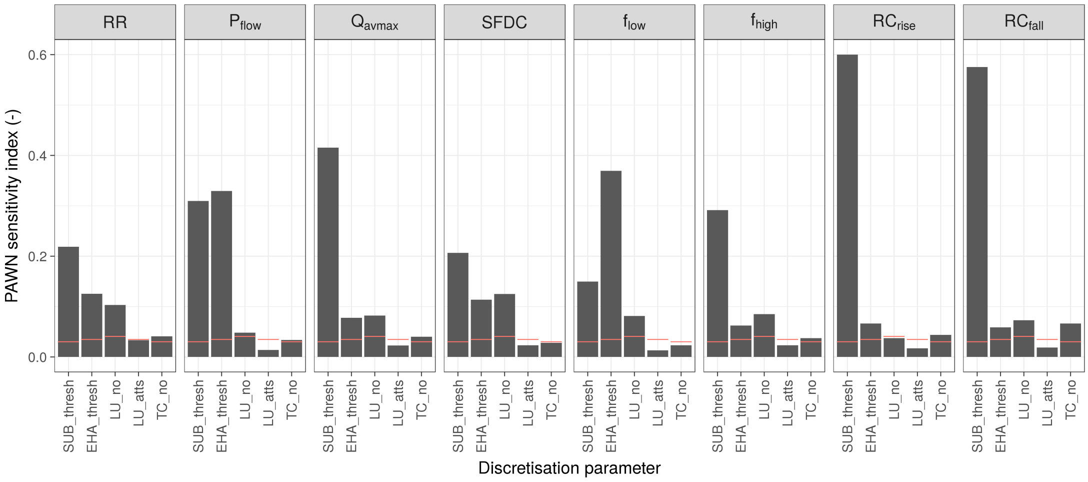

Figure 3.6: Barplots of PAWN sensitivity indices T_j for each streamflow index and landscape discretisation parameter. Red lines indicate critical Kolmogorov–Smirnov values (KS_{crit}); i.e. parameters, for which T_j is not greater than this value, can be regarded as insignificant for the respective streamflow index.

Peak flows (Q_{avmax}) as well as runoff concentrations (characterised by RC_{rise} and RC_{fall}) vary considerably between the realisations also exhibiting multi-modal distributions.

Screening and ranking of landscape discretisation parameters is illustrated in Fig. 3.6. The size of subbasins (`SUB_thresh`) is the most influential parameter for all indices except for those being related to low flow characterisation (f_{low} and P_{flow}) which are dominated by the size of EHAs (`EHA_thresh`). The number of LUs (`LU_no`) can be regarded as the third important parameter being especially of relevance for high flow-related indices (f_{high} and Q_{avmax}) and, to some extent, flow regime (SFDC) and runoff concentration (RC_{fall}). The least important parameters are the number of TCs (`TC_no`), except for runoff concentration (RC_{rise} and RC_{fall}), and the number of attributes considered for LU classification (`LU_atts`), which is insignificant for all streamflow indices.

More information on parameter influences on the various streamflow indices can be obtained from Fig. 3.7. The more influential a parameter, the larger the deviations of the conditional empirical cumulative distribution functions from the unconditional one. Many of the diagrams show a clear relationship between parameter realisation and streamflow index value. The larger the subbasins (i.e. the larger `SUB_thresh` and the lower the number of subbasins) the smaller the generated amount of runoff (RR gets smaller) and the smaller the probability of significant runoff (P_{flow}). On the other hand, peak discharges (Q_{avmax}) increase and the catchment appears to produce more rapid runoff responses (higher values of RC_{fall} and RC_{rise}). Furthermore it can be seen that `SUB_thresh` is responsible for the multi-modal distributions of Q_{avmax}, f_{high}, RC_{fall}, and RC_{rise} as the conditional distribution functions show a less stepped shape than the unconditional functions. The influence on f_{low} appears to be less clear. It can be seen, however, that larger values of `SUB_thresh` result in a more pronounced bimodal distribution of that index. Low flows are otherwise more dominated by the size of EHAs (`EHA_thresh`). The larger the `EHA_thresh` (i.e. the larger the EHAs and the lower their number) the less the probability for significant streamflow (P_{flow}) and the less the number of low flow events per year (f_{low}) while the flow regime becomes more variable (higher values of SFDC). Higher numbers of LUs (`LU_no`) result in more generated runoff (RR increases) with a tendency to both higher frequencies of low flow and high flow events, and generate more variable streamflow regimes.

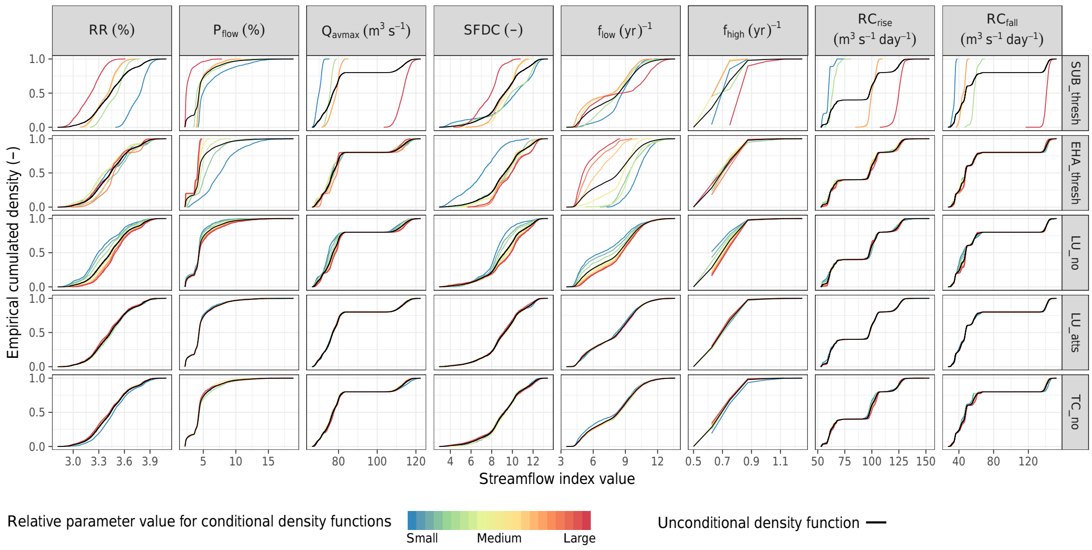

Figure 3.7: Unconditional (black lines) and conditional (rainbow-coloured lines) empirical cumulative distribution functions for each streamflow index–discretisation parameter combination.

3.5 Discussion

3.5.1 lumpR: features, benefits, limitations

It is still common practice for many researchers in the field of hydrological modelling to not automate their pre-processing steps. Even if they do, the related scripts are rarely published along with the studies. Furthermore, common limitations of existing software are that they are often model specific, and/or perform only certain steps of pre-processing. Some tools are commercial or can only be used along with commercial software (e.g. ArcGIS or MATLAB).

With lumpR, a package for the free and open-source programming language R has been developed that addresses these limitations and build on the philosophy of FOSS. So far, the software has been tested under Windows as well Linux-based operating systems (openSUSE and Ubuntu). lumpR interacts with the GIS GRASS and thus allows for graphical investigation and manual correction of outcomes. As R is a widespread scripting language, model pre-processing in that way can easily be customised, automated, and reproduced. Via the database tools, the software allows one to keep the output directories clearly arranged by putting all information into a database. Several database systems are supported.

In this study, lumpR's functionality was demonstrated together with the model WASA-SED. However, the package provides a function to create the model's input files, which can by easily adapted to the requirements of other hillslope-based models. First tests employing the simulation environment ECHSE (Kneis, 2015) revealed the package to be easily adaptable for producing the input files needed by a different model. Other candidate models are, in principal, all hydrological models with similarly complex spatial aggregation schemes as used by WASA-SED that can make use of the information provided by lumpR, such as the WEPP model (Flanagan and Nearing, 1995).[2]

The centrepiece of the package is the LUMP algorithm introduced by Francke et al. (2008) for the calculation of hillslope properties and the delineation of representative LUs together with the subdivision into TCs. In contrast to simple GIS overlay techniques, as usually employed for the delineation of HRUs, it preserves information on the distribution of hillslope parameters and their relative topographic position, i.e. their downslope connection. In that way, the task of hillslope-based landscape discretisation and parameterisation can be purposefully directed to landscape properties dominating hydrological response. Via its integration into lumpR, the application of the algorithm could be further simplified and harmonised as the original LUMP consisted of

[2]Readers considering lumpR for use with their model are encouraged to contact the corresponding author of this paper for support.

a loose collection of less user-friendly scripts, partly also relying on non-free software such as MATLAB. In that way, the package hides unnecessary detail from the user while at the same time ensuring a certain level of control over the discretisation process. For instance, in the present study, the use of morphological parameters was limited to shape, horizontal and vertical extension, and hillslope width. The lumpR package is, however, flexible enough that a user can include further parameters as supplemental information for the classification process. This might include the use of other factors relevant for hillslopes characterisation such as, for instance, contour curvature being related to the convergence or divergence of flow paths and as such being of hydrological relevance (Bogaart and Troch, 2006). Furthermore, lumpR comes with a lot more additional functionalities than the mere LU and TC deviation.

Thanks to lumpR allowing for a high degree of automation, for the first time a multi-hierarchy sensitivity analysis of discretisation parameters in a hillslope model could be conducted. Within this analysis, a high number of discretisations of varying complexity were easily produced. Thus, a user can experiment and find out the optimal degree of complexity for a certain catchment and a specific objective, e.g. by systematically employing a multiple hypothesis framework. For the presented case study, the analysis revealed the pronounced influence of the size of the subbasins and the EHAs on various aspects of the hydrograph. All other discretisation parameters showed no or considerably less influence.

During the testing phase, some shortcomings of lumpR were identified. With regards to the applicability over large datasets, i.e. when applying the package to large areas in the order of > 100,000 km^2 and/or when employing high-resolution DEMs, time consumption might pose a restriction, although lumpR already uses parallelised code in the most critical steps. Therefore, future enhancements also need to include further improvements regarding computational efficiency.

A limitation more related to the algorithm is that the software is not able to automatically distinguish and account for artificial hydrological discontinuities. This includes, e.g. ditches and field boundaries or other problematic formations such as large flat areas in a DEM as produced by lakes. While the former pose restrictions on the general applicability of the hillslope approach, the latter need to be masked in GRASS before the analysis. In addition, some of the pre- and post-processing steps within lumpR (i.e. functions *lump_grass_prep()* and *lump_grass_post()*) still employ rather simplistic approaches. This affects in particular the deviation of the river network, subbasin delineation, and the approximation of streamflow routing parameters (the latter tailored to the rather simplistic unit hydrograph approach of WASA-SED). In this respect, future enhancements should also include a review on latest advancements of terrain analysis and parameterisation and the refinement of employed algorithms.

3.5.2 On the sensitivity analysis of discretisation parameters

As in science reproducibility and objectivity are primary criteria for any investigation, it has to be noted that any model discretisation is subject to a certain degree of subjectivity. Especially for hillslope-based discretisation, this can cover several hierarchy levels. Consequently, the effects of these choices on the model output have been assessed via the sensitivity analysis within the example application and its results shall be discussed in the following.

The results from the 12,250 realisations of landscape discretisation show only little difference with respect to water storage dynamics of the Banguê reservoir (Fig. 3.4). The small variation of the runoff coefficient (see Fig. 3.5) further supports the conclusion that decisions on landscape discretisation parameterisation only have a minor impact on simulated runoff volume for the given case. On the other hand, the influence on other indices describing runoff concentrations and dynamics, and the frequency of flood or drought events is much more obvious.

The hydrological regime of the study area is primarily influenced by precipitation, which is characterised by a high temporal concentration and a large temporal and spatial variability. A comparison with the data reported in Medeiros and de Araújo (2014) further supports that uncertainties regarding the precipitation input to the model have a much larger impact on simulation results than the discretisation parameters. In their study, Medeiros and de Araújo (2014) used a set of raw station data in contrast to the pre-processed and gridded dataset by Xavier et al. (2016) used for our experiments and their runoff values have been assessed by taking the Benguê reservoir inflows computed from water balance calculations. Their reported runoff coefficients are mostly lower than ours, even when precipitation is higher, and shows less inter-annual variation (see Fig. 3.A1 in Appendix).

The precipitation forcing for the current study was implicitly slightly influenced by variable subbasin sizes and numbers, as the precipitation was specified at the subbasin level. The variability in precipitation input among the realisations, however, appears to be negligible as it is generally less than 6 mm for daily values and less than 10 mm for yearly sums (see Fig. 3.A2 in Appendix).

Our simulation results show a general overestimation in comparison to measured values (Fig. 3.4). Despite the mentioned uncertainties arising from the precipitation input, there are some other possible factors that could have led to the observed mismatch: uncertainties in the reservoir parameterisation in the model (e.g. we use a static parameterisation of reservoir abstractions which are, in reality, dynamic); uncertainties in the observations (e.g. due to deficiencies or changes of the rating curve); model parameterisation uncertainty (the model has been run with standard parameterisation for the area without further calibration, see Sect. 3.4.3). Regarding the parameterisation it should furthermore be noted that the different discretisations did not directly affect soil nor land-cover parameters. They merely modified the fractions of soil and vegetation types that are assigned to the spatial units.

Considering the method of sensitivity analysis it can be concluded that the choice for a density-based approach was reasonable as most of the analysed streamflow indices exhibit multi-modal or skewed value distributions. A drawback of the analysis approach is that only first-order effects of parameter sensitivities have been quantified while interactions among parameters have been neglected. It might thus be that insignificant parameters (i.e. the parameter `LU_atts`) have significant higher-order effects due to parameter correlations. With respect to the discretisation parameterisations, it can be argued that the chosen parameter realisations are both subjective and case study specific. On the other hand, all parameter realisations show a monotonous effect on the streamflow indices; i.e. when increasing a sensitive parameter the streamflow index values increase or decrease monotonously (see Fig. 3.7). This suggest a continuous response of the parameters, facilitating some transferability of the results.

Overall, the example study and sensitivity analysis is catchment and model specific. Strictly speaking, conclusions are thus limited to applications of the same model under similar hydro-climatic conditions, i.e. semi-arid areas without substantial groundwater influences mainly characterised by spatially and temporally heterogeneous precipitation patterns. It remains an open question whether the use of a different model and/or the application in a catchment with distinct environmental and climatological characteristics and/or different dominant runoff generation mechanisms would lead to other conclusions. This paper presents a novel framework along with an example application to address these questions in future studies.

3.6 Conclusions

The goal of this study was to introduce a new software for landscape discretisation in semi-distributed hydrological modelling. Thereby, three objectives have been pursued.

First, we provided a short review of existing landscape discretisation algorithms and software solutions. The number of existing concepts and corresponding tools was found to be large, making it a difficult task to choose a specific approach and software. Besides grid-based approaches, the most common strategies for semi-distributed hydrological modelling focus on the delineation of spatial entities with homogeneous process dynamics, such as the frequently implemented HRU approach. Approaches directly concentrating on the description of a hillslope as central modelling unit or pursuing hierarchical multi-scale frameworks as efficient solutions for large-scale application are less in number. In addition, existing programmes implementing a specific discretisation often exhibit various limitations, e.g. they are model specific, commercial or employ commercial back-end software, or allow only a limited or no automation of workflows.

Second, we developed and presented a new software called lumpR as a package for the open-source environment R. It was designed to implement a hillslope-based hierarchical multi-scale discretisation of landscapes, including the delineation of subbasins, the derivation and lumping of hillslopes, and the subdivision of the latter into terrain and soil–vegetation components. The package thereby connects to GRASS GIS, directly using prepared spatial information and writing spatial output into an initialised location for immediate inspection. Furthermore, database functionalities have been included to manage the outcomes of the discretisation process. lumpR overcomes existing limitations in a way that it easily allows one to include different hillslope-based models, it is completely free and open source, and it facilitates the automation of workflows. At the same time, however, it is retaining a sufficient degree of freedom to the user via the selection of parameters, and the inclusion of expert knowledge and additional information.

Third, the functionality of the package was shown in a case study in the semi-arid northeastern Brazil, employing the hydrological model WASA-SED. Thereby, the workflow automation allowed a systematic sensitivity analysis of crucial landscape discretisation parameters. Regarding multiple streamflow metrics, the model appeared to be reasonably robust to the dicretisation parameters. The size of subbasins and delineated hillslopes were found to be the most influential factors. The number of landscape units (i.e. lumped hillslopes) and the further subdivision into terrain components appeared to be less important, making the amount of information included in the hillslope lumping process being even completely insignificant.

The R package turned out to be an efficient and user-friendly tool for the automation of landscape discretisation for hillslope-based large-scale hydrological models. Future work, on the one hand, will focus on comparing uncertainties arising from discretisation to other sources of uncertainties. On the other hand, in order to obtain more general conclusions, the presented sensitivity analysis of landscape discretisation parameters needs to be extended to other catchments within different environmental and hydro-meteorological conditions as well as other hillslope-based models. Technical extensions will include the integration of further models, improvement of time consumption and memory handling for application in large areas $> 100{,}000\,\mathrm{km}^2$, consideration of artificial discontinuities and mechanisms for large flat areas, refinement of certain parameter estimation approaches, and testing the package for other hydro-meteorological and environmental conditions.

Code availability

Code for lumpR is freely available at `https://github.com/tpilz/lumpR`. The Latex code to reproduce this paper including R code to reproduce all analyses and figures is available at `https://github.com/tpilz/lumpr_paper`.

Data availability

Meteorological data are available from `http://careyking.com/data-downloads/`. DEM raw data can be obtained via `http://srtm.csi.cgiar.org/SELECTION/inputCoord.asp` by selecting tile 28/14 (horizontal/vertical). Reservoir data and the geology map have been processed within the SESAM project and are not publicly available. For more information and contact details see `http://www.uni-potsdam.de/sesam`. Land cover and soil raster maps are not publicly available.

Author contributions

All authors contributed to the idea and methodology. T. Pilz and T. Francke developed the lumpR package (more specific contribution information are included in the source code files). Experiments and analyses have been conducted by T. Pilz with support by T. Francke. The manuscript was prepared by T. Pilz with contributions from all co-authors.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

Authors thank Fernanda Scholz (former M.Sc. student at the Leibniz University of Hannover) for the preparation of soil data. Student Lisa Berghäuser from the University of Potsdam is acknowledged for testing lumpR and giving valuable feedback and bug reports. Furthermore, we appreciate the feedback and suggestions on sensitivity analysis by Dr. Gabriele Baroni. Tobias Pilz acknowledges funding by the Helmholtz graduate school GeoSim.

Edited by: Bethanna Jackson
Reviewed by: three anonymous referees

3.A Appendix

Precipitation uncertainty and comparisons with other studies

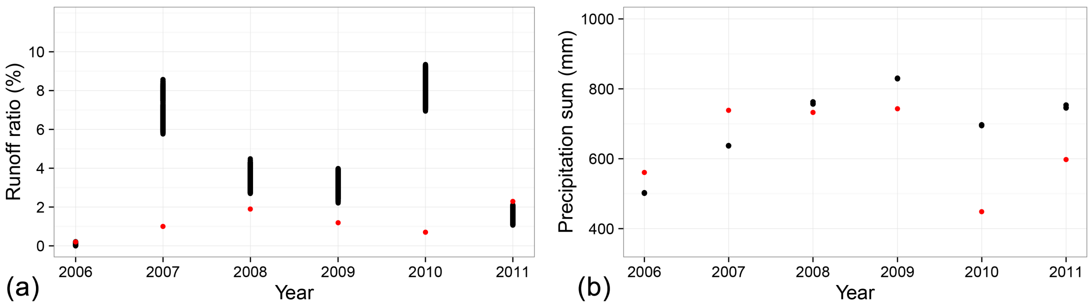

Figure 3.A1: Comparison of simulated yearly runoff coefficients (a) and precipitation forcing (b) for all 12,250 discretisations (black dots) with values reported by Medeiros and de Araújo (2014) (red dots).

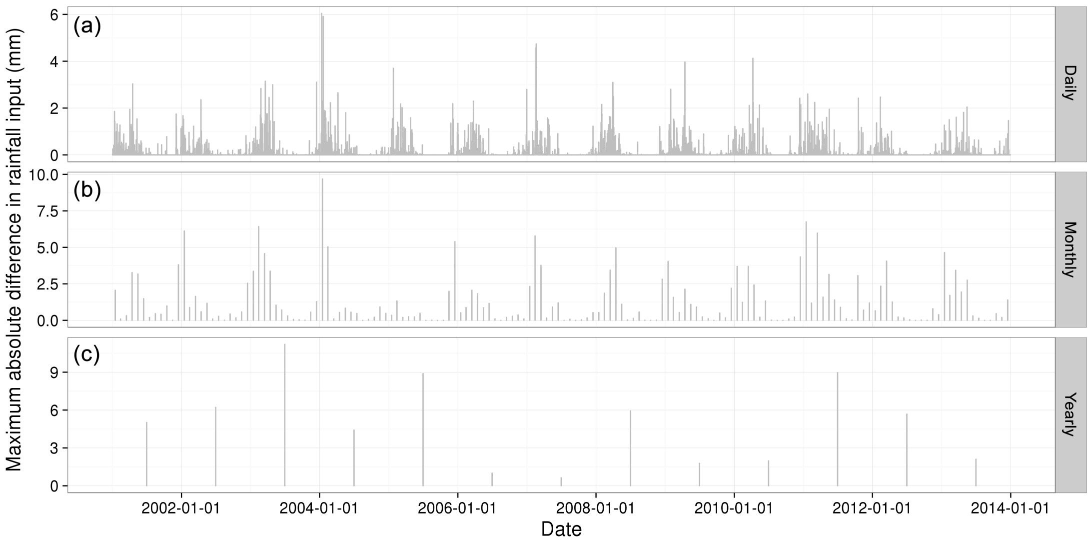

Figure 3.A2: Maximum of absolute differences in rainfall input between the 12,250 discretisations for daily, monthly, and yearly aggregated time steps.

4. How to tailor my process-based model? Dynamic identifiability analysis of flexible model structures

Abstract

In the field of hydrological modeling, many alternative mathematical representations of natural processes exist. To choose specific process formulations when building a hydrological model is therefore associated with a high degree of ambiguity and subjectivity. Identifiability analysis may provide guidance by constraining the a priori range of alternatives based on observations. In this work, a flexible simulation environment is used to build a process-based hydrological model with alternative process representations, numerical integration schemes, and model parametrizations in an integrated manner. The flexible simulation environment is coupled with an approach for dynamic identifiability analysis. The objective is to investigate the applicability of the coupled framework to identify the most adequate model structure. It turned out that identifiability of model structure varies in space and time, driven by the meteorological and hydrological characteristics of the study area. Moreover, the most accurate numerical solver is often not the best performing solution. This is possibly influenced by correlations and compensation effects among process representation, numerical solver, and parametrization. Overall, the proposed coupled framework proved to be applicable for the identification of adequate process-based model structures and is therefore a useful diagnostic tool for model building and hypotheses testing.

Key Points:

- A flexible simulation environment coupled with dynamic identifiability analysis forms a diagnostic tool for process-based model building
- The most adequate model structure in terms of process representation, numerical solver, and its parametrization is identified
- Identifiability of model structures varies in space and time driven by the current hydro-meteorological conditions

Submitted manuscript:

Pilz, T., Francke, T., Baroni, G., and Bronstert, A. (2019b). “How to Tailor my Process-based Model? Dynamic Identifiability Analysis of Flexible Model Structures”. Submitted to Water Resources Research

4.1 Introduction

Computer models are imperfect abstractions and simplifications of the real world transferred into computer code. As such, they necessarily impose uncertainties when simulating a certain processes or the evolution of target variables. In surface hydrology, one main objective of modeling is to transfer a precipitation signal into a discharge hydrograph at a certain river section. However, the diversity of landscapes, data sets, and specific research objectives led to the development of a large number of different hydrological models. These can vary in their conceptualization, how and to which degree of realism hydrological processes are represented, model runtime, initialization efforts, the number of parameters and if they need to be calibrated, or under which environmental conditions they are appropriate simulation tools (e.g., Clark et al., 2011b; Fenicia et al., 2016; Weiler and Beven, 2015).

Uncertainties are commonly associated with different input factors of a model, such as parametrization, forcing, or the selected equations for process representation (Pianosi et al., 2016). The impact of such input factors on simulations and predictions varies among different models and applications. This study is specifically focusing on parametric and structural model uncertainty.

Parametric uncertainty refers to uncertainty stemming from the parametrization of a model. Parameters can be determined by measurement and/or calibration. Model calibration aims at the estimation of effective model parameters by fitting the model output to observations. This imposes large uncertainties, primarily related to the phenomena of equifinality (different parameter realizations end up with equal model performances) and overparametrization (a larger number of parameters results in superior calibration performance but less predictive power). Such uncertainties are difficult to quantify and understand (Beven, 1993, 2006a; Her and Chaubey, 2015; Schoups et al., 2008).

Structural uncertainties are associated with the general conception of a model, incorporated mathematical equations, and computer code (Gupta et al., 2012). The structure of a model is basically related to the perceptual model of the real-world system and therefore reflects the system understanding of the model developer, which is in turn based on evidence from observations and experience (Beven, 2009; Wrede et al., 2015). Furthermore, model developers are usually confronted with a number of ambiguities, such as multiple equally plausible equations for a certain process. In addition, input and output of the model, state variables, effective calibration parameters, and the scale of operation need to be defined. Eventually, the functionality of the model needs to be proven in case studies. Model development is therefore often problem or even catchment specific, where usually no straightforward solutions exist and compromises need to be made (Fenicia et al., 2016; Höge et al., 2018; Jakeman et al., 2006). Consequently, structural uncertainties can be attributed to misconceptions of the general system, a lack of process understanding, spatial and temporal scaling issues, subjectivity and ambiguity regarding important decisions during model building, and random programming errors.

Hydrological model structures commonly comprise a set of Ordinary Differential Equations (ODEs) to describe the evolution of state variables. These ODEs need to be integrated over discrete time steps along a model application. However, complex hydrological models typically contain nonlinear ODEs, which are analytically intractable. Consequently, numerical approximation methods, also referred to as *ODE solvers*, need to be employed, which raise a number of mathematical issues that need to be considered, such as convergence (the solver needs to converge to a solution), order (how well solutions are approximated), and stability (the solution needs to be stable and must not oscillate). In surface hydrology, this problem has been ignored by many model developers and was handled negligently by applying the explicit Euler method, often along with operator splitting and solution constraints; i.e., water fluxes are calculated from the process equations in a predefined order, multiplied by time step length, added to the current values of state variables, and the solutions are eventually adapted to fit into physical constraints. It has been shown that this procedure can induce high uncertainties and lead to wrong conclusions (Gupta et al., 2012; Kavetski and Clark, 2011; Schoups et al., 2010).

Several approaches have been developed to disentangle the role of different input factors and support the identification of the best model structure and/or parametrization (Pianosi et al., 2016). Among others, in the last decades approaches of global sensitivity analysis (GSA) have been identified as important tools for model assessment and improvement (e.g., Pianosi and Wagener, 2016; Savage et al., 2016). Complementary, frameworks of identifiability analyses have been used to identify adequate model parametrizations (Herman et al., 2013; Wagener et al., 2003). Identifiability analysis relates to sensitivity analysis in a way that it tries to reduce uncertainty of model output by constraining the a priori range of sensitive input factors based on additional information, such as observations (Ghasemizade et al., 2017; Guillaume et al., 2019).

In the context of model structure analysis and identification, some former studies compared few different alternatives directly with each other (e.g., Fenicia et al., 2016; Kavetski and Fenicia, 2011). However this type of analysis has found some limitations due to interactions between the input factors and nonlinearities in the model response (Saltelli and Annoni, 2010). In contrast, simultaneous comparison of several model structures in a variety of Monte Carlo (MC) based approaches were developed, which avoid these limitations (e.g., Ajami et al., 2007; Baroni and Tarantola, 2014). For instance, by means of GSA it is possible to investigate the role of different factors (Günther et al., 2019; Savage et al., 2016; Stahn et al., 2017). Similarly, identifiability analyses can be performed to assess the capability of observational data to support model development (Coxon et al., 2014; Guillaume et al., 2019). These studies showed, how structured comparisons can account for the complex relation between model structures and parameters to support model improvement and testing.

In the present study, we further explore the use of identifiability analysis by coupling it with a flexible model environment. In specific, the aim of this study is to identify the optimal model structure with respect to uncertainties arising from parametrization, process representation, and numerical ODE solvers. By employing a flexible modeling platform, process equations can be easily exchanged and the integration of ODEs is separated from the process implementations. Moreover, instead of using simplified conceptual approaches as many other studies, more complex process-based representations are employed. In addition, spatial and temporal variability will be considered by employing dynamic identifiability analysis over catchments with different hydro-meteorological characteristics. It is hypothesized that this will provide more insights on model functioning and process behavior from a set of equally plausible process-based model structures tailored to the hydro-meteorological conditions of the area under investigation. The following specific research questions are addressed:

1. How well identifiable is a set of equally likely process representations, numerical ODE solvers, and parameter realizations?
2. How does identifiability vary over time with different meteorological conditions?
3. How consistent is the pattern for different parts of a catchment with varying hydrological conditions?

The paper is structured as follows: section 4.2 introduces the flexible model environment and framework for identifiability analysis. In section 4.3, a case study is presented to evaluate the proposed framework of model identification. The results of the case study and a discussion of their implications are given in sections 4.4 and 4.5, respectively. Eventually, the final conclusions are presented in section 4.6.

4.2 Framework for Process-based Model Identification

In order to determine the most adequate model structure, multiple alternatives need to be tested against each other. This is best achieved by implementing them in a single environment, which offers the infrastructure for a straightforward exchange of model structures, while preparation of input data and handling of output files are independent from the specific model. In the past, several such model platforms have been developed (Clark et al., 2015b, 2008b; Fenicia et al., 2011; Kneis, 2015).

To identify the most plausible model, the alternative model structures need to be evaluated. Instead of directly comparing the alternative model structures with each other, MC based strategies exist, which account for correlations and provide more robust results. In the following, the chosen flexible model environment and the algorithm for identifiability analysis are introduced.

4.2.1 The Flexible ECHSE Environment

The ecohydrological simulation environment (ECHSE) is a software designed for flexible model building (Kneis, 2015). The environment consists of a generic part and the model engines. The former is the basis of each ECHSE-based model and defines the format and general structure of input and output files, provides data types for model development (state variables, parameters, input, and output variables), and contains methods for the actual simulation, including a number of ODE solvers. The latter is the actual model and consists of code provided by the user, i.e., the actual process formulations.

For an application, first a user needs to declare generic model classes (such as a river, lake, subbasin, or soil–vegetation class), including state variables (e.g., soil moisture), parameters (e.g., hydraulic conductivity), input (e.g., rainfall), and output (e.g., lateral surface runoff), and include the process equations for each class (e.g., evapotranspiration, infiltration, runoff generation, and water movement for the soil–vegetation class). This defines the actual model. The user can choose from a pool of already existing classes and process formulations generated by other users or contribute own code for example applications using the ECHSE see Abon et al., 2016; Kneis et al., 2014; Kneis et al., 2017. During model initialization, the classes of the model need to be associated with real-world objects and the relations between the objects need to be defined (e.g., which subbasins drain into which downstream subbasin).

For this analysis, a new model engine has been developed, which is oriented at the process-based hydrological and sedimentological model WASA-SED (Güntner and Bronstert, 2004; Mueller et al., 2010). The model comprises an efficient hierarchical landscape disaggregated scheme over multiple scales. These spatial scales include subbasins, which contain landscape units (LUs) as representative hillslopes, further characterized by different terrain components (TCs), which in turn consist of several soil–vegetation components (SVCs) described by a characteristic soil profile with specific vegetation cover. Explicitly represented processes include evapotranspiration, infiltration, both infiltration-excess and saturation-excess runoff, soil water movement, exfiltration as well as lateral runoff redistribution. Groundwater is represented by a simplified linear storage approach. River flow is described by a simple unit hydrograph routing with evaporation as only source for transmission losses. The model has been applied in several dryland regions in Spain (Bronstert et al., 2014; Francke et al., 2018b; Mueller et al., 2009, 2010), Brazil (de Araújo and Medeiros, 2013; Krol et al., 2011; Medeiros et al., 2014, 2010; Pilz et al., 2019a), and India (Jackisch et al., 2014).

4.2.2 Dynamic Identifiability Analysis (DYNIA)

To identify the most adequate model structure, the dynamic identifiability analysis (DYNIA) approach by Wagener et al. (2003) was adopted. It is based on the regional SA (RSA, also called MC filtering) approach by Spear and Hornberger (1980) by partitioning an ensemble of model runs into behavioral (acceptable model performance) and nonbehavioral sets. The DYNIA framework uses the same basis of RSA with the aim to dynamically assess the information content of input factors over moving time windows. In that way the influence of varying hydro-meteorological conditions, i.e., wet and dry periods, on the identifiability of model structures can be assessed. Instead of employing a formal Bayesian approach for the partition of model results, Wagener et al. (2003) employed an informal method as in GLUE (Beven and Binley, 1992) by selecting the best performing realizations in an arbitrary manner (e.g., the best 10 %), which was also used in this work.

In general, DYNIA consists of the following steps: (i) calculation of a performance metric for each model run; (ii) filtering of the best model runs; (iii) calculation of an identifiability measure for each input factor; (iv) determination of the posterior distribution of each input factor from filtered model results.

Originally, DYNIA has been applied to parameters, for which a distribution was defined and realization were sampled (Wagener et al., 2003). To account for discrete nonscalar input factors (e.g., different model structures), the analysis is modified following the approach described by

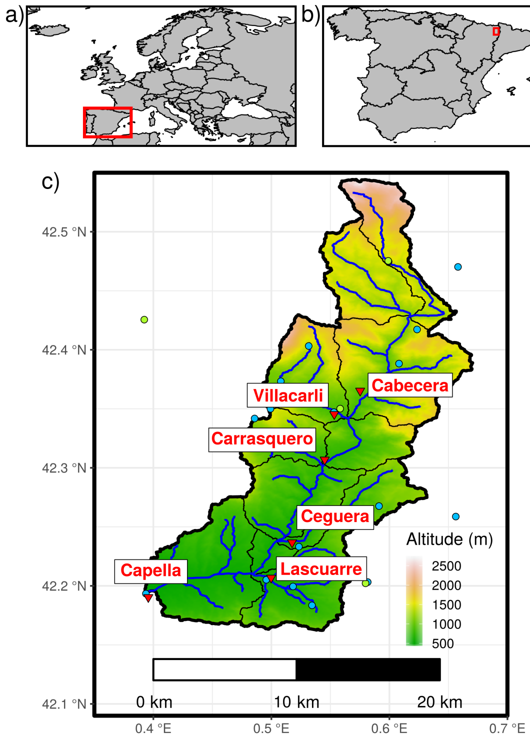

Figure 4.1: Overview over the Isábena catchment (c) and its location within Spain (b) and Europe (a). In panel (c), background colour is based on the DEM used for model initialization, thin black lines within the catchment outline subbasins (delineated with the lumpR software), red triangles mark the position of discharge gauges, blue and green points show gauges of rainfall and other meteorological variables, respectively. Note that not all stations used in this study are visible due to feature overlays and because some stations are located slightly outside the plot.

Baroni and Tarantola (2014) within their general probabilistic framework (GPF). In specific, each possible realization of model structure–parameter combinations is associated with a scalar value. The number of possible realizations characterizes the discrete uniform distribution, from which is sampled. This is a generic strategy for the quantification and ranking of different sources of uncertainty in the context of environmental model application and allows for the quantification of both numeric and nonscalar input factors. The framework requires that all possible combinations of the input factors are assessed.

4.3 Case Study

4.3.1 Study Area

For the model simulations, the Isábena catchment in northeastern Spain was selected (Figure 4.1). The watershed comprises an area of about 425 km^2 and is located at the southern edge of the Pyrenees. It is a small headwater catchment of the Ebro river basin. Therefore, the mountainous topography is characterized by a heterogeneous relief with altitudes ranging from about 500 m to 2700 m. Consequently, also precipitation is spatially heterogeneous with annual sums ranging from 450 mm in the lowlands up to 1600 mm in the upland parts of the catchment and a spatial average of about 770 mm. The climate is characterized by Atlantic and Mediterranean influences and is generally wet and cold (García-Ruiz et al., 2001).

The hydrological regime is influenced by rain and snow. Floods may occur in spring as a consequence of precipitation events amplified by snowmelt, or in late summer and autumn caused by thunderstorms. Mean annual discharge at the catchment outlet is 4.1 m^3/s, while a maximum instantaneous value of 370 m^3/s has been observed so far. Minimum discharge can be less than 1 m^3/s, but the river never falls completely dry. The study area is not regulated, thus the hydrological regime is determined by natural factors only. It is mainly composed of deciduous woodland, agriculture, pasture, and bushes in the valley bottoms with evergreen oaks and pines.

