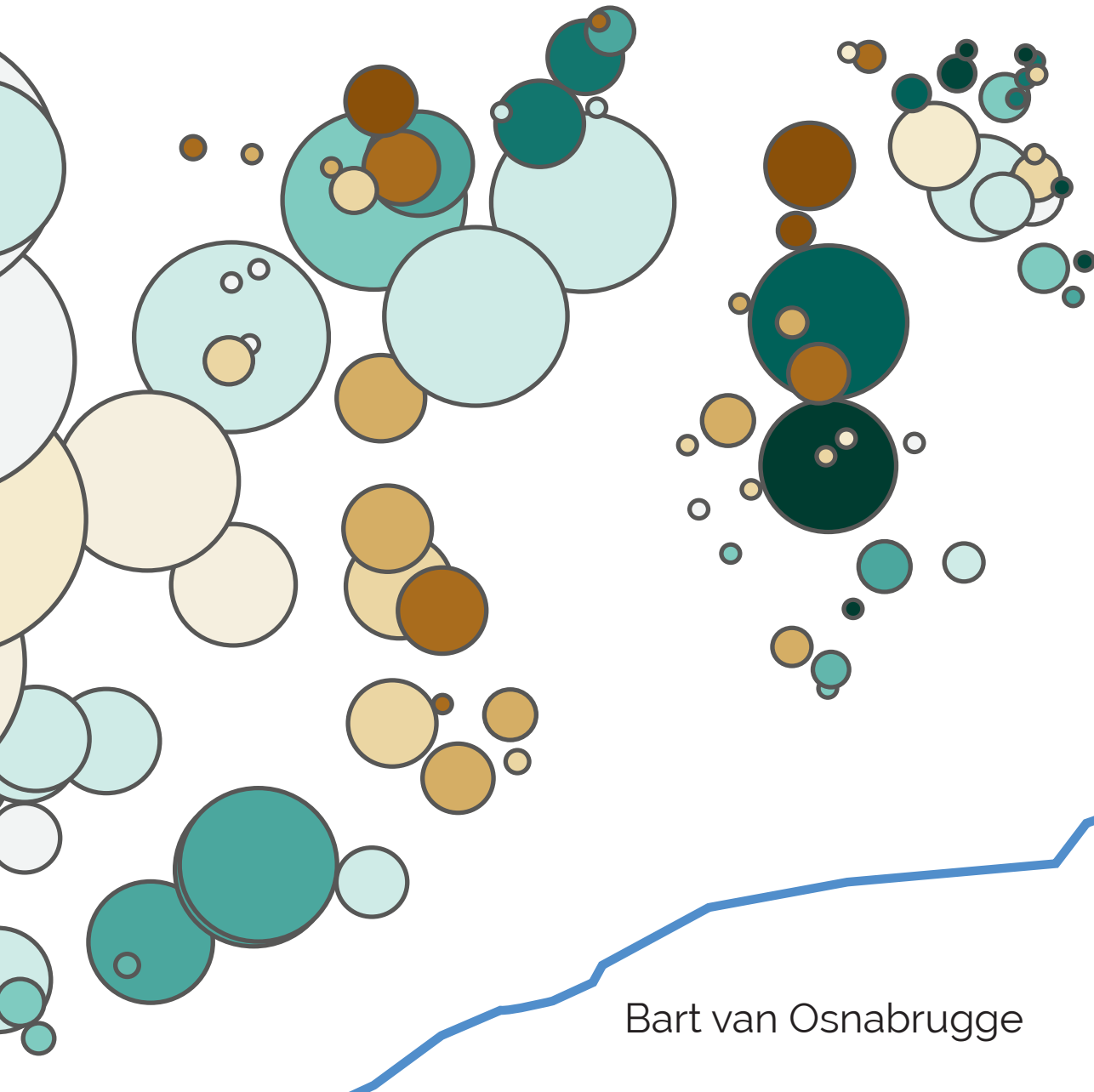# Interpolate, Simulate, Assimilate
operational aspects of improving
hydrological forecasts in the Rhine basin

Bart van Osnabrugge

Interpolate, Simulate, Assimilate: operational aspects of improving hydrological forecasts in the Rhine basin

Bart van Osnabrugge

Thesis committee

Promotors
Prof. Dr A.H. Weerts
Special professor of Hydrological Predictability
Wageningen University & Research

Prof. Dr R. Uijlenhoet
Professor of Hydrology and Quantitative Water Management
Wageningen University & Research

Other members
Prof. Dr G.B.M. Heuvelink, Wageningen University & Research
Prof. Dr D. Solomatine, IHE Delft Institute for Water Education
Prof. Dr B.J.J.M. van den Hurk, Vrije Universiteit Amsterdam
Dr M.H. Ramos, IRSTEA, Antony, France

Interpolate, Simulate, Assimilate: operational aspects of improving hydrological forecasts in the Rhine basin

Bart van Osnabrugge

"Great minds against themselves conspire."

- from *Dido and Aeneas*, H. Purcell

Foreword

In the opera 'Dido and Aeneas,' composed by the English composer Henry Purcell, the Queen of Carthage (Dido) and the Prince of Troy (Aeneas) share a brief romance, before being driven apart by malicious forces (witches) pretending to represent the will of the Gods, leading to the death of Dido. However, it is made clear by the chorus that all of this suffering was unnecessary had either of them had the clarity of mind to look beyond their own grievances: *"Great minds against themselves conspire, and shun the cure they most desire."* When I participated in a performance of this opera during the final stage of my PhD research, I felt a certain resonance between the words sung and the process of completing a PhD, or even, *by Jove*, with research in general.

There is a risk of a PhD project becoming a lone struggle against unaccomplishable goals and high standards, most of them self-inflicted. Mental health statistics do not look friendly on PhD candidates: it seems we do 'against ourselves conspire'. As with Dido and Aeneas, outside influences play an important role, but let's hope that supervisors and colleagues do not have to be compared with the foul witches from the opera who introduce themselves with "Harm's our delight, and mischief all our skill."

The cure is out there. My cure has been in the great support of friends, fellow PhD candidates in Wageningen and at Deltares, doing small side projects and presentations that showed me how much I had grown in knowledge, conferences, deep and shallow science discussions, a certain night in Stockholm, understanding supervisors, performing in a science related music performance and in the EGU poetry slam, and many other things.

I remember a conversation at the start of my thesis about 'not wanting to write papers that are just another paper on an endlessly growing pile of papers.' My growth as a scientists might be that I now realize the folly of that thought and can look at the pile of papers not as an enemy, but as confirmation that what I do matters not only to me, but for other people as well.

Still, I have good hope that this thesis will, just like a skilful forecast, make some impact before becoming part of history. It is for that reason that I have chosen to dedicate a specific section of the synthesis to concrete advise to improve the forecasting system used by the Water Management Centre for the Netherlands (WMCN).

So without further ado, I present the result of four years of conspiracy and cure, a thesis that can proudly claim its place on the grand pile of PhD theses, and I am glad for that.

Contents

Chapter 1

Introduction

1.1 Why this work?

The threat of increased extreme weather events and their negative consequences, such as floods and droughts under climate change, has increased the awareness of such devastating events. It is no longer enough to know how likely a flood or drought that has just occurred was, but the new question is if future events will happen more often and are more extreme (*van den Hurk et al.*, 2016). However, it is not true that extremes exist only because of climate change. Processes in the hydrological cycle (e.g. precipitation, temperature, streamflow, etc.) are well known to have statistical distributions that show enormous natural variability with 'casual' extreme events that can easily differ one or more orders of magnitude in size compared to average conditions (e.g. *Katz et al.*, 2002; *Katz*, 2010).

It is the role of environmental scientists to help society to deal with environmental risks, both current risks and future risks. This societal request has led to a shift towards more 'actionable climate research' (*Asrar et al.*, 2013). Research should not only answer questions about the future climate, but also inform people about how to respond and to act (*Palmer*, 2012). It is not enough to generate knowledge; the gained insights have to be communicated with practitioners and science communication has become an integral part of being a good scientist (*Peters et al.*, 2008).

This thesis has been written in the framework of the IMPREX (IMproving PRedictions and management of hydrological EXtremes) project. The rationale behind the IMPREX project is that 'present prediction and projection systems, and the present use of the information derived from these systems, are a starting point to understand, put in context and make progress on the impacts of a future climate' (*van den Hurk et al.*, 2016). With the motto 'learn from today to anticipate tomorrow' it takes into account two points that have been briefly introduced above. First, the realization that the best way (or at least a good and effective way) to prepare for future extremes is to draw from, and build on, the knowledge and practices that exist in managing and estimating the hydrological extremes of today. Second, that actionable research is best achieved by inviting local partners that have to deal with climate and water resource problems on a daily basis into the scientific process.

The local partner of the research presented in this thesis is the 'Watermanagementscentrum Nederland' (WMCN, translation: Water Management Centre for the Netherlands). The WMCN is a part of the Ministry of Infrastructure and Water Management, that is tasked with supplying actual information on the current status of all water bodies and rivers relevant to stakeholders in the Netherlands. The WMCN provides crisis management in the event of floods, droughts or heavy pollution, but also acts as a place to share knowledge about water management (*Rijkswaterstaat*, n.d.). In particular, the WMCN runs and manages a forecasting system for the main rivers entering the Netherlands: the Meuse and the Rhine. The latter of these two rivers is taken as test bed for the studies presented in this thesis.

In the remainder of the introduction, I first briefly explain the role of hydrological forecasts and what type of hydrological forecasts are the target subject of this thesis (1.2). This is followed by a walk-through of such a short-to-medium term forecasting system, to give some insight on what type of components are involved in making such a forecast (1.3). Next, I narrow down on the research topic of this thesis by explaining what I mean with 'operational aspects' (1.5). This results in the research questions (1.6), followed by the thesis outline (1.7).

1.2   What are forecasts and how are they used?

A hydrological forecast concerns the estimation of future states of hydrological phenomena (*WMO*, 2009). Hydrological forecasts are made to aid decision making in water management, in particular to manage adverse consequences of water shortages (droughts) and excess of water resources (floods). Hydrological forecasts are part of risk management strategies. In general, (flood) risk management incorporates the following steps (*Alfieri et al.*, 2012):

1. Prevention

2. Protection

3. Preparedness

4. Emergency response

5. Recovery and lessons learned

Forecasts provide early warning and are an essential part of the preparedness phase of hydrological risk management.

The applications of hydrological forecasts are many, dependent on which hydrological states are forecasted and how far into the future the forecasted event will occur. The time between that the forecast is issued and the moment that the event occurs is called the *lead time* of the forecast. Several fields of application of hydrological forecasts and their relation with lead time are shown in figure 1.1. Based on the lead time, forecasts can be categorised into groups, each with associated methods:

Short-term forecasts have a lead time up to two days. Very short-term forecasts that forecast up to a couple of hours are also referred to as (radar)nowcasting. Nowcasting models predict precipitation essentially based on extrapolation from a series of consecutive weather radar scans up to a few hours (*Liguori and Rico-Ramirez*,
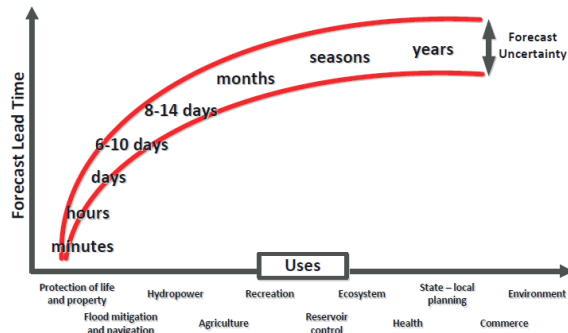
Figure 1.1: Different uses of hydrological forecasts and associated lead times and uncertainty. From Verkade (2015).

2014). Short-term streamflow forecasts in larger basins can be reliably achieved by extrapolating from currently observed upstream streamflow. Short-term forecasts are mostly relevant on a small spatial scale, as local conditions can change most quickly from safe to dangerous and include hazards such as land slides and flash floods.

Medium-range forecasts have a lead time ranging from two to ten days. This lead time is achieved by using numerical weather prediction (NWP) systems. Specifically, limited area NWPs are used. Limited area NWP are higher resolution models that take the results of a global NWP as boundary conditions to resolve the local weather at more detail than is possible on the global scale (*Baldauf et al.*, 2011). Medium-range forecasts target rivers at the (sub)basin scale (*Renner et al.*, 2009; *Jaun and Ahrens*, 2009; *van den Bergh and Roulin*, 2010; *Addor et al.*, 2011). The accuracy of NWP models becomes questionable beyond 10 days (*Bartholmes et al.*, 2009). Skilful hydrological forecasts can occasionally be made for larger basins (e.g. the whole Rhine) beyond 10 days when the initial conditions are the main influence on the future state, for example under very stable weather conditions with little rainfall, or large dependency on spring snow-melt (*Wanders et al.*, 2019).

Long-range or seasonal forecasts deal with monthly to seasonal time scales. Seasonal forecasts are different from the aforementioned forecasts, in that seasonal forecasts rely on a combination of initial conditions (e.g. snow storage) and climate precursors such as the El Niño Southern Oscillation (ENSO) to predict monthly aggregated values of, for example, streamflow. Whereas historically seasonal forecasting models relied on statistical relations between climate precursors and observations, currently seasonal forecasts are also produced based on climate models (*Yuan et al.*, 2015).

Climate projections deal with expected meteorological and hydrological changes on the scale of years to decades. A remaining issue is how to effectively translate these long-term changes into actual future weather (*Hazeleger et al.*, 2015). Climate pro-

jections are not usually grouped as operational forecast. Nonetheless, climate projections clearly aim to improve preparedness and (emergency) response and therefore fit the broader picture of operational forecasting, albeit on a timescale that is more associated with policy making.

This thesis deals specifically with medium-range forecasts. For the Rhine basin, 10 days is at the limit of skilful prediction of precipitation from NWPs (*Renner et al.*, 2009; *van Osnabrugge et al.*, 2019). The next paragraph gives a technical overview of the medium-range forecasting system under study.

1.3 Medium-range hydrological forecasting systems

Different components that are necessary to produce medium-range hydrological forecasts can best be explained by describing an actual forecasting system. The forecasting system that is repeatedly evaluated in this thesis is the forecasting system RWsOS Rivers, the forecasting system that monitors the two largest rivers entering the Netherlands: the Rhine and the Meuse. The RWsOS Rivers forecasting system is based on Delft-FEWS. The Delft-FEWS software provides a platform by means of which operational forecasting systems can be constructed (*Werner et al.*, 2013). Delft-FEWS is not a model, but takes care of all the processes in between: data import from observation networks or external forecasting models, pre- and postprocessing of these data, and scheduling and running the models that perform the forecast. Delft-FEWS is used in over 40 forecasting systems worldwide (*Werner et al.*, 2013).

To start the forecasting process for the Rhine, hydrometeorological data is imported in FEWS from several observation networks: precipitation and temperature from the SYNOP (a standardized protocol to share gauge data) network and from national agencies such as the Koninklijk Nederlands Meteorologisch Instituut (KNMI), Deutsche Wetter Dienst (DWD), the Swiss Federal Office of the Environment (FOEN) and MeteoFrance; water levels and discharges from several river gauge networks; numerical weather predictions from several sources such as the ECMWF, KNMI and DWD; precipitation radar data (RADOLAN) from the DWD; and forecasts that are provided by external parties such as the streamflow forecast for the Rhine at Basel from the FOEN and snow cover from again the DWD. As of yet, not all data that are imported are used in the production of the forecast. With many more possible sources of information available such as satellite-based observations of, for example, soil moisture and radiation, and more local observations of streamflow throughout the basin, there is certainly room to improve the forecasting process by including additional data.

Before the forecasting models can be run, it is often necessary to preprocess the imported data. Most notably, the raw measurements need to be validated by an automatic quality control procedure so as to limit errors in the forecast due to obvious errors in the imported data (*Blenkinsop et al.*, 2017). Another preprocessing step is the spatial and temporal interpolation of data to provide areal estimates and to fill gaps in the time series. Other preprocessing could include bias-correction of NWP forecasts data (*Verkade et al.*, 2013). After preprocessing, all necessary data is exported to the forecasting models in a data format that the models can understand. To make the forecast for the Rhine, multiple models are being used. First, a hydrological rainfall-runoff model is used to translate observed

and forecasted precipitation and temperature into runoff for 148 subbasins in the Rhine. Second, the calculated runoff is fed into a hydrodynamic model that models the propagation of the water through the main course of the Rhine river.

The first model, and the modelling step that will be a focus in this thesis, is the hydrological model that translates observed and forecasted precipitation and temperature into runoff for 148 subbasins in the Rhine. The hydrological model is the center of hydrological forecasting systems (*Werner et al.*, 2013). At the start of this thesis project (Oct 2016), the operational hydrological model was the HBV96 model. The HBV96 model lacks the flexibility to make use of recent advancements, such as the rapid expansion of available satellite derived spatial information (see e.g. *Peng et al.*, 2017; *Guo et al.*, 2017; *Zhang et al.*, 2017; *Li et al.*, 2016; *Lettenmaier et al.*, 2015; *Thakur et al.*, 2017; *Balsamo et al.*, 2018; *McCabe et al.*, 2017, for some recent reviews), standardized interfaces such as OpenMI (*Gregersen et al.*, 2007), and the call for open and reproducible science through shared and open models (*Hutton et al.*, 2016). It has a closed model base and does not share common data sharing conventions, which makes coupling with other applications, that can improve in the forecast, unnecessarily difficult. Additionally, partial updating of parameters for different applications such as for low flows, high flows, hourly and daily time step and for policy use, including corrections on precipitation, makes interpretation of results more difficult. This has led to the conclusion that the HBV96 model has reached its end of life status for operational hydrological applications in the Rhine basin. In this thesis, attention is given to a possible replacement. Firstly in the form of wflow_hbv, which is a translation of the old model concept to the wflow hydrological modelling framework (*Schellekens et al.*, 2019a) and secondly in the form of wflow_sbm, which is a different model concept altogether (see Chapter 5).

Before future run-off can be calculated by the hydrological model, first the current conditions, the so-called initial conditions, need to be determined by running the model based on historical observations until the time of forecast, referred to as $t0$. Better initial conditions improve the subsequent forecasts. State updating is a data assimilation method that can improve the initial state estimate (*Sun et al.*, 2016). This was not used in RWsOS Rivers yet for the hydrological model and is thus another possibility for improving the forecasts. State updating is investigated in Chapters 6 and 7.

After the initial state is determined by the historical update run of the model, the actual forecast model run is made. The forecast run is done with the same model as the historical update, starting from the initial state and forced by the forecasts. For RWsOS Rivers, several forecasts are made with the same model and initial state, but forced with data from different limited area NWP models. Furthermore, the forecast of a NWP does not consist of a single value for each lead time (deterministic forecast), but consists of a collection of possible future states (ensemble forecast) as expressed by a fixed number of ensemble members in which each member is a possible future state. A calculation is made for each ensemble member, leading to an ensemble forecast of discharge for each of the subbasins.

After the run of the hydrological model, the results can be further adjusted by applying postprocessing techniques. Postprocessing in RWsOS Rivers is done with ARMA correction (see Chapter 7) and quantile-to-quantile mapping (*Verkade*, 2015).

The second model that is run is a hydrodynamic SOBEK model that models the water flow
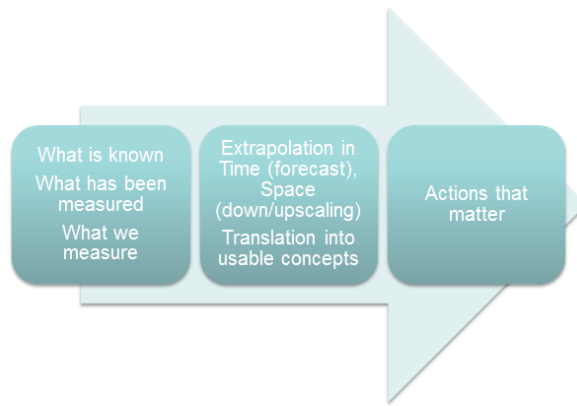
Figure 1.2: Arrow representing the rationale behind operational forecasting. The transition from left to right is accomplished using operational forecast services.

in the main course of the Rhine river. The SOBEK model takes the observed (for the historical update run) and forecasted discharge from the hydrological model as input for the final forecast of discharge and water level at Lobith, which is where the Rhine enters the Netherlands. The SOBEK model and hydraulic modelling is outside of the scope of this thesis.

As a last but not to be underestimated step, the results have to be presented to the forecasters at the WMCN and disseminated to other users of the forecasts. The results of the forecast for the Rhine (and Meuse) are used as boundary conditions to run models for the Rhine delta and water allocation in the Netherlands. Effective dissemination of forecasts is a subject that has been gaining increased attention of the forecasting research community in recent years, especially what the communication about forecast uncertainty is concerned (*Ramos et al.*, 2010).

1.4 Operational hydrological forecasting

Operational forecasting systems refer to systems that continually issue forecasts with a fixed schedule, for example two times a day at 00:00 and 12:00, 24/7. The goal of operational forecasting is to transform 'raw' (hydro-meteorological) information into operationally useful concepts and knowledge with the goal of taking informed action, in real time. During the forecast exercise our knowledge of the system and measurements of input and output variables are, by means of statistical methods and/or process-based models, extrapolated in both time and space, as shown in Figure 1.2. Because the exact meaning of 'operationally useful concept' is end-user defined, hydrologic operational modelling is a separate niche within the larger hydrological research community. The field of meteorological forecasting provides an example of a field where studies on the numerical computation and data challenges of operational systems form their own sub-discipline.

Traditionally hydrological research was developed to cater client needs (*Freeze and Harlan*, 1969). However, with the advancement of process understanding and challenging research questions, involving change to the studied system, it became apparent that simply being right (i.e. matching the observed hydrograph) was not enough. One has to be right for the right reasons (*Kirchner*, 2006), meaning that for a justifiable use of a hydrological model the model should not only replicate the variable that has to be forecasted over a historic period (e.g. river flow at Lobith), but the model's internal distribution of flows should also match the hydrologists understanding of the basin (*Klemeš*, 1986). While there a broad agreement on this point, it has to be recognized that 1) for most operational systems it is currently impossible to verify that modelled processes correspond to real-world processes because of experimental data scarcity, and 2) it might not be possible at all to run models that resolve all processes on the micro-scale operationally because of the mentioned data constraint and computational constraints (*Beven*, 1989).

The art of operational forecasting is focused on translating measured and predicted observables into an end-user defined product. The hydrological model in this case is not used for 'truth-finding,' but to enable the modeller to answer a user-defined question. Historically, this would involve classic inquiries regarding streamflow characteristics such as 'What will the discharge be next day, week or month?', 'What is the expected frequency of a certain event?' Today, it is accepted that hydrological systems are under constant change (*Montanari et al.*, 2013). While the current response of the system to hydrometeorological forcings might be quite well understood, new questions arise with respect to the future (*Wagener et al.*, 2010). These questions ask for new operational models which give the modeller access to the perceived controls on streamflow generation, such as landuse, climate, and the change thereof.

Operational hydrological forecasting is more data-oriented than (flow) process-oriented. Operational forecasting is not about truth-finding, although it is logical that answers that closely resemble reality are more useful. Independent of the modelling philosophy used in setting up the forecasting models, the data available in real time is the starting point, which is then used in two ways (Fig. 1.3):

On one hand, all available historical data can be used to set-up and calibrate a model which translates measurable and predictable observables into a forecast of a requested variable. Within hydrology, this is the traditional art of building a rainfall-runoff model.

On the other hand, all currently available measurements of observables can be used to adjust the model prediction to more likely values using data assimilation techniques, making use of all sorts of real-time available data (e.g. from remote sensing) besides only the required input data of a model.

The future of operational forecasting is then in combining these two pathways to make predictions as good as possible, warranted by the quality of the model, the quality of the data, and the efficiency to extract relevant information from those sources, in real-time.

1.5   What is meant with 'operational aspects'

Operational aspects deal with challenges that arise on scientific topics when interacting with operational practice. It is not only about knowledge on the topic, but also about gen-

erating knowledge about the use of knowledge and methods in practice. For example, as discussed in Chapter 3, the topic of precipitation interpolation, or the estimation of areal precipitation in general, is a vivid research topic (one you could easily spend multiple PhDs on) producing many methods to complete the task of estimating spatial precipitation from a network of precipitation gauges and in combination with other sources of information (e.g. *Brauer*, 2014; *Hazenberg*, 2013; *Rios Gaona*, 2017). The goal is usually to find the optimal method. Intersecting this topic with forecasting practice, new constraints are introduced, which limit the freedom to search for such an optimal method. The operational aspect of precipitation interpolation can be found in the suitability of a method to be used in an operational context.

The operational context adds additional requirements. Operational forecasting is set apart from hydrological science because of a set of constraints particular to the operational forecaster:

Near real time availability of data Lack of data is a major constraint in all hydrological research. This holds from the practical viewpoint that many catchments are ungauged (*Hrachowitz et al.*, 2013), but it has also been realized for a long time that it is impossible to measure hydrological fluxes and stores on each relevant scale (*Klemeš*, 1983). Point measurements are not representative for the heterogeneity of input observables, while spatially gridded products are too coarse to capture all relevant detail.

The additional data constraint for operational forecasting is that data sets have to be available in (near) real-time. There is a constant need for input data, without which the operational model cannot run. In contrast, for other purposes such as frequency analysis or off-line historical systems analysis it is less relevant when the measurements become available.

Not all data is available in (near) real time. For example, certain remote sensing products can have delays for up to a couple of days or weeks; Only a limited set of all available observations might be shared with the forecaster through data sharing agreements.

Continuous operation Operational systems are in constant operation. The repetitiveness of the task calls for a high level of automatization, but also for robustness in used methods. In particular, methods should be robust against errors in the data and missing data.

Processing time There is a strict constraint on processing time determined by the required update schedule. For example it may be required that a new forecast is issued every day, or every six hours, or even every hour. This limits the processing time.

Additionally, processing time is time that is 'borrowed' from lead time as the lead time is decreased by the time it takes to make the forecast. A longer lead time increases the positive impact a forecast can have in mitigating negative consequences. This is most limiting for (very) short term forecasts. As lead time is an important indicator of the quality of a forecasting system, efficient algorithms are more than just a convenience and cost savers, but do directly impact the forecast quality by
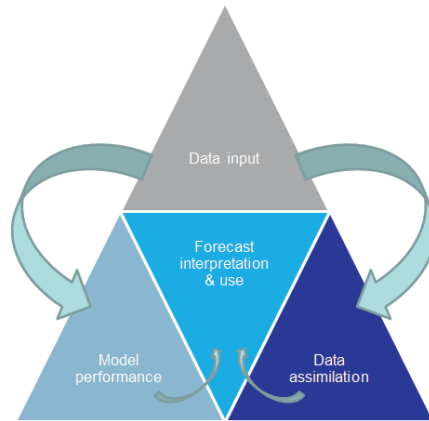
Figure 1.3: Representation of the steps in operational forecasting. The triangles represent the different focus areas, while the arrows show the flow of information between different aspects. The starting point is the available data on top. Information from this data is used together with expert knowledge to build a model which, based on a certain amount of input data, provides forecasts. The same or other data sources can be used simultaneously to reflect on the perceived quality of the generated forecast and make updates where necessary through data assimilation schemes.

influencing the lead time. Alternatively, delays in data availability can negatively impact forecast quality.

Desicion making under uncertainty Forecasting systems are tools to aid decision making. Presenting the forecast, including its uncertainty, to the decision maker is therefore an integrated part of the forecast chain. To be able to do this, not only must the sources of uncertainty in the model be well understood, the interaction between the presented information and the user must also be taken into account (*Ramos et al.*, 2010). The degree of uncertainty is not just a measure of model performance, as is often used in model comparison studies, but is important information for the decision maker (e.g. *Weijs et al.*, 2010). Based on the estimate of predictive uncertainty and objective guidance on the level of confidence in the forecast, the end user can decide to take action based on their risk tolerance (*Demargne et al.*, 2014).

Due to the predictive uncertainty increasing with lead time, situations can arise a specific trade-off between predicting an event with a long lead time, but a low predictability, and one with a short lead time, but high predictability. Because both might be required by the water manager, and the consequences might be considerable, the meaning and degree of uncertainty must be thoroughly understood.

Operational aspects include then all steps between and including the observations and gathering of the input data up to and including the dissemination of results to the end-user, under the constraints presented here.

1.6   Research questions

As laid out above, hydrological forecasts are a useful and cost-effective tool to aid decision making. Hydrological forecasts are based on a set-up consisting of several model and data components which need to be integrated for an effective forecast. Part of this model train is the hydrological model. Regarding hydrological models, there is a current trend to move towards high-resolution spatially distributed gridded models, both in the wider literature and in the specific case of the forecasting system RWsOS Rivers of Rijkswaterstaat. Concurrently, the forecasting models are used for ensemble forecasts and are further improved with data assimilation techniques. The combination of those two movements, gridded models and ensemble forecasts including data assimilation, leads to new challenges, conceptually, but even more so when intersected with operational practice. This has led to the following research questions about these operational aspects of hydrological forecasts:

Research questions

- · To what extent is hydrological operational forecasting a separate scientific subfield in hydrology and to what extent is it the summation of other determined subfields?
- · What are the outstanding challenges in operational forecasting related to the change from lumped to gridded hydrological models?
- · In which manner does case-specific actionable research contribute to the science of hydrological modelling and hydrological forecasting?

1.7   Thesis outline

The outline of this thesis is as follows. The thesis consists of five chapters (Ch. 3-7) containing original research under the theme of operational aspects of hydrological forecasts in the Rhine basin. The ordering is determined logically from the order of the modelling/forecasting process as reflected in the title of this thesis: interpolate, simulate, assimilate. The chapters are presented in such a way that they can be read stand-alone.
Chapter 2 provides an overview of the Rhine basin. Both Chapter 3 and Chapter 4 focus on the foring used in hydrological models.
Chapter 3 describes the genRE method: a method to interpolate precipitation based on a limited amount of rain gauge data under operational conditions.
Chapter 4 presents the results of an experiment with different types of potential evaporation forcing. It also describes the production of the temperature and potential evaporation gridded dataset that complement the precipitation dataset from the first chapter to form a complete set of forcing data to perform distributed hydrological model experiments.
Chapter 5 is all about simulation as a new hydrological model concept for modelling the Rhine is introduced: wflow_sbm. It is the only chapter were this model concept is used, as later chapters return to the wflow_hbv model. The model development is an important part of the improving predictions narrative and showcases the interrelation between forecasting issues and model development.

Chapter 6 and 7 then continue with data assimilation. Lake level measurements are used to improve the initial conditions in the Swiss part of the Rhine basin (Chapter 6) and a comparison is made between currently implemented post processing techniques and data assimilation techniques in a large scale assimilation experiment including all the major subbasins of the Rhine.

The last chapter, Chapter 8, synthesises on the conclusions from the earlier chapters and extrapolates these conclusions together with other lessons-learned into comments both on the scientific context and specifically also into advice related to operational forecasting. It is this latter part which I hope will be read by the operational forecasters at the forecasting centre at Rijkswaterstaat and other operational hydrological forecasting centres and I hope it will form a basis to further improve predictions in the Rhine basin and beyond.

Chapter 2

Study area

2.1 The Rhine basin

The river Rhine (figure 2.1) runs from the Swiss Alps along the French-German border, through Germany and enters the Netherlands near Lobith, where the Rhine river enters into the Rhine delta. At Lobith, the basin area equals approximately 160,000 km^2 (*Verkade et al.*, 2013).

Sub-basins and topography

The Rhine basin can be partitioned in many sub-basins. In particular the main tributaries of the Rhine are treated as separate river systems: the Aare and Swiss Rhine in Switzerland, the Neckar and Main in Germany and the Moselle in the French-German border region.

The river is partitioned in six different reaches:

Alpine Rhine from its source in the Alps to the outlet in Lake Constance.

High Rhine from Lake Constance untill Basel (Switzerland). Main tributary is the river Aare.

Upper Rhine from Basel untill Bingen (Germany). The rivers Elz, Kinzig, Murg, Ill, Neckar and Main are important tributaries in the Upper Rhine.

Middle Rhine from Bingen untill Bonn (Germany). Here the tributaries Lahn and Moselle join the Rhine.

Lower Rhine from Bonn untill Pannerdense Kop (The Netherlands). Tributaries in this part are the Sieg, Ems and Lippe.

Rhine Delta from the bifurcation at Pannerdense kop towards the sea, where each bifurcation carries a different name.

Verkade et al. (2013) used three spatial scales when analyzing ECMWF precipitation and temperature ensemble reforecasts in the Rhine basin: headwaters (defined as gauged sub-basins smaller than 2500 km^2), main tributaries, and the whole basin as represented by the flow into the Rhine delta at Lobith.
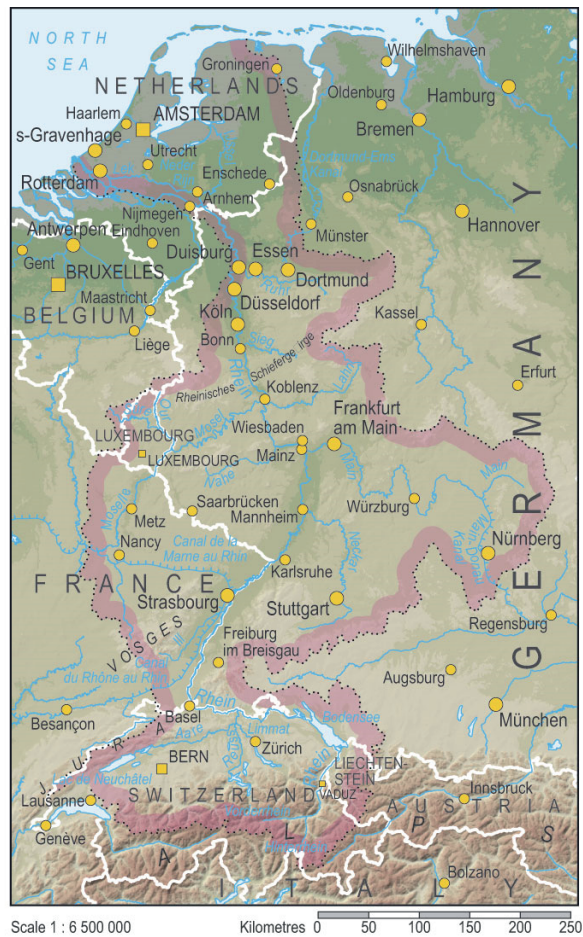
Figure 2.1: *Map of the Rhine basin (UNEP/DEWA/GRID-Europe, Retrieved on 20-07-16 from* `http: // grid. unep. ch/ / products/ 4_ Maps/ basin_ rhineb. gif)`
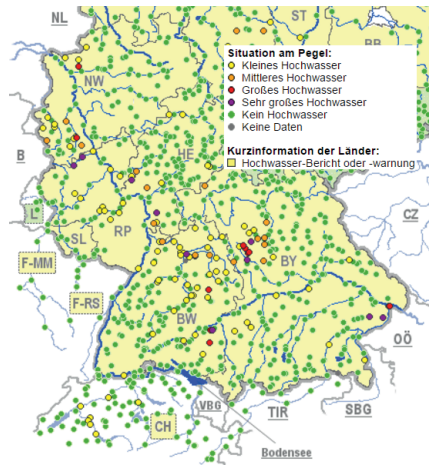
Figure 2.2: Screenshot of the 'Hochwasserportal,' a website providing an overview of near real-time water level measurements (www.hochwasserportal.de). The situation at the gauges (Situation am Pegel) is the highest recorded water level in the period 30-05-2016 untill 03-06-2016, during which a high water level alert was issued.

Runoff regimes

The Rhine has an average discharge of 2200 $m^3 s^{-1}$ at Lobith. Snow and snowmelt have a large effect on the river Rhine's discharge dynamics. During spring and early summer, more than half of the river's flow at the outlet originates from snowmelt in the Swiss Alps (*Verkade et al.*, 2013). River flow is less prone to fluctuations than that of other European rivers of comparable dimensions because of the different hydrological characteristics, which show opposing factors in seasonality. In the Upper Rhine, discharge is lower during the cold season because a large part of winter precipitation is transferred towards the warm season in the form of snow. The lower Rhine, on the other hand, has a reversed relation, because in the low mountains and lowlands snow storage is negligible and summer precipitation largely evaporates. The Alpine Rhine shows a very irregular discharge regime, but this is filtered by Lake Constance, downstream of which discharge is relatively steady (*Kalweit et al.*, 1993).

Floods in the Rhine basin

The classification of a certain event as a flood event is a matter of perspective. For example, June 2016 saw local flood events due to extensive and intensive rainfall in the German states of Baden-Württemberg and Bayern. Also, some of the tributaries recorded very high water levels. However, the flood peak in the main river was not threatening (Fig. 2.2).

The largest recent flooding event with a large risk of flooding in the Netherlands happened around Christmas 1993 and during the last ten days of January 1995. The 1995 ex-

traordinary flood was caused by a number of record-breaking rainfall events, spaced over the time period of a week, partly augmented by snow-melt. A long duration of extreme flood levels on the river Moselle led to a situation where interference with any surge coming down the Rhine was extremely likely, which is what happened. Finally, a smaller, but ill-timed, rainfall event leading to additional inflow from the river Sieg atop of the already extreme water levels led to the observed record discharges (*Finkl et al.*, 1996).

Anthropogenic influence on landscape and flow regime

A good overview of anthropogenic influences on the Rhine river can be found in *Kalweit et al.* (1993). The river Rhine serves many purposes for the over 60 million inhabitants of the basin. The enormous rise in economic productivity produces stress on the landscape. Wood and wasteland have been replaced by agricultural land, which has been made possible because of flood protection, drainage and irrigation measures. Besides agricultural land, also the area occupied with paved surfaces and the like has increased dramatically. Different sectors of industry play an important role in the basin and have, in some areas, a large influence on river flow and more strongly on water quality and water temperature. For example, open brown coal mining in the Lower Rhine area required to lower groundwater levels by hundreds of meters. Now that mining is coming to an end, it is a great challenge to restore the area to 'natural' groundwater levels. Other direct influences are the use of Rhine water as cooling water for nuclear power plants and hydro-power production in the High Rhine, Upper Rhine and numerous tributaries.

The Rhine is an important transportation artery for the European main land. For this purpose the river Rhine has been made navigable with locks and weirs and connected to canalized rivers and canals. Relatively new factors in water management planning concerning the Rhine are recreational possibilities and ecology. For the latter fish population is the most important indicator of the ecological quality of the Rhine river.

Through the years the demand to control the waters of the Rhine river have led to many interventions. During Roman times engineering works mainly served shipping, as this was of exceptional importance as a means of transport, in those days. Early settlements and agricultural valuable lands were already protected by dikes.

Regulation of the river started during the Middle Ages. The first river diversions were carried out and flood protection, drainage and irrigation works were created. The use of water power was introduced and extended throughout the basin. Two notable advanced measures were the diversion of the Kanderbach into the Lake of Thun in the Alps, which resulted in strong attenuation of flood peaks. Along the Upper Rhine successful, and lasting, cut-offs were realized.

Regulation measures have become more coordinated in the 19th and 20th centuries, including cooperation between riparian states. Stretches of the river and its tributaries were provided with a stable and regular bed, routed through natural lakes in the Alps. Reservoirs were constructed in the higher parts of the basin and an extensive network of dikes was constructed in the lower reaches.

These structures substantially impacted flood trends. It was found that the construction of the Rhine weir cascade and other river training measures were responsible for up to 10% change in observed flood trends (*Vorogushyn and Merz*, 2013). However, time series

corrected for this river training still exhibit strong significant trends for a number of time periods with climate variability and/or change, land use change and river engineering in tributaries as possible perpetrators. This shows the difficulties that arise when assessing the Rhine as a pure free-flowing river.

Chapter 3

A spatial hourly precipitation dataset for model experiments in the Rhine basin

To enable operational flood forecasting and drought monitoring, reliable and consistent methods for precipitation interpolation are needed. Such methods need to deal with the deficiencies of sparse operational real-time data compared to quality-controlled offline data sources used in historical analyses. In particular, often a fraction of the measurement network reports in near real-time. For this purpose we present an interpolation method, generalized REGNIE (genRE), which makes use of climatological monthly background grids derived from existing gridded precipitation climatology datasets. We show how genRE can be used to mimic and extend climatological precipitation datasets in near real-time using (sparse) real-time measurement networks in the Rhine basin upstream of the Netherlands (approx. 160.000 km^2). In the process, we create a 1.2x1.2 km transnational gridded hourly precipitation dataset for the Rhine basin. Precipitation gauge data is collected, spatially interpolated for the period 1996–2015 with genRE and inverse-distance squared weighting (IDW), and then evaluated on the yearly and daily timescale against the HYRAS and EOBS climatological datasets. Hourly fields are compared qualitatively with RADOLAN radar based precipitation estimates. Two sources of uncertainty are evaluated: station density and the impact of different background grids (HYRAS vs EOBS). The results show that the genRE method successfully mimics climatological precipitation datasets (HYRAS/EOBS) over daily, monthly and yearly time frames. We conclude that genRE is a good interpolation method of choice for real-time operational use. genRE has the largest added value over IDW for cases with a low real-time station density and a high resolution background grid.

3.1 Introduction

Accurate and reliable areal gridded precipitation is one of the most important hydrom-eteorological input variables for hydrological model development and calibration, extreme value estimation, as well as for operational hydrological forecasting. With the advent of high temporal and spatial resolution hydrological models (e.g. *Bierkens et al.*, 2015; *Wood et al.*, 2011) there is an increasing need for estimates of precipitation at equally high temporal and spatial resolutions (*Euser et al.*, 2015; *Melsen et al.*, 2016). For operational hydrological forecasting, high resolution forcing data is needed to derive the best estimates of the initial conditions at the start of the forecast. Subsequently, data assimilation techniques (*Liu et al.*, 2012; *Rakovec et al.*, 2012, 2015) may be used to further improve these initial state estimates. High resolution historical precipitation data is also needed to resolve the model tendencies of the weather forecasts with equally high accuracy.

In Europe, current and past efforts to construct transnational (gridded) datasets of precipitation have mainly focused on climatological analyses (*Haylock et al.*, 2008; *Rauthe et al.*, 2013; *Photiadou et al.*, 2011; *Isotta et al.*, 2014). The derived datasets have space-time resolutions appropriate for climatological analyses, where trends encompassing periods of decades are investigated. However, they do not meet the high resolutions needed for local modeling of streamflow, where relevant processes need to be resolved at sub-daily timescales (*Ficchì et al.*, 2016; *Lobligeois et al.*, 2014), and local spatial differences in precipitation can be large, especially in mountainous areas (*Tobin et al.*, 2011). Such areas are not covered in sufficient detail in the current gridded datasets.

Currently available gridded datasets are not available in near real-time, needed for hydrological forecasting, either. Deriving high quality near real-time gridded forcing data for operational hydrological forecasting is a challenge. While some European national meteorological agencies provide this kind of information in near real-time, for many national hydrological agencies this information is not available. As a consequence, the task of deriving aerially averaged rainfall in near real-time is often conducted by the responsible hydrological forecasting agencies. For transboundary catchments (e.g. Danube, Elbe, Oder, Rhine, Sava etc.), this is an even greater challenge.

Additionally, hydrological models are calibrated to a certain spatial distribution of precipitation (*Euser et al.*, 2015). For example, a structural overestimation of precipitation in one region might be compensated for in the model with larger soil stores to dampen the peak flow resulting in good calibration scores, but for the wrong reasons (*Kirchner*, 2006). An important requirement for the hydrological forecasting agencies is therefore that the real-time precipitation estimates should reflect the climatology of the data used for calibration.

The main issue hampering the derivation of high quality precipitation information in near real-time is the limited availability of rain gauge data, either due to restrictions in the equipment and infrastructure used, or due to restrictions placed on the data by the data provider. Many more historical observations are often available for deriving climatological daily precipitation grids, providing important information on orographic and other effects on areal precipitation which is difficult to derive in near real-time.

It seems therefore attractive to capture this hard-won information from climatological gridded precipitation datasets and use it to emulate and extend these high quality data-

sets in near real-time for hydrological forecasting purposes, but based only on operationally available data.

With this goal in mind, we describe in this paper an interpolation method for operational use called genRE. We propose that genRE is a suitable candidate to bridge the gap between high quality offline climatological datasets and the real-time operational datasets needed by hydrological forecasting agencies.

The genRE method is evaluated regarding its capacity to emulate the HYRAS climatological dataset (*Rauthe et al., 2013*) for hydrological forecasting purposes in the Rhine basin in western Europe upstream of the Dutch border. However, the method is not only applicable in the Rhine basin but can be employed in any (transnational) basin to derive gridded areal precipitation datasets for operational streamflow forecasting if a suitable gridded climatological precipitation dataset is available.

The aim of this work is therefore (1) to investigate if the genRE method indeed mimics and extends climatological precipitation datasets in near real-time using (sparse) real-time rainfall measurement networks, (2) to create a high resolution (near real-time) gridded precipitation dataset for the purpose of high resolution hydrological modeling of streamflow and (re)forecasts in the Rhine basin, and (3) to evaluate the precipitation interpolation procedure used, in a hydrological sense.

Considering our aims we answer the following research questions:

1. Does the use of a background grid based on a reference climatological gridded dataset for interpolation of hourly rain gauge observations result in comparable climatology and realistic areal precipitation fields at different (yearly, daily, hourly) temporal scales for operational forecasting?

2. What is the effect of station density on the quality of the obtained gridded datasets, and how is the background grid affected by the length of the employed climatological reference?

3. Does the use of the genRE method lead to modelled discharges which are consistent with those obtained with the reference climatology as forcing for a model that is calibrated on the latter?

The data and study area are presented in Section 2. Sections 3.1 and 3.2 describe the genRE interpolation method and the employed hydrological model. The evaluation setup is explained in Section 4. The results are presented in Section 5, followed by discussion and conclusion in Sections 6 and 7.

3.2 Study area and Data

Rain gauge data

The Rhine basin upstream of the Dutch border (approx. 160.000 km^2) is part of eight different countries (Fig.3.1, Chapter 5.2). To ensure a smooth transition over the basin borders, data was also collected for The Netherlands. Consequently, precipitation data was collected from seven different providers. Some of the data is freely available (Netherlands, Germany). Other data could be accessed after subscription (Switzerland), or was provided without fee after contacting the provider (Luxembourg, Belgium, France). Data
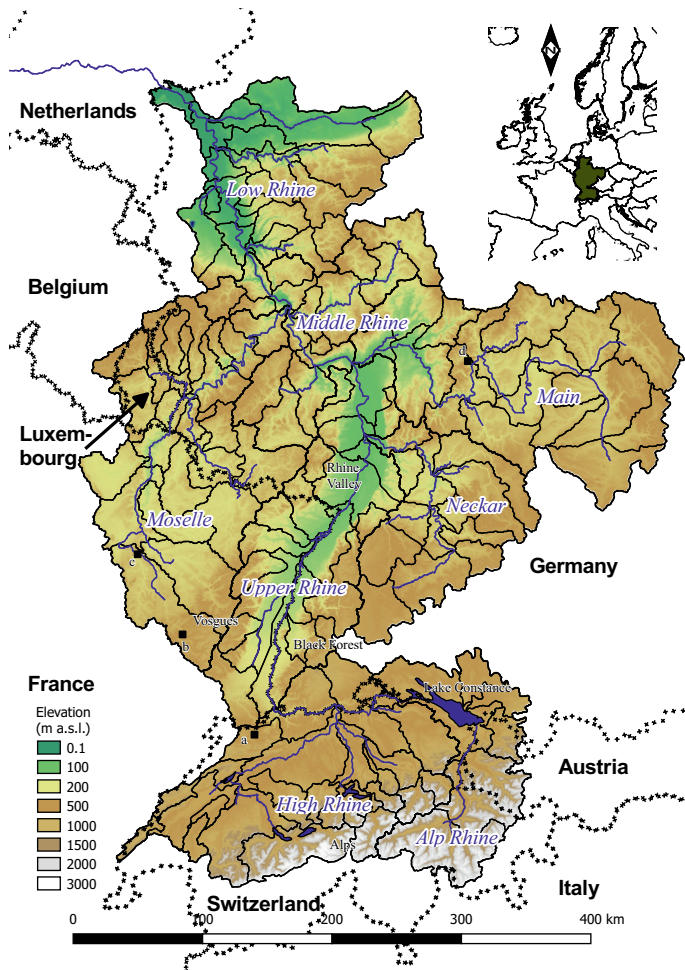
Figure 3.1: Overview of the Rhine basin in UTM32N projection with Digital Elevation Model as background. The thick black lines delineate the eight major sub-catchments with names in italic blue. The thin grey lines represent 148 smaller sub-catchments. The black squares show the locations of four streamflow measurement stations used in Figure 12.

for the Austrian and Liechtensteinian part of the basin (5 stations) was collected from an operational archive of SYNOP precipitation stations provided by the Bundesanstalt für Gewässerkunde (BfG, Germany). Although a tiny part (51 km^2) of the Rhine basin can be found in Italy, no data was collected from that country.

All available stations with hourly resolution have been used in the interpolation scheme. It was decided to start interpolation in June 1996 as this was the start time of the hourly precipitation gauge series made available on the DWD climate data server which covers

Germany.

The total number of stations contributing to the data set at each time step and their spatial distribution are shown in Figure 3.2. Two things are worth noticing. First, the number of stations available in near real-time steadily increases over time, with the quickest increase between 2004 and 2008. To stress the point of the limited availability of real-time data, keep in mind that the complete observation network includes more stations than in the 2015 figure, for the whole period. Second, the number of stations with measured data at each individual (hourly) timestep fluctuated. Large downward peaks resembled moments in time where, for example, no data was recorded for any of the German stations. The loss of station coverage for a large part of the domain is incidental and is dealt with through a preprocessing step in the interpolation procedure, detailed in the methods section. For clarity those outings are not shown in Figure 3.2a.

The data was checked for outliers. The check relies on simple thresholds based mainly on record values measured in Germany and expert opinion of operational forecasters in the Rhine basin (*Weerts et al.*, 2008). Individual hourly values above 90mm and below 0mm were removed from the dataset, which resulted in the removal of only seven occurrences of values above this threshold and two instances of erroneous reporting of negative precipitation. As a secondary check, time series were checked against multiple thresholds for different aggregation times: 50mm, 90mm, 120mm, 150mm and 250mm for 1hr, 3hr, 6hr, 12hr and 24h, respectively. Instances where the threshold was exceeded were flagged and investigated manually to assess if the recording was plausible.

It should be noted that no check was performed on the long-term consistency (temporal homogeneity) of the time series and that artefacts of changing measurement conditions are to be expected. Together with the changing measurement density this makes the created dataset unsuitable for trend analysis (*Peterson et al.*, 1998).

Reference precipitation climatology datasets

In this study we compare our gridded precipitation estimates with three gridded precipitation datasets (Table 3.1).

- The HYRAS (HYdrologische RASterdatensätze) v2.0 (*Rauthe et al.*, 2013) dataset, which was obtained via the Bundesanstalt für Gewässerkunde (BfG).
- The E-OBS v13.1 (*Haylock et al.*, 2008) gridded dataset, which was retrieved from the European Climate Assessment and Dataset website.
- RADOLAN (*Bartels et al.*, 2004) gauge adjusted precipitation estimates, taken from the data feed of the experimental operational forecasting system for the Rhine running at Deltares.

HYRAS

The HYRAS dataset has been developed by the German Weather Service (DWD) (*Rauthe et al.*, 2013). The dataset covers the so-called KLIWAS domain, which comprises a number of transboundary river basins that flow through Germany, including the Rhine. The gridded estimates are created using the REGNIE method, which is explained in section
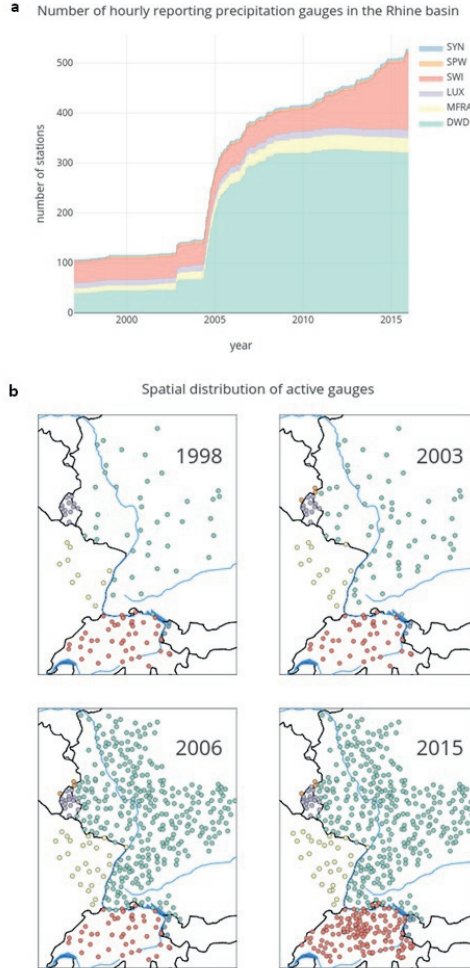
Figure 3.2: ***a***, *Number of available precipitation gauges for each time step inside the Rhine basin upstream of the Dutch border. DWD: Deutscher Wetterdienst, MFRA: Météo France, LUX: Administration de la gestion de l'eau du Grand-Duché de Luxembourg, SWI : Meteo Swiss, SYN: SYNOP stations from operational archive;* ***b***, *Spatial distribution of hourly reporting gauges for four years during the interpolation period. The black lines represent country borders and the blue lines show the location of two major rivers: the Rhine, from south to north and the Danube from west to east.*

Table 3.1: Properties of the reference data sets as compared to genRE

	genRE	HYRAS	EOBS	RADOLAN
Time step	Hour	day	day	15 min
Grid size	1.2x1.2 km	5x5 km	0.25 deg	1x1 km
Period	1996–current	1951–2006	1950–current	2005–current
Domain	Rhine basin	KLIWAS	Europe	Germany
Updates	Near real-time	Sporadic	Monthly	Near real-time
Reference	(This paper)	(*Rauthe et al.*, 2013)	(*Haylock et al.*, 2008)	(*Bartels et al.*, 2004)

3.3, from a data base of at maximum 6000 stations spread over the larger KLIWAS domain. Based on *Rauthe et al.* (2013, Fig. 2) we estimate that the station density is at least comparable with the state of the hourly real-time reporting network of figure 3.2b for the year 2015.

EOBS

The European Daily High-Resolution Observational Gridded Dataset (E-OBS) (*Haylock et al.*, 2008) is based on daily precipitation observations from the European Climate Assessment and Dataset and interpolated with a three step method. At the time of writing, 62 stations were active inside the Rhine basin out of a total of 11383 stations in the complete database (*European Climate Assessment & Dataset*, 2017). Firstly, monthly sums are interpolated using trivariate thin-plate splines with latitude, longitude and elevation as variables. Secondly, the daily anomalies are interpolated using universal Kriging. Last, the monthly and daily steps are combined. The method provides daily standard errors for every grid square.

RADOLAN

RADOLAN (Radar-Online-Aneichung) (*Bartels et al.*, 2004) is another precipitation product from the DWD. It merges gauge measurements from 1300 stations with radar-derived precipitation from 16 weather radars covering Germany. The merge is based on interpolating gauge precipitation and anomalies between gauge and radar data in different ways and then taking a weighted sum of the different components. Due to the range of the rain radars, the French part of the Rhine basin is also included. The Swiss part of the basin is excluded, however.

3.3 Methods

genRE interpolation method

The genRE method was inspired by a method from the DWD (German weather service) called REGNIE ('REgionalisierte NIEderschlagshöhe', which translates as 'regionalized precipitation amounts') (*Rauthe et al.*, 2013). Because we deviate in a number of ways from the original REGNIE method to make it more generally applicable and applicable in real-time, we refer to the method as used in this paper as *generalized* REGNIE or *gen*RE.

The genRE and REGNIE interpolation method are two-step methods. First, monthly precipitation background grids (MBGs) are derived, one for each month of the year, with the expected precipitation amount for each grid cell for that month. In particular, the MBGs can be derived from existing gridded climatological datasets by calculating the average monthly precipitation sum.

Second, the monthly precipitation background grids are used in inverse-distance squared weighting as follows:

To calculate precipitation P at a given grid cell x from a number of n surrounding gauges, first calculate the ratio between the value of the monthly background grid at location x and the value of the monthly background grid at the gauge location for every gauge i to create a set of multipliers $m_{i,x}$ from the gauges to the given grid cell. Hence $m_{i,x}$ would change based on the month:

$$m_{i,x} = \frac{MBG_x}{MBG_i} \tag{3.1}$$

Next, determine a set of weights based on inverse-distance squared weighting between all gauges (typically the n closest gauges) and the grid cell. This step can have a threshold for maximum distance as well. $d_{i,x}$ is the distance between gauge i and cell x:

$$w_{i,x} = \frac{1/d_{i,x}^2}{\sum_{i=1}^{n} 1/d_{i,x}^2} \tag{3.2}$$

Finally, use the weights and the multipliers to predict the precipitation at grid cell x:

$$P_x = \sum_{i=1}^{n} P_i m_{i,x} w_{i,x} \tag{3.3}$$

This formulation makes the role of the background grid clear. Precipitation measurements at locations that are expected to be dry get amplified for cells that are expected to be wet and measurements at expected wet locations are dampened for cells expected to be dry. Note that since $m_{i,x}$ has a range of $[0, \infty)$, $m_{i,x}$ is indeed a multiplier and not a weight.

For implementation we use a slightly different form. Substituting Eq. (3.1) in Eq. (3.3) and taking MBG_x out of the summation gives:

$$P_x = MBG_x \sum_{i=1}^{n} \frac{P_i}{MBG_i} w_{i,x} \tag{3.4}$$

In this form, genRE can be easily implemented with an existing inverse-distance squared weighting package. First relative precipitation, P_i/MBG_i, is interpolated with inverse-distance squared weighting and then the resulting relative precipitation field can be multiplied with the MBG to get the desired precipitation fields.

The main difference between REGNIE (*Rauthe et al.*, 2013) and genRE is that the REG-NIE method specifically uses Multiple Linear Regression (MLR) based on five explanatory variables: geographical longitude and latitude, height above sea level, exposition, and mountain slope at the stations, to derive the monthly precipitation background grids. In genRE, the background grids are calculated from existing climatological datasets.

There are two reasons for using existing climatological datasets over deriving monthly climatological datasets from raw data independently. First, as shown in this paper, using an existing climatology forces the interpolation to spatially mimic the reference climatology, which is a quality sought after by flood forecasters as calibrated models might be 'tuned' to a certain areal distribution of rainfall (*Andréassian et al.*, 2001). Second, many flood forecasting agencies are tasked with deriving areal rainfall for their domains, but have only access to a limited part of the available measurements. In many cases the resources (time, people, money to purchase data) are lacking and thus this approach offers an alternative to other makeshift solutions.

Two additional, smaller, differences between REGNIE and genRE are that we extend the use of genRE to also interpolate on an hourly basis, where REGNIE was developed to interpolate daily accumulations, and that we interpolate to each grid square in the field and do not copy station data directly to the related grid square as is done in REGNIE. The first is not a methodological difference, but an extension of the use of the algorithm for use at shorter time scales. The second decision was made because we did not see any reason to give special treatment to cells containing stations, favouring a continuous approach. An additional advantage is that the original measurements cannot be retrieved from the interpolated results, which can help navigating data sharing agreements that prohibit sharing station data, but allow sharing of the interpolated product.

Monthly background grids

The background grids are calculated from the HYRAS 2.0 gridded data set (*Rauthe et al.*, 2013). This is done simply by summation of the daily precipitation amounts for each calendar month over the full time span of the HYRAS dataset (1951–2006) and then dividing this by the total number of days. This results in one background field per calendar month, twelve in total, at the grid used in HYRAS (ETRS89-LCC with 5 x 5 km spatial resolution). The background grids are included in the Supplementary Information (S01). To calculate $m_{(i,x)}$ for each grid cell, we resampled the background fields with nearest neighbour resampling to our model grid (1.2 x 1.2 km spatial resolution and projection in UTM32N). This was repeated to calculate background grids using EOBS (*Haylock et al.*, 2008) data for the period 1995-2016 to investigate the effect of the choice of background grid on interpolation results.

The nearest neighbour method for resampling the grids to our model grid was chosen to prevent smoothing and to ensure that the background values at the station's locations would not change. However, as a consequence, multiple pixels from the 1.2 x 1.2 km grid

that fall within one pixel of the lower resolution climatological grids receive the same background grid value.

Implementation of genRE for the Rhine basin

The method is already effectively employed as interpolation method in the operational flood forecasting and drought monitoring system called 'RWsOS Rivieren', part of a clustered multi-hazard early warning system operated by the Dutch Ministry of Infrastructure and the Environment (in Dutch: Rijkswaterstaat, RWS) (*Weerts et al.*, 2008). This genRE set-up, however, was not yet evaluated for its intended purpose, namely to closely mimic the HYRAS climatology, over a longer time period.

genRE was implemented with the Delft-FEWS (*Werner et al.*, 2013) operational forecasting platform. The inverse-distance squared weights were calculated from the eight closest stations (n=8) within a maximum distance of 100km. A distance of 100km was chosen to ensure that there was full coverage for the whole basin at any time step. This is acceptable as it is close to the decorrelation distance of hourly precipitation in the Netherlands which is 90 km in winter and 30 km in summer (*van de Beek et al.*, 2012).

Specific to this Delft-FEWS implementation of genRE is how gaps in station data were handled. A gap filling step was necessary because whole sections of the observation network did fail at the same time and secondly because of how the inverse-distance squared weighting was implemented in FEWS. In our FEWS implementation the active gauge network and related weights were calculated once for each day. A gauge was included for interpolation for all days between the station's first and last record in the database. During pre-processing it was verified for each station visually if there were any large gaps during this period. If necessary, the active period was adjusted to remove unnecessary gaps. Still, stations could have missing data, reflecting operational circumstances. Incidental gaps in station data were then filled in a two step hierarchical procedure: 1) genRE itself is used to assign precipitation values for the station without data; 2) if gaps remained, relative precipitation values are interpolated linearly in time with a maximum gap of 1 time step.

Hydrological model of the Rhine basin

The hydrological model used is the HBV (Hydrologiska Byråns Vattenbalansavdelning) model concept (*Lindström et al.*, 1997a) applied to a 1.2 x 1.2 km grid as implemented in the open source hydrological modelling framework wflow (*Schellekens*, 2016). The generated runoff is routed through the river network with a kinematic wave approach. In the following we refer to this model as wflow_hbv.

The parameters for wflow_hbv were derived from an earlier HBV model, HBV-148, that is used in the daily operational forecasting suite for the river Rhine at RWS. HBV-148 consists of 148 sub-catchments where each sub-catchment is a hydrological response unit (HRU). The HRUs were parameterized through calibration with a Generalised Likelihood Uncertainty Estimation (GLUE) like procedure (*Beven and Binley*, 1992), using HYRAS precipitation as forcing data (*Winsemius et al.*, 2013a,b). To transfer parameters from HBV-148 to wflow_hbv, every grid cell in wflow_hbv inherited the parameters of the sub-catchment

in HBV-148 the cell belongs to. During preliminary testing, the resulting streamflow generation of wflow_hbv was shown to be very close to the streamflow generation of HBV-148. This gave confidence that the model behaves well enough to use it for our current purpose of investigating the influence of different precipitation forcing data on streamflow generation without the need to spatially aggregate the input forcing to the 148 sub-catchments or costly recalibration of wflow_hbv. This is not too surprising, as spatial proximity is known as a strong parameter regionalization method (*Merz and Blöschl*, 2004).

3.4 Experimental set-up

To answer our research questions we constructed several experiments. Concerning the first question, we hypothesized that the climatology of interpolated precipitation gridded datasets made with genRE resemble their donor climatology. This implies that the genRE climatology resembles the HYRAS climatology where HYRAS is used as reference and that it resembles the EOBS climatology if EOBS is used as reference. If no background grid is used, in which case genRE defaults to inverse-distance squared weighting, information on orographic effects, etc. is lost.

To test this hypothesis, genRE was applied with background grids derived from the HYRAS and EOBS climatologies, and the obtained genRE climatologies (yearly sums and monthly averages) were compared with plain inverse-distance squared weighting and the HYRAS and EOBS climatologies for the period 1997–2006. Furthermore, we compared genRE with the HYRAS dataset for daily accumulations and with the RADOLAN gauge adjusted radar product for hourly fields. The latter comparison is only qualitatitve to reflect on the spatial detail found with the present day gauge network with genRE.

We expect that as the station density increases, the influence of the background grid will diminish. For this purpose, the obtained spatial patterns of yearly precipitation accumulations are compared for two time periods of interpolation: 1997–2006 and 2006–2014. During the latter period the station density is much higher as the number of stations has increased from roughly 100 stations to 500 stations inside the basin (see also Fig. 3.2 in the data section).

To investigate the effect of gauge density in a quantitative manner, the Pearson correlation coefficient was calculated between the yearly sum grids from plain inverse-distance squared weighting and the genRE method for each year. The correlation is expected to increase with station density.

The variability in the background grid was investigated by taking a 9-year moving average of monthly precipitation amounts from the HYRAS data record of 56 years resulting in 47 possible background grids for each month. These were normalized by dividing them by the value of the background grid obtained by averaging all 56 years. Then, the 47 normalized values were used to determine the standard deviation for each grid cell for each month. The 9-year moving average was chosen to reflect the time span over which the results were analysed, which is 10 years (1997–2006) and 9 years (2006–2014).

Additionally, minimum and maximum background grids were calculated for each month by taking the minimum and maximum of all 47 possible background grid values for each pixel, respectively.

Finally, the hydrological model is fed with both genRE precipitation and HYRAS precipitation. The modelled discharge for each location is extracted from the model grid by determining on which pixel the streamflow gauge is located. Where the gauge coordinates did not coincide with the right streamflow pixel, as determined by visual comparison between the locations of streams in openstreetmap and the stream network from the model, the gauge coordinates were adjusted to match the right stream in the model. The hydrological impact of choosing the HYRAS or genRE data set as precipitation forcing over the period 1997–2006 is expressed by calculating the Kling-Gupta Efficiency (KGE) (*Gupta et al.*, 2009). To exclude sites where the model performed non-behaviorally and thus to limit the effect of model set-up on our analysis of the effect of precipitation forcing, we only analyze pairs for which the model results were deemed behavioral, expressed as a KGE > 0.5 for at least one of the applied forcings. The model runs start on 01-06-1996 and the first half year is used as spin-up period.

The Kling-Gupta efficiency is calculated by:

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \tag{3.5}$$

With,

$$\alpha = \frac{\sigma_m}{\sigma_o}$$

$$\beta = \frac{\mu_m}{\mu_o}$$

In which r is the linear correlation coefficient between modelled and observed values, σ the standard deviation and μ the mean of modelled (m) and observed (o) flows.

3.5 Results

Climatological evaluation of genRE approach

To show the influence of the genRE method, three interpolation schemes were used: genRE with HYRAS as background grid, genRE with EOBS as background grid and plain inverse-distance squared weighting. The results were compared with their respective reference climatologies. Following the rationale behind the genRE approach, the expectation was that the interpolated products would closely match the spatial patterns found in their respective climatology reference.

The genRE products match indeed their reference climatologies, especially when compared with plain inverse-distance squared weighting. The influence of the genRE method is clear in that Figures 3.3b and 3.3e resemble their reference climatologies in Figures 3.3a and 3.3b. There are clear differences in spatial patterns of yearly average precipitation accumulations between methods and between reference climatologies.

The genRE, IDW and EOBS climatologies are significantly different from the HYRAS climatology (Fig. 3.4). At first sight, Fig. 3.3 seemed to indicate that, even for the period (1997-2006) with relatively low network density, the yearly patterns found in HYRAS are well emulated with the genRE method. However, the genRE method results in 4% less rainfall over the whole Rhine basin compared to HYRAS over the period 1997–2006 (Fig. 3.4).
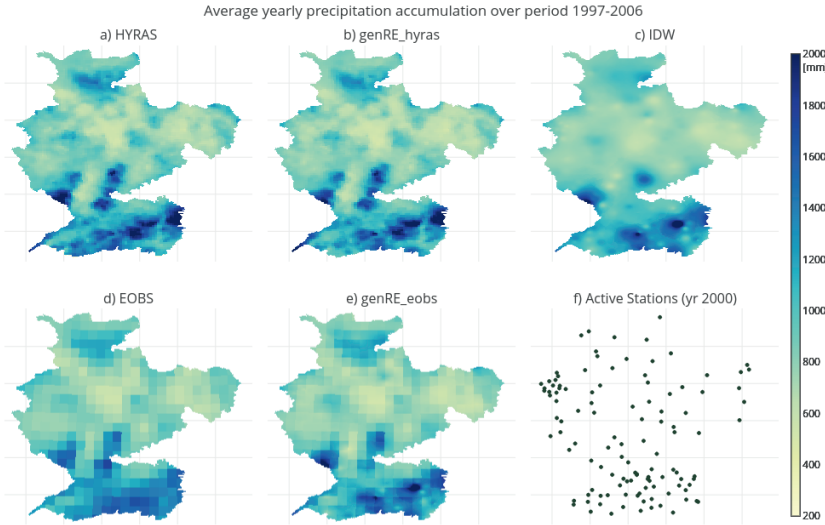
Figure 3.3: *Spatial distribution of average yearly accumulations over the period 1997-2006 obtained with:* **a**, *HYRAS reference climatology;* **b**, *genRE approach with HYRAS as background grid;* **c**, *plain inverse-distance squared weighting;* **d**, *EOBS reference climatology;* **e**, *genRE approach with EOBS as background grid. The last sub figure (***f***) shows the state of the gauge network for a representative year (year 2000).*

The largest biases are concentrated in a number of regions. Firstly, large biases are seen in Luxembourg and southern Switzerland (Fig. 3.4a), where genRE yields up to 30% less precipitation than HYRAS. Secondly, there are a number of distinct points (e.g. clear red and blue dots in Switzerland) where genRE underestimates or overestimates precipitation compared to HYRAS, which seem to be related to locations where individual stations in genRE have a large weight because of the lesser station density compared to HYRAS. Although locally the differences between genRE and HYRAS are substantial, those differences are smaller than those between IDW and HYRAS (Fig. 3.4b).

The EOBS climatology has the same spatial distribution of precipitation as the HYRAS climatology and genRE, but much more diluted due to the coarser spatial grid as rain rates are a function of scale (*Tan et al.,* 2017). The effect of averaging over larger regions becomes especially significant when the EOBS climatology precipitation values are projected on the higher resolution grid of genRE (as would be the case when using EOBS data with a high resolution hydrological model). Although the bias of the EOBS climatology compared to the HYRAS climatology is small, large local differences due to the spatial averaging are visible, following a raster-like pattern (Fig. 3.4c). Local differences between the HYRAS and EOBS reference climatologies are much larger than between HYRAS and 'emulated HYRAS' with the genRE method (Fig. 3.4a). A raster-like pattern is also, but less, visible in the other sub-figures, which is caused by resolution differences between the genRE and climatological HYRAS grid.
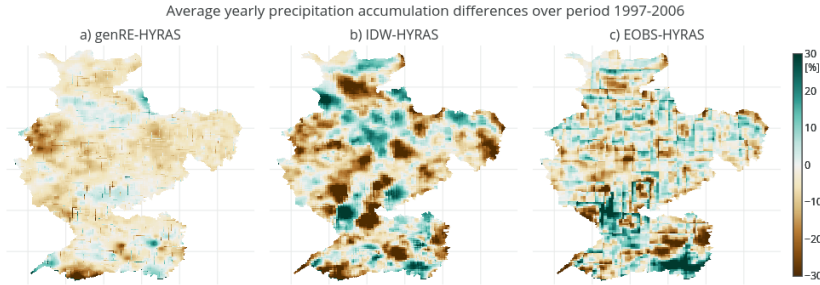
Figure 3.4: *Differences (%) between methods in terms of average yearly climatology projected on the high resolution (1.2x1.2km) model grid.* **a**, *genRE (with HYRAS) − HYRAS* **b**, *IDW − HYRAS* **c**, *EOBS − HYRAS.*

The differences in climatology between HYRAS and genRE persist also when evaluating the grids on multi-year monthly averages (Fig. 3.5), with the largest average biases during autumn and winter. Based on the spatial pattern we noted structural underestimation or over-estimation of precipitation in the same areas as before.

For the 1997–2006 period genRE underestimates precipitation sums compared to both HYRAS and EOBS in most sub-basins for January, February, October–December. Additionally, genRE underestimates precipitation sums compared to HYRAS in the Moselle sub-basin for the whole year.

In Figure 3.6 we compare seven of the most extreme rainfall events, which gave the highest areally averaged precipitation over the different main sub-basins according to HYRAS in the period 1996–2006. The matching patterns are evident, although HYRAS has a sharper delineation of local peaks and between precipitating and non-precipitating areas. For example, for the event of 18-sep-2006 genRE shows clearly two maxima in the Vosges area, while HYRAS shows a more homogeneous field, due to the higher station density underlying the latter. Likewise, for the event of 07-jul-1999 HYRAS distinguishes between two local peaks in the northern part, whereas genRE merges these peaks into one general area of intense rainfall.

The differences between genRE and HYRAS are quantified for three events in Figure 3.7: two of the six extreme events from Figure 3.6, and one randomly chosen less extreme precipitation event (05-oct-2002). Evaluating on a pixel by pixel basis yields a high coefficient of determination (ρ^2) and a low coefficient of variation (CV) of the residuals for the two extreme events. For less extreme events the coefficient of determination decreases and the coefficient of variation increases.

Large deviations from the 1:1 line can be traced back to inclusion or exclusion of individual gauges. For example, locations where HYRAS includes gauge locations with high intensity rainfall which are not available on an hourly basis and thus not included in genRE, lead to horizontal stripes in the upper left-hand part of the scatter density plots. The horizontal stripes are caused by the resolution difference; genRE produces 25 values for each HYRAS value.

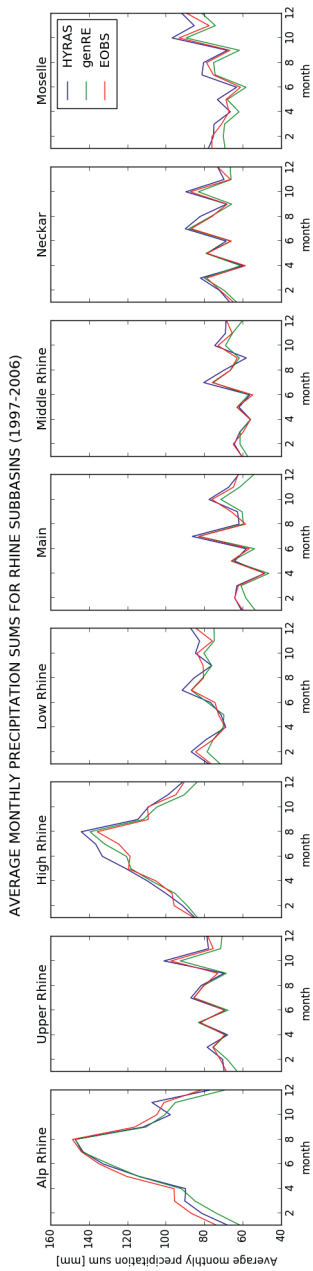In a similar manner, genRE incorporates a smoothing effect, which causes the precipita-

Figure 3.5: Average monthly precipitation accumulations for major Rhine sub-basins for genRE, HYRAS and EOBS.
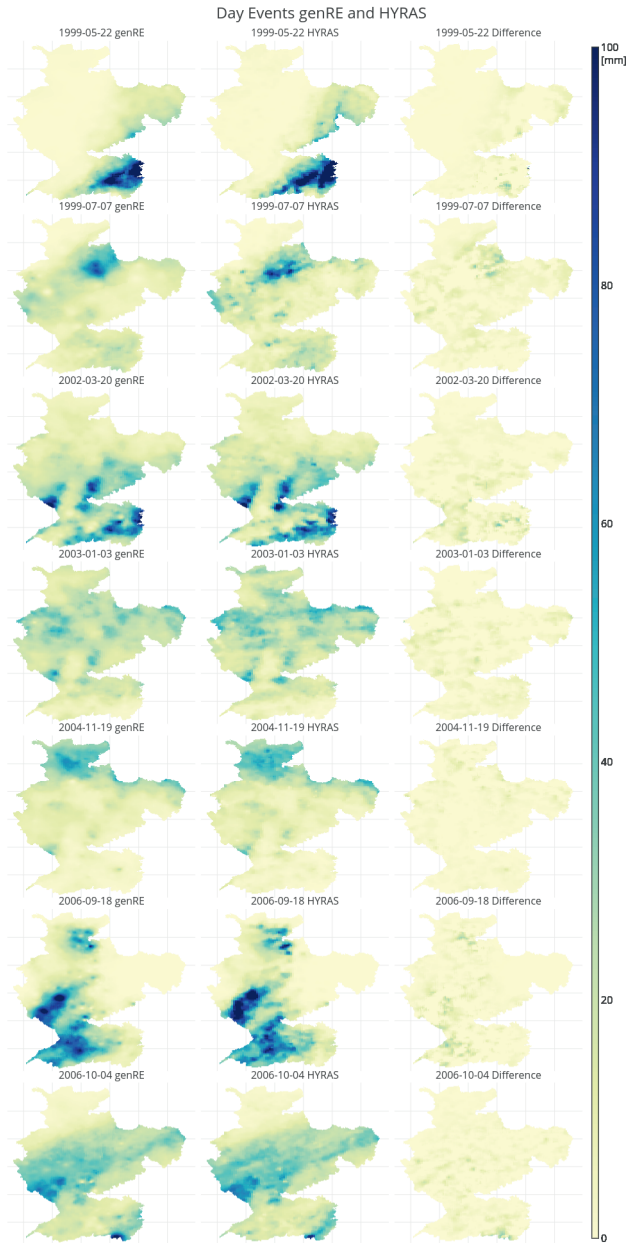
Figure 3.6: Daily accumulations of seven extreme precipitation events from genRE (left) and HYRAS (right).
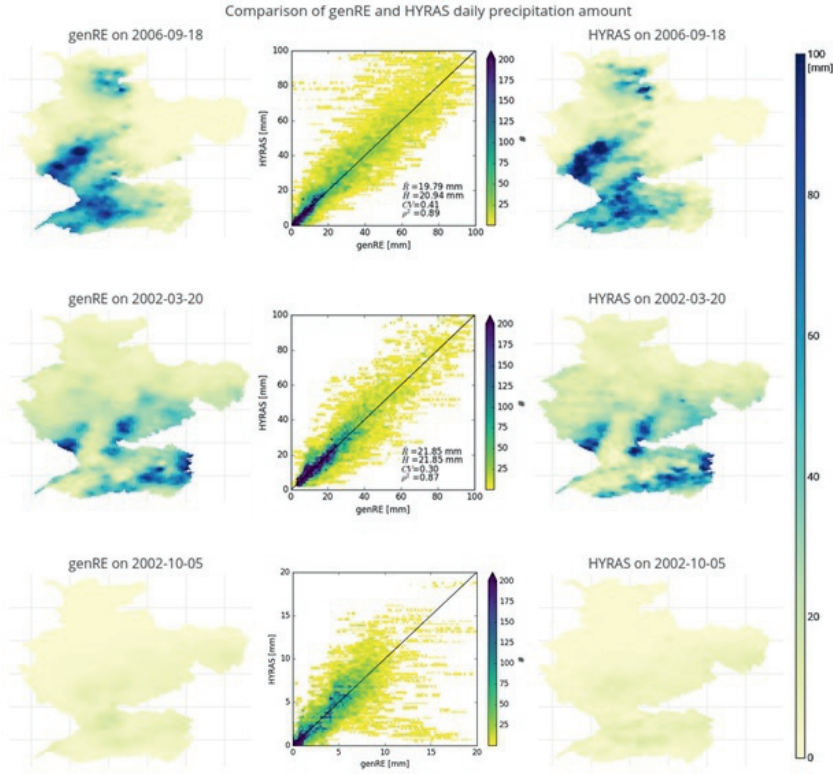
Figure 3.7: Maps and scatter density plots of areal precipitation according to genRE and HYRAS for three days. The shown statistics are the basin average for genRE and HYRAS (R and H), the coefficient of variation (CV) and the coefficient of determination (ρ^2).

tion to extend over a larger area. In the scatter density plot for the event of 05-oct-2002 (Fig. 3.7), we see most of the deviation from the 1:1 line in the lower right-hand part of the figure, which means that there are areas which yield a relatively high precipitation value in genRE, but not in HYRAS. This smoothing behaviour is a trait of inverse-distance squared weighting. The difference in smoothing can be attributed to the number of stations on one hand, and the fact that we use more stations for each weighting in genRE (eight closest stations) than in HYRAS (four closest stations) on the other, which we did because of the limited station density of our base data.

So far, genRE has been compared to the existing HYRAS and EOBS climatological data sets. However, genRE has been designed to provide hourly precipitation fields for supporting operational forecasting in the Rhine basin. To evaluate these fields, genRE fields were qualitatively compared with RADOLAN raingauge corrected weather radar based imagery.

Figure 3.8 shows the temporal evolution of a precipitation field passing over the Rhine

basin in 2014. The location and timing of genRE and RADOLAN precipitation coincide. The precipitative areas in genRE do not extend over a larger area than in RADOLAN. This suggests that the station density underlying genRE has reached a point where it is sufficient (in relation to the areal scale of the Rhine basin) to capture the most relevant details of areal precipitation as shown by the RADOLAN merged gauge-radar product.

Influence of station density on genRE and variability of the background grid

To estimate the influence of the station density on the genRE method the spatial patterns were compared for two time periods of interpolation: 1997–2006 (Fig. 3.3) and 2006–2014 (Fig. 3.9). During the latter period the station density is much higher as the number of available stations jump from 150 to 400 during the years 2005-2006. As indicated by Fig. 3.10, there is no change in spatial distribution of rainfall between the two periods.

There is notably less difference between the spatial patterns found with the genRE approach and inverse-distance squared weighting than in the previous period with lower station density (Fig. 3.9). Some features, however, remain: the second precipitation maximum in the Vosges area (annotated with 1 in Fig. 3.9b) and two areas in the Alps (annotated 2 and 3).

To quantify this, the Pearson correlation coefficient was calculated between the grids calculated with plain IDW and the genRE method (with HYRAS as background grid) for each year. The correlation between IDW and genRE did indeed increase steadily with the number of stations included in the interpolation, from 0.85 with 110 stations to 0.96 with 480 stations between the years 1997 and 2015 (Fig. A.7).

The large variability of monthly precipitation amounts in the Rhine basin, as recorded in the reference climatology, makes that the choice of years to calculate the background grid has significant impact on the absolute values of the background grid. The maximum range, i.e. the difference between the minimum and maximum value for one grid point over all 47 background grids, equaled 1.42 times the average background value. In general, the range was between 0.3 and 1 times the average value (Figures A.1-A.6 in the appendix).

The genRE method, however, as shown in Eq. (3.1), is only affected by the local spatial gradients of precipitation in the background grid and not by absolute background values. For each monthly background grid from any 9-year average the spatial pattern, indicating where it is relatively wet and where it is relatively dry, remained the same (Fig. 3.10). However, there are differences in local gradients leading to multipliers that have the same direction (amplifying or dampening) for each possible background grid but different magnitudes.

Resulting effects on discharge simulations

The choice of forcing dataset has distinct implications on model performance as expressed by the KGE and its decomposition in correlation (cc), bias of the variance (alpha) and bias of the mean (beta) (Figure 3.11). Notably, HYRAS scores better for Switzerland, while genRE leads to unexpected improvements in the Moselle. These are the same areas as those where genRE and HYRAS showed the strongest systematic differences in yearly
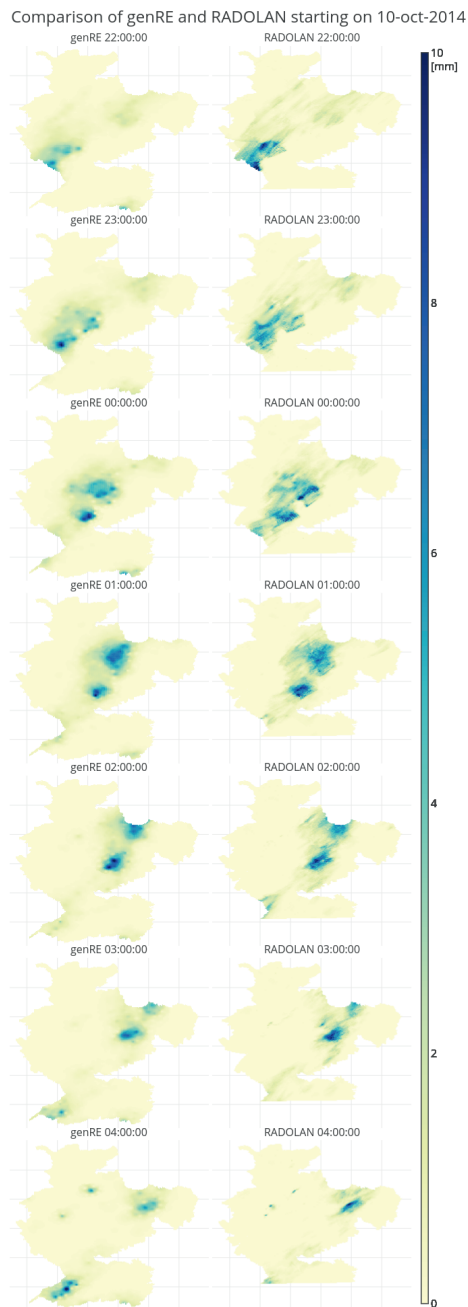
Figure 3.8: The passing of a precipitation event on 10-11 October 2014 as detected by genRE (left) and RADOLAN (right).

Average yearly precipitation accumulations over period 2006-2014
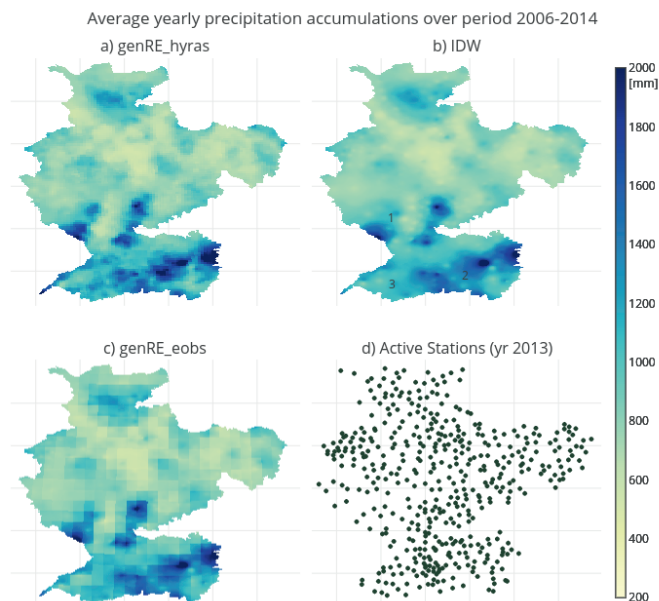
Figure 3.9: Spatial distribution of average yearly accumulations over the period 2006–2014 obtained with: **a***, genRE approach with HYRAS as background grid;* **b***, plain inverse-distance squared weighting;* **c***, genRE approach with EOBS as background. The last sub-figure shows the current state of the gauge network* **d***.*

precipitation amounts. The difference between HYRAS and genRE KGE scores for the downstream reaches of the Rhine are small (δKGE $\leq$ 0.05).

Looking at the KGE decompostition, there is little difference in correlation score. Larger differences are found in the relative variability (Alpha) and relative Bias (Beta). In the main reach of the Rhine the variability of flows is slightly underestimated by genRE and slightly overestimated with HYRAS leading to mixed results in the difference column. Variability is too low in Switzerland and too high in the Moselle area, for both forcing datasets (Fig. 3.11, Alpha).

There is a negative bias in mean discharge for both forcing datasets for the Rhine river and Switzerland, but this negative bias is smaller for HYRAS than for genRE. Streamflow mean is also negatively biased in Luxembourg for both forcing datasets. HYRAS leads to overestimation of mean streamflow in the Moselle area and the Lower Rhine basin (Fig. 3.11, Beta).

The ambiguity of the calculated difference scores (right column, Figure 3.11) is exemplified by the four hydrographs shown in Figure 3.12. For the same basin and in a short time period, both datasets can result in under and overshoot of modelled discharge peaks, with little to distinguish between the two. Note that these stations were not used for model calibration.
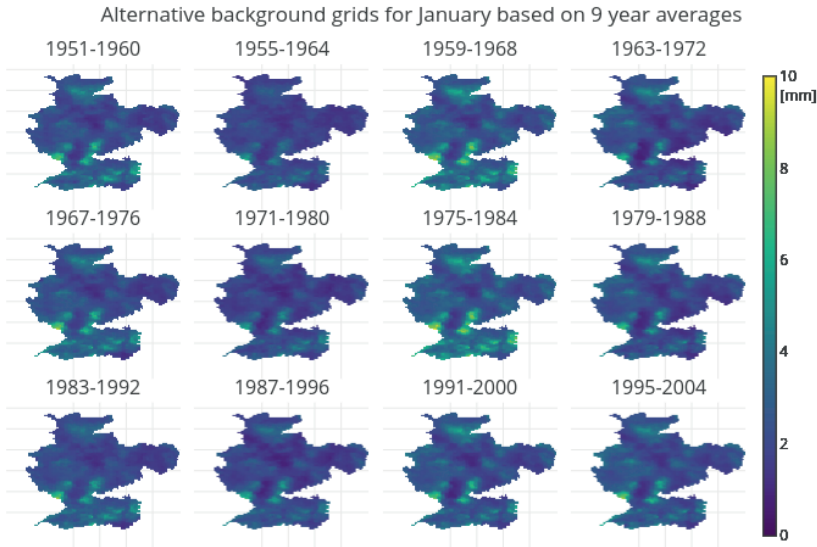
Figure 3.10: 12 out of 47 background grids for January calculated from the HYRAS climatological dataset.

3.6   Discussion

Discussion of experimental set-up and results

Table 3.1 gave an overview of some important properties of the HYRAS and EOBS dataset in comparison with the dataset created with the genRE approach. As can be seen, the reference datasets differed in temporal and spatial resolution. Therefore a choice had to be made on which resolution the comparisons were made.

A choice often made in this is to aggregate the higher resolution product to the lower resolution product (*Rios Gaona et al.*, 2016, 2015). *Chen and Knutson* (2008) in particular argue that comparing precipitation datasets at different resolutions is unfair. Aggegrating to the native resolutions of the reference datasets would have meant, however, that any difference due to the higher resolution of genRE would be ignored, especially in regions with strong topography. We focused here on the application of the high resolution product in a similar high resolution model, and thus evaluated the performance of the gridded rainfall products on the hydrological model grid. As a result of this approach, we noted that the EOBS spatial resolution is too coarse for high resolution hydrological modelling, with very large local differences which exceeded the differences between genRE and HYRAS.

Concerning the differences between genRE and HYRAS, two methodological factors were investigated, but not mentioned in the paper in more detail so far, which could contribute to the large biases in the affected areas: the choice of using the eight instead of the four
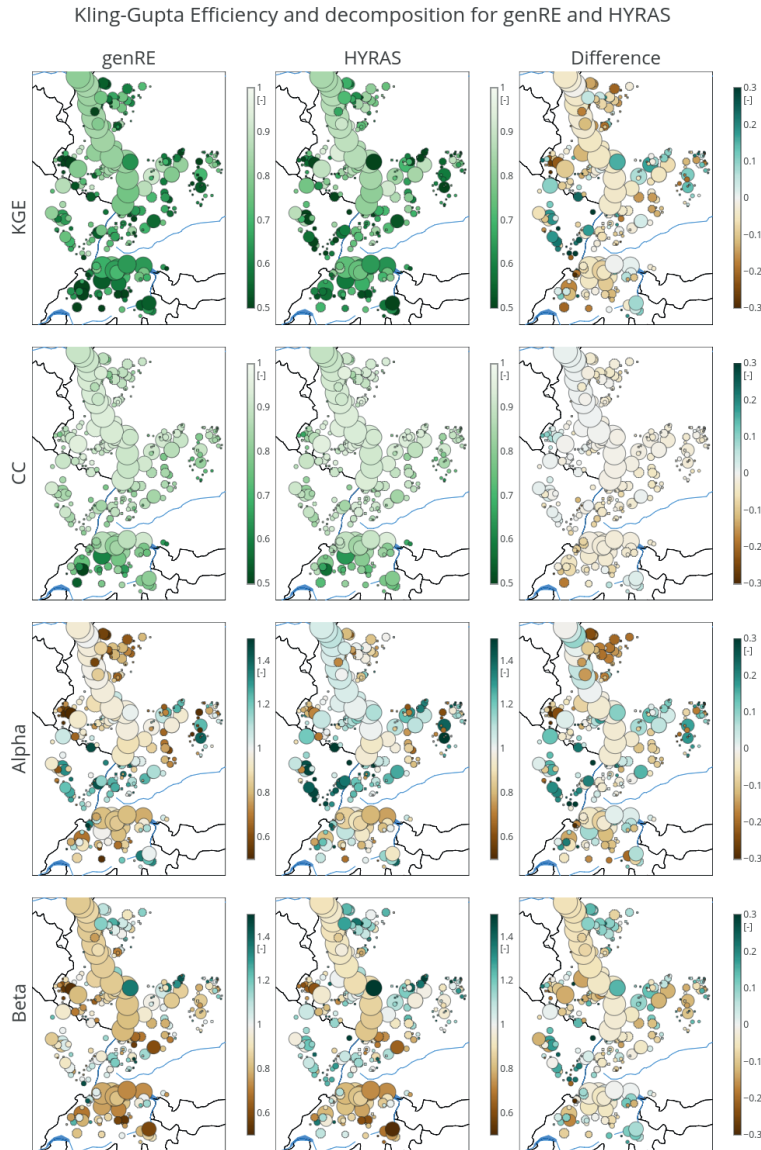
Figure 3.11: *Kling-Gupta Efficiency (KGE) scores and their decomposition (CC: correlation coefficient, Alpha: measure of relative variability, Beta: measure of relative bias) for 346 discharge gauging stations for two interpolated precipitation forcings (genRE and HYRAS) and their difference. In the difference column positive values indicate more optimal scores for genRE over HYRAS and negative scores the opposite.*
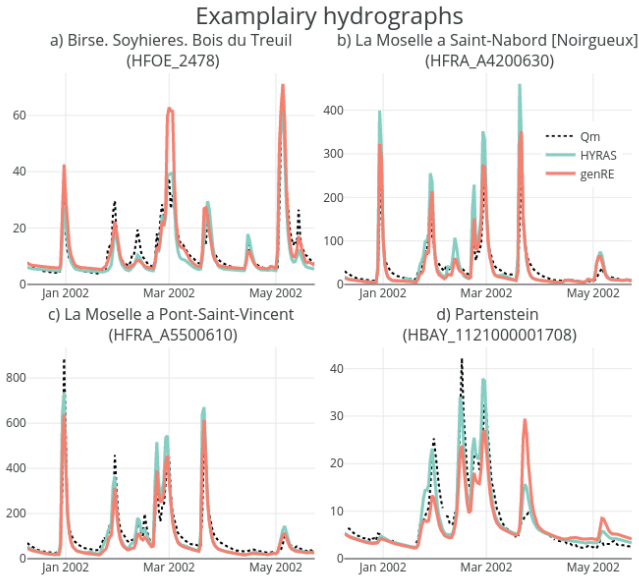
Figure 3.12: Four exemplary hydrographs of measured (black dots) and modelled streamflow in the Rhine basin for two different forcings (genRE: orange, HYRAS: green). The locations (a,b,c,d) are notated in Figure 3.1.

closest stations for inverse distance interpolation and the inclusion of a large number of stations just outside the Rhine basin domain which were not included in the HYRAS interpolation. Applying these changes led to even larger negative biases in Luxembourg and to insignificant change elsewhere. The lack of evidence that methodological differences are the cause of the found biases combined with the fact that the largest negative bias is during winter hint that the biases might be partially snow related as snow measurements are very sensitive to the equipment used (*Rasmussen et al.*, 2015). More generally, malfunctioning gauges can lead to underestimation of precipitation means as malfunctioning gauges mostly report less rainfall, adding to the need for automated quality control (*Steiner et al.*, 1999).