The area has been investigated in many research projects, including intensive hydro-sedimentological monitoring see Bronstert et al., 2014, and references therein. Consequently, a rich dataset exists with relatively high spatial coverage of meteorological and discharge measurements. In addition, the mountainous catchment consists of several subcatchments with varying hydrological and meteorological conditions, which provide different settings to apply the proposed framework of this study and investigate the influence of hydro-meteorological conditions on the identification of model structures. Recorded time-series data have recently been published and described by Francke et al. (2018a). The dataset will be described in more detail along with model initialization in the following subsection.

4.3.2 Data and Model Initialization

All model simulations are based on the same initialization procedure, including the discretization of the study area into model units, parametrization of soil and vegetation, and the preparation of meteorological inputs. Hillslope-based landscape discretization was performed using the lumpR software, a package for the free and open-source environment R (Pilz et al., 2017). The basis of the algorithms were a 15 m x 15 m DEM processed from ASTER raw data, a soil type, and a land-use map including parametrizations of the soil and vegetation types. The initialization procedure comprised the delineation of the catchment and subbasins (outlined in Figure 4.1), the derivation of further model units (the LUs, TCs, and SVCs described in 4.2.1), calculation and checking of parameters, and the generation of model input files.

Meteorological and discharge data were obtained from the dataset of Francke et al. (2018a). This includes rainfall data from 18 stations, temperature from 9 stations, and solar radiation and air humidity data from 2 stations within or in close vicinity to the study area. Gaps in the time-series were interpolated with data from neighbouring stations. Furthermore, station data needed to be interpolated to the centroids of the subbasins by employing inverse distance interpolation (IDW), which was realized using the geostat R package of the ECHSE toolbox (Kneis, 2012). For model evaluation, spatially distributed discharge measurements from the Isábena river at the catchment outlet (Capella) and from five subcatchments were used (Table 4.1).

All experiments were conducted under the same experimental design. The general simulation and analysis period covered three years from 1 January 2013 to 31 December 2015 with a temporal resolution of one day. This decision is based on a compromise between data availability and computational feasibility. On the one hand, the period should be sufficiently long to cover the hydrological catchment dynamics under different conditions. On the other hand, although subdaily measurements are available and numerical ODE solvers would be expected to be more reliable under hourly resolution, model runtimes with hourly resolution would be too long to achieve results in acceptable time, when applied in the presented framework. To bring model states into equilibrium and avoid artificial effects on outputs, a warm-up was conducted prior to any simulation run. This warm-up run consisted of iterations over one year (1 January 2012 to 31 December 2012) until convergence was achieved, i.e., until the sum of hydrological storages at the end of a warm-up iteration deviated by less than 0.1 % from the sum of storages at the end of the previous iteration.

4.3.3 Input Factor Definition

In this study, five input factors reflecting different sources of uncertainty were distinguished: (Ia) structural uncertainty with respect to evapotranspiration subprocesses; (Ib) uncertainty in the representation of soil water processes; (Ic) runoff concentration; (II) numerical integration of underlying ODEs; (III) parametrization. In the following it will be described, how these have been implemented in the ECHSE environment.

Table 4.1: Characteristics of the study area and delineated subcatchments referring to the analysis period 2013 to 2015.

Gauge	A (km^2)	Gauge elevation (m a.s.l.)	P ($mm\,year^{-1}$)	Q ($m^3\,s^{-1}$)	RC (%)
Villacarli	41	866	673	0.30	27
Cabecera	145	841	901	2.09	46
Carrasquero	25	762	673	0.44	68
Ceguera	29	598	599	0.13	21
Lascuarre	44	565	572	0.06	6
Capella	424	490	743	4.57	40

Note: Capella is located at the outlet of the study area, while all other gauges refer to individual subcatchments; A: catchment area referring to the model setting of this study; P: average annual rainfall (spatially interpolated from station data, see text); Q: average discharge; RC: runoff coefficient.

Process Representations

For nine hydrological (sub-)processes, two alternatives each were implemented in ECHSE (Table 4.2). These alternatives consist of the respective representation copied from the WASA-SED model (denoted by approach 1 in Table 4.2) and an alternative (approach 2). In most of the cases, approaches are similar but may vary in degree of detail. For instance, the evapotranspiration approach by Shuttleworth and Wallace (1985) is deviated from the Penman-Monteith formula, but consists of a more detailed conception of resistances to account for patches of bare soil instead of assuming a homogeneous vegetation cover.

ODE Solvers

An aspect of this study was to account for more appropriate numerical integration of the model's ODEs than it is commonly done in the field of surface hydrology (Kavetski and Clark, 2011). Therefore, several ODE solvers were implemented in ECHSE using the external *GNU Scientific Library* (Galassi et al., 2017). Four different solvers were considered in this study, varying in accuracy and stability (Table 4.3). More detailed information about ODE solvers in general and the selected implementation are provided in the Supporting Information, Section 4.A.

Parametrization

To reflect uncertainty in the parametrization, seven sensitive and commonly uncertain parameters were considered (Table 4.4). These parameters influence different components of the model including evapotranspiration, soil water movement, groundwater recharge, infiltration, and runoff concentration. Parameters denoted as factors are multiplied by the a priori estimated parameter value for each spatial unit. In case of absolute parameters, the value is directly inserted into the model equation. Consequently, all parameters are globally effective parameters and independent from the spatial model set-up.

4.3.4 Implementation of the Analysis Framework

This section describes how the dynamic identifiability analysis was coupled with the flexible simulation environment ECHSE for the specific case study.

Input Factor Realizations (Prior Distribution)

Regarding process representation (Ia to Ic), the realizations for each input factor were defined by all possible combinations of the two alternative representations for each subprocess. This resulted in 2^n realizations for each input factor, where n is the number of considered subprocesses; i.e., 5, 3, and 1 subprocesses for evapotranspiration, soil, and runoff concentration, respectively, resulting in 32, 8, and 2 realizations for input factors Ia, Ib, and Ic, respectively (for an overview see Supporting Information Table 4.7). The ODE solvers (input factor II) were combined with the possibility of constrained or freely evolving solutions, which resulted in eight realizations for this input factor (4 solver variants times 2 variants of constraints). To represent

Table 4.2: Overview over considered alternative process representations implemented in the ECHSE engine.

	Approach 1 (as in WASA-SED)		Approach 2	
Process	ID	Reference	ID	Reference
Evapotranspiration				
Evapotranspiration	**SW**	Shuttleworth and Wallace (1985)	**PM**	Penman-Monteith
Bulk stomatal resistance of canopy	**SK**	Saugier and Katerji (1991) eq. 4	**SW19**	Shuttleworth and Wallace (1985) eq. 19
Roughness length	**SG43**	Shuttleworth and Gurney (1990) eq. 43	**BT**	Brutsaert (1975)
Displacement height	**SG42**	Shuttleworth and Gurney (1990) eq. 42	**SG41**	Shuttleworth and Gurney (1990) eq. 41
Clear sky radiation	**AN**	Ångström formula	**AL**	Allen et al. (2005) eq. 19
Soil water				
Infiltration	**GA**	Modified Green–Ampt (Güntner, 2002)	**PH**	Philip (1957)
Percolation	**PS**	SWAT approach (described in Güntner, 2002)	**PR**	Simplified Richards' equation
Soil water retention	**VG**	van Genuchten (described in Maidment, 1993)	**CB**	Campbell (described in Maidment, 1993)
Runoff concentration				
Runoff concentration	**RW**	Lateral re-distribution along a hillslope as developed for the WASA-SED model (Güntner, 2002)	**RS**	Additional delay of WASA-SED based runoff concentration by linear storage approach (requires calibration)

Note: equations see Supporting Information, Tables 4.5 and 4.6.

the parameter space (input factor III), 1000 realizations were taken by Latin hypercube sampling from the log-transformed and uniformly distributed parameter spaces (except for Phil_cal, which was not transformed). In line with the GPF (Baroni and Tarantola, 2014), each realization of the input factors was associated with a number in order to obtain numerical input factors with discrete distributions (see also Supporting Information Table 4.7).

Because of computational constraints, of the resulting 4,096,000 possible input factor combinations, 12,000 samples were randomly drawn. The model was subsequently evaluated for each sample. It should be noted that this framework involves no explicit parameter calibration of the model. The resulting ensemble of model runs was then analyzed following the DYNIA framework as follows.

Posterior Distribution and Dynamic Identifiability Measure

As performance metric the root mean square error was selected. For the dynamic analysis a performance value was calculated for each simulation day d over a moving window of the width $2w + 1$ as follows:

$$RMSE(d) = \sqrt{\frac{1}{2w+1} \sum_{i=d-w}^{d+w} \left(q_s(i) - q_o(i) \right)^2} \tag{4.1}$$

where q_s is the simulated and q_o the observed discharge, and w a parameter defining the window size. For this study, w was set to a value of 15 days, which resulted in a total moving window size of 31 days.

Identifiability in general can be quantified by comparing the prior and posterior distributions for each input factor. In this context, different quantification strategies exist (e.g., Pianosi and Wagener, 2016; Wagener et al., 2003). In this study, as dealing with discrete input factor distributions, identifiability was defined by comparing the frequency of occurrences of all

Table 4.3: Overview over ODE solvers.

ID	Description	Accuracy	Time step
EU	Explicit Euler	First order	Fixed (daily)
RK	Explicit Runge–Kutta Cash–Karp	Fourth order	Fixed (daily)
RKST	Explicit Runge–Kutta Cash–Karp	Fourth order	Adaptive
BDF	Backward differentiation formula in Nordsieck form (predictor–corrector)	Dynamic (first to fifth order)	Multistep

Note: Each solver can be applied with (CONS) or without (FREE) solution constraints. Stability of solutions generally increases from EU to BDF.

Table 4.4: Overview over parameters and sampling ranges.

Symbol	Description	Component	Type	Lower range	Upper range
cal_wind	Windspeed correction	Evapotranspiration	Factor	0.1	5
cal_ks	Correction of saturated hydraulic conductivity	Soil water movement	Factor	0.01	100
cal_kfbed	Correction of bedrock conductivity	Groundwater recharge	Factor	0.01	100
Phil_cal	Parameter for Philip's equation ($\mathrm{m\,s^{-1}}$)	Infiltration	Absolute	0.2	1
str_surf	Surface runoff retention parameter	Surface runoff	Factor	0.001	0.1
str_inter	Subsurface runoff (interflow) retention parameter	Interflow	Factor	0.01	0.5
str_base	Groundwater runoff (baseflow) retention parameter	Baseflow	Factor	0.1	10

realizations for a certain input factor before (referring to the prior) and after the filtering (referring to the posterior distribution). As proposed by Wagener et al. (2003), the posterior distributions were obtained by arbitrarily selecting the best 10 % of model evaluations with respect to RMSE measures. A specific realization of an input factor was considered negligible if the number of occurrences in the posterior distribution was less than 10 % the number of occurrences in the prior distribution. Eventually, the posterior distribution generically consisted of 1200 values out of the 12,000 prior samples.

For a specific input factor, the identifiability measure was defined as

$$IM = 1 - \frac{n_{post} - 1}{n_{prior} - 1}, \tag{4.2}$$

where n_{prior} is the number of realizations in the prior distribution, and n_{post} is the number of remaining realizations in the posterior distribution of that input factor after the filtering. Thus, IM ranges between zero and one, where the former indicates no identifiability at all and the latter indicates perfect identifiability (i.e., the posterior distribution of that input factor comprises of only a single realization).

Note that both n_{prior} and n_{post} in this context are positive integers. A value of zero for n_{post} would indicate that the model is not able to produce acceptable performances. Furthermore, when comparing the input factors, the different characteristics and number of realizations have to be kept in mind, which define each input factor.

The aforementioned steps of DYNIA were applied over both the full simulation period (static identifiability analysis) in order to identify the optimal model structure, and a moving window over the simulation period (dynamic identifiability analysis) to study the influence of varying meteorological conditions. Spatial variability, and hence the influence of hydrological characteristics, was investigated by analyzing the full study area and individual subcatchments with distinct characteristics

Credibility Assessment

Credibility of the processing framework was assessed by means of: (i) analysis of convergence (is a sample size of $N = 12,000$ realizations sufficient?); and (ii) analysis of robustness (is the sampled posterior distribution independent of specific samples?). Convergence is typically analyzed by subsampling, i.e., the computation of the target variable for increasing N from the original sample. This was coupled with a bootstrapping approach to analyse robustness, which consists of sampling from the N model results N_b times with replacement and subsequent re-calculation of the target variable. The range over the values of the target variable from the N_b bootstrapping procedures is a measure of robustness. In that way credibility analysis does not require any additional model evaluations (Pianosi et al., 2016).

In this study, N_b was set to a value of 1000. The target variable, for which credibility was assessed, was the RMSE value of the posterior distribution of model results. This also includes the filtering step, i.e., in each bootstrap iteration the filtering step was applied and subsequently the minimum, maximum, and median RMSE value of the resulting posterior distribution calculated.

4.4 Results

4.4.1 Model Simulations

Model Errors

To carry out the experiments with the different model structures and parametrizations, the ECHSE environment was run 12,000 times according to the number of sampled realizations of Process representations, ODE solvers, and parametrizations. Individual simulation runtimes varied between less than a minute and more than one hour, mainly depending on the ODE solver and workload of the high performance cluster, where the simulations were carried out. During model evaluation, 32 runtime errors occurred, i.e., for the further analyses only 11,968 realizations could be considered. The runtime errors occurred exclusively when employing the backward differentiation formula (BDF) without solution constraints as ODE solver. This is not surprising as it is the only one of the family of implicit solvers considered in this analysis, which are generally known to be limited by the given constraints of accuracy and maximum number of iterations. However, the number of errors is low, even in comparison to all runs with the BDF solver (approx. 3000), and thus no impact on the general conclusions of this study are expected.

Discharge Simulation

Observed discharge values for most time steps fall into the 90 % probability range of the 11,968 model realizations (Figure 4.2). However, especially large discharge peaks are often underestimated by the model ensemble as well as by the single best model runs. Nash–Sutcliffe values range from 0.25 for the Villacarli headwater catchment to 0.67 for the basin outlet at Capella. The RMSE for the best performing realization generally lies in the order of magnitude of average discharge (compare with Table 4.1). However, especially for the smaller subcatchments (Villacarli, Carrasquero, Ceguera) RMSE is relatively high. In general, the model is performing better for the larger (sub-)catchments (Cabecera, Capella).

Credibility Assessment

The sample of realizations is adequate for the further processing steps (Figure 4.3). For all subcatchments, characteristic values of the posterior distribution of model performances (i.e., the minimum, maximum, and median RMSE) each converge with increasing sample size. Besides, the characterizations of the posterior distributions for each subcatchment appear to be robust as the uncertainty ranges are small.

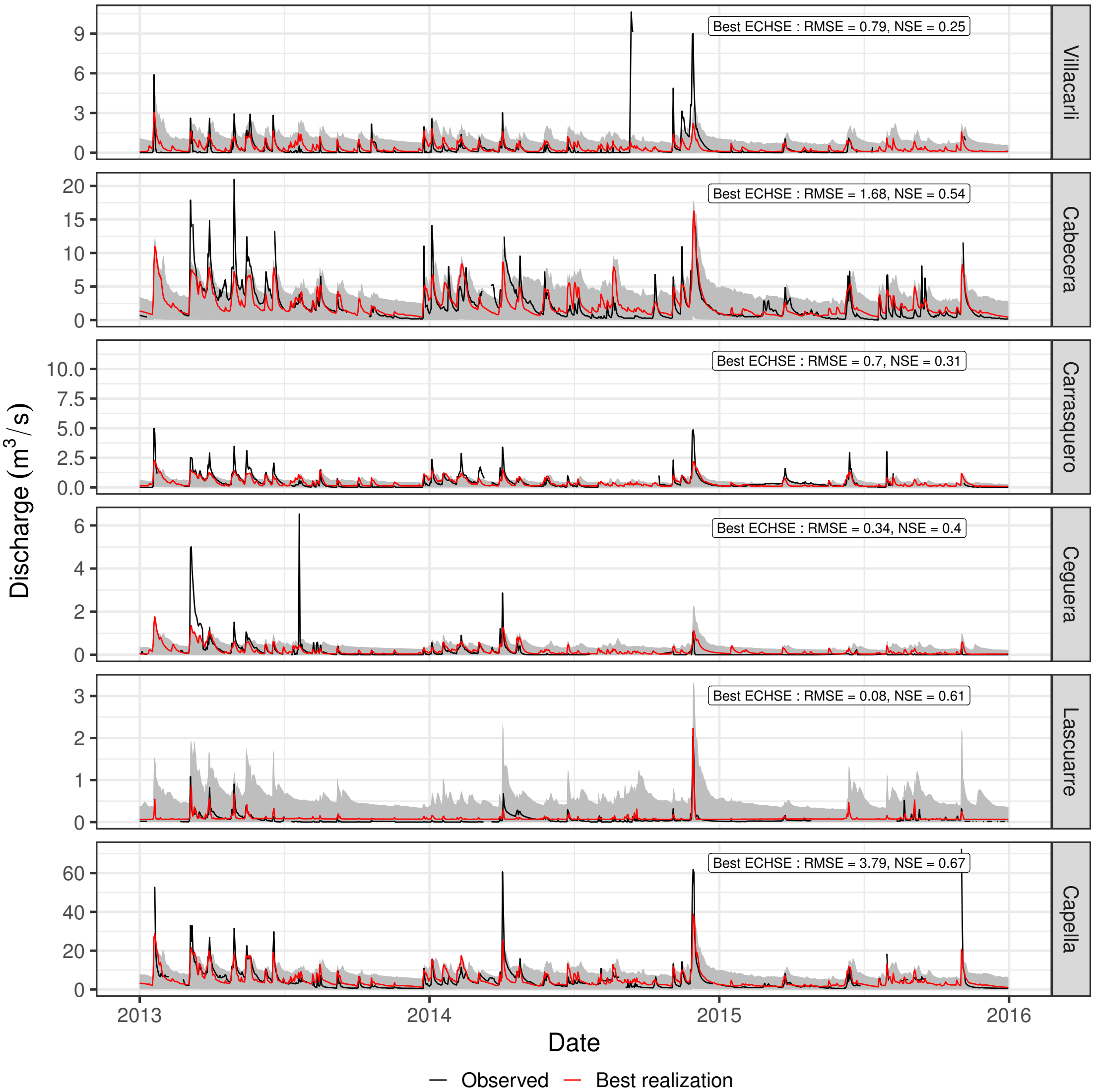

Figure 4.2: Discharge simulations in comparison to observations for each studied subcatchment and the whole study area (Capella). The gray area illustrates the 90 % probability range of the model realizations (prior distribution).

4.4.2 Static Identifiability Analysis

Parametrization is the best identifiable input factor for all gauges, followed by evapotranspiration structures (Figure 4.4). For the latter, however, spatial differences exist. At gauge Lascuarre, with its comparatively low runoff coefficient (see Table 4.1), no behavioral model structures with respect to evapotranspiration processes could be distinguished from the set of a priori structures. For the other input factors, all realizations of the prior distribution also occur in the posterior distribution resulting in identifiability measures of zero for all gauges.

More detailed information on identifiability can be obtained when analyzing the posterior distributions of each input factor (Figure 4.5). For evapotranspiration, not only the identifiability measure but also the occurrences of realizations (n_{post}) are similar for all gauges except Lascuarre. In general, the Penman–Monteith approach (PM, odd-numbered realizations, see Supporting Information Table 4.7) has the highest occurrence values in the posterior distribution of that input factor. For the other evapotranspiration subprocesses, no obvious patterns can be seen. In contrast to the other gauges, at Lascuarre the Shuttleworth–Wallace approach (SW, even-numbered realizations) achieves high values of n_{post}. That means that the PM approach (almost exclusively constituting the best 10 % of realizations) is clearly the superior

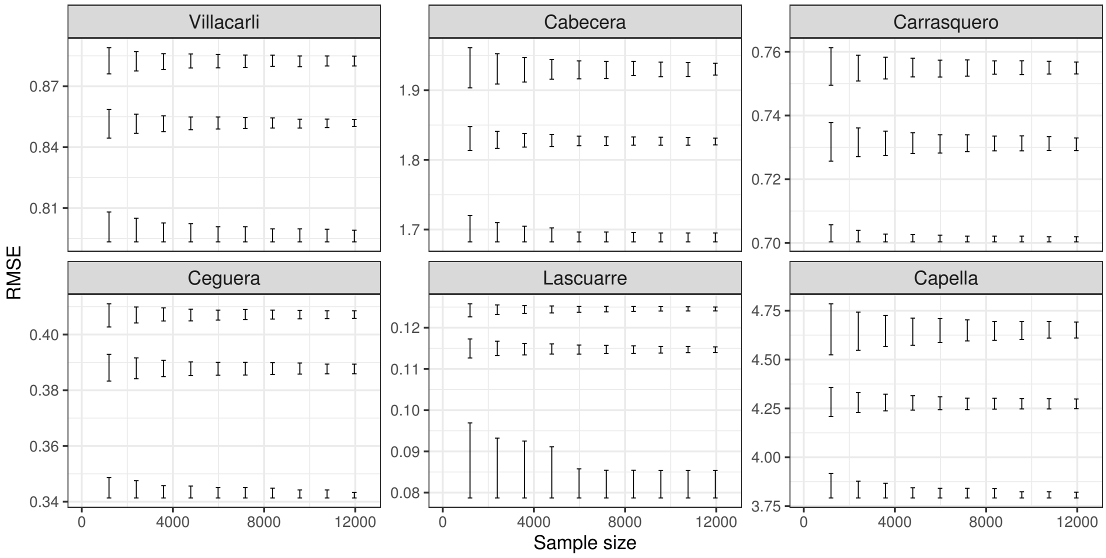

Figure 4.3: Credibility of the sampled posterior distribution of model performances. Vertical bars represent the 90 % confidence intervals of minimum (lower row), median (middle), and maximum (upper) RMSE values of the bootstrap samples and are a measure of robustness of the sample. Decreasing bar size with increasing sample size indicates sampling convergence.

evapotranspiration model for most parts of the study area, regardless of the choices for the other evapotranspiration subprocesses, while for Lascuarre the SW approach excels.

Regarding the soil water processes, the approach by van Genuchten (VG, IDs 1 and 4) obtained the highest posterior occurrences at most gauges. Again, there is a different picture for gauge Lascuarre, where no clear pattern can be distinguished. For runoff concentration, ID 2 (approach with empirical delay factors in addition to the physically based approach of WASA-SED) achieved slightly more occurrences for all gauges but for Lascuarre.

The ODE solvers without solution constraints (odd-numbered IDs) exhibit the highest n_{post} values for all gauges except Villacarli and Lascuarre. Moreover, solvers with higher accuracy tend to have obtained slightly more occurrences in the posterior distribution. This, however, is opposed to the findings for Lascuarre, where solvers with solution constraints (even-numbered IDs) dominate and the explicit Euler approach even achieved most occurrences.

It should be noted that configurations with the highest values of n_{post} do not necessarily comprise the best performing realizations (see red bars in Figure 4.5, representing the red line in Figure 4.2).

Despite the high identifiability measure, the posterior distributions of input factor parametrization show distinct peaks only for three of the seven parameters (cal_kfbed (groundwater recharge), cal_ks (soil water movement), and cal_wind (evapotranspiration); Figure 4.6). The other parameters are relatively equally distributed and show little distinction from the prior distribution (uniform distribution within the parameter ranges). Differences among gauges are small. The only exception is parameter cal_wind, for which at gauge Villacarli a peak in the posterior distribution for values greater zero (resulting in increasing values of evapotranspiration) can be seen, while for the other gauges peaks are at smaller values (translating into less evapotranspiration). Parameters cal_kfbed and cal_ks show a tendency towards smaller values, which results in less groundwater recharge and delayed subsurface runoff generation.

4.4.3 Dynamic Identifiability Analysis

Identifiability may change over time depending on the current boundary conditions (Figure 4.7). This holds especially true for input factor evapotranspiration, which is varying considerably, while for the other input factors identifiability remains more ore less constant. There is some evidence that identifiability of evapotranspiration is enhanced during wet periods. However, differences exist among subcatchments, as for Lascuarre the pattern seems to be reversed and

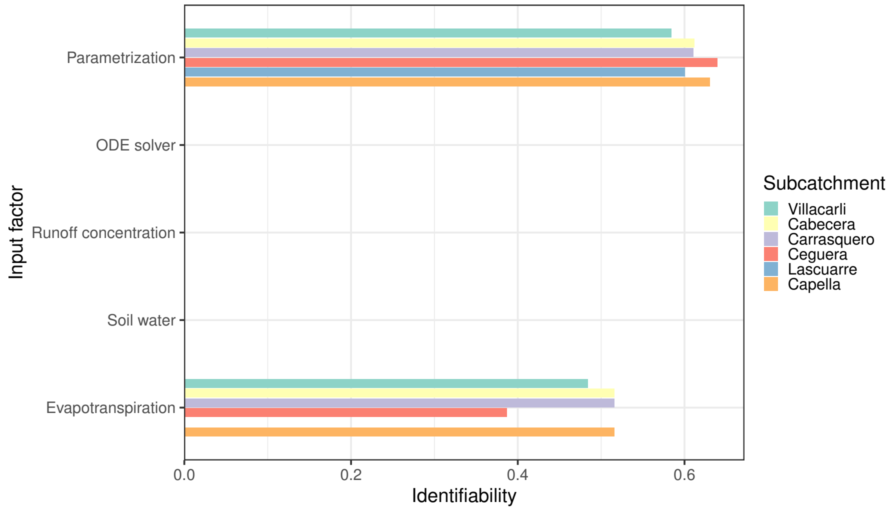

Figure 4.4: Identifiability measures for the different input factors and subcatchments.

for Ceguera there are periods of enhanced identifiability during wet and dry periods. For the other input factors no relationships become visible.

The dynamic analysis of posterior frequencies, however, provides more detailed insights and reveals patterns even for input factors other than evapotranspiration (Figure 4.8). It shows that the clear pattern of high posterior values of odd-valued realizations of evapotranspiration (i.e., preference of the PM formula) is blurred or even reversed during periods of low flows, where even-valued realizations (the SW formula) dominate the posterior distribution (e.g., end of 2013 or beginning of 2015). In contrast, for soil water, such low flow periods lead to high posterior values of IDs 7 and 8 (soil water retention model after Campbell and percolation modeled by simplified Richards' approach), while during high flows ID 1 with completely different equations is favored (see Supporting Information Table 4.7). A model structure for runoff concentration is best identifiable during peak flows (ID 2, with additional calibration parameters), but for most of the simulation period posterior values were close to prior occurrences. The pattern for ODE solvers is mostly blurred but shows a small tendency towards even-numbered realizations (constrained solvers) during low flows. During high flows, however, although the pattern is still blurred, odd-numbered (unconstrained) solvers with higher accuracy are more frequent in the posterior. However, it should be kept in mind that absolute identifiability measures for soil water, runoff concentration, and ODE solver are low and small absolute values of posterior frequencies are amplified due to the scaling in Figure 4.8.

For input factor parametrization (Figure 4.9), clear signals can only be seen for parameters cal_kfbed (groundwater recharge) and cal_ks (soil water movement). Parameter cal_kfbed appears to be best identifiable during dry periods with a tendency towards smaller values (less groundwater recharge, more soil moisture and interflow). In contrast, cal_ks is best identifiable during wet periods with a tendency towards smaller values (reduced conductivity of soil, more surface runoff). For cal_wind (evapotranspiration) and str_base (baseflow), patterns can be distinguished as well, but are more difficult to generalize. Both during peak discharge and low flow periods, cal_wind is well identifiable with tendency towards smaller values (less evapotranspiration), while in times of intermediate flow identifiability tends towards larger values (more evapotranspiration). Parameter str_base shows a tendency towards small values (small retention, quick release of flows) at the beginning of flow events with values increasing over the discharge event (increasing retention and prolonged release of flows; e.g., at the end of year 2013 or several times in 2015).

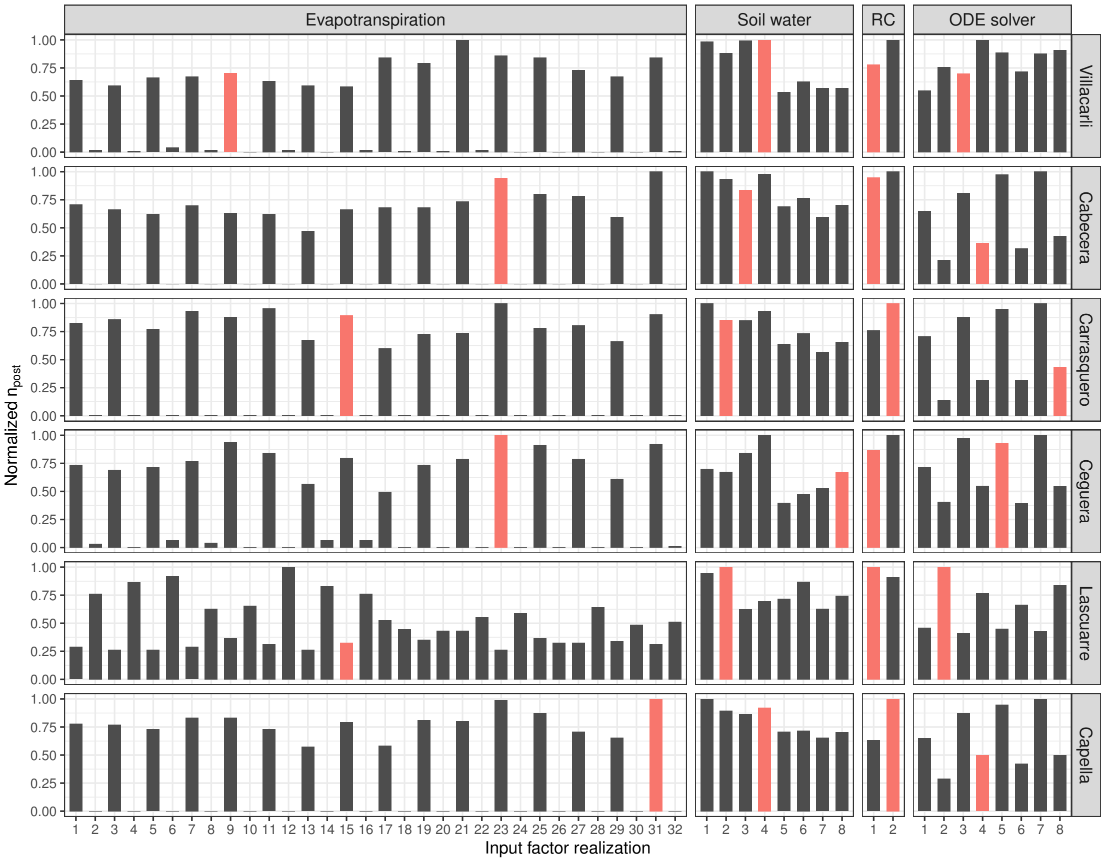

Figure 4.5: Normalised occurrences of realizations in the posterior distribution (n_{post}) for the various input factors and subcatchments. A red bar marks the realization for each input factor related to the best model run (red line in Figure 4.2).

4.5 Discussion

4.5.1 Evaluation of Model Performance

There are a number of factors influencing the performance of the hydrological model and, hence, the subsequent identifiability analysis. It was shown that the model, overall, performs acceptable but is often missing the large peak flow events (Figure 4.2). Such behavior can often be attributed to uncertainties in the precipitation input. This arises from limited coverage with station data, especially in mountainous areas, where rainfall is typically unevenly distributed in space. Moreover, convective rainfall events in summer, which are characterized by high intensity, short duration, and small spatial coverage, can produce sharp peaks in the discharge hydrograph but might be entirely undetected by rainfall stations. Such an event presumably occurred in summer 2013 in subcatchment Ceguera, which resulted in the highest peak flow of the simulation period (Figure 4.2), but no such event appears in the rainfall dataset (not shown). Moreover, the highest influential rainfall station of this subcatchment did not record within this particular period, which leads to missing data due to station failure as yet another important source of uncertainty. This issue could generally be addressed by using rainfall products derived from remotely sensed data (e.g., radar data), which provide a better spatial coverage as well as spatial and temporal resolution. Yet a combination with station measurements is still crucial as radar products do not yet provide rainfall information with sufficient quality (Abon et al., 2016; Kneis et al., 2014).

Apart from the limited number of stations, interpolation of station data to computational model units (in this case subbasins, thin black lines in Figure 4.1) results inevitably in a smoothed input signal for the model. On the one hand, single events of high intensity detected by just a single station might be obscured when merged with data of other stations. On the other

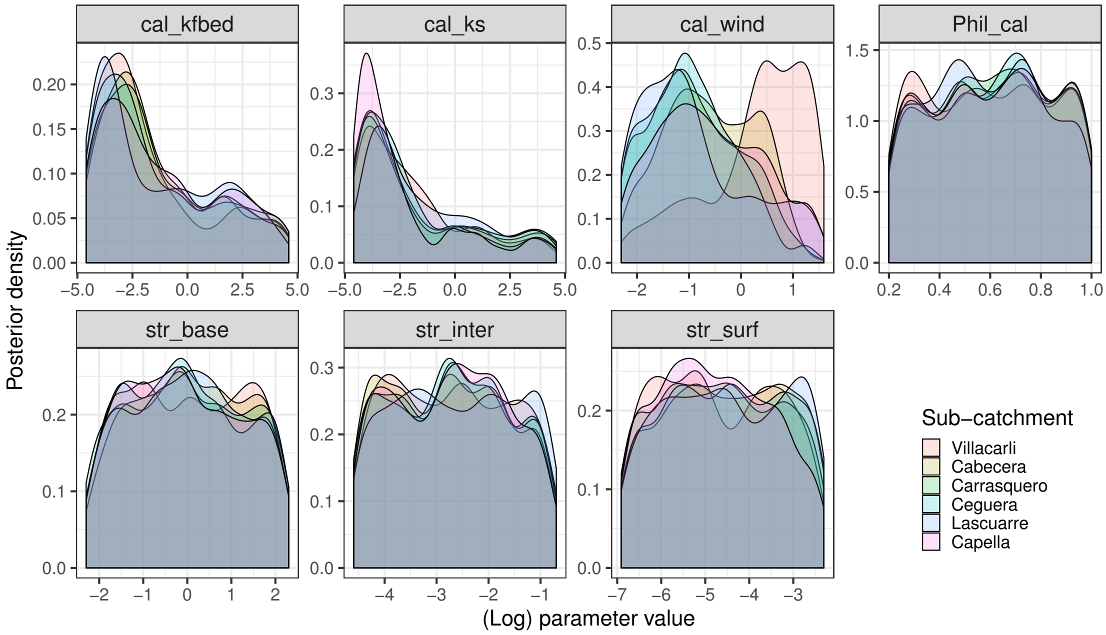

Figure 4.6: Density plots of the posterior distributions of parameters. Shown are log-transformed parameter spaces except for Phil_cal. For parameter symbols see Table 4.4.

hand, the size of such computational units, typically defined as subjective decision during model initialization, dictates the spatial resolution of rainfall input and is again of high relevance for convective rainfall events (Pilz et al., 2017). In contrast, events of long duration and large spatial coverage are often better detected by stations and directed into models, and therefore usually result in better model performance. One such event occurred in November 2014, where several weeks of recurring days with high precipitation resulted in high streamflow peaks in all subcatchments. This event was much better reproduced by the model, especially at gauges Cabecera and Lascuarre, where the match of observed and simulated discharge hydrographs is almost perfect.

4.5.2 Methodology and Identifiability Measure

In general it has to be noted that the employed model structures were not calibrated in terms of explicit parameter optimization. Instead, the presented framework refers to the 10 % best of the 12,000 model structure and parameter realizations for each gauge. Besides, the model structures used in this study are largely derivatives of the WASA-SED model, which was transferred into the ECHSE environment and enhanced by additional process representations and ODE solvers. This accounts for structural uncertainties in hydrological models, but introduces further uncertainties related to (possibly) random programming errors and ambiguities during model building due to subjective decisions, e.g., which parameters are hard coded, which need to be calibrated, and which can be derived from measurements.

The results of parameter identification and sensitivity analyses are influenced by the employed performance metric (Francke et al., 2018b; Guse et al., 2017). For this study, the RMSE was selected as in the studies of Wagener et al. (2003) and Pianosi and Wagener (2016). Experiments with the Nash–Sutcliffe index gave very similar results (not shown). Yet, RMSE, as well as the Nash–Sutcliffe index and other metrics using squared residuals, is biased towards higher values and the timing of the hydrograph while deficiencies of a model in reproducing low flows are less strictly penalized. This aspect is relevant when aggregating over long periods and may therefore influence the findings of the static identifiability analysis of this work.