Routing the genRE and HYRAS precipitation fields through a gridded hydrological model showed some complex interaction between forcing and model bias. Even though we assumed that the HYRAS reference precipitation climatology is more reliable than our genRE product, the hydrological model is less biased and scored higher KGE scores for the Moselle area. Overall, the wflow_hbv model, here derived without recalibration from the HBV-148 model, can be improved upon.

Outlook

There are multiple deterministic and (geo)statistical methods to choose from for interpolating precipitation gauge data. After an extensive literature review *Ly et al.* (2013) concluded that geostatistical methods seem preferable for the interpolation of monthly and annual rainfall, but that for daily rainfall geostatistical methods and IDW provide comparable results. This conclusion supports the rationale behind the genRE method, namely to derive the spatial relationships between grid cells on the monthly time scale and use IDW for the hourly interpolation. It is also the approach used in the method behind EOBS. The novelty of the genRE method is not the interpolation method itself, but the use of existing climatological grids as secondary variable to guide the interpolation, thereby extending those climatological grids in real-time without the need for additional data. In that respect, there are multiple ways to build upon both the idea of combining monthly climatologies with daily or hourly anomalies, and the idea of extracting geostatistical information from climatological datasets. For example, the climatological grids could be used to derive the parameters for geostatistical methods, or the background climatology could be used directly as a secondary variable in, for example, Kriging with External Drift (KED). Still, the simplicity of the genRE approach is favorable for operational circumstances. For example, the European Flood Awareness System (EFAS) chose IDW over geostatistical methods to maintain their operational gridded precipitation product for robustness (*Ntegeka et al.*, 2013).

There are also several ways to further improve the dataset we created with the genRE method. Firstly, there is no substitute for more, and more accurate, measurements, including automated verification. The latter is difficult as a distinction needs to be made between real outliers (extreme events) and errors. Here, static thresholds were used, but there are methods in development for more dynamic automated verification (*Blenkinsop et al.*, 2017).

A legitimate concern is the validity of the background grids under changing atmospheric circulation patterns. For western Europe, both measurement and model based studies have shown evidence of changing circulation patterns (*Bardossy and Caspary*, 1990) and precipitation amounts (*Frei et al.*, 2006, e.g.). For the Rhine basin in particular, different magnitudes of change of seasonal average precipitation were found for topographical features in the basin, most notably for the Rhine valley (*Hurkmans et al.*, 2009a; *Photiadou et al.*, 2016, Fig. 6). As HYRAS is meant to be updated in the near future, it would be wise to calculate new background fields when these become available.

The variability in the background grid derived from different subsets of the original data could be used to construct ensembles, although this would probably count as 'adding uncertainty in an ad-hoc manner' as criticized in (*Newman et al.*, 2015). In this respect it should be noted that the whole genRE method is an 'ad-hoc' method and is not meant to replace established methods to create climatological datasets, but rather to fill the gap between those relatively static exercises and the more dynamic reality found in transboundary operational hydrological forecasting.

Finally, the usefulness of the proposed method could be further demonstrated by applying the method in other (transboundary) river basins and with other climatological gridded datasets. For the Rhine basin, progress is underway to deliver interpolated products for temperature and potential evaporation in the same spatial and temporal resolution.

3.7   Conclusions

In this paper a method to extend high quality gridded precipitation climatology datasets in near real-time for hydrological forecasting purposes was described. It is called generalized REGNIE (genRE), linking its origins to the original REGNIE method used in creating the HYRAS climatology dataset (*Rauthe et al.,* 2013). The aims of this paper were: (1) to investigate if the genRE method can indeed be used to mimic and extend climatological precipitation datasets in near real-time using (sparse) real-time measurements; (2) create a high resolution (near real-time) gridded precipitation dataset for the purpose of high resolution hydrological modeling of streamflow and (re)forecasts in the Rhine basin; (3) evaluate the precipitation interpolation procedure used, in a hydrological sense.

The genRE method yields interpolated fields that closely match the reference climatology used in the method on the yearly, monthly, daily and hourly time scale. The method has the largest added value when employed with sparse gauge networks, which suits the purpose of the method. The variation in the background grid for HYRAS was estimated to be of the same order of magnitude as the relative differences found between the genRE method and the HYRAS climatology. This shows the dependence of the genRE method on the selected reference climatology and the associated natural variability.

Our work resulted in a gridded 1.2x1.2 km resolution hourly precipitation dataset for the period 1996–2016, to be updated in real-time, covering the entire Rhine basin. The data set is unique in spatial and temporal resolution compared to climatological data sets such as HYRAS and EOBS. Furthermore, it uses a simple interpolation method and can be updated in real-time. This precipitation dataset successfully mimics HYRAS in terms of matching spatial patterns, showing a low yearly bias and a high correlation coefficient for daily accumulations. Differences in simulated discharges using HYRAS and genRE as expressed in KGE scores are overall relatively small, but can be significant for individual peak flow simulations.

Chapter 4

Contribution of potential evaporation forecasts to 10-day streamflow forecast skill for the Rhine river

Medium term hydrologic forecast uncertainty is strongly dependent on the forecast quality of meteorological variables. Of these variables, the influence of precipitation has been studied most widely, while temperature, radiative forcing and their derived product potential evapotranspiration (PET) have received little attention from the perspective of hydrological forecasting. This study aims to fill this gap by assessing the usability of potential evaporation forecasts for 10-day-ahead streamflow forecasting in the Rhine basin, Europe. In addition, the forecasts of the meteorological variables are compared with observations.

Streamflow reforecasts were performed with the daily wflow_hbv model used in previous studies of the Rhine using the ECMWF 20-year meteorological reforecast dataset. Meteorological forecasts were compared with observed rainfall, temperature, global radiation and potential evaporation for 148 subbasins. Secondly, the effect of using PET climatology versus using observation-based estimates of PET was assessed for hydrological state and for streamflow forecast skill.

We find that: (1) there is considerable skill in the ECMWF reforecasts to predict PET for all seasons, (2) using dynamical PET forcing based on observed temperature and satellite global radiation estimates results in lower evaporation and wetter initial states, but (3) the effect on forecasted 10-day streamflow is limited. Implications of this finding are that it is reasonable to use meteorological forecasts to forecast potential evaporation and use this is in medium-range streamflow forecasts. However, it can be concluded that an approach using PET climatology is also sufficient, most probably not only for the application shown here, but for most models similar to the HBV concept and for moderate climate zones.

4.1   Introduction

Hydrologic forecasting has the aim to predict the future state of important hydrologic fluxes, most notably streamflow. Throughout the process of forecasting, from model set-up via initial state estimation to forecast run, meteorological forcing is a key component. Precipitation is known to be the main driver of hydrological processes and most of the forecast uncertainty is attributed to inaccurate precipitation forcing (*Cuo et al.*, 2011; *Pappenberger et al.*, 2005). As a consequence, most attention has been given to the accuracy of precipitation forecasts. See for example the review of *Cloke and Pappenberger* (2009).

Evaporation is a result of the interaction between meteorological forcing and, physical and physiological, processes at the land surface. Meteorological forcing provides the potential energy (potential evaporation or PET) for evaporative processes to take place. There are many formulas to estimate the potential energy available for evaporation, which can be divided in three types of formulas based on their data requirements (*Xystrakis and Matzarakis*, 2011; *Xu and Singh*, 2002): Temperature-based (e.g. Hargreave equation, Hammon's equation), radiation-based, and combined methods (e.g. Hansen's equation, Turc's equation, Makkink's equation). From an operational viewpoint the different types of formulas result in different demands on data availability.

Constraints on data availability have led to additional approximations for potential evaporation. A common approximation is the calculation of a monthly potential evaporation climatology or PET demand curves (*Bowman et al.*, 2016). This climatology is then used as driver for both historic potential evaporation and future potential evaporation.

Hydrological models have proven to be insensitive to the difference between variable potential evaporation forcing and climatological monthly potential evaporation forcing with respect to the model's potential to estimate streamflow after calibration (*Andréassian et al.*, 2004; *Oudin et al.*, 2005a,b). However, in forecasting, different choices in the handling of forcing data can be made between the historic update step and the forecast step, while the hydrological model, as a rule, remains the same. It therefore remains relevant to understand how a single model reacts to potential evaporation forcing. Insensitivity to the type of potential evaporation during the process of calibration does not mean that a model is insensitive to the form of potential evaporation input.

As mentioned above, there has been little attention of the forecast skill of the secondary forcing variables temperature and radiation in the hydrological context of potential evaporation. Furthermore, there is an easy and often used practice of avoiding potential evaporation forecasts by using a potential evaporation climatology. Therefore, the objective of this study is to assess to what extent potential evaporation forecasts can contribute to streamflow forecast skill.

This question is evaluated for the Rhine basin in Europe (Fig. 4.1). The Rhine is one of the basins currently employed as case study for the IMproving PRedictions of EXtremes (IMPREX) project, which aims to improve predictions and management of hydrological extremes through climate services (*van den Hurk et al.*, 2016).

Several studies already directly addressed some aspects of operational ensemble flow forecasts in the Rhine. *Renner et al.* (2009) showed that at the time meaningful hydrological ensemble forecasts could be produced up to a 9 day leadtime for the Rhine river based on ECMWF ensemble meteorological forecasts. *Reggiani et al.* (2009) used a Bayesian en-

semble uncertainty processor to improve the assessment of uncertainty in the ensemble forecast. *Terink et al.* (2010) applied downscaling techniques to ERA15 ECMWF reanalysis data to develop a downscaling strategy for regional climate models (RCMs). *Verkade et al.* (2013) developed post-processing techniques to improve the precipitation and temperature ECMWF forecasts for the hydrological model. *Photiadou et al.* (2011) compared two historical precipitation datasets and assessed the influence of precipitation datasets on model results. Recently, *van Osnabrugge et al.* (2017) developed a high resolution hourly precipitation dataset for use with gridded hydrologic models.

To answer the research question model experiments are performed, but first the data and hydrological model are presented (Section 2). Second, the model experiments are described, which also partitions the main question into three subquestions (Section 3) which are subsequently answered (Section 4). The paper concludes with a discussion on the results in the wider context of evaporation modelling in hydrologic forecasting and the conclusions (Section 5).

4.2 Data and model

Observational data has been preprocessed for use with a grid based hydrological model. The data was processed with hourly time resolution, on a 1.2x1.2 km grid spatial resolution, and for the period mid 1996 through 2015. All source data to derive the gridded estimates comes from sources that supply their data in near real-time making the datasets suitable for operational forecasting. For this study all data was aggregated to a daily time step. The hourly datasets are downloadable through the 4TU data centre (*van Osnabrugge*, 2017, 2018).

Precipitation

For this study the precipitation dataset from *van Osnabrugge* (2017) is used. The precipitation data set has been derived using the genRE interpolation method based on ground measurements and the HYRAS (*Rauthe et al.*, 2013) climatological precipitation dataset (*van Osnabrugge et al.*, 2017).

Temperature

Temperature observations (1996-2016) are interpolated on the same 1.2x1.2 km grid as the precipitation data. Temperature is derived from interpolation of ground measurements with correction for elevation using the SRTM digital elevation model (*Farr et al.*, 2007) and standard lapse rate as follows:

To calculate temperature T_x at a given grid cell x from a number of n surrounding stations, determine a set of weights based on inverse-distance squared weighting between all stations (typically the n closest stations) and the grid cell. This step can have a threshold for maximum distance. $d_{i,x}$ is the distance between station i and cell x:

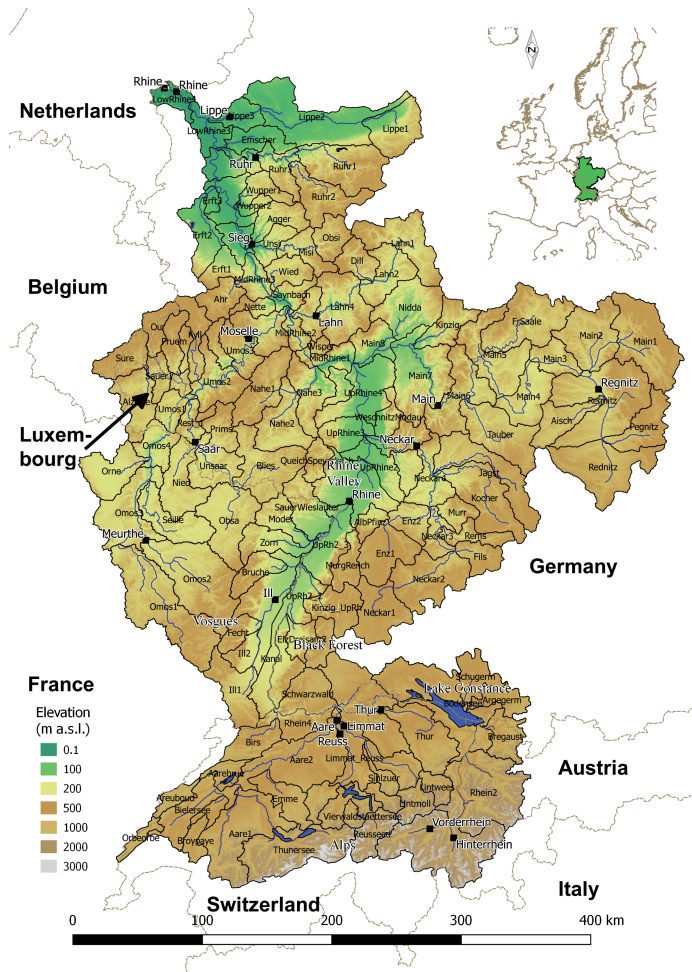$$w_{i,x} = \frac{1/d_{i,x}^2}{\sum_{i=1}^{n} 1/d_{i,x}^2} \qquad (4.1)$$

Figure 4.1: Map of the Rhine basin, Europe. Black lines delineate 148 subbasins used in the analysis of the meteorological forecast skill. Square markers show the locations used for forecast skill analysis.

Second, interpolate the measured temperature $T_{m,i}$ with the weights as with standard inverse-distance squared interpolation:

$$T_{m,x} = \sum_{i=1}^{n} T_{m,i} w_{i,x} \tag{4.2}$$

Third, calculate the temperature lapse correction term $T_{\gamma,x}$ as the weighted difference between the height of the grid cell H_x and the height of the considered stations H_i multiplied with the lapse rate γ.

$$T_{\gamma,x} = \gamma \left(\sum_{i=1}^{n} (H_i - H_x) w_{i,x} \right) \tag{4.3}$$

Note that $T_{\gamma,x}$ is static for a fixed configuration of the measurement network if γ is taken to be a constant. In this study the configuration of the measurement network changed based on the number of reporting stations at each time step. A constant lapse rate was assumed: $\gamma = 0.0066[°Cm^-1]$.

The final temperature estimate for grid cell x is obtained by adding $T_{\gamma,x}$ and $T_{m,x}$:

$$T_x = T_{\gamma,x} + T_{m,x} \tag{4.4}$$

Downwards shortwave surface radiation flux

The availability of solar radiation measurements at the surface has proven to be spatially and temporally inadequate for many applications, with remotely sensed solar radiation products having the largest potential to remedy this (*Journée and Bertrand*, 2010). Remotely sensed solar radiation estimates from the Land Surface Analysis Satellite Application Facility (LSA-SAF) were found to be in closer agreement with ground observations than reanalysis datasets such as the Système d'Analyse Fournissant des Renseignements Atmosphériques à la Neige (SAFRAN) reanalysis (*Carrer et al.*, 2012) and ERA-Interim (*Jedrzej et al.*, 2014).

For this study, downward shortwave radiation is resampled and merged from the EUMET-SAT Surface Incoming Solar Radiation (SIS) (*Mueller et al.*, 2009) and Downward Surface Shortwave Flux (DSSF) (*Trigo et al.*, 2011) products from the Climate Monitoring Satellite Application Facility (CM-SAF) and LSA-SAF, respectively. Gaps in the satellite data are filled with the ERA5 surface solar radiation downwards (ssrd) parameter from the 4D-VAR reanalysis (Copernicus Climate Change Service, 2018). ERA5 was found to have comparable mean bias with satellite-derived products for inland stations (*Urraca et al.*, 2018).

In earlier research it has been shown that LSA-SAF (2005-current) and CM-SAF (1983-2005) can consistently be merged into one time series (*Jedrzej et al.*, 2014). The products of the different SAFs are comparable in terms of bias and standard deviation (*Ineichen et al.*, 2009).

Makkink potential evaporation

There are different approaches in making use of remotely sensed data to calculate evaporation. One branch of research aims to calculate actual evapotranspiration directly from

satellite imagery (*Su*, 2002). Applications range from estimating the global evaporation flux (*Mu et al.*, 2011), water resources management (*Bastiaanssen et al.*, 2005) and constraining model parameters for a gridded model (*Immerzeel and Droogers*, 2008).

For operational use, PET estimates can be derived from satellite data only, or from a combination of satellite imagery and ground measurements. *Bowman et al.* (2017) explored the use of MODIS to provide a daily PET, both as dynamic PET (*Spies et al.*, 2015) and PET climatology (*Bowman et al.*, 2016) for a gridded and lumped version of the Sacramento Soil Moisture Accounting (SAC-SMA) model. The model was recalibrated for each PET input. No configuration with MODIS derived PET showed consistent improvements across all basins in their case study. Still, it was concluded that the combination of dynamic PET in combination with a gridded model had the best overall results (*Bowman et al.*, 2017).

A disadvantage of using satellites such as MODIS is their temporal coverage which is restricted to a single overpass at a set time each day giving one instantaneous value. This can be resolved by assuming a sinusoidal development of PET over the day (*Kim and Hogue*, 2008), but the limitation is clear. This disadvantage is resolved by using geostationary satellites. For example, *Jacobs et al.* (2009) used solar radiation from NOAA GOES geostationary satellite in combination with ground observations to calculate daily PET with the Penmann-Montheith equation.

Here, potential evaporation is calculated from geostationary satellite radiation estimates and ground observations of temperature with the method proposed by Makkink (*Makkink*, 1957), which is applicable with remote sensed radiation estimates (*de Bruin et al.*, 2016). PET calculated with Makkink's equation is a reference crop evapotranspiration, which means that crop factors apply determined by the hydrological model. In the set-up of our hydrological model the crop factor was determined by land-use. A crop factor of 1.15 is applied to the forested areas and 1.0 to all others.

The reasons for choosing the Makkink equation are that 1) it only needs radiation and temperature, for which gridded estimations are available and 2) the Makkink equation is used by the Royal Netherlands Meteorological Institute (KNMI) so that the work presented here is compatible with ongoing local research (*Hiemstra and Sluiter*, 2011).

The potential evaporation is calculated based on air temperature T [$^\circ C$] and downward shortwave radiation R_g [Wm^{-2}] for accumulation period t [s] (*Hiemstra and Sluiter*, 2011):

$$PET = 1000 \cdot 0.65 \frac{\Delta}{\Delta + \psi} \cdot \frac{tR_g}{\lambda \rho_w} [mm] \qquad (4.5)$$

with, ψ the psychrometric constant, λ the latent heat of water, Δ the slope of the saturation vapor pressure curve and ρ_w the density of water calculated by:

$$\psi = 0.646 + 0.0006T [hPa^\circ C^{-1}] \qquad (4.6)$$

$$\lambda = 1000(2501 - 2.38T)[Jkg^{-1}] \qquad (4.7)$$

$$\Delta = \frac{6.107 \cdot 7.5 \cdot 273.3}{(273.3 + T)^2} e^{\frac{7.5T}{273.3+T}} [hPa^\circ C^{-1}] \qquad (4.8)$$

$$\rho_w = 1000 [kg\, m^{-3}] \qquad (4.9)$$

The Makkink Potential Evaporation calculated for each time step is called 'near real-time' (PET_{NRT}). The potential evaporation climatology (PET_{Clim}) was calculated by averaging over the full time period (20yr) for each day (Fig. 4.2).
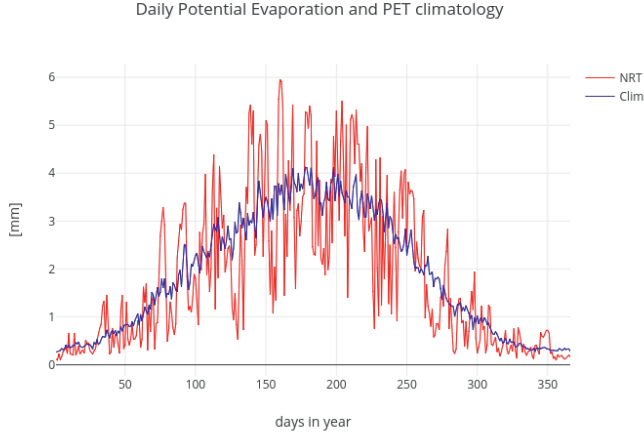
Figure 4.2: *Difference between climatology and near real time potential evaporation. Shown for the year 2004 for grid-cell x:200, y:200.*

ECMWF reforecast

The European Center for Medium-Range Weather Forecasts (ECMWF) issues hindcasts produced with the current model cycle for certain days for the last 20 years. The reforecast obtained for this study was produced with model cycle 43r1 (*Buizza et al.*, 2017). The first forecast is on 1996-03-10 and the last forecast on 2015-12-29 with reforecasts alternating every three or four days.

Forecasted Makkink potential evaporation (PET_{Fcast}) is calculated based on the $t2m$ (T) and $ssrd$ (R_g) variables using equations 4.5-4.9. Temperature was first downscaled to the model resolution using the standard lapse rate as used in the interpolation of the temperature observations as follows:

$$T_x = \overline{T} + \left(\overline{h} - h_x\right)\gamma \tag{4.10}$$

With, $\overline{T}$ the temperature given by the ECMWF forecast on the ECMWF resolution, $\overline{h}$ the average height of the DEM corresponding to the footprint of the ECMWF grid cell, h_x the height of cell x in the model, and γ the lapse rate.

Hydrological model

wflow is a modular hydrological modelling framework that allows for easy implementation and prototyping of regular grid hydrological model concepts in python-pcraster (*Schellekens et al.*, 2017). The hydrological model concept used is the HBV (Hydrologiska Byråns Vattenbalansavdelning) model concept (*Lindström et al.*, 1997a) applied on a grid basis. The generated runoff is routed through the river network with a kinematic wave approach (*Schellekens et al.*, 2017). In the following this model is referred to as wflow_hbv.
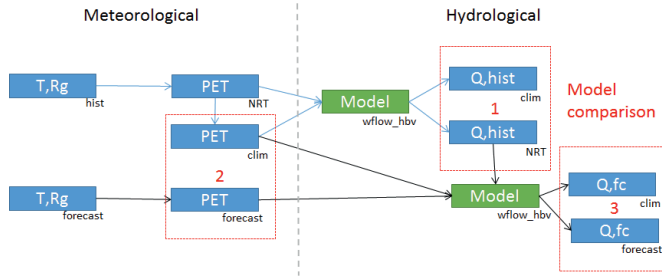
Figure 4.3: Flow chart of the model experiment. Blue boxes represent data products. Green boxes depict modeling activities. Arrows represent the flow of data for historical runs (blue lines) and forecast runs (black). The red boxes indicate the areas for analysis of the results, each box targeting a research subquestion.

The set-up of the hydrological model is the same as used in assessing the validity of the genRE precipitation data set (*van Osnabrugge et al.*, 2017). The model was parameterized through calibration with a Generalised Likelihood Uncertainty Estimation (GLUE) like procedure (*Beven and Binley*, 1992), using HYRAS precipitation as forcing data (*Winsemius et al.*, 2013b,a). The model is taken 'as is' and is not recalibrated for each PET forcing, the effect of which has been studied extensively elsewhere (e.g. *Bowman et al.*, 2017; *Oudin et al.*, 2005a).

4.3 Experimental set-up

The analysis consists of a meteorological part and an hydrological part (Fig. 4.3).

Analysis of meteorological forecast skill

In this analysis we aim to answer the following question: What is the forecast skill of temperature, radiation and potential evaporation compared to precipitation? For this purpose the observations and forecasts are spatially averaged over 148 subbasins Fig. (4.1). Time series of observations and forecasts are then used to calculate the Mean Continuous Ranked Probability Skill Score (CRPSS) for each basin and each season (MAM, JJA, SON, DJF).
The Mean Continuous Ranked Probability Score (CRPS) is an overall measure of forecast quality and is calculated by:

$$\overline{CRPS} = \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \left(F_y(y) - H(y \geq x) \right) \, \mathrm{d}y \tag{4.11}$$

In which $F_y(y)$ is the cumulative distribution function of the forecast variable and $H(y \geq x)$ the Heaviside step function that assumes probability 1 for values greater than or equal to the observation and 0 otherwise (*Brown et al.*, 2010). Interpretation of the mean CRPS is similar to interpretation of a Root Mean Square Error. Both scores have no fixed upper

bound, their magnitude is determined by the variable, and lower scores are better, with zero the perfect score.

The limits of the mean CRPS vary depending on the basin and season and it is therefore difficult to compare between basins and season. For this reason the CRPS is translated into the Continuous Ranked Probability Skill Score, which measures the performance of a forecasting system relative to a reference forecast. The reference forecast here is seasonal climatology. As such the CRPSS equals 1 for a perfect forecast and 0 when the forecast ensemble does not score a better CRPS than the CRPS calculated for the climatological distribution.

$$\overline{CRPSS} = \frac{\overline{CRPS_{REF}} - \overline{CRPS}}{\overline{CRPS_{REF}}} \tag{4.12}$$

Additionally the Relative Mean Error (RME) is calculated for the mean of the forecasts $\overline{Y_i}$ to detect relative biases in the mean:

$$RME = \frac{\sum_{i=1}^{n}(\overline{Y_i} - x_i)}{\sum_{i=1}^{n} x_i} \tag{4.13}$$

In which $\overline{Y_i}$ is the mean of the ensemble for forecast i and x_i the corresponding observation.

The above scores are calculated with the Ensemble Verification System (EVS), a software package to calculate ensemble verification metrics (*Brown et al.*, 2010).

Analysis of the effect of PET forecasts on streamflow predictions

In this second part of the analysis we aim to answer the following questions:

1. To what extent are initial states affected by the use of climatological versus near real time potential evaporation?

2. To what extent can potential evaporation forecasts contribute to streamflow forecast skill?

To answer the first question, the wflow_hbv model is consecutively forced with PET_{Clim} and PET_{NRT}. Four states and two fluxes are exported for analysis: 1) upper soil reservoir, 2) lower soil reservoir, 3) interception storage, 4) soil moisture store; and fluxes 5) discharge and 6) actual evaporation. For the different states and fluxes the Mean Difference (MD) is calculated for each grid cell. This is done for each season to investigate seasonality of differences. The MD is calculated as:

$$MD = \frac{\sum_{i=1}^{n}(STATE_{NRT,i} - STATE_{CLIM,i})}{n} \tag{4.14}$$

To answer the second question two hindcast runs are performed with PET_{Fcast} and PET_{Clim} as PET forcing, respectively. To avoid effects caused by the initial state all forecasts start from the initial states derived from the PET_{NRT} simulation. Forecast skill scores are calculated as for the meteorological variables for 20 discharge gauges and for

each season. Different from the meteorological verification exercise, the metrics are calculated for the forecasts with reference to the model output and not compared with observations. The reason for this was that differences between observation and forecast stem from many different sources, including errors in the initial state. Subsequently, a forecast that is 'too wet' might compensate in the 10 day forecast for initial states that were 'too dry'. For this reason the effect of the meteorological forecast was isolated by calculating the verification metrics against modeled streamflow. This also avoids issues of perceptive bias due to the model being calibrated on another PET forcing; One of the PET types might simply perform better because it is more like the original PET used in calibration.

Streamflow gauges for analysis were selected such that:

1. Only gauges were chosen for which the model was deemed behavioral as expressed by a KGE score threshold of 0.5.

2. Only one gauge was selected for each stream in the basin, except for the Rhine river itself, for which 2 additional gauges were chosen. If multiple gauges in the same stream were present the gauge most downstream was chosen. The gauge 'most downstream' was selected by sorting on mean yearly discharge and picking the highest.

3. From the then remaining list, the largest 20 streams were selected for analysis.

The streamflow locations are shown in figure 4.1 as black squares including the name of the river.

4.4 Results

Analysis of meteorological forecast skill

The forecast skill is assessed for all catchments and for each season. Seasons are northern hemisphere seasons spring (MAM), summer (JJA), autumn (SON), and winter (DJF). Fig. 4.4 shows the Mean Continuous Ranked Probability Skill Score (CRPSS) calculated for subsamples of all forecast-observation pairs for different levels of exceedance, $P(X \geq x)$, for each variable. Simply put, the CRPSS value at $P(X \geq x) = 0.1$ is calculated for the top 10% of observations and the CRPSS value at $P(X \geq x) = 0.7$ is calculated for the highest 70% of the observations. $P(X \geq x)$ is calculated over all observations from all seasons. This means that for some seasons, for example temperature in winter, there is an lower limit in $P(X \geq x)$, because the highest temperatures do not occur during winter. On the other hand, the response of the CRPSS curve is flat for high $P(X \geq x)$ for temperature during summer, as all summer temperatures fall in the highest 60% of temperatures of the whole year. The same is shown for the Continuous Ranked Probability Score (CRPS), Fig. 4.5, and the Relative Mean Error (RME), Fig. 4.6.

There is no skill in the ECMWF forecast beyond 10 days for daily precipitation. This is consistent with the 9-day leadtime in streamflow forecasts found by *Renner et al.* (2009). The skill is best in winter and worst in summer, which is expected based on the dominating meteorological processes (frontal systems in winter and convective events in summer). The total amount of precipitation is underestimated after one-day lead time (Fig. 4.6).
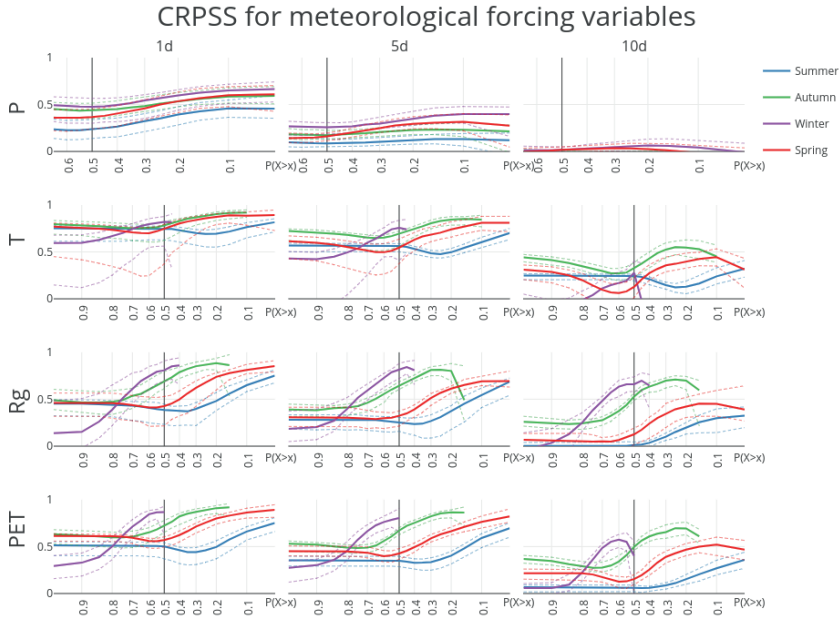
Figure 4.4: *Continuous Ranked Probability Skill Score (CRPSS) for the four forcing variables benchmarked against sample climatology for the 148 HBV subbasins. CRPSS scores are aggregated into mean (solid), 10th and 90th percentile (dashed). Note that the CRPSS score at $P(X \geq x) = 0.1$ resp. 0.7 are calculated over respectively the 10% and 70% highest observation-forecast pairs, conditioned on the observations.*

There is more skill in the forecast for the variables temperature and incoming shortwave radiation. Likewise, there is considerable skill remaining in the potential evaporation forecast. For temperature the one-day forecast is close to perfect for autumn and spring. The skill in temperature forecast is similar for spring, summer and autumn, but worse during winter. The spread, the difference in skill between basins, is also largest during winter and spring. The RME shows that there is a small negative bias in the temperature forecasts. The RME for winter is largests, however it should be noted that the RME is the mean difference weighed by the mean of the observations (Eq. (4.13). As the mean temperature in winter is closer to zero, this results in larger RME. Still, also when expressed in absolute values, the error for temperature during winter is larger than for other seasons (Fig. 4.5).

For radiation there is already quite a considerable loss in skill after one day, but then the CRPSS remains quite stable for longer forecasts, notably during spring and autumn. There is a larger decline in skill for summer and for extreme low radiation values in winter. In absolute terms, the CRPS is related to the magnitude of the average radiation for each season, with the smallest absolute errors for winter and the largest during summer (Fig. 4.5).

CRPS for meteorological forcing variables
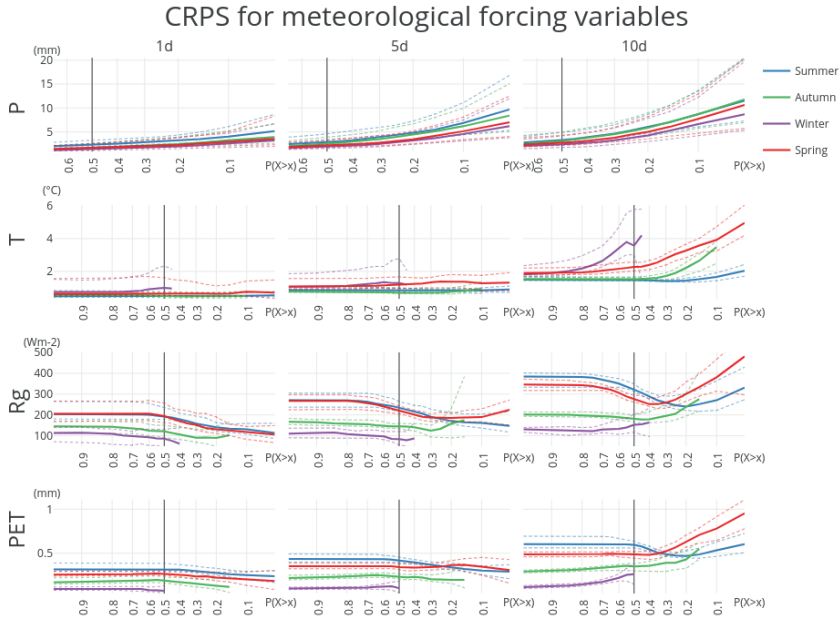
Figure 4.5: Continuous Ranked Probability Score (CRPS) for the four forcing variables for the 148 HBV subbasins for the whole year. CRPS is aggregated into mean (solid), 10th and 90th percentile (dashed). Note that the CRPS score at $P(X \geq x) = 0.1$ resp. 0.7 are calculated over respectively the 10% and 70% highest observation-forecast pairs, conditioned on the observations.

In terms of bias, we see that the relative mean error increases with lower $P(X \geq x)$ (Fig. 4.6, row 3). This indicates that low values are slightly overestimated while high values are slightly underestimated, making the radiation forecasts slightly less extreme than the observations. This is further demonstrated in Fig. B.1 in the appendix, which plots the RME for different levels of non-exceedance ($P(X \leq x)$), as opposed to exceedance in Fig. 4.6.

The skill of the potential evaporation forecast is closely tied to the skill in radiation forecast, both because Makkink potential evaporation is directly proportional to radiation and because the larger uncertainty in the radiation forecast. The forecast skill has the same properties as those found for the radiation forecast. A small difference is that part of the forecast skill in temperature is found back in a slightly improved forecast skill after 10 days for PET compared to radiation in summer.

Overall, there is relatively little spread in skill between basins, with the 10th and 90th percentile close to the mean and following the same trajectory. The difference in skill between the different seasons is larger than the spread between basins, especially for the variables temperature, radiation and potential evaporation. This difference in skill between seasons is partly misleading. For example, the forecast skill for radiation in winter
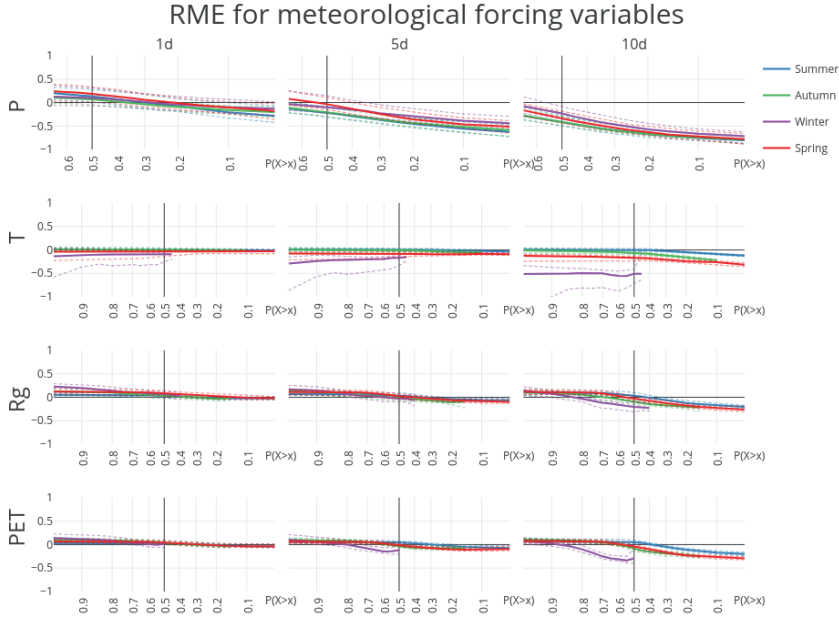
Figure 4.6: Relative Mean Error (RME) for the four forcing variables for the 148 HBV subbasins for the whole year. RME is aggregated into mean (solid), 10th and 90th percentile (dashed). Note that the CRPSS score at $P(X \geq x) = 0.1$ resp. 0.7 are calculated over respectively the 10% and 70% highest observation-forecast pairs, conditioned on the observations.

(Fig. 4.4, purple line) appears to be an outlier. However, the whole range of occurrences of extreme high and low radiative forcing is compressed in a limited part of $P(X \geq x)$. Although the forecast over the whole range of winter radiative forcing is lower than that for the other seasons, the top 10% of winter radiative forcings are actually among the best predicted.

Likewise, high temperatures receive higher skill scores than low season temperatures. This is even more distinct in the radiation forecasts. This does, however, not mean that the forecasts of such rare events are more accurate: both RME (Fig. 4.6) and CRPS (Fig. 4.5) are larger for high extremes, meaning larger errors for those forecasts. Still, taking into account the rarity of the event by calculating the CRPSS, which is the skill of the forecast relative to the skill of a random draw from the climatology, temperature, radiation and potential evaporation forecasts are found to add most information for extreme high values, even though the error of those forecasts is larger than for more 'average' values (values with higher probability of occurrence).
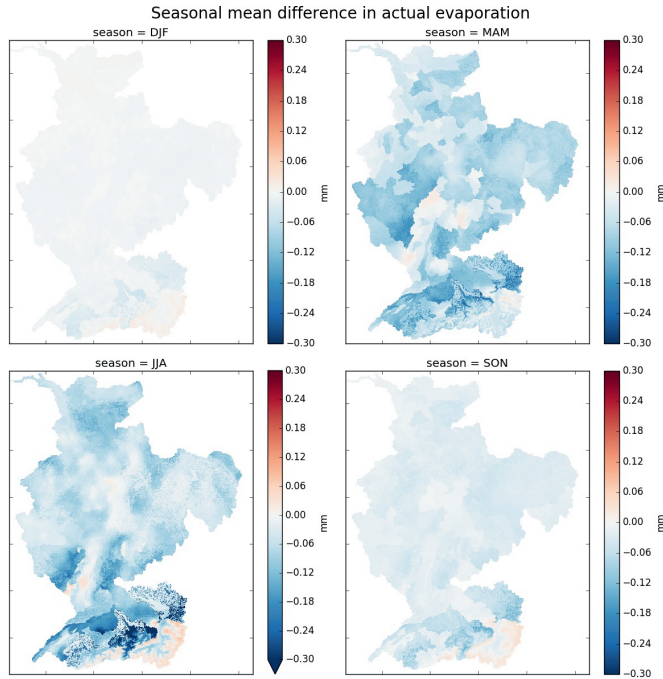
Figure 4.7: *Seasonal mean difference in calculated actual evaporation (AET) for each season. Actual evaporation includes evaporation from interception.*

Influence of dynamic PET on initial states

Dynamic potential evaporation leads to lower actual evaporation (AET). The difference is largest for summer and spring (Fig. 4.7). Part of this lower evaporation is from a reduction in interception as the interception storage is more filled on average under dynamical forcing. This can be explained by the correlation between precipitation events and low potential evaporation. On rainy days the dynamic potential evaporation is generally lower, which decreases the amount of interception evaporation. Under climatological forcing the energy available is not reduced and thus more water evaporates from the interception store. The latter is sometimes taken into account in hydrological models by adding a potential evaporation reduction function dependent on the intensity of precipitation to correct the PET climatology. For example, the HBV model has this option (*Schellekens et al.*, 2017).

The lower evaporation with dynamic PET forcing cascades through the different model storages, accumulating in a mostly wetter lower zone (LZ) storage under dynamic forcing. Finally, the lower evaporation results in higher discharge throughout the Rhine basin (see Fig. B.2-B.7 in the appendix). Exceptions are the high Rhine during spring and to a lesser extent during autumn, and several areas during winter when there is very little effect overall. The wetter conditions also result in higher peak discharges. As these higher

discharges are a result of the temporal dynamics of the potential evaporation input, we expect to find a similar effect on forecasted discharges. As will be shown later (Fig. 4.9), this is indeed the case, albeit very limited.

Influence of PET forecast on streamflow forecast

The CRPSS for streamflow forecast is hardly influenced by potential evaporation forcing type. At first sight, the skill scores obtained with dynamic or climatological PET are identical. Small differences only become visible when taking a close up of the differences by subtracting one from the other (Fig. 4.8). However, the small difference in skill grows with lead-time. The influence of PET forcing type becomes more intuitive when looking at the relative mean error (RME). Visible is an increasing drift with lead-time between PET forcing types (Fig. 4.9). Interestingly, this drift in RME is almost uniform over all subsets of predicted discharge. The drift is positive, which means that forecasted PET leads to slightly higher forecasted discharges, as expected based on the results of the influence of variable PET on the initial states.

Analyzed for each season separately, there is a little more to discover about the role of potential evaporation forecasts and the sensitivity of forecast skill to the meteorological forecast in general. The contribution of the meteorological forecast to streamflow forecast uncertainty is largest for summer, as shown by the largest decrease in CRPSS for the 10-day forecast in summer compared to the other seasons. The CRPSS especially 'dips' for the most extreme discharges, which is not as strong for spring and autumn, and especially compared to the flat response of the CRPSS for the highest 30% of discharges in winter.

In terms of the effect of potential evaporation climatology versus forecasted potential evaporation, the influence is largest (but still quite small) for summer and spring. This is tied to the potential evaporation being of larger magnitude; there is hardly a response for winter where there is lowest potential evaporation.

The influence of PET forecasts on low flow prediction is further examined by calculating the scores for different levels of non-exceedance $P(X \leq x)$, instead of exceedance, so that the score value at $P(X \leq x) = 0.1$ is calculated for the 10% smallest observations and the score value at $P(X \leq x) = 0.7$ is calculated for the lowest 70% of the observations. Not only has the choice of PET forcing for the forecast hardly any effect on the forecasted streamflow (Fig. 4.10, bottom row), but also the forecast skill of low discharge is affected only slightly by the skill of the meteorological forecast in general. The meteorological forecast skill declines with lead time (e.g. Fig. 4.4), but the forecast skill of low percentile discharge remains almost perfect (very close to one) compared to the model under perfect forcing.

4.5 Conclusions

This paper presented a simple and straightforward investigation with an operational forecasting practice perspective. First, observation data was preprocessed for used in the gridded wflow_hbv model. Second, the wflow_hbv model was subjected to dynamical and climatological PET forcing. Three aspects were analyzed: 1) the skill in meteorological

Figure 4.8: *Continuous Ranked Probability Skill Score (CRPSS) for forecast runs (forecasted PET, climatological PET) and their difference benchmarked against model output for the 20 largest sub-catchments in the Rhine basin. CRPSS scores are aggregated into mean (solid), 10th and 90th percentile (dashed). Note that the CRPSS score at $P(X \geq x) = 0.1$ resp. 0.7 are calculated over respectively the 10% and 70% highest observation-forecast pairs, conditioned on the observations.*
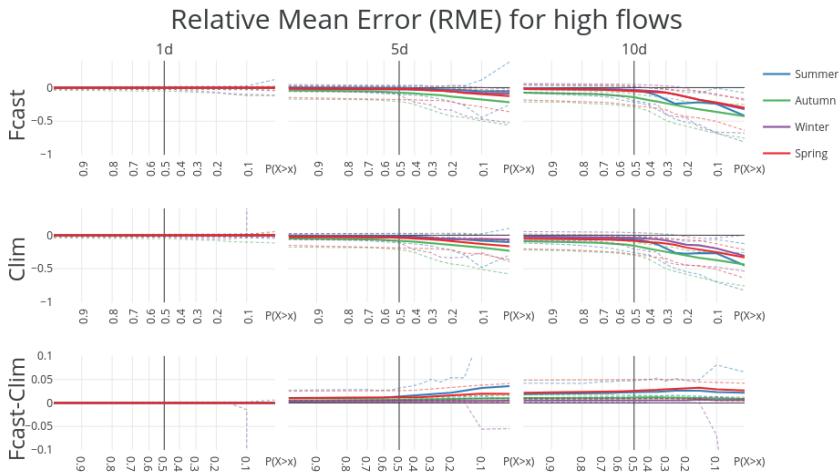
Figure 4.9: *Relative Mean Error (RME) for forecast runs (forecasted PET, climatological PET) and their difference for the 20 largest streams in the Rhine basin. RME scores are aggregated into mean (solid), 10th and 90th percentile (dashed). Note that the CRPSS score at $P(X \geq x) = 0.1$ resp. 0.7 are calculated over respectively the 10% and 70% highest observation-forecast pairs, conditioned on the observations.*
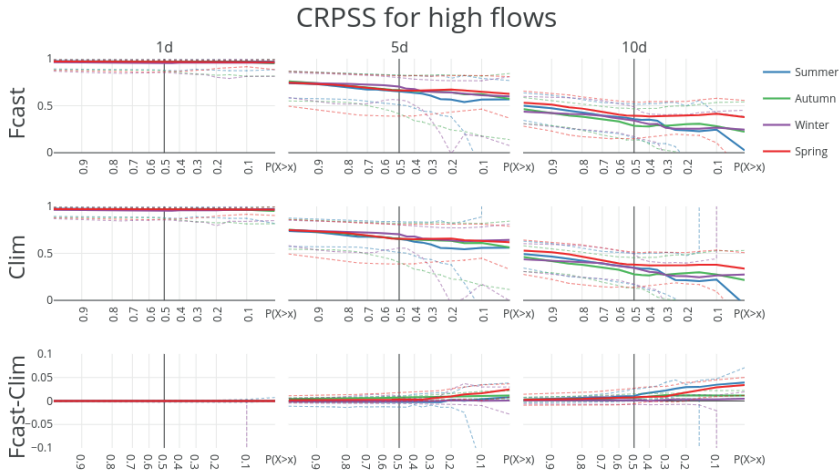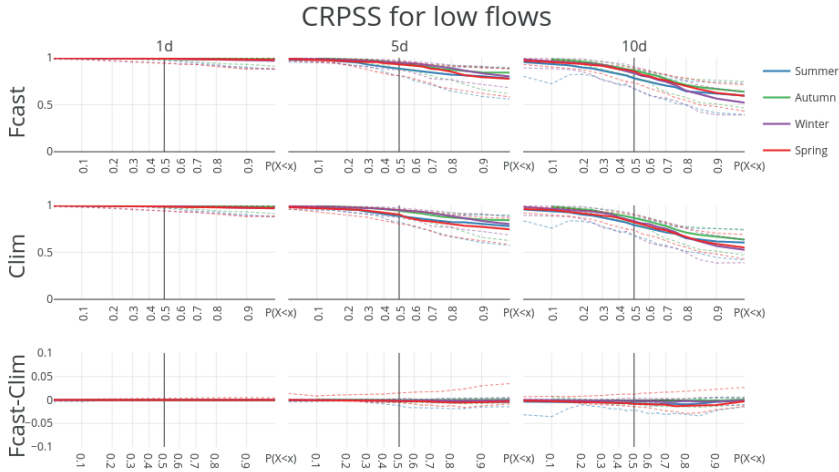
Figure 4.10: *Continuous Ranked Probability Skill Score (CRPSS) for forecast runs (forecasted PET, climatological PET) and their difference benchmarked against model output for the 20 largest streams in the Rhine basin. CRPSS scores are aggregated into mean (solid), 10th and 90th percentile (dashed). Note that the CRPSS score at $P(X \leq x) = 0.1$ resp. 0.7 are calculated over respectively the 10% and 70% lowest observation-forecast pairs, conditioned on the observations.*

forecast, 2) the effect of PET forcing on initial states and 3) the effect of PET forcing on forecast skill.

Nine to ten days is the upper limit on forecast lead time for daily precipitation for the ECMWF forecast in the Rhine basin, with only very little skill remaining compared to climatology. There is considerable skill in daily temperature, radiation and potential evaporation forecasts, also after ten days. Variable PET forcing resulted in lower evaporation and to wetter initial states and higher modeled discharges.

The main result of this study is that potential evaporation forecasts improved streamflow forecasts only slightly. This confirms earlier results that the influence of random errors on estimated streamflow was generally not measurable when comparing model runs directly, needing a 20% systematic bias in PET to influence model outcomes significantly (*Parmele*, 1972). Likewise, *Fowler* (2002) concluded that climatological PET estimates produced a soil water regime very similar to that derived with actual daily PET values, including extreme periods, for a site in Auckland, New Zealand.

There is a wider discussion on evaporation modeling in hydrological models (*Andréassian et al.*, 2004; *Oudin et al.*, 2005a,b) to which the results here might add a new perspective: that of evaporation as a process relevant for medium term forecasts. This is directly also a limitation of this research; Only the influence on forecasts up to 10 days was investigated. The influence on seasonal forecasting might be more substantial, considering that the modeling of evaporation strongly influences the partitioning between runoff and evaporation in the longer term water balance (*Bai et al.*, 2016).

Further limitations are that only one model was tested (wflow_hbv) and for one climate

zone (moderate temperate). The model was calibrated originally on a different PET climatology than studied here and was not recalibrated. The latter is not seen as a limitation. Deliberately not recalibrating the model enabled to focus on the changes in modeled processes instead of comparably vague assessments based on model performance expressed in efficiencies, with the effects brought forward by the PET forcing somewhere hidden in the parameter space.

In the analysis, forecasting metrics were calculated over subsets of observation-forecast pairs conditioned on the observations. Alternatively, the subsets could have been conditioned on the mean of the forecasts. This would present more intuitive information for a forecaster at the time of a forecast when the observation is by definition not yet known (*Lerch et al.*, 2015).

The idea to look at potential evaporation forecast was instigated as part of a program to improve forecasts of low flows. Indeed, it is a recurring hypothesis that potential evaporation forecasts should aid especially in making low flow predictions. The uniform response of several skill scores for different sub-sets of observed discharge does not support this idea; there is no special gain for low flows.

Instead, from our model results it follows that the correct prediction of a drought is firstly dependent on a correct forecast of no-rain. Low flow recession is subsequently determined, in the absence of further feedback mechanisms, solely by the storage-discharge relationship of, in this case, the lower zone representing the saturated zone as well as the routing of surface water.

The follow-up question then is: Is this true in reality, or is this a model deficiency? Should we rethink hydrological modeling to incorporate more feedbacks on evaporation? Certainly there are models with more complex representation of evaporative processes. These are valid and important questions especially in the light of hydrologic response to change of climate drivers. However, from the results presented here, it should not be expected that a better understanding of evaporative processes and feedbacks will result directly in a significant increase in 10-day predictive skill for streamflow.

Chapter 5

Scaling point-scale (pedo)transfer functions to seamless large-domain parameter estimates for high-resolution distributed hydrologic modelling

Moving towards high-resolution gridded hydrologic models asks for novel parametrization approaches. A high-resolution conceptual hydrologic model (wflow_sbm) was parameterised for the Rhine basin in Europe based on point-scale (pedo)transfer-functions, without further calibration of effective model parameters on discharge. Parameters were estimated on the data resolution, followed by upscaling of parameter fields to the model resolution. The method was tested using a 6-hourly time step at four model resolutions (1.2, 2.4, 3.6 and 4.8 km), followed by a validation with discharge observations and a comparison with LSA SAF actual evapotranspiration (ET_{act}) estimates. Additionally, the scalability of parameter fields and simulated fluxes was tested. Validation of simulated discharges yielded KGE-values ranging from 0.6 to 0.9, except for the Alps where a volume bias caused lower performance. Catchment averaged temporal ET_{act} dynamics were comparable with LSA SAF (KGE ≈ 0.7), although wflow_sbm model simulations were on average 100 mm y^{-1} higher. Spatially, the two models were less in agreement (SPAEF = 0.51), especially around the Rhine valley. Consistent parameter fields were obtained and by running the model at the different resolutions, preserved ET_{act} fluxes were found across the scales. For recharge, fluxes were less consistent with relative errors around 30% for regions with high drainage densities. However, catchment averaged fluxes were better preserved. Routed discharge in headwaters was not consistent across scales, although simulations for the main Rhine River were. Better processing (scale independent) of the river and drainage network may overcome this issue.

5.1   Introduction

In the European Rhine river, higher winter discharges are expected as a result of climate change (*Middelkoop et al.*, 2001; *Görgen et al.*, 2010; *Hurkmans et al.*, 2009a; *Krysanova et al.*, 2017). For both policy and climate change impact studies (e.g. *Wit and Buishand*, 2007; *Te Linde et al.*, 2010; *Photiadou et al.*, 2011), and for operational flood forecasting (*Reggiani and Weerts*, 2008; *Renner et al.*, 2009; *Verkade et al.*, 2017) by the Dutch Ministry of Infrastructure and Water Management, the hydrologic behavior of the Rhine is currently modeled with a semi-distributed HBV96 model (*Lindström et al.*, 1997b). Model selection and development originate from the late 90's, following the extreme high water events in 1993 and 1995.

Currently, a transition takes place from lumped to distributed hydrologic models for operational and policy purposes. Within the context of the ongoing H2020 project IMPREX (*van den Hurk et al.*, 2016), we aim to improve the physical realism of the hydrologic model and process descriptions for the Rhine, to obtain more trust in the outcomes of such a model. A transition from lumped to gridded models enables the model developer to take advantage of present day availability of spatial data (either for model building, forcing or constraining). High-resolution spatial data enables the modeler to represent land use and vegetation (e.g. LAI), as they are observed from space, in spatially distributed hydrologic models. For example, estimates of (shortwave downward) radiation fluxes from geostationary satellites make it possible to estimate potential evapotranspiration at high spatial and temporal resolution (*de Bruin et al.*, 2016). Moreover, high-resolution satellite data provide new ways to test and possibly constrain model simulates, particularly the spatial pattern and magnitude of snow, overland flow, soil moisture and actual evapotranspiration.

As such, it is clear that a new generation of dynamic and spatially explicit hydrologic models should take advantage of present day availability of high-resolution spatial data sources. However, the step towards a distributed hydrologic model that takes into account the natural spatial variability within catchments, does not automatically lead to more physical realism. Within the ongoing scientific debate, three main topics can be identified: (1) the importance of and link between temporal and spatial scales (see (*Melsen et al.*, 2016; *Clark et al.*, 2017) and references therein), (2) Model structure and in particular process representation of these spatially distributed models ((*Clark et al.*, 2017) and reference therein), and (3) deriving/estimating model parameters for these models (*Beven*, 2006; *Samaniego et al.*, 2010; *Archfield et al.*, 2015; *Bierkens et al.*, 2015; *Paniconi and Putti*, 2015; *Clark et al.*, 2016; *Mizukami et al.*, 2017). Within the bottom-up modeling approach, it is generally believed that if we solve physical hydrologic processes at the right scale and if linked to correct landscape indicators, less (or no further) calibration will be needed. Moving to a distributed model, which resolves relevant processes on high-resolution, has consequences and comes at a prize regarding parameter estimations (*Beven*, 2006; *Samaniego et al.*, 2010; *Archfield et al.*, 2015; *Bierkens et al.*, 2015; *Paniconi and Putti*, 2015; *Clark et al.*, 2016; *Mizukami et al.*, 2017).

In (*Samaniego et al.*, 2017), an overview is given of the state-of-the-art regarding parametrization and regionalization techniques. They propose the Multiscale Parameter Regionalization (MPR) technique (*Samaniego et al.*, 2010) as a suitable and practical way to ob-

tain seamless parameters across scales. A fundamental step in MPR is the selection of regionalization functions and scaling operators to ensure the transferability of global parameters across spatial scales and to guarantee the seamlessness of parameter fields. A systematic approach of determining these functions is in its infancy and a generalized approach for determining the scaling operators is not yet available. In the MPR methodology, many transfer-function coefficients need to be estimated by calibration (∼53 for the mHM model, see (*Rakovec et al.*, 2016a)). These so-called global parameters include the coefficients of the transfer-functions that link soil and vegetation properties, such as LAI, sand and clay content, to physical or model properties (e.g. residual water content and saturated hydraulic conductivity).

To lower the number of calibrated global parameters in the MPR methodology, or minimize calibration, one could use (pedo)transfer-functions (PTFs) from literature instead of calibrated relations (for an overview of PTFs developed by the pedometric community, see *van Looy et al.*, 2017). A large advantage of such an approach is that the transfer-functions are not constrained to the model, but to actual field measurements. Note, that those PTFs are point-scale relationships and that the scale at which these functions can be used, remains an issue (*van Looy et al.*, 2017; *Samaniego et al.*, 2017). However, the availability of high-resolution soil data (e.g. 250m by *Hengl et al.*, 2017) in combination with the right selection of regionalization functions and scaling operators, ensuring flux preservation on different scales, make it worthwhile to investigate if point-scale PTFs can be used for hydrologic modeling. This, by using the philosophy of MPR: first derive model parameters on the highest available resolution of the data, followed by scaling to the desired model resolution.

Therefore, the objective of this study is to investigate the applicability of point-scale (pedo) transfer-functions in combination with suitable upscaling operators for deriving seamless hydrologic model parameters and the preservation of fluxes across scales for a multi-scale hydrologic model for the Rhine River.

To test the applicability of point-scale PTFs we use the spatially distributed hydrologic model wflow_sbm. In this model, the soil part is largely based on topog_sbm (*Vertessy and Elsenbeer*, 1999), and it uses a kinematic wave approach for channel, overland and lateral subsurface flows similar to TOPKAPI (*Benning*, 1994; *Todini and Ciarapica*, 2002) and G2G (*Bell et al.*, 2007). The evaluation is based on the following sub-questions: (1) Which (pedo)transfer-functions used in wflow_sbm cover the sensitive parameters of this model? (2) What is the quality of the resulting wflow_sbm model on the highest resolution of interest (1200m)? (3) How scale (in)dependent are the resulting wflow_sbm modeled fluxes? Note that to the authors' knowledge, the upscaling methodology advocated in MPR was so far only tested using models without subsurface lateral exchange.

This paper is set up as follows: in section 5.2, the data and study area of this study are presented. Section 5.3 describes the used methodology, with in 5.3, a description of the wflow_sbm model and parametrization, and in 5.3, a description of the experimental setup of this study. The results are presented in section 5.4, followed by the discussion and conclusions in respectively section 5.5 and 5.6.
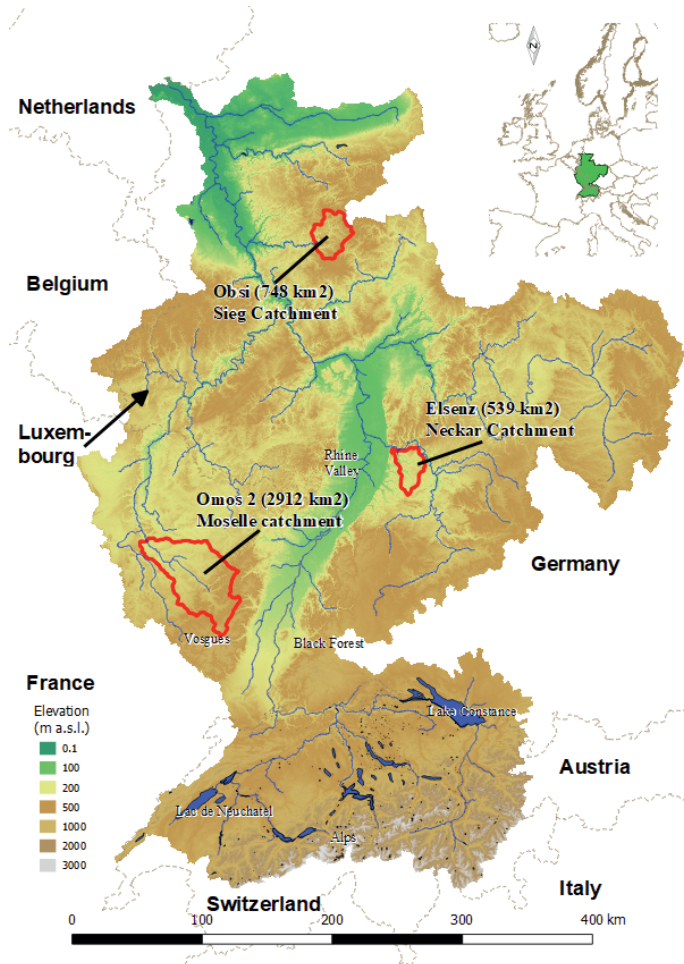
Figure 5.1: Elevation map, based on the STRM digital elevation model (Farr et al., 2007), and overview of the Rhine basin with three chosen sub-basins as study area (in red).

5.2 Study area and data

Study area

The studied basin is the Rhine basin upstream of the Dutch border, which has a surface area of approximately 160 000 km^2 (Figure 5.1, Chapter 5.2). In this figure, we have indicated three sub-basins of the Rhine, which are used for the sensitivity analysis and later on to highlight some results. The sub-basins are: Obsi (748 km^2), part of the Sieg catchment, Elsenz (539 km^2), part of the Neckar catchment and Omos 2 (2912 km^2), which is

part of the Moselle catchment. These three are all headwaters and located in different regions of the Rhine basin. As such, they are a sample out of the variety of e.g. discharge regimes, soil types, local climates, vegetation properties and slopes in the Rhine basin.

Available data

In this section, we introduce the most important datasets used in the study. A complete list of all datasets can be found in Table C.1 in the appendix.

Forcing

A meteorological forcing dataset is available for the Rhine basin, consisting of gauged precipitation interpolated with the genRE method (*van Osnabrugge et al.*, 2017), potential evapotranspiration (PET) and temperature data (*van Osnabrugge et al.*, 2019), on a 1200 m grid. Since not all meteorological stations were continuously delivering high quality data, we refer to (*van Osnabrugge et al.*, 2017; *van Osnabrugge et al.*, 2019) for an indication of the yearly availability and quality of the forcing data. This forcing data is available for the period 1996 through 2015.

DMET LSA SAF evapotranspiration

For the comparison of simulated actual evapotranspiration (hereinafter referred to as ET_{act}) with model simulations from a land-surface model, the DMET product is used from the Land Surface Analysis Satellite Application Facility (LSA SAF) (*Trigo et al.*, 2011). This data is available from 2011 and onward.

LSA SAF DMET (from here-on referred to as DMET) is a product with daily ET_{act} estimates for terrestrial surfaces and inland water systems, following from physically-based Soil Vegetation Atmosphere Transfer (SVAT) models, notably a modified version of the Tiled ECMWF Surface Scheme for Exchange Processes over Land (TESSEL and H-TESSEL, *Viterbo and Beljaars*, 1995; *van den Hurk et al.*, 2000; *Balsamo et al.*, 2009; *Albergel et al.*, 2012). The daily values are an integration over instantaneous 30 min ET_{act} estimates. In their approach, (*Trigo et al.*, 2011) combine radiative data from the Meteosat Second Generation geostationary satellites with ECMWF meteorological forecasts and land cover information as input for these physically-based land surface models (step 1). DMET then uses tiles within grid cells with a mixture of plant functional types. For each tile, the SVAT-model solves the surface energy balance and the resulting ET_{act} value per grid cell is then a weighted contribution of all tiles in that cell (step 2). Finally, a post-processing module is used to create the daily product (step 3).

Grid cell resolution of the original data

The grid cell sizes of the used datasets differ. Whereas it is 1200 m for model and forcing, it is 3000 m for DMET, 250 m for soil data of ISRIC SoilGrids (*Hengl et al.*, 2017), 500 m for monthly averages of Leaf Area Index (LAI) (*Myneni et al.*, 2015) and 100 m for the CORINE land cover (*European Environment Agency*, 2018). Topographic characteristics are based on

a 30 m rescaling of the SRTM digital elevation model (*Farr et al.*, 2007). The highest spatial resolution determines the base resolution before upscaling, also referred to as 'level-0' in the MPR framework (*Samaniego et al.*, 2010). The handling of the resolution differences between the model resolution, and the soil and vegetation properties (the upscaling step), as used for the parameter transfer-functions, is described in section 5.3. For the comparison between DMET product and the wflow_sbm results, model simulations are upscaled to the DMET resolution following an arithmetic mean.

5.3   Methodologies

In this section, we briefly introduce the wflow_sbm model (5.3), its model setup (5.3) and parameterization (5.3), followed by a description of the experimental setup (5.3). The experimental setup is subdivided in three sections describing in order: the parameter sensitivity analysis (5.3), the validation of the model setup on the highest model resolution (5.3), and the scalability of the model (5.3).

The wflow_sbm concept and modeling setup

The conceptual bucket model wflow_sbm is based on topog_sbm (*Vertessy and Elsenbeer*, 1999) with a kinematic wave approach for lateral subsurface, overland and river flow processes that are similar to TOPKAPI (*Benning*, 1994; *Todini and Ciarapica*, 2002) and G2G (*Bell et al.*, 2007). It is available in the wflow open source modeling framework (*Schellekens et al.*, 2019b), which is based on PCRaster (*Karssenberg et al.*, 2010) and python. In contrast to many conceptual models, wflow_sbm models lateral subsurface flows explicitly and it has a simplified physical basis with parameters that represent physical characteristics, leading to (theoretically) an easy linkage of the parameters to actual physical properties *Vertessy and Elsenbeer* (1999). While topog_sbm performs best when simulating fast runoff processes in small catchments (*Vertessy and Elsenbeer*, 1999), wflow_sbm can be applied on a wider variety of catchments (*Schellekens et al.*, 2019b). Notable properties of the wflow_sbm model are:

1. A spatially distributed gridded cell network for both surface and subsurface flow routing, with the presence of lateral subsurface flow based on a D8-network flow routing network (*Karssenberg et al.*, 2010).

2. Parameters that represent environmental physical characteristics such as vegetation and soil properties.

3. A soil divided in saturated and unsaturated store(s). the transfer of water from unsaturated to (un)saturated zones is simulated with the Brook-Corey equation for the hydraulic conductivity as a function of normalized volumetric water content (*Brooks and Corey*, 1964).

4. A kinematic wave module for river, surface and subsurface lateral flow (comparable to *Todini and Ciarapica*, 2002; *Bell et al.*, 2007).

5. Evapotranspiration and interception losses. The latter via an analytical Gash model on daily time steps or a modified Rutter model on sub-daily time steps (*Rutter et al.*, 1971, 1975; *Gash*, 1979).
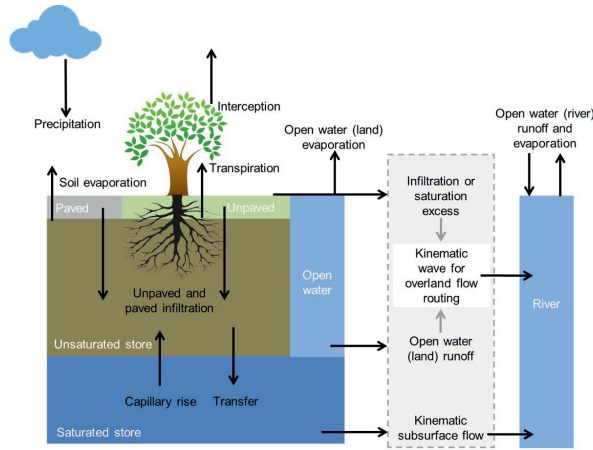
Figure 5.2: An overview of the processes and fluxes represented by the spatially distributed wflow_sbm model (Schellekens et al., 2019b).

6. Capillary rise to the unsaturated store.

7. The option to divide the soil depth in any number of layers instead of one soil column.

8. Reduction coefficients for ET_{act} from the unsaturated zone based on the Feddes curve (*Feddes et al.*, 1978).

9. Glacier build-up and melting processes.

Model setup

Since a meteorological forcing on a cell size of 1.44 km^2 (1200$\times$1200 m) is available for this study (see section 5.2), the same spatial resolution is used as the highest resolution for model runs (level-1). The model setup in this study has four soil layers: 0–10 cm, 10–40 cm, 40–120 cm and 120 cm up to the depth of an impermeable layer or bedrock. The first layer, 0–10 cm, is comparable to the depth remotely sensed soil moisture data is thought to represent (e.g *Dorigo et al.*, 2017), which should enable spatial comparisons with soil moisture products. The next layer (10–40 cm) represents a layer, dense in plant roots, which reacts rapidly to changes in the meteorological forcing. The layer 40–120 cm represents the depth up to the average end of the soil profile for shallow soils. Below this last layer, wflow_sbm computes a variable depth per grid cell, i.e. the layer depth is limited to the maximum soil depth that is set for that grid cell. In this study, we base this depth on the occurrence of an impermeable layer or the presence of bedrock. This is in the current setup never deeper than 2 meters, because of dataset limitations (*ESDAC*, 2004; *Hengl et al.*, 2017).

The model is run on a six hourly time step. A brief sensitivity analysis (not shown) revealed that at time-steps of more than six hours, a deterioration of the model results will

start, while at a six hourly time step and smaller, model results were more or less comparable. An hourly time step is possible with the available meteorological forcing, but coarser temporal resolutions are favorable to decrease run times, especially on the scale of the Rhine basin.

The Rhine basin consists of several large lakes and reservoirs in the Alpine regions. The largest lakes in Switzerland are represented in the model by a lake module, which is also used in other hydrologic models within the wflow framework (*Lindström et al.*, 1997b; *van Osnabrugge et al.*, 2017; *Schellekens et al.*, 2019b). This module requires the reservoir locations, their surface areas, initial water levels and lake operation schemes, which are available from the operational model used for hydrologic forecasting at the Dutch water authority Rijkswaterstaat.

In addition, Alpine glaciers are modeled on all four resolutions. To do this in wflow_sbm, the initial fraction per grid cell covered by a glacier and the initial storage in these glaciers (in mm) needs to be estimated. The global RGI (*RGI Consortium*, 2017), GLIMS (*Raup et al.*, 2007) and Swiss GLAMOS (*Fischer et al.*, 2014) databases are used for this, together with glacier volume estimations from (*Grinsted*, 2013). The modeling of glaciers introduces three extra parameters in the model, notably:

1. G_SIfrac, the fraction of the snow pack on top of the glacier which is converted into ice per time step. This parameter is set to 0.002 (typically ranges from 0.001 to 0.006 *Luo et al.*, 2013).

2. G_TT, threshold temperature at which precipitation falls as snow on the glacier. Set to 1.3 °C (*Tobin et al.*, 2013).

3. G_fmax, the melting factor (mm °C^{-1} day^{-1}). Variable from glacier to glacier, but set to 5.3 mm °C^{-1} day^{-1} here based on reported values by (*Hock*, 2003).

Parametrization and upscaling rules

Model parameters are estimated with available point-scale PTFs from literature (Table 5.1), leading to seamless parameter maps for these parameters over the entire Rhine basin. Following the MPR methodology, we have estimated the parameters on the original data resolution (often referred to as 'level-0'), followed by upscaling procedures to upscale the parameters to the model resolutions ('level-1'). These procedures should enable us to use the derived parameter sets on every desired model resolution (*Samaniego et al.*, 2010; *Kumar et al.*, 2013a; *Samaniego et al.*, 2017).

Universal scaling rules for hydrologic model parameters are not available yet, as they also depend on model structure (*Nijzink et al.*, 2016). However, the right procedure is found when parameter fields remain seamless with a constant mean and standard deviation on various spatial resolutions. The same holds for resulting fluxes and states on various spatial resolutions. That this is possible with the MPR approach, was already illustrated for the mHM-model (*Samaniego et al.*, 2010; *Kumar et al.*, 2013a; *Samaniego et al.*, 2017).

Here, we follow some of the upscaling procedures as used by *Zhu and Mohanty* (2002), *Samaniego et al.* (2010), *Kumar et al.* (2013a), *Mizukami et al.* (2017) and *Samaniego et al.* (2017) to estimate parameter values on four model resolutions: 1200, 2400, 3600 and 4800 m. Generally, these studies slightly differ in the applied upscaling operators, resulting from differences in the model structures and the types of parameters. The used

Table 5.1: List of parameters estimated with a PTF and the used upscaling operators. Only the abbreviated parameter names are illustrated, see Table 5.2 for the interpretation of the parameters. Upscaling operators are abbreviated as follows: A - arithmetic mean and log A - arithmetic mean of the natural logarithm.

Parameter	Pedo-transfer function by	Upscaling operator	Additional notes
c	(*Rawls and Brakensiek*, 1989)	log A	λ upscaled with log A, c determined from λ at model resolution
Kext	(*van Dijk and Bruijnzeel*, 2001)	A	Look-up table from land cover
KsatVer	(*Brakensiek et al.*, 1984)	log A	
LAI	(*Myneni et al.*, 2015)	A	
M			Derived as exponential decay-function from KsatVer at 7 depths
N	(*Engman*, 1986) and (*Kilgore*, 1997)	A	Look-up table from land cover
N_River	(*Liu et al.*, 2005)		Derived at model resolution, depends on Strahler order
RootingDepth	(*Schenk and Jackson*, 2002; *Fan et al.*, 2016b)	A	d_{75} rooting depth
Sl	(*Pitman*, 1989) and (*Liu*, 1998)	A	Look-up table from land cover
Slope	(*Farr et al.*, 2007)	A	PCRaster-functionality (*Karssenberg et al.*, 2010), based on DEM
SoilThickness	(*Hengl et al.*, 2017) and (*ES-DAC*, 2004)	A	
Swood	(*Pitman*, 1989) and (*Liu*, 1998)	A	Look-up table from land cover
thetaR	(*Tóth et al.*, 2015)	A	
thetaS	(*Tóth et al.*, 2015)	A	

PTFs and upscaling operators per parameter are listed in Table 5.1. The equations of the transfer-functions are shown in appendix Table C.2 and a more detailed description of the parameter estimations is present in the supplementary material (Appendix C and Tables C.3–C.5).

In the approach taken here, no further calibration on observations (e.g. discharge) was applied. Note, however that for two model parameters, notably the lateral/horizontal hydraulic conductivity and the day-degree parameter for snow and the glaciers, no suitable PTF is available yet. We applied a constant factor (KsatHorFrac) of 250 on the derived vertical hydraulic conductivity at different scales and 1.3 °C for the day-degree (TT and G_TT) parameters. The sensitivity of the model results on this choice is discussed in the results and discussion.

Experimental setup

Parameter sensitivity analysis

Table 5.2 lists all used parameters of the wflow_sbm model. A sensitivity analysis is performed to assess which of the wflow_sbm parameters have a pronounced influence on both discharge and ET_{act} simulations. With the outcomes, we can determine if all sensitive parameters are estimated with a tranfer-function, while the less sensitive parameters can be set to a default-value in the absence of a PTF. The sensitivity analysis is applied on three sub-basins (Figure 5.1), which are chosen based on their variety in: soil type, geographical location and catchment size to have a representation of the catchment variety within the Rhine basin.

A Latin Hypercube One-factor-At-a-Time (LH-OAT) sensitivity analysis is applied for the period 2014-10-01 until 2016-01-01 with N=100 LH-points. Hence, 100 × 23 (number of used parameters + 1 reference) model runs are necessary for this sensitivity analysis. The LH-OAT analysis uses loops with each loop having one of the LH-points. Each LH-point j then has a partial effect ($S_{i,j}$) for each parameter e_i (*Van Griensven et al.*, 2006):

$$
\begin{aligned}
S_{i,j} &= \frac{100 \cdot \left(\dfrac{g \cdot a - g \cdot b}{[g \cdot a + g \cdot b]/2} \right)}{f_i} \\
a &= (e_1, ..., e_i \cdot (1 + f_i), ..., e_p) \\
b &= (e_1, ..., e_i, ..., e_p)
\end{aligned}
\tag{5.1}
$$

In these equations, g refers to the model functions and f_i is the fraction by which parameter e_i is changed (*Van Griensven et al.*, 2006). This fraction is 0.02 in this study. The partial effect is determined for model simulations of discharge and ET_{act}. Finally, the mean partial effect per parameter is ranked to give an indication of the relative parameter sensitivity.

Validation of the wflow_sbm model for the Rhine

The model is run for the Rhine basin for the period 1 January 1998 until 31 December 2015, with the first year as warm-up period. In the following paragraphs, the model setup validation for this model run is described.

Discharge validation on various resolutions

Validation of simulated discharge on 1200 m takes place by a comparison with discharge observations at 174 gauging stations. The metrics Kling-Gupta Efficiency (KGE) and Nash-Sutcliffe Efficiency (NSE) are used for this (*Nash and Sutcliffe*, 1970; *Gupta et al.*, 2009). Moreover, the lakes in Switzerland and Southern Germany have a pronounced effect on the behavior of the Alpine basins and the main course of the Rhine river. The lake operation procedures of the eight largest lakes are also modeled in this study. Therefore, the simulated lake levels of these eight lakes are validated with observed lake levels.

Validation of evapotranspiration estimates

Current practice to assess model performance is often to solely use discharge observations as a means for the overall performance. Since we are dealing with a gridded hydro-

Table 5.2: *wflow_sbm parameters that are part of the sensitivity analysis. The right column indicates whether these parameters have a PTF or not. Also the parameter value ranges for the sensitivity analysis are displayed.*

Parameter	Parameter interpretation	Lowest value in range	Highest value in range	PTF
Soil parameters				
c	Exponent in the Brooks-Corey equation to calculate the hydraulic conductivity as a function of normalized volumetric water content [-]. This equation is used to simulate the transfer of water from unsaturated to (un)saturated zones.	1.0	20	✔
KsatHorFrac	A multiplication factor applied to the KsatVer parameter that gives the horizontal saturated conductivity of the soil at the surface [-].	0.1	10 000	
KsatVer	Vertical saturated conductivity of the soil at the surface [mm d^{-1}].	1.0	10 000	✔
M	Decay of KsatVer with depth [-].	1.0	3 000	✔
SoilThickness	Depth of the upper aquifer [mm].	100	5 000	✔
thetaR	Residual water content [-].	0.01	0.25	✔
thetaS	Saturated water content [-].	0.25	0.95	✔
Transpiration				
CapScale	Scaling factor in the capillary rise calculations [-].	0.50	1.50	
rootdistpar	Curvature value for root connection with the groundwater table [mm].	-0.01	-10 000	
RootingDepth	Length of the vegetation roots [mm].	100	5 000	✔
Interception				
Kext	Extinction coefficient in the canopy gap fraction equation [-].	0.48	0.96	✔
Sl	Specific leaf storage [mm].	0.02	0.2	✔
Swood	The fraction of wood in the vegetation/plant [-].	0.0	0.5	✔
Flux partitioning				
InfiltCapPath	Infiltration capacity of the compacted soil/paved [mm d^{-1}].	5.0	500	
InfiltCapSoil	Infiltration capacity of the non-compacted soil [mm d^{-1}].	10.0	2 500	
MaxLeakage	Maximum leakage [mm d^{-1}].	0.0	2.0	
PathFrac	Fraction of compacted or urban area per grid cell [-].	0.0	0.2	✔
Routing				
N	Manning's roughness coefficient for overland flow [m$^{-\frac{1}{3}}$ s].	0.008	0.96	✔
N_River	Manning's roughness coefficient for river flow [m$^{-\frac{1}{3}}$ s].	0.003	0.7	✔
Snow				
TT	Threshold temperature at which precipitation falls as snow (HBV parameter) [°C].	-2.0	2.0	
TTI	Critical snowmelt and refreezing temperature (HBV parameter) [°C].	0.0	3.0	
WHC	Fraction of water stored in snow volume (HBV parameter) [-].	0.01	0.3	

logic model, spatial model outputs can also be assessed for model performance. Here, we focus next to discharge on simulated ET_{act} (as is also done in e.g. *Stisen et al.*, 2011; *Wanders et al.*, 2014; *Kunnath-Poovakka et al.*, 2016). For the assessment of ET_{act}, *Demirel et al.* (2018) and *Koch et al.* (2018) have introduced a new efficiency metric, the Spatial Efficiency Metric (SPAEF), which resembles the KGE in that it equally weighs multiple components. It is formulated as:

$$SPAEF = 1 - \sqrt{(\phi - 1)^2 + (\chi - 1)^2 + (\psi - 1)^2} \tag{5.2}$$

with:

$$\phi = \rho(A, B) \tag{5.3}$$

$$\chi = \frac{\left(\frac{\sigma_A}{\mu_A}\right)}{\left(\frac{\sigma_B}{\mu_B}\right)} \tag{5.4}$$

$$\psi = \frac{\sum_{j=1}^{n} min(K_j, L_j)}{\sum_{j=1}^{n} K_j} \tag{5.5}$$

In these equations, ϕ is the Pearson correlation coefficient between 'observed' and simulated ET_{act}, χ is the fraction of coefficient of variations, which represents the spatial variability, and ψ is the percentage of histogram intersection between observations and simulations, with K the histogram of the DMET simulations and L the histogram of the model simulations with n bins. The difference from other spatial metrics is especially pronounced in ψ, because this term is sensitive to clusters in spatial patterns (*Demirel et al.*, 2018; *Koch et al.*, 2018). Hence, this assures that the efficiency metric not only performs on the pixel scale, but also on a larger scale. This is a useful addition, as the model is compared with another model and 'observed' pixel values are not per definition the true values.

We asess modeled ET_{act} with the estimations from DMET by means of the attained SPAEF as a measure for the spatial validation of temporally averaged ET_{act} fluxes. The KGE is used for a temporal assessment of sub-basin averaged daily ET_{act} simulations.

Validation of snow water equivalent estimates

Validation of simulated snow water equivalents (SWE) takes place by a comparison with modeled SWE from SLF (*Jörg-Hess et al.*, 2014). SLF simulations are available for the same period as the model runs and for all Swiss basins on a sub-basin level (∼1000 km^2) and a weekly temporal resolution. The results of this validation are not shown here, but are available in Figures C.13–C.15 of the supplement.

Scalability of the model

To evaluate the outcomes of the parameter estimation and upscaling procedures, we have used the obtained parameter maps and the results from the model runs to assess whether parameter maps, states and fluxes remain constant on different resolutions. First, mean values and standard deviations in parameter maps are compared to show how the parameters are influenced by spatial scaling.

Secondly, the relative error in modeled fluxes and states are compared to those on the highest model resolution (1200 m). This cross-spatial comparison of modeled fluxes is

Table 5.3: Results of the parameter sensitivity analysis. Shown are the rankings from highest to lowest sensitivity on both modeled discharge (Q) and evapotranspiration (ET) per sub-basin, following Van Griensven et al. (2006). The top 5 most sensitive parameters per sub-basin and flux type, are highlighted in blue. Parameters with the same sensitivity receive the same ranking.

Parameter name	Elsenz Q	Elsenz ET	Obsi Q	Obsi ET	Omos 2 Q	Omos 2 ET	PTF
Soil parameters							
c	15	5	15	8	15	8	✔
KsatHorFrac	6	11	7	9	5	12	
KsatVer	6	12	9	11	9	14	✔
M	4	1	3	1	6	1	✔
SoilThickness	10	8	12	6	12	9	✔
thetaR	13	10	14	12	13	10	✔
thetaS	8	4	10	5	11	7	✔
Transpiration							
CapScale	17	17	17	17	17	17	
rootdistpar	16	16	16	16	16	16	
RootingDepth	14	2	13	3	14	2	✔
Interception							
Kext	5	9	5	10	4	6	✔
Sl	3	6	4	7	2	5	✔
Swood	1	3	2	2	1	3	✔
Flux partitioning							
InfiltCapPath	17	17	17	17	17	17	
InfiltCapSoil	17	17	17	17	17	17	
MaxLeakage	9	13	11	15	10	15	
PathFrac	17	17	17	17	17	17	✔
Routing							
N	17	17	17	17	17	17	✔
N_River	17	17	17	17	17	17	✔
Snow							
TT	2	7	1	4	3	4	
TTI	11	14	8	13	7	11	
WHC	12	15	6	14	8	13	

also called flux-matching and is used as a measure of the scale-insensitivity of MPR based hydrologic models (e.g. *Kumar et al.*, 2013a; *Samaniego et al.*, 2017).

Thirdly, the performance of the model across the different scales is investigated by comparing the KGE for simulated discharge on the four resolutions with observations. For this, 20 (of the 174) gauging stations across the Rhine basin are used.

5.4 Results

Parameter sensitivity analysis

The top 5 most sensitive parameters per sub-catchment and per efficiency metric are colored blue in Table 5.3. The nine parameters that are once or more often part of this top 5 are: *c, Kext, KsatHorFrac, M, RootingDepth, Sl, Swood, thetaS* (θ_s) and *TT*. These parameters have a shared effect on both discharge and ET_{act} simulations, although *RootingDepth*

mostly affects the ET_{act} simulations (Table 5.3), as it directly influences the maximum depth that can be used for water uptake in the evapotranspiration module.

c, *KsatHorFrac*, M and θ_s are sensitive parameters as they influence either the available water storage or the fluxes between layers and cells. From that perspective, we expected *KsatVer* to be part of these sensitive parameters as well. *Kext*, *Sl* and *Swood* are parameters that are used in the interception module by making use of the LAI maps. This affects ET_{act} and through ET_{act} the runoff ratio is altered affecting discharge. The sensitivity of this parameter is thus also related to the land cover. Since all three sub-basins have a high fraction of forest, interception starts to play a bigger role leading to an expectedly higher sensitivity of these parameters.

Most of the sensitive parameters have an applied PTF to spatially estimate the parameter values. The two remaining sensitive parameters without a PTF are *KsatHorFrac* and *TT*. Both parameters have a higher degree of conceptualization than the more physically based parameters that do have a PTF. Because of this, it is challenging to find a point-scale PTF from literature with which the parameters can be estimated. Ideally, proper transfer-functions are derived for these parameters. As this is outside the scope of this study, default parameter values are used for the model runs: 250.0 for *KSatHorFrac*, a value in between the range of values generally used for TOPKAPI and G2G (*Todini and Ciarapica*, 2002; *Bell et al.*, 2007), and 1.3 °C for TT (*Tobin et al.*, 2013). Hence, with respect to the first sub-question, most sensitive parameters of wflow_sbm are covered by a PTF, but not all. In that case, a default value could be applied based on literature.

Validation of wflow_sbm model for the Rhine

Discharge

For most gauging stations in the central and northern part of the basin, KGE values range from 0.60 to 0.90 (Figure 5.3a). In this figure, an overview is given of the reached KGE (including r, α and β) per river in the Rhine basin. Most of the Rhine tributaries have more than one discharge station. For each tributary, only the station with the largest standard deviation in discharge is displayed in order to have only one station per tributary in the figure, as these locations generally correspond to the river outlets.

Over the main course of the Rhine, KGE values fluctuate between 0.57 in Switzerland to 0.82 close to the Dutch-German border. The lower KGE values upstream in the main course of the Rhine are caused by a bias towards too low simulated discharge volumes in this region (Figure 5.3a), something we also see in the hydrograph for 2011–2015 of the Upper Rhine (KGE = 0.57 at Basel; Figure 5.4), a location which transports all water of the Alpine basins. Especially during spring and summer, the discharge volume is highly underestimated with approximately a difference of 500–1,000 m^3-1. Going downstream, KGE values increase to 0.78 for the Lower Rhine at Emmerich (0.82 for the full period, see Figure 5.3). The volume difference of the Upper Rhine remains noticeable with slightly underestimated spring and summer discharges. However, due to the addition of water from non-Alpine headwaters, which do not have the aforementioned discharge volume bias, this effect is less pronounced in the results downstream.

This effect in the Alps already starts in the simulations for the headwaters in this area, with KGE values < 0.4 in the southwest of the area. For these headwaters, simulated dis-
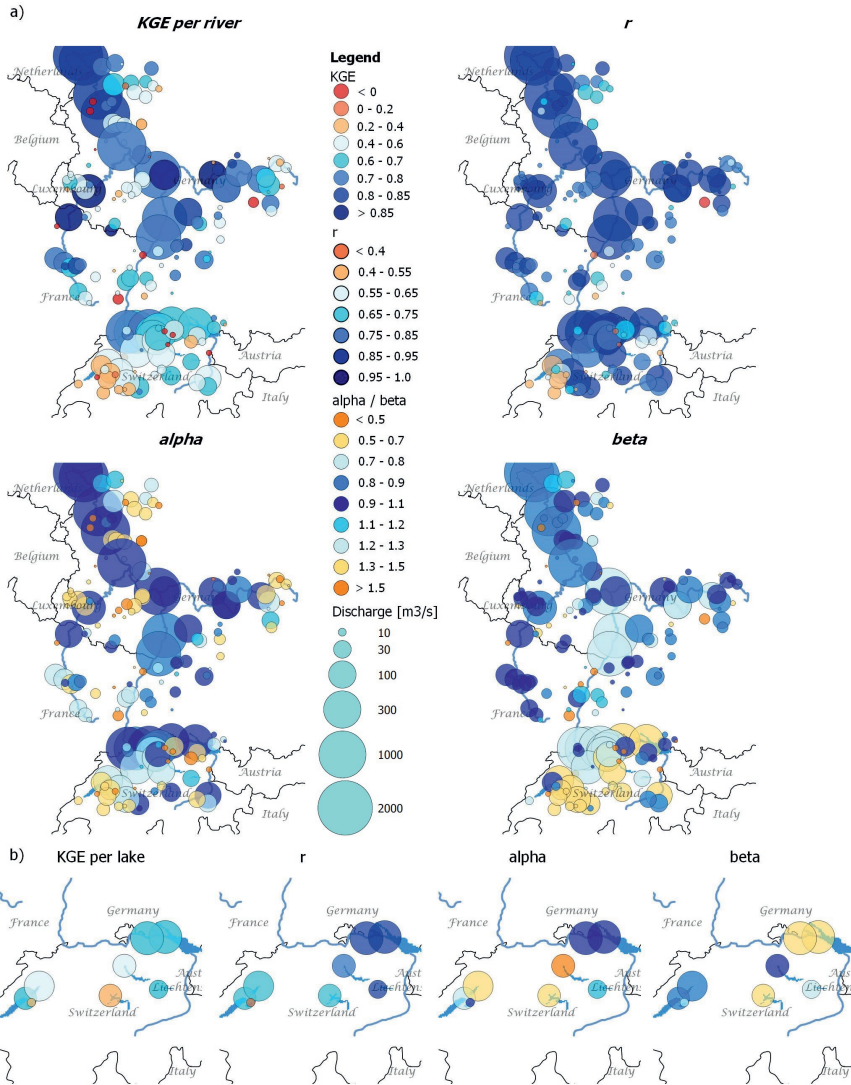
Figure 5.3: (a) Attained KGE per discharge station, split in the r, α and β terms as described by Gupta et al. (2009). The size of the circles gives an indication of the mean discharge flux measured at that discharge station, while the colors display the reached values. Per tributary and if available, one location is displayed; this is the station with the highest standard deviation in observed discharge. KGE values and their decomposition are based on a comparison of the modeled discharge with discharge observations for the period 1999–2015. (b) Similar to a, but for the simulated lake levels (bias-term is based on subtracting the minimum lake level of the dataset).

Figure 5.4: *Hydrographs of discharge simulations with wflow_sbm for four sub-basins of the Rhine (Elsenz, Agger, Omos 2 and the Alpine Thur) and two locations in the main course of the upper and lower Rhine (Basel and Emmerich). Hydrographs and metrics are shown for the period 2011–2016. Model simulations are indicated by blue lines and observations in black. Attained KGE and NSE values are displayed in the top left corner.*

charge volumes are generally too low, as pronounced in the bias-term (< 0.7). Also discharge peaks are often underestimated. For instance, 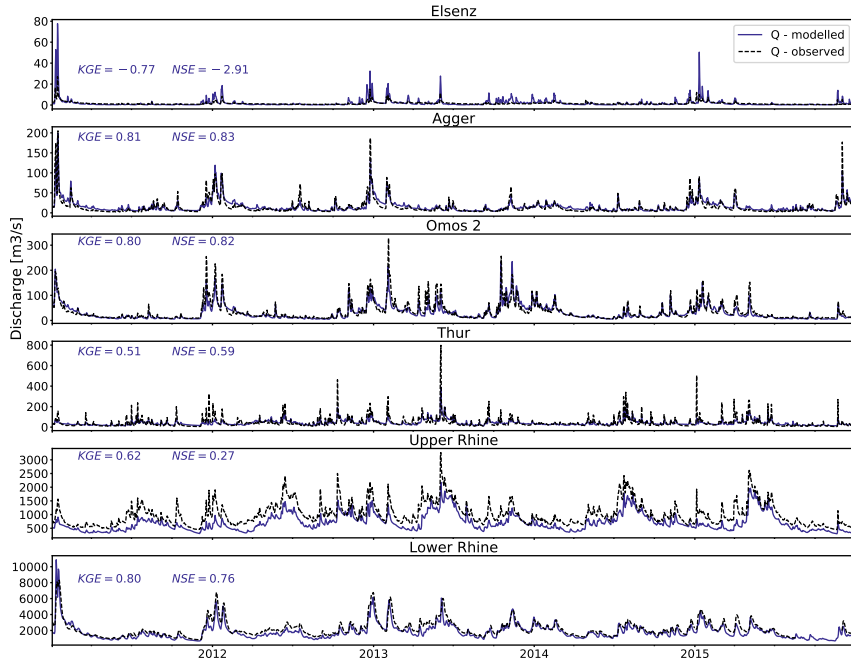the Thur basin in the Alpine region (KGE = 0.52, Figure 5.4), has often highly underestimated peaks, and similar to the results in the Omos 2, the recession after a peak is slower than in the observations.

Most of the Alpine headwaters are supplying water to the reservoirs in the south of the Rhine basin, so the simulations of the lake levels should be influenced by the upstream simulations. This persists into the main course of the Rhine, as the reservoirs are an important source of water to the Rhine in this region. For the lakes in the east, a bias is present indicating on the volume difference we also saw in the discharge simulations (Figure 5.3b). This volume difference is also present in the SWE estimations for snow and glaciers in the Alps (Figures C.13–C.15 in the supplement). In the southwest, however, the bias in the lake level simulations is less. Here, not only a volume difference, but also mismatches in the lake operation schemes (see Figure C.3 in the supplement), and in the timing and magnitude of the river discharges into these lakes (Figures C.10 to C.12) can cause the differences in simulated lake levels.