The dynamic analysis, in turn, is influenced by the decision on the length of the moving window. This in particular is the case if sensitivity of the considered input factors varies in time, e.g., because they represent processes occurring within larger or smaller characteristic time

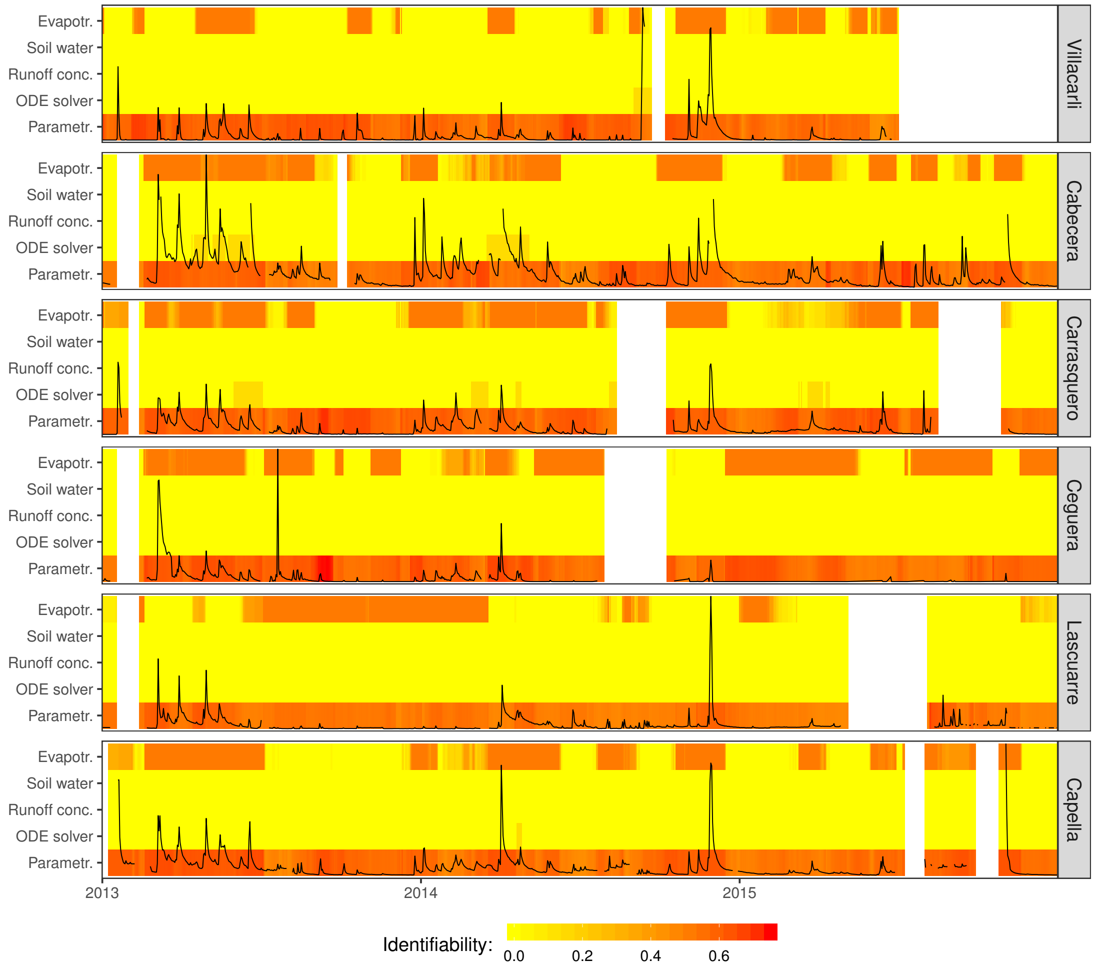

Figure 4.7: Identifiability measures over the simulation period. Black lines represent the observed discharge hydrograph at each gauge.

scales than the chosen window width (Massmann et al., 2014). However, objective selection criteria are missing and it was decided to use a window of one month. In that way measurement uncertainties are less likely to dominate the results (as would be the case for a small window size), but the window is still small enough to identify relevant process realizations at varying hydrological conditions. Experiments with window size (not shown) revealed that a smaller window size generally leads to a more blurred pattern while with larger windows patterns can be more easily detected. However, the general conclusions which can be derived from Figures 4.8 and 4.9 (i.e., the dependency of posterior values from wetness conditions) do not change. On the other hand, a small window might aid in more detailed analyses of process dynamics. For instance, identifiability patterns changed slightly for Evapotranspiration when using a smaller window, especially for gauge Villacarli, where more periods of high identifiability appeared, where nothing could be seen with the original window size. For the other gauges, patterns became more diverse as well, but not relevant regarding the general conclusions of this study.

This work addressed the identifiability of different input factors reflecting uncertainty in the choice of equally likely process representations, ODE solvers, and parameter realizations. However, the identifiability measure was defined in a way that each input factor has to be analyzed separately. This resulted from the discrete distribution of input factors and the highly different numbers of realizations. To overcome this issue, Pianosi and Wagener (2016) sampled the prior and posterior distributions with an alternative approach and were able to compare time-varying sensitivity indices of discrete input factors. In that way they also applied a mathematically more rigorous approach than the informal MC filtering applied in this study to derive the posterior

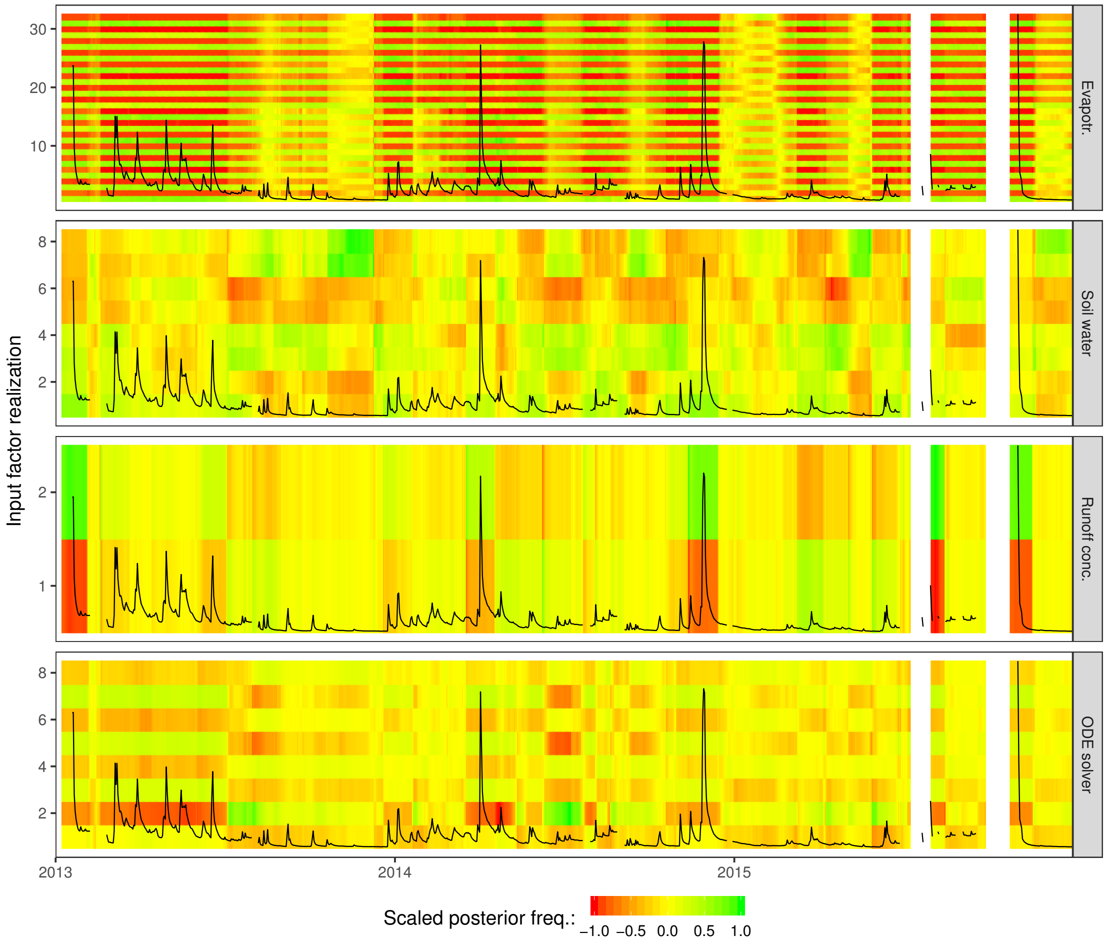

Figure 4.8: Posterior frequencies over the simulation period for gauge Capella (the catchment outlet) scaled to $[-1, 1]$. Negative (positive) values indicate posterior frequencies less (larger) than prior frequency values. Black lines represent the observed discharge hydrograph.

distribution. Their method, however, requires a larger sample size and, hence, more model runs, which was not achievable with the complex model structures employed in this work. Other approaches, such as global sensitivity analysis, rely on assumptions that input factors are uncorrelated or at least continuously distributed. For the future, in order to directly compare the impacts of different input factors for process-based models, an approach dealing with correlated input factors of discrete distribution is required, which is also computationally feasible.

4.5.3 Spatiotemporal Patterns of Identifiability

Static analysis over the whole simulation period as well as the dynamic approach using a moving window found evapotranspiration and parametrization to be the only input factors showing a certain degree of identifiability. Yet the temporal analysis provides much more insights into model functioning. It allows for a more detailed comparison of prior and posterior distributions and reveals temporal patterns also for the other input factors, which are overall poorly identifiable. In addition, the analysis shows that the time-varying dominance of certain input factors is to a large degree driven by meteorological conditions. This conclusion is well in line with other studies emphasizing the added value of a temporal analysis of sensitivity or identifiability (e.g., Ghasemizade et al., 2017; Guse et al., 2014; Herman et al., 2013; Pianosi and Wagener, 2016; Reusser and Zehe, 2011; Savage et al., 2016).

It was found that during wet periods the PM approach clearly dominated the posterior distribution and the parameter cal_wind was directed towards reduced evapotranspiration amounts. During dry periods, the SW approach was dominant with a less clear pattern

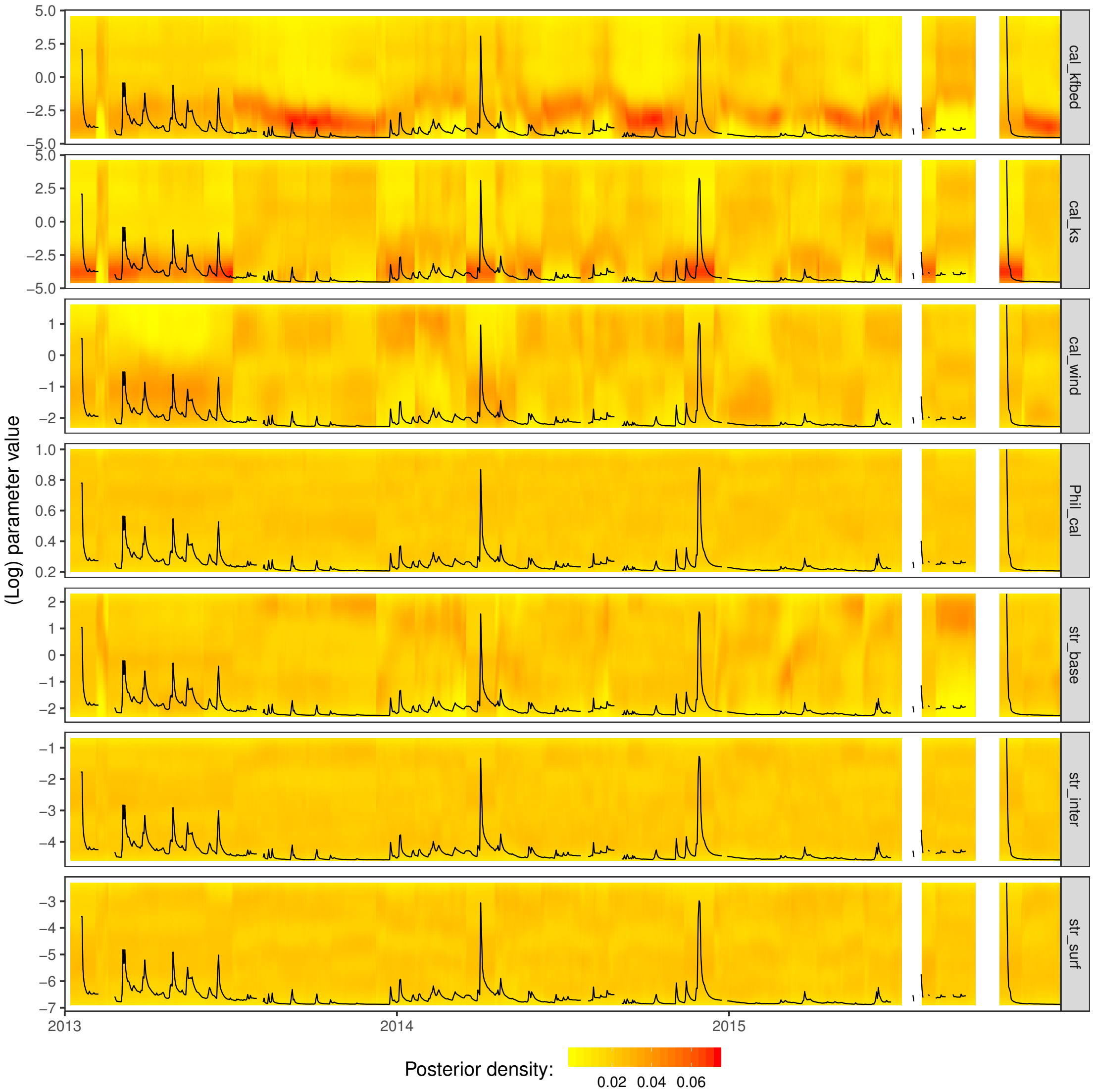

Figure 4.9: Posterior densities of model parameters derived via kernel density estimation and scaled to sum up to one at each time step for each parameter. Parameter values are log-transformed, except for Phil_cal. Shown is only gauge Capella, the catchment outlet. Black lines represent the observed discharge hydrograph.

for cal_wind. This is supported by the fact that the SW approach was most dominant for subcatchment Lascuarre, which is the driest part of the study area. The SW formula relaxes the big-leaf assumption of the PM approach in a way that it accounts for bare soil and is therefore a more sophisticated approach for sparse crops and patchy vegetation (Shuttleworth and Wallace, 1985). While subcatchment Lascuarre has the largest fraction of cropland, it is not clear whether this really translates to a more patchy vegetation and thus no better suitability of either approach can be inferred a priori. It rather seems that moisture condition is the most influential factor for the selection of an evapotranspiration model in comparison to landscape characteristics (full vs. patchy vegetation cover).

An interesting finding is that unconstrained ODE solvers with high accuracy perform better during wet periods. A possible explanation is that implausible model states, which are likely caused by unconstrained solvers under rainfall conditions, can still produce a more realistic streamflow dynamic. This can be attributed to the faster soil water fluxes with characteristic time scales less than the model's temporal resolution. Under such circumstances ODE solvers without solution constraints could serve as compensation for the rather coarse daily resolution of the model runs. For instance, consider a high amount of daily precipitation, which occurs in fact just within a few hours rather than equally over a full day. A large signal of rainfall input within a single model time step would likely cause a high amount of surface runoff, while if the signal would be distributed over several sub steps, it would rather result in less overland flow and a less sharp runoff signal. In that way, the temporary storage of water in the soil, although exceeding the physical boundaries, is delaying the runoff signal and could therefore still result in a more realistic system behavior. In contrast, under dry conditions, constrained ODE solvers are favored as they keep the model states within physical limits, which eventually results in a more realistic streamflow dynamic.

Apart from the temporal patterns, some differences among the subcatchments could be found. This is especially true for gauge Lascuarre in comparison to the rest of the catchment. Apart from the issue of the evapotranspiration model, which already has been discussed, constrained ODE solvers are more clearly favored at this than at the other gauges. In general, this can be attributed to the distinct hydrological and meteorological conditions in this subcatchment (see Table 4.1). In contrast to the other subcatchments, Lascuarre is characterized by a very low runoff coefficient, sharper discharge peaks, less precipitation, less steep topography, and more agricultural areas. This supports the findings of van Werkhoven et al. (2008) who found distinct patterns of parametric controls in dry and wet catchments. Their findings for parametric controls can therefore be extended to certain process realizations and even ODE solvers.

4.5.4 Is There an Optimal Model Structure?

The most straightforward approach to address the question of the optimal model structure would be to select the best performing realization. On the other hand, it was shown that the best performing model run does not necessarily refer to the highest occurrences in the posterior distributions of the analyzed input factors (Figure 4.4). This suggests a high influence of parametrization, i.e., only very specific parameter values result in a good performance of a certain model structure, while changes in the parameters may significantly deteriorate model behavior.

It was found that spatial variability, even in small catchments such as investigated in this work, can be substantial and lead to contrasting conclusions in neighboring subcatchments. In combination with the identified temporal patterns, these findings allow for more general conclusions. For instance, it was consistently found that under dry conditions the SW model is a more plausible evapotranspiration model, while under wet conditions the PM approach was favored. In future studies, more advanced methods, such as machine learning techniques, could be employed to derive relationships between catchment characteristics and meteorological conditions as predictors and certain model process formulations as response variable. This would allow to design the most likely model configuration prior to an application based on the characteristics of the catchment to be investigated. Flexible simulation environments such as ECHSE (among others) enable such a task and serve as a toolbox for the modeller (Clark et al., 2015b, 2008b; Fenicia et al., 2011; Kneis, 2015).

It should be kept in mind, after all, that the complexity of the problem makes it difficult to draw general conclusions. Quite surprisingly, ODE solvers of low accuracy achieved rather high rankings in the posterior distribution. This suggests that model deficiencies can be easily compensated by, albeit unrealistic, parametrizations or process formulations. Consequently, there is a high chance to obtain the right answers for the wrong reasons, a phenomenon resulting from, e.g., overparametrization of a model, which has been acknowledged in numerous studies in the field of hydrological modeling (e.g., Fenicia et al., 2016; Kavetski and Clark, 2010; Kirchner, 2006; Samaniego et al., 2010; Schoups et al., 2008). As a consequence, the use of unconstrained ODE solvers should be avoided, even if they achieve good model performances. In general, ODE solvers of high accuracy should be used. However, this as well calls for finer discretizations (in space and time) than they were used in this study. Separate analysis of the soil water module in ECHSE revealed that simulations with accurate ODE solvers are of much more use when running the model with shorter temporal resolution (not more than one hour) and more soil horizons with gradually varying soil parameters (not shown in this paper). On the other hand, model runtimes increase dramatically and prevent detailed analyses of ensemble-based approaches as presented in this study.

4.6 Conclusions

This study investigated the spatiotemporal identifiability of multiple model structures. The experiments were conducted in a small mountainous catchment in northeastern Spain. To carry out simulations, the flexible simulation environment ECHSE was used, which enabled the rapid implementation of different alternatives for process representation (with respect to subprocesses of evapotranspiration, soil water movement, and runoff concentration) and ODE solvers. Model configurations were in general based on the process-based hydrological and sedimentological model WASA-SED, whose process formulations have been transferred into ECHSE and extended by alternatives. This was the first approach analyzing complex process representations, ODE solvers, and parametrizations in a fully integrated manner by coupling the flexible model environment with dynamic identifiability analysis (DYNIA).

Overall, the approach proved to be useful for the identification of complex process-based model structures. With respect to the initially stated research questions, the main findings shall be briefly summarized.

1. Parametrization and subprocesses of evapotranspiration turned out to be the only identifiable input factors. Yet different patterns could be identified in the posterior distributions also for the other input factors. Surprisingly, ODE solvers without solution constraints achieved better simulation performances than those with constraints.
2. Identifiability patterns vary over time, especially for the input factor of evapotranspiration routines. As such, the Penman–Monteith approach appeared to be superior during wet periods while during dry periods the Shuttleworth–Wallace approach led to better model performances. Moreover, unconstrained ODE solvers with high accuracy performed better during wet periods, while solvers with solution constraints obtained better model performance during dry periods.
3. The results of model identification are clearly influenced by hydrological characteristics. While identifiability patterns are relatively consistent over areas with similar hydrological characteristics, identified model structures are most distinct for the subcatchment with the most diverging characteristics with respect to land use, topography, rainfall sum, and runoff coefficient.

It should be noted that identifiability patterns might be influenced by correlation among input factors and compensation effects, which can distort the general findings. This could explain why sometimes ODE solvers of low accuracy achieved good model results and therefore high rankings in the posterior distribution. Therefore, it is difficult to decide for a specific model configuration, as the model obtaining the best performance metrics might be influenced by such compensation effects. Consequently, the following questions could be addressed by future research:

1. How can the compensation effects be eliminated and model identification made more robust? How do temporal resolution and ODE solvers dictate these issues and how do they interact?
2. If temporal resolution and ODE solvers are crucial, how can they be addressed while maintaining feasible model runtimes?
3. Could data science (e.g., machine learning) be combined with process knowledge to determine the most adequate model structure for a study area before conducting time consuming model evaluations?

Acknowledgments

The data, R scripts, and ECHSE model, on which this study is based, were published online. Reviewers may use the following link to anonymously access and review the data and R scripts to reproduce identifiability analysis: `https://tinyurl.com/DataScripts`. The ECHSE framework with extensions as described in the manuscript is available from: `https://tinyurl.com/EchseModel`. In case the manuscript is published, the temporary links will be made permanent with associated DOI provided by GFZ Data Services (a repository complying the FAIR Data Principles). The authors thank Thorsten Wagener for his valuable suggestions on the general methodology and the DYNIA framework. Tobias Pilz acknowledges funding by the Helmholtz graduate school GeoSim and University of Potsdam.

4.A Supporting Information

Background information on ODE solvers

In many hydrological models, fluxes r are calculated based on the model state S of the current time step i:

$$r_i = f(S_i) \tag{4.3}$$

Model states for the next time step are then simply updated by multiplication of fluxes with time step length h (the so-called explicit Euler method):

$$S_{i+1} = S_i + r_i h \tag{4.4}$$

This, however, may lead to discrepancies as simulated processes (such as evapotranspiration or infiltration) often change within time scales much smaller than h (e.g., minutes or hours while h, in this study as in many others, is one day).

To catch unrealistic results, in many models calculated fluxes and model states are checked for physical realism (e.g., soil moisture should range between saturated and residual water content defined by soil texture) and adjusted if necessary. This adaptation is referred to as *solution constraint*.

To account for subscale temporal dynamics and possible feedbacks between fluxes and states, equations 4.A and 4.A are typically applied for each process individually in a prescribed order, such that state variables are updated successively within a time step (e.g., in WASA-SED: lateral inflow → precipitation and interception → infiltration → evapotranspiration → percolation and lateral re-distribution). In a mathematical sense, this approach is referred to as *operator splitting* and is popular as the problem can be subdivided into parts which are easier to integrate (Schoups et al., 2010). This, however, comes at the cost of a "splitting error" and is unsatisfying from a physical point of view, as processes in reality do not occur sequentially but take place concurrently. Therefore, the ECHSE-based model structures built for this study do not contain operator splitting.

In ECHSE, several ODE solvers were implemented from the *GNU Scientific Library* (Galassi et al., 2017). Explicit Runge–Kutta methods calculate state S_{i+1} from S_i as a weighted average of k increments within the fixed interval h. The explicit Euler method is a special case where $k = 1$ resulting in first-order accurate solutions, while for fourth-order methods $k = 4$ with fourth-order accurate results. Adaptive step size methods differ from methods with fixed step size in a way that h is adaptively split into subintervals until the estimated local error of the integration is smaller than a predefined tolerance. The Backward Differentiation Formula (BDF) is another approach considered in this work, which belongs to the family of linear multistep methods. Instead of discarding all previous information before taking the next time step integration, as is done by Runge–Kutta methods, linear multistep algorithms use a linear combination of derivatives from previous time steps. Implicit methods as another group first estimate S_{i+1} in an explicit manner, as illustrated by equation 4.A, followed by a corrector step, where an improved estimation of S_{i+1} is derived implicitly as (in case of the implicit Euler method)

$$S_{i+1} = S_i + hf(S_{i+1}). \tag{4.5}$$

Such methods generally have the advantage over explicit methods that they provide more stable solutions, but require, especially in case of stiff equations (equations that may cause rapid variation in the solution), much more computation time. This is the reason why in this study no implicit methods (e.g., implicit Runge–Kutta solvers) except for BDF (which also contains a predictor–corrector method) could be considered, although they can be used within ECHSE. Test runs revealed that they cannot produce solutions within acceptable computation time within this framework.

Equations of Implemented Process Representations

Table 4.5: Equations of evapotranspiration processes.

ID	Equation
SW	$et = \frac{1}{E_{wat}}\left[C_c PM_c + C_s PM_s\right]$ $PM_c = \frac{sA + (\rho_{air} C_{air} D - s r_{ca} A_s)/(r_{aa}+r_{ca})}{s+\gamma\left[1+r_{cs}/(r_{aa}+r_{ca})\right]}, \quad PM_s = \frac{sA + \left[\rho_{air} C_{air} D - s r_{sa}(A-A_s)\right]/(r_{aa}+r_{sa})}{s+\gamma\left[1+r_{ss}/(r_{aa}+r_{sa})\right]}$ $C_c = \frac{1}{1+R_c R_a/R_s(R_c+R_a)}, \quad C_s = \frac{1}{1+R_s R_a/R_c(R_s+R_a)}$ $R_a = (s+\gamma) r_{aa}, \quad R_c = (s+\gamma) r_{sa} + r_{ss}, \quad R_s = (s+\gamma) r_{ca} + r_{cs}$
PM	$et = \frac{1}{E_{wat}}\left[\frac{s(R_{net}-G_{soil}) + \rho_{air} C_{air}(E-e)/r_{aa}}{s+\gamma(1+r_{cs}/r_{aa})}\right]$
SK	$r_{cs} = r_{l,act}\epsilon / ln\left[\frac{g_{srad}+\epsilon R_{inS}}{g_{srad}+\epsilon R_{inS} exp(-\epsilon L)}\right]$
SW19	$r_{cs} = \frac{r_{l,act}}{2L}$
SG43	$z = \begin{cases} z_{r,s} + 0.3 h_c (c_d L)^{0.5} & \text{if } 0 < c_d L < 0.2 \\ 0.3 h_c (1 - \frac{d}{h_c}) & \text{if } 0.2 < c_d L < 1.5 \end{cases}$
BT	$z = \begin{cases} 0.123 h_c & \text{if } h_c \leq 2 \\ 0.058 h_c^{1.19} & \text{otherwise} \end{cases}$
SG42	$d = 1.1 h_c ln(1 + (c_d L)^{0.25})$
SG41	$d = \frac{2}{3} h_c$
AN	$R_{inS,cs} = (a_s + b_s)\, R_{ex}$
AL	$R_{inS,cs} = \left(0.75 + 2 \times 10^{-5} h\right) R_{ex}$

Symbols: E_{wat}: latent heat of water evaporation; s: slope of saturation vapour pressure curve; γ: psychrometric constant; ρ_{air}: density of air; C_{air}: specific heat of moist air; D: vapour pressure deficit at canopy source height; e: vapour pressure; E: vapour pressure at saturation; $A_s = R_{net,soil} - G_{soil}$: total energy available at soil surface; $R_{net,soil}$: incoming net (short- and long-wave) radiation hitting the soil surface; G_{soil}: soil heat flux; $A = R_{net} - G_{total}$: total energy available at measurement height (should be above canopy); R_{net}: incoming net (short- and long-wave) radiation at measurement height; G_{total}: heat flux into soil, vegetation, and air below measurement height; r_{aa}: aerodynamic resistance; r_{ss}: soil surface resistance; r_{ca}: bulk boundary layer resistance of the vegetative elements in the canopy; r_{sa}: aerodynamic resistance between soil and canopy source height; $r_{l,act}$: surface resistance of a single leaf; ϵ: canopy extinction coefficient; g_{srad}: solar radiation for which stomatal conductance is half of its maximum value; R_{inS}: incoming short-wave radiation (above canopy); L: leaf area index; $z_{r,s}$: roughness length of bare soil; h_c: canopy height; c_d: drag coefficient of vegetative elements; R_{ex}: extraterrestrial radiation; a_s, b_s: Ångström coefficients

Table 4.6: Equations of soil water processes.

ID	Equation
GA	Solve iteratively: $F_{\Delta t} = k_f(\Delta t - t_{sat}) + n_r\psi_f \cdot ln\left(\frac{F_{\Delta t}+n_r\psi_f}{F_s+n_r\psi_f}\right) + F_s$
	$F_s = D_{wet}n_r, \quad D_{wet} = \frac{\psi_f}{R_f/k_f-1}, \quad t_{sat} = \frac{F_s}{R_f}$
PH	$F = St^{1/2} + At, \quad S = \sqrt{2k_f(\theta_s-\theta)\psi_f}$
PS	$q_{perc} = SW_{perc}\left[1 - exp\left(\frac{-\Delta t}{TT}\right)\right]/\Delta t$
	$SW_{perc} = \begin{cases} 0 & \text{if } \theta \leq \theta_{fc} \\ D_{hor}(\theta-\theta_{fc}) & otherwise \end{cases}, \quad TT = \frac{SW_{perc}}{k_u}$
PR	$q_{perc} = \begin{cases} k_u(\Delta\psi/D_{hor}+1) & \text{if } \Delta\psi > 0 \\ 0 & \text{otherwise} \end{cases}$
VG	$S_e = \frac{\theta-\theta_r}{\theta_s-\theta_r} = \left[\frac{1}{1+\left(\psi/\psi_b\right)^{\lambda+1}}\right], \quad k_u = k_f S_e^{1/2}\left[1-\left(1-S_e^{\frac{\lambda+1}{\lambda}}\right)^{\frac{\lambda}{\lambda+1}}\right]^2$
CB	$S_e = \frac{\theta}{\theta_s} = \left(\frac{\psi_b}{\psi}\right)^{\lambda}, \quad k_u = k_f\left(S_e\right)^{3+2/\lambda}$

Symbols: k_f: saturated hydraulic conductivity; Δt: time step length; n_r: refillable porosity; ψ_f: capillary suction at the wetting front; R_f: surface water flux for infiltration; θ: actual water content; θ_s: water content at saturation; θ_{fc}: water content at field capacity; θ_r: residual water content; t: time since start of infiltration event; A: calibration parameter (Phil_cal in Table 4.4); D_{hor}: thickness of soil column; k_u: unsaturated hydraulic conductivity; ψ: capillary suction; ψ_b: bubbling capillary pressure; λ: pore size index

Input factor realizations

Table 4.7: Input factor realizations.

ID	Realization				
Ia: Evapotranspiration					
1	PM	SW19	BT	SG41	AN
2	SW	SW19	BT	SG41	AN
3	PM	SK	BT	SG41	AN
4	SW	SK	BT	SG41	AN
5	PM	SW19	SG43	SG41	AN
6	SW	SW19	SG43	SG41	AN
7	PM	SK	SG43	SG41	AN
8	SW	SK	SG43	SG41	AN
9	PM	SW19	BT	SG42	AN
10	SW	SW19	BT	SG42	AN
11	PM	SK	BT	SG42	AN
12	SW	SK	BT	SG42	AN
13	PM	SW19	SG43	SG42	AN
14	SW	SW19	SG43	SG42	AN
15	PM	SK	SG43	SG42	AN
16	***SW***	***SK***	***SG43***	***SG42***	***AN***
17	PM	SW19	BT	SG41	AL
18	SW	SW19	BT	SG41	AL
19	PM	SK	BT	SG41	AL
20	SW	SK	BT	SG41	AL
21	PM	SW19	SG43	SG41	AL
22	SW	SW19	SG43	SG41	AL
23	PM	SK	SG43	SG41	AL
24	SW	SK	SG43	SG41	AL
25	PM	SW19	BT	SG42	AL
26	SW	SW19	BT	SG42	AL
27	PM	SK	BT	SG42	AL
28	SW	SK	BT	SG42	AL
29	PM	SW19	SG43	SG42	AL
30	SW	SW19	SG43	SG42	AL
31	PM	SK	SG43	SG42	AL
32	SW	SK	SG43	SG42	AL

ID	Realization		
Ib: Soil			
1	PH	PS	VG
2	***GA***	***PS***	***VG***
3	PH	PR	VG
4	GA	PR	VG
5	PH	PS	CB
6	GA	PS	CB
7	PH	PR	CB
8	GA	PR	CB
Ic: Runoff concentration			
1	***RW***		
2	RS		
II: ODE solvers			
1	EU	FREE	
2	***EU***	***CONS***	
3	RK	FREE	
4	RK	CONS	
5	RKST	FREE	
6	RKST	CONS	
7	BDF	FREE	
8	BDF	CONS	

III: Parametrization

ID	cal_wind	cal_ks	cal_kfbed	Phil_cal	str_surf	str_inter	str_base
1	2.130	1.980	0.091	0.455	0.003	0.147	5.720
2	0.635	15.911	3.374	0.796	0.025	0.176	0.980
3	1.844	0.029	0.022	0.290	0.022	0.065	0.161
...	...	...	...	...	...	...	...
998	0.185	0.021	3.823	0.226	0.093	0.290	2.514
999	0.400	1.451	0.136	0.433	0.011	0.030	4.773
1000	4.455	0.458	0.147	0.702	0.045	0.076	2.340

Note: Highlighted realizations mark the WASA-SED implementations.

5. Seasonal drought prediction for semiarid northeast Brazil: what is the added value of a process-based hydrological model?

Abstract

The semiarid northeast of Brazil is one of the most densely populated dryland regions in the world and recurrently affected by severe droughts. Thus, reliable seasonal forecasts of streamflow and reservoir storage are of high value for water managers. Such forecasts can be generated by applying either hydrological models representing underlying processes or statistical relationships exploiting correlations among meteorological and hydrological variables. This work evaluates and compares the performances of seasonal reservoir storage forecasts derived by a process-based hydrological model and a statistical approach.

Driven by observations, both models achieve similar simulation accuracies. In a hindcast experiment, however, the accuracy of estimating regional reservoir storages was considerably lower using the process-based hydrological model, whereas the resolution and reliability of drought event predictions were similar by both approaches. Further investigations regarding the deficiencies of the process-based model revealed a significant influence of antecedent wetness conditions and a higher sensitivity of model prediction performance to rainfall forecast quality.

Within the scope of this study, the statistical model proved to be the more straightforward approach for predictions of reservoir level and drought events at regionally and monthly aggregated scales. However, for forecasts at finer scales of space and time or for the investigation of underlying processes, the costly initialisation and application of a process-based model can be worthwhile. Furthermore, the application of innovative data products, such as remote sensing data, and operational model correction methods, like data assimilation, may allow for an enhanced exploitation of the advanced capabilities of process-based hydrological models.

Published as:

Pilz, T., Delgado, J. M., Voss, S., Vormoor, K., Francke, T., Costa, A. C., Martins, E., and Bronstert, A. (2019a). “Seasonal drought prediction for semiarid northeast Brazil: what is the added value of a process-based hydrological model?” In: *Hydrol. Earth Syst. Sc.* 23.4, pp. 1951–1971. DOI: 10.5194/hess-23-1951-2019

5.1 Introduction

Drought is a type of natural hazard characterised by meteorological, hydrological, and water management conditions, affecting many regions around the globe. Generally, it arises due to a shortage of water availability. A general valid or comprehensive definition, however, is hardly achievable due to many different possible causes, complex relationships, and feedbacks among its determining factors and, consequently, different impacts on nature, society, and economy. As such, different categories can be distinguished ranging from meteorological (lack of rainfall) and hydrological (shortage of consumable water resources) to agricultural (water deficit for crops or husbandry) and socio-economic droughts (not enough of income to pay water price). For the characterisation of droughts, different statistics can be computed describing duration, frequency, and severity based on various predictors and thresholds (Mishra and Singh, 2010).

The semiarid northeast of Brazil (NEB) is one of the world's most densely populated dryland regions (Marengo et al., 2017). Its climate is characterised by a short rainy season with high interannual variability. As a consequence, already since the colonisation in the 16th century, regularly occurring severe droughts causing famine and mass exodus have been reported. Drought occurrence is primarily driven by sea surface temperature (SST) anomalies in the eastern Pacific, i.e. the El Niño–Southern Oscillation (ENSO), and the northern tropical Atlantic region (i.e. the Tropical Atlantic SST Dipole) influencing the location of the Intertropical Convergence Zone (ITCZ), which is the main source of rain during the rainy season in the NEB area (Hastenrath, 2012). Profound governmental actions for drought mitigation since the late 19th century resulted, among others, in the construction of thousands of small reservoirs and several large dams for water storage and provision within the dry season and during dry spells. Still, severe drought events might endanger water supply, as has happened in the current series of drought years since 2012. Even the regular years (in terms of rainfall amount) of 2017 and 2018 were not able to eliminate or significantly alleviate water scarcity, resulting in filling states of the largest reservoirs of less than 10 % (for the current state of water provision and statistics of the state of Ceará see `http://www.hidro.ce.gov.br`; last access: 6 April 2019 and the drought monitor `http://msne.funceme.br`; last access: 6 April 2019). In addition, climate change is likely to aggravate water scarcity, calling for efficient strategies in the management of water storages (de Araújo et al., 2004; Braga et al., 2013; Krol et al., 2006).

Reliable seasonal forecasting, i.e. forecasts of streamflow and reservoir storages for the upcoming rainy season, can be of significant value for water managers (Sankarasubramanian et al., 2009). Accurate precipitation forecasts over several months are still a challenge for dynamical climate models. However, many dryland regions are located in areas with distinct dry and rainy seasons, the latter often connected to large-scale atmospheric circulation patterns. Therefore, statistical models relating meteorological or SST indices with streamflow or a combination of statistical and process-based models are applied in many dryland regions in the world to provide seasonal forecasts (e.g. Schepen and Wang, 2015; Seibert et al., 2017; Sittichok et al., 2018).

For the northern NEB region, the high correlation of rainfall and droughts with SST anomalies in the eastern Pacific and tropical Atlantic, together with correlation of preseason rainfall, offers a favourable setting for seasonal prediction (Hastenrath, 2012; Souza Filho and Lall, 2003; Sun et al., 2006). Several studies exist for the area, typically employing one or several (realisations of) general circulation models (GCMs) driven by SST predictions, downscaled to a finer scale by statistical or dynamical downscaling approaches, whose meteorological (especially rainfall) outputs are eventually used as forcing in a hydrological model producing streamflow and/or reservoir level forecasts. For instance, Galvão et al. (2005), Block et al. (2009), and Alves et al. (2012) employed different hydrological models of varying complexity to generate streamflow and/or reservoir level predictions. While model performance over daily timescales was generally reported to be low, over longer aggregation periods, such as at a monthly or seasonal scale, acceptable results could be achieved.

In a recent study, Delgado et al. (2018b) investigated the use of a statistical relationship to provide seasonal reservoir level predictions. They used the two GCMs ECHAM4.6 and ECMWF, with the meteorological output of each downscaled by three different statistical approaches, generating ensembles of wet-season (i.e. January to June) hindcasts for each year in the period 1981 to 2014. Based on these meteorological hindcasts, they calculated a number of meteorological drought indices which are compared with observations to evaluate the skill of the predictions. Using reservoir storage as a target variable, they further computed hydrological drought indices and fitted a multivariate linear regression to predict these indices using the meteorological indices as predictors. Even though there was variation among the GCM and downscaling combinations, the occurrence of meteorological drought could mostly be predicted with skill. Furthermore, their relatively simple statistical model was able to predict also hydrological droughts with skill. However, the absolute hindcast error was often not appreciably better than climatology, i.e. the observed long-term average of a variable.

While being straightforward to apply and computationally advantageous, such statistical relationships, in contrast to process-based hydrological models, do not represent underlying processes and are less flexible in terms of the output variable and their spatial and temporal resolution. However, what remains is the question of how to balance accuracy, operability, and usability from the perspective of water managers and stakeholders. As such, this study complements the work of Delgado et al. (2018b), employing a process-based hydrological model instead of a statistical model. Thus, the aim is to present and evaluate a forecasting system, predicting seasonal reservoir levels and the occurrence of hydrological droughts for the Jaguaribe River basin, located within the NEB region. Three different objectives are put into focus: first, the process-based hydrological model and the statistical model of Delgado et al. (2018b) shall be evaluated and compared in terms of reservoir level simulation performance. Second, the process-based hydrological model as an operational forecasting tool is to be verified in a hindcast experiment. Third, major sources of prediction and simulation errors in the modelling system are to be investigated. Thereby, the question of whether the costly initialisation and use of a complex hydrological model is worthwhile in comparison to a much simpler statistical relationship is to be answered, and guidelines for further research and the improvement of the forecasting system shall be given.

This study touches on issues of atmospherical sciences, hydrology, and water resources management. As terminology partially differs, a clarification on certain terms used throughout the paper can be consulted in Appendix, Sect. 5.A.1.

5.2 Study site

The study area comprises the Jaguaribe River basin in the state of Ceará, northeast Brazil (see Fig. 5.1). The catchment is of crucial importance in terms of water supply for the whole state and has been intensively investigated in numerous studies (e.g. de Araújo et al., 2004; Bronstert et al., 2000; de Figueiredo et al., 2016; Gaiser et al., 2003; Krol et al., 2006; Mamede et al., 2012; van Oel et al., 2012). It covers an area of about 70 000 km^2 with a rural population of 2.7 million. Additionally, it is the source of water for the metropolitan area of Fortaleza with 2.6 million people (IPECE, 2016). Annual precipitation sums up to, on average, 755 mm per year, whereas 90 % of rainfall occurs within the rainy season between January and June. Potential evapotranspiration is high with more than 2000 mm per year. The mean annual temperature is about 25 °C with little variation. Rainfall, however, is mostly convective with only a few events of high intensity per year and a strong inter-annual variation caused by SST anomalies resulting in a northward shift of the ITCZ inducing recurrent droughts that can last over several years (see also Sect. 5.1; Hastenrath, 2012; Marengo et al., 2017). As the geology is characterised by a primarily crystalline basement with low-density fractures, water supply needs to be secured by surface water resources. Accordingly, thousands of small and several large reservoirs were constructed. The small reservoirs are typically bordered by uncontrolled earth dams, mainly serving for water provision of rural population and livestock. Conversely, large so-called *strategic reservoirs* contain a barrage with intake devices for active regulation, are sometimes also used for hydropower production, and serve as water resources for larger towns and cities and industrial farming. These settings cause meteorological droughts (lack of precipitation) and hydrological droughts (lack of surface water) to be often out of phase (de Araújo and Bronstert, 2016; van Oel et al., 2018).

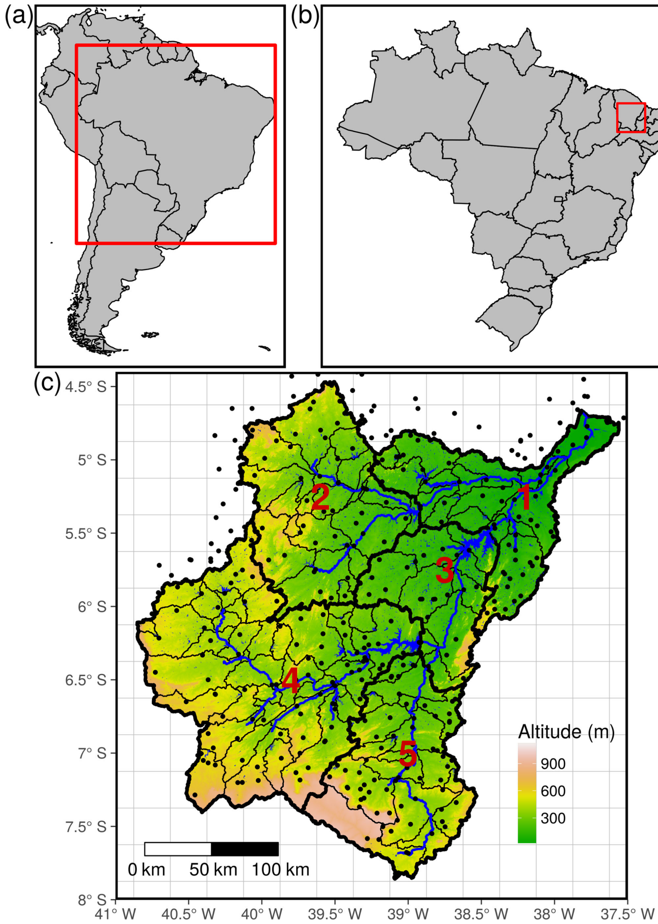

Figure 5.1: Overview over the Jaguaribe watershed **(c)** and location within Brazil **(b)** and South America **(a)**. The five regions of interest (red numbers) are (1) Lower Jaguaribe, (2) Banabuiú, (3) Castanhão, (4) Orós, and (5) Salgado. Thin black lines in **(c)** outline subbasins, which are the computational units within the model. Black dots are rainfall stations considered within the study. Background grid lines refer to the gridded meteorological dataset of Xavier et al. (2016).

For the present study, the Jaguaribe catchment was subdivided into five subregions, named after the main tributary river or the major reservoir at its outlet: Banabuiú, Orós, Salgado, Castanhão, and Lower Jaguaribe (see Fig. 5.1 for their location).

5.3 Data and methods

5.3.1 General workflow

The aim of this study is to elucidate the application potential of a process-based hydrological model for water resources and drought prediction. Consequently, hindcasts of reservoir volumes and hydrological drought indices shall be produced, driving the model by meteorological hindcasts. The general workflow is illustrated in Fig. 5.2.

A process-based hydrological model was first calibrated to observations and an initial model run conducted for the period of 1980 until 30 June 2014. This initialisation run was driven by observed meteorology and at each 1 January the storage volume of each strategic reservoir was replaced by the observed value. Furthermore, if available, measured reservoir releases through a dam's intake devices were fed into the model in order to make use of as much information as available to produce simulations as realistic as possible. The first year of the run was used as a warm-up to bring the model states into equilibrium. At each end of year, the model's state variables, including soil moisture, groundwater, river, and small (i.e. non-strategic)

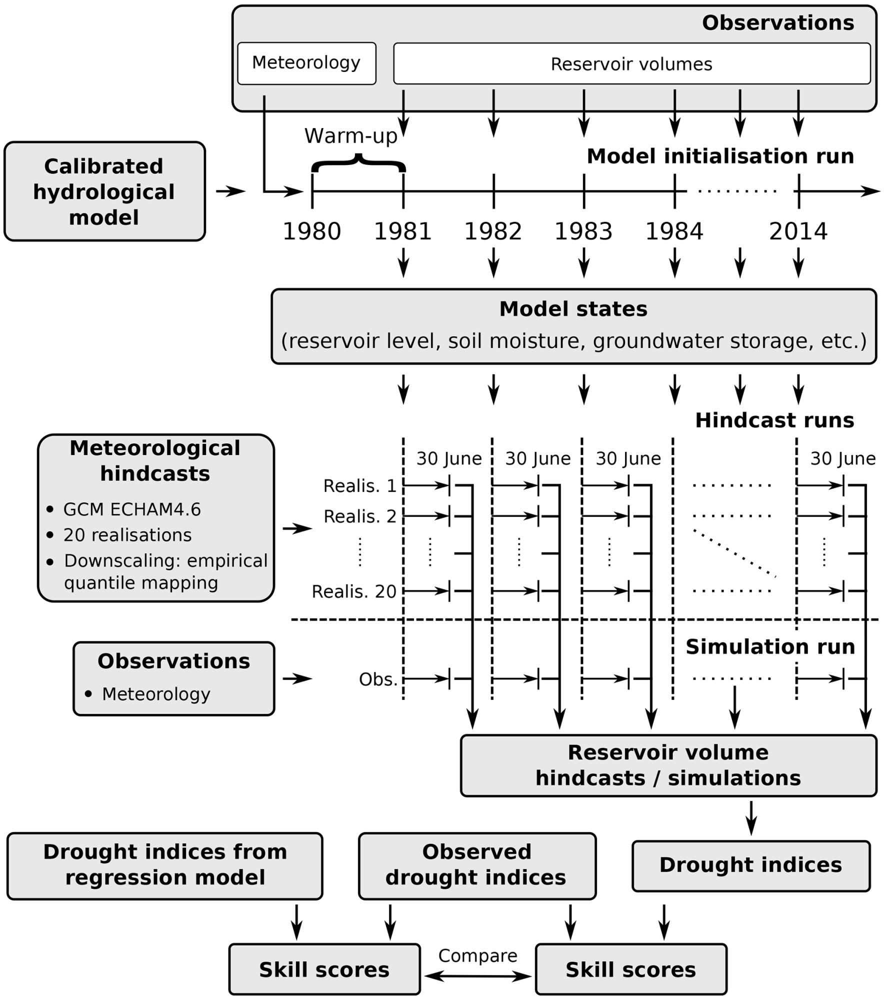

Figure 5.2: Workflow for the generation and evaluation of hindcasts of hydrological drought indices.

reservoir storages, were stored. This entire procedure is intended to mimic the conditions in a real forecast situation.

In a specific hindcast run, the model was then re-initialised with the saved model states and driven by hindcast meteorology. These runs were conducted successively for the wet seasons (1 January to 30 June) of 1981 to 2014. The resulting strategic reservoir volumes were used to infer drought indices which were evaluated employing verification metrics. To distinguish uncertainties from the meteorological hindcasts and in order to investigate mere model performance, the model runs were performed in two ways: driven by observations (*simulation mode*) and meteorological hindcasts (*hindcast mode*).

The runs were conducted for both the process-based model initialised and calibrated within this study and a statistical model, which is a regression approach derived by Delgado et al. (2018b) for the same study area. Consequently, verification metrics were calculated and analysed for both model approaches and both forcing modes.

In order to identify the strengths and weaknesses of the process-based model, the results of the simulation runs were further analysed. In this context, the model output (reservoir storage) was stratified. The details of the individual processing steps are described in the following.

5.3.2 Data

To parametrise the hydrological model, various spatial data were obtained including a 90 m $\times$ 90 m SRTM digital elevation model (DEM), a soil map provided by the Research Institute for Meteorology and Water Resources of the state of Ceará (FUNCEME) along with soil parameters from a local database (Jacomine et al., 1973) from which the necessary model parameters were calculated employing pedotransfer functions, a land cover map from the Brazilian Ministry of the Environment with parametrisations assembled by Güntner (2002), and a map of small and strategic reservoirs provided by FUNCEME. Reservoir parameters were made available by the Company for Water Resources Management of Ceará (COGERH) and FUNCEME and include the year of dam construction, storage capacity, and water-level–lake-area–storage-volume relationships along with daily resolution time series of water levels and artificial water release. A time series of daily precipitation for 380 stations within and in close vicinity around the study area were provided by FUNCEME. Other daily meteorological time series needed by the model (relative humidity, air temperature, and incoming shortwave radiation) were derived from the gridded dataset ($0.25° \times 0.25°$ resolution) of Xavier et al. (2016).

5.3.3 Meteorological hindcasts

Daily meteorological hindcast data for the period 1981 to 2014 used as input into the hydrological model stem from an ensemble (20 members) of ECHAM4.6 GCM runs (Roeckner et al., 1996) which were bias corrected by empirical quantile mapping (EQM) (Boé et al., 2007; Gudmundsson et al., 2012). Although Delgado et al. (2018b) identified some deficiencies regarding this product, there was no clear better-performing alternative. In addition, ECHAM4.6 is already employed operationally by the local water authority FUNCEME (Sun et al., 2006), and, in contrast to other seasonal forecast systems like those by ECMWF, it comes without further costs for operational use, making it the candidate for future operational application. The 20-member ensemble runs of ECHAM4.6 were conducted and results provided by FUNCEME. More information is given in Delgado et al. (2018b).

5.3.4 The process-based model

Introduction to WASA-SED

The hydrological model WASA-SED, version rev_257, was employed for the process-based hindcasts of reservoir volumes. WASA-SED is a deterministic, process-based, semi-distributed, time-continuous hydrological model. The representation of hydrological processes focuses on dryland environments. A complex but efficient hierarchical spatial disaggregation scheme allows for application over large scales up to an order of magnitude of 100 000 km^2 (Güntner and Bronstert, 2004; Mueller et al., 2010). Reservoirs can be simulated by treating large strategic reservoirs in an explicit manner while representing smaller ones as lumped water bodies of different size classes to efficiently account for water retention of many small reservoirs in a study region (Güntner et al., 2004). The model was developed for and successfully applied in the semiarid areas of northeastern Brazil (de Araújo and Medeiros, 2013; Krol et al., 2011; Medeiros et al., 2014, 2010) and used for other dryland regions, such as in India (Jackisch et al., 2014) and Spain (Bronstert et al., 2014; Mueller et al., 2009, 2010).

Model parametrisation and calibration

The model was parametrised using the lumpR package for the statistical environment R (Pilz et al., 2017). This included the delineation of catchment and model units, assembly, calculation, and checking of parameters, and the generation of the model's input files. Meteorological data were interpolated to the respective spatial units (sub-basins). For rainfall, this step used the Thiessen polygon method as implemented in the Information System for Water Management and Allocation (SIGA) (Barros et al., 2013). For the other meteorological variables, inverse distance weighting (IDW) from the R package geostat (Kneis et al., 2012) was used. Reservoir data were processed and prepared for the model. A total of 36 strategic reservoirs within the study area was selected for explicit treatment in the model according to their size and importance for water management.

The model was calibrated independently for each of the five regions in the study area (see Fig. 5.1). Calibrated output of upstream regions was used as boundary condition for downstream regions. Due to lack of data for Lower Jaguaribe, the calibrated parameters of Castanhão were transferred. However, sufficient data were available for further analyses. The calibration period spanned 2003 to 2010, which includes both wet (2004 and 2009) and dry (2005, 2007, and 2010) years.

Daily reservoir volume increase in the strategic outlet reservoir of a specific region was used as a target variable as reservoir level measurements were assumed to be more reliable than streamflow observations. Streamflow in the area is highly variable and rivers, especially in the downstream part of the catchment, are characterised by broad and dynamic cross sections and dense riparian vegetation inducing large uncertainties in streamflow measurements derived from rating curves. However, reservoir management has a strong impact on reservoir dynamics and only a limited number of data on artificial releases were available, while there was even no information on overspill (which does not often occur at the large strategic reservoirs) and only rough estimates of withdrawals. To minimise the impact on calibration, only positive volume variations (i.e. net reservoir volume gain), which are effectively caused by runoff draining into the reservoirs, were considered for calibration. Therefore, daily net losses of volume, which are largely determined by such management influences, were set to zero and therefore effectively excluded from the calibration. However, for the region of Salgado, streamflow measurements had to be used as this specific region does not contain a strategic reservoir at its outlet.

In total, 15 parameters were chosen for calibration. As objective function, a modified version of the Nash–Sutcliffe efficiency (NSE) called benchmark efficiency (BE) following Schaefli and Gupta (2007) was employed. It is calculated as

$$\mathrm{BE} = 1 - \frac{\sum_{t=1}^{N} \left(q_{\mathrm{obs}}(t) - q_{\mathrm{sim}}(t)\right)^2}{\sum_{t=1}^{N} \left(q_{\mathrm{obs}}(t) - q_{\mathrm{bench}}(t)\right)^2}, \tag{5.1}$$

with t being the index of time containing N time steps within the calibration period; q_{obs} represents the observations; q_{sim} represents the simulations; and q_{bench}, instead of being the average of the observations as in the traditional NSE, represents the mean of the observations for every Julian day over all years within N (i.e. the mean annual cycle). In this way, a value of BE > 0 means the model is able to reproduce the average yearly dynamics better than simply using statistics. Consequently, a value of BE = 1 signifies perfect agreement of simulations with measurements. Eventually, BE as a performance measure employs a much stricter criterion on simulated hydrological dynamics compared to using the NSE measure.

For calibration, the dynamically dimensioned search (DDS) algorithm (Tolson and Shoemaker, 2007) implemented in the R package ppso (Francke, 2017) was used. Since DDS was developed for computationally demanding hydrological models it is able to obtain satisfying results within the order of 1000 to 10 000 model calls. For this study, the number of calls was limited to 5 000 for every region, which resulted in about 10 000 h of CPU core processing time on a high-performance cluster.

Analysis of simulation performance and influencing factors

An objective of this study is to analyse the simulation performance of the process-based model in more detail and to identify possible influencing factors. Instead of using a single goodness of fit measure, as for automated calibration, different aspects of model performance should be investigated. Therefore, the Kling–Gupta efficiency (KGE) was chosen as a performance measure along with its three components correlation, bias, and deviation of standard deviations of simulations and observations (see upper part of Table 5.1). Like NSE and BE, KGE scales from minus infinity to one where one is the optimum value achieved for maximum correlation (i.e. COR = 1) and no deviation of means and standard deviations. To assess which factors influence the model performance, several candidate descriptors where selected, which are presented in the lower section of Table 5.1. These descriptors were tested for their capability to explain model performance in time and space in a regression approach by using these descriptors as predictors and the performance metrics as the response variable.

For the analysis, the calibration period 2003 to 2010 was used. Each response variable (i.e. performance metric) was calculated for each of the 36 strategic reservoirs located in the study area. Furthermore, each year was divided into a falling period, where the difference

Table 5.1: Response and predictor variables used for the analyses of the process-based model performance.

Abbrev.	Explanation
Responses	
KGE	Kling–Gupta efficiency (Gupta et al., 2009): $1 - \sqrt{(\mathrm{COR}-1)^2 + \mathrm{BIAS}^2 + \mathrm{VAR}^2}$
COR	Pearson correlation of simulations y and observations o: $\frac{\mathrm{cov}(y,o)}{\sigma_y \sigma_o}$ with cov being their covariance and σ their standard deviations
BIAS	Deviation of means μ: $\frac{\mu_y}{\mu_o} - 1 \in [-1, \infty)$
VAR	Deviation of variability: $\frac{\sigma_y}{\sigma_o} - 1 \in [-1, \infty)$
Predictors	
A_{up}	Upstream catchment area of the reservoir (km^2)
V_{cap}	Reservoir volume capacity (hm^3)
n_{resup}	Number of upstream reservoirs (–)
Δ_{vol}	Rising or falling period of reservoir volume (–)
P_{max}	Maximum regional daily precipitation sum over rising/falling period of a year (mm)
P_{reg}	Regional precipitation sum over rising/falling period of a year (mm)
P_{12}	Regional precipitation sum over the entire previous year (mm)
P_{36}	Regional precipitation sum over 36 months of the preceding years (mm)

of reservoir levels for two consecutive days was negative, and a rising period, where the difference was greater than or equal to zero. For each reservoir, year, and period, the respective performance was computed and analysed separately. This resulted in a total of 32 reservoirs times 8 years times two periods minus some missing observations, i.e. 484 values to be aggregated for each response variable. The predictors were either static and unique for each reservoir (upstream catchment area A_{up}, reservoir capacity V_{cap}, number of upstream reservoirs n_{resup}), region-specific and dynamic as aggregation over a certain amount of time (maximum daily precipitation P_{max}, regional precipitation sum P_{reg}, regional precipitation over the last 12 months P_{12}, and over the last 36 months P_{36}), or a grouping variable by itself (reservoir level is currently rising or falling Δ_{vol}) (see also Table 5.1 for more information).

To identify predictor importances and their specific influence on the performance measures, a random forest analysis was conducted using the R package party (version 1.3-1). In general, random forests consist of an ensemble of regression trees, where each tree is fitted using a bootstrap sample of the training dataset and only a subsample of all available predictors. This eliminates typical problems of traditional regression tree approaches, such as a high sensitivity to small changes in the data and the likelihood of overfitting (Breiman, 2001). For this study, a refined random forest algorithm was employed, which is better suited for predictors of different types (e.g. mixed categorical and continuous) and produces more robust measures of predictor importance in the case of correlated predictor variables (Hothorn et al., 2006; Strobl et al., 2008, 2007).

For each response variable, an individual random forest was built. Except for Δ_{vol} (categorical), each predictor and response variable was treated as numerical. To generate robust estimates of predictor importance, 1000 regression trees were built per forest (otherwise standard parameter values of the algorithm were used). The most influential predictors for a certain response were then distinguished by an importance measure, which in this study was derived by permuting the values of each predictor and measuring the difference in prediction accuracy of the random forest before and after permutation (also termed *permutation importance* in contrast to the often used *Gini importance* or *mean decrease in impurity*). In addition, the permutation of predictor values was done by accounting for potential correlation among predictor variables (hence termed *conditional permutation importance*) as suggested by Strobl et al. (2008).

In order to get an impression of the concrete effect of each predictor instead of the mere variable importance, the two leaf nodes with the highest and lowest median response values for each tree were identified. For these two nodes, the ranges of each numerical predictor

Table 5.2: Regional equations used for the calculation of monthly volume changes with the statistical approach. Table extracted and extended from Delgado et al. (2018b) (Table A1). For details and abbreviations see text.

Region	Formula	R^2
Lower Jaguaribe	$0.689 + 2.222\mathrm{SPI}_1 + 0.035\frac{\mathrm{SPI}_{36}}{\mathrm{SPI}_{12}} + 2.116\mathrm{SPI}_1\mathrm{SPI}_{12} + 1.075\mathrm{SPI}_{12}\frac{\mathrm{SPI}_{12}}{\mathrm{SPEI}_{12}} + 0.286\mathrm{SPI}_1\frac{\mathrm{SPI}_{12}}{\mathrm{SPEI}_{36}}$	0.38
Orós	$0.416 + 2.428\mathrm{SPEI}_1\mathrm{SPI}_1 + 2.233\mathrm{SPI}_1\mathrm{SPI}_{12} - 0.173\mathrm{SPI}_{36}\frac{\mathrm{SPEI}_{36}}{\mathrm{SPEI}_{12}} + 4.197\mathrm{SPEI}_1\frac{\mathrm{SPI}_{12}}{\mathrm{SPEI}_{12}} - 0.003\frac{\mathrm{SPEI}_{36}}{\mathrm{SPEI}_{12}}\frac{\mathrm{SPI}_{36}}{\mathrm{SPI}_{12}}$	0.36
Salgado	$-0.377 + 2.002\frac{\mathrm{SPI}_{12}}{\mathrm{SPEI}_{12}} + 2.454\mathrm{SPEI}_1\mathrm{SPI}_1 + 2.600\mathrm{SPEI}_1\mathrm{SPI}_{36} + 4.20\frac{\mathrm{SPI}_{12}}{\mathrm{SPEI}_{12}}\mathrm{SPEI}_1 + 0.314\mathrm{SPI}_1\frac{\mathrm{SPI}_{36}}{\mathrm{SPEI}_{12}}$	0.45
Castanhão	$2.947 + 3.468\mathrm{SPI}_1 - 1.147\frac{\mathrm{SPI}_1}{\mathrm{SPEI}_1} - 1.270\frac{\mathrm{SPI}_1}{\mathrm{SPEI}_1}\mathrm{SPEI}_{36} - 0.791\mathrm{SPI}_1\frac{\mathrm{SPI}_{36}}{\mathrm{SPEI}_{12}} + 1.412\mathrm{SPEI}_{12}\frac{\mathrm{SPI}_{36}}{\mathrm{SPEI}_{36}}$	0.21
Banabuiú	$4.812 + 4.638\mathrm{SPI}_1 - 13.853\mathrm{SPI}_{12} - 2.293\mathrm{SPI}_{12}\mathrm{SPEI}_{36} + 15.317\mathrm{SPEI}_{12}\frac{\mathrm{SPI}_{36}}{\mathrm{SPEI}_{36}} - 0.341\frac{\mathrm{SPI}_{36}}{\mathrm{SPEI}_{36}}\frac{\mathrm{SPI}_{12}}{\mathrm{SPEI}_{36}}$	0.23

(except Δ_{vol}) were classified into four groups ranging from small to large to facilitate visual investigation.

5.3.5 The statistical model: a regression approach

A goal of this study is to answer the question of whether the application of a complex process-based simulation model is worthwhile in comparison to a much more convenient statistical approach to generate seasonal forecasts of reservoir storage and drought indices. To achieve this, the regression model of Delgado et al. (2018b), which was developed for the same study area, was employed. They fit a multivariate linear regression (MLR) model individually for each of the subregions also defined in this study. As a response variable, regional volume changes were used (approach M2 in Delgado et al., 2018b). As possible predictors, meteorological drought indices (standardised precipitation index, SPI; and standardised-precipitation–evapotranspiration index, SPEI) aggregated over time periods of 1, 12, and 36 months, respectively, were considered in their study. To account for correlation among predictors, ratios of predictors exhibiting significant correlation to each other were used. Genetic optimisation with respect to the Akaike information criterion (AIC) was employed to determine the specific predictors for each subregion. To enforce model parsimony, not more than five predictors should be used in the regression equation. For the model fit, all available observations within the analysis period were used (monthly values from 1986 to 2014, less a few missing values). The resulting equations are presented in Table 5.2.

To generate hindcasts, the predictors of the equations (the SPI and SPEI values over different time horizons) were calculated on a monthly scale to obtain monthly forecasts of regional reservoir volume changes for each rainy season of the hindcast period. Regional storage volume values could then be obtained by successively adding predicted volume changes to the measured value of December of the previous year, which served as a base value for each rainy season. Even though the shown model fits for monthly volume changes were rather poor (low R^2 values in Table 5.2), the derived absolute reservoir level values were in good agreement with measurements (Delgado et al., 2018b).

To compare the mere simulation performances, both the process-based and the statistical model were first driven by observed meteorology to exclude the effect of the downscaled GCM runs. In a second step, the two approaches were evaluated for real hindcasts.

5.3.6 Drought hindcasting

Hydrological drought quantification

As water stored in surface reservoirs is of primary importance to water supply, hydrological drought indices based on surface reservoir filling level appear to be the most adequate choices to identify and characterise hydrological droughts in the study area. Thus, in line with Delgado et al. (2018b), for the quantification of hydrological droughts the regionally and monthly aggregated reservoir storage was defined as a drought indicator:

$$I_t = \frac{\sum_{i=1}^{R_j} V_t^i}{\sum_{i=1}^{R_j} V_{\text{cap}}^i}, \tag{5.2}$$

with t being the time index, V_t^i the volume stored in reservoir i of a certain region R_j (i.e. one of the five subregions of interest illustrated in Fig. 5.1), and V_{cap}^i the storage capacity of that reservoir. This metric was calculated for each of the five regions of interest (R) and each month of the hindcast period (wet seasons, January to June, of 1981 to 2014). For each month the last daily value was taken.

A drought was then defined as

$$D_t = \begin{cases} 1 & \text{if} \quad I_t < q_{0.3} \\ 0 & \text{if} \quad I_t \geq q_{0.3} \end{cases}, \tag{5.3}$$

where $D = 1$ denotes drought, $D = 0$ indicates no drought, and $q_{0.3}$ is the 0.3 quantile of I over the hindcast period. The definition of $q_{0.3}$ is based on the choice of local decision makers who defined this value as a warning threshold for reservoir scarcity. In Sect. 5.5.1 the impact of this decision will be discussed. The threshold was applied to each region individually and, thus, resulted in regionally different drought thresholds. As such, the results of this study will be comparable to the work of Delgado et al. (2018b).

Verification of drought hindcasts

Hindcasts of reservoir volumes (V_t^i) and, consequently, the hydrological drought index (I_t) were verified employing the root mean square error (RMSE), the relative operating characteristic skill score (ROCSS), and the Brier skill score (BSS). Definitions and discussions of the various forecast verification metrics can be found in textbooks such as Wilks (2005). In the following, short explanations for each selected measure shall be given.

The RMSE is a deterministic measure and was derived by calculating the root of squared differences of hindcasts and observations averaged over all values of the hindcast period:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(I_t^{\text{f}} - I_t^{\text{o}}\right)^2}, \tag{5.4}$$

where N is the number of forecasted time steps, and the superscripts "f" and "o" denote forecast and observation, respectively. It was calculated multiple times by using as I_t^{f} each GCM member individually and, in addition, the median of members as the deterministic value. The metric quantifies the average magnitude of hindcast errors in units of the target variable, i.e. in this case regional reservoir storage in percent points, and is therefore useful for the interpretation of the suitability of the model for water managers who rely on accurate forecasts of volumes to coordinate reservoir operation. As such, the RMSE refers to the attribute of accuracy. The lower the RMSE, the lower the forecast error and the higher the accuracy.

The ROCSS quantifies the ability of a model to correctly discriminate between events and non-events. In this context, an event is defined as a hydrological drought which, in turn, is distinguished by the drought index falling below the 0.3 quantile ($q_{0.3}$) as explained above. The ROCSS is based on the ROC curve which plots the probability of event detection against the false alarm rate for different thresholds of forecast probability defining an event. Taking the area under the curve (AUC) of this graph, the skill score can be calculated as

$$\text{ROCSS} = 2 \cdot \text{AUC} - 1. \tag{5.5}$$

The value ranges between -1 and 1 with values lower than or equal to zero indicating the false alarm rate being greater than or equal to the probability of event detection and, thus, the model having no skill. A value of one represents the highest score, i.e. the model is able to predict every event and non-event correctly. As such, the ROCSS is a measure for event resolution of probabilistic forecasts.

The Brier score (BS) measures the mean squared error of probabilistic forecasts and indirectly contains information about reliability, resolution, and the variability of observations (the latter being commonly referred to as uncertainty). As such it can be calculated as

$$\mathrm{BS} = \frac{1}{N}\sum_{t=1}^{N}\left(D_t^{\mathrm{f}} - D_t^{\mathrm{o}}\right)^2 . \tag{5.6}$$

The corresponding skill score (BSS) compares the BS of a forecast model with that of a simple reference forecast, in our case climatology:

$$\mathrm{BSS} = 1 - \frac{\mathrm{BS}}{\mathrm{BS}_{\mathrm{reference}}}, \tag{5.7}$$

with $\mathrm{BS}_{\mathrm{reference}} = 0.3$, corresponding to $q_{0.3}$, the initially defined long-term average probability of drought occurrence (as described above). It follows that $\mathrm{BSS} \in (-\infty, 1]$ and a forecast model having skill relative to the reference model if $\mathrm{BSS} > 0$.

5.4 Results

5.4.1 Comparison of model performance in simulation mode

Figure 5.3 compares the performances of the process-based and statistical model in simulating relative regional reservoir storage driven by observed meteorology. The regional RMSE varies between 5 % and 18 %, whereas for the whole catchment both modelling approaches achieve a result of about 13 %. Overall, the performance differences between the two models are small for all regions. Only for Salgado the statistical model shows a lower RMSE compared to the process-based model, and the difference among the two approaches is largest (6 %). For all other regions, the process-based model exhibits a slightly higher accuracy, and the interregional ranking is equal for both approaches. Generally speaking, both models show a comparable performance, suggesting they are equally suitable for their application in hindcast mode.

5.4.2 Comparison of model performance in hindcast mode

The uppermost panel of Fig. 5.4 shows that, in hindcast mode, the accuracy in terms of RMSE considerably decreases when compared to simulation mode for both types of models. However, in contrast to the situation in simulation mode, the statistical approach outperforms the process-based model for all regions. While for the statistical model deterioration in terms of RMSE is generally less than 10 %, the process-based model achieves significantly lower accuracy with increasing RMSE by up to almost 30 %. This degradation of model performance in hindcast mode for the process-based model is especially pronounced for the Banabuiú region.

The lower two panels of Fig. 5.4, however, demonstrate that both approaches are able to generate drought hindcasts with skill. The resolution of event hindcasting of the two models (i.e. the ROCSS) is very similar when it is combined over the whole catchment. Regional differences are more pronounced but still negligible. For some regions the process-based model performs better, but for other regions the statistical model performs slightly better. The BSS, while also indicating skill, shows lower performance values which can be attributed to lack of accuracy (as already indicated by RMSE) and reliability.

An attribute plot, as the one presented in Fig. 5.5, can reveal more details on that issue. Therein, the predicted probability of drought occurrence (obtained from the outcomes of individual ensemble members) is plotted against the relative frequency of observed drought occurrences (solid lines) together with the relative prediction frequency of a certain forecast probability interval (dotted lines). It demonstrates several verification attributes including resolution (the flatter the solid lines, the less resolution), reliability (agreement with the gray diagonal line), sharpness (dotted coloured lines), and skill (values within the gray region contribute

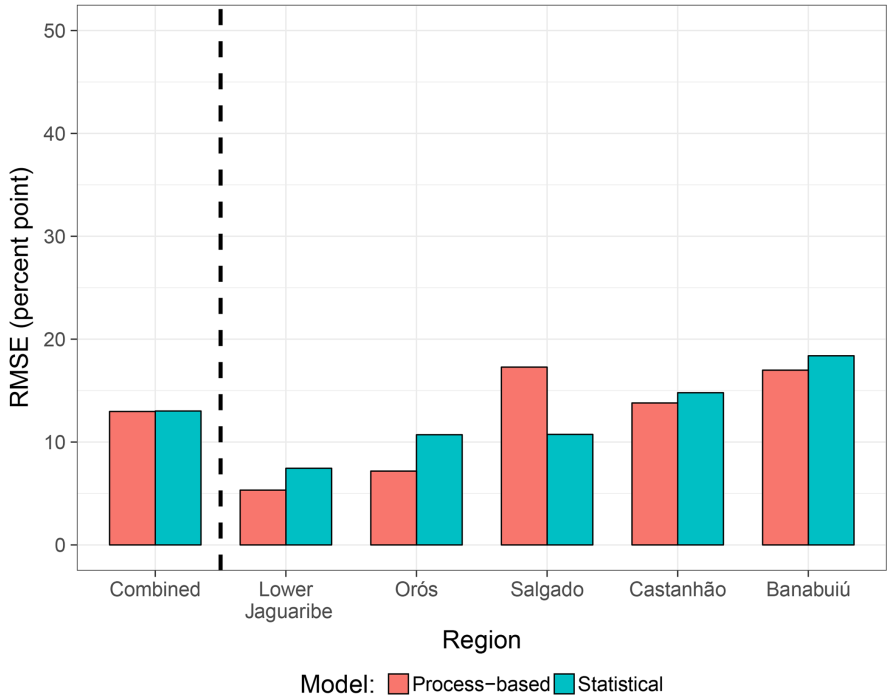

Figure 5.3: Comparison of accuracy in predicting relative reservoir storage for the process-based and statistical model in simulation mode (i.e. run with observed forcing). The underlying analysis period comprises monthly values of the rainy season (January to June) over the hindcast period (observations available since 1986 until 2014 with some data gaps in between, resulting in a maximum of 174 values for each subregion).

positively to BSS; for unclear terms see Appendix Sect. 5.A.1 or consult textbooks such as Wilks, 2005). Apparently, predictions from both models contain skill except for low forecast probabilities where both models contribute negatively to BSS. Furthermore it can be seen that both approaches exhibit problems in terms of reliability. Specifically, forecast probabilities are too low compared to observed occurrences, which is generally denoted as *underforecasting*. This observation appears to be a bit more pronounced for the process-based model than for the statistical model. That also holds true for sharpness, as the statistical approach shows slightly more confidence for higher forecast probabilities; i.e. the relative frequency of maximum forecast probability is higher.