Model simulations for the river Moselle and its tributaries (west of the figure and mostly
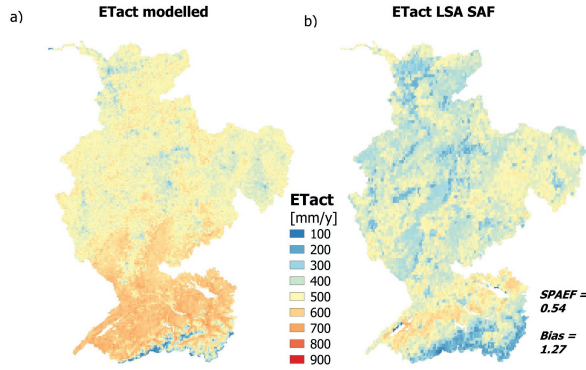
Figure 5.5: *A comparison between modeled yearly averaged ET_{act} simulations (upscaled to a resolution of 3 km) and DMET ET_{act} simulations (Trigo et al., 2011).*

in France), and part of the Main plus its tributaries (east of the basin), give relatively high efficiencies with KGE values often in the range 0.7–0.9. In the aforementioned regions, smaller tributaries sometimes have lower KGE-values (< 0.6). This is mostly pronounced in the α term of the KGE, with values often < 0.7, indicating on a difference in streamflow variability between observations and simulations. At these locations, the discharge simulations generally do not fully capture the quick discharge peaks after a rainfall event, while the baseflow is higher than observed. The values of r and β on the other hand are generally closer to 1.0, indicating on both a high correlation and a low bias between observed and simulated discharge. Especially in the Moselle, most model simulations are unbiased.

Focusing on a couple of the headwaters, discharge simulations for both the Agger (Sieg catchment) and Omos 2 (Moselle catchment) result in KGE values of respectively 0.79 and 0.77 (NSE = 0.83 in both cases). Summer baseflow is often well captured for these basins, but discharge peaks are somewhat underestimated for both catchments, generally followed by a too slow recession. For the Elsenz (Neckar catchment), the performance is less with a KGE of -0.20 (NSE = -1.02). The latter sub-basin, Elsenz, does not have a station at the outlet, but it has a gauging station in the middle of the catchment at a tributary of the main watercourse, which makes the contributing area even smaller, while the catchment is already small with 539 km^2. The discharge simulations for this basin generally contain overestimated discharge volumes and peaks.

Hence, with the current model setup, we were able to reach average to good model performance in the middle and downstream region of the Rhine River basin. The Alpine region, in the south of the basin, is an exception to this due to a clear bias in the estimated discharge amounts and to a lesser extent due to not well captured discharge variability of the peaks and recession rates.

Actual evapotranspiration

For the entire Rhine basin, the spatial patterns of ET_{act} simulations directly reveal that modeled ET_{act} by wflow_sbm is systematically higher ($527\,mm\,y^{-1}$) than the DMET modeled ET_{act} ($427\,mm\,y^{-1}$), with a relative 'bias' of 1.23 (Figure 5.5). Spatial patterns shows a degree of correspondence in the higher Alps in the south and close to the catchment outlet. A band of higher ET_{act} in the flatter valley in northern Switzerland is also present in both simulations, although the simulation of wflow_sbm is higher for a larger region than just a narrow band in the north of Switzerland, as is the case for DMET.

Between other regions, there are more differences, leading to a SPAEF of 0.51. Although the overall correlation and spatial variability are quite similar ($\phi = 0.93$ and $\gamma = 0.95$), the clusters of spatial patterns differs (the histogram intersection, $\psi = 0.52$). Especially the Rhine valley (see also Figure 5.1) shows a clear difference between the two models. While for DMET, yearly ET_{act} simulations are higher in the Rhine valley than on the two forested ridges next to it. For the wflow_sbm simulations, this is exactly the opposite.

In Figure 5.6, a time series of sub-basin spatially averaged daily ET_{act} simulations are compared to DMET. For all basins, winter ET_{act} derived with wflow_sbm is higher than DMET. For the remaining seasons, the differences between the two models are smaller. Omos 2 (KGE = 0.68) is an exception to this, with almost continuously higher ET_{act} by wflow_sbm (bias = 1.33). Other than that, day to day catchment averaged patterns are somewhat more in agreement than it was the case for the spatial comparison in Figure 5.5, with KGE-values around 0.7 ($\rho \approx 0.8$).

A clear discrepancy is present between modeled ET_{act} and DMET simulates for all three basins from July to August 2013. During a drier period in these months, ET_{act} simulated by wflow_sbm shows reduced ET_{act} for most of the period. The simulated ET_{act} from the DMET does not show this behavior. Either the wflow_sbm model overestimates the reduction in evaporation as a result of limited soil moisture, or the DMET model did not apply a suitable reduction. If the difference stems from the former, (in this case) likely either the water storage was too low, e.g. due to the limited soil depth of 2 m, or the rooting depth was too shallow. In section 5.5, we elaborate on the different processes between wflow_sbm and DMET used for estimating the soil moisture content.

Scalability of the model

Parameter estimates on four resolutions

Figure 5.7 illustrates obtained parameter maps for the saturated conductivity (*KsatVer*) and the saturated water content (θ_s), in combination with probability density distributions of the obtained parameter values. Both the resulting parameter maps of the saturated conductivity (a–h) and the saturated water content (i–p) indicate on a preservation of the spatial parameter fields on the four resolutions. This is also illustrated by the mean values and standard deviations of the probability density distributions, which are unchanged when moving to coarser resolutions: $\mu = 6.31$ and $\sigma = 1.01$ ($\log mm\,d^{-1}$) for the saturated conductivity, and $\mu = 0.46$ and $\sigma = 0.03$ for the saturated water content). However, for *KsatVer*, a small decrease of μ (from 6.31 to 6.30 $\log mm\,d^{-1}$) and of σ (from 1.01 to 1.00 $\log mm\,d^{-1}$) is present.
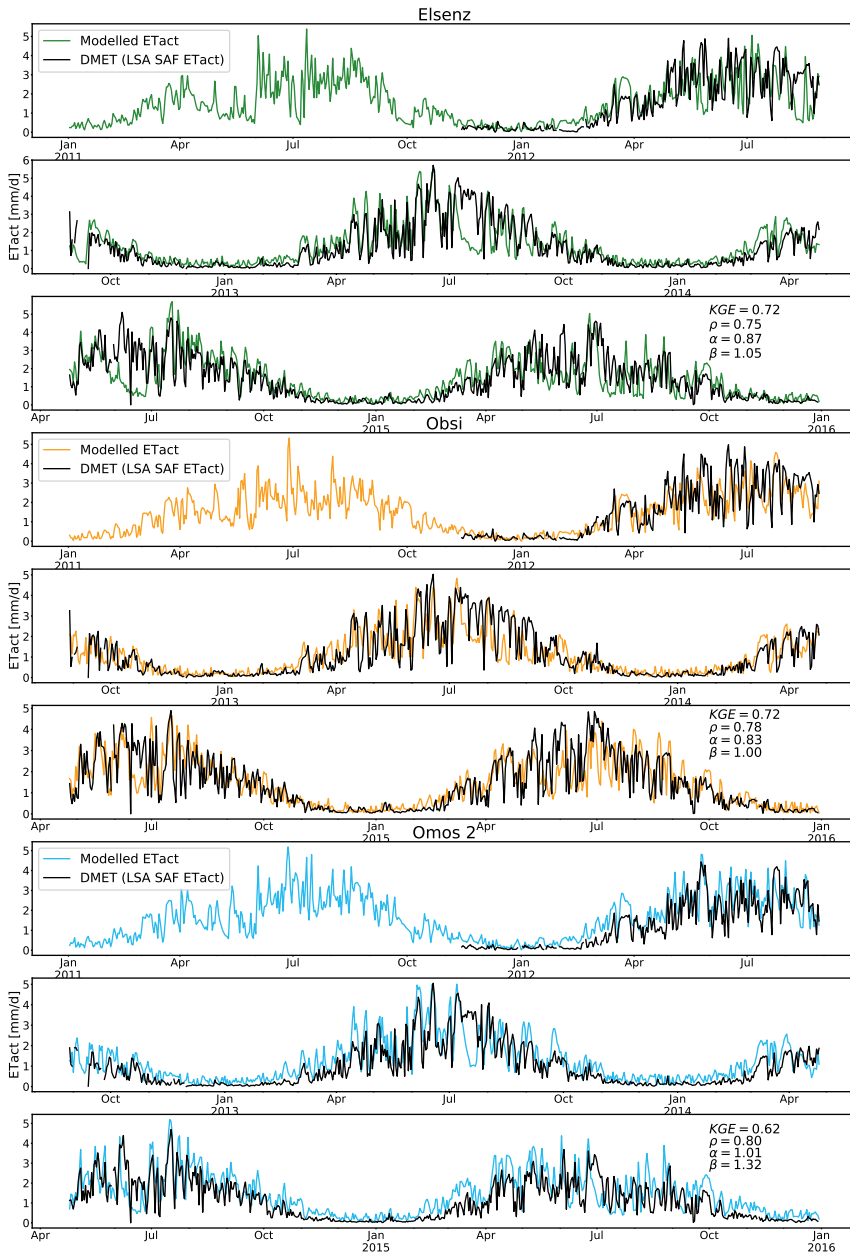
Figure 5.6: *Catchment averaged daily ET*$_\mathrm{act}$ *simulated with wflow_sbm for three sub-basins of the Rhine for the period 2011–2016. These aggregated daily ET*$_\mathrm{act}$ *simulations are compared to DMET ET*$_\mathrm{act}$ *(black line). Green lines indicate the modeled ET*$_\mathrm{act}$ *for the Elsenz, in orange for Obsi and in blue for Omos 2.*

83
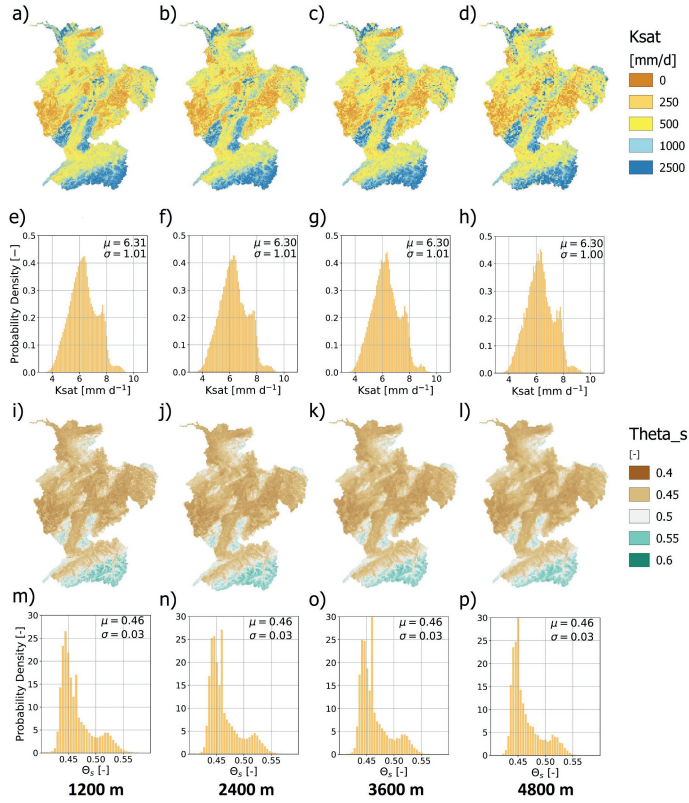
Figure 5.7: *Parameter estimates at four resolutions for KsatVer and θ_s. (a)–(d), parameter esti-
mates of KsatVer with increasing resolution from 1.2 km to 4.8 km. (e)–(h), histograms of the
natural logarithm of the KsatVer parameter values; histograms match with the parameter maps
on top of them. (i)–(l), parameter estimates of θ_s at the same four resolution. (m)–(p), similar to
the previous histograms, but for θ_s.*

Consistency of fluxes and states on four resolutions

ET_{act} fluxes are well preserved on all four resolutions (Figure 5.8, a–j), with relative errors between the highest resolution and the coarser resolution that are generally $<$ 10% (e–g). Occasionally, relative errors of +/– 30% are present. This mostly occurs in regions with an average daily ET_{act} flux around 0.5 mm, e.g. the higher Alps and the northern Rhine valley. For the Alps, these differences can be caused by glaciers and snow processes that are not well scaled, due to the absence of PTFs for the parameters of these modules. As most variability seems to take place locally (from pixel to pixel) and not necessarily per region, figure 5.8 (h–j) illustrates the relative errors of the catchment averaged ET_{act} simulations. The relative errors in these figures is for all sub-basins close to zero, indicating that ET_{act} fluxes remain preserved on all four resolutions.

Flux preservation is certainly not everywhere the case for the recharge fluxes (the net flux of downward transfer and capillary rise, see Figure 5.2) in Figure 5.8 (k–n). Especially in regions with recharge fluxes just above or below 0.0 mm d^{-1}, the relative errors in (o–q) often exceed the +/– 50%. However, absolute errors are generally low (0.0–0.1 mm d^{-1}) in those regions and mostly take place close to river cells. This is likely caused by the scaling of river cells, which is well visible between the maps of Figure 5.8k (1200 m) and n (4800 m). Around these river cells, the relative error is highest (Figure 5.8q), while in other regions the relative error is minimal. The resulting speckled pattern, however, reduces when the catchment averaged fluxes are compared in (r–t). There, relative errors are more often $<$ 10% and around 30% in the sub-basins around the main course of the Rhine, where the recharge fluxes are closer to 0.0 mm d^{-1}. This mainly applies to the central Rhine valley, the river Main and the course of the Rhine after the tributaries of the Main and Moselle have joined it. These are regions where the drainage density is relatively high.

In Figure 5.9, performances of modeled discharge, indicated with KGE, are given at the four model (level-1) resolutions as an average performance for the main course of the Rhine (dark blue), for thirteen headwaters in the middle and downstream part of the basin (blue) and for three Alpine headwaters (light blue). Discharge simulations in the main course of the Rhine remain constant with an average KGE over three gauges of 0.78 on all resolutions. This is, however, not the case for simulations in the (Alpine) headwaters. The averaged attained KGE for the thirteen sub-basin decreases from on average 0.65 to 0.48 between the finest (1.2 km) and the coarsest (4.8 km) resolution. For the three Alpine headwaters, this decrease is almost absent until a resolution of 3.6 km, followed by a quick decrease towards a resolution of 4.8 km. Hence, although simulated discharge scales well for the larger rivers, discharges simulates are not fully consistent for the headwaters with wflow_sbm and the approach in this study.

5.5   Discussion

The transition from lumped to gridded models enables us to take advantage of spatial data and is important for a variety of modeling applications, in particular to study climate and land use changes (*Fatichi et al.*, 2016). As stated in the introduction, moving to distributed modeling concepts has consequences and comes at a price regarding parameter
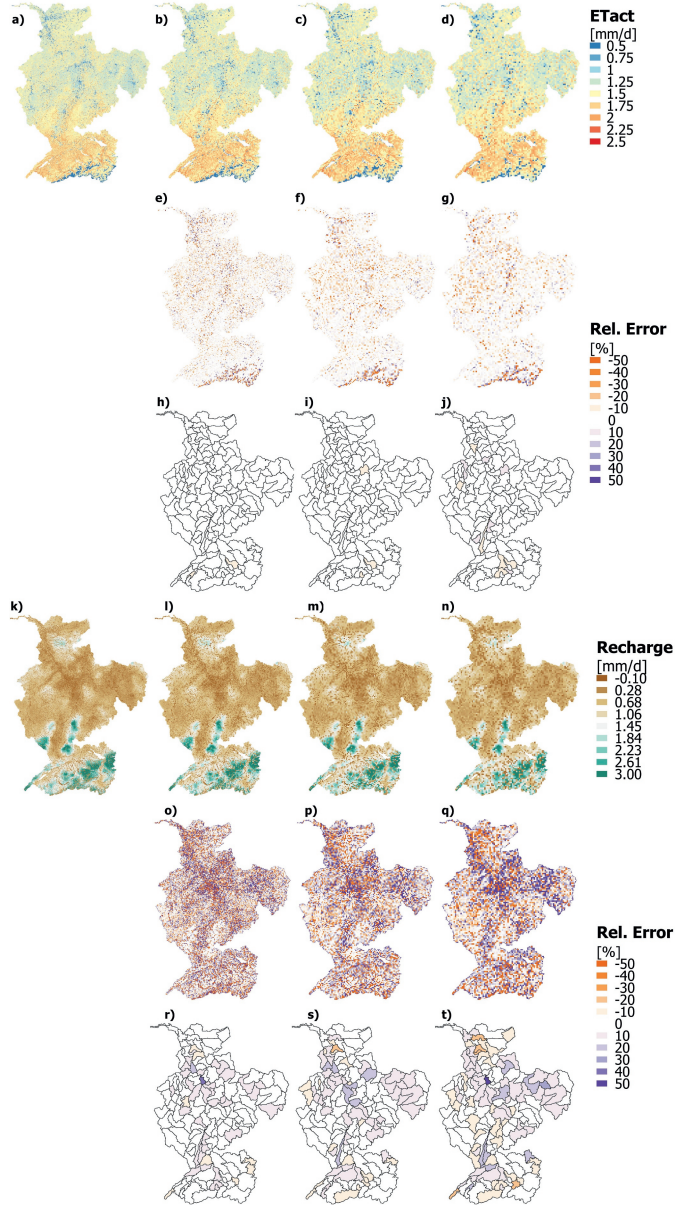
Figure 5.8: Simulations of daily averaged ET$_{act}$ on four resolutions, (a) 1.2 km, (b) 2.4 km, (c) 3.6 km and (d) 4.8 km. (e) - (g) illustrate the relative error per grid cell between one of the coarser resolutions (b–d) and the simulations on 1.2 km (a) after upscaling these simulations to one of the coarser resolutions. (h) - (j), same as (e–g), but catchment averaged fluxes are used instead of fluxes per grid cell. (k) - (n) illustrate the daily averaged recharge fluxes on the four resolutions. (o) - (t), same as (e–j) but then for the recharge fluxes.
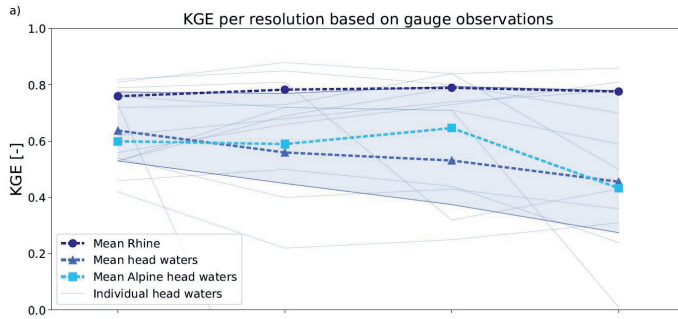
Figure 5.9: KGE per resolution as derived with model simulations and observations from 19 gauges. Shown are the mean KGE for: 3 gauges in the main course of the Rhine (dark blue), 3 gauges in Alpine head waters (light blue) and 13 gauges in headwaters in the middle and downstream area of the Rhine basin (blue). The shaded blue area is the area in between the 25% and 75% percentile (the interquartile range - IQR) for the middle and downstream headwaters.

estimations (*Beven*, 2006; *Samaniego et al.*, 2010; *Archfield et al.*, 2015; *Bierkens et al.*, 2015; *Paniconi and Putti*, 2015; *Clark et al.*, 2016; *Mizukami et al.*, 2017). In this study, we were able to parameterize most sensitive components of the wflow_sbm model using litera-ture PTFs derived from laboratory experiments with point-scale samples in a bottom-up approach. Still, in the current scientific debate, there is no consensus on the matter of using either a bottom-up or top-down approach in setting up hydrologic models. Argu-ments against a bottom-up approach are that this results in overly complex models, and that processes on the data-driven conceptualization scale, e.g. the PTFs derived on the point scale, may not be representative on the actual modeling scale (*Beven*, 1989; *Blöschl and Sivapalan*, 1995; *Sivapalan et al.*, 2003). With the increased availability of (remotely sensed) spatial data, meteorological forcing with both high spatial and temporal resolu-tion, new upscaling procedures (e.g. *Samaniego et al.*, 2010), and high-resolution models, it can be argued that this makes a bottom-up approach more feasible.

However, the use of so-called point-scale PTFs, as used in this study, is still debatable. *Samaniego et al.* (2017) even state that these PTFs should not be used at all beyond their derivation scale (often 100 cm^3). They support this statement based on an analysis with PCR-GLOBWB (*Bierkens and Van Beek*, 2009; *van Beek et al.*, 2011) and Noah-MP (*Niu et al.*, 2011), which are parameterized with ad hoc implementations of PTFs. An important dif-ference with this study, is that the model parameters are not well upscaled (Noah-MP) or even derived at the model resolution (PCR-GLOBWB). In this study, we have focused on deriving the parameters on the original data resolution (level-0) followed by upscal-ing with suitable upscaling operators to the model resolution (level-1). In essence, when this two step approach takes place in the appropriate way, the process of upscaling the parameter estimates should be independent of the chosen parametrization (i.e., calibra-tion, use of PTFs, etc.) on the level-0 scale. That does not mean that calibrated transfer-functions can not lead to better model performance, when focusing on simulated fluxes as compared to observations or other models, but this is something that will follow from

the model validation and not from flux- and state-matching on different resolutions.
The question that then remains, is whether point-scale PTFs can be used to obtain parameters at the current level-0 scale (e.g. 250 m for SoilGrids; *Hengl et al.*, 2017). We expect this to be parameter specific, e.g. the saturated conductivity varies more on small scales than the soil depth. A focus on the parametrization (level-0) scale is therefore recommended in future studies with point-scale PTFs. Point-scale PTFs have the advantage that they require no further model calibration and therewith leaving out any model and forcing related dependency of the results. In fact, these transfer-functions are derived from laboratory experiments with soil samples and are therefore still closely linked to soil properties. A comparison between derived transfer-functions with the MPR approach and point-scale PTFs (for one model) would be of interest in our opinion.

We think that with the increasing spatial resolution of soil and vegetation datasets, in combination with high (spatial and temporal) resolution models that move towards better physical representations of hydrologic processes, the use of these point-scale PTFs becomes more feasible. Certainly, with the results obtained in this study, we see a clear reason to keep exploring the potential of such an approach.

An advantage of this approach is that it shifts the focus towards the model processes and its limitations. This is something that, in our opinion, is not present yet in other papers that included the MPR approach. This focus, makes it possible to specify improvements for model parameters and model structure for both increased model performance and a better scalable model. In the following sections of the discussion, we focus on these improvements by reflecting on the sub-questions that were posed in the introduction.

Parameter sensitivity, estimates and upscaling rules

In this section, we focus on the (pedo)transfer-functions used in wflow_sbm that cover the sensitive parameters of this model and the parameters that are not covered by a transfer-function (5.5). Possible improvements in the parameter estimates are discussed in section 5.5.

Parameter sensitivity analysis

In the sensitivity analysis, we have found nine sensitive parameters that should have a parameter estimation, either via PTFs or via calibrated transfer-functions. Two of them did not have an applied (pedo)tranfer-function. Although this is outside the scope of the current study, a transfer-function for these parameters is needed in the absence of a point-scale PTF. The search for and calibration of transfer-functions for the snow module parameters is then an option. That this is possible for these parameters, has already been shown by *Samaniego et al.* (2010) and *Kumar et al.* (2013a).

The other sensitive parameter without a PTF is *KsatHorFrac*. This multiplication factor applied to the vertical saturated conductivity, which is generally larger than 1, has a pronounced effect on the lateral subsurface flows and via that way on modeled discharges (see also section 5.5). Although multiple PTFs are available for the vertical saturated conductivity in literature (e.g. *van Looy et al.*, 2017), there are no PTFs available in literature for this multiplication factor, that may compensate for anisotropy resulting from the point-scale PTF to estimate *KsatVer*. The estimation of the horizontal saturated conductivity is

therefore likely not representative for larger scale horizontal saturated hydraulic conductivities at the model resolution (with in reality smaller hill slope flow length scales). To allow for no further calibration, we have chosen to fix this parameter to a default value. Nevertheless, for future model improvement, a transfer-function or calibration of this parameter is recommended.

Kext, *Sl* and *Swood*, parameters that are part of the interception module, were found to be among the most sensitive parameters of the model. In section 5.4, we gave a reason for this sensitivity, as it highly impacts evapotranspiration and with that discharge. In addition, the temporal scalability of these parameters (here run on a 6-h time step) is possibly an issue too. The wflow_sbm model switches from the analytical Gash model (*Gash*, 1979) on daily time steps to a modified Rutter (*Rutter et al.*, 1971, 1975) model on sub-daily time steps. This can already cause non-scalable behavior, especially for interception related parameters (*Ficchì et al.*, 2019), possibly leading to a higher sensitivity to these parameters on a 6-h time step, as this is also at the transition between the two approaches. An improved interception module, that uses a constant approach on different time scales, is therefore recommended.

Parameter estimates

Using point-scale PTFs from literature, a large part of the measured variability (expressed in relatively high KGE/NSE values) is explained. This minimizes the need for further calibration, although of course an MPR calibration could potentially further improve the outcomes. This is left for future work. The advantage of the approach taken here is that the transfer-functions are no longer constrained to the model and regional discharge data, giving, in theory, even more potential for regionalization. However, also PTFs have a constraint, which are the field samples they are based on. These field samples are generally limited to certain regions in the world, which makes the empirical relationships derived in the laboratory not necessarily representative for all river basins. For the estimation of especially θ_s, θ_r, *KsatVer* and *c*, many more PTFs are available. So, the choice of a modeler for a certain point-scale PTF can then depend on the area of interest.

In addition, some of the point-scale PTF implementations can still be improved. By making use of *KsatVer* estimations at seven depths and a simple exponential fitting, a rough estimation of the M-parameter is made. Conceptually, M is the exponential decay of *KsatVer* with depth, so this approach suits the concept. However, with the wealth of soil information on seven depths and therewith also *KsatVer* estimations on seven depths, the M parameter is not strictly necessary. For shallow soils, the saturated conductivity with depth can also be estimated with the available information and in that way also allow for e.g. increasing saturated conductivity values with depth, instead of only a decay.

Moreover, N_{river}, Manning's roughness coefficient for water flow in watercourses, has not been found to be a sensitive parameter, but it may become more important for larger river systems such as the main course of the Rhine. In this study, we have used a method which relates the Strahler-order of a river cell to a coefficient value for N_{river} (Table 5.1) based on *Liu et al.* (2005). This method is, however, not often tested and therefore it is uncertain whether these values are universally applicable or only locally applicable.

Validation of wflow_sbm model for the Rhine

In this section, we discuss the resulting quality of the wflow_sbm model on the highest resolution of interest (1200 m) as assessed on simulated discharge (5.5) and evapotranspiration (5.5) after the parameter estimations and upscaling procedures were applied.

Discharge

The results at Lobith, at the Dutch-German border, are comparable to results found with the (MPR calibrated) mHM-model in an earlier study for the Rhine by *Rakovec et al.* (2016a) (supplement Figure C.8; note that the used forcing is different). With a KGE just under 0.9 for both models, discharge is generally well simulated. One difference is that the first peaks during fall seem to be better captured by wflow_sbm, as the results with mHM give too flashy responses with a high recession rate afterwards. Winter peaks are reasonably well captured by both models in the comparison.

The results are also comparable to discharge simulations by *van Osnabrugge et al.* (2017) with wflow_hbv (a distributed version of HBV within the wflow framework), where five model parameters were derived with a basin by basin GLUE analysis approach. In that study, the same meteorological forcing was used. In the Alpine headwaters, lower KGE values were attained as well. Discharge performance in the river Main is better than for wflow_hbv, while the opposite is true for the river Ahr (central region of the basin) and the river Ruhr (northeast). Other regions are similar in terms of attained KGE for discharge simulations, see also (*van Osnabrugge et al.*, 2017).

Although discharge simulations and therewith model performance are promising with this modeling approach, discharge simulations are insufficient in the Alpine region. We see five likely reasons for this behavior of the model: (1) incorrect lake 'operation' representation, (2) used model parameters, mainly *KsatHorFrac* and snow parameters, (3) the model forcing in this area, (4) the chosen model resolution and (5) the river and drainage network derivation procedure.

In Figure 5.3b, we saw that some of the lakes in Switzerland are not well modeled, resulting in lower KGE values. Additional time series of lake levels are present in the supplementary material (Figure C.3). Part of this can be attributed to a mismatch in simulated discharge volumes and timing from the upstream headwaters, but e.g. the Murtensee is an example of a reservoir for which the release of water clearly occurs in the wrong season leading to a clear mismatch between the observed and modeled lake levels. Improving the reservoir operation schemes, which were obtained from the operational Rhine model, is therefore a necessary additional improvement.

It is obvious that the used approach of a uniformly applying default snow parameter values is insufficient, as also the SWE estimations are on average a factor two lower than the SLF estimates (see also Figures C.13–C.15). Ideally, these parameters are estimated with transfer-functions. Day-degree parameters are hard to relate to physical characteristics (*Seibert*, 1999), although some studies have tried to link the day-degree factor to land cover, elevation, slope and aspect (e.g. *Kuusisto*, 1984; *Semádeni-Davies*, 1997; *Dunn and Colohan*, 1999). As a calibrated transfer-function, relating the day-degree factor to land cover was already done by *Kumar et al.* (2013a).

In addition, *KsatHorFrac* was found to be a sensitive parameter. This parameter has a pronounced effect on the partitioning of discharge in quick runoff and baseflow. A higher value for *KsatHorFrac* leads to a higher baseflow, lower peaks and a somewhat slower recession after a discharge peak. Besides the need for a calibrated transfer-function for this parameter, the uniformly applied value of 250 may have been too high for many regions in the Rhine basin (see e.g. Figure 5.4). Especially for the Alps, quick responding catchments are often not well modeled with too high baseflows and too low or even missed discharge peaks. In this region, a lower value for this parameter can increase model performance. This is tested for the Thur basin (supplementary Figure C.9), where four different *KsatHorFrac* values (10, 100, 250 and 1 000) are implemented. With decreasing *KsatHorFrac*, peaks get higher with steeper recession curves and generally a lower baseflow. For a *KsatHorFrac* of 10, the peaks that were highly underestimated in Figure 5.4, are even overestimated. With a calibration procedure, we expect that a value in between 10 and 100 will be found, leading to increased model performance for this (and similar Swiss) basin(s).

Mainly in the south west of the Alps, forcing errors played a role. In this region, mostly located in Switzerland, the observation network used for the forcing is not dense enough to capture the high rainfall variability in this mountainous region. The result is an underestimation of precipitation for this region compared to other precipitation data sets (*van Osnabrugge et al.*, 2017). Hence, besides the parameterization, this is a second reason for the mismatch between modeled and observed discharge in this region, especially concerning the volume bias during the winter period (October - March).

Within this approach, a high resolution model on 1200×1200 m was used which led to discharge simulations that well resembled the observations in most of the Rhine basin. However for the Alpine region, this grid cell resolution may still be too large to well capture processes such as the lateral subsurface flow representation. It is possible that other parameters, e.g. *KsatHorFrac*, are compensating for this now.

Another issue related to resolution, is the derivation of the river and drainage direction and network. This derivation was conducted on the model resolution, while ideally this takes place on the DEM resolution, followed by an upscaling procedure keeping the subgrid information available on the model resolution. What we see with the current model, is that discharge and subsurface lateral flow simulations are not scale independent (Figures 5.9 and C.7). This is certainly related to the fact that the river network is not properly scaled. Section 5.5 focuses further on this scalability. For the Rhine and especially the Alps, it is needed to improve the derivation of river and drainage direction networks. This is currently under development.

Actual evapotranspiration

In this study, we have compared modeled ET_{act} values with LSA SAF DMET estimates by *Trigo et al.* (2011). Note that this is a comparison of two models and care should be taken when interpreting the results or drawing hard conclusions on the basis of this comparison. In our opinion, however, this is a valuable addition to the validation procedure of hydrologic models since both models estimate ET_{act} in a different manner. An important difference is that DMET estimates the plant available water with remotely sensed soil moisture status estimations, while in wflow_sbm, plant available water is the remaining precipitation from the forcing in the unsaturated buckets that are reachable by the

plants roots. Hence, the driving boundary conditions for root-zone water availability, i.e. remotely sensed soil moisture or precipitation estimates, can already be substantially different, leading to e.g. the bias between the two model estimates.

Simulated average yearly ET_{act} differed considerably between the models with 427 mm a year for DMET and 527 mm a year with wflow_sbm. *Hurkmans et al.* (2008, 2009b) have found a yearly average of 659 mm from lysimeters for the Rheindahlen (near the Dutch border), 541 for the Rietholzbach (part of the Thur basin), and on average for the land use types deciduous forest, grassland and crops values of: 491, 659 and 398 mm. This indicates that DMET estimates are too low for the area. Nevertheless, day to day variability of ET_{act} for the three sub-basins is comparable for both models (Figure 5.6).

Scalability of the model

In this section, we focus on the scale (in)dependency of the resulting wflow_sbm model. How important were the upscaling rules (5.5) and what is still necessary to reach flux preservation for especially the discharge related fluxes (5.5)?

Upscaling rules

The application of appropriate parameter-specific upscaling rules are of key importance within this approach. Both in Figure 5.7 and in the supplementary Figures C.4–C.6, some examples are given of the effects of the upscaling rules on obtained parameter fields. The results show a high degree of consistency in the parameter fields, comparable to e.g. Figure 3 in the study by *Samaniego et al.* (2017). Therefore, we argue that consistent parameter fields on different resolutions are reachable with a combination of a-priori PTFs and the right upscaling rules.

The results of this study have shown that saturated hydraulic conductivity (and lambda) fields, are well upscaled with an arithmetic mean of their natural logarithm. This choice has followed from their log-normal distribution and the used PTFs which are formulated as an exponential function (Table C.2). In practice, however, the difference between upscaling with the arithmetic mean of the natural logarithm or just the arithmetic mean of the normal values is small (but present) (Figure C.6). Possibly, this is caused by the high standard deviation in the parameter estimations, partly caused by the high natural spatial variability in the saturated conductivity, which makes the effect of different upscaling operators less pronounced.

In addition, saturated and residual water content parameters are upscaled with a harmonic mean in the mHM model (*Samaniego et al.*, 2010; *Kumar et al.*, 2013a,b; *Rakovec et al.*, 2016b,a; *Samaniego et al.*, 2017). Within our approach, upscaling of these parameters has taken place with an arithmetic mean. In the supplement (Figure C.4), we have compared both approaches and they give similar consistencies in parameter fields on different resolutions, although with a slightly different mean for the two approaches. Hence, it seems that the chosen upscaling procedure (i.e. harmonic or arithmetic) does not matter that much for these two parameters.

Flux preservation on different resolutions

In theory, the choice of the right upscaling procedure per parameter should lead to the preservation of modeled fluxes and states on different spatial resolutions (*Samaniego et al.*, 2010; *Kumar et al.*, 2013a,b; *Samaniego et al.*, 2017). In this study, similar results are found for ET_{act} flux estimates on the four applied spatial resolutions (*Samaniego et al.*, 2017, e.g. 0–10% in Figure 7 of). Only on a sub-basin scale, we have found similar results for recharge fluxes, while on the grid cell scale, the variability is higher.

In contrast to those results, we have not found flux preservation for discharge related fluxes (e.g. river runoff and subsurface lateral flows), especially for headwaters (Figures 5.8 and C.7). This is particularly the case for small basins (mostly headwaters) that rely on well scaled slopes, and well scaled river and drainage networks. In areas where the network is dense and not well scaled to the model resolution, this leads to a river network and lateral subsurface flows that have a slightly different direction or location (in case of the network) than at higher resolution, which can lead to large relative errors in spatially accumulating fluxes. This is well visible in Figure C.7 of the supplement.

For mHM, however, discharge simulations have given quite consistent results on both finer and coarser spatial resolutions, with a maximum efficiency at a simulation and calibration scale of 8 km (*Samaniego et al.*, 2010; *Kumar et al.*, 2013a). Hence, this shows that it is possible to obtain discharge flux preservation on different spatial scales. However, that this is possible with a model that allows for lateral subsurface flows has not been shown yet. For wflow_sbm, it requires a structural change in the (current) automated process of slope, river and drainage network derivation from e.g. DEM information. Ideally, this takes place on a sub-grid scale, as the local drainage directions, determined at the model resolution, probably do no longer capture all drainage processes well on the coarser model resolution, leading to erroneous representations of rivers and drainage networks.

Finally, we have not tested the approach on finer resolutions than 1200×1200 m. *Kumar et al.* (2013a) have shown that fluxes remain more or less preserved when moving towards higher resolutions than the used calibration scale. Although this is not tested here, we expect to obtain similar results or even increasing model performance when it comes to simulation of discharge or lateral subsurface flows with the current version of wflow_sbm. This is based on the aforementioned representation of the river and drainage network, which is expected to be better on higher resolutions.

5.6 Conclusions

Moving towards high-resolution spatially distributed hydrologic models asks for a different approach in estimating parameter values. In this paper, we combined parameter estimates from point-scale (pedo)transfer-functions (PTFs) with a high spatial and temporal resolution model (wflow_sbm) for the Rhine basin. Following previous studies by among others *Samaniego et al.* (2010), *Kumar et al.* (2013a), *Kumar et al.* (2013b), *Rakovec et al.* (2016a), *Mizukami et al.* (2017) and *Samaniego et al.* (2017), we set up the parameter estimates on the original data resolutions and upscaled derived parameter fields to various model resolutions by making use of suitable upscaling operators.

The objective of this study was to investigate the applicability of point-scale (pedo)transfer-functions in combination with suitable upscaling operators for deriving seamless hydrologic model parameters and the preservation of fluxes across scales for a multi-scale hydrologic model for the Rhine River. We focused on three aspects: (1) Can we cover the sensitive parameters of the wflow_sbm model with point-scale PTFs from literature? (2) To what model performance does this setup lead on the highest resolution of interest (1200 m)? (3) How scale (in)dependent is the resulting model on four resolutions (1.2, 2.4, 3.6 and 4.8 km)?

With point-scale PTFs from literature, we were able to estimate almost all sensitive parameters of the wflow_sbm model. The two remaining parameters, together with the relatively insensitive parameters, were set to default-values and were applied homogeneously over the entire Rhine basin. This gave us a model setup with parameter values, either estimated or fixed, which made it possible to model the hydrology of the Rhine basin without any further calibration, while obtaining similar performance as calibrated models (i.e., mHM and wflow_hbv).

The used parametrization of the model led to promising results. For most discharge gauging stations in the central and northern part of the Rhine basin, KGE values ranged from 0.6 to 0.9. An exception to this result was the Alpine region, where lower KGE-values were attained, especially in the southwest of the basin. For the greatest part, this could be attributed to a volume bias in the simulated discharge. This volume bias in the upper part of the basin transferred downstream in the main course of the Rhine and we attribute this to a biased forcing in this region.

The validation of evapotranspiration simulates took place with a comparison with LSA SAF DMET estimates. Model simulates were systematically higher than for DMET with 100 mm y^{-1}, although DMET was found to have too low estimates for the Rhine region compared to values from literature. Spatial patterns between the two models were somewhat similar for the Alps, northern Switzerland and the downstream area of the basin. Between the other regions, however, many differences were present, especially in the Rhine valley and the surrounding hills (SPAEF = 0.51). Here, spatial patterns of higher and lower evapotranspiration were often the opposite of the DMET estimates. Contrarily, catchment averaged temporal dynamics for three sub-basins were better with KGE values around 0.7. Hence, although the temporal dynamics were comparable to the DMET estimates, fluxes were highly different between various regions in the Rhine basin.

Results indicated that consistent parameter fields on various resolutions could be obtained with the used method. Modeled actual evapotranspiration fluxes remained preserved on coarser spatial resolutions, which is in agreement with the findings of *Samaniego et al.* (2010), *Kumar et al.* (2013a) and *Samaniego et al.* (2017). Recharge fluxes were only preserved for regions where the fluxes were considerably higher than 0. In the other regions, local relative errors of ± 30% (sometimes up to 50%) occurred on a pixel to pixel scale. Catchment averaged fluxes were, however, better preserved on the four resolutions.

Moreover, flux preservation was not found for routed discharge in smaller rivers and headwaters on coarser resolutions, while the routed discharge for the main course of the Rhine was consistent on all four resolutions. The headwaters rely most on well scaled slopes, river and drainage networks. In areas where the network is dense and not well scaled to

the model resolution, this leads to a river network and lateral subsurface flows that have slightly different directions and locations than on the highest resolution. For wflow_sbm, a model that allows for lateral subsurface flows, it requires a structural change in the model process of slope, river and drainage network derivation from spatial information, such as the DEM. Ideally, this takes place on a sub-grid scale in order to better represent the drainage processes on smaller scales. Research in this direction is currently underway.

Chapter 6

Assimilation of multiple lake levels in an operational integrated catchment model of the Swiss Rhine basin

Data assimilation methods have the potential to improve hydrological forecasts by reducing errors in the initial state of the model at the time of forecast. In this paper the potential of using lake level measurements in state updating is explored. By means of a synthetic model experiment and a real world case applying the Ensemble Kalman Filter (EnKF) to a hydrological model of the Swiss Rhine, it was found that lake level measurements contain information that can be related to upstream hydrological states. Hence, there is potential for state updating. For the real world case the results were mixed. Updating multiple states, most notably the store related to fast runoff processes, led to overshoots in the low flow period. Restricting the updates to the lake level and the state related to baseflow processes gave the best results, closely followed by direct insertion of the lake level observations.

6.1   Introduction

Operational streamflow forecasts are an integral part of water management. Operational forecasts are not only set-up for flood early warning (*Reggiani and Weerts*, 2008), but are also beneficial for a number of other purposes, such as providing water level information to the shipping industry (*Meißner et al.*, 2017; *Meißner and Klein*, 2016) and for water allocation issues during dry periods (*Gijsbers et al.*, 2017). A clear example of daily use of streamflow forecasts is hydropower production, where a good forecast can result in improved reservoir management (*Hamlet et al.*, 2002; *Chen et al.*, 2011; *Fan et al.*, 2016a).

Forecasts are inherently uncertain. This uncertainty comes from multiple sources in the forecast model train. Precipitation estimates are the largest contributer to streamflow forecast uncertainty (*Cloke and Pappenberger*, 2009), with only a secondary influence of other meteorological variables such as potential evaporation (*van Osnabrugge et al.*, 2019). In hydrological modelling three sources of model uncertainty are generally identified: model structural errors, model parameter uncertainty and uncertainty in the model's initial state. Although different papers have addressed each of those sources of uncertainty separately (*Wagener et al.*, 2004), in practice the different sources of uncertainty are difficult to separate (*Beven and Binley*, 2014), an issue that accidentality has led to one of the more extended polemics in hydrology (*Mantovan and Todini*, 2006; *Beven et al.*, 2007; *Mantovan et al.*, 2007; *Beven et al.*, 2008).

The influence of state errors on total model uncertainty has long been neglected in hydrological modelling. The reason for this is that hydrological systems are damped dynamical systems, which means that any distribution of initial conditions will eventually converge over time towards a single state, in contrast to meteorological systems where state errors lead to strong divergence. Hence, for hydrological models it is possible to run lengthy calibration runs without reinitializing the model for years. The initial state error is incorporated in the parameter uncertainty as errors from the previous time step need to be corrected by the parametrization of the model for future time steps. The parameters of the model then do not describe the propagation of the model from one time step to the next, but the average propagation over the calibration period as the model has to adjust continuously for any possible past error that has led to an incorrect state (*Vrugt et al.*, 2005).

Theoretical issues aside, in hydrological operational forecasting practice model structure and model parameters are determined offline and are generally fixed for several years due to the costs of the exercise of updating parameters. At the same time the meteorological forecast is usually provided by an external agency. This basically leaves mostly only the initial state of the model as playground for the hydrological forecaster to improve the quality of the forecast. This can be achieved with state updating. State updating is a form of data assimilation (DA) were the discrepancy between observations and model results are mapped onto the model state in near real-time. The model state is then adjusted in such a way that would ideally result in the best possible estimate based on the respective uncertainty of the model and the observations. A growing number of studies have shown that state updating is effective in improving forecasts (e.g. *Xie et al.*, 2014; *Rakovec et al.*, 2012; *Seo et al.*, 2009; *Clark et al.*, 2008).

One difficulty in the data assimilation of hydrological models, is that most hydrological

model states are not directly observable. This is partly because of the difficulty of measuring the quantities involved (e.g. soil moisture), but also largely because of the fact that in almost every hydrological model the model states are highly conceptual, even if they are denoted by physical characteristics (*Beven*, 1989). Still, there is a growing list of observational variables that have been used for state updating of hydrological models, including snow cover, surface brightness, soil moisture and streamflow. For a review and references, see *Sun et al.* (2016).

Another problem when using data assimilation is the ambiguity in the technical implementation of any DA filter, referring to how the outcome of a DA algorithm can be tuned by setting DA parameters such as the ensemble size and, most notably, the presumed uncertainty in the assimilated observations (*Thiboult and Anctil*, 2015). Last, it is uncommon (or unwise) to update all hydrological states. Instead, careful consideration should be given to which states are updated, and which states are ignored by the filter. However, in practice the above settings, which have a large influence on the post-ensemble, are set by means of trial and error, until the modeller is satisfied with the outcome.

As mentioned above, the nature of hydrological models is such that any ensemble of initial states will converge towards one future state without external perturbation. However, the rate of this convergence can differ greatly between different types of hydrological storages. The time it takes for a model to converge after an initial perturbation of states is called the memory of the system. The memory determines the expected lead time of the expected gain, with improvements to be expected to be longer with a larger memory of the system. From this perspective it is most interesting to update states that have the largest contribution to system memory. Snow and groundwater states are states with notably long memories, that have predictive skill on the seasonal time scale (*Koster et al.*, 2010).

The influence of different storages and fluxes on streamflow forecast skill differs from upstream to downstream locations of interest in a river basin. In upstream or headwater areas, future streamflow is dependent on local meteorological conditions and local initial conditions. Further downstream future streamflow also becomes more dependent on the amount of water already in the river system as the concentration time, the time needed for all water from a basin to reach its outlet, increases. Under dry conditions, the prediction of streamflow becomes even more strongly dependent on the determination of the amount and distribution of current water.

Most hydrological storages have been considered for state updating, such as snow, groundwater, and river water level. One source of storage that seems to have been overlooked is the storage of water in natural and artificial lakes. Lake levels have not been assimilated extensively in an operational system, but show clear potential. Lake levels generally change slower than discharge, indicating system memory. Lake levels are easily measurable, in-situ, but also using remote sensing, which could enable application of assimilation of lake levels in global models. For these reasons, we investigate the potential of assimilating lake levels and compare this to the better studied approach of assimilating streamflow.

In the following, first the Ensemble Kalman filter (EnKF) is explained and the issues briefly presented above are addressed with some insights from literature (6.2). Next, an overview is given of the study area and related data (6.2), followed by the experimental set-up de-

scribing the model experiments (6.3). Results are given in Section 6.4, followed by Discussions (6.5) and Conclusions (6.6).

6.2   Data and methods

Formulation of the EnKF

One-dimensional problem

To understand the working and results of the EnKF it is useful to start with the most basic understanding of the data assimilation algorithm. The basic concept of data assimilation is well explained by *Reichle* (2008):

Let m be a scalar model variable with uncertainty (or error variance) σ_m^2, with corresponding scalar observation o with uncertainty σ_o^2. m is called the *prior*. The model estimate m is the output of the model at the moment observation o becomes available. The goal is then to find the most likely estimate $\check{x}$ of the *true state* x based on the available information from m and o.

To do this one has to define an *objective function* J which quantifies the misfit between the true state x, the model state and the observation. If the difference between m and o is quantified by the least-square-error, with the assumption that all errors are random and Gaussian, the objective function becomes:

$$J = \frac{(x - m)^2}{\sigma_m^2} + \frac{(x - o)^2}{\sigma_o^2} \tag{6.1}$$

This can be solved to get estimate $\check{x}$ by taking $dJ/dx = 0$:

$$\check{x} = (\sigma_m^2 + \sigma_o^2)^{-1}(\sigma_o^2 m + \sigma_m^2 o) \tag{6.2}$$

This is typically rewritten as:

$$\check{x} = (1 - K)m + Ko \tag{6.3}$$

or

$$\check{x} = m + K(o - m) \tag{6.4}$$

where $K = \frac{\sigma_m^2}{\sigma_m^2 + \sigma_o^2}$

The term K is called the *(Kalman) gain* and it is now understood that the best estimate $\check{x}$ is the weighted average between the model outcome m and the observation o weighed by their respective uncertainties as expressed by the gain K (note that $0 \leq K \leq 1$). The best estimate $\check{x}$ is also referred to as the *analysis* or *posterior*. If the observation error variance σ_o^2 is *small* compared to the model uncertainty σ_m^2, the gain will be large, and the resulting estimate will be close to the observation. If the observation error variance σ_o^2 is *large* compared to the model uncertainty σ_m^2 the estimate will be close to the model. Equal model and observation error variances produce equal weights ($K = 0.5$), reflecting equal trust in the model and the observation.

Mentioning 'level of trust' as a measure that defines the results of an otherwise mathematical method may seem unscientific, where it would be expected that σ_m^2 and especially σ_o^2 are based on a 'true uncertainty' derived from statistical time series analysis

or first principles. However, the true uncertainty is always unknown for the complex dynamical heterogeneous systems dealt with in hydrology, which makes σ_o^2 and σ_m^2 subjective choices reflecting the trust the user has in the observations and model, respectively. In practice, values for σ_o^2 are based on prior data assimilation experience. For example, the error standard deviation for streamflow observations is often set to $\sigma_Q = 0.1Q \cdot \mathcal{N}(\mu, \sigma^2)$. This setting can be traced back to (*Georgakakos*, 1986) who substantiate their assumption only with: "the standard deviation of the observation noise corresponding to the discharge rate was taken to be time varying as a function of the observation measurement according to $\sigma_Q = 0.1Q$, implying higher measurement error for higher flows", without reference given. Since then, the assumption of $\sigma_Q = 0.1Q$ has propagated through literature to become an accepted reasonable estimate that results in decent data assimilation performance (*Weerts and El Serafy*, 2006; *Rakovec et al.*, 2012, 2015; *Mazzoleni et al.*, 2018).

Rewriting Eq. (6.3) as

$$\check{x} - m = K(o - m),$$ (6.5)

two more definitions can be added: the *increment* $\check{x} - m$ and the *innovation* $o - m$. This formulation shows that the increment is proportional to the innovation. In other words, the largest potential for updates in absolute terms is when the difference between o and m is large, but note that $\check{x}$ will always be in between the values of o and m. Equation 6.5 is also called the *update equation* as the model estimate is updated with information from the observation. In forecasting the whole activity of propagating a model forward in time can be referred to as *update step*, which includes running the model to generate m and running a data assimilation scheme to produce $\check{x}$.

Finally, if the errors in the model output and the observation are uncorrelated the error variance of the assimilation estimate for this example with the least-square-error is:

$$\sigma_{\check{x}}^2 = (1 - K)\sigma_m^2 = K\sigma_o^2$$ (6.6)

Hence, $\sigma_{\check{x}}^2$ is smaller than the variances of the model and observation (as $0 \leq K \leq 1$), reflecting how data assimilation has increased knowledge about the true state x.

EnKF matrix formulation

Environmental problems usually have more than one model outcome and more than one observation. However, Eq. (6.3) can be easily expanded by writing it as a matrix multiplication:

$$\check{\mathbf{x}} = \mathbf{m} + \mathbf{K}(\mathbf{o} - \mathbf{m})$$ (6.7)

In which $\check{\mathbf{x}}$, $\mathbf{m}$, $\mathbf{o}$ and $\mathbf{K}$ are know vectors of length [mx1], with m the number of states, which equals here the number of observations. The Kalman gain becomes:

$$\mathbf{K} = \mathbf{P_m}(\mathbf{P_m} + \mathbf{R_o})^{-1}$$ (6.8)

$\mathbf{P_m}$ and $\mathbf{R_o}$ are the covariance matrices for the error variances of the model and observations, respectively.

In the above case, there is a direct observation available for every model state. In any environmental case the number of model states is much higher than the number of observations.

$$\mathbf{\Phi}_i^a = \mathbf{\Phi}_i^f + \mathbf{K}(\mathbf{d}_i - \mathbf{H}\mathbf{\Phi}_i^f), i = 1, \cdots, N \tag{6.9}$$

Here, $\mathbf{\Phi}_i^a$ [mx1] is the analysis and $\mathbf{\Phi}_i^f$ the model forecast [mx1]. The update procedure is still the same as in Eq. (6.7), with the difference that now the perturbed observations $\mathbf{d}_i$ [nx1] and the state vector $\mathbf{\Phi}_i^f$ [mx1] have different dimensions. Hence, a transformation is needed such that only the states with matching observations are selected. This is done with the observation matrix $\mathbf{H}$ [nxm]. Because discharge is not a linear function of the model states, the discharge is added to the state vector when discharge is assimilated. This is known as state augmentation. Now that the discharge is treated as a state, the observation matrix is a simple matrix filled with ones and zeros to match the observed discharge with the modelled discharge at certain points.

The Kalman gain [mxn] also needs to be adjusted to take into account the fact that there are only limited observations:

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1} \tag{6.10}$$

$\mathbf{P_m}$ from Eq. (6.8) has now been replaced with two terms: $\mathbf{P}^f\mathbf{H}^T$ and $\mathbf{H}\mathbf{P}^f\mathbf{H}^T$. $\mathbf{P}^f\mathbf{H}^T$ is the covariance between all states in the full forecast vector ($\mathbf{\Phi}_i^f$) and the limted number of those states that make up the forecasted observation ($\mathbf{H}\mathbf{\Phi}_i^f$). $\mathbf{H}\mathbf{P}^f\mathbf{H}^T$ is the covariance of the forecasted observations with themselves. This is expressed in the formulas (*Houtekamer and Mitchell*, 2001):

$$\mathbf{P}^f\mathbf{H}^T \equiv \frac{1}{N-1}\sum_{i=1}^N (\mathbf{\Phi}_i^f - \overline{\mathbf{\Phi}^f})(\mathbf{H}\mathbf{\Phi}_i^f - \overline{\mathbf{H}\mathbf{\Phi}^f})^T \tag{6.11}$$

$$\mathbf{H}\mathbf{P}^f\mathbf{H}^T \equiv \frac{1}{N-1}\sum_{i=1}^N (\mathbf{H}\mathbf{\Phi}_i^f - \overline{\mathbf{H}\mathbf{\Phi}^f})(\mathbf{H}\mathbf{\Phi}_i^f - \overline{\mathbf{H}\mathbf{\Phi}^f})^T \tag{6.12}$$

With these equations for $\mathbf{P}^f\mathbf{H}^T$ and $\mathbf{H}\mathbf{P}^f\mathbf{H}^T$ the Kalman gain can be calculated directly from the ensemble forecast ($\mathbf{\Phi}_i^f$) with ensemble mean $\overline{\mathbf{\Phi}^f}$.

OpenDA

Implementation of the EnKF formula is quite easy, but optimizing the ensemble calculation process is much more complex. The open data assimilation library (OpenDA) is a toolbox to link models with a library of data assimilation techniques (*Ridler et al.*, 2014). The OpenDA library also has methods to perform the perturbation of spatial forcing fields and observations (*OpenDA Assiociation*, 2016).

Study area, data and model

Study area

The Rhine river (area approx. 160.000 km^2, see Chapter 5.2) has a mixed pluvial-snowmelt flow regime (Fig. 6.1). The largest contribution of snow melt to river flow is from the Swiss
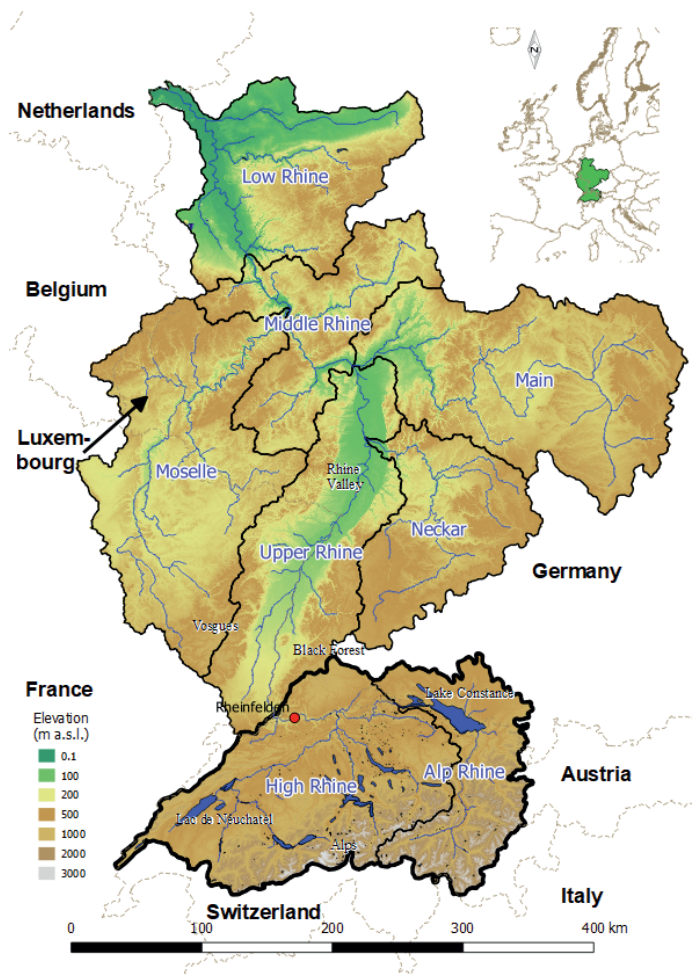
Figure 6.1: Map of the Rhine basin, Europe. Thick black line delineates the Swiss part of the Rhine basin. Rheinfelden is the main station at the outlet of the Swiss Rhine basin.

Alps area in the south of the basin. Another notable feature of this area are the large natural lakes. Furthermore, the Swiss region of the Rhine basin has the highest yearly precipitation. During the months June-July the Swiss part of the Rhine contributes on average almost 70% to total streamflow at the Dutch-German border (Fig. 6.2). Almost all of the water that leaves the sub-basin at Rheinfelden is routed through one of the Swiss lakes.

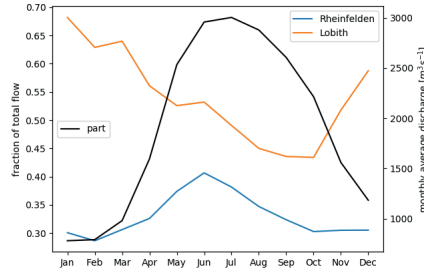*Figure 6.2: Percentile contribution (black line, left y-axis) of the flow of the Rhine near the Swiss border at Rheinfelden (blue line, right y-axis) to the total monthly streamflow at the Dutch-German border (Orange line, right y-axis). 20 year average 1996-2016.*

Data

Observational data has been preprocessed for use with a grid-based hydrological model. The data was processed with hourly time resolution, on a 1.2×1.2 km grid spatial resolution, for the period mid 1996 through 2015. All source data to derive the gridded estimates comes from sources that supply their data in near real-time, making the datasets suitable for operational forecasting. For this study all data was aggregated to a daily time step. The hourly datasets are downloadable through the 4TU data centre (*van Osnabrugge*, 2017, 2018).

The precipitation data set has been derived using the genRE interpolation method based on ground measurements and the HYRAS (*Rauthe et al.*, 2013) climatological precipitation dataset (*van Osnabrugge et al.*, 2017). The temperature is interpolated using a fixed lapse rate to correct for height differences between stations and grid cells, and potential evaporation is calculated with the Makkink formula based on those temperature fields, in combination with satellite-based solar radiation estimates from the CM-SAF (Climate Monitoring Satellite Application Facility) and LSA-SAF (Land Surface Analysis Satellite Application Facility) (*van Osnabrugge et al.*, 2019).

Forecast data is obtained from the European Center for Medium-Range Weather Forecasts (ECMWF). ECMWF issues hindcasts produced with the current model cycle for certain days for the last 20 years. The reforecast obtained for this study was produced with model cycle 43r1 (*Buizza et al.*, 2017). The first forecast is on 1996-03-10 and the last forecast on 2015-12-29 with reforecasts alternating every three or four days.

Forecasted Makkink potential evaporation was calculated based on the 2-meter temperature, $t2m$ (T), and the surface solar radiation downwards flux, $ssrd$ (R_g), variables. Temperature was first downscaled to the model resolution using the standard lapse rate, similar to the interpolation of the observed temperature fields. The process is the same as in *van Osnabrugge et al.* (2019).
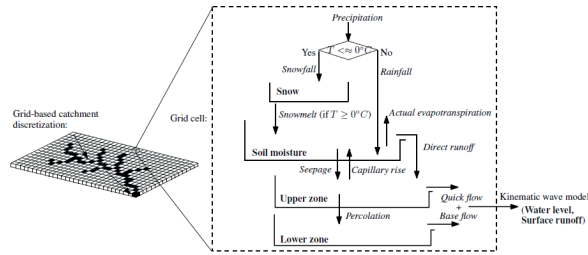
Figure 6.3: Catchment discretization and schematic structure of the HBV-96 model for each grid cell. Model states are in bold and model fluxes in italics (Rakovec et al., 2015).

Hydrological model

Wflow is a modular hydrological modelling framework that allows for easy implementation and prototyping of regular grid hydrological model concepts in python-pcraster (*Schellekens et al.*, 2018). The hydrological model concept used is the HBV96 (Hydrologiska Byråns Vattenbalansavdelning) model concept (*Lindström et al.*, 1997a) applied on a grid basis (Fig. 6.3). The generated runoff is routed through the river network with a kinematic wave approach (*Schellekens et al.*, 2018). In the following this model is referred to as wflow_hbv. The set-up of the hydrological model is the same as used in assessing the validity of the genRE precipitation data set (*van Osnabrugge et al.*, 2017), and in assessing the influence of potential evaporation forecasts (*van Osnabrugge et al.*, 2019). The model was parameterized through calibration with a Generalised Likelihood Uncertainty Estimation (GLUE)-like procedure (*Beven and Binley*, 1992), using HYRAS precipitation as forcing data (*Winsemius et al.*, 2013a,b).

Lake modelling

Of special attention is the modeling of the Swiss lakes in the hydrological model. Eight lakes are modelled (Fig. 6.4), of which specific properties are listed in Table 6.1.
Lakes are initialized based on two spatial maps: 1) lake area, and 2) lake location. The lake area map is simply a pixel map in which the pixels that make up the surface area of the lake are numbered with each number corresponding to a lake. The lake location map shows a single pixel for each modelled lake. This shows the model at which locations in the stream network the lake model routine has to be run. The lake surface area is used to calculate lake evaporation and direct precipitation into the lake, but for water balance purposes the lake is compressed in one pixel of the stream network. This is most logically the most downstream area of the lake where the outlet is located. Lakes are modelled in different ways depending on the available information.
The first difference is the link between water level and water volume stored in the lake. For lakes for which no information is available on the relation between water level and surface area (denoted in Table 6.1 with ResStorFunc=SIMPLE), the assumption is of a simple rectangular reservoir and $S = A * H$, with S the storage, A the surface area, and H the water level. Note that, since the water levels are given in height above mean sea
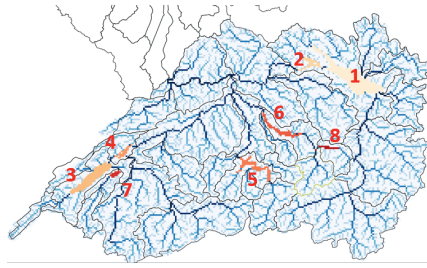
Figure 6.4: Location of the eight modelled lakes on top of the stream network. 1. Bodensee (Upper), 2. Bodensee (Lower), 3. Lac de Neuchâtel, 4. Bielersee, 5. Vierwaldstättersee, 6. Zürichsee, 7. Murtensee, 8. Walensee.

Table 6.1: Table listing the model properties of the modelled lakes. The columns list which type of functions are used to model the surface area as function of water level (ResStorFunc), the outflow as function of water level (ResOutflowFunc) and which lakes are linked with each other and with which lake (Linked).

#	name	ResStorFunc	ResOutflowFunc	Linked
1	Bodensee (Upper)	LUT	SDC	2
2	Bodensee (Lower)	LUT	LUT	1
3	Lac de Neuchâtel	SIMPLE	SDC	4
4	Bielersee	SIMPLE	LUT	3
5	Vierwaldstättersee	SIMPLE	LUT	-
6	Zürichsee	SIMPLE	LUT	-
7	Murtensee	SIMPLE	LUT	-
8	Walensee	SIMPLE	SDC	-

level, the calculated volume is not an indication of the actual volume of water stored in the lake, but a practical way to implement $dH = dS((H)/A$. For some lakes (denoted in Table 6.1 with ResStorFunc=2) the relationship between water level and storage is given in a lookup table, $S = LUT(H)$, which should reflect actual stored volumes.

The second difference is how the outflow of the lake is calculated based on the water level. For most lakes (ResOutflowFunc=LUT in Table 6.1), the outflow is determined with a lookup table $Q_{out} = LUT(H, JDAY)$, with $JDAY$ the day of the year. This enables simple control rules to be implemented in the modelling. For example, one water level can lead to different outflows depending on water level targets and human control. The other lakes (ResOutflowFunc=SDC in Table 6.1) are modelled with a stage-discharge relationship:

$$Q_{out} = a(H - H_0)^n, \tag{6.13}$$

with H_0 the outflow threshold and a and n parameters that are fitted to observed water levels and observed downstream discharge.

In the case of linked reservoirs, the water levels of the linked reservoirs are compared. If the water level of the downstream lake is higher than the upstream lake level, Eq. 6.13 is then used to calculate a flow from the lower reservoir to the upper reservoir. H is then the

water level of the lower lake (which is higher) and H_0 is the threshold of the upper reservoir. The upper reservoir then has no outflow (but additional inflow, as negative outflow Q_{out}).

After determining all fluxes the storage is calculated with the water balance:

$$S_{t+1} = S_t + Q_{in}\Delta t + P_{lake}\Delta t - PET_{lake}\Delta t - Q_{out}\Delta t \qquad (6.14)$$

With the mentioned considerations, the computation algorithm is as follows:

For a location x, y on the stream network that is designated a lake, the following steps are performed:

- determine the inflow from upstream into the lake node;

- determine water level differences with any linked reservoirs;

- calculate lake level storage from the water level state, based on either a rectangular assumption or a lookup table, linking water level with storage if available;

- calculate the outflow based on a) a lookup table linking water level and discharge or b) a stage-discharge relationship. Based on the water level differences with linked reservoirs water can also flow back into upstream lakes using the same stage discharge relationship with the water level between the lakes as input;

- calculate the new storage based on the water balance;

- calculate the new water level state using the same relationships as earlier when determining the storage;

- add the calculated outflow to the downstream pixel in the stream network.

Before the algorithm above is started, a check is done on the time step. For any time step above 6 hours the time step for the lake level calculation is reduced to prevent instabilities related to the linked reservoirs, which can show numerically induced oscillation if the time step is too large.

6.3 Experimental set-up

To test the usefulness of assimilating lake level data in our hydrological model, a twin experiment is performed. With 'useful' we mean here that 1) the algorithm works properly and a better estimate of the true state is obtained as described by Eq. 6.6, and 2) the algorithm is able to update states with longer memories such as upstream groundwater related states. In this twin experiment the application is first investigated using a complete model based example (synthetic) so that all states are known beforehand. Only after the added value of assimilation of lake levels has been evaluated in the idealised modelled case, a real world experiment is performed in which actual water levels are assimilated.

Synthetic experiment

The synthetic case is run for one of the eight modelled lakes. For the synthetic case the Walensee was chosen as subbasin, since it is located upstream and small, but still has all the characteristics which complicate hydrological modelling in the Swiss Alps, such as high spatial gradients, snow, and glaciers. In addition, the lake is fed by two tributaries,

one of which is routed through by means of a channel. Since the synthetic case is meant to show the workings of the filter, the run time is limited to one year with an additional two months of spin up (2002-10-01 till 2004-10-01). The analysis period is 2003-01-01 till 2004-10-01, which is known as a particularly dry year.

The set-up is as follows (Fig. 6.5):

A 16 member ensemble is created by perturbing the forcing fields with the openDA software. Precipitation is perturbed with a multiplicative noise with $\sigma_P = 0.5$, zero time correlation between days and a horizontal correlation scale of 10000 m. Temperature is perturbed with an additive noise with $\sigma_T = 2°C$, zero time correlation between days and a horizontal correlation scale of 5000 m. The short decorrelation distances are chosen to reflect the large spatial variability in forcing errors in a mountainous region. Potential evaporation is not perturbed. The ensemble of meteorological forcings is then fed to the wflow_hbv model to generate an ensemble of hydrological states. From the ensemble then a single member is chosen to present the model-truth, including discharge, lake level and all other internal states of the model.

Next, a new ensemble is formed with the same uncertainty settings and also run through wflow_hbv to create a model ensemble forecast. The forecast is then updated to the posterior using the EnKF in openDA, using observations derived from the model-truth. This then leads to the posterior ensemble, which is analysed with respect to the model truth and the original ensemble simulation of the model-truth. The observations are derived from the model-truth by adding some noise. Discharge is perturbed with noise $\sigma_Q = 0.1Q$, and water level with $\sigma_W = 0.02m$. The synthetic case is an idealized case, not only because all uncertainties are known, but also because the data assimilation update is performed with the true uncertainties.

To investigate the usefulness of data assimilation of lake levels, control runs are performed in which downstream discharge is assimilated so that the results can be compared. Additionally, two different sets of states are selected to be updated. In the first only the lake level state is updated, while in the second the upper zone store, lower zone store and river water level model states are also updated. Additionally the experiment is repeated to test the data assimilation approach for robustness with respect to errors in the stage discharge relationship of the lake. In total 10 runs are performed: 2×2 for the combination of assimilated variable with both sets of updatable states, and 3×2 to test the robustness of the result under a specified error in the H_0, n and a parameter, respectively, and again for both the assimilation of lake level and downstream discharge.

Real world experiment

For the real world experiment the assimilation scheme is applied to the Swiss part of the Rhine basin (Fig. 6.1). The uncertainty settings are kept the same, with exception of $\sigma_W = 0.01m$ after it became clear (see Fig. 6.7) that this is closer to the assumed error in the discharge of $\sigma_Q = 0.1Q$ in terms of discharge. The number of ensemble members is doubled to 32. To show the added value of the data assimilated scheme, the data assimilation case is compared with a benchmark. The benchmark is not only the open loop case, but a second baseline scenario is generated in which the lake level states are directly replaced by the perturbed observations (direct insertion). This gives more insight in the
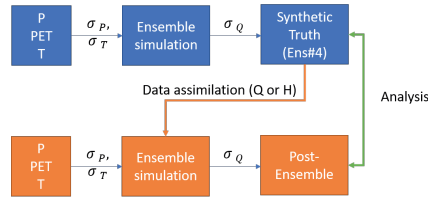
Figure 6.5: Flow diagram of the synthetic experiment set-up. Blue squares give the set-up of the idealised model-truth, while the orange blocks show the steps to regain the idealised truth by means of data assimilation.

added value of using a computationally expensive data assimilation scheme in comparison with simpler methods.

Data assimilated are the lake levels for all eight modelled lakes. Validation takes place based on the downstream discharge at Rheinfelden. Additionally the states of the lakes themselves are evaluated. Forecasts are produced from the updated states with observed forcing (perfect forecast).

6.4 Results

Synthetic experiment

The posterior ensemble converges strongly towards the model-truth in all cases for both the simulation of downstream discharge and for lake level, with strongly decreased RMSE of the ensemble mean and reduced variance. Apparent is the difference in the degree of convergence between the assimilation with lake level (Fig. 6.6, top row) and the assimilation with discharge (Fig. 6.6, bottom row), in which the posterior ensemble is much narrower when discharge is assimilated.

Choices on which states are updated by the filter have effects on data assimilation performance. Differences in behaviour are found between the cases updating the lake level states and three other upstream states (right column in Fig. 6.6) and the cases updating only the lake levels (left column). When only the lake level is included in the state update, the part of the Kalman gain that relates to non-updated states is neglected. This results in a less reduced variance in the posterior for both the assimilated lake level measurements and discharge data. Reducing the degrees of freedom to only the most governing state for the outflow and updating this state directly with an observation of that state resulted in the lowest RMSE error (upper left), while other combinations did improve the RMSE, but to a lesser extent. The reason for this can be understood by comparing the position of the synthetic truth in the posterior ensemble, which is close to the mean for the lake level-lake level quadrant, but is on the bottom for the other cases, resulting in a slight bias.

The difference in convergence between updating discharge data or lake level measurements, has to do with how the observation uncertainties translate into discharge through the stage-discharge relationship. The propagation of the error from the water level obser-
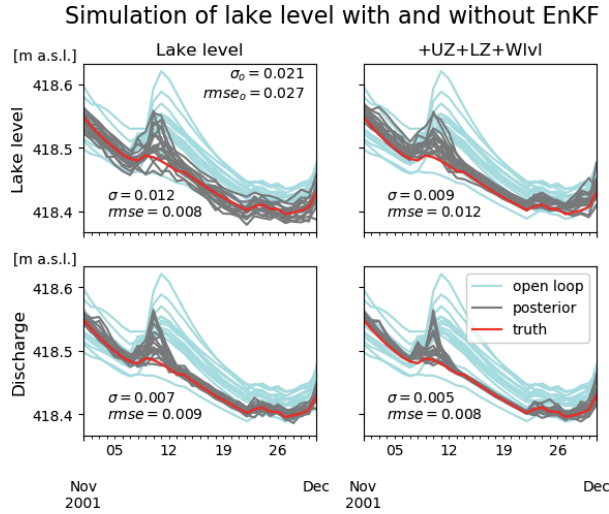
Figure 6.6: Exemplary hydrograph of modeled lake level for different set-ups of the update. The upper row shows the results for updates with the lake level as assimilated variable and the bottom row for the discharge. The left column shows the results for the case where only the lake level state was updated, while the right column shows the results for the case where multiple states were updated.

vation to the calculated discharge can be calculated by applying the rules of error propagation to the stage-discharge formula (Eq. 6.13):

$$\delta Q = |A||n|\frac{\delta H}{|H - H_0|}|Q| \qquad (6.15)$$

Although the standard error on the lake level observation of $\sigma_W = 0.02m$ seems small, a 2 cm difference in water level means a $Q = a(H + \sigma_W - H_0)^n$ difference in discharge. Not only is this of different magnitude (2 cm difference for the Walensee equals +/- 25-30 $m^3 s^{-1}$), but also has non-linear properties as compared to the applied error to the discharge, $\sigma_Q = 0.1Q$, which is assumed to be linearly increasing with the discharge (Fig. 6.7). This leads to the conclusion that the errors on the water level measurement have to be very small indeed ($\sim 0.005m$) to be able to compete with the assumed accuracy of the discharge measurement, hence the larger spread in the posterior ensemble for the case with lake levels.

Propagation of information to upstream states

Data assimilation of lake levels has a positive effect on the modelled water level and the downstream discharge, however, ideally the correction extends to all hydrological model states. Figure 6.8 shows the prior and posterior ensembles for the lower zone storage, which is related to slow flow processes. The update extends to the lower zone storage, as
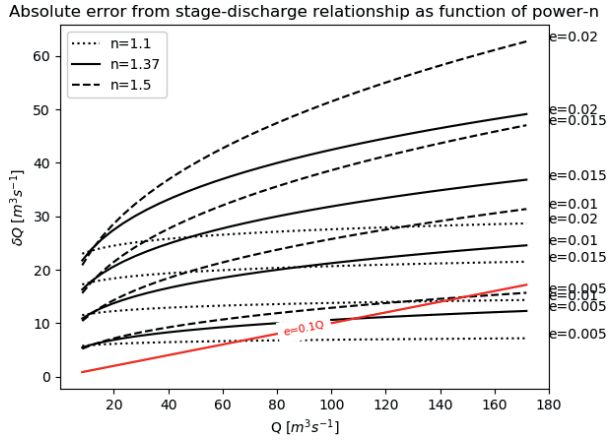
Figure 6.7: *Difference in conceptualization of the measurement error between a widely used assumption, $\sigma_Q = 0.1Q$ (red line), and levels of additive error (e) added to the water level in a stage-discharge relationship, $Q_{out} = a(H - H_0)^n$, for different values of n. δQ refers to the resulting standard deviation as function of the discharge Q.*

the posterior ensemble differs from the prior ensemble significantly. The spread of the ensemble is reduced everywhere in the catchment, but this does not lead to a better representation of the model truth by the mean of the ensemble or as expressed in the mean continuous ranked probability score (MCRPS) overall. There is no defining spatial feature that seems to determine if a single pixel benefited from the update, although there are some areas were improved and non-improved RMSE scores are clustered together. A higher number of ensemble members might lead to a clearer image of good and bad performing areas as random factors are reduced.

Time series for the lower zone storage and snow storage were extracted for six points distributed throughout the basin. For all locations the posterior ensemble is narrower than the prior. However, unlike for the lake water level and the discharge earlier, the model-truth is not always captured by the posterior ensemble, indicating over-estimated certainty in the posterior ensemble.

For updates of the snow store, the results are largely the same as for the lower zone store in that the average RMSE over the basin increases for the data assimilation run. Different is that in the updates for the snow store, a clear pattern emerges where successful updates, marked by positive values in Fig. 6.9, are close to the stream network. It is not possible to say if this is because of the pixels' faster drainage or because the snow pack is of intermediate height as deeper snow packs are found further up the slope.

Robustness of the assimilation

So far the usability of water level information for state updating looked promising: data assimilation of lake levels moved the posterior ensemble closer towards the model-truth for both modelled water level and discharge. An important aspect of the usability is the

State updates of the LowerZoneStorage
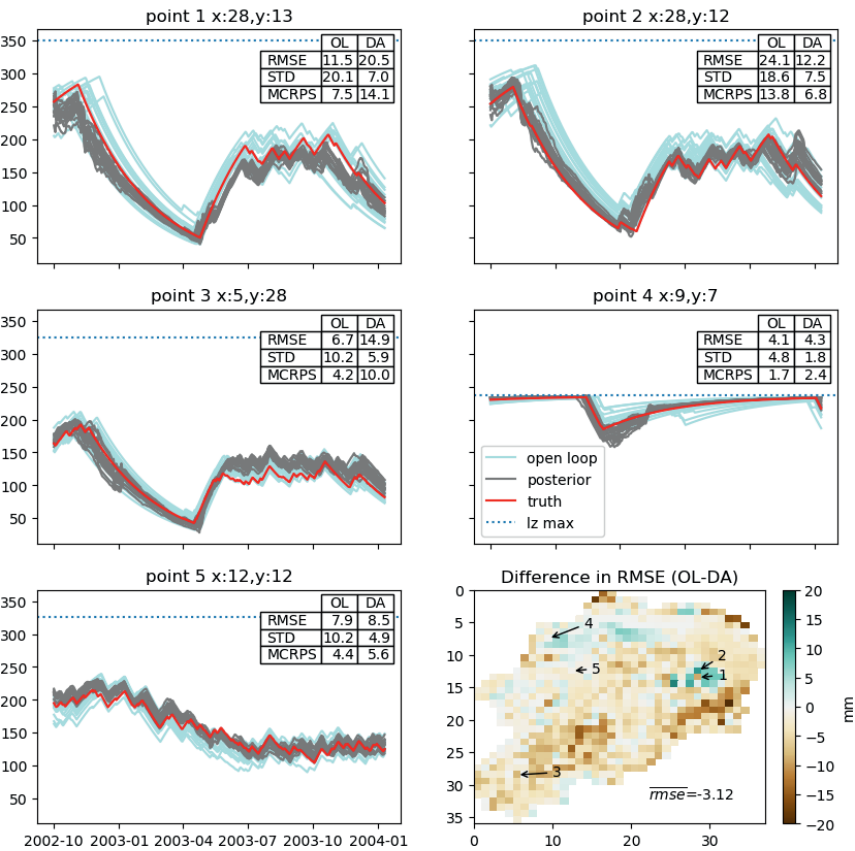
Figure 6.8: State update of the lower zone store by the EnKF. Spatial representation of the gain in RMSE and visually for 5 representative pixels (location shown in the spatial display). Observations (red), open loop (grey) and posterior (light blue). The lz max line denotes the maximum volume the lower zone store can take for each pixel. The tables inside the plots list the Root Mean Square Error (RMSE), Standard Deviation (STD) and Mean Continuous Ranked Probability Score (MCRPS) for the open loop (OL) and the data assimilation (DA) run.

Figure 6.9: State update of the snow store state by the EnKF. Spatial representation of the gain in RMSE and visually for 5 representative pixels (location shown in the spatial display). Observations (red), open loop (grey) and posterior (light blue). The tables inside the plots list the Root Mean Square Error (RMSE), Standard Deviation (STD) and Mean Continuous Ranked Probability Score (MCRPS) for the open loop (OL) and the data assimilation (DA) run.

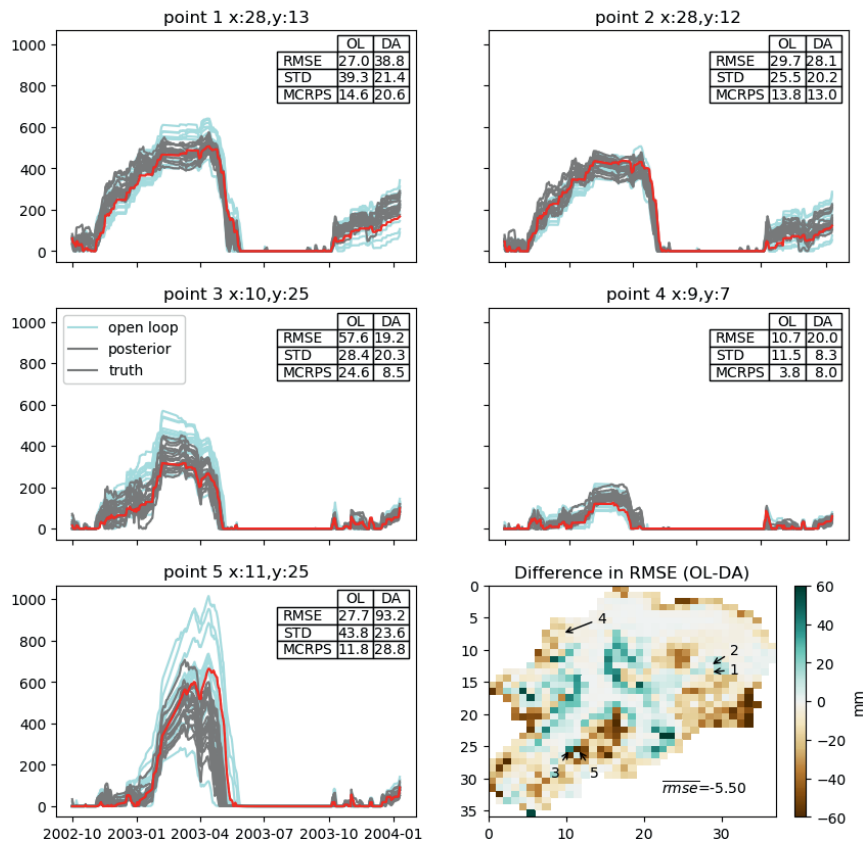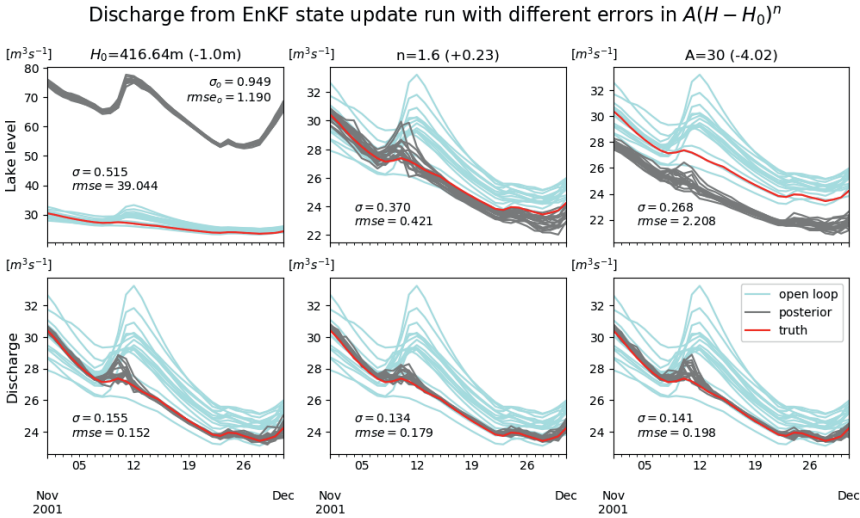Figure 6.10: *Robustness test for data assimilation of lake water levels with respect to modelled discharge. The column headers show perturbations of parameters in the stage-discharge relationship $Q = a(H - H_0)^n$, while the row headers denote the data type that is used for assimilation: lake level (top row) or discharge (bottom row).*

robustness. In particular, it is important to verify if there are risks of unexpected behaviour that might decrease the efficiency of the filter or even lead to worse results than the open loop simulation. The efficiency of data assimilation with lake levels is dependent on the quality of the modelling of the lake (Fig. 6.10). From the three parameters in the stage discharge relationship (Eq. 6.13), the posterior ensemble is most sensitive to the H_0 parameter. Errors in this parameter immediately lead to biases in the estimation of discharge (when assimilating lake water level) or in modelled lake level (when assimilating discharge). The reason for this bias is clear: the filter tries to match the modelled value with the observation. Because in the case shown H_0 is too low this leads to a too large $(H - H_0)$ term and consequently a too high discharge. For the next step this water is again added to the lake in the model to correct for a now too low water level (Fig. 6.10, top row). Importantly, the same is true for water level in case discharge is assimilated.

Also for biases in the other terms of the stage-discharge relationship, A and n, the performance of the data assimilation is more stable when discharge is assimilated. The mean of the posterior ensemble is a better estimation of the true value than the mean of the prior ensemble for all the disturbed cases, whereas this is not the case when lake level is assimilated.

Real world experiment

In the real world experiments the update is performed with actual observed data. The assumed uncertainty in the observations and the perturbations of the forcing to form the

ensembles is the same as in the synthetic case, but now the true uncertainty is of course unknown. Additionally there are no model structural errors in the synthetic experiment.

The posterior ensemble is highly influenced by the assimilation and is distinctly different from the open loop (Fig.6.11). First to note is that the model is 'too dry' and water levels are consequently too low for parts of the year. The quality of the open loop modelling also varies between lakes. This is not necessarily a reflection of the quality of the modelling of the lake itself, but is also influenced by the modelling of the inflow. For the upper and lower Bodensee the modeled water level is too low, but otherwise follows the dynamics of the observations. The linkage of Lac de Neuchâtel and the Bielersee can be clearly seen by the matching open loop simulation of the two lakes. The observations show that the nearby Murtensee is also linked with those lakes, as the dynamics of that lake match the dynamics of the former lakes almost perfectly. This is, however, not represented in the model, and the modelled behaviour of the Murtensee deviates most from the observations of all lakes. That there is no free flow from the Lac de Neuchâtel and Murtensee can also be seen from the observed outflow of those lakes, which at times is negative. The Zürichsee and the Walensee are modelled quite well. The simulation of the water level of the Vierwaldstättersee is also chronically too low.

For the Bodensee (upper and lower), the EnKF filter works as expected as the posterior ensembles are in between the observation and the open loop, reflecting their respective uncertainties. The same works for the Vierwaldstättersee, but the posterior is much closer to the model than to the observation. This has to do with large uncertainties in the lake modelling. As can be seen in Fig. 6.12, the modelled discharge is higher than the observed discharge for most of the year based on direct insertion (the downstream discharge is calculated directly from observed water levels). The positive bias in simulated discharge leads to lower lake levels for the one-day ahead forecast of water level. The issue here is not the data assimilation scheme, but the inadequate modelling of the lake. It is important to note that, just as in the synthetic experiment when errors were introduced in the stage-discharge relationship of the lake (Fig. 6.10), perfectly adjusted water levels would not lead to improved discharges. Instead, errors in lake level can compensate for errors in discharge modelling.

That the assimilation can have large influence is shown for the Murtensee, where the open loop model result has a completely different dynamic, but the updated result follows the general trend of the observations, pushing the modelled lake in line with its neighbouring lakes Lac de Neuchâtel and Bielersee. The reaction of the latter lakes to the EnKF filter is similar.

A large drawback of the assimilation results were the observed high peaks in the posterior ensemble when the upper zone storage is included in the updates (EnKF_W+LZ+UZ). The peaks are most notable in the Murtensee, Bielersee and Walensee. The peaks in water levels also result in peak discharges from the lakes, and at the outlet (see Fig. 6.12,9. Rheinfelden, around 2003-08). The peaks are forced because the filter adds water to the (unbounded) upper zone of the model upstream of the lake. This causes the next precipitation event to not infiltrate but produce direct runoff into the routing and into the lake. This excess of water is then not corrected for in the following update, but needs to be drained by the hydrological model itself, which it does in the form of a peak discharge. The peaks mainly occur in the smaller lakes.
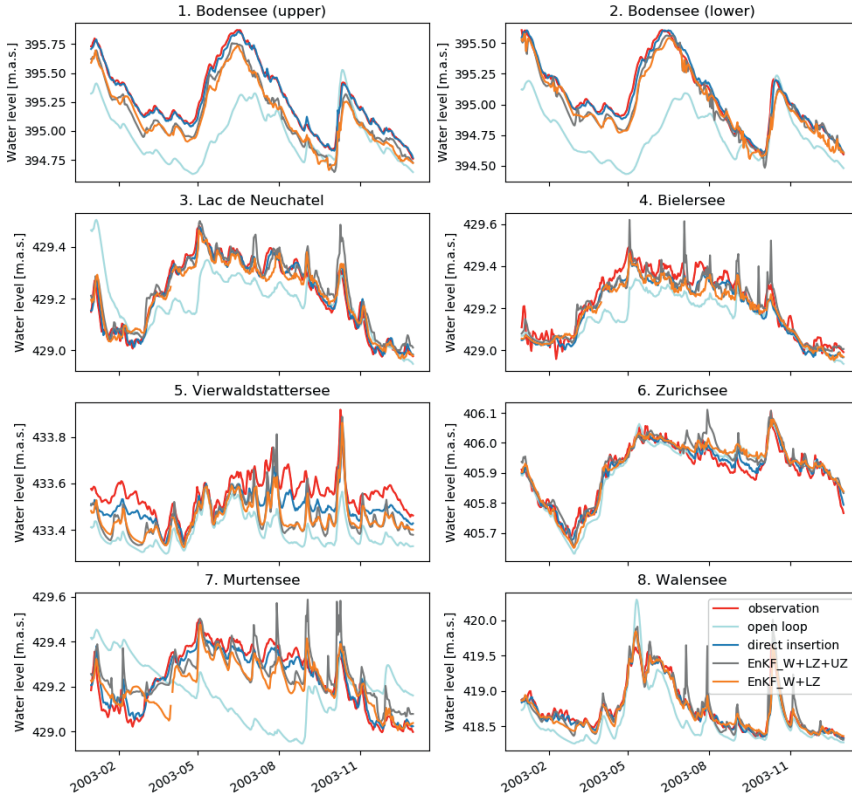
Water level in 8 lakes

Figure 6.11: *Modelled and observed water levels of the eight Swiss lakes. The observed water levels are the one step ahead forecast. Only the means of the ensembles are shown. The labels EnKF_W+LZ+UZ and EnKF_W+LZ refer to the update runs including and excluding the uppder zone store for state updating, respectively.*

Downstream discharge from 7 lakes and Rheinfelden

Figure 6.12: Downstream discharge of 7 lakes and the outlet at Rheinfelden. The labels refer to the lakes, but the measurements are from independent observation stations downstream. Only the mean of the ensembles are shown. The labels EnKF_W+LZ+UZ and EnKF_W+LZ refer to the update runs including and excluding the upper zone store for state updating, respectively.

When the upper zone is not included in the updates (EnKF_W+LZ), but only the lake water level and the lower zone were updated, the peaks were reduced. For the outflow at Rheinfelden, updating only the lower zone improved the update for the peaks during low flow, but at the cost of some loss of being able to follow the observations into local peaks during intermittend flow conditions, for example during April and June.

The direct insertion run reflects the quality of the modelling of the lakes. Looking at the outflow of the lakes (Fig. 6.12), the outflow modelled with direct insertion should match the observations closely. Combined, the calculated discharge from observed water levels does not add up to the observed discharge at Rheinfelden for the period of lowest flow from August till October. For that period the outflow calculated with direct insertion of the lake levels leads to higher streamflows than observed.

Table 6.2: Nash-sutcliffe efficiency scores for modelling water levels (left) and discharges (right). The NSE scores are calculated for daily aggregations. The columns list the model set-up variants: open loop (OL), direct insertion (DI), EnKF updating four states (LZUZ) and EnKF updating with only the lower zone (LZ). Row numbers list the number of the lake corresponding to Fig. 6.11, with the exception of number 9 which represents the outflow at Rheinfelden, corresponding to Fig. 6.12.

Water level					Downstream discharge			
#	OL	DI	LZUZ	LZ	#	OL	DI	LZUZ	LZ
1	-0.67	0.99	0.76	0.71	1	-0.48	0.98	0.57	0.56
2	-0.88	0.98	0.90	0.89	2	-0.64	0.96	0.90	0.89
3	0.38	0.98	0.89	0.92	3	0.24	-0.25	-0.06	-0.04
4	0.50	0.90	0.88	0.86	4	-0.12	0.70	0.64	0.60
5	-3.97	0.30	-1.13	-1.10	5	0.30	-0.91	0.40	0.47
6	0.83	0.95	0.85	0.86	6	0.55	0.89	0.40	0.59
7	-1.92	0.94	0.57	0.55	7	0.13	-1.08	-0.74	-0.53
8	0.53	0.95	0.85	0.93	9	0.24	0.82	0.57	0.84

Deviations between model and observations have a number of sources: bias in the observations, uncertainty of the measurements of the lake levels or their representativeness, or model errors in the lake modelling. As mentioned above, the Vierwaldstättersee shows a systematic difference between the outflow modelled with direct insertion and the observed discharge in particular. The higher discharge from the Vierwaldstättersee is the main contributer to the overestimation of low flows at Rheinfelden, indicating that solving this model issue has a good chance of improving the final modelling of discharge at Rheinfelden.

The performance of the different set-ups is summarized in Table 6.2 by a comparison of Nash-Sutcliffe efficiencies. The scores underline the results described above, for example the fact that both water level and discharge are only marginally adequately modelled without data assimilation for the Zürichsee (#6) and that the relation between water level and lake outflow is badly modelled for Lac de Neuchâtel, Vierwaldstättersee and the Murtensee (#3,#5,#7, right column DI).

There is little difference in NSE scores between the update with the LowerZone and the update also including the UpperZone. The only notable (larger than 0.1) differences are for the water level of the Walensee (#8), the discharge from the Zürichsee (#6) and the discharge at the outlet at Rheinfelden (#9). The latter show (bottom line) that EnKF updates targeting the lake level and LowerZone give the best results in terms of NSE, closely followed by direct insertion.

So far, the effect of the update procedures has been demonstrated by considering the updated initial states at the time of forecast, with similar performance of EnKF_LZ and direct insertion. For the EnKF it is expected that updates to upstream states result in a longer retention of the advantage gained by state updates for the forecast, as states are updated which have a longer memory such as the lower zone store that represents longer stored groundwater. This was tested by calculating the Relative Absolute Mean Distance between the open loop forecast and the posterior forecasts. With l the index for a partic-
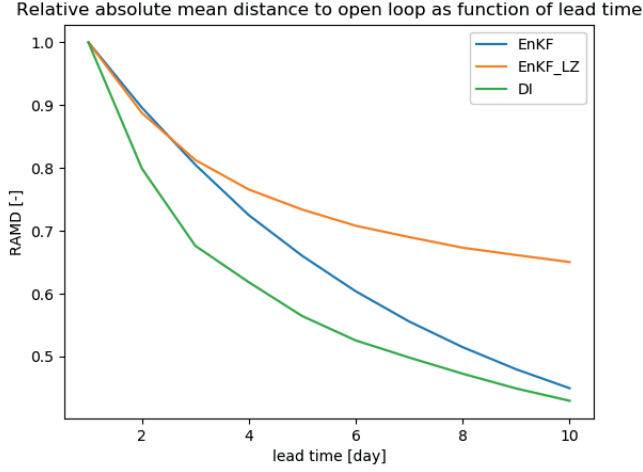
Figure 6.13: The relative absolute mean distance (RAMD) expresses the convergence between the updated forecasts and the open loop as function of lead time. A RAMD of 1 equals an absolute mean difference equal to the absolute mean difference at lead time one. A RAMD of 0.5 means that the absolute mean difference has since been halved.

ular lead time and n pairs of forecasts, the RAMD for lead time l equals:

$$RAMD_l = \frac{1}{n} \sum_{i=1}^{n} \frac{p_{n,l} - f_{n,l}}{f_{n,l}} \Big/ \frac{p_{n,1} - f_{n,1}}{f_{n,1}} \ , \qquad (6.16)$$

with f the forecast from an open loop initial state and p a forecast from an updated initial state. The RAMD equals 1 for the first lead time, and subsequently indicates to what extent both forecasts have converged.