5.4.3 Model performance attribution

Hindcasts

The monthly aggregated accuracy of the hindcasts, i.e. performance with increasing lead time, is shown in Fig. 5.6. Overall, the hindcast error (i.e. RMSE) increases with lead time (i.e. progression of the wet season), even when using observed forcing. The statistical approach generally produces better hindcasts. Its RMSEs differ only little from runs with observed forcing. Also the increase in RMSE with lead time is very similar. For the process-based model, hindcasts deviate clearly from observation-based results (as was already shown in Fig. 5.4). The error increases much more strongly over the hindcast horizon. However, its RMSE values reach a plateau at about 40 % in March. Generally, it can be seen that aggregating the ensemble members by using the median of reservoir storage hindcasts (solid lines) is usually a better choice than most of the single ensemble members (distributions shown as box plots). The spread of ensemble member results differs for the two approaches. These ranges are clearly larger for the process-based model in January and February but comparable for the other months.

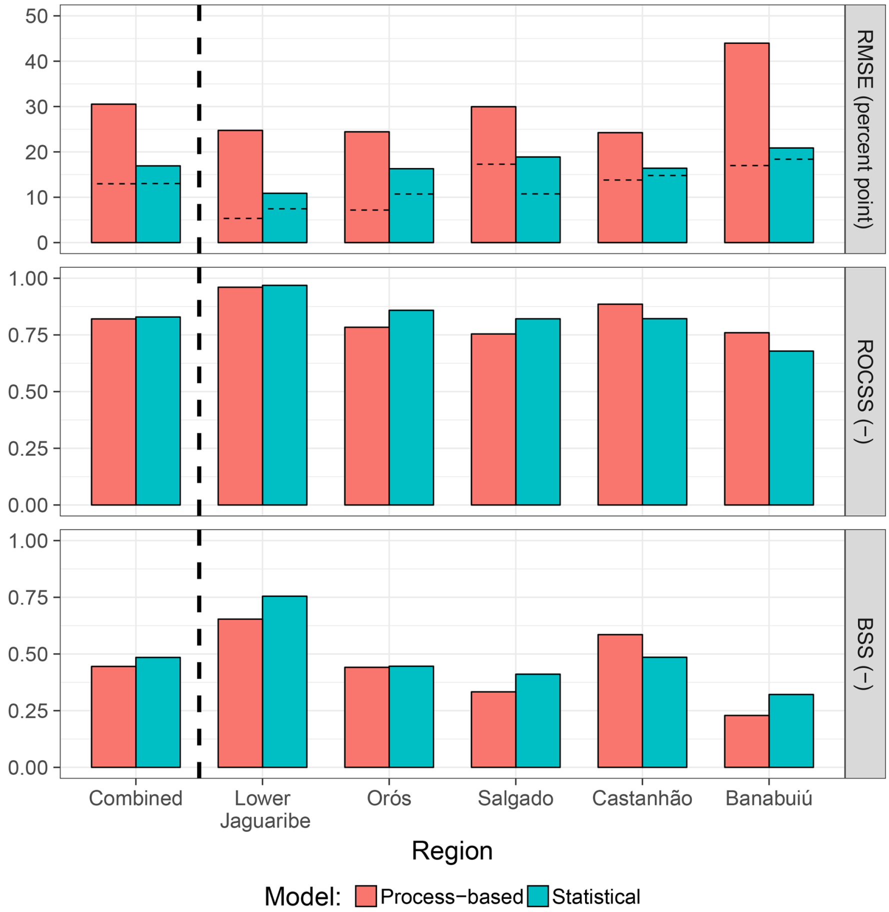

Figure 5.4: Model performance in hindcast mode for the two model approaches. In the top panel, horizontal dashed lines in the bars mark the results obtained with observed forcing (simulation mode, as in Fig. 5.3). Note that for RMSE low values indicate a better performance while for ROCSS and BSS higher values are favoured.

In Fig. 5.7 prediction accuracy is assessed for different wetness conditions (i.e. dry, normal, wet) over different accumulation time periods for rainfall. Again, when driven by meteorological hindcasts, the statistical approach performs best with relatively small differences compared to results obtained using observed forcing. Under wet conditions, irrespective of the rainfall accumulation period, the error is highest for most settings. The only exception from this pattern shows the process-based model driven by hindcasts. Here, the error under dry preconditions increases with increasing rainfall accumulation length while the performance under wet preconditions improves with longer accumulation length.

Process-based simulation performance

In the preceding subsections it was shown that the process-based model does not outperform the statistical approach. Moreover, in hindcast mode, the process-based model often achieved worse performance measures, especially in terms of accuracy. This subsection therefore aims at the identification of deficit causes by analysing the results of process-based model calibration and potential influencing factors of simulation performance in more detail.

Regional calibration performance of the process-based model is summarised in Table 5.3. A good overall agreement of simulated and observed reservoir dynamics in terms of BE values could be achieved during calibration. However, percent bias (PBIAS) as a performance metric not used in the calibration shows, on the one hand, acceptable values of no more than 12 % but, on the other hand, a consistent slight overestimation of reservoir level dynamics. It can be further observed that a good BE value does not correlate with a low PBIAS.

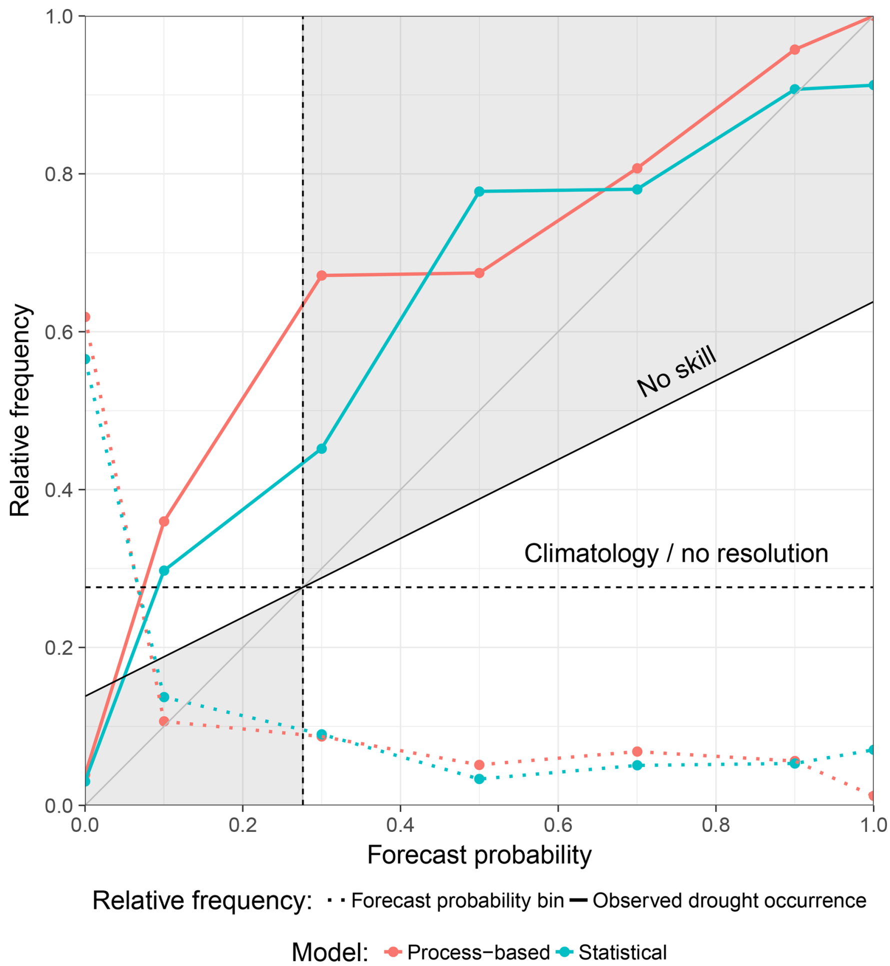

Figure 5.5: Attribute plot of drought hindcasts aggregated over the whole study area. Values within the gray region contribute positively to the Brier skill score (BSS). For details on the interpretation of the plot see text.

The random forest analysis brought more insight into process-based model performance and its influencing factors. Figure 5.8 illustrates the importance of each potential predictor for different performance metrics. Apparently, overall model performance (here measured via KGE) primarily depends on the wetness preconditions (P_{36}). While reservoir size (V_{cap}) plays only a minor role for the overall performance metric KGE, it clearly affects correlation and bias. (Mis)match of standard deviation (VAR), however, is mainly determined by both wetness conditions and reservoir size. Overall, long-reaching antecedent wetness condition (P_{36}) is more important than the conditions of the preceding 12 months (P_{12}), and reservoir capacity (V_{cap}) is dominant over upstream catchment area (A_{up}), although the latter is not negligible. The current rainfall conditions, in terms of intensity (P_{max}) and sum over a rising/falling period (P_{reg}), whether it is a reservoir level increase or decrease period (Δ_{vol}), and the number of upstream reservoirs (n_{resup}) show little or no explanatory value for any of the performance measures.

To analyse the specific influence of predictors on the response variables, Fig. 5.9 relates the values of the most influential predictors to the corresponding performance measures. This is done by plotting the occurrences of predictor categories in the highest and smallest valued leaf nodes of all regression trees within the random forest. It shows that under dry preconditions (P_{36} = min) there is a tendency for underestimation of standard deviations (VAR = min), i.e. a less variable reservoir storage series than observed, but a better a better overall performance (KGE = max). On the other hand, under wet conditions, especially for larger upstream catchments (A_{up} = high), results tend to show an overestimation of variability (VAR = max), whereas under dry conditions in small catchments variability is more often

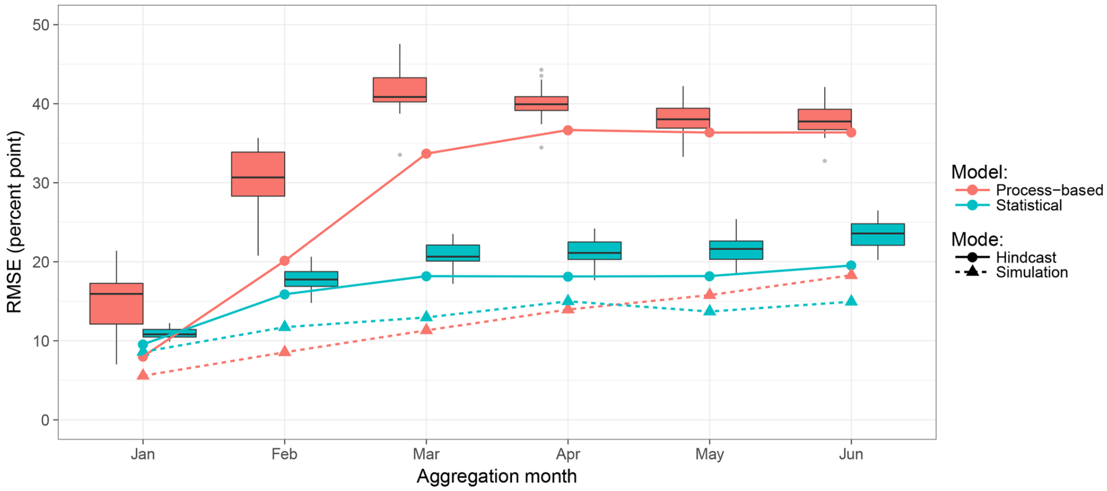

Figure 5.6: RMSE of regional reservoir storage hindcasts with increasing forecast horizon and lead time. Monthly values are obtained by aggregation over the full analysis period (1986 to 2014) for a specific month. Each box reflects the distribution of the 20 ensemble members. Coloured solid lines refer to the ensemble median taken as the deterministic forecast and analysed individually.

Table 5.3: Results of regional calibration of the process-based model. BE refers to benchmark efficiency (Eq. 5.1) and was used for calibration; PBIAS is percent bias, i.e. the average tendency for over- or underestimation of simulations in comparison to observations. For Lower Jaguaribe, no observations at the catchment outlet were available.

Region	BE	PBIAS (%)
Banabuiú	0.84	11.78
Orós	0.76	0.92
Salgado	0.79	7.37
Castanhão	0.76	6.93
Lower Jaguaribe	[no obs.]	[no obs.]

underestimated. For small reservoirs correlation is mostly low. It should be noted, however, that relationships cannot always be clearly distinguished. For instance, a low precipitation sum over the preceding year (P_{12}) may result in both a high and a low KGE value, whereas very low precipitation over the preceding 3 years (P_{36}) only led to a high KGE. Furthermore, there is no relationship between reservoir capacity and KGE.

5.5 Discussion

5.5.1 Robustness of performance metrics

There are two important algorithmic parameters affecting drought predictions of this study. One is the threshold of drought definition, i.e. the quantile of drought index observations specifying a drought, which was set to 0.3 as commonly used in the study area among water managers. This choice affects the performance values of BSS and ROCSS. The other is the number of probability bins into which hindcasts are grouped for further analysis, affecting ROCSS but not BSS as BS was herein calculated without probability binning (see Eq. 5.6). Figure 5.10 illustrates the sensitivity of verification attributes to the two parameters. It shows that a higher drought threshold results in a more evenly running curve while a smaller threshold of 0.2 tends to be better oriented towards the reliability line and appears more variable (Fig. 5.10a). This might result from the necessarily lower number of values of smaller thresholds. However, altogether general conclusions remain untouched, namely underforecasting and the statistical being

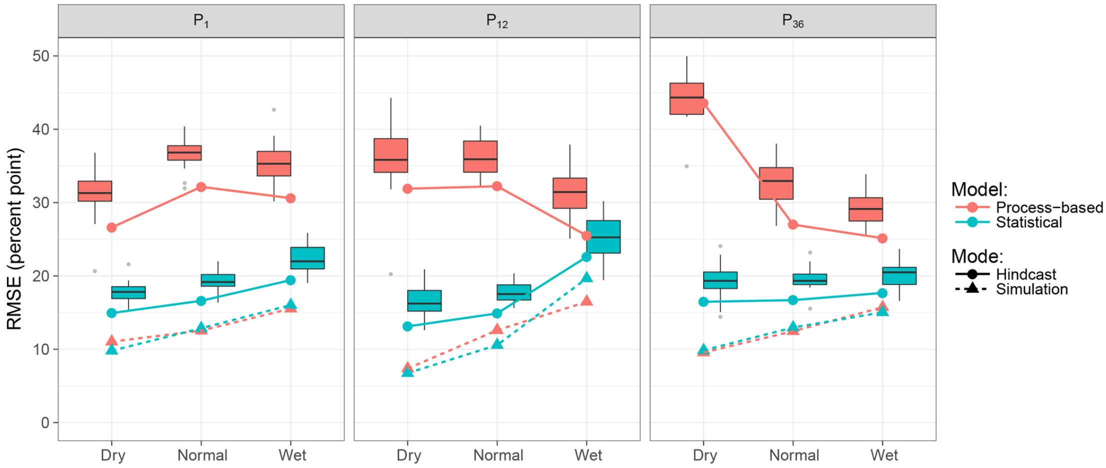

Figure 5.7: RMSE of regional reservoir level hindcasts for different antecedent wetness conditions. Wetness is expressed by three different accumulation horizons of rainfall (1, 12, and 36 months; left, centre, right). Each box reflects the distribution of the 20 ensemble members. Coloured solid lines refer to the ensemble median taken as the deterministic forecast and analysed individually.

superior to the process-based model. Regarding the number of probability bins (Fig. 5.10b), a larger value leads to a more variable curve. This effect can be attributed to the decreasing number of values per bin with increasing number of bins. For this study, it was decided to use a value of seven as it appears to be the best compromise between sufficient data availability per bin and an adequate number of bins for further calculations (namely ROCSS). Even affecting the values of ROCSS and partly BSS (not shown), it can be concluded that the somewhat arbitrary decision on a certain drought threshold and the number of bins, as long as reasonable values are chosen, does not affect the general results of the analysis.

The RMSE as an accuracy measure is free of such decision parameters but is admittedly influenced in a different way. With the target variable (relative regional reservoir filling) ranging from 0 % to 100 %, the actual maximum value of the metric tends to be smaller during wet periods: the observed value (which is usually greater than zero) effectively causes the RMSE to be limited to about 40 % to 50 %. This effect is reflected as the apparent performance plateau for the process-based model in Fig. 5.6 and is also likely to affect the results presented in Fig. 5.7. The effect that, when driven by hindcasts, the process-based model exhibits larger errors under dry than under wet conditions can be at least partially attributed to this issue. In contrast, when models are driven by observations, it seems reasonable that model simulation performance is generally better under dry conditions (Fig. 5.7). However, as no threshold effects can be observed and the RMSE values are always considerably lower for the statistical model, this effect should not influence general conclusions of the model comparison.

5.5.2 Model comparison

In terms of simulation accuracy when driven by observations and for drought event prediction in the hindcast mode, both models perform equally well. Hindcast accuracy, however, is substantially lower for the process-based approach. This result is well in line with findings of other studies that simple statistical model approaches often perform equally well or even better than complex process-based prediction systems, especially in tropical regions due to exploitable correlations among meteorological and hydrological variables (Block and Rajagopalan, 2009; Hastenrath, 2012; Sittichok et al., 2018). It has to be noted, however, that the process-based approach with the WASA-SED model achieved acceptable results on monthly (hindcasts) and even daily (calibration metrics) timescales whereas former studies in NEB reported passable results only aggregated over seasonal scales (Alves et al., 2012; Block et al., 2009; Galvão et al., 2005).

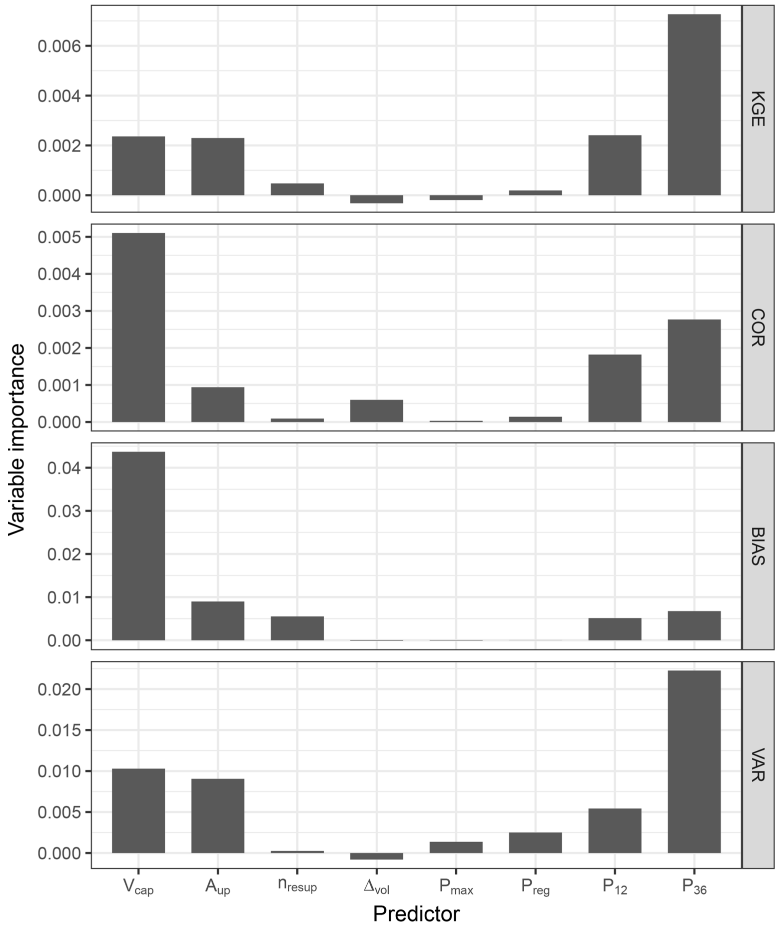

Figure 5.8: Predictor importances for each response variable determined by the random forest approach. In this case conditional permutation importance was used (see Sect. 5.3.4).

The reason for the discrepancy of model ranking between simulation and hindcast mode can be attributed to the different model structures. To illustrate this, Fig. 5.11 shows the average monthly changes of regional reservoir storage for the different models and modes in comparison to observations. For the simulation mode (dashed lines) it can be seen that the process-based model, though exhibiting a constant overestimation, all in all is well in line with observations. The statistical model, however, shows a more or less constant storage change over the whole simulation horizon, resulting in over- and underestimations and, eventually, a good overall simulation performance (see Fig. 5.3). In hindcast mode (solid lines), for the process-based model the overestimation of storage change is much more pronounced and the peak shifted from April to March. Although the statistical model now more realistically exhibits seasonal dynamics, the general pattern still appears too smooth, which effectively results in less deviation from observations than the output of the process-based model (Fig. 5.4). This indicates a strong influence of precipitation forcing on the process-based model while the statistical approach shows less pronounced reactions to changes in the rainfall input. Consequently, deficiencies in this forcing affect the process-based model much more. This, in addition to the plateau effect discussed in Sect. 5.5.1, explains the more diverse RMSE values among the hindcast realisations for the process-based model at the beginning of the rainy season when reservoirs are filling up (Fig. 5.6) and the higher RMSE under antecedent dry conditions (Fig. 5.7 middle and right panels). In contrast, for the statistical model, the general patterns of RMSE over different lead times or under different antecedent moisture conditions do not change in hindcast mode when compared to simulation mode. The issue of uncertainties arising from defective precipitation forcing will be discussed in more detail later on.

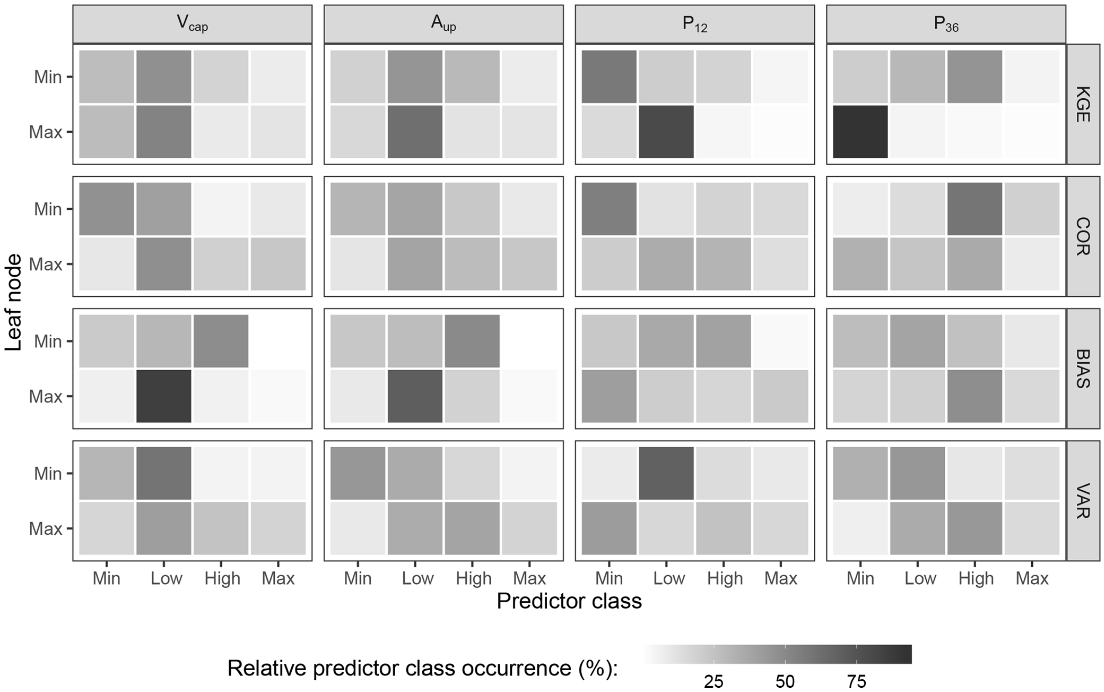

Figure 5.9: Relative distribution of predictor class occurrences within the largest and smallest valued leaf nodes aggregated over all regression trees of the random forest for each response variable. Shown are only the most important predictors.

Despite the lower prediction performance, the process-based approach still provides benefits over the statistical model. This includes the potential access and investigation of multiple spatially distributed hydrological variables with daily resolution, such as evapotranspiration, runoff generation, or streamflow, which were generated during the model runs. This clearly excels over the statistical model, which only yielded predictions of a single target variable. Another advantage is that model output is not only provided in a regionally and monthly aggregated manner, as for the statistical approach, but for all individual strategic reservoirs in the area as daily time series. Figure 5.12 illustrates that accuracies of individual reservoirs exhibit a slightly larger variation, but the RMSEs of individual reservoirs are at a similar level as when regionally aggregated. This suggests that most of the single reservoirs can be modelled with a comparable performance to the regionally aggregated values.

A further advantage of a model such as WASA-SED is that underlying processes are directly represented. As such it can be of higher value to water managers interested not only in streamflow or reservoir level forecasts but also in the investigation of process behaviour or assessments under changing boundary conditions. Therein the model could be used in scenario analyses, such as climate change impact assessment, or sensitivity analyses of, for instance, uncertain meteorological input to detect critical streamflow or reservoir stages. Furthermore, the model is transferable and can be easily applied in different regions and over different spatial and temporal scales, only limited by computational resources and available input data.

5.5.3 Deficiencies of the process-based simulation approach

To improve the performance of the process-based model it is first necessary to identify sources for simulation inaccuracies. It was shown that the process-based model achieved regionally different performances. A comparison of Fig. 5.3 and Table 5.3 reveals that regional bias during the calibration period is in compliance with the ranking of regional simulation errors. Moreover, although exhibiting the highest BE value, the region of Banabuiú is characterised by the largest bias during calibration and highest simulation and hindcast errors. As the latter is observed for both the process-based and statistical approaches, the reason is suspected to originate from

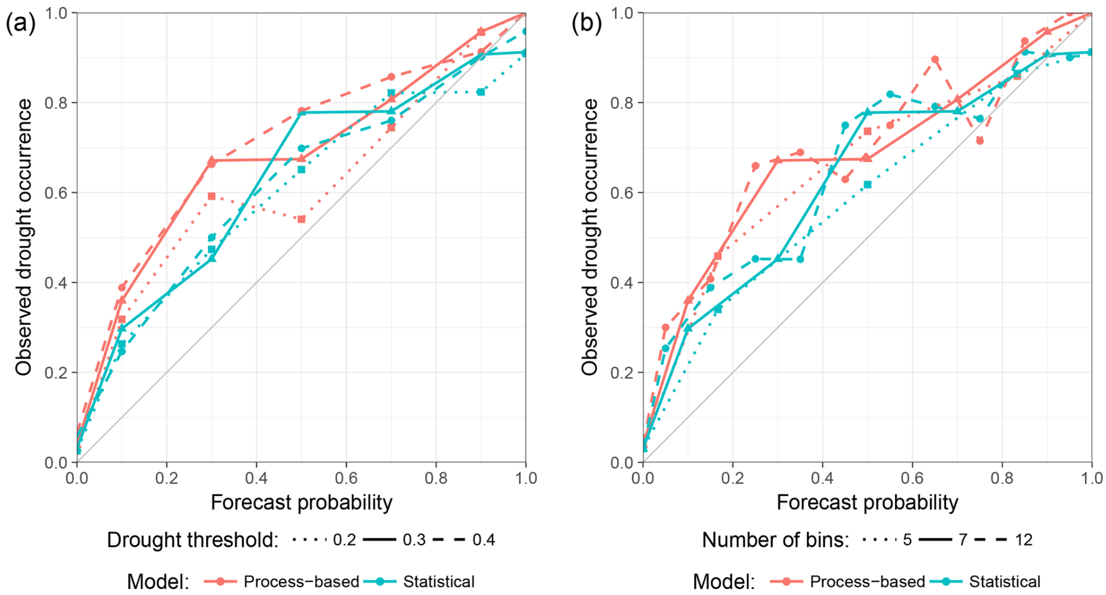

Figure 5.10: Reliability plots for different settings of **(a)** drought thresholds and **(b)** number of probability bins. Solid lines refer to the values used in this study. The gray 1 : 1 lines in each plot illustrate perfect reliability for comparison.

uncertainties in observations, i.e. precipitation measurements within the region, or defective reservoir level acquisition. The reason for the Salgado region being out of the general pattern for the process-based model certainly originates from the different calibration procedure applied here, namely the use of streamflow measurements in contrast to reservoir dynamics as for the other regions. In addition, the region is distinct from other parts of the catchment in terms of environmental settings such as larger groundwater influence and sedimentary plateaus in the headwater area. Conversely, the transfer of the calibrated parameters from Castanhão to the Lower Jaguaribe region seems justifiable as the simulation error was small. Overall, reservoir size largely influences both simulated storage time series and bias. Model performance, however, appears to not be superior for large reservoirs. Moreover, wetness condition in terms of antecedent rainfall sums over the last 36 months is of major importance, i.e. dry conditions lead to the best model performance in terms of KGE. The latter is not surprising as rainfall in the study area is extremely heterogeneous both in space and time, usually characterised by convective heavy precipitation events with short durations. Thus, prolonged periods without rain constitute a spatially more homogenous input. Conversely, the aggregation of rainfall to daily sums and interpolation over subbasin units, on average covering an area of about 700 km^2, must necessarily induce uncertainties. The assimilation of observed reservoir filling states at the beginning of each hindcast season is therefore a reasonable approach to improve predictions and compensate for preceding rainfall input uncertainties during the initialisation run.

5.5.4 Potential improvements

There are several options to make use of the findings of this study and improve the forecast system in upcoming applications. In the presented study, observed reservoir level data were assimilated into the process-based model to correct the initial conditions for the hindcast runs by simply replacing model states by measurements. For assimilation, more formal approaches already exist, such as the rich families of Kalman and particle filtering approaches (e.g. Komma et al., 2008; Liu and Gupta, 2007; Sun et al., 2016; Vrugt et al., 2013; Yan et al., 2017). These, however, require a profound quantification of both simulation and observation uncertainties and, thus, much additional information and, moreover, significantly higher expenses in terms of data preparation, processing, and model application. Nevertheless, they hold the potential to better account for uncertainties in the observations, which were disregarded in this study, despite being considerable.

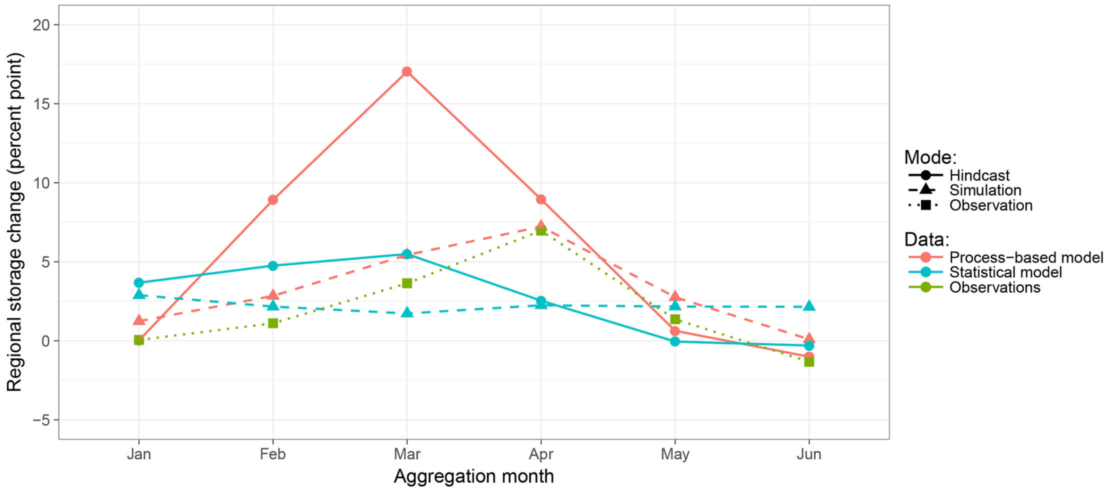

Figure 5.11: Monthly changes of regional reservoir storage averaged over all regions, years, and hindcast members for the two models and application modes in comparison to observations.

Preprocessing schemes in the context of hydrological forecasting usually focus on the improvement of rainfall predictions used as main drivers for hydrological models (e.g. Kelly and Krzysztofowicz, 2000; Reggiani and Weerts, 2008; Verkade et al., 2013). This is partly already included in the downscaling scheme applied to GCM products but may as well be further extended. The importance of rainfall forcing on model results, especially for the process-based approach, was already addressed above. A further comparison of the statistical properties (distribution of daily sums, dry/wet spell lengths) of rainfall hindcasts used in this study with observations revealed large discrepancies. Some preliminary tests suggested these to be responsible for the decreased accuracy of the process-based model hindcasts (not shown). In comparison to observations, the hindcasts contain (i) a general shift of rainfall seasonality towards the first months of the rainy season; (ii) a much lower frequency of both wet and dry periods for spell lengths up to 4 days; (iii) a lower frequency of low daily rainfall values while the number of large precipitation events is overestimated and daily extreme values are much higher; and (iv) a much higher probability that a dry day follows a dry day, and the probability that a wet day follows a wet day is often underestimated. These findings indicate a high potential for improvement in future applications in the study area. As a first starting point, monthly bias of hindcasts per region was corrected and both models were rerun. Figure 5.13 shows that this relatively simple procedure already results in a considerable decrease in RMSE for the process-based model, even though it is still higher than for the statistical approach. The improvement of drought forecast performance in terms of BSS and ROCSS is thereby less pronounced than the increase in accuracy. For the statistical model, performance metrics hardly change, which can be attributed to the smoothing effects of its model structure on regional reservoir storage identified in a previous subsection (Fig. 5.11).

In addition to preprocessing, post-processing approaches directly tackle the correction of streamflow forecasts by statistical means including bias correction or the estimation of an error model applied to predictions (e.g. Bourdin et al., 2014; Krzysztofowicz and Kelly, 2000; Roulin and Vannitsem, 2014; Todini, 2008). Especially when focussing on extreme events, such as floods or droughts, the adequate characterisation of model residuals exhibits a large potential when incorporated into the correction of simulations and predictions (Farmer and Vogel, 2016). While being still an active field of research, such means are routinely applied in operational streamflow forecasting and, in addition to rainfall correction, could further improve model performance.

The parametrisation of the process-based model could be further improved by the use of more and different data sources. This includes, for instance, the use of satellite data to infer spatially distributed reservoir information with greater detail and more accuracy than currently available. The study area has already been of interest in ongoing research (Delgado et al.,

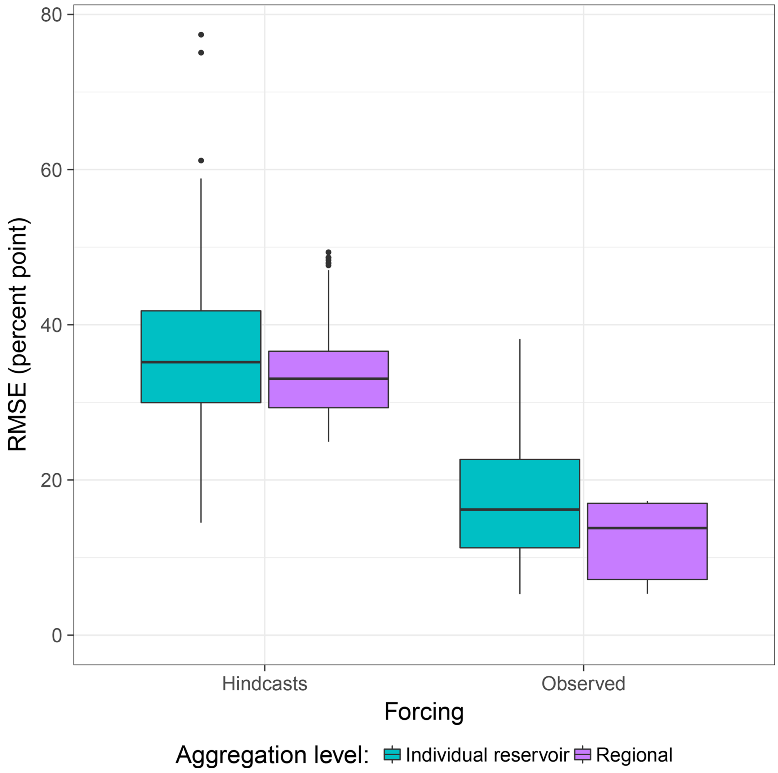

Figure 5.12: Comparison of accuracies of the process-based model for different spatial aggregation levels on a monthly timescale.

2018a) and past studies (Heine et al., 2014) addressing that issue. In addition, management plans as well as data on water abstraction and reallocation from the larger reservoir should be included in the model but were not available for the present study. Another opportunity is to increase rainfall input resolution in the model to better account for sub-daily and spatially heterogeneous precipitation. This could be done by improving the current spatial scaling of rainfall in the model to account for heterogeneous patterns and to make use of radar rainfall data recently made available in the area.

The combination of multiple models may provide further benefits in cases where different models show strengths in different aspects of performance (e.g. Block and Rajagopalan, 2009; Schepen and Wang, 2015). However, within this work the two employed model approaches, with respect to simulation performance, achieved almost equal results and did not diverge in aspects such as lead time and antecedent moisture conditions. Thus, the combination of the two models analysed in this study is not expected to provide benefits.

5.5.5 Generally valid features and broader implications of results

The possibilities to generalise and transfer insights gained for one particular environment and a specific hydro-climatological setting will always be limited to questions of transferability of model approaches and the applicability of particular methods. In other word, one cannot expect that the actual numbers of model performance, be it statistical or process-based, have similar values under differing conditions. Adapting the two forecast setups presented here to other regions will presumably not lead to the same ranking of the statistical vs. the process-based model approach. In order to achieve an optimal forecast score, it is obvious that the forecast design, e.g. selection of target variables, models, and input data, needs to be tailored to the region of interest and to the purpose of the forecast including the designated forecast lead time. The design of the statistical forecast approach presented here is particularly tailored to the rather specific hydro-climatological and landscape conditions of this semiarid region. In our case, the statistical model relies on antecedent wetness conditions of the basin, quantified by

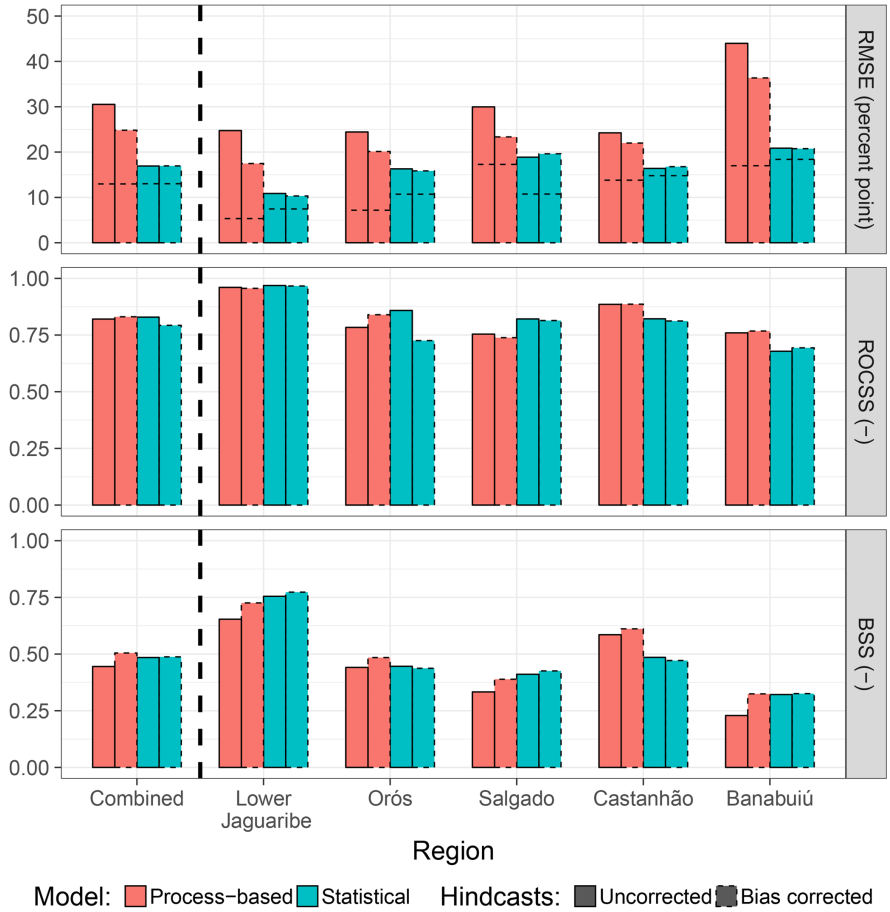

Figure 5.13: As Fig. 5.4 but driving hindcasts with additional bias correction of precipitation on the monthly scale (dotted boxes).

precipitation indices (SPI and SPEI) aggregated over several months (see regression equations in Table 5.2). This satisfactorily describes the attenuated hydrological behaviour of this particular semiarid region, which is further influenced by many large and small reservoirs, in order to obtain forecasts over several months. This adaptation to the regional peculiarities does not per se allow a spatial transfer of the derived statistical model. However, the application of the underlying statistical model principles to other regions is generally possible, given that appropriate regional information and forcing data are available.

The transfer of a process-based hydrological model to another region is, generally speaking, more straight forward, because one can rely on the fact that the underlying physical assumptions and process descriptions of the model are valid for different environments and hydro-climatological conditions. In other words, the hydrological model implicitly contains the representation of general hydrological processes and, therefore, represents an adequate working hypothesis of the hydrological system at other target locations. Still, process-based models also require a sufficient data availability and quality at the region of interest while forecast performance for both model types primarily depends on the quality of rainfall forecasts.

The question of which modelling approach will finally yield a better performance score cannot universally be answered. On the one hand, this is dependent on the predictive power (i.e. the strength of correlation between predictors and target variable) of the statistical model and, on the other hand, on the validity of the physical assumptions and governing equations of the process-based model. Both model strategies benefit from good data conditions, while the required type of data is rather different. The advantage of a statistical model approach as presented here is that less distributed data of the catchment are required and the model is easier to establish and needs much less computational power. On the other hand, the process-based model allows us to simulate a rather large variety of hydrological variables in a spatially distributed manner. This enables independent verification of process-based models and leads further to a high explanatory power of this model type, which are important advantages, even for cases where the overall performance of a process-based model might be lower compared to a statistical model approach.

5.6 Conclusions

The aim of this work was to explore options for a seasonal forecasting system of regional reservoir volume and drought occurrence with lead times up to 6 months for the semiarid northeast of Brazil. In this context, the performance of a complex process-based hydrological model was evaluated against a much simpler statistical model developed by Delgado et al. (2018b) given the same meteorological forcing. The study pursued three objectives.

First, the two modelling approaches were to be investigated in terms of mere simulation performance, i.e. when driven by meteorological observations. It turned out that both models performed almost equally well. However, regional differences exist where the process-based model achieved slightly better results in four out of five subregions. Furthermore, regional performance ranking of both models was equal in four regions. This suggests that data uncertainty of meteorological input or reservoir level observations exceeds model structural uncertainties and dictates simulation performance in the study area.

Second, the process-based model was to be verified as a prediction tool in a hindcast experiment and evaluated against the statistical approach. In comparison to simulation runs with observed forcing, hindcast performance of the process-based model dropped significantly while the performance of the statistical approach decreased only to a small degree. This can be explained by the structure of the statistical approach which is less sensitive to rainfall forecasts. Although this exhibits less realistic intra-seasonal dynamics than for the process-based model, performance metrics were eventually superior as uncertainties from precipitation hindcasts could not propagate as much to the model output. However, apart from reservoir level predictions, forecasting of mere drought occurrence works almost equally well for both approaches. The two models exhibit satisfying event resolution while slight deficiencies in terms of underforecasting were detected regarding the reliability of the hindcasts.

The third and last objective was to identify the major sources for simulation and hindcast deficiencies and provide guidelines for further improvement. In general, both models achieve better results under dry than under wet (pre)conditions. An attempt to identify potential predictors of model performance for the process-based model revealed that reservoir size and antecedent rainfall conditions explain most of the variance of the performance metrics while variables such as current precipitation amount and daily precipitation intensity are of surprisingly low importance. However, hardly any clear patterns through which these predictors contribute to model performance could be identified. Consequently, no direct means through which the process-based model could be improved to achieve better simulation results could be derived. Also regarding the hindcasts, precipitation was identified as the most significant source of uncertainty. It was found that rainfall hindcasts from the downscaled GCM show statistical properties significantly distinct from observations. Therefore, simple approaches, such as the tested monthly regional bias correction, already result in improved hindcast accuracies. Future studies should also consider the use of more sophisticated means of preprocessing as well as post-processing approaches, such as forecast error modelling, or innovative data assimilation and data fusion approaches to correct erroneous model states.

So, what is the added value of a process-based hydrological model? In our case, when it comes to reservoir level or mere drought event prediction on regionally and monthly aggregated scales, a statistical model proved to be the better option, as computational effort is much lower and the model is easier to apply. Nevertheless, we advocate the application of an appropriate process-based hydrological model in cases where predictions on finer spatial (e.g. for individual reservoirs) and temporal scales are desirable. In cases for which information on more hydrological processes or variables, such as evapotranspiration or various runoff generation and concentration variables, is required, a process-based model is the right choice. As such, due to the explanatory power of process-based hydrological models, decision makers and stakeholders can be supported to detect and understand hydrological changes in their catchments in order to derive reasonable and sustainable decisions.

These conclusions are likely to differ among study sites, primarily depending on the characteristics and the degree of stochasticity of the hydro-meteorological system and available data. In regions where hydrological variables can be described by covariates, which in turn can be forecast with high reliability, statistical models might be advantageous. This would more likely be the case in climatic regions that are characterised by a high degree of seasonality, such as in many semiarid regions around the globe. In contrast, process-based models are independent of such correlation patterns and are more generically applicable due to the underlying process representations. They can be applied under any environmental and hydro-climatological conditions, for which the incorporated process formulations constitute valid working hypotheses, and thereby primarily depend on rainfall input. Consequently, they can also be used in regions characterised by more chaotic weather and low predictability over seasonal scales, such as in temperate climatic zones, provided adequate rainfall forecasts can be delivered.

Therefore, further research is needed to increase the accuracy of important model drivers, i.e. for many regions as well as in our case of dry northeastern Brazil, first and foremost precipitation. We expect that the use of new data products, such as rainfall radar and satellite data along with conventional data from rainfall stations with sub-daily resolution, in combination with innovative methods of data assimilation and data fusion, may provide opportunities to improve forecast accuracy, in particular for process-based hydrological models. In that respect, the time and effort of their application can be justified and allow for the exploitation of their advanced capabilities.

Code and data availability

Meteorological observations (except precipitation) are available from `http://careyking.com/data-downloads/`. Precipitation and raw data of meteorological hindcasts need to be requested from FUNCEME. DEM raw data can be obtained via `http://srtm.csi.cgiar.org/SELECTION/inputCoord.asp` (tiles (horizontal/vertical): 28/13, 28/14, 29/13, and 29/14). Reservoir data can be accessed at `http://www.hidro.ce.gov.br` or requested from FUNCEME. Land cover and soil maps are not publicly available. The WASA-SED model is available at `https://github.com/TillF/WASA-SED`. Scripts to investigate or reproduce experiments, analyses, and compilation of plots can be accessed at `https://github.com/tpilz/paper_drought_prediction_brazil`.

Author contributions

All authors contributed to the experiment design. Meteorological hindcasts were preprocessed by KV and JMD. Experiments with the statistical model were conducted by SV and JMD. All other experiments and analyses were conducted, the figures were generated, and the paper was written by TP with support by the co-authors.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

Authors thank Gerd Bürger for his comments and discussions. Tobias Pilz acknowledges funding by the Helmholtz graduate school GeoSim. We acknowledge the support of the Deutsche Forschungsgemeinschaft (German Research Foundation) and Open Access Publication Fund of Potsdam University.

This paper was edited by Shraddhanand Shukla and reviewed by two anonymous referees.

5.A Appendix

5.A.1 Terminology

The word *forecast* generally refers to model-based estimations of future meteorological or hydrological variables such as precipitation, streamflow, or reservoir level. The term *prediction*, in this article like in many others, can be used synonymously to forecast. With *hindcast* we specifically denote retrospective forecasts, i.e. predictions issued for a period in the past building only on data available up to the time of start of the model run. The results are then compared with observations. In some occasions, the terms forecast and hindcast might be used interchangeably. In contrast to predictions or hindcasts, we denote *model simulations* as model runs driven by observations instead of forecasts of model forcing.

Many of the notions discussed in this article refer to the field of forecast verification. While being standard in atmospherical sciences, some terms are less common for the hydrological community and thus will be briefly explained in the following. For more information, the reader is generally referred to textbooks such as Wilks (2005). The analysis of drought hindcasts will focus on their *quality*, i.e. the correspondence of such hindcasts with observations. This quality as defined by Murphy (1993) can be described in terms of nine different aspects of which five will be addressed explicitly in this study: *accuracy* as the average agreement of forecast–observation pairs which is as such inversely proportional to the *error*; *reliability* which, in the case of probabilistic drought forecasts, quantifies the average correspondence of forecast probabilities and observed drought occurrences; *resolution* evaluating the ability of a model to correctly predict an event; *sharpness* describing the variability of forecasts of a model; and *skill* comparing the ability of a model with a much simpler reference model, such as climatology (which is the observed long-term average of a specific variable) or persistence (i.e. no change of a variable or the pattern of a quantity over the forecast period).

Furthermore, we distinguish *process-based* from *statistical* models. The former are rather complex computer programs combining a set of mathematical equations (simple, linear up to complex differential equations), which can be derived from first-order principles, e.g. conservation of mass and energy. The aim here is to represent, up to a certain degree of abstraction, the governing subprocesses of the hydrological cycle and their interactions. They compute estimates of the unknown variables (e.g. river discharge, soil moisture, reservoir storage) as a reaction to a set of input or *driving* variables (e.g. precipitation, solar radiation, water abstraction). In this paper, the underlying process-based hydrological model refers to the WASA-SED model which is described in Sect. 5.3.4. The latter, on the other hand, relies on purely empirical relationships between one or more predictors and the target variable, often consisting of only a single equation, typically obtained by regression. Consequently, the regression model of Delgado et al. (2018b), which is used for model intercomparison in this study, is referred to as the statistical model or statistical approach throughout this work.

6. Discussion and conclusions

6.1 Summary of results

The general aim of this thesis was to pursue the understanding of uncertainties in hydrological modelling. To achieve this, existing research gaps have been identified and analysed in more detail. In the following, the general results shall be summarised with respect to the specific research questions formulated at the outset of this thesis.

What are the strengths and deficits of frequently applied hydrological models? What are the reasons for deficits and how can they be alleviated?

The diversity of landscapes and hydrological conditions, datasets, and research objectives led to a large variety of hydrological models with different conceptualisations and complexity. To capture this complexity and explore common strengths and weaknesses of hydrological models, an online survey addressed to German-speaking model developers and users was conducted (Chapter 2). This survey, though not fully representative, constitutes a starting point for future model enhancement. First of all, it was found that there is a high degree of subjectivity in the classification of models into conceptual groups, as some models were classified differently by different participants of the survey. This can be explained by the fact, that different process representations can be implemented with a varying degree of complexity within one model. Nevertheless, evapotranspiration processes are usually represented in a complex and more physically based manner, while others, especially processes related to groundwater, are often simplified. As a result, the participants of the survey were especially calling for a revision of soil and groundwater processes in their models. However, there was a great diversity in the specifications of strengths and deficits of models and their causes. In general, flexibility of models, for instance because they include a selection of alternative process representations or are relatively independent of parameter calibration, was seen as a strength by the participants.

How does the methodology of discretisation of landscapes into spatial model units influence simulation results?

Landscapes and environmental conditions are parametrised in many different ways in computer models. However, it turned out that not only different philosophies of landscape discretisation exist, but also different algorithms and computer tools, even for the same discretisation approach (Chapter 3). Yet many approaches require user interactions and allow only limited or no workflow automation at all. Consequently, the influence of typical decisions along the discretisation process has hardly been investigated. Therefore, a new software package has been developed to largely automatise the processing steps of hillslope-based landscape parametrisation. This enabled the investigation of crucial discretisation parameters via Monte Carlo-based sensitivity analysis. Within the performed case study, model output appeared to be relatively independent from such decisions on discretisation. The highest impact could be directed towards spatial resolution, represented by subbasin size as well as the number of elementary hillslope units, which formed the basis for the derivation of landscape units, the spatial entities in the employed hydrological model.

What is the adequate structure for a process-based hydrological model?

To identify the most adequate model structure for a hydrological simulation, several alternative structures have been transferred into a flexible model environment (Chapter 4). Model structure identification was performed by conducting dynamic, i.e. time varying, identifiability analysis. Model structures in this study consisted of alternative process representations of evapotranspiration, soil water movement, and runoff concentration. In addition, different numerical solvers for the temporal integration of the underlying ordinary differential equations were tested, which has not been done before in the field of process-based hydrological modelling. Different parametrisations of effective model parameters were further included and analysed in an integrated way together with the choices for specific model structures. It turned out that process formulations of evapotranspiration are better identifiable, i.e. specific evapotranspiration formulas consistently achieve acceptable model performance while others do not, in comparison to the other process groups, where often no superior formulation could be detected. However, the optimal combination of model structure and parametrisation varied in space and time. There is strong evidence that identifiability of certain model configurations is primarily driven by meteorological conditions, especially rainfall, and hydrological characteristics of the landscape. Surprisingly, the most accurate numerical solver is not always leading to the best simulation results. There are indications that this can be attributed to compensation effects due to correlations among the numerical solver, process formulations, and parametrisation, where imperfect model structures result in optimal model performance with a certain parametrisation. Eventually it has to be noted that there is no single best performing combination of process representation, numerical solver, and parametrisation that gives optimal results under all conditions, but some combinations are more suitable than others.

Are process-based models suitable tools for operational forecasting? What are the deficits and how can forecasts be improved?

In operational forecasting, fast computation time is crucial. Rapidly growing computational resources more and more allow for fast initialisation and application of even complex process-based models. Therefore, the suitability of a process-based model has been analysed in comparison to a simpler statistical regression approach in an experiment of seasonal drought prediction in northeast Brazil (Chapter 5). Regarding the forecasts of regional reservoir filling, which was defined as drought proxy, the statistical approach delivered results superior to the process-based model. However, both models achieved acceptable and comparable results in terms of prediction performance of mere drought occurrence. Moreover, when driving both models with measured instead of predicted rainfall, the process-based model achieved a simulation performance superior to the regression. Generally speaking, uncertainty in the rainfall prediction could be identified as the largest source of uncertainty. This is more relevant for the process-based than for the statistical approach. Consequently, the improvement of hydrological forecasts should primarily focus on the correction of rainfall prediction as main driver for subsequent hydrological predictions, for instance by employing bias correction or other means of pre-processing. Post-processing methods may further improve potentially erroneous hydrological predictions. Even though they demand more computational resources and greater efforts in their initialisation, advantages of process-based models are that they can readily produce hydrological predictions at finer spatial and temporal scales, and provide information for other variables such as evapotranspiration or runoff. It can be concluded, that, as long as sufficient computational resources are available, the application of complex process-based models in operational forecasting should not be limited by computer power anymore. However, in order to obtain adequate predictions and to be able to use the full capacity of process-based models, rigorous initialisation as well as pre- and post-processing are vital.

6.2 Discussion and directions for further research

6.2.1 Advances on model deficits and methods for their identification

Different strategies were pursued in this thesis to explore the deficits and uncertainties of hydrological models. These can be roughly classified into informal and formal approaches.

The online survey presented in Chapter 2 can be regarded as informal approach. It directly aims at collecting and analysing the needs and concerns of modellers, but is associated with a high degree of subjectivity as the results are influenced by the perceptions on hydrological processes, experiences, and subjective opinions of the survey participants and the different purposes of models. Sensitivity and identifiability analysis are more formal approaches to assess the impact of uncertain input factors and determine the behavioural range, where input factors produce adequate model performances. Thereby they support uncertainty analysis and can as well be used to identify deficits in model structures. Such analyses were presented in Chapter 3, where landscape discretisation was investigated, and Chapter 4, which focused on model structures. Other formal approaches exist, treating individual models as source of uncertainty in an integrated way, but thereby obscuring the various uncertainties associated with a specific model, such as is done in Bayesian model averaging (Duan et al., 2007; Hoeting et al., 1999). In general, the identification and elimination of model deficits can be conducted by rejection of certain models or model structures (Graeff et al., 2009), stepwise improvement (Fenicia et al., 2008), or even mixed approaches (Francke et al., 2018b).

In this work it turned out that in many hydrological models evapotranspiration processes are implemented in a more complex, physically based manner while processes of groundwater and soil water movement are often simplified. The latter is seen as a deficit by many modellers, who therefore call for a revision of these process implementations. On the other hand, even so-called physically based evapotranspiration approaches rely on strong simplifications of real-world processes. An example is the frequently used Penman-Monteith formula, which is based on the so-called big-leaf approach, where the land surface is reduced to a homogeneous plant cover, and water and energy fluxes are described by the concept of resistances (for a comprehensive discussion see Shuttleworth, 2007). Consequently, the experiments shown in Chapter 4 demonstrate that even process-based formulations of evapotranspiration processes may lead to different model performances under contrasting hydrological and meteorological conditions, depending on the assumptions and simplifications of a specific approach. Moreover, the selection of a specific evapotranspiration model had a larger influence on model results than a specific approach for soil water movement.

Deficiencies of model structures can often be compensated by their parametrisation. On the other hand, such parametrisation might by physically unrealistic and limit the predictability of models and their transferability to (even slightly) different environments. Moreover, compensation effects can distort the findings of formal approaches, eventually leading to surprising results (e.g. the low identifiability of accurate numerical solvers shown in Chapter 4) and consequently impeding model improvement.

Model resolution is another important aspect. In Chapter 3, the definition of the sizes of computational units was identified as the most influential factor during model initialisation. Temporal resolution, on the other hand, affects the integration of model equations. This becomes relevant if water fluxes occur over characteristic time scales different than the size of model time steps. However, increasing model resolution is often computationally not feasible. Therefore, this issue adds a further dimension to the compensation effects and hinders model structure identification (Chapter 4).

In future studies, it will be essential to reduce undesired side-effects to a minimum. However, strategies need to be found to increase model resolution, while keeping runtimes computationally feasible and/or applying statistical methods (such as sensitivity analysis) with a small number of model evaluations. To determine the optimal model structure, data science (e.g. machine learning) could be combined with process knowledge before conducting time consuming model runs. In order to advance process understanding and implementation into models, workshops of model application, such as presented by Holländer et al. (2014), could provide further insights by comparing model results and opinions of modellers and thereby accounting for the impact of subjective decisions. To achieve such goals, flexible model frameworks as employed in Chapter 4 are promising tools to conduct experiments in a controlled framework by following a predefined protocol, reducing the chances of random errors and undesired side-effects.

6.2.2 Software solutions for explorative uncertainty analysis

To study uncertainties in the field of hydrological modelling, a large number of model runs and repeated model application is essential to solve the involved complex differential equations, which are analytically intractable. Therefore, an important focus of this thesis was the development and enhancement of efficient computer tools to automatise workflows. The lumpR package for the statistical software environment R is such a tool, which serves the discretisation of landscapes into model units and the initialisation of hydrological models (Chapter 3). As such it complements the numerous existing software solutions for landscape discretisation (see Section 3.2.3 for a review) and integrates features that these do not offer or contain only partially. It is free and open source and is flexible in a way that it allows for complete workflow automation, which enables the efficient use in Monte Carlo simulation, but still allows user interaction, such as to include additional information depending on data availability, and therefore enables optimal characterisation of spatial model units.

Flexible hydrological models have been proven as efficient tools to study process dynamics in catchments and refine process descriptions. As such, a selection of alternative process implementations was seen as a strength of models by participants of the survey presented in Chapter 2. During the work on this thesis, the eco-hydrological simulation environment (ECHSE) introduced by Kneis (2015) has been extended and employed (Chapter 4). It builds upon the philosophy of other flexible frameworks such as FLEX (Fenicia et al., 2008), SUPERFLEX (Fenicia et al., 2011), or FUSE (Clark et al., 2008b). These, however, are solely based on simplified conceptualisations of hydrological processes. The SUMMA framework by Clark et al. (2015b) already contains more rigorous physically based formulations of fluxes of mass and energy, comes with a range of alternative process formulations, and includes a more advanced numerical integration of fluxes than is commonly used, but only supports less complex landscape discretisation. ECHSE, on the other hand, is more flexible in a way that any kind of complexity regarding spatial discretisation or process formulation can be included. The HYPSO-RR engine, for instance, is a less complex conceptual model, while the WASA-SED engine, which was developed and used along this work, involves more complex process formulations and spatial disaggregation. A drawback still of all such hydrological model environments focussing on (near) surface processes on the catchment scale is that no partial differential equations can be solved, i.e. no explicit 2-D or 3-D integration of fluxes is possible. Moreover, even with today's capabilities, computational burden is still a major drawback when applying process-based hydrological models with advanced numerical solvers over catchment scales larger than 1000 km^2. However, as for ECHSE, the extensive computational load could partly be attributed to input–output operations, which could be optimized by reorganising the internal data handling.

Having this in mind, clear recommendations about future advances are difficult. A common model for the whole community would certainly be advantageous as individual efforts could be better consolidated than it is the case nowadays, where individual groups of modellers tend to work on their own advances. This leads to parallel developments and therefore wasting of efforts and potentially oversight of significant ideas. On the other hand, there are yet unresolved questions, such as how should generically applicable process representations and closure schemes be defined (Weiler and Beven, 2015). In the meantime, the use of flexible environments is certainly advantageous over static model configurations, as process formulation can be integrated, exchanged, refined, and tested in a straightforward manner.

6.2.3 What is the largest source of uncertainty?

The preceding discussion focussed on model architecture and process formulations. But is this really the key towards optimal model simulations? Many studies (see for instance McMillan et al., 2012) as well as the experiments of this thesis identified model forcing as the primary source of uncertainty, i.e. first and foremost precipitation. For instance, in Chapter 4 deviations of streamflow simulations from observations were largest in case of station failure. This is further supported by Bárdossy and Das (2008) and Francke et al. (2018b) who found that the number and distribution of rain gauges affect simulation results, and Yatheendradas et al. (2008) who identified rainfall volume bias as the most significant source of uncertainty in flash flood forecasts.

A common strategy, especially in flood forecasting, is to account for potential deficits in precipitation forcing by pre-processing of the rainfall dataset and error modelling (e.g. Kavetski et al., 2006; Kelly and Krzysztofowicz, 2000; Renard et al., 2010). Yet, Fraga et al. (2019) concludes that accounting for rainfall uncertainty only has little effect, while the correct reproduction of spatial and temporal evolutions of rainfall fields is more important. This could be realised by increasing the spatial resolution of the model entities, which direct precipitation into the model. In the WASA-SED model, for instance, these are the subbasins and, indeed, in Chapter 3 subbasin size has been identified as the most important discretisation parameter. Moreover, during the implementation of process representations into the ECHSE framework, synthetic experiments for the spatio-temporal investigation of soil water movement in the model showed that there is a high influence of spatial and temporal resolution on adequate model functioning (not shown in this work). On the other hand, the higher the model resolution, the more important is a high station density, which is often not given (Bárdossy and Das, 2008). To overcome this issue, remote sensing data, such as satellite or radar measurements, can be used as they provide a high spatial and temporal resolution. This usually comes at the cost of a high uncertainty in rainfall volume estimates, which is why remote sensing data are commonly used in combination with station-based information (Abon et al., 2016; Kneis et al., 2014; Yatheendradas et al., 2008).

Considering all the potential means for improvement of both the model and its forcing, it should be kept in mind that calibration is still often able to compensate for many deficits in a way that simulated discharge dynamics adequately resembles observations. At the same time, however, internal sub-processes such as soil moisture dynamics are often unrealistic, e.g. exceeding physical limits when integrating the differential equations without solution constraints. This may influence the predictive capabilities of models and potentially limit the ability of models to be applied under changing boundary conditions, e.g. in studies of climate change impacts. In the end it can be concluded that the improvement of models and their input data as well as the elimination of deficits strongly depends on the field of study and associated objectives. For instance, an improved representation of underlying processes may lead to a better representation of water fluxes while model performance is decreasing, possibly because rainfall input is the main source of uncertainty that can no longer be compensated by parametrisation.

6.3 Conclusion

This thesis provides an overview of uncertainties associated with the application of hydrological models and aims at filling existing gaps of knowledge. A study of existing literature reveals that many studies on uncertainties as well as computer tools for their quantification already exist. However, there is still a large gap with respect to structural uncertainties and more complex process-based models, which was therefore the focus of this work. Specifically, a survey among modellers was conducted to collect and analyse their perspectives and concerns, new software tools have been developed, and computer experiments were carried out to analyse model uncertainties. The main conclusions and associated needs for further research are summarised in the following.

- Hydrological modellers have distinct perceptions of the various conceptions of hydrological models that exist and how models should be applied. As a consequence, results of the various models applied by different modellers are often hardly comparable, which hampers the discovery of common deficits and subsequent improvement of models.

- The software lumpR for automated landscape discretisation and model initialisation was developed and its applicability proven in a case study. The software was further used for sensitivity analysis of user decisions and to initialise all further model applications presented in this thesis. It turned out that spatial resolution in terms of subbasin size and the number of spatial model units are the most sensitive parameters.

- Flexible model environments were identified as useful tools for the rapid development, exchange, refinement, and testing of alternative model structures. Their use instead of static models is therefore recommended.

- The flexible model environment ECHSE has been extended by a new model engine providing a number of complex, process-based model structures. Together with dynamic identifiability analysis it was used a diagnostic tool for process-based model building. So far, research in this field focussed on less complex conceptual model structures and often neglected the impact of inaccurate numerical solvers to avoid excessive computational burden. However, the presented approach proved to be efficient for the exploration of complex process formulations, numerical solvers, and associated deficits and uncertainties.

- It was found that the performance of specific model structures and therefore their identifiability is determined by the current meteorological conditions and hydrological characteristics. Consequently, no optimal model structure exists and compromises are inevitable during model selection. However, in future studies, data science and process knowledge could be combined in order to determine the optimal model structure before conducting time consuming model runs.

- In general, rainfall input was the largest source of uncertainty. The impact was found to be more relevant for complex process-based than for statistical models. Innovative data sources (such as remote sensing products) in combination with well maintained local stations and processing algorithms are required to remedy this deficit.

- Even though modellers generally call for a revision and more detailed description of soil and groundwater processes in their models, this study found a larger impact of alternative (but similarly complex and process-based) evapotranspiration approaches on general model performance. Although they are often regarded as physically based, even complex approaches such as the Penman–Monteith formula are based on a number of simplifications that should be revised by future studies.

- Parametrisation can compensate for deficits in model structures. However, this effect may as well obscure the results of sensitivity and identifiability analyses and therefore hinder model selection and improvement. Consequently, as the best performance metric might not be related to the most plausible model structure, the capabilities are limited to transfer a model into different environments or apply them under changing boundary conditions, such as in studies of climate change impacts.

- In addition to parametrisation, insufficient model resolution can induce further undesired side-effects. For instance, the temporal resolution of one day, common in most hydrological models, is much larger than the characteristic time scale of soil water processes, which affects the integration of process equations and lead to unrealistic soil moisture states. Therefore, investigations are needed, how finer model resolution can be achieved while keeping computational burden within acceptable limits.

Bibliography

Abbott, M., Bathurst, J., Cunge, J., O'Connell, P., and Rasmussen, J. (1986). "An introduction to the European Hydrological System — Systeme Hydrologique Europeen, "SHE", 1: History and philosophy of a physically-based, distributed modelling system". In: *J. Hydrol.* 87.1, pp. 45–59. DOI: `https://doi.org/10.1016/0022-1694(86)90114-9`.

Abon, C. C., Kneis, D., Crisologo, I., Bronstert, A., David, C. P. C., and Heistermann, M. (2016). "Evaluating the potential of radar-based rainfall estimates for streamflow and flood simulations in the Philippines". In: *Geomat. Nat. Haz. Risk* 7.4, pp. 1390–1405. DOI: `10.1080/19475705.2015.1058862`.

Ajami, H., Khan, U., Tuteja, N. K., and Sharma, A. (2016). "Development of a computationally efficient semi-distributed hydrologic modeling application for soil moisture, lateral flow and runoff simulation". In: *Environ. Modell. Softw.* 85, pp. 319–331. DOI: `10.1016/j.envsoft.2016.09.002`.

Ajami, N. K., Duan, Q., and Sorooshian, S. (2007). "An integrated hydrologic Bayesian multi-model combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction". In: *Water Resour. Res.* 43.1, W01403. DOI: `10.1029/2005wr004745`.

Allen, R. G., Walter, I. A., Elliott, R., Terry Howell, D. I., and Jensen, M. (2005). *The ASCE standardized reference evapotranspiration equation*. Report prepared by the Task Committee on Standardization of Reference Evapotranspiration. Environmental and Water Resources Institute of the American Society of Civil Engineers (ASCE).

Alves, J. M. B., Campos, J. N. B., and Servain, J. (2012). "Reservoir Management Using Coupled Atmospheric and Hydrological Models: The Brazilian Semi-Arid Case". In: *Water Resour. Manage.* 26.5, pp. 1365–1385. DOI: `10.1007/s11269-011-9963-2`.

de Araújo, J. C., Doll, P., Gúntner, A., Krol, M., Abreu, C. B. R., Hauschild, M., and Mendiondo, E. M. (2004). "Water scarcity under scenarios for global climate change and regional development in semiarid northeastern Brazil". In: *Water Int.* 29.2, pp. 209–220.

de Araújo, J. C. and Medeiros, P. H. A. (2013). "Impact of Dense Reservoir Networks on Water Resources in Semiarid Environments". In: *Aust. J. Water Resour.* 17.1, pp. 87–100. DOI: `10.7158/13241583.2013.11465422`.

de Araújo, J. C. and Bronstert, A. (2016). "A method to assess hydrological drought in semi-arid environments and its application to the Jaguaribe River basin, Brazil". In: *Water Int.* 41.2, pp. 213–230. DOI: 10.1080/02508060.2015.1113077. eprint: http://dx.doi.org/10.1080/02508060.2015.1113077.

Band, L. E., Tague, C. L., Brun, S. E., Tenenbaum, D. E., and Fernandes, R. A. (2000). "Modelling Watersheds as Spatial Object Hierarchies: Structure and Dynamics". In: *Trans. GIS* 4.3, pp. 181–196. DOI: 10.1111/1467-9671.00048.

Bárdossy, A. and Das, T. (2008). "Influence of rainfall observation network on model calibration and application". In: *Hydrol. Earth Syst. Sc.* 12.1, pp. 77–89. DOI: 10.5194/hess-12-77-2008.

Baroni, G. and Tarantola, S. (2014). "A General Probabilistic Framework for uncertainty and global sensitivity analysis of deterministic models: A hydrological case study". In: *Environ. Modell. Softw.* 51, pp. 26–34. DOI: 10.1016/j.envsoft.2013.09.022.

Barros, F. V. F., de Maria Alves, C., Martins, E. S. P. R., and Reis Jr., D. S. (2013). "The Development and Application of Information System for Water Management and Allocation (SIGA) to a Negotiable Water Allocation Process in Brazil". In: *World Environmental and Water Resources Congress*. DOI: 10.1061/9780784412947.128.

Bergström, S. (1992). *The HBV model – its structure and applications*. SMHI Reports Hydrology 4. Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden.

Beven, K. (2007). "Towards integrated environmental models of everywhere: uncertainty, data and modelling as a learning process". In: *Hydrol. Earth Syst. Sc.* 11.1, pp. 460–467. DOI: 10.5194/hess-11-460-2007.

Beven, K. (1989). "Changing ideas in hydrology – The case of physically-based models". In: *J. Hydrol.* 105.1, pp. 157–172. DOI: 10.1016/0022-1694(89)90101-7.

Beven, K. (1993). "Prophecy, reality and uncertainty in distributed hydrological modelling". In: *Adv. Water Resour.* 16.1, pp. 41–51. DOI: 10.1016/0309-1708(93)90028-E.

Beven, K. (1995). "Linking parameters across scales: Subgrid parameterizations and scale dependent hydrological models". In: *Hydrol. Process.* 9.5-6, pp. 507–525. DOI: 10.1002/hyp.3360090504.

Beven, K. (2006a). "A manifesto for the equifinality thesis". In: *J. Hydrol.* 320.1, pp. 18–36. DOI: 10.1016/j.jhydrol.2005.07.007.

Beven, K. (2006b). "Searching for the Holy Grail of scientific hydrology: $Q_t = H(S, R, \Delta t)A$ as closure". In: *Hydrol. Earth Syst. Sc.* 10.5, pp. 609–618. DOI: 10.5194/hess-10-609-2006.

Beven, K. (2009). *Environmental modelling: an uncertain future?* Routledge, London, UK.

Beven, K. and Binley, A. (1992). "The future of distributed models: Model calibration and uncertainty prediction". In: *Hydrol. Process.* 6.3, pp. 279–298. DOI: 10.1002/hyp.3360060305.

Beven, K., Calver, A., and Morris, E. M. (1987). *The Institute of Hydrology distributed model*. IH Report 98. Institute of Hydrology, Wallingford, UK.

Block, P. J., Souza Filho, F. A., Sun, L., and Kwon, H.-H. (2009). "A Streamflow Forecasting Framework using Multiple Climate and Hydrological Models". In: *J. Am. Water Resour. As.* 45.4, pp. 828–843. DOI: 10.1111/j.1752-1688.2009.00327.x.

Block, P. and Rajagopalan, B. (2009). "Statistical–Dynamical Approach for Streamflow Modeling at Malakal, Sudan, on the White Nile River". In: *J. Hydrol. Eng.* 14.2, pp. 185–196. DOI: 10.1061/(ASCE)1084-0699(2009)14:2(185).

Blöschl, G., Nester, T., Parajka, J., and Komma, J. (2014). "Hochwasserprognosen an der österreichischen Donau und Datenassimilation". In: *Hydrol. Wasserbewirtsch.* 58.2, pp. 64–72.