For direct insertion, on average only 43% of the original gain remains after 10 days, measured as the relative distance between the forecasted value from the updated initial state with the open loop. As differences in initial condition converge over time this measure will converge for each version at a different rate from one to zero. Updates with the EnKF have a longer memory than direct insertion. Updates with and without the upper zone included converge at the same rate for the first two days, after which the update with the upper zone included converges faster (45% after 10 days) than with only updates to the lower zone (65% after 10 days).

6.5 Discussion

The results described above need to be understood within the following limitations. The experiment was conducted for a single hydrological model (wflow_hbv). In particular, this model has an unbounded upper zone store. Unbounded states are tricky to update

as there is no physical control on the values the states will take. The conclusion that the best results were obtained when only the lower zone store of the wflow_hbv model was updated is model specific. Other models might benefit from including fast runoff components in the update. The modelling of the lakes themselves is also a fundamental part underlying this study.

Smaller fluctuations and deviations and irregularities in the posteriors, such as the fluctuations on the reclining limb of the lower Bodensee, were partly caused by spurious correlations between lakes. The influence of spurious correlations was investigated by updating all lake levels with only the observation of the Vierwaldstättersee. It showed that other lakes that are not connected to this lake still were affected greatly. Doubling the ensemble size and adding the localization function developed by *Zhang and Oliver* (2011) reduced the spurious correlations. Even though the spurious correlations were only secondary factors contributing to the issues seen in the assimilation, as the influence of the direct observation of each lake far outweighed the influence of other observations, the used number of ensemble members remains small compared to other studies utilizing the EnKF.

It is known that management rules are imperfect and a source of uncertainty for the modelling of the behaviour of lakes. A trade-off in the real-world experiment was between focussing on a single lake or the combined effect of multiple lakes on the discharge from an integrated catchment model. Focussing on a single lake might have resulted in better performance as more settings and combinations could have been tried as one of the conclusions is that each lake needs to be investigated separately to obtain optimal data assimilation parameters for each lake. The advantage of the integrated catchment approach is that 1) data assimilation in models that span multiple tributaries is on the advancing edge of hydrological data assimilation itself, and 2) the issues of spurious correlations as well as the differences between lakes would not have come to the fore while focussing on a single lake.

This was an exploratory study in the use of lake water levels for data assimilation. More complex methods for analysis could have been used in several parts of this study, but it was judged that this would not change the lessons learnt. For example, forecasting metrics based on multiple years of forecasts could have further quantified the difference in forecast between open loop, EnKF and direct insertion, but this would not change the underlying mechanics which have already been made visible.

This study is a starting point for further studies, both theoretical and practical. Theoretically it would be of interest to further investigate the discrepancy between the simple linear observation uncertainty used for discharge, and the complexity of the real errors, including errors in the stage-discharge curve, used to calculate observed streamflow. Another interesting study would be to replace the in-situ measurements with estimates from satellite imagery, possibly introducing global wide potential. Practical pathways to improve further on the data assimilation for this area of the Rhine would be to add a diverse mix of observations of different variables. Snow measurements and gridded snow estimates are available, as are plenty of discharge measurements. The openDA software makes it easy to add additional variables.

6.6 Conclusion

The goal of this paper was to investigate the usefulness of integrating lake water level observations into a data assimilation scheme to improve low flow forecasts for the Rhine. To gather insight in the particularities of assimilating lake level data, first a synthetic model experiment was performed to test the data assimilation set-up under ideal circumstances.

The synthetic experiment showed that assimilating lake levels did not only improve the lake level and downstream discharge simulations, as expected, but also influenced upstream hydrological states. This was not a given due to the delaying function a lake has on the flow of water and thus upstream processes and the lake level might not be correlated due to a time shift (*Rakovec et al.*, 2015). In this case, two aspects contributed to the correlation between lake level and upstream states: 1) the daily time step is relatively coarse for the basin tested, neglecting for a large part the delay between upstream and downstream due to routing that is seen at small time steps (*Rakovec et al.*, 2015); 2) specifically for the relation between water level and the snow state, the temperature was perturbed, so that the snow pack during snow melt is related in the same way to the water level in the lake as precipitation would have been.

Special attention has been given to general aspects of assimilating lake water level data. The main lessons learnt are:

First, observation uncertainty needs to be determined for each lake separately, as any uncertainty can be relatively small or large based on the dynamic range of the lake. This is different from, for example, discharge, where a single blanket observation uncertainty suffices in most cases. This could be remedied by describing the observation uncertainty as a function of dynamic range. Still, it was also shown that the functional form of the widely used linear observation uncertainty for discharge, $\sigma_Q = 0.1Q$, is different from any possible additive error to the lake level, which is distinctly non-linear when transformed into error in discharge through the stage-discharge curve of the lake. This should give pause for thought about the correctness of $\sigma_Q = 0.1Q$ for discharge, as discharge is also in fact produced based on a measured water level and a stage-discharge curve.

Second, assimilating lake water levels might not be the most robust option when the goal of the model is to predict streamflow. In the case of errors in the stage-discharge relationship of the lake, assimilating downstream discharge will adjust the lake model state toward a level that corresponds with the measured discharge, which will give better results for streamflow even if the lake model state is incorrect. Assimilating lake water level can lead to biases in such cases. Logically, this is the other way around when the focus is on modeling the lake levels as accurately as possible, for example for lake management.

When applied to the actual model of the Swiss Rhine, results were mixed. The assimilation did reduce the dry bias in the model for the update. However, the update also resulted in overestimation of streamflow at the outlet for peaks during the dry season. Better results for the dry period were achieved if the updates were limited to the lake level and the lower zone. The effect of the update was also the most persistent for the EnKF, with only the lower zone and the lake water level as updated states having positive effects on longer lead times.

Direct insertion proved to be a computationally lean alternative to the EnKF for updat-

ing lake levels. Short term forecasts were still improved as the dry bias in the model was corrected by replacing the modelled lake levels with observed ones. Direct insertion was compatible with how the lakes are modelled as point states and hence no spatial estimate has to be made based on a single observation.

The three methods showed clear trade-offs in performance. First, direct insertion is computationally lean and not prone to unwanted side-effects caused by the update, such as the observed high peaks for the update with the upper zone. The drawback of the direct insertion is that due to modelling errors in the lake modelling, direct insertion still leads to temporal biases based on those errors at times when the EnKF alternatives did make successful corrections. The direct insertion did also converge most quickly to the open loop run for forecasts.

Second, the EnKF is computationally heavy and 32 ensemble members is still at the lower end to prevent spurious correlations and other artefacts that can inhibit a good EnKF result. Some experimentation was needed to decide which states were suitable for update, with the case excluding the upper zone showing the better performance for low flows. In particular, the updates with only the lower zone updated did not show overshoots during the low flow period and the update had a longer persistence in the forecast. .

Chapter 7

Assimilation of streamflow in a distributed hydrological model for Rhine tributaries: comparison of forecast skill between state updating and post-processing

In this study, a large scale data assimilation and reforecast experiment was conducted for the twelve main tributaries of the river Rhine. The effect on forecast skill of state updating with the Asynchronous Ensemble Kalman Filter (AEnKF) and ARMA error correction are compared for medium-term (15-day) forecasts over a period of 20 years (1996 till 2016). State updating improved the initial state for all subbasins and resulted in lasting skill score increase. ARMA also improved the forecast skill, but the forecast skill with ARMA does not always converge towards the uncorrected model skill, but instead can deteriorate for longer lead times. ARMA correction outperformed the AEnKF state updating for the first two days, after which state updating became more effective and outperformed ARMA. We conclude that state updating has more potential for medium-term hydrological forecasts than ARMA.

7.1   Introduction

There is an ongoing trend in hydrological forecasting towards both spatially distributed (gridded) models, ensemble forecasting and data assimilation techniques to improve forecasts' initial states. While in the last years those different aspects have been investigated separately, there are only few studies where the three techniques are combined: ensemble forecasts with state updating of a gridded hydrological model.

State updating has te ability to improve hydrological forecasts by adjusting model states towards values that better match the observed (diagnostic) values (see also previous chapter). A remaining challenge is to apply data assimilation methods to large scale integrated gridded hydrological models. Such models have an extremely large number of states (e.g. all model states for each grid cell in the model) and correlations between the innovation (observed error between model and observation) and model states are separated in both space and time (*Rakovec et al.*, 2012, 2015). There are still only few studies that apply data assimilation to fully distributed hydrological models. In general, studies addressing data assimilation in hydrology focus also on relatively short time spans, ranging from a single event to two or three years.

We here aim to fill this knowledge gap with a 20-year data assimilation and ensemble reforecast experiment with a high resolution gridded hydrological model (wflow_hbv, 1200x1200 m) of the full Rhine basin (160 000 km^2). To put the impact of state updating in an operational forecasting context, the data assimilation results are compared with AR post-processing as used by the Dutch Forecasting Centre (WMCN).

State updating is not the only way to improve a model forecast. Error correction and other post-processing techniques are available to improve the forecast skill and to give insights into forecast uncertainty (*Verkade*, 2015). Post processing techniques have a longer history in operational hydrological forecasting and are therefore more often found in operational practice (e.g. *van Andel et al.*, 2013). The operational system for hydrological forecasts of the Rhine river at the Dutch border of Lobith uses an Auto Regressive Moving Average (ARMA) error correction to correct errors in the predictions of streamflow from upstream subbasins, that serve as input for the Rhine hydrodynamic model. As part of current changes to the hydrological modelling environment, state updating is one of the options to replace, or augment, the current ARMA error correction. To make a decision about which method to pursue for the new modelling system, it is vital to better understand the relative advantages and disadvantages of both methods.

In this study, we compare data assimilation state updating, with the Asynchronous Ensemble Kalman Filter (AEnKF), with ARMA error correction for medium-term (15-day) forecasts using a fully distributed hydrological model and over a long (20-yr) time-period. We compare the forecast skill of two scenarios: one set of forecasts with state updating and one set of forecasts with ARMA correction. To extract generalities about the difference between the two methods in general, a reforecast data set is used to make forecasts for the last 20 years over the period 1996 till 2016 and for multiple subbasins.
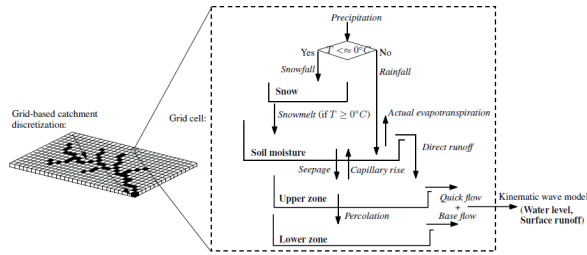
Figure 7.1: Catchment discretization and schematic structure of the HBV-96 model for each grid cell. Model states are in bold and model fluxes in italics Rakovec et al. (2015).

7.2 Methods

Data

Observational data has been preprocessed for use with a grid based hydrological model (*van Osnabrugge et al.*, 2017; *van Osnabrugge et al.*, 2019). The data was processed with hourly time resolution, on a 1.2x1.2 km grid spatial resolution, and for the period mid 1996 through 2015. All source data to derive the gridded estimates comes from sources that supply their data in near real-time making the datasets suitable for operational forecasting. For this study all data was aggregated to a daily time step. The hourly datasets are downloadable through the 4TU data centre (*van Osnabrugge*, 2017, 2018).

For the meteorological forecast data reforecast data from the European Center for Medium-Range Weather Forecasts (ECMWF) is used. The ECMWF issues hindcasts produced with the current model cycle for certain days for the last 20 years. The reforecast obtained for this study was produced with model cycle 43r1 (*Buizza et al.*, 2017). The first forecast is on 1996-03-10 and the last forecast on 2015-12-29 with reforecasts alternating every three or four days.

Hydrological model

wflow is a modular hydrological modelling framework that allows for easy implementation and prototyping of regular grid hydrological model concepts in python-pcraster (*Schellekens et al.*, 2018). The hydrological model concept used is the HBV96 (Hydrologiska Byråns Vattenbalansavdelning) model concept (*Lindström et al.*, 1997a) applied on a grid basis (Fig. 7.1). The generated runoff is routed through the river network with a kinematic wave approach (*Schellekens et al.*, 2018). In the following this model is referred to as wflow_hbv. The set-up of the hydrological model is the same as used in assessing the validity of the genRE precipitation data set (*van Osnabrugge et al.*, 2017), and in assessing the influence of potential evaporation forecasts *van Osnabrugge et al.* (2019). The model was parameterized through calibration with a Generalised Likelihood Uncertainty Estimation (GLUE) like procedure (*Beven and Binley*, 1992), using HYRAS precipitation as forcing data (*Winsemius et al.*, 2013a,b).

Asynchronous Ensemble Kalman Filter (AEnKF)

A distinction can be made between 'three-dimensional' (3-D) and 'four-dimensional' (4-D) data assimilation schemes. In 3-D data assimilation, all assimilated observations are assumed to be synchronous with the time of updating. 4-D data assimilation takes the temporal dimension into account, hence 4-D, by assimilating also observations that are asynchronous and taken at times different than the time of updating.

One solution to the problem of asynchronous updating could be to run the updating filter every time a new observation becomes available. However, this is unfeasible in any real operational setting as assimilation filters are computationally expensive. It is therefore preferable to be able to assimilate over a certain assimilation window, reducing the number of updates but taking into account all observations.

The AEnKF was therefore developed as an extention of the Ensemble Kalman Filter (EnKF) to deal with asynchronous observations so that the popular EnKF algoritmn now also could be used for 4-D updating (*Sakov et al.*, 2010). It is observed that the AEnKF should not be seen as a new method, but as a simple modification of the EnKF (*Rakovec et al.*, 2015).

Additionally, for hydrological simulations, assimilating over a longer assimilation window can provide improved representation of the time-lag between internal model states and the catchment response in the form of discharge. This idea was also investigated by *Li et al.* (2013) for the Ensemble Kalman Smoother (EnKS), which is an algorithm very similar to the AEnKF in effect.

As with other assimilation filters, most of the published results are in the fields of meteorology and oceanography. A number of studies discuss the use of the AEnKF for use in hydrology. *Rakovec et al.* (2015) applied the AEnKF for flood forecasting in the Upper Ourthe catchment ($1600 km^2$) in the Belgian Ardennes with a grid-based spatially distributed version of the HBV hydrological model with a 1x1km spatial and hourly temporal resolution. This is supposed to be the first use of AEnKF in flood forecasting (*Rakovec et al.*, 2015).

At the time of writing, about ten research papers have cited *Rakovec et al.* (2015). However, only two of those studies actually use the AEnKF. *Tao et al.* (2016), use the AEnKF in comparison with the EnKF and the fixed-lag EnKS to assess the potential of data assimilation for an operational system of three basins in the Southern Appalachian mountains, USA. Discharge at the outlet is assimilated in the fully distributed physically based hydrological model (DCHM). The DCHM model is run with a spatio-temporal resolution of 250x250m and 5-min time-step. The experiment is performed for a single high water event during an incentivised measurement campaign and only the discharge after updating is evaluated. The AEnKF and EnKS outperform the EnKF. Testing different assimilation frequencies and time windows leads to the conclusion that the optimal assimilation time window varies between basins and that this asks for context-aware configuration of the DA system. *Mazzoleni et al.* (2018), compared five DA techniques, amongst which the AEnKF to assimilate streamflow observations into a 3-parameter Muskingum routing model for three river stretches of the Trinity and Sabine rivers in Texas, USA. Regarding the AEnKF, they found that the performance of the AEnKF was less sensitive to suboptimal choices in the settings governing the assimilation, that ensemble methods provided smooth updates along the stream direction, and that the AEnKF was deemed the preferable method for distributed models. A review of all papers citing *Sakov et al.*

(2010) did not unearth other studies where the AEnKF is used for land surface hydrology at the time.

The AEnKF algorithm is one of the data assimilation algorithms available in the openDA software (*Ridler et al.*, 2014). For the mathematical formulation in a hydrological context we refer to *Rakovec et al.* (2015).

ARMA post processing

An Autoregressive Moving Average model (ARMA) is used to describe weakly stationairy stochastic time series in terms of two polynomial functions that represent the AutoRegressive (AR) and Moving Average (MA) part, respectively:

$$X_t = c + \epsilon_t + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \qquad (7.1)$$

with, ϕ the autoregressive model's parameters, θ the moving average model's parameters, c a constant and ϵ white noise error terms.

The lengths of the sums is determined by the order of the ARMA model, written as ARMA(p,q). So an ARMA(3,4) model takes into account the last three previous model outcomes for the AR part and last four model outcomes for the MA part, and has 7 parameters. ARMA models can be used for forecasting by running the ARMA equation sequentially for each future time step.

When ARMA is used for error correction, the ARMA model is set-up to forecast the future residual error, which is subsequently added to the forecast of the forecast model to produce a corrected forecast. X_t from Eq. (7.1) is then the error of the forecasting model (hydrological model) at time t: $X_t = Q_{m,t} - Q_{o,t}$.

The coefficients ϕ and θ need to be fitted and the order (p,q) needs to be determined so that best fit with historic data is obtained. To do this automatically, *Broersen and Weerts* (2005) developed the ARMASA algorithm to fit the coefficients as well as determining the optimal order. An equivalent to the ARMASA algorithm has been incorporated in the Delft-FEWS forecasting platform (*Werner et al.*, 2013), which is used in this research. In the operational system that is investigated, the maximum order for the AR part is set to 3. The MA part is not used in the current operational system.

7.3 Experimental set-up

The domain of the Rhine is split in 12 subbasins (tributaries), with corresponding discharge measurement stations (see Fig. 7.2). The outlets of the subbasins correspond with the locations where the calculated discharge is used to feed a hydrodynamic model that covers the modelling of the flow in the main course of the Rhine. The AEnKF is applied to update the initial state prior to the forecast for 11 of the subbasins. Splitting the domain in multiple separate subbasins has two advantages: 1) there is no danger of spurious crosscorrelations between basins and 2) the smaller domains enable the computations to be run on relatively 'normal' machines (12GB RAM, 8 cores). In this way, the manual splitting of the whole domain in subdomains is a form of manual parallelization. The AEnKF is not
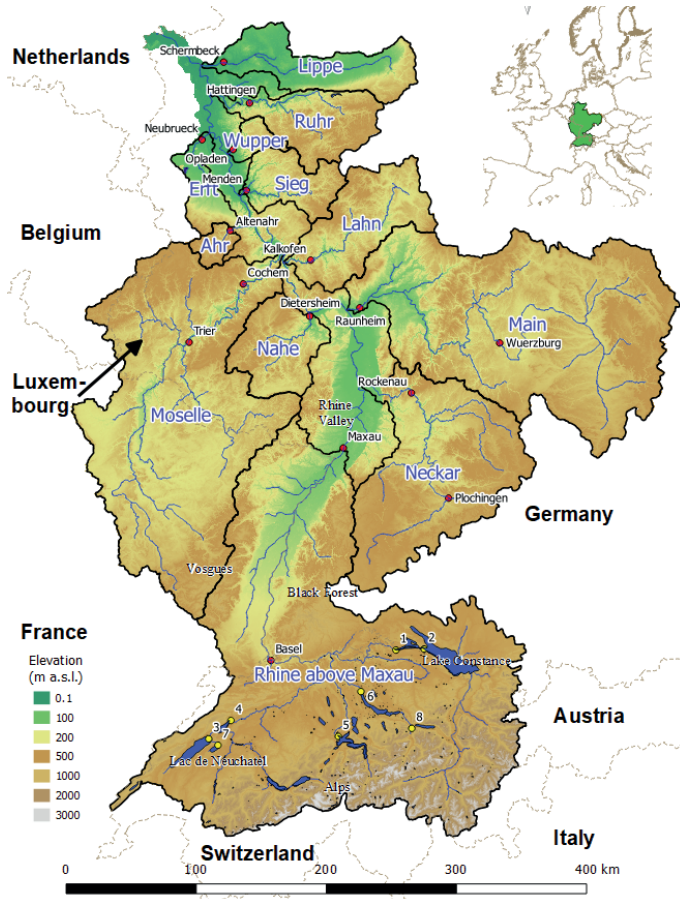
Figure 7.2: Location of the subbasins in the Rhine basin. The red dots show the location of the discharge gauges used in this study, while the yellow dots show the locations of the stations that measure the lake levels.

applied for the subbasin 'Rhine above Maxau.' Instead, the modelled lake levels were replaced by observed water levels (direct insertion, or DI) as was investigated in Chapter 6.

The AEnKF update is carried out with global settings for all basins. The assumed error on the discharge equals $\sigma_Q = 0.1Q(m^3s^{-1})$ (Chapter 6 *Weerts and El Serafy*, 2006; *Rakovec et al.*, 2012, 2015; *Mazzoleni et al.*, 2018). The ensembles are formed by perturbing the input precipitation and temperature fields. The precipitation fields are perturbed with $\sigma_P = 0.5P(mm)$, with no time correlation between hours and a spatial correlation distance of the error of 10 km, and the temperature fields with $\sigma_T = 2(°C)$, with no time correlation between hours and a spatial correlation distance of 5 km. The wflow_hbv mo-

del states that are updated are the LowerZone (associated with slow-flow processes), Up-perZone (associated with fast-flow processes) and the water level state of the kinematic wave routing. The soil moisture state is not updated as this deteriorated performances for discharge forecasting using wflow_hbv (*Rakovec et al.*, 2015).

The model is run with an hourly time step. The initial state was updated every 6 hours over the period 1996-06-16 till 2015-12-31, which matches the available forcing data. Forecasts were done every three to four days matching the availability of forecast data from the ECMWF reforecast. The ensemble forecasts depart once from the ensemble mean of the AEnKF update and once from the open loop initial state without updating. The ensemble mean of the AEnKF is used to compare the forecast based on the AEnKF initial states with the forecasts based on the deterministic open loop run. After the forecast run, the ARMASEL algorithm is used to apply an AR correction to the open loop ensemble forecast with AR parameters estimated from the deterministic open loop run.

Results are evaluated in update mode and forecast mode, focussing solely on effects on modelled discharge. The analysis is based on well known scores: the KGE and NSE for assessing the initial states, and the CRPS and CRPSS to assess the forecast runs. A one year spin-up period is excluded from the analysis. The scores are calculated over the period 1996-06-11 till 2015-12-28 for the initial states assessment and starting from 1998-01-01 for the assessment of the forecasts.

7.4 Results

Initial state updates

Applying the AEnKF improves the fit of the discharge to the observed discharge for all basins (Fig. 7.3). The largest gains are for the basins that are modelled worst (Erft, Wup-per), which is not unexpected as the potential for updates is also largest when the inno-vation ($o - m$) is largest. Hydrographs were plotted for each basin for the best and worst performing year based on the open loop model run and the KGE metric to get insight in the model performance. For the poor performing basins, there is a large negative bias in the Erft basin (gauge Neubrueck). For the Wupper (gauge Opladen), discharge is missing for mainly the discharge peaks (Fig. 7.4.

Effect of state updating and ARMA correction on forecast skill

Forecast skill was calculated for all subbasins and methods. The Mean Continuous Ranked Probability Skill and Skill Score (MCRPS and MCRPSS, respectively) are plotted as func-tion of lead time and for three different quantiles for all basins. The chosen quantiles represent the high flows (top 25%), low flows (bottom 25%) and overall behaviour (all data). The quantiles are based on the observations. Each basins differs in its response to the AEnKF and ARMA correction, but some general trends are clear. For this purpose, Fig-ure 7.5 is chosen to show a number of typical aspects of the open loop, AEnKF and ARMA forecast. Figure for all basins can be found in the appendix, Figures D.1-D.5.

First, the absolute errors, as expressed in the MCRPS, are small for the low flows and high for the high flows, with the average flows in-between (Fig. 7.5, left panel). This is impor-tant to keep in mind when assessing the relative skill score metric later, as small absolute
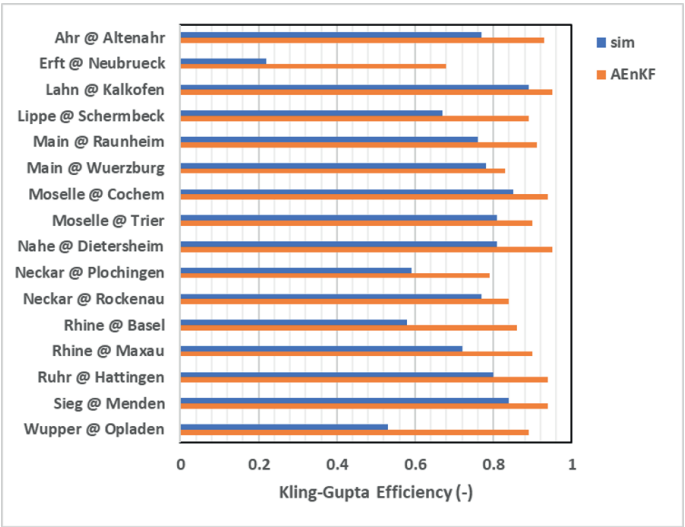
Figure 7.3: KGE for all subbasins with and without state updating with the AEnKF.
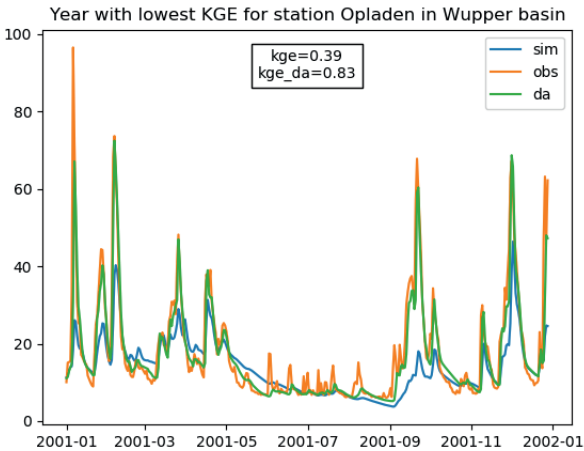

Figure 7.4: Hydrograph for Opladen measurement station in the Wupper subbasin, with and whithout state updating with the AEnKF. The year shown is the year with the lowest KGE score of the open loop simulation.

changes in forecasts can lead to large changes in MCRPSS. The forecast error is smaller for AEnKF and ARMA for short lead times, with the error for ARMA for the first lead time almost zero. Both the AEnKF and ARMA forecast then converge towards the open loop forecast, but where the error for the AEnKF based forecast remains smaller than the open loop error (although negligibly smaller for longer lead times), the error for the ARMA corrections becomes slightly larger than the open loop forecast.

Second, there is considerable skill for all quantiles for each model set-up (Fig. 7.5, middle panel). The positive skill score extends up to the total forecast length of 15 days (360 hours), which is longer than the skill found in the precipitation forecast (*van Osnabrugge et al.*, 2019), although the latter paper takes seasonal variability into account in the sample climatology, setting a stronger benchmark for the MCRPSS. The MCRPSS for low flows does not decrease with lead time (for the case shown, but the other basins show similar behaviour) . This is due to a number of factors. To begin, the sampling based on observations means that peaks caused by future precipitation are excluded from the analysis for the low flow quantile. After all, a considerable precipitation event shifts the observed flow quickly to the intermediate range. Therefore, it is likely that the low flow quantile is made up almost exclusively of observed recession periods (dry spells). The only injections of uncertainty are precipitation events that are predicted, but that do not show up in the observed streamflow. This makes the forecast period assessed much like a normal model run and hence the skill does stay the same for all lead times. Additionally, the open loop model shows negative biases for low flow periods almost exclusively. Now that there is a positive sampling bias in forecasted precipitation, due to the hits and misses being filtered out by the above mechanism, the positive bias compensates for the negative bias, thus even increasing forecast skill for the low flow quantile. This slight increase in skill is also seen in a number of basins.

Third, ARMA improves the forecast skill compared to the open loop forecast skill the most for short lead times (first two days), however for longer lead times ARMA is not consequently improving the forecast. Forecasts based on the AEnKF show less improvement compared to ARMA for the short lead times, but then perform consistently better for longer lead times (Fig. 7.5, right panel).

The advantages and disadvantages in terms of forecast skill for ARMA compared to AEnKF state updating are shown in Figure 7.6. Most basins follow the pattern as described above: ARMA has a better skill for the first 24-50 hours and then the AEnKF has higher skill. The outliers are shown to be basins were the model has poor performance: a red coloured line means that the model without state updating or correction has worse performance than sample climatology.

For the low flows (right panel), the outliers are the Main@Raunheim and Lippe@Schermbeck. For both the Main@Raunheim and Lippe@Schermbeck the peaks are well modelled, but biases exist for low flows that leads to bad performance for uncorrected forecasts. The AEnKF weighs the respective uncertainties of the observations and the model and ends up in between the model and the observation. The ARMA model corrects the forecast with the complete bias over the full forecasting window and therefore shows better skill.

For the high flows (middle panel), the forecast for Erft@Neubrueck is the outlier. Already in discussing the updates, it was mentioned that the model does model the dynamics
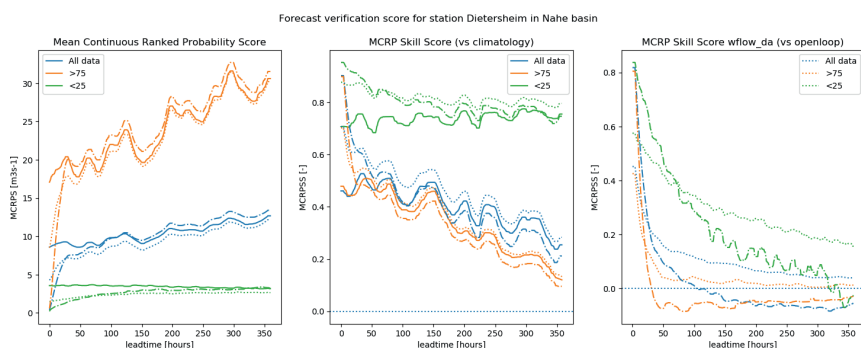
Figure 7.5: *Forecast verification scores for station Dietersheim MCRPS (left panel) and MCRPSS (middle,right). The scores are shown for three scenarios: open loop (solid line), AEnKF (dotted line) and open loop with AR correction (stripe dot line), and for three quantiles (see legend). The difference between the middle and right panel is that for the first the MCRPSS is calculated with sample climatology as benchmark forecast and for the right panel the MCRPSS is calculated with the forecast from the open loop initial state as benchmark.*
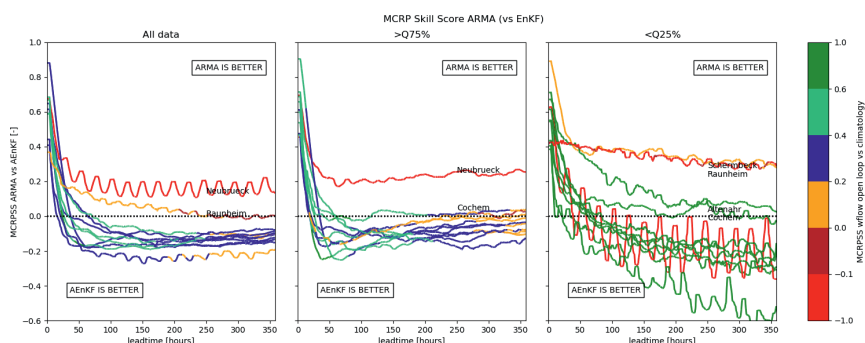
Figure 7.6: *MCRPSS of ARMA posprocessing versus AEnKF state updating for all basins, for three quantiles and as function of lead time. The colors of the lines indicate the overall quality of the forecast itself as expressed in the MCRPSS of that forecast with sample climatology as benchmark.*

of the basin very badly, in addition to a large negative bias. The AEnKF algorithm only corrects partly for the bias, and the model deteriorates quickly after the update as it is not able to properly model the catchment. As for the low flows, the ARMA model corrects for this lack of dynamics for the whole forecast period.

7.5 Discussion

This discussion consists of three parts. First, the computational approach in running the model experiment is discussed. Second, the results obtained here are compared with those found in *Bourgin et al.* (2014), who did a similar exercise. Third, advantages and dis-

advantages of using a 20-year reforecast are discussed and future research is presented.

Data assimilation in large scale distributed hydrological models

The performance of spatially-distributed hydrologic simulations has historically been hindered by high computational demands (*Wood et al.*, 2011). Although computational power is increasing, the limitations are still present. The focus shifts from simply running the models at the desired scale, towards performing meaningful experiments with it, such as data assimilation, ensemble uncertainty and long-period climate modelling, all of which require many model runs. For this study, the domain of the Rhine basin was partitioned into several subbasins, for which subsequent steps were then applied individually. This manual parallelization limits the relevance of applying the AEnKF to a large river basin, as this approach has effectively turned the problem into 12 smaller independent assimilation experiments. Still, attempts at parallelization for distributed hydrological models are all about how to partition a region into independent sub-basins in a way that optimally distributes the computational load (*Vivoni et al.*, 2011), a process which is further complicated when coupled with distributed data assimilation. Hence manual, or parallelization based on a preprocessing step (*Yalew et al.*, 2013), is still a common approach to solve computational problems related to large domains.

Acknowledging that the experiment here is similar to 12 independent data assimilation experiments, it is good to question the optimality of the used settings of the AEnKF. Different basins can have different optimal (uncertainty) settings (*Thiboult and Anctil*, 2015; *Tao et al.*, 2016) and this could influence the conclusion that ARMA is consequently outperformed by the AEnKF for longer lead times, but that ARMA outperforms the AEnKF for shorter lead times. An example of how this result would change under different AEnKF settings, is the case where the observation uncertainty is set to be negligibly small. In that case, the updated initial state would practically match the observed discharge value at the start of the forecast, as did ARMA in the current set-up, enabling the AEnKF to compete with ARMA for short lead times. As for longer lead times, under all AEnKF settings the influence of the in initial state on the forecast quality will diminish with lead time and the forecast of based on the AEnKF initial state will always converge with the forecast from the not-updated state. As the ARMA correction can actually make forecasts worse for longer lead times, it is safe to conclude that the behaviour of the AEnKF is more consequent and reliable for longer lead times than ARMA, as any auto-correlations of the error for longer lead times can become entirely fictional (*Broersen and Weerts*, 2005).

A similar question can be posed about the optimality of the implementation of the ARMA correction. A problem for the ARMA correction is that the hourly observations are noisy and small temporary fluctuations can hide the more relevant signal. Furthermore, the adaptive way in which the ARMASEL tries to optimize the order of the AR function makes it difficult to assess exactly why the ARMA correction operates the way it does. However, it is no surprise that the auto-correlation of the error between observation and model is limited in time and that as a consequence the effectiveness of ARMA correction is limited for longer lead time (*Broersen and Weerts*, 2005).

More discharge stations are available for the 12 subbasins. Additional stations can be easily added to the state updating scheme due to the flexibility of the combination of a dis-

tributed model and openDA. For the large basins (Moselle, Main) this could stretch the computation resources as the memory requirements of the AEnKF (as with the EnKF) is dependent on the number of observations assimilated. For this study, assimilating mainly the station at the outlet assured a fair comparison with the ARMA correction as both methods received the same information.

Post-processing and data assimilation

The interaction between state updating and post-processing has been previously investigated by *Bourgin et al.* (2014). *Bourgin et al.* (2014) compared four set-ups employing either state updating, post-processing, neither, or combined, for a 17-month period and 202 unregulated catchments in France with the lumped GRP model. Short-range (48-hour) forecasts were investigated. As post-processing technique, they employed a hydrological uncertainty processor that adds uncertainty to the hydrological forecast based on empirical uncertainty bounds, also known as ensemble dressing. It is not clear exactly what data assimilation procedure was employed. The explanation given, is that it consists of a two-step procedure that first exploits the last observed discharge to update the routing store and then applies a multiplicative correction. The given reference, *Berthet et al.* (2009), only adds that the first step is not of the Kalman family and that the second step is, in fact, an ARIMA correction step.

Bourgin et al. (2014) found that state updating and post-processing strategies had complementary effects. State updating influenced the forecast accuracy, while post-processing had a positive and longer lasting effect on forecast reliability. A combination of techniques is recommended. It is important to note that for their experiment state updating consists of two steps, the second of which is actually an ARIMA correction. ARIMA is a variant of the ARMA error correction, a step that we call post-processing in this study.

This is not to discuss semantics, but to point out that our results of the first 48 hours should only be compared with their data assimilation results. When we applied ARMA corrections to the DA results (not presented), we did see that this led to results in skill that were broadly an average of the skill for when the two methods were applied individually. Our results for data assimilation are comparable with those found in *Bourgin et al.* (2014), in that state updating has a strong positive impact on the ensemble mean. Adding ensemble dressing to the state updating approach presented here for medium-term forecasts would be an interesting further study.

Further research

The approach for data assimilation that was used in this study is not new, in fact is an upscaling of the method of *Rakovec et al.* (2015) from a single basin towards multiple basins, and from individual events to a consecutive period of 20 years. This could be improved upon further, by adding more measurements to the data assimilation scheme. For a true 'large scale' data assimilation experiment it would be interesting to assimilate not a dozen observations, but the 700 or so discharge observations that are routinely done in the Rhine basin.

For this research, a 20-year hindcast was performed with a forecast every three to four days. Pursuing such reforecasts is a significant exercise, which raises the question if long

hindcasts truly add insight to the performance of our methods compared to the more usual hindcasts of a couple of years or even individual events.

While the value of hindcasts with the ECMWF reforecast data set is understandable for experiments as presented here, namely to work towards comparisons between methods that are statistically significant, it is less clear if the absolute values of the metrics calculated have any real value for operational forecasters working with the system. The main issue, assuming that the skills themselves are perfectly interpretable, is that the actual forecasts are made on the basis of limited area NWPs, which have different uncertainty characteristics than the ECMWF reforecasts. The question then is if a short hindcast with the latest system is not more insightful than a long hindcast based on different data.

Further research could, by using the data created in here, investigate the variability of the several verification metrics between years, and conclusions could be compared between event based analysis, seasonal, yearly and for the whole reforecast period. This would shed light on the issue above.

7.6   Conclusions

A large scale data assimilation and reforecast experiment was conducted for twelve main tributaries of the river Rhine. The aim of this research was to combine ensemble forecast with data assimilation and a spatially distributed model and to evaluate the performance over a long time period (20-year) and on a scale that is relevant operationally (the Rhine river). To put the impact of state updating in an operational forecasting context, the data assimilation results were compared with AR post-processing as used by the Dutch Forecasting Centre (WMCN).

Two sets of forecasts were performed for multiple subbasins in the Rhine basin: one with state updating with the AEnKF and one with ARMA error correction. For both scenarios the same data was used, namely the discharge at the outlet of the subbasin.

State updating improved the initial state for all subbasins and results in lasting skill score increase, in particular for low flows. ARMA also improved the forecast skill, but it was found that the forecast skill with ARMA does not always converge towards the uncorrected model skill, but can deteriorate for longer lead times. ARMA correction outperformed the AEnKF state updating for the first two days, after which state updating became more effective and outperformed ARMA. We conclude that state updating has more potential for medium-term hydrological forecasts.

Further research aims to mine the 20 year reforecast that was produced for more insights in the value of such long-term reanalysis in the light of the effort that is involved and the limited availability of data sets that are long enough to perform such an exercise.

Chapter 8

Synthesis

8.1 Introduction to synthesis

In this Chapter, I work towards answering the research questions formulated the introduction. First, I discuss how a definition of operational aspects as set out in the introduction defines a research field in hydrological forecasting and how different operational aspects were investigated in this thesis. Second, I give special attention to the hydrological modelling that was done in Chapter 5, and how this modelling relates to operational aspects. Third, I focus on the operational challenges that are introduced by moving from lumped to gridded models. Fourth, I accumulate the results of this thesis to give advice for advancing the operational forecasting system operated by Rijkswaterstaat for the Rhine. This is followed by an argument that such a practical advice is a meaningful part of a scientific thesis as part of actionable research. Last, I present my view on what is the current state-of-the art in hydrological forecasting systems, and my view on future hydrological forecasting systems.

8.2 Operational aspects of hydrological forecasts

Forecasts are an important aid in making decisions under uncertainty. Many different components are involved in making a forecast, and all of those components, from precipitation interpolation to model simulation and data assimilation, have their own field of research. For the subfield of operational hydrological forecasting then to deserve to exist, it needs to focus on challenges that are unique to the operational domain (*Pagano et al.*, 2014).

Luckily, at least for someone who would want to write a PhD thesis on operational forecasting, this field has shown to have its own challenges (see also *Pagano et al.*, 2014), which I categorized here by means of a number of specific constraints. When these constraints are applied to methods and insights from the various subfields that are used in making an operational forecast, this leads to new challenges. In this thesis, I have called this *operational aspects*.

In Chapter 3, the operational aspect studied was precipitation interpolation under the constraint of limited data availability. Specifically, limited network density of hourly real-time reporting gauges and the constraint of reliability in the sense that any interpolation

139

method should also work in case of missing data. Collected rain gauge data were spatially interpolated using the genRE method, which uses climatological grids to inform the interpolation about the spatial distribution of precipitation. It was shown that improved hourly interpolation results could be obtained with the operational interpolation method by using climatological grids which were derived from non-operationally available daily data with a higher quality and quantity, thereby successfully using non real-time data to enhance the information available in real-time.

The interpolation exercise also led to a much needed hourly gridded interpolation set, which was subsequently used throughout this thesis. Regarding the existence of this dataset, and the availability of this dataset to other researchers, one needs to be aware that the genRE precipitation dataset was made with operational methods, studying operational aspects. It is therefore definitely not the end-all of precipitation datasets for the Rhine basin. In particular, it has been shown in later Chapters (Ch. 5 for example), that there are known negative biases in the dataset for large parts of the Alpine region.

The subject of Chapter 4 was the treatment of potential evaporation (PET) in hydrological forecasting. The operational aspect that was studied was the choice between using average PET climatology based on a long time series of offline data, or the use of near real-time available data to calculate PET in near real-time, including calculating PET from forecast data for use in hydrological forecasts. A 20-year reforecast was done and the resulting skill scores with online and offline PET data were compared. It was shown that for our case there was a negligible difference in discharge forecast skill between using the offline or online PET data for forecasts up to 10 days.

Chapter 5 reported the set-up of the wflow_sbm model concept without calibration. Parameters for the model were derived based on pedotransfer functions found in literature and open access spatial data, such as soil properties. The model parameters were subsequently scaled from the highest data resolution to several coarser model resolutions. It was shown that this resulted in adequate modelling results of discharge throughout the Rhine basin, as well as flux conservation between modelling on different spatial scales.

Although the scalability and uncalibrated properties of the model are interesting for use in an operational context, Chapter 5 did not directly deal with operational aspects, as the hydrological modelling exercise was investigated on its own and not as part of the constraints that define operational aspects. However, it delivers interesting context as the development of gridded high-resolution models drive the operational aspects in the other chapters. The relation of the main properties of the wflow_sbm model, namely the fact that it is spatially scalable and uncalibrated, with operational aspects of hydrological modelling is further discussed in Sections 8.3 and 8.4.

Last, Chapters 6 and 7 showed the results of two experiments with state updating. Chapter 6 employed the Ensemble Kalman Filter (EnKF) to investigate if measured lake water levels could be used to improve downstream discharge forecasts. The results showed that indeed state updating with lake level measurements can aid in making discharge forecasts, but that the robustness of this state updating is highly dependent on the quality of the lake modelling.

Chapter 7 applied the Asynchronous Ensemble Kalman Filter (AEnKF) to assimilate discharge into the gridded hydrological model at the subbasin level. The results of the assimilation were compared with ARMA postprocessing. It was shown that the ARMA cor-

rection was very strong for the first two days, but that the AEnKF provided more consistent improvements for longer lead times.

Some operational aspects of state updating are quite intuitive as state updating is used to deal with correcting errors while taking into account multiple sources of uncertainty, concerning observations and model alike. A less intuitive operational aspect of state updating comes to the fore when data assimilation research is intersected with the constraint that a decision has to be made. As explained in Chapter 6, data assimilation methods need to be instructed regarding the relative uncertainty of the model and observations in question. There is no unambiguous way to set the model uncertainty and observational uncertainty in a real practical situation, but the parameters governing the respective model and observational uncertainty can be used to adjust the reliability (sharpness) and the uncertainty (ensemble spread) of the forecast according to the wishes of the forecaster.

To conclude, there are many more ways in which established research fields are confronted with operational constraints and create operational challenges that deserve the attention of the research community, than addressed in this thesis. I argue that addressing and including those operational aspects can aid in closing the gap between researcher and practitioner.

8.3 Hydrological modelling at the forecasting interface

An operational aspect that can very well be discussed based on the results of the wflow _sbm model that is developed in Chapter 5, is how calibrated hydrological models decrease the modularity of hydrological forecasting systems by tying forcing and hydrological model strictly together.

During calibration of a hydrological model, the parameters of the model are adjusted to match a set of observations, mostly discharge. In automatic calibration an objective function is maximized and the resulting model with calibrated parameters is mathematically optimal to describe the observations that have been used to calibrate it. As a result of this procedure, the parameters of the model are dependent on the employed forcing data. Calibrated models are known to be very sensitive to changes in forcing data through calibration of parameters (*Andréassian et al.*, 2004, 2001; *Melsen et al.*, 2016), which makes this issue non-trivial.

As a calibrated model is only optimal for a given forcing data set, it becomes increasingly difficult to make changes to the forcing part of the hydrological forecasting chain. Changes in the spatial distribution of precipitation alter the optimal parameter sets of hydrological models (*Melsen et al.*, 2016; *Euser et al.*, 2015). Likewise, a change in model forcing might easily decrease the metrics that are used to quantify the quality of the model set-up. Calibration techniques that do not aim for a single optimal parameter set, but instead aim to constrain the model parameter space, give more freedom to the model to reflect the uncertainty in the model, which adds flexibility, but the essential problem of calibration, namely that the model is conditioned on a fixed set of observations and forcing data, remains (*Andréassian et al.*, 2012). In effect, this makes recalibration an essential part of any change to the forecasting system. Thus it is more difficult to study parts of the forecasting system in isolation, which reduces the modularity of the forecasting system and makes research to improve the whole forecasting system more expensive.

The wflow_sbm model was set up without calibration on streamflow. As a disadvantage, this means that the wflow_sbm model does not score as well as calibrated models in simulating discharge at specific points in the basin. In particular, it is sensitive to the negative precipitation bias in the Alps (Ch. 3). On the other hand, this highlights exactly that better forecasts can be obtained by improving the precipitation estimates in these areas, which might be obscured in calibrated models.

8.4   Operational challenges for gridded models

The move towards gridded hydrological models for operational forecasting purposes interferes with the continuous operation and processing time operational constraints due to the added computational costs and the quadratic increase in the volume of produced data.

There are three consecutive developments that claim computational resources: ensemble forecasts, data assimilation methods and high resolution gridded models, with the latter having by far the most impact as ensembles and data assimilation methods are multipliers on the base cost of running the model. The impact of computational resources propagates throughout the forecasting process, including increased computation time, but also increased data costs for preparation of forcing data, model data, and ensemble and data assimilation applications. Computational power remains one of the limiting factors in pursuing full uncertainty analysis of integrated forecasting systems (*Pappenberger et al.*, 2005).

Parts of this study were hampered by the added load of working with gridded models. A six-hour time step for the wflow_sbm model was chosen over an hourly time-step partly based on run-time considerations. The daily time-step for the investigation of the assimilation of lake levels was chosen for the same reasons. Additionally, the number of ensemble members was kept limited.

Given a fixed amount of computational resources and the processing time constraint in operational forecasting, a valid question is then how to distribute those computational resources between uncertainty analysis of the integrated forecasting system, finer spatial distribution of the hydrological model including more detailed processes, and updating the model through data assimilation techniques. This question on effective attribution of resources cannot be solved without integrating the view of the user.

8.5   Advice related to operational forecasting for the Rhine

The previous sections have put the results of this thesis in the broader context of operational forecasting science, focussing on operational aspects. In this section, the focus is on the actionable part; it is specific advice on how to advance the short-to-medium term forecasts in the Rhine basin and sets out further research.