Blöschl, G. and Sivapalan, M. (1995). "Scale issues in hydrological modelling: a review". In: *Hydrol. Process.* 9.3-4, pp. 251–290. DOI: 10.1002/hyp.3360090305.

Boé, J., Terray, L., Habets, F., and Martin, E. (2007). "Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies". In: *Int. J. Climatol.* 27.12, pp. 1643–1655. DOI: 10.1002/joc.1602. eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.1602.

Bogaart, P. W. and Troch, P. A. (2006). "Curvature distribution within hillslopes and catchments and its effect on the hydrological response". In: *Hydrol. Earth Syst. Sc.* 10.6, pp. 925–936. DOI: 10.5194/hess-10-925-2006.

Bormann, H., Breuer, L., Giertz, S., Huisman, J. A., and Viney, N. R. (2009a). “Spatially explicit versus lumped models in catchment hydrology – experiences from two case studies”. In: *Uncertainties in Environmental Modelling and Consequences for Policy Making*. Ed. by Baveye, P., Mysiak, J., and Laba, M. Springer Netherlands, pp. 3–26. DOI: 10.1007/978-90-481-2636-1_1.

Bormann, H., Breuer, L., Graeff, T., Huisman, J., and Croke, B. (2009b). “Assessing the impact of land use change on hydrology by ensemble modelling (LUCHEM) IV: Model sensitivity to data aggregation and spatial (re-)distribution”. In: *Adv. Water Resour.* 32.2, pp. 171–192. DOI: 10.1016/j.advwatres.2008.01.002.

Bourdin, D. R., Nipen, T. N., and Stull, R. B. (2014). “Reliable probabilistic forecasts from an ensemble reservoir inflow forecasting system”. In: *Water Resour. Res.* 50.4, pp. 3108–3130. DOI: 10.1002/2014wr015462.

Braga, C., De Nys, E., Frazao, C., Martins, P., and Martins, E. S. P. R. (2013). *Impacto das mudanças do clima e projeções de Demanda sobre o processo de alocação de água em Duas bacias do nordeste semiárido (Portuguese)*. Agua Brasil series 8. Washington, DC, USA: World Bank Group.

Breiman, L. (2001). “Random Forests”. In: *Mach. Learn.* 45.1, pp. 5–32. DOI: 10.1023/A:1010933404324.

Breuer, L., Huisman, J., Willems, P., Bormann, H., Bronstert, A., Croke, B., Frede, H.-G., Gräff, T., Hubrechts, L., Jakeman, A., Kite, G., Lanini, J., Leavesley, G., Lettenmaier, D., Lindström, G., Seibert, J., Sivapalan, M., and Viney, N. (2009). “Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM). I: Model intercomparison with current land use”. In: *Adv. Water Resour.* 32.2, pp. 129–146. DOI: 10.1016/j.advwatres.2008.10.003.

Bronstert, A., Jaeger, A., Güntner, A., Hauschild, N., Doll, P., and Krol, M. (2000). “Integrated modelling of water availability and water use in the semi-arid Northeast of Brazil”. In: *Phys. Chem. Earth Pt. B* 25.3, pp. 227–232. DOI: 10.1016/S1464-1909(00)00008-3.

Bronstert, A. (1994). “Modellierung der Abflussbildung und der Bodenwasserdynamik von Hängen”. In: *Mitteilungen des Instituts für Hydrologie und Wasserwirtschaft*. 46. Universität Karlsruhe.

Bronstert, A. (1999). “Capabilities and limitations of detailed hillslope hydrological modelling”. In: *Hydrol. Process.* 13.1, pp. 21–48. DOI: 10.1002/(SICI)1099-1085(199901)13:1<21::AID-HYP702>3.0.CO;2-4.

Bronstert, A., de Araújo, J.-C., Batalla, R. J., Costa, A. C., Delgado, J. M., Francke, T., Foerster, S., Guentner, A., López-Tarazón, J. A., Mamede, G. L., Medeiros, P. H., Mueller, E., and Vericat, D. (2014). “Process-based modelling of erosion, sediment transport and reservoir siltation in mesoscale semi-arid catchments”. In: *J. Soils Sediments* 14.12, pp. 2001–2018. DOI: 10.1007/s11368-014-0994-1.

Bronstert, A. and Bárdossy, A. (1999). “The role of spatial variability of soil moisture for modelling surface runoff generation at the small catchment scale”. In: *Hydrol. Earth Syst. Sc.* 3.4, pp. 505–516. DOI: 10.5194/hess-3-505-1999.

Brutsaert, W. (1975). “Comments on Surface Roughness Parameters and the Height of Dense Vegetation”. In: *J. Meteorolog. Soc. Jpn.* 53.1, pp. 96–98. DOI: 10.2151/jmsj1965.53.1_96.

Bürger, G., Reusser, D., and Kneis, D. (2009). “Early flood warnings from empirical (expanded) downscaling of the full ECMWF Ensemble Prediction System”. In: *Water Resour. Res.* 45.10, W10443. DOI: 10.1029/2009WR007779.

Chen, H., Yang, D., Hong, Y., Gourley, J. J., and Zhang, Y. (2013). “Hydrological data assimilation with the Ensemble Square-Root-Filter: Use of streamflow observations to update model states for real-time flash flood forecasting”. In: *Adv. Water Resour.* 59, pp. 209–220. DOI: 10.1016/j.advwatres.2013.06.010.

Clark, M. P., Bierkens, M. F. P., Samaniego, L., Woods, R. A., Uijlenhoet, R., Bennett, K. E., Pauwels, V. R. N., Cai, X., Wood, A. W., and Peters-Lidard, C. D. (2017). “The evolution of process-based hydrologic models: historical challenges and the collective quest for physical realism”. In: *Hydrol. Earth Syst. Sc.* 21.7, pp. 3427–3440. DOI: 10.5194/hess-21-3427-2017.

Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., Hooper, R. P., Kumar, M., Leung, L. R., Mackay, D. S., Maxwell, R. M., Shen, C., Swenson, S. C., and

Zeng, X. (2015a). “Improving the representation of hydrologic processes in Earth System Models”. In: *Water Resour. Res.* 51.8, pp. 5929–5956. DOI: 10.1002/2015WR017096.

Clark, M. P., Kavetski, D., and Fenicia, F. (2011a). “Pursuing the method of multiple working hypotheses for hydrological modeling”. In: *Water Resour. Res.* 47.9, W09301. DOI: 10.1029/2010wr009827.

Clark, M. P., Kavetski, D., and Fenicia, F. (2011b). “Pursuing the method of multiple working hypotheses for hydrological modeling”. In: *Water Resour. Res.* 47.9, W09301. DOI: https://doi.org/10.1029/2010wr009827.

Clark, M. P., McMillan, H. K., Collins, D. B. G., Kavetski, D., and Woods, R. A. (2011c). “Hydrological field data from a modeller’s perspective: Part 2: process-based evaluation of model hypotheses”. In: *Hydrol. Process.* 25.4, pp. 523–543. DOI: 10.1002/hyp.7902.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Brekke, L. D., Arnold, J. R., Gochis, D. J., and Rasmussen, R. M. (2015b). “A unified approach for process-based hydrologic modeling: 1. Modeling concept”. In: *Water Resour. Res.* 51.4, pp. 2498–2514. DOI: 10.1002/2015WR017198.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Gochis, D. J., Rasmussen, R. M., Tarboton, D. G., Mahat, V., Flerchinger, G. N., and Marks, D. G. (2015c). “A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies”. In: *Water Resour. Res.* 51.4, pp. 2515–2542. DOI: 10.1002/2015WR017200.

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J., and Uddstrom, M. J. (2008a). “Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model”. In: *Adv. Water Resour.* 31.10, pp. 1309–1324. DOI: 10.1016/j.advwatres.2008.06.005.

Clark, M. P., Schaefli, B., Schymanski, S. J., Samaniego, L., Luce, C. H., Jackson, B. M., Freer, J. E., Arnold, J. R., Dan Moore, R., Istanbulluoglu, E., and Ceola, S. (2016). “Improving the theoretical underpinnings of process-based hydrologic models”. In: *Water Resour. Res.* 52.3, pp. 2350–2365. DOI: 10.1002/2015WR017910.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., and Hay, L. E. (2008b). “Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models”. In: *Water Resour. Res.* 44.12, W00B02. DOI: 10.1029/2007WR006735.

Cloke, H. and Pappenberger, F. (2009). “Ensemble flood forecasting: A review”. In: *J. Hydrol.* 375.3–4, pp. 613–626. DOI: 10.1016/j.jhydrol.2009.06.005.

Cochrane, T. and Flanagan, D. (2003). “Representative hillslope methods for applying the WEPP model with DEMs and GIS”. In: *T. ASAE* 46.4, pp. 1041–1049.

Costa-Cabral, M. C. and Burges, S. J. (1994). “Digital Elevation Model Networks (DEMON): A model of flow over hillslopes for computation of contributing and dispersal areas”. In: *Water Resour. Res.* 30.6, pp. 1681–1692. DOI: 10.1029/93WR03512.

Coxon, G., Freer, J., Wagener, T., Odoni, N. A., and Clark, M. (2014). “Diagnostic evaluation of multiple hypotheses of hydrological behaviour in a limits-of-acceptability framework for 24 UK catchments”. In: *Hydrol. Process.* 28.25, pp. 6135–6150. DOI: http://dx.doi.org/10.1002/hyp.10096.

Dawes, W. R. and Short, D. (1994). “The significance of topology for modeling the surface hydrology of fluvial landscapes”. In: *Water Resour. Res.* 30.4, pp. 1045–1055. DOI: 10.1029/93WR02479.

DeChant, C. M. and Moradkhani, H. (2012). “Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting”. In: *Water Resour. Res.* 48.4, W04518. DOI: 10.1029/2011WR011011.

Delgado, J. M., Voss, S., Bürger, G., Vormoor, K., Murawski, A., Rodrigues, M., Martins, E., Vasconcelos Júnior, F., and Francke, T. (2018a). “Seasonal Drought Prediction for Semiarid Northeast Brazil: Verification of Six Hydro-Meteorological Forecast Products”. In: *Hydrol. Earth Syst. Sc.* 22.9, pp. 5041–5056. DOI: 10.5194/hess-22-5041-2018.

Delgado, J., Zhang, S., Schuettig, M., and Foerster, S. (2018b). “Quantifying Distributed Water Availability in Small Dams in the State of Ceará, Brazil”. In: *2nd Mapping Water Bodies from Space Conference, ESA-ESRIN, Frascati, Italy*. Poster presentation.

DeVantier, B. A. and Feldman, A. D. (1993). "Review of GIS Applications in Hydrologic Modeling". In: *J. Water Res. Pl.-ASCE* 119.2, pp. 246–261. DOI: 10.1061/(ASCE)0733-9496(1993)119:2(246).

Di Luzio, M., Srinivasan, R., and Arnold, J. G. (2004). "A GIS-Coupled Hydrological Model System for the Watershed Assessment of Agricultural Nonpoint and Point Sources of Pollution". In: *Trans. GIS* 8.1, pp. 113–136. DOI: 10.1111/j.1467-9671.2004.00170.x.

Dooge, J. C. I. (1959). "A general theory of the unit hydrograph". In: *J. Geophys. Res.* 64.2, pp. 241–256. DOI: 10.1029/JZ064i002p00241.

Duan, Q., Ajami, N. K., Gao, X., and Sorooshian, S. (2007). "Multi-model ensemble hydrologic prediction using Bayesian model averaging". In: *Adv. Water Resour.* 30.5, pp. 1371–1386. DOI: 10.1016/j.advwatres.2006.11.014.

Dyck, S. and Peschke, G. (1995). *Grundlagen der Hydrologie*. 536 pages. VEB Verlag für Bauwesen.

Efstratiadis, A. and Koutsoyiannis, D. (2010). "One decade of multi-objective calibration approaches in hydrological modelling: a review". In: *Hydrolog. Sci. J.* 55.1, pp. 58–78. DOI: 10.1080/02626660903526292.

Elfert, S. and Bormann, H. (2010). "Simulated impact of past and possible future land use changes on the hydrological response of the Northern German lowland 'Hunte' catchment". In: *J. Hydrol.* 383.3–4, pp. 245–255. DOI: 10.1016/j.jhydrol.2009.12.040.

Ercolani, G. and Castelli, F. (2017). "Variational assimilation of streamflow data in distributed flood forecasting". In: *Water Resour. Res.* 53.1, pp. 158–183. DOI: 10.1002/2016WR019208.

Euser, T., Hrachowitz, M., Winsemius, H. C., and Savenije, H. H. (2015). "The effect of forcing and landscape distribution on performance and consistency of model structures". In: *Hydrol. Process.* 29.17, pp. 3727–3743. DOI: 10.1002/hyp.10445.

Evensen, G. (1994). "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics". In: *J. Geophys. Res.* 99.C5, p. 10143. DOI: 10.1029/94jc00572.

Farmer, W. H. and Vogel, R. M. (2016). "On the deterministic and stochastic use of hydrologic models". In: *Water Resour. Res.* 52.7, pp. 5619–5633. DOI: 10.1002/2016WR019129. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2016WR019129.

Fatichi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., Downer, C. W., Camporese, M., Davison, J. H., Ebel, B., Jones, N., Kim, J., Mascaro, G., Niswonger, R., Restrepo, P., Rigon, R., Shen, C., Sulis, M., and Tarboton, D. (2016). "An overview of current applications, challenges, and future trends in distributed process-based models in hydrology". In: *J. Hydrol.* 537, pp. 45–60. DOI: 10.1016/j.jhydrol.2016.03.026.

Fenicia, F., Kavetski, D., and Savenije, H. H. G. (2011). "Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development". In: *Water Resour. Res.* 47.11, W11510. DOI: 10.1029/2010WR010174.

Fenicia, F., Kavetski, D., Savenije, H. H. G., and Pfister, L. (2016). "From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions". In: *Water Resour. Res.* 52.2, pp. 954–989. DOI: 10.1002/2015WR017398.

Fenicia, F., Savenije, H. H. G., Matgen, P., and Pfister, L. (2008). "Understanding catchment behavior through stepwise model concept improvement". In: *Water Resour. Res.* 44.1, W01402. DOI: 10.1029/2006WR005563.

de Figueiredo, J. V., de Araújo, J. C., Medeiros, P. H. A., and Costa, A. C. (2016). "Runoff initiation in a preserved semiarid Caatinga small watershed, Northeastern Brazil". In: *Hydrol. Process.* 30.13, pp. 2390–2400. DOI: 10.1002/hyp.10801.

Flanagan, D. C. and Nearing, M. A. (1995). *USDA Water Erosion Prediction Project – Hillslope Profile and Watershed Model Documentation*. NSERL Report 10. USDA-ARS National Soil Erosion Research Laboratory, West Lafayette, Indiana, USA.

Flanagan, D. C., Frankenberger, J. R., Cochrane, T. A., Renschler, C. S., and Elliot, W. J. (2013). "Geospatial Application of the Water Erosion Prediction Project (WEPP) Model". In: *T. ASABE* 56.2, pp. 591–601. DOI: 10.13031/2013.42681.

Flügel, W.-A. (1995). "Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the River Bröl, Germany”. In: *Hydrol. Process.* 9.3–4, pp. 423–436. DOI: 10.1002/hyp.3360090313.

Fraga, I., Cea, L., and Puertas, J. (2019). “Effect of rainfall uncertainty on the performance of physically-based rainfall-runoff models”. In: *Hydrol. Process.* 33.1, pp. 160–173. DOI: 10.1002/hyp.13319.

Francke, T., Foerster, S., Brosinsky, A., Sommerer, E., Lopez-Tarazon, J. A., Güntner, A., Batalla, R. J., and Bronstert, A. (2018a). “Water and sediment fluxes in Mediterranean mountainous regions: comprehensive dataset for hydro-sedimentological analyses and modelling in a mesoscale catchment (River Isábena, NE Spain)”. In: *Earth Syst. Sci. Data* 10.2, pp. 1063–1075. DOI: 10.5194/essd-10-1063-2018.

Francke, T., Güntner, A., Mamede, G., Müller, E. N., and Bronstert, A. (2008). “Automated catena-based discretization of landscapes for the derivation of hydrological modelling units”. In: *Int. J. Geogr. Inf. Sci.* 22.2, pp. 111–132. DOI: 10.1080/13658810701300873.

Francke, T. (2017). *ppso: Particle Swarm Optimization and Dynamically Dimensioned Search, optionally using parallel computing based on Rmpi*. R package version 0.9-9994.

Francke, T., Baroni, G., Brosinsky, A., Foerster, S., López-Tarazón, J. A., Sommerer, E., and Bronstert, A. (2018b). “What did really improve our meso-scale hydrological model? A multi-dimensional analysis based on real observations”. In: *Water Resour. Res.* 54.11, pp. 8594–8612. DOI: 10.1029/2018WR022813.

Freeze, R. and Harlan, R. (1969). “Blueprint for a physically-based, digitally-simulated hydrologic response model”. In: *J. Hydrol.* 9.3, pp. 237–258. DOI: http://dx.doi.org/10.1016/0022-1694(69)90020-1.

Freitas, H. R. d. A., Freitas, C. d. C., Rosim, S., and Oliveira, J. R. d. F. (2016). “Drainage networks and watersheds delineation derived from TIN-based digital elevation models”. In: *Comput. Geosci.* 92, pp. 21–37. DOI: 10.1016/j.cageo.2016.04.003.

Gaiser, T., Krol, M., Frischkorn, H., and de Araújo, J. C., eds. (2003). *Global Change and Regional Impacts – Water Availability and Vulnerability of Ecosystems and Society in the Semiarid Northeast of Brazil*. Springer-Verlag Berlin Heidelberg.

Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Alken, P., Booth, M., Rossi, F., and Ulerich, R. (2017). *GNU Scientific Library release 2.4*. http://www.gnu.org/software/gsl/.

Gallant, J. C. and Wilson, J. P. (1996). “TAPES-G: A grid-based terrain analysis program for the environmental sciences”. In: *Comput. Geosci.* 22.7, pp. 713–722. DOI: 10.1016/0098-3004(96)00002-7.

Galvão, C. d. O., Nobre, P., Braga, A. C. F. M., de Oliveira, K. F., da Silva, R., Silva, S., Gomes Filho, M. F., Santos, C., Lacerda, F. F., and Moncunill, D. F. (2005). “Climatic predictability, hydrology and water resources over Nordeste Brazil”. In: *Regional Hydrological Impacts of Climatic Change: Impact assessment and decision making*. Ed. by Wagener, T., Franks, S., Gupta, H., Bogh, E., Bastidas, L., Nobre, C., and Galvão, C. d. O. Vol. 295. IAHS Publications, pp. 211–220.

Garbrecht, J. and Martz, L. W. (1999). *An Overview of TOPAZ: An Automated Digital Landscape Analysis Tool for Topographic Evaluation, Drainage Identification, Watershed Segmentation, and Subcatchment Parameterization*. Report no. GRL 99-1. Grazinglands Research Laboratory, USDA, Agricultural Research Service, El Reno, Oklahoma, USA.

García-Ruiz, J. M., Beguería, S., López-Moreno, J. I., Lorente, A., and Seeger, M. (2001). *Los recursos hídricos superficiales del Pirineo aragonés y su evolución reciente*. Geoforma, Logroño, Spain.

Gerstengarbe, F.-W. and Werner, P. C. (1999). “Estimation of the beginning and end of recurrent events within a climate regime”. In: *Clim. Res.* 11.2, pp. 97–107.

Ghasemizade, M., Baroni, G., Abbaspour, K., and Schirmer, M. (2017). “Combined analysis of time-varying sensitivity and identifiability indices to diagnose the response of a complex environmental model”. In: *Environ. Modell. Softw.* 88, pp. 22–34. DOI: 10.1016/j.envsoft.2016.10.011.

González, V. I., Carkovic, A. B., Lobo, G. P., Flanagan, D. C., and Bonilla, C. A. (2016). “Spatial discretization of large watersheds and its influence on the estimation of hillslope sediment yield”. In: *Hydrol. Process.* 30.1, pp. 30–39. DOI: 10.1002/hyp.10559.

Graeff, T., Zehe, E., Reusser, D., Lück, E., Schröder, B., Wenk, G., John, H., and Bronstert, A. (2009). "Process identification through rejection of model structures in a mid-mountainous rural catchment: observations of rainfall–runoff response, geophysical conditions and model inter-comparison". In: *Hydrol. Process.* 23.5, pp. 702–718. DOI: 10.1002/hyp.7171.

Grayson, R. B., Moore, I. D., and McMahon, T. A. (1992a). "Physically based hydrologic modeling: 1. A terrain-based model for investigative purposes". In: *Water Resour. Res.* 28.10, pp. 2639–2658. DOI: 10.1029/92WR01258.

Grayson, R. B., Moore, I. D., and McMahon, T. A. (1992b). "Physically based hydrologic modeling: 2. Is the concept realistic?" In: *Water Resour. Res.* 28.10, pp. 2659–2666. DOI: 10.1029/92WR01259.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., and Engen-Skaugen, T. (2012). "Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods". In: *Hydrol. Earth Syst. Sc.* 16.9, pp. 3383–3390. DOI: 10.5194/hess-16-3383-2012.

Guillaume, J. H., Jakeman, J. D., Marsili-Libelli, S., Asher, M., Brunner, P., Croke, B., Hill, M. C., Jakeman, A. J., Keesman, K. J., Razavi, S., and Stigter, J. D. (2019). "Introductory overview of identifiability analysis: A guide to evaluating whether you have the right type of data for your modeling purpose". In: *Environ. Modell. Softw.* 119, pp. 418–432. DOI: 10.1016/j.envsoft.2019.07.007.

Günther, D., Marke, T., Essery, R., and Strasser, U. (2019). "Uncertainties in Snowpack Simulations—Assessing the Impact of Model Structure, Parameter Choice, and Forcing Data Error on Point-Scale Energy Balance Snow Model Performance". In: *Water Resour. Res.* 55.4, pp. 2779–2800. DOI: https://doi.org/10.1029/2018WR023403.

Güntner, A. (2002). *Large-scale hydrological modelling in the semi-arid North-East of Brazil*. PIK Report 77. Potsdam, Germany: Potsdam Institute for Climate Impact Research.

Güntner, A. and Bronstert, A. (2004). "Representation of landscape variability and lateral redistribution processes for large-scale hydrological modelling in semi-arid areas". In: *J. Hydrol.* 297.1–4, pp. 136–161. DOI: 10.1016/j.jhydrol.2004.04.008.

Güntner, A., Krol, M. S., De Araújo, J. C., and Bronstert, A. (2004). "Simple water balance modelling of surface reservoir systems in a large data-scarce semiarid region". In: *Hydrolog. Sci. J.* 49.5, pp. 901–918. DOI: 10.1623/hysj.49.5.901.55139.

Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G., and Ye, M. (2012). "Towards a comprehensive assessment of model structural adequacy". In: *Water Resour. Res.* 48.8, W08301. DOI: 10.1029/2011WR011044.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F. (2009). "Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling". In: *J. Hydrol.* 377.1–2, pp. 80–91. DOI: 10.1016/j.jhydrol.2009.08.003.

Guse, B., Pfannerstill, M., Gafurov, A., Kiesel, J., Lehr, C., and Fohrer, N. (2017). "Identifying the connective strength between model parameters and performance criteria". In: *Hydrol. Earth Syst. Sc.* 21.11, pp. 5663–5679. DOI: 10.5194/hess-21-5663-2017.

Guse, B., Pilz, T., Stoelzle, M., and Bormann, H. (2019). "Charakterisierung und Analyse hydrologischer Modelle im deutschsprachigen Raum". In: *Wasser und Abfall* 5.2019, pp. 43–52.

Guse, B., Reusser, D. E., and Fohrer, N. (2014). "How to improve the representation of hydrological processes in SWAT for a lowland catchment – temporal analysis of parameter sensitivity and model performance". In: *Hydrol. Process.* 28.4, pp. 2651–2670. DOI: 10.1002/hyp.9777.

Haghnegahdar, A., Tolson, B. A., Craig, J. R., and Paya, K. T. (2015). "Assessing the performance of a semi-distributed hydrological model under various watershed discretization schemes". In: *Hydrol. Process.* 29.18, pp. 4018–4031. DOI: 10.1002/hyp.10550.

Han, J.-C., Huang, G.-H., Zhang, H., Li, Z., and Li, Y.-P. (2014). "Effects of watershed subdivision level on semi-distributed hydrological simulations: case study of the SLURP model applied to the Xiangxi River watershed, China". In: *Hydrolog. Sci. J.* 59.1, pp. 108–125. DOI: 10.1080/02626667.2013.854368.

Hastenrath, S. (2012). "Exploring the climate problems of Brazil's Nordeste: a review". In: *Climatic Change* 112.2, pp. 243–251. DOI: 10.1007/s10584-011-0227-1.

Haude, W. (1958). “Über die Verwendung verschiedener Klimafaktoren zur Berechnung der potentiellen Evaporation und Evapotranspiration”. In: *Meteorol. Rundschau* 11.3, pp. 96–99.

Haverkamp, S., Fohrer, N., and Frede, H.-G. (2005). “Assessment of the effect of land use patterns on hydrologic landscape functions: a comprehensive GIS-based tool to minimize model uncertainty resulting from spatial aggregation”. In: *Hydrol. Process.* 19.3, pp. 715–727. DOI: 10.1002/hyp.5626.

Hazenberg, P., Fang, Y., Broxton, P., Gochis, D., Niu, G.-Y., Pelletier, J. D., Troch, P. A., and Zeng, X. (2015). “A hybrid-3D hillslope hydrological model for use in Earth system models”. In: *Water Resour. Res.* 51.10, pp. 8218–8239. DOI: 10.1002/2014WR016842.

Heine, I., Francke, T., Rogass, C., Medeiros, P. H. A., Bronstert, A., and Foerster, S. (2014). “Monitoring Seasonal Changes in the Water Surface Areas of Reservoirs Using TerraSAR-X Time Series Data in Semiarid Northeastern Brazil”. In: *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 7.8, pp. 3190–3199. DOI: 10.1109/JSTARS.2014.2323819.

Her, Y. and Chaubey, I. (2015). “Impact of the numbers of observations and calibration parameters on equifinality, model performance, and output and parameter uncertainty”. In: *Hydrol. Process.* 29.19, pp. 4220–4237. DOI: 10.1002/hyp.10487.

Herman, J. D., Reed, P. M., and Wagener, T. (2013). “Time-varying sensitivity analysis clarifies the effects of watershed model formulation on model behavior”. In: *Water Resour. Res.* 49.3, pp. 1400–1414. DOI: 10.1002/wrcr.20124.

Herr, H. D. and Krzysztofowicz, R. (2015). “Ensemble Bayesian forecasting system Part I: Theory and algorithms”. In: *J. Hydrol.* 524, pp. 789–802. DOI: 10.1016/j.jhydrol.2014.11.072.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). “Bayesian model averaging: a tutorial”. In: *Stat. Sci.* 14.4, pp. 382–401.

Höge, M., Wöhling, T., and Nowak, W. (2018). “A Primer for Model Selection: The Decisive Role of Model Complexity”. In: *Water Resour. Res.* 54.3, pp. 1688–1715. DOI: 10.1002/2017WR021902.

Holländer, H. M., Blume, T., Bormann, H., Buytaert, W., Chirico, G. B., Exbrayat, J.-F., Gustafsson, D., Hölzel, H., Kraft, P., Stamm, C., Stoll, S., Blöschl, G., and Flühler, H. (2009). “Comparative predictions of discharge from an artificial catchment (Chicken Creek) using sparse data”. In: *Hydrol. Earth Syst. Sc.* 13.11, pp. 2069–2094. DOI: 10.5194/hess-13-2069-2009.

Holländer, H. M., Bormann, H., Blume, T., Buytaert, W., Chirico, G. B., Exbrayat, J.-F., Gustafsson, D., Hölzel, H., Krauße, T., Kraft, P., Stoll, S., Blöschl, G., and Flühler, H. (2014). “Impact of modellers’ decisions on hydrological a priori predictions”. In: *Hydrol. Earth Syst. Sc.* 18.6, pp. 2065–2085. DOI: 10.5194/hess-18-2065-2014.

Hothorn, T., Buehlmann, P., Dudoit, S., Molinaro, A., and Van Der Laan, M. (2006). “Survival Ensembles”. In: *Biostatistics* 7.3, pp. 355–373. DOI: 10.1093/biostatistics/kxj011.

Hrachowitz, M. and Clark, M. P. (2017). “HESS Opinions: The complementary merits of competing modelling philosophies in hydrology”. In: *Hydrol. Earth Syst. Sc.* 21.8, pp. 3953–3973. DOI: 10.5194/hess-21-3953-2017.

IPECE (2016). *Anuário Estatístico do Ceará*. pblished online: http://www.ipece.ce.gov.br/index.php/anuario-estatistico-do-ceara.

Ivanov, V. Y., Vivoni, E. R., Bras, R. L., and Entekhabi, D. (2004). “Catchment hydrologic response with a fully distributed triangulated irregular network model”. In: *Water Resour. Res.* 40.11. W11102, W11102. DOI: 10.1029/2004WR003218.

Jackisch, C., Zehe, E., Samaniego, L., and Singh, A. K. (2014). “An experiment to gauge an ungauged catchment: rapid data assessment and eco-hydrological modelling in a data-scarce rural catchment”. In: *Hydrolog. Sci. J.* 59.12, pp. 2103–2125. DOI: 10.1080/02626667.2013.870662.

Jacomine, P. K. T., Almeida, J. C., and Medeiros, L. A. R. (1973). *Levantamento exploratorio – Reconhecimento de solos do Estado do Ceará*. Vol. 1. Recife, Brazil: DNPEA, DRN-SUDENE.

Jakeman, A. J. and Hornberger, G. M. (1993). “How much complexity is warranted in a rainfall-runoff model?” In: *Water Resour. Res.* 29.8, pp. 2637–2649. DOI: 10.1029/93WR00877.

Jakeman, A., Letcher, R., and Norton, J. (2006). “Ten iterative steps in development and evaluation of environmental models”. In: *Environ. Modell. Softw.* 21.5, pp. 602–614. DOI: 10.1016/j.envsoft.2006.01.004.

Jiménez Cisneros, B. E., Oki, T., Arnell, N. W., Benito, G., Cogley, J. G., Döll, P., Jiang, T., and Mwakalila, S. S. (2014). “Freshwater resources”. In: *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by Field, C., Barros, V., Dokken, D., Mach, K., Mastrandrea, M., Bilir, T., Chatterjee, M., Ebi, K., Estrada, Y., Genova, R., Girma, B., Kissel, E., Levy, A., MacCracken, S., Mastrandrea, P., and L.L.White. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 229–269.

Kalman, R. E. (1960). “A New Approach to Linear Filtering and Prediction Problems”. In: *J. Basic Engineering* 82.1, pp. 35–45. DOI: 10.1115/1.3662552.

Kavetski, D. and Clark, M. P. (2010). “Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction”. In: *Water Resour. Res.* 46.10, W10511. DOI: 10.1029/2009WR008896.

Kavetski, D. and Clark, M. P. (2011). “Numerical troubles in conceptual hydrology: Approximations, absurdities and impact on hypothesis testing”. In: *Hydrol. Process.* 25.4, pp. 661–670. DOI: 10.1002/hyp.7899.

Kavetski, D. and Fenicia, F. (2011). “Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights”. In: *Water Resour. Res.* 47.11, W11511. DOI: http://dx.doi.org/10.1029/2011WR010748.

Kavetski, D., Kuczera, G., and Franks, S. W. (2006). “Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory”. In: *Water Resour. Res.* 42.3, W03407. DOI: 10.1029/2005wr004368.

Kavetski, D., Qin, Y., and Kuczera, G. (2018). “The fast and the robust: Trade-offs between optimization robustness and cost in the calibration of environmental models”. In: *Water Resour. Res.* 54.11, pp. 9432–9455. DOI: 10.1029/2017WR022051.

Kelly, K. S. and Krzysztofowicz, R. (2000). “Precipitation uncertainty processor for probabilistic river stage forecasting”. In: *Water Resour. Res.* 36.9, pp. 2643–2653. DOI: 10.1029/2000WR900061.

Khan, U., Tuteja, N. K., Ajami, H., and Sharma, A. (2014). “An equivalent cross-sectional basis for semidistributed hydrological modeling”. In: *Water Resour. Res.* 50.5, pp. 4395–4415. DOI: 10.1002/2013WR014741.

Kiang, J. E., Gazoorian, C., McMillan, H., Coxon, G., Le Coz, J., Westerberg, I. K., Belleville, A., Sevrez, D., Sikorska, A. E., Petersen-Øverleir, A., Reitan, T., Freer, J., Renard, B., Mansanarez, V., and Mason, R. (2018). “A Comparison of Methods for Streamflow Uncertainty Estimation”. In: *Water Resour. Res.* 54.10, pp. 7149–7176. DOI: 10.1029/2018WR022708.

Kinner, D., Mitasova, H., Stallard, R., Harmon, R. S., and Toma, L. (2005). “GIS-Based Stream Network Analysis for the Upper Río Chagres Basin, Panama”. In: *The Río Chagres, Panama: A Multidisciplinary Profile of a Tropical Watershed*. Ed. by Harmon, R. S. Springer Netherlands. Chap. 6, pp. 83–95. DOI: 10.1007/1-4020-3297-8_6.

Kirchner, J. W. (2006). “Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology”. In: *Water Resour. Res.* 42.3, W03S04. DOI: 10.1029/2005WR004362.

Kite, G. W. (1995). “Scaling of Input Data for Macroscale Hydrologic Modeling”. In: *Water Resour. Res.* 31.11, pp. 2769–2781. DOI: 10.1029/95WR02102.

Kneis, D., Bürger, G., and Bronstert, A. (2012). “Evaluation of medium-range runoff forecasts for a 50km2 watershed”. In: *J. Hydrol.* 414–415, pp. 341–353. DOI: 10.1016/j.jhydrol.2011.11.005.

Kneis, D., Chatterjee, C., and Singh, R. (2014). “Evaluation of TRMM rainfall estimates over a large Indian river basin (Mahanadi)”. In: *Hydrol. Earth Syst. Sc.* 18.7, pp. 2493–2502. DOI: 10.5194/hessd-11-1169-2014.

Kneis, D. (2012). *geostat: Utilities for spatial interpolation*. R package version 0.1. https://github.com/echse/echse_tools/tree/master/R/packages/geostat.

Kneis, D. (2015). “A lightweight framework for rapid development of object-based hydrological model engines”. In: *Environ. Modell. Softw.* 68, pp. 110–121. DOI: 10.1016/j.envsoft.2015.02.009.

Kneis, D., Abon, C., Bronstert, A., and Heistermann, M. (2017). “Verification of short-term runoff forecasts for a small Philippine basin (Marikina)”. In: *Hydrolog. Sci. J.* 62.2, pp. 205–216. DOI: 10.1080/02626667.2016.1183773.

Kokkonen, T. S. and Jakeman, A. J. (2001). “A comparison of metric and conceptual approaches in rainfall-runoff modeling and its implications”. In: *Water Resour. Res.* 37.9, pp. 2345–2352. DOI: 10.1029/2001WR000299.

Komma, J., Blöschl, G., and Reszler, C. (2008). “Soil moisture updating by Ensemble Kalman Filtering in real-time flood forecasting”. In: *J. Hydrol.* 357.3–4, pp. 228–242. DOI: 10.1016/j.jhydrol.2008.05.020.

Kouwen, N. (2016). *WATFLOOD / CHARM Canadian Hydrological And Routing Model*. Department of Civil Engineering, University of Waterloo, Waterloo, Ontario, Canada.

Kouwen, N., Soulis, E. D., Pietroniro, A., Donald, J., and Harrington, R. A. (1993). “Grouped Response Units for Distributed Hydrologic Modeling”. In: *J. Water Res. Pl.-ASCE* 119.3, pp. 289–305. DOI: 10.1061/(ASCE)0733-9496(1993)119:3(289).

Krol, M. S., de Vries, M. J., Oel, P. R., and de Araújo, J. C. (2011). “Sustainability of Small Reservoirs and Large Scale Water Availability Under Current Conditions and Climate Change”. In: *Water Resour. Manage.* 25.12, pp. 3017–3026. DOI: 10.1007/s11269-011-9787-0.

Krol, M., Jaeger, A., Bronstert, A., and Güntner, A. (2006). “Integrated modelling of climate, water, soil, agricultural and socio-economic processes: A general introduction of the methodology and some exemplary results from the semi-arid north-east of Brazil”. In: *J. Hydrol.* 328.3–4, pp. 417–431. DOI: 10.1016/j.jhydrol.2005.12.021.

Krysanova, V., Müller-Wohlfeil, D.-I., and Becker, A. (1998). “Development and test of a spatially distributed hydrological/water quality model for mesoscale watersheds”. In: *Ecol. Model.* 106.2–3, pp. 261–289. DOI: 10.1016/S0304-3800(97)00204-4.

Krzysztofowicz, R. (1999). “Bayesian theory of probabilistic forecasting via deterministic hydrologic model”. In: *Water Resour. Res.* 35.9, pp. 2739–2750. DOI: 10.1029/1999wr900099.

Krzysztofowicz, R. and Kelly, K. S. (2000). “Hydrologic uncertainty processor for probabilistic river stage forecasting”. In: *Water Resour. Res.* 36.11, pp. 3265–3277. DOI: 10.1029/2000wr900108.

Kumar, R., Samaniego, L., and Attinger, S. (2010). “The effects of spatial discretization and model parameterization on the prediction of extreme runoff characteristics”. In: *J. Hydrol.* 392.1–2, pp. 54–69. DOI: 10.1016/j.jhydrol.2010.07.047.

Lacroix, M. P., Martz, L. W., Kite, G. W., and Garbrecht, J. (2002). “Using digital terrain analysis modeling techniques for the parameterization of a hydrologic model”. In: *Environ. Modell. Softw.* 17.2, pp. 125–134. DOI: 10.1016/S1364-8152(01)00042-1.

Lagacherie, P., Rabotin, M., Colin, F., Moussa, R., and Voltz, M. (2010). “Geo-MHYDAS: A landscape discretization tool for distributed hydrological modeling of cultivated areas”. In: *Comput. Geosci.* 36.8, pp. 1021–1032. DOI: 10.1016/j.cageo.2009.12.005.

Leavesley, G. H., Lichty, R. W., Troutman, B. M., and Saindon, L. G. (1983). *Precipitation-Runoff Modeling System: User’s manual*. Water-Resources Investigations Report 83-4238. United States Department of the Interior, Denver, Colorado, USA.

Liu, Y. and Gupta, H. V. (2007). “Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework”. In: *Water Resour. Res.* 43.7, W07401. DOI: 10.1029/2006wr005756.

Maidment, D. R., ed. (1993). *Handbook of Hydrology*. McGraw-Hill, Inc., New York, USA.

Mamede, G. L., Araújo, N. A. M., Schneider, C. M., de Araújo, J. C., and Herrmann, H. J. (2012). “Overspill avalanching in a dense reservoir network”. In: *Proc. Natl. Acad. Sci. U.S.A.* 109.19, pp. 7191–7195. DOI: 10.1073/pnas.1200398109.

Manguerra, H. B. and Engel, B. A. (1998). “Hydrologic parameterization of watersheds for runoff prediction using SWAT”. In: *J. Am. Water Resour. As.* 34.5, pp. 1149–1162. DOI: 10.1111/j.1752-1688.1998.tb04161.x.

Marengo, J. A., Torres, R. R., and Alves, L. M. (2017). “Drought in Northeast Brazil—past, present, and future”. In: *Theor. Appl. Climatol.* 129.3, pp. 1189–1200. DOI: 10.1007/s00704-016-1840-8.

Markstrom, S., Niswonger, R., Regan, R., Prudic, D., and Barlow, P. (2008). *GSFLOW–Coupled ground-water and surface-water flow model based on the integration of the Precipitation-Runoff Modeling System (PRMS) and the Modular Ground-Water Flow Model (MODFLOW-2005)*. U.S. Geological Survey Techniques and Methods 6-D1, 240 p.

Massmann, C., Wagener, T., and Holzmann, H. (2014). "A new approach to visualizing time-varying sensitivity indices for environmental model diagnostics across evaluation time-scales". In: *Environ. Modell. Softw.* 51, pp. 190–194. DOI: 10.1016/j.envsoft.2013.09.033.

Maunder, C. J. (1999). "An automated method for constructing contour-based digital elevation models". In: *Water Resour. Res.* 35.12, pp. 3931–3940. DOI: 10.1029/1999WR900166.

Maurer, T. (1997). "Physikalisch begründete, zeitkontinuierliche Modellierung des Wassertransports in kleinen ländlichen Einzugsgebieten". Dissertation, Mitteilungen Inst. f. Hydrologie u. Wasserwirtschaft, 61. Universität Karlsruhe.

McMillan, H., Hreinsson, E. O., Clark, M. P., Singh, S. K., Zammit, C., and Uddstrom, M. J. (2013). "Operational hydrological data assimilation with the recursive ensemble Kalman filter". In: *Hydrol. Earth Syst. Sc.* 17.1, pp. 21–38. DOI: 10.5194/hess-17-21-2013.

McMillan, H., Krueger, T., and Freer, J. (2012). "Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality". In: *Hydrol. Process.* 26.26, pp. 4078–4111. DOI: 10.1002/hyp.9384.

McMillan, H., Seibert, J., Petersen-Overleir, A., Lang, M., White, P., Snelder, T., Rutherford, K., Krueger, T., Mason, R., and Kiang, J. (2017). "How uncertainty analysis of streamflow data can reduce costs and promote robust decisions in water management applications". In: *Water Resour. Res.* 53.7, pp. 5220–5228. DOI: 10.1002/2016WR020328.

McMillan, H. and Westerberg, I. K. (2015). "Rating curve estimation under epistemic uncertainty". In: *Hydrol. Process.* 29.7, pp. 1873–1882. DOI: 10.1002/hyp.10419.

Medeiros, P. H. A. and de Araújo, J. C. (2014). "Temporal variability of rainfall in a semiarid environment in Brazil and its effect on sediment transport processes". In: *J. Soils Sediments* 14.7, pp. 1216–1223. DOI: 10.1007/s11368-013-0809-9.

Medeiros, P. H. A., de Araújo, J. C., Mamede, G. L., Creutzfeldt, B., Güntner, A., and Bronstert, A. (2014). "Connectivity of sediment transport in a semiarid environment: a synthesis for the Upper Jaguaribe Basin, Brazil". In: *J. Soils Sediments* 14.12, pp. 1938–1948. DOI: 10.1007/s11368-014-0988-z.

Medeiros, P. H. A., Güntner, A., Francke, T., Mamede, G. L., and de Araújo, J. C. (2010). "Modelling spatio-temporal patterns of sediment yield and connectivity in a semi-arid catchment with the WASA-SED model". In: *Hydrolog. Sci. J.* 55.4, pp. 636–648. DOI: 10.1080/02626661003780409.

Merz, R. and Blöschl, G. (2004). "Regionalisation of catchment model parameters". In: *J. Hydrol.* 287.1, pp. 95–123. DOI: 10.1016/j.jhydrol.2003.09.028.

Metz, M., Mitasova, H., and Harmon, R. S. (2011). "Efficient extraction of drainage networks from massive, radar-based elevation models with least cost path search". In: *Hydrol. Earth Syst. Sc.* 15.2, pp. 667–678. DOI: 10.5194/hess-15-667-2011.

Miller, S. N., Semmens, D. J., Goodrich, D. C., Hernandez, M., Miller, R. C., Kepner, W. G., and Guertin, D. P. (2007). "The Automated Geospatial Watershed Assessment tool". In: *Environ. Modell. Softw.* 22.3, pp. 365–377. DOI: 10.1016/j.envsoft.2005.12.004.

Miller, S. N., Semmens, D., Miller, R., Hernandez, M., Goodrich, D., Miller, W., Kepner, W., and Ebert, D. (2002). "GIS-based Hydrologic Modeling: The Automated Geospatial Watershed Assessment Tool". In: *Second Federal Interagency Hydrologic Modeling Conference*. Las Vegas, Nevada, USA.

Mishra, A. K. and Singh, V. P. (2010). "A review of drought concepts". In: *J. Hydrol.* 391.1–2, pp. 202–216. DOI: 10.1016/j.jhydrol.2010.07.012.

Molnár, D. K. and Julien, P. Y. (2000). "Grid-Size Effects on Surface Runoff Modeling". In: *J. Hydrol. Eng.* 5.1, pp. 8–16. DOI: 10.1061/(ASCE)1084-0699(2000)5:1(8).

Moore, I. D., Grayson, R. B., and Ladson, A. R. (1991). "Digital terrain modelling: A review of hydrological, geomorphological, and biological applications". In: *Hydrol. Process.* 5.1, pp. 3–30. DOI: 10.1002/hyp.3360050103.

Moore, I. D. and Grayson, R. B. (1991). “Terrain-based catchment partitioning and runoff prediction using vector elevation data”. In: *Water Resour. Res.* 27.6, pp. 1177–1191. DOI: 10.1029/91WR00090.

Moore, I. D., O’Loughlin, E. M., and Burch, G. J. (1988). “A contour-based topographic model for hydrological and ecological applications”. In: *Earth Surf. Proc. Land.* 13.4, pp. 305–320. DOI: 10.1002/esp.3290130404.

Moradkhani, H., Hsu, K.-L., Gupta, H., and Sorooshian, S. (2005). “Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter”. In: *Water Resour. Res.* 41.5, W05012. DOI: 10.1029/2004WR003604.

Moretti, G. and Orlandini, S. (2008). “Automatic delineation of drainage basins from contour elevation data using skeleton construction techniques”. In: *Water Resour. Res.* 44.5, W05403. DOI: 10.1029/2007WR006309.

Moriasi, D. N., Arnold, J. G., Liew, M. W. V., Bingner, R. L., Harmel, R. D., and Veith, T. L. (2007). “Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations”. In: *T. ASABE* 50.3, pp. 885–900. DOI: 10.13031/2013.23153.

Moussa, R., Voltz, M., and Andrieux, P. (2002). “Effects of the spatial organization of agricultural management on the hydrological behaviour of a farmed catchment during flood events”. In: *Hydrol. Process.* 16.2, pp. 393–412. DOI: 10.1002/hyp.333.

Mueller, E. N., Francke, T., Batalla, R. J., and Bronstert, A. (2009). “Modelling the effects of land-use change on runoff and sediment yield for a meso-scale catchment in the Southern Pyrenees”. In: *CATENA* 79.3, pp. 288–296. DOI: 10.1016/j.catena.2009.06.007.

Mueller, E. N., Güntner, A., Francke, T., and Mamede, G. (2010). “Modelling sediment export, retention and reservoir sedimentation in drylands with the WASA-SED model”. In: *Geosci. Model Dev.* 3.1, pp. 275–291. DOI: 10.5194/gmd-3-275-2010.

Murphy, A. H. (1993). “What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting”. In: *Weather Forecast.* 8.2, pp. 281–293. DOI: 10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2.

Neteler, M., Bowman, M. H., Landa, M., and Metz, M. (2012). “GRASS GIS: A multi-purpose open source GIS”. In: *Environ. Modell. Softw.* 31, pp. 124–130. DOI: 10.1016/j.envsoft.2011.11.014.

Nijzink, R. C., Samaniego, L., Mai, J., Kumar, R., Thober, S., Zink, M., Schäfer, D., Savenije, H. H. G., and Hrachowitz, M. (2016). “The importance of topography-controlled sub-grid process heterogeneity and semi-quantitative prior constraints in distributed hydrological models”. In: *Hydrol. Earth Syst. Sc.* 20.3, pp. 1151–1176. DOI: 10.5194/hess-20-1151-2016.

Noël, P., Rousseau, A. N., Paniconi, C., and Nadeau, D. F. (2014). “Algorithm for Delineating and Extracting Hillslopes and Hillslope Width Functions from Gridded Elevation Data”. In: *J. Hydrol. Eng.* 19.2, pp. 366–374. DOI: 10.1061/(ASCE)HE.1943-5584.0000783.

O’Callaghan, J. F. and Mark, D. M. (1984). “The extraction of drainage networks from digital elevation data”. In: *Comput. Vision Graph.* 28.3, pp. 323–344. DOI: 10.1016/S0734-189X(84)80011-0.

Ocio, D., Le Vine, N., Westerberg, I., Pappenberger, F., and Buytaert, W. (2017). “The Role of Rating Curve Uncertainty in Real-Time Flood Forecasting”. In: *Water Resour. Res.* 53.5, pp. 4197–4213. DOI: 10.1002/2016WR020225.

van Oel, P. R., Krol, M. S., and Hoekstra, A. Y. (2012). “Application of multi-agent simulation to evaluate the influence of reservoir operation strategies on the distribution of water availability in the semi-arid Jaguaribe basin, Brazil”. In: *Phys. Chem. Earth.* 47-48, pp. 173–181. DOI: 10.1016/j.pce.2011.07.051.

van Oel, P. R., Martins, E. S. P. R., Costa, A. C., Wanders, N., and van Lanen, H. A. J. (2018). “Diagnosing drought using the downstreamness concept: the effect of reservoir networks on drought evolution”. In: *Hydrolog. Sci. J.* 63.7, pp. 979–990. DOI: 10.1080/02626667.2018.1470632.

Oxford dictionaries (2018). *Definition of* water *in English*. Oxford University Press, published online, https://en.oxforddictionaries.com/definition/water, accessed 11 October 2018.

Passalacqua, P., Do Trung, T., Foufoula-Georgiou, E., Sapiro, G., and Dietrich, W. E. (2010). "A geometric framework for channel network extraction from lidar: Nonlinear diffusion and geodesic paths". In: *J GEOPHYS RES-EARTH* 115.F1, F01002. DOI: 10.1029/2009JF001254.

Perrin, C., Michel, C., and Andréassian, V. (2003). "Improvement of a parsimonious model for streamflow simulation". In: *J. Hydrol.* 279.1, pp. 275–289. DOI: 10.1016/S0022-1694(03)00225-7.

Philip, J. R. (1957). "The theory of infiltration: 4. Sorptivity and algebraic infiltration equations". In: *Soil Sci.* 84.3, pp. 257–264.

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., and Wagener, T. (2016). "Sensitivity analysis of environmental models: A systematic review with practical workflow". In: *Environ. Modell. Softw.* 79, pp. 214–232. DOI: 10.1016/j.envsoft.2016.02.008.

Pianosi, F. and Wagener, T. (2015). "A simple and efficient method for global sensitivity analysis based on cumulative distribution functions". In: *Environ. Modell. Softw.* 67, pp. 1–11. DOI: 10.1016/j.envsoft.2015.01.004.

Pianosi, F. and Wagener, T. (2016). "Understanding the time-varying importance of different uncertainty sources in hydrological modelling using global sensitivity analysis". In: *Hydrol. Process.* 30.22, pp. 3991–4003. DOI: 10.1002/hyp.10968.

Pilz, T., Delgado, J. M., Voss, S., Vormoor, K., Francke, T., Costa, A. C., Martins, E., and Bronstert, A. (2019a). "Seasonal drought prediction for semiarid northeast Brazil: what is the added value of a process-based hydrological model?" In: *Hydrol. Earth Syst. Sc.* 23.4, pp. 1951–1971. DOI: 10.5194/hess-23-1951-2019.

Pilz, T., Francke, T., and Bronstert, A. (2017). "lumpR 2.0.0: an R package facilitating landscape discretisation for hillslope-based hydrological models". In: *Geosci. Model Dev.* 10.8, pp. 3001–3023. DOI: 10.5194/gmd-10-3001-2017.

Pilz, T., Francke, T., Baroni, G., and Bronstert, A. (2019b). "How to Tailor my Process-based Model? Dynamic Identifiability Analysis of Flexible Model Structures". Submitted to Water Resources Research.

Pinzinger, R., Klein, B., Meißner, D., and Lisniak, D. (2014). "Auswahl Pareto-optimaler Modelle zur statistischen Korrektur von Abflussvorhersagen". In: *Hydrol. Wasserbewirtsch.* 58.2, pp. 128–137.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. accessed 5 July 2016. R Foundation for Statistical Computing. Vienna, Austria.

Refsgaard, J. C. (1997). "Parameterisation, calibration and validation of distributed hydrological models". In: *J. Hydrol.* 198.1–4, pp. 69–97. DOI: 10.1016/S0022-1694(96)03329-X.

Refsgaard, J. C. and Knudsen, J. (1996). "Operational Validation and Intercomparison of Different Types of Hydrological Models". In: *Water Resour. Res.* 32.7, pp. 2189–2202. DOI: 10.1029/96WR00896.

Reggiani, P. and Weerts, A. H. (2008). "Probabilistic Quantitative Precipitation Forecast for Flood Prediction: An Application". In: *J. Hydrometeorol.* 9.1, pp. 76–95. DOI: 10.1175/2007JHM858.1.

Reggiani, P., Sivapalan, M., and Hassanizadeh, S. M. (1998). "A unifying framework for watershed thermodynamics: balance equations for mass, momentum, energy and entropy, and the second law of thermodynamics". In: *Adv. Water Resour.* 22.4, pp. 367–398. DOI: 10.1016/S0309-1708(98)00012-8.

Reichle, R. H. (2008). "Data assimilation methods in the Earth sciences". In: *Adv. Water Resour.* 31.11, pp. 1411–1418. DOI: 10.1016/j.advwatres.2008.01.001.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W. (2010). "Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors". In: *Water Resour. Res.* 46.5, W05521. DOI: 10.1029/2009WR008328.

Renschler, C. S. (2003). "Designing geo-spatial interfaces to scale process models: the GeoWEPP approach". In: *Hydrol. Process.* 17.5, pp. 1005–1017. DOI: 10.1002/hyp.1177.

Reusser, D. E. and Zehe, E. (2011). "Inferring model structural deficits by analyzing temporal dynamics of model performance and parameter sensitivity". In: *Water Resour. Res.* 47.7, W07550. DOI: 10.1029/2010WR009946.

Roeckner, E., Arpe, K., Bengtsson, L., Christoph, M., Claussen, M., Dümenil, L., Esch, M., Giorgetta, M., Schlese, U., and Schulzweida, U. (1996). *The atmospheric general circulation model ECHAM-4: Model description and simulation of present-day climate.* MPI-Report 218. Max-Planck-Institut für Meteorologie, Hamburg, Germany.

Roulin, E. and Vannitsem, S. (2014). "Post-processing of medium-range probabilistic hydrological forecasting: impact of forcing, initial conditions and model errors". In: *Hydrol. Process.* 29.6, pp. 1434–1449. DOI: 10.1002/hyp.10259.

Saltelli, A. and Annoni, P. (2010). "How to avoid a perfunctory sensitivity analysis". In: *Environ. Modell. Softw.* 25.12, pp. 1508–1517. DOI: https://doi.org/10.1016/j.envsoft.2010.04.012.

Samaniego, L., Kumar, R., and Attinger, S. (2010). "Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale". In: *Water Resour. Res.* 46.5, W05523. DOI: 10.1029/2008WR007327.

Sangireddy, H., Stark, C. P., Kladzyk, A., and Passalacqua, P. (2016). "GeoNet: An open source software for the automatic and objective extraction of channel heads, channel network, and channel morphology from high resolution topography data". In: *Environ. Modell. Softw.* 83, pp. 58–73. DOI: 10.1016/j.envsoft.2016.04.026.

Sankarasubramanian, A., Lall, U., Souza Filho, F. A., and Sharma, A. (2009). "Improved water allocation utilizing probabilistic climate forecasts: Short-term water contracts in a risk management framework". In: *Water Resour. Res.* 45.11, W11409. DOI: 10.1029/2009wr007821.

Sanzana, P., Jankowfsky, S., Branger, F., Braud, I., Vargas, X., Hitschfeld, N., and Gironás, J. (2013). "Computer-assisted mesh generation based on hydrological response units for distributed hydrological modeling". In: *Comput. Geosci.* 57, pp. 32–43. DOI: 10.1016/j.cageo.2013.02.006.

Saugier, B. and Katerji, N. (1991). "Some plant factors controlling evapotranspiration". In: *Agric. For. Meteorol.* 54.2, pp. 263–277. DOI: 10.1016/0168-1923(91)90009-F.

Savage, J. T. S., Pianosi, F., Bates, P., Freer, J., and Wagener, T. (2016). "Quantifying the importance of spatial resolution and other factors through global sensitivity analysis of a flood inundation model". In: *Water Resour. Res.* 52.11, pp. 9146–9163. DOI: 10.1002/2015WR018198.

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G. (2011). "Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA". In: *Hydrol. Earth Syst. Sc.* 15.9, pp. 2895–2911. DOI: 10.5194/hess-15-2895-2011.

Schaefli, B. and Gupta, H. V. (2007). "Do Nash values have value?" In: *Hydrol. Process.* 21.15, pp. 2075–2080. DOI: 10.1002/hyp.6825.

Schepen, A. and Wang, Q. (2015). "Model averaging methods to merge operational statistical and dynamic seasonal streamflow forecasts in Australia". In: *Water Resour. Res.* 51.3, pp. 1797–1812. DOI: 10.1002/2014WR016163.

Schöniger, A., Wöhling, T., and Nowak, W. (2015). "A Statistical Concept to Assess the Uncertainty in Bayesian Model Weights and its Impact on Model Ranking". In: *Water Resour. Res.* 51.9, pp. 7524–7546. DOI: 10.1002/2015WR016918.

Schöniger, A., Wöhling, T., Samaniego, L., and Nowak, W. (2014). "Model selection on solid ground: Rigorous comparison of nine ways to evaluate Bayesian model evidence". In: *Water Resour. Res.* 50.12, pp. 9484–9513. DOI: 10.1002/2014WR016062.

Schoups, G., van de Giesen, N. C., and Savenije, H. H. G. (2008). "Model complexity control for hydrologic prediction". In: *Water Resour. Res.* 44.12, W00B03. DOI: 10.1029/2008WR006836.

Schoups, G., Vrugt, J. A., Fenicia, F., and van de Giesen, N. C. (2010). "Corruption of accuracy and efficiency of Markov chain Monte Carlo simulation by inaccurate numerical implementation of conceptual hydrologic models". In: *Water Resour. Res.* 46.10, W10530. DOI: 10.1029/2009WR008648.

Schulla, J. and Jasper, K. (2007). *Model Description WaSiM-ETH.* Technical report 2007, pp. 181. ETH Zürich.

Schwartze, C. (2008). "Deriving Hydrological Response Units (HRUs) using a Web Processing Service implementation based on GRASS GIS". In: *Geoinformatics FCE CTU* 3, pp. 67–78.

Seibert, M., Merz, B., and Apel, H. (2017). “Seasonal forecasting of hydrological drought in the Limpopo Basin: a comparison of statistical methods”. In: *Hydrol. Earth Syst. Sc.* 21.3, pp. 1611–1629. DOI: 10.5194/hess-21-1611-2017.

Seo, D.-J., Koren, V., and Cajina, N. (2003). “Real-Time Variational Assimilation of Hydrologic and Hydrometeorological Data into Operational Hydrologic Forecasting”. In: *J. Hydrometeorol.* 4.3, pp. 627–641. DOI: 10.1175/1525-7541(2003)004<0627:RVAOHA>2.0.CO;2.

Sherman, L. K. (1932). “Streamflow from rainfall by the unit-graph method”. In: *Eng. News-Rec.* 108, pp. 501–505.

Shuttleworth, W. J. (2007). “Putting the "vap" into evaporation”. In: *Hydrol. Earth Syst. Sc.* 11.1, pp. 210–244. DOI: 10.5194/hess-11-210-2007.

Shuttleworth, W. J. and Gurney, R. J. (1990). “The theoretical relationship between foliage temperature and canopy resistance in sparse crops”. In: *Quart. J. Roy. Meteor. Soc.* 116.492, pp. 497–519. DOI: 10.1002/qj.49711649213.

Shuttleworth, W. J. and Wallace, J. S. (1985). “Evaporation from sparse crops-an energy combination theory”. In: *Quart. J. Roy. Meteor. Soc.* 111.469, pp. 839–855. DOI: 10.1002/qj.49711146910.

Sittichok, K., Seidou, O., Djibo, A. G., and Rakangthong, N. K. (2018). “Estimation of the added value of using rainfall–runoff transformation and statistical models for seasonal streamflow forecasting”. In: *Hydrolog. Sci. J.* 63, pp. 630–645. DOI: 10.1080/02626667.2018.1445854.

Smith, T., Marshall, L., McGlynn, B., and Jencso, K. (2013). “Using field data to inform and evaluate a new model of catchment hydrologic connectivity”. In: *Water Resour. Res.* 49.10, pp. 6834–6846. DOI: 10.1002/wrcr.20546.

Souza Filho, F. A. and Lall, U. (2003). “Seasonal to interannual ensemble streamflow forecasts for Ceara, Brazil: Applications of a multivariate, semiparametric algorithm”. In: *Water Resour. Res.* 39.11, p. 1307. DOI: 10.1029/2002wr001373.

Spear, R. and Hornberger, G. (1980). “Eutrophication in peel inlet—II. Identification of critical uncertainties via generalized sensitivity analysis”. In: *Water Res.* 14.1, pp. 43–49. DOI: 10.1016/0043-1354(80)90040-8.

Stahn, P., Busch, S., Salzmann, T., Eichler-Löbermann, B., and Miegel, K. (2017). “Combining global sensitivity analysis and multiobjective optimisation to estimate soil hydraulic properties and representations of various sole and mixed crops for the agro-hydrological SWAP model”. In: *Environmental Earth Sciences* 76.10, p. 367. DOI: https://doi.org/10.1007/s12665-017-6701-y.

Stephenson, G. R. and Freeze, R. A. (1974). “Mathematical simulation of subsurface flow contributions to snowmelt runoff, Reynolds Creek Watershed, Idaho”. In: *Water Resour. Res.* 10.2, pp. 284–294. DOI: 10.1029/WR010i002p00284.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., and Zeileis, A. (2008). “Conditional Variable Importance for Random Forests”. In: *BMC Bioinf.* 9.307. DOI: 10.1186/1471-2105-9-307.

Strobl, C., Boulesteix, A.-L., Zeileis, A., and Hothorn, T. (2007). “Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution”. In: *BMC Bioinf.* 8.25. DOI: 10.1186/1471-2105-8-25.

Sulis, M., Paniconi, C., and Camporese, M. (2011). “Impact of grid resolution on the integrated and distributed response of a coupled surface–subsurface hydrological model for the des Anglais catchment, Quebec”. In: *Hydrol. Process.* 25.12, pp. 1853–1865. DOI: 10.1002/hyp.7941.

Sun, L., Seidou, O., Nistor, I., and Liu, K. (2016). “Review of the Kalman-type hydrological data assimilation”. In: *Hydrolog. Sci. J.* 61.13, pp. 2348–2366. DOI: 10.1080/02626667.2015.1127376.

Sun, L., Moncunill, D. F., Li, H., Moura, A. D., Filho, F. D. A. D. S., and Zebiak, S. E. (2006). “An Operational Dynamical Downscaling Prediction System for Nordeste Brazil and the 2002–04 Real-Time Forecast Evaluation”. In: *J. Climate* 19.10, pp. 1990–2007. DOI: 10.1175/jcli3715.1.

Tague, C. and Band, L. (2004). “RHESSys: Regional Hydro-ecologic simulation system: An object-oriented approach to spatially distributed modeling of carbon, water and nutrient cycling”. In: *Earth Interact.* 8.19, pp. 1–42.

Tarboton, D. (2003). "Terrain Analysis Using Digital Elevation Models in Hydrology". In: *23rd ESRI International Users Conference, San Diego, California, USA*.

Tarboton, D., Bras, R. L., and Rodriguez-Iturbe, I. (1991). "On the extraction of channel networks from digital elevation data". In: *Hydrol. Process.* 5.1, pp. 81–100. DOI: 10.1002/hyp.3360050107.

Thiboult, A. and Anctil, F. (2015). "On the difficulty to optimally implement the Ensemble Kalman filter: An experiment based on many hydrological models and catchments". In: *J. Hydrol.* 529.3, pp. 1147–1160. DOI: 10.1016/j.jhydrol.2015.09.036.

Tian, S., Tregoning, P., Renzullo, L. J., van Dijk, A. I., Walker, J. P., Pauwels, V. R., and Allgeyer, S. (2017). "Improved water balance component estimates through Joint Assimilation of GRACE water storage and SMOS soil moisture retrievals". In: *Water Resour. Res.* 53.3, pp. 1820–1840. DOI: 10.1002/2016WR019641.

Todini, E. (2008). "A model conditional processor to assess predictive uncertainty in flood forecasting". In: *Int. J. River Basin Management* 6.2, pp. 123–137. DOI: 10.1080/15715124.2008.9635342.

Tolson, B. A. and Shoemaker, C. A. (2007). "Dynamically dimensioned search algorithm for computationally efficient watershed model calibration". In: *Water Resour. Res.* 43.1, W01413. DOI: 10.1029/2005wr004723.

Troch, P. A., Paniconi, C., and Emiel van Loon, E. (2003). "Hillslope-storage Boussinesq model for subsurface flow and variable source areas along complex hillslopes: 1. Formulation and characteristic response". In: *Water Resour. Res.* 39.11, p. 1316. DOI: 10.1029/2002WR001728.

Tucker, G. E., Lancaster, S. T., Gasparini, N. M., Bras, R. L., and Rybarczyk, S. M. (2001). "An object-oriented framework for distributed hydrologic and geomorphic modeling using triangulated irregular networks". In: *Comput. Geosci.* 27.8, pp. 959–973. DOI: 10.1016/S0098-3004(00)00134-5.

Uhlenbrook, S. and Leibundgut, C. (2002). "Process-oriented catchment modelling and multiple-response validation". In: *Hydrol. Process.* 16.2, pp. 423–440. DOI: 10.1002/hyp.330.

Verkade, J., Brown, J., Reggiani, P., and Weerts, A. (2013). "Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales". In: *J. Hydrol.* 501, pp. 73–91. DOI: 10.1016/j.jhydrol.2013.07.039.

Viviroli, D., Zappa, M., Gurtz, J., and Weingartner, R. (2009). "An introduction to the hydrological modelling system PREVAH and its pre- and post-processing-tools". In: *Environ. Modell. Softw.* 24.10, pp. 1209–1222. DOI: 10.1016/j.envsoft.2009.04.001.

Vivoni, E. R., Ivanov, V. Y., Bras, R. L., and Entekhabi, D. (2004). "Generation of Triangulated Irregular Networks Based on Hydrological Similarity". In: *J. Hydrol. Eng.* 9.4, pp. 288–302. DOI: 10.1061/(ASCE)1084-0699(2004)9:4(288).

Volpi, E., Schoups, G., Firmani, G., and Vrugt, J. A. (2017). "Sworn testimony of the model evidence: Gaussian Mixture Importance (GAME) sampling". In: *Water Resour. Res.* 53.7, pp. 6133–6158. DOI: 10.1002/2016WR020167.

Vrugt, J. A., ter Braak, C. J., Diks, C. G., and Schoups, G. (2013). "Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications". In: *Adv. Water Resour.* 51, pp. 457–478. DOI: 10.1016/j.advwatres.2012.04.002.

Vrugt, J. A., Gupta, H. V., Bouten, W., and Sorooshian, S. (2003). "A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters". In: *Water Resour. Res.* 39.8, p. 1201. DOI: 10.1029/2002WR001642.

Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V., and Sorooshian, S. (2001). "A framework for development and application of hydrological models". In: *Hydrol. Earth Syst. Sc.* 5.1, pp. 13–26. DOI: 10.5194/hess-5-13-2001.

Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S., and Gupta, H. V. (2003). "Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis". In: *Hydrol. Process.* 17.2, pp. 455–476. DOI: 10.1002/hyp.1135.

Wagener, T. and Gupta, H. V. (2005). "Model identification for hydrological forecasting under uncertainty". In: *Stoch. Env. Res. Risk A.* 19.6, pp. 378–387. DOI: 10.1007/s00477-005-0006-5.

Walker, A. P., Ye, M., Lu, D., De Kauwe, M. G., Gu, L., Medlyn, B. E., Rogers, A., and Serbin, S. P. (2018). "The multi-assumption architecture and testbed (MAAT v1.0): R code for generating ensembles with dynamic model structure and analysis of epistemic uncertainty from multiple sources". In: *Geosci. Model Dev.* 11.8, pp. 3159–3185. DOI: 10.5194/gmd-11-3159-2018.

Weiler, M. and Beven, K. (2015). "Do we need a Community Hydrological Model?" In: *Water Resour. Res.* 51.9, pp. 7777–7784. DOI: 10.1002/2014WR016731.

van Werkhoven, K., Wagener, T., Reed, P., and Tang, Y. (2008). "Characterization of watershed model behavior across a hydroclimatic gradient". In: *Water Resour. Res.* 44.1, W01429. DOI: 10.1029/2007WR006271.

Westerberg, I. K. and Birkel, C. (2015). "Observational uncertainties in hypothesis testing: investigating the hydrological functioning of a tropical catchment". In: *Hydrol. Process.* 29.23, pp. 4863–4879. DOI: 10.1002/hyp.10533.

Whittaker, G., Confesor, Jr., R., Di Luzio, M., and Arnold, J. G. (2010). "Detection of overparameterization and overfitting in an automatic calibration of SWAT". In: *T. ASABE* 53.5, pp. 1487–1499.

Wilks, D. S. (2005). *Statistical Methods in the Atmospheric Sciences, Volume 100, Second Edition (International Geophysics)*. Academic Press.

Winter, T. C. (2001). "The concept of hydrologic landscapes". In: *J. Am. Water Resour. As.* 37.2, pp. 335–349. DOI: 10.1111/j.1752-1688.2001.tb00973.x.

Wood, E. F. (1995). "Scaling behaviour of hydrological fluxes and variables: Empirical studies using a hydrological model and remote sensing data". In: *Hydrol. Process.* 9.3–4, pp. 331–346. DOI: 10.1002/hyp.3360090308.

Wood, E. F., Sivapalan, M., Beven, K., and Band, L. (1988). "Effects of spatial variability and scale with implications to hydrologic modeling". In: *J. Hydrol.* 102.1, pp. 29–47. DOI: 10.1016/0022-1694(88)90090-X.

Woolhiser, D. A. (1996). "Search for Physically Based Runoff Model—A Hydrologic El Dorado?" In: *J. Hydraul. Eng.* 122.3, pp. 122–129. DOI: 10.1061/(ASCE)0733-9429(1996)122:3(122).

Woolhiser, D. A., Smith, R. E., and Goodrich, D. C. (1990). *KINEROS, A Kinematic Runoff and Erosion Model: Documentation and User Manual*. USDA, Agricultural Research Service, ARS-77, 130 pp.

Wrede, S., Fenicia, F., Martínez-Carreras, N., Juilleret, J., Hissler, C., Krein, A., Savenije, H. H. G., Uhlenbrook, S., Kavetski, D., and Pfister, L. (2015). "Towards more systematic perceptual model development: a case study using 3 Luxembourgish catchments". In: *Hydrol. Process.* 29.12, pp. 2731–2750. DOI: 10.1002/hyp.10393.

Xavier, A. C., King, C. W., and Scanlon, B. R. (2016). "Daily gridded meteorological variables in Brazil (1980–2013)". In: *Int. J. Climatol.* 36.6, pp. 2644–2659. DOI: 10.1002/joc.4518.

Yan, H., Moradkhani, H., and Zarekarizi, M. (2017). "A probabilistic drought forecasting framework: A combined dynamical and statistical approach". In: *J. Hydrol.* 548, pp. 291–304. DOI: 10.1016/j.jhydrol.2017.03.004.

Yang, D., Herath, S., and Musiake, K. (2000). "Comparison of different distributed hydrological models for characterization of catchment spatial variability". In: *Hydrol. Process.* 14.3, pp. 403–416. DOI: 10.1002/(SICI)1099-1085(20000228)14:3<403::AID-HYP945>3.0.CO;2-3.

Yang, D., Herath, S., and Musiake, K. (2002). "A hillslope-based hydrological model using catchment area and width functions". In: *Hydrol. Sci. J.* 47.1, pp. 49–65. DOI: 10.1080/02626660209492907.

Yatheendradas, S., Wagener, T., Gupta, H., Unkrich, C., Goodrich, D., Schaffner, M., and Stewart, A. (2008). "Understanding uncertainty in distributed flash flood forecasting for semiarid regions". In: *Water Resour. Res.* 44.5, W05S19. DOI: 10.1029/2007WR005940.

Zehe, E., Ehret, U., Pfister, L., Blume, T., Schröder, B., Westhoff, M., Jackisch, C., Schymanski, S. J., Weiler, M., Schulz, K., Allroggen, N., Tronicke, J., van Schaik, L., Dietrich, P., Scherer, U., Eccard, J., Wulfmeyer, V., and Kleidon, A. (2014). "HESS Opinions: From response units to functional units: a thermodynamic reinterpretation of the HRU concept to link spatial organization and functioning of intermediate scale catchments". In: *Hydrol. Earth Syst. Sc.* 18.11, pp. 4635–4655. DOI: 10.5194/hess-18-4635-2014.

Zehe, E., Maurer, T., Ihringer, J., and Plate, E. (2001). “Modeling water flow and mass transport in a loess catchment”. In: *Phys. Chem. Earth Part B* 26.7, pp. 487–507. DOI: 10.1016/S1464-1909(01)00041-7.

Zhang, L., Dawes, W. R., Hatton, T. J., Reece, P. H., Beale, G. T. H., and Packer, I. (1999). “Estimation of soil moisture and groundwater recharge using the TOPOG_IRM Model”. In: *Water Resour. Res.* 35.1, pp. 149–161. DOI: 10.1029/98WR01616.

Zhang, W. and Montgomery, D. R. (1994). “Digital elevation model grid size, landscape representation, and hydrologic simulations”. In: *Water Resour. Res.* 30.4, pp. 1019–1028. DOI: 10.1029/93WR03553.

Author's declaration

I prepared this dissertation without illegal assistance. The work is original except where indicated by special reference in the text and no part of the dissertation has been submitted for any other degree.

This dissertation has not been presented to any other University for examination, neither in Germany nor in another country.

Tobias Pilz

Potsdam, November 2019