Available data and data structuring

To start with the most practical recommendation, it is advised to implement the methods that have been used in this thesis to derive the spatial forcing estimates (Ch. 3 and 4). This

includes importing the downward short-wave surface radiation (ssrd) product from the Land Surface Analysis Satellite Application Facility (LSA SAF) and calculating near real-time PET. Although the comparison between forecasts based on real-time estimated PET and climatological PET did show only a negligible improvement for 10-day forecasts, it is still advised to make the change to real-time estimated PET for future model implementations, as preparation for models than can take advantage of the improved spatial and temporal information (Ch. 4).

For the purpose of this thesis, a large number of discharge gauge measurements, far exceeding the number of currently operationally available gauges, was requested and obtained from throughout the Rhine basin. The usefulness of this data was not fully explored. The additional data was mainly used to show a spatial assessment of the hydrological model performance, which might become part of a standardized benchmarking strategy. It is recommended to still investigate what value those additional measurements, mostly on smaller streams, can bring when used in data assimilation techniques. A practical experiment would be to broadly repeat the experiment performed in Ch. 7, but with many more discharge measurements included for each subbasin.

It is recommended to keep an up-to-date database containing all the data that is needed for a reforecast analysis. The Delft-FEWS software has the necessary functionality to keep and maintain such an archive. The acquisition of data for the Rhine basin was, due to the many parties involved, time consuming and therefore an expensive part of each analysis. This is especially the case when this exercise needs to be repeated for each research project separately, severely limiting the scope of small research projects in particular. Additionally, there is a delay between current events and the moment at which the data necessary to properly study them becomes available. A good recent example of this is the drought of 2018. There was a clear incentive to perform the assimilation study of Ch. 6 for this recent year to also contribute to the discussion on this recent drought. However, the datasets built for this thesis work do not extend past December 2016. The time involved in requesting the necessary data was the reason why the dataset could not be extended, and the 2018 drought could not be studied. A good database of necessary base variables enables and promotes research in the Rhine basin. That good databases of basic variables promote research can be seen in the popularity and success of for example data collections such as MOPEX (*Schaake et al.*, 2006), CANOPEX (*Arsenault et al.*, 2016) and CAMELS (*Addor et al.*, 2017). Having a readily available archive of base data is also a corner stone of the automated benchmarking system proposed.

Data assimilation

It is advised to implement the data assimilation set-up as used in Chapter 7. This set-up improves the forecasts at the inflow points of the hydrodynamic SOBEK model. It was shown to be more reliable and consistent for short-to-medium range forecasts than the ARMA correction that is currently employed. As mentioned in the relevant Chapters (6,7), the state updating system is not yet optimized. Further steps would include adding more discharge measurements to the updating procedure and the assimilation of spatial data, such as snow cover and spatially sensed soil moisture data.

Hydrological modelling and the wflow framework

As argued in the introduction (Section 1.3), the currently used HBV96 model is out-of-date. The work presented in this thesis (Ch. 5) is not the first exploration of alternative models to use in the forecasting system RWsOS Rivers either. Previously, a TOPOFLEX model was explored for the Meuse basin based on the thesis work of *Euser et al.* (2015); *de Boer-Euser* (2016). This model was implemented in the wflow modelling framework. The recent model developed by *Buitink et al.* (2019) is also currently being implemented in wflow. I argue that it is of paramount importance to choose a hydrological modelling framework first, and a preferred model concept second.

The wflow framework has now been tested with different model concepts and has been integrated with the Delft-FEWS and openDA software. Therefore, it is advised to adopt the wflow framework as hydrological modelling framework. The old HBV96 model can be replaced directly by the wflow_hbv model. Despite the promising features of the wflow_sbm concept as parameterised in this thesis, the work in this thesis does not support a conclusive advice on which model concept to use eventually, especially when discussing the use of such a model for both short-to-medium range forecasts and policy analysis, which have different requirements. Instead, I hope that, by focussing on operational aspects, ideal conditions can be created to perform future model experiments to determine which model concept to use, starting with fully benchmarking the wflow_sbm concept against wflow_hbv. The following paragraph describes how to set-up this benchmarking system, which incorporates all components that have been studied in this thesis.

Automated benchmarking system

Forecasting systems are not static, but should be able to accommodate changes in data availability, changes in external models (NWPs in particular) and new insights from the scientific community. However, the effects of those changes on the forecast quality, especially from the latter source, can be difficult to quantify. Testing the quality of the forecast should be standard routine; testing and improving the forecasting system should be performed themselves as operational routines. The key to this is twofold:

First, it is important to specify how the quality is assessed: how do I know if my forecasts are better? The paper by that name by *Pappenberger et al.* (2015) can be taken as a guide to choose appropriate benchmarks. The benchmarks should reflect the priorities of the forecaster and as such are an excellent starting point to start a discussion on what the forecaster's expectations are of the forecasting system. Part of this conversation is the trade-off between reliability (sharpness) and uncertainty.

Second, after the benchmarking procedure has been defined, it should be set-up such that performing a benchmark test is a small effort, not taking into account calculation times, which still can be considerable. This can be done with Delft-FEWS. Delft-FEWS was used as data manager for all research shown in this thesis, and has also shown to be an excellent research tool for reanalyses elsewhere (*Werner et al.*, 2013, and references therein).

The basis of a Delft-FEWS based reforecasting system would be that historical operational data is stored in a FEWS operational archive. Workflows then can be set-up to run the calculations necessary to produce the historical reanalysis and reforecast data of the new

to-be-tested set-up. Due to the modularity of Delft-FEWS, small singular changes to the input data are easy to implement. With the combination of the wflow framework and openDA software for data assimilation, changes in hydrological model or data assimilation are also easily implemented. The calculation of benchmark metrics and other verification statistics does not have to be done in the Delft-FEWS system, but can be coded separately and then run as part of the workflow after generating the data.

8.6 The role of case-based actionable research

The previous section can be read as a specific set of advice to a specific client. It is not unlogical to question such a specific advice in a scientific thesis. However, in this paragraph I will argue not only why it is justified to add this advice to this thesis, but also why I would challenge other PhD students to translate their thesis work into specific advice as an integral part of reflecting on the work done.

As discussed in *Palmer* (2012), for science that is effective in influencing policy decisions, not only broad multidisciplinary teams are needed that go beyond the natural sciences, but those teams also need to turn the tables and ask policy makers and policy influencers what they need from the scientific community. Engaging with policy and active interaction with users is therefore a skill that is important to learn. Looking at these requirements, which are laid out for current and future scientists, it is good to question if a thesis should only focus on the scientific community, e.g. a discussion on papers. Instead, targeting a broader audience, i.e. specifically addressing the needs of a user, can increase the impact of thesis work as well as prepare the PhD candidate to work in the multidisciplinary user question driven scientific environment that is necessary to tackle outstanding environmental problems.

Translating conclusions into specific advice is also a good reflection method. The possibility for action puts an increased weight on each word written. It is relatively safe to discuss, lets say, the advantages and disadvantages of gridded models in a scientific context, but it is a different matter to ask an organization to change their operations based on your work. The scientific discourse is by nature well equipped to identify the gaps in knowledge and the unknowns. There is no time limit either. Translating scientific discussions into specific time-bound advice asks for a broader vision that also reflects on non-scientific aspects that influence the decision making.

8.7 The future of operational hydrological forecasting

To conclude this thesis, I will give my view on the future of hydrological forecasting, by discussing some properties that any operational hydrological forecasting system should possess:

Modular Hydrological forecasting systems used to be build around specific hydrological models, but have in many instances already been replaced by software that recognizes the many components that make up a forecasting system (*Werner et al.*, 2013). This modular approach is critical to develop forecasting systems, as it allows forecasting centres to incorporate new techniques and new data more rapidly into

their forecasting system. Part of this modular philosophy is a clear separation between developing forecasting systems and developing hydrological models.

In my opinion, a quest for better hydrological model concepts can be a bit of a red herring, distracting from other components that can improve forecasts and the impact that those forecasts can have. The type of hydrological model concept does not make or break what is a state-of-the-art forecasting system. Instead, from an operational perspective, developing hydrological models should focus on expanding hydrological modelling frameworks to integrate the hydrological model with other forecasting components, such as data assimilation software.

Integrated and coupled models Paradoxically, forecasting systems also need to become more integrated, especially regarding the propagation of uncertainty through all model components (*Pappenberger et al.*, 2005). This is possible in a modular framework, but requires standard formats to exchange uncertainty information between components. Two-way coupled hydrological and meteorological models are a possible way forward to increase forecast skill (*Seuffert et al.*, 2002).

Integrated social elements and user interaction The integrated perspective should not only include technical elements. Recently, growing attention has also been given to the interaction between forecast and forecaster (e.g. *Demeritt et al.*, 2007; *Ramos et al.*, 2010; *Arnal et al.*, 2016). For state-of-the-art forecasting systems, the human aspects of forecasting need to be treated with the same attention as other, more technical, parts (*Pagano et al.*, 2016). Future forecasting systems move from portraying technical information towards the forecasting of impacts that better support decision making by including social-economic factors (*Terti et al.*, 2015).

Improved integration of data sources There is still much to be gained by combining different sources of information. This does not only hold on a component by component basis, such as improving the spatial estimation of precipitation through a combination of gauge data, radar data, satellite data and other novel sources such as telecom links (*Rios Gaona et al.*, 2015) and in combination with crowdsourced data (*de Vos et al.*, 2018), but also regarding the combination of data from throughout the forecasting chain. In an integrated uncertainty framework, uncertainty is not only a feed-forward property that propagates from forcing and forecast to the model and onwards, but model outcomes and 'downstream' information have equal strength to inform about uncertainty in the 'upstream' information (*Vrugt et al.*, 2008). In this way, errors in precipitation can be corrected based on the response of streamflow, for example.

Multi-purpose The uses of hydrological forecasting systems are increasingly multi-purpose. As the number and the diversity of the users of forecast information grows, so do the requirements of hydrological forecasting systems, as skilful forecasts are required for different variables, scales and under varying circumstances. It is an open question whether those different needs can be better catered for with multiple specialized models, or with coupled models of everything, aiming to model the whole environment (*Blair et al.*, 2019).

Automated learning systems State-of-the-art forecasting systems include the process of model improvement and system updates into the operational procedure itself.

Clear benchmarks that inform about what a forecasting system can, and cannot, forecast are part of communicating uncertainty. Changes to the system are directly tested on benchmark performance, based on which a well informed decision can be made. A possible drawback of the use of standardized benchmarks is that benchmark results could become indecisive when results show different signs of change for different locations or variables. The new approach should then be applied for some locations, but not for others. A possible solution could be to apply multi-model ensembles with weights determined by benchmarks. New ideas that work locally can be added and the new system has only to adjust the weights to generate new optimal multi-model ensembles for each location.

Consolidated There might be only a few organisations that will be able to combine the expertise and resources necessary to provide forecasting systems that combine the points above. Therefore, for the changes above to materialize, a step has to be made to form consortia that will deliver forecasts on a transnational scale, but still can beat local forecasts. In this respect, it will be very interesting to see how the European Flood Awareness System (EFAS) will develop, and if EFAS forecasts, or future counterparts, can first beat and then replace locally produced forecasts (*Bartholmes et al.*, 2009). This does not only ask for steps in technology, but also in international cooperation between forecasting centres.

Appendices A-C

A   Supplemental figures to Chapter 3
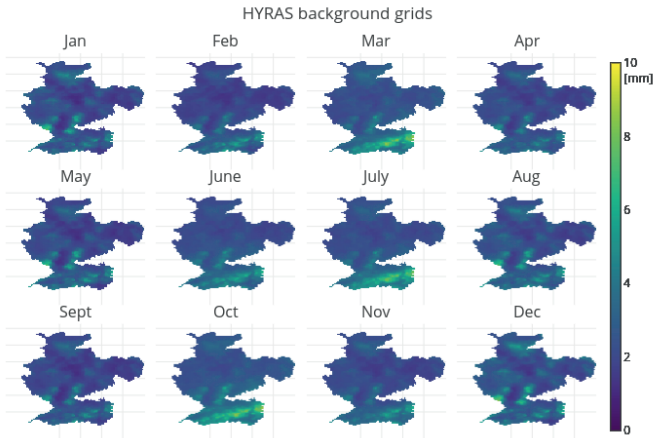
Figure A.1: *Background grids for the genRE method derived from the HYRAS climatology.*
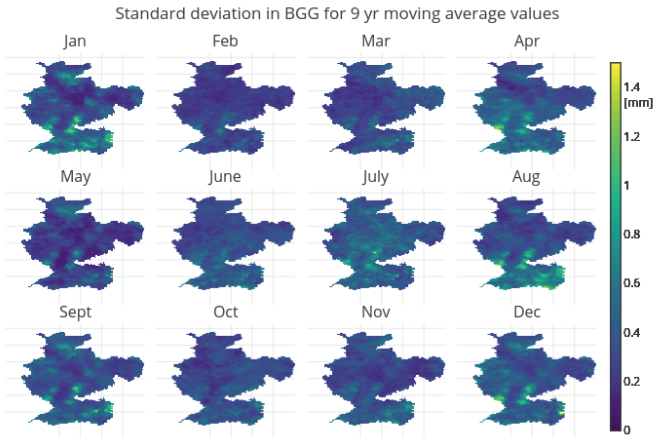
Figure A.2: *Standard deviation (not normalized) in the background grid derived from HYRAS cal-culated from 47 9-year averages.*
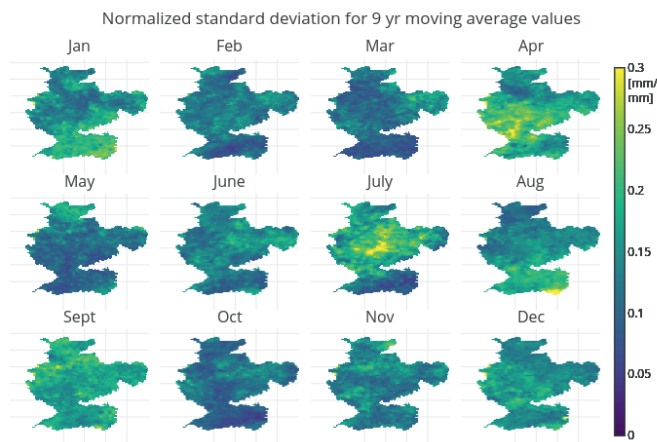
Figure A.3: Standard deviation (normalized) in the background grid derived from HYRAS calculated from 47 9-year averages.
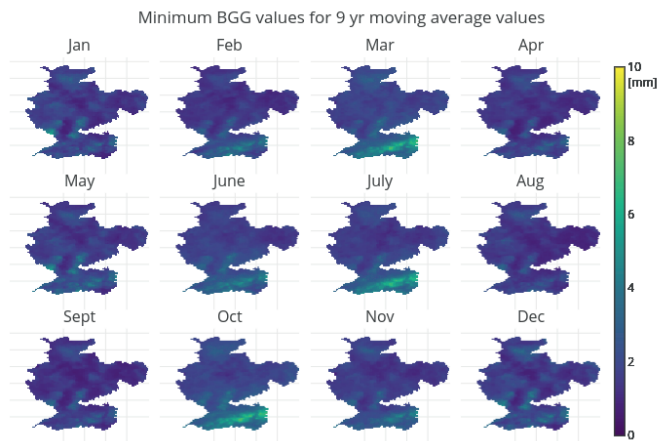
Figure A.4: Minimum background grid values derived from HYRAS out of the 47 9-year periods.

Maximum BGG values for 9 yr moving average values
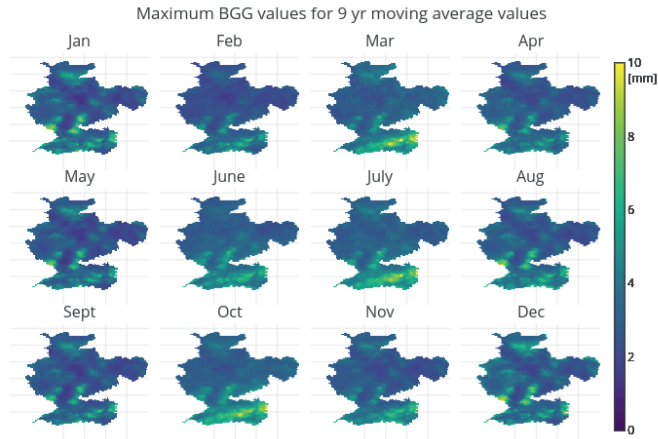
Figure A.5: Maximum background grid values derived from HYRAS out of the 47 9-year periods.

Difference between max and min BGG values
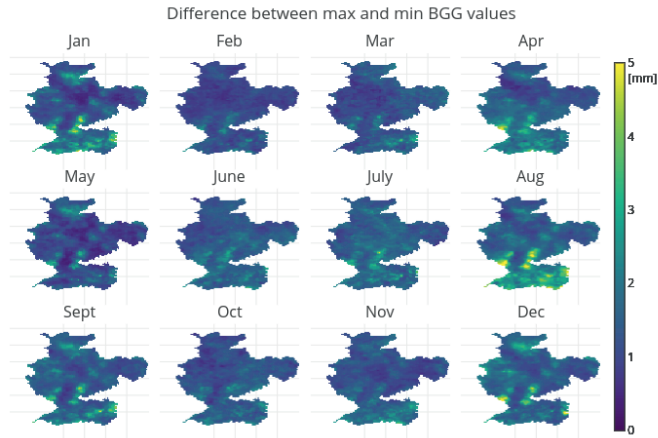
Figure A.6: Difference, or range, between maximum and minimum background grid values derived from HYRAS out of the 47 9-year periods.
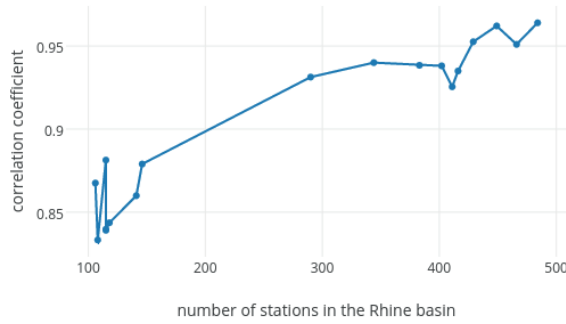
Figure A.7: *Correlation between yearly precipitation accumulations calculated with IDW and with genRE_hyras as a function of the number of stations in the Rhine basin.*

B   Supplemental figures to Chapter 4
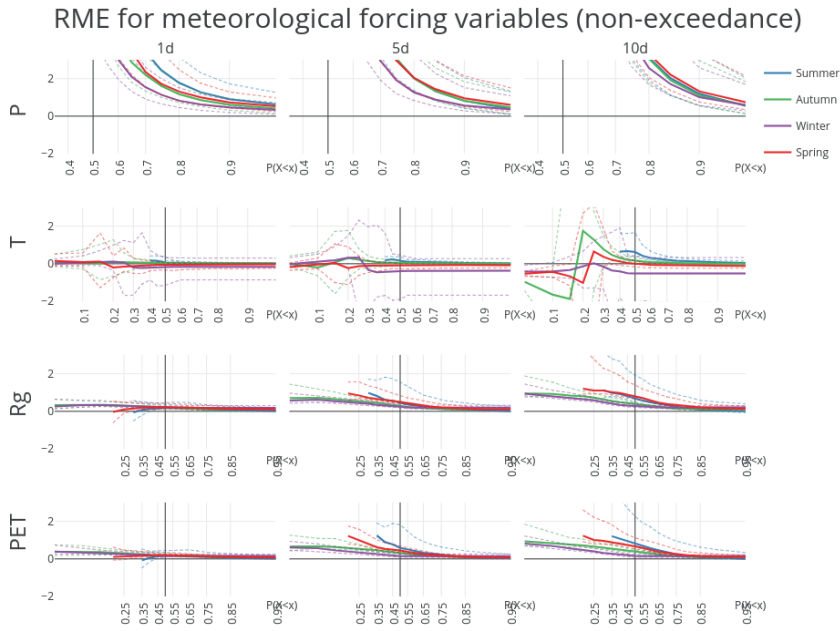
Figure B.1: Relative Mean Error (RME) for the four forcing variables benchmarked against sample climatology for the 148 HBV subbasins for the whole year. RME is aggregated into mean (solid), 10th and 90th percentile (dashed). The CRPSS score at $P(X \leq x) = 0.1$ resp. 0.7 are calculated over respectively the 10% and 70% lowest observation-forecast pairs, conditioned on the observations. Note that radiation (Rg third row) is indeed overestimated for low extremes as presented in the main text. Additionally, the asymptotic behaviour of the RME of precipitation (P, first row) is caused by the large number of events with zero or close to zero precipitation, so that the relative error grows without bounds. In the inverse figure (for P(X>x), Fig. 6) those values were automatically excluded. For temperature (T, second row), the RME is unstable for values around zero, but since actual temperatures of exactly zero are rare, this remains within bounds, albeit with a jump from positive to negative due to sign differences between observed and forecasted values.
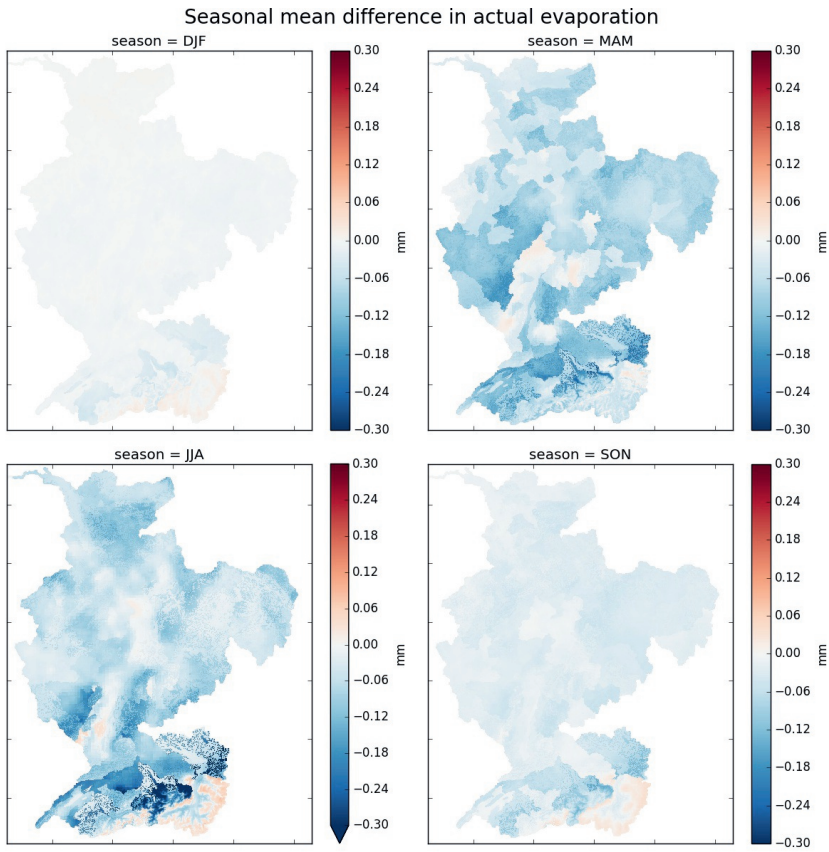
Figure B.2: *Seasonal Mean Difference in calculated actual evaporation (aevap) for each season.
Actual evaporation includes evaporation from interception.*
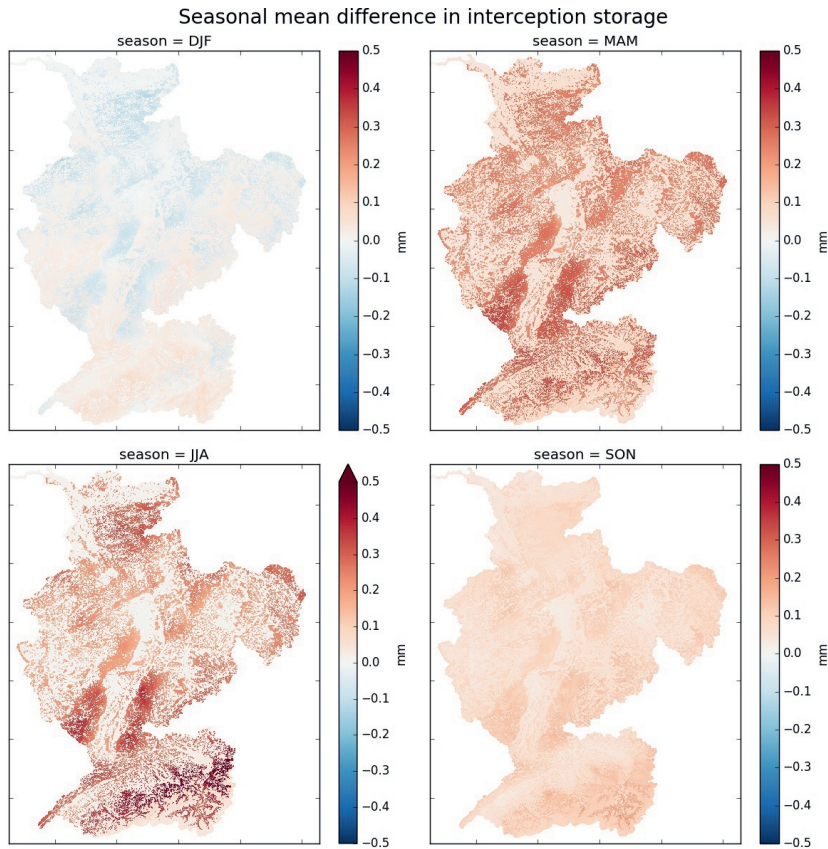
Figure B.3: Seasonal Mean Difference in calculated interception storage (ic) for each season. This is not the average interception flux, but the average storage in the interception reservoir so that wetter interception stores means less interception.
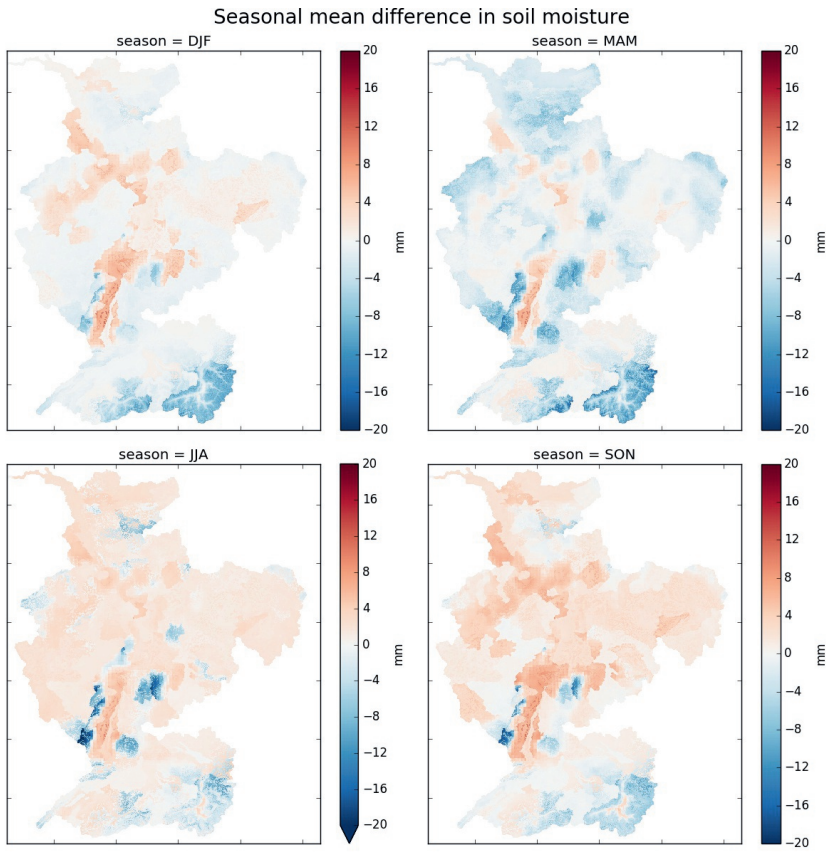
Figure B.4: *Seasonal Mean Difference in calculated soil moisture (sm) for each season.*
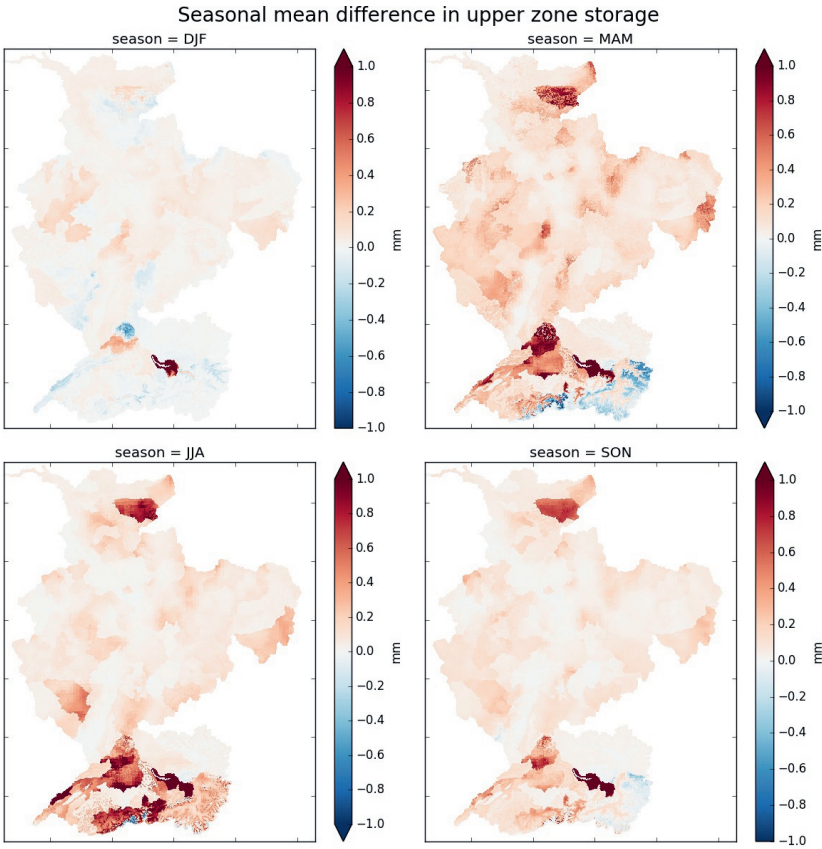
Figure B.5: Seasonal Mean Difference in calculated upper zone storage (uz) for each season.

Seasonal mean difference in lower zone storage
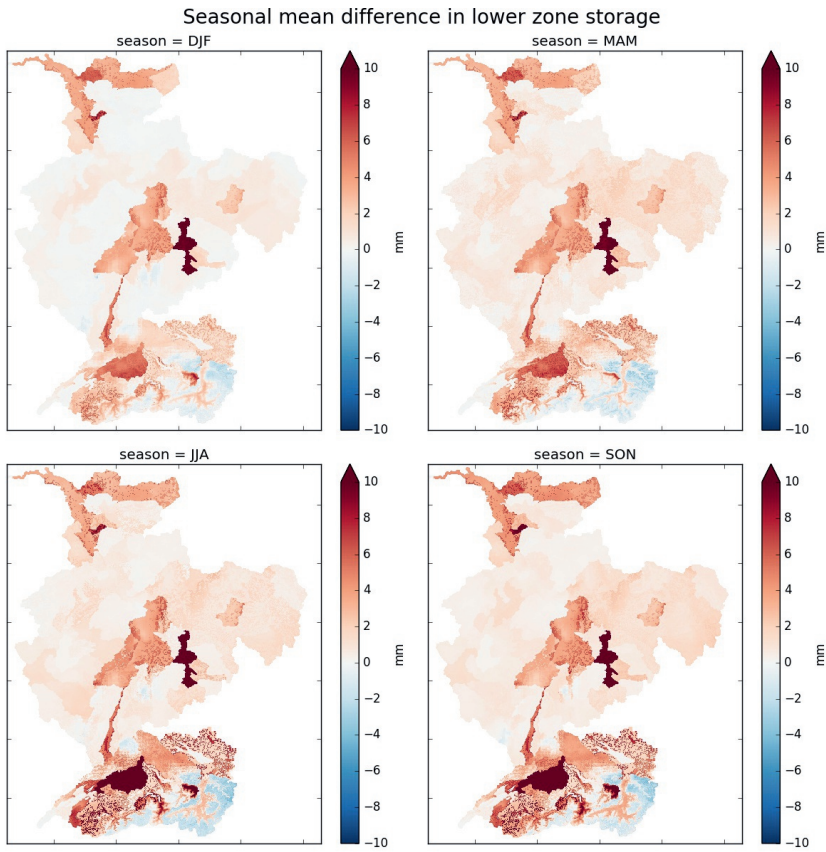
Figure B.6: *Seasonal Mean Difference in calculated lower zone storage (lz) for each season.*
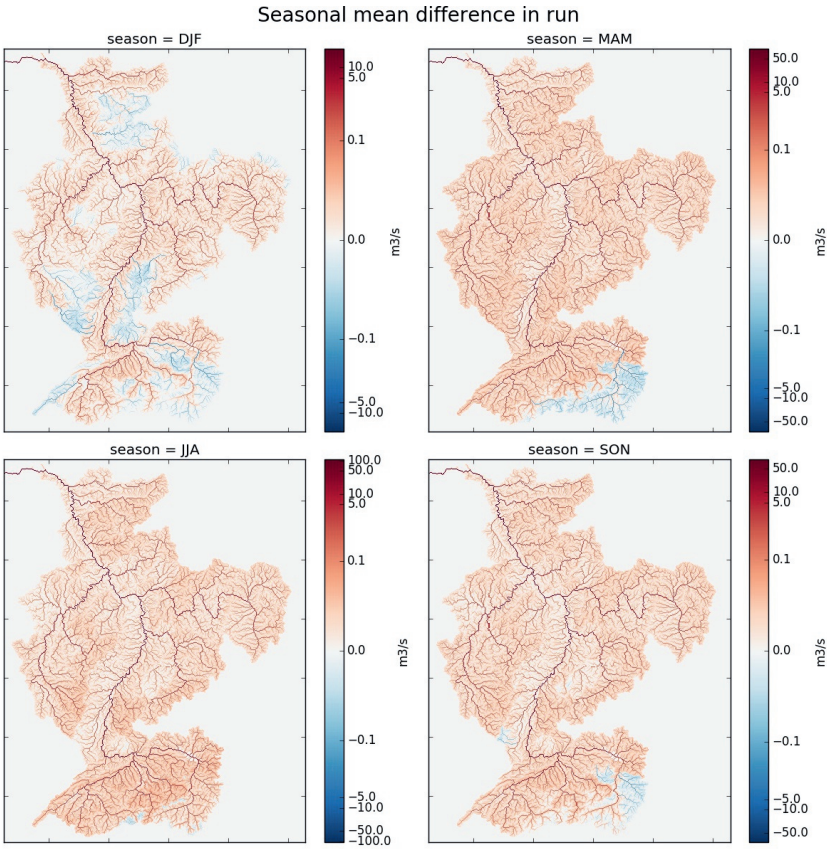
Figure B.7: Seasonal Mean Difference in calculated discharge (run) for each season.

C   Supplemental tables and figures to Chapter 5

In this section the implementation of the PTFs from Chapter 5 is described in more detail.

SoilThickness
The SoilThickness parameter describes the depth of the upper (phreatic) aquifer. An estimation of this parameter (per grid cell) is possible and purely depends on the depth to either an impermeable layer or bedrock. Depths to bedrock are globally available from *Hengl et al.* (2017) and depths to impermeable layers are available for Eurasia from the *ESDAC* (2004) dataset. Note, however, that both datasets are limited to a depth of 2 meters, which is acceptable for shallow soils, but will result in an underestimation for deeper phreatic aquifers.

Saturated water content
θ_r and θ_s are parameter values for respectively the residual and saturated water content. Per grid cell, both parameters can be estimated by two empirical functions as proposed by *Tóth et al.* (2015). The equations are present in Table B1 of the main paper's appendix. Both parameters thus depend on soil properties, which are available at seven depths from the 250m SoilGrids dataset as described by *Hengl et al.* (2017).

Manning parameters
Manning's coefficient consists of two different parameters in wflow_sbm: N for overland flow (non-river cells) and N_river for water flow through watercourses (river cells). It is possible to estimate the Manning parameter, N, for all non-river cells, by making use of the land cover. *Engman* (1986) and *Kilgore* (1997) made a list of Manning values for various land covers (i.e. a look-up table for these values). Hence, N can be determined when a land cover map, such as Corine (*European Environment Agency*, 2018), is present.
The estimation of the N_river parameter is possible via a similar method which is based on the Strahler order instead of the land cover (Table C.5), as proposed and used by *Liu et al.* (2005) for the TOPKAPI model. Note, however, that this relationship between Strahler order and Manning's roughness coefficient is purely empirical and is not often tested.

Saturated conductivity
The parameter *KsatVer* is a representation of the saturated conductivity of the soil. It is hard to estimate the saturated conductivity due to the spatial variability of this soil property. A couple of empirical equations are, however, available to estimate this parameter. A study by *Wagner et al.* (2001) concluded that the method of *Wösten* (1997) gave the best results, but this method can not be used when the percentage of organic matter in the soil exceeds 15%. For these cases, the method by *Brakensiek et al.* (1984) is recommended. For consistency, we only use the method by *Brakensiek et al.* (1984) in this approach. The PTF is shown in Table B1 of the paper's appendix. Since the soil data of *Hengl et al.* (2017) is available at seven depths, we have also estimated *KsatVer* at seven depths.

M
The model only requires the *KsatVer* of the top layer and it will approximate the satu-

rated conductivity of the deeper layers with the *M*-parameter. *M* governs the (often exponential) decay of the saturated conductivity with depth (*Vertessy and Elsenbeer*, 1999; *Schellekens et al.*, 2019b). With the assumption that this decay is often exponential, *KsatVer* at depth z_i is estimated in wflow_sbm with the *M*-parameter. Since we have calculated *KsatVer* at seven depths, it is possible to fit the decay-function of *M* through the values of *KsatVer*. The used equations for this are also shown in Table B1 of the appendix of the paper. Hence, with the present data, we have calculated *KsatVer* at seven depths and we have determined *M* by making use of the estimated *KsatVer* at those seven layers.

Water fraction
The parameter *WaterFrac* represents the fraction of water per pixel, which is used for the open water evaporation estimations in the model. Grid cells have a value between 0 and 1, with 1 representing a cell with only water and 0 a cell with no open water at all. This parameter is estimated with the used land cover map in this study (*European Environment Agency*, 2018).

Rooting depth
RootingDepth is a parameter which gives the effective depth of vegetation roots on a particular pixel. The estimation of this parameter is based on the work by *Schenk and Jackson* (2002), who estimated rooting depth for different vegetation classes by making use of the following fitted curve per vegetation type:

$$R(d) = \frac{R_{max}}{[1 + (\frac{d}{d_{50}})^c]} \, .$$
(8.1)

Here, $R(d)$ is the cumulative root amount above depth d, R_{max} is the total amount of roots, d_{50} is the depth at which the cumulative root amount (R) equals $0.5 \times R_{max}$ and c is a dimensionless coefficient which is fitted to the available vegetation data.
The required data to determine the depth of the roots for a certain vegetation type and cumulative root fraction, are the d_{50} length and coefficient c, which are available for a variety of vegetation types in *Schenk and Jackson* (2002) and *Fan et al.* (2016b). In this study, we have related these vegetation types to the used land cover dataset by *European Environment Agency* (2018). And with the available datasets, we have calculated the d_{75} length, i.e. the depth up to which 75% are present. The reason for this depth is twofold: (1) The maximum rooting depth (d_{100}) is much longer than the effective rooting depth as used by wflow_sbm, and (2) at the d_{75} length, to be precise in between 60 and 80%, there is a tipping point where the steepness of the rooting depth as a function of cumulative root density percentage highly increases. This point, we assume to be the end of the effective rooting depth.

c and λ
The parameter c is the value for the exponent in the wflow_sbm vertical transfer equation which determines the transfer of water from the unsaturated to the saturated zone per soil layer. This parameter has its origin in the Brooks-Corey equations and we can estimate c when we know the value for λ (*Brooks and Corey*, 1964). These equations are also illustrated in appendix Table B1 of the paper. When λ is estimated and upscaled, c is cal-

culated from the λ parameter fields on the model resolution, following *Rawls and Brakensiek* (1989).

Leaf Area Index
Another option in the model, is to make use of the Leaf Area Index (LAI) in order to estimate interception and evaporation losses. This requires estimations of LAI and introduces three new parameters (see below). In this study, we used LAI estimates from the MODIS/Terra Leaf Area Index global dataset by *Myneni et al.* (2015). These estimates are averaged per month over the years in the run period, resulting in one map per month.

The introduced parameters are Specific leaf storage (Sl), the woody fraction of the vegetation (Swood) and an extinction coefficient (Kext). These three parameters can be related to land cover according to look-up tables as derived for wflow_sbm by *Schellekens et al.* (2019b). The look-up tables are determined from *Pitman* (1989) and *Liu* (1998) for Sl and Swood, and determined from *van Dijk and Bruijnzeel* (2001) for Kext.

Table C.1: Complete List of datasets used in Chapter 5.

Name dataset	Unit	Additional information	Source
Precipitation (genRE)	mm	Time period: 1996-2015	*van Osnabrugge et al.* (2017)
Air temperature	°C	Time period: 1996-2015	*van Osnabrugge et al.* (2019)
Potential Evapotranspiration	mm	Time period: 1996-2015	*van Osnabrugge et al.* (2019)
LSA SAF DMET (ET$_{act}$, 9km^2)	mm	Time period: starting from 2011-11-14	*Trigo et al.* (2011)
Discharge at multiple locations in the Rhine basin	m^3 s^{-1}	Time period: 1997-2015	Same as used in *van Osnabrugge et al.* (2017)
Leaf Area Index	m^2 m^{-2}	Time period: monthly 2000-02 until 2017-02	*Myneni et al.* (2015)
ISRIC SoilGrids 250m	-	See below	*Hengl et al.* (2017)
	kg m^{-3}	Bulk Density	
	-	Soil Types	
	kg m^{-3}	Soil Organic Carbon	
	%	Sand content	
	%	Clay content	
	%	Silt content	
	cm	Depth to bedrock	
ESDAC Depth to Impermeable layers	Classes [cm]		*ESDAC* (2004); *Panagos et al.* (2012)
Global Land Cover	-		*European Environment Agency* (2018)

Table C.2: List of parameters estimated with a PTF consisting of an (empirical) equation. Their equations and the source of the PTF are stated below. Only the abbreviated parameter names are illustrated, see Table 5.2 for the interpretation of the parameters.

Parameter	Equation[a]	Source
c	$$c = 3 + \frac{2}{\lambda}$$ (8.2)	(Rawls and Brakensiek, 1989)
	$$\lambda = exp[-0.784 + 0.018 \cdot Sa - 1.062 \cdot \theta_s - Sa^2 \cdot (5 \cdot 10^{-5}) - 0.003 \cdot Cl^2$$ $$+1.111 \cdot \theta_s^2 - 0.031 \cdot Sa \cdot \theta_s + (3 \cdot 10^{-4}) \cdot Sa^2 \cdot \theta_s^2 - 0.006 \cdot Cl^2$$ $$\cdot \theta_s^2 - (2 \cdot 10^{-6}) \cdot Sa^2 \cdot Cl + 0.008 \cdot Cl^2 \cdot \theta_s - 0.007 \cdot \theta_s^2 \cdot Cl]$$ (8.3)	
KsatVer	$$K_s = 240.19 \cdot exp[19.52348 \cdot \theta_s - 8.96847 - 0.028212 \cdot Cl + (1.8107 \cdot 10^{-4})$$ $$\cdot Sa^2 - (9.4125 \cdot 10^{-3}) \cdot Cl^2 - 8.395215 \cdot \theta_s^2 + 0.077718 \cdot Sa \cdot \theta_s$$ $$-0.00298 \cdot Sa^2 \cdot \theta_s^2 - 0.019492 \cdot Cl^2 \cdot \theta_s^2 + (1.73 \cdot 10^{-5}) \cdot Sa^2 \cdot Cl$$ $$+0.02733 \cdot Cl^2 \cdot \theta_s + 0.001434 \cdot Sa^2 \cdot \theta_s - (3.5 \cdot 10^{-6}) \cdot Cl^2 \cdot Sa]$$ (8.4)	Brakensiek et al. (1984)
M	$$K_s(z_i) = K_s(0) \cdot e^{-f \cdot z_i}$$ (8.5) $$f = \frac{\theta_s - \theta_r}{M}$$ (8.6)	
θ_r & θ_s	$$\theta_r = 0.09878 + 0.002127 \cdot Cl - (8.366 \cdot 10^{-4}) \cdot Si - \frac{0.0767}{OC + 1} + Si \cdot Cl$$ $$\cdot (3.853 \cdot 10^{-5}) + \frac{0.00233 \cdot Cl}{OC + 1} + \frac{(9.498 \cdot 10^{-4}) \cdot Si}{OC + 1}$$ (8.7)	Tóth et al. (2015)
	$$\theta_s = 0.6819 + \frac{0.06480}{OC + 1} - 0.119 \cdot BD^2 - 0.02668 + (8.031 \cdot 10^{-4}) \cdot Si + \frac{0.02321 \cdot BD^2}{OC + 1}$$ $$+Cl \cdot 0.001489 + 0.01908 \cdot BD^2 - 0.001109 \cdot Cl - (2.315 \cdot 10^{-5})$$ $$\cdot Si \cdot Cl - 0.001197 \cdot Si \cdot BD^2 - (1.068 \cdot 10^{-4}) \cdot Cl \cdot BD^2$$ (8.8)	

[a] Sa is % sand, Cl % clay, Si % silt, BD is the bulk density [g cm^{-3}], OC % organic carbon, OM % organic matter (approximated as 1.724 * OC (van Bemmelen, 1890)). These basin predictors are indicated with the general term u^0.

Table C.3: Corine land cover codes(European Environment Agency, 2018) and derived wflow_sbm parameters based on empirical relationships. Manning N parameter is based on Engman (1986) and Kilgore (1997), and assigned to the land cover. Specific leaf storage (Sl) and the woody fraction of the vegetation (Swood) are based on Pitman (1989) and Liu (1998). The extinction coefficient (Kext) is based on van Dijk and Bruijnzeel (2001). Rooting depth, see also text S1, are based on Schenk and Jackson (2002) and Fan et al. (2016b). Part 1 of the table, part 2 on the next page.

Code	Land cover	N $[m^{-\frac{1}{3}} s]$	Sl [mm]	Swood [-]	Kext [-]	Rooting Depth[mm]
111	Continuous urban fabric	0.011	0.04	0.0	0.6	256
112	Discontinuous urban fabric	0.011	0.04	0.01	0.6	256
121	Industrial or commercial units	0.011	0.04	0	0.6	0
122	Road and rail networks	0.011	0.04	0.01	0.6	107
123	Port areas	0.011	0	0	0.6	0
124	Airports	0.011	0.04	0.01	0.6	107
131	Mineral extraction sites	0.011	0	0	0.6	0
132	Dump sites	0.011	0	0	0.6	0
133	Construction sites	0.011	0.04	0	0.6	0
141	Green urban areas	0.15	0.04	0.05	0.6	256
142	Sport and leisure facilities	0.15	0.04	0.01	0.6	107
211	Non-irrigated arable land	0.2	0.1272	0	0.6	390
212	Permanently irrigated land	0.2	0.1272	0	0.6	249
213	Rice fields	0.2	0.1272	0	0.6	107
221	Vineyards	0.5	0.07	0.2	0.7	432
222	Fruit trees and berry plantations	0.5	0.07	0.2	0.7	430
223	Olive groves	0.5	0.07	0.2	0.7	432
231	Pastures	0.15	0.1272	0.01	0.6	107
241	Annual crops associated with permanent crops	0.2	0.1272	0	0.6	390
242	Complex cultivation patterns	0.2	0.1272	0	0.6	390
243	Land occupied by agriculture with significant natural vegetation areas	0.44	0.1272	0	0.6	284
244	Agro-forestry areas	0.5	0.03926	0.5	0.8	406
311	Broad-leaved forest	0.6	0.036	0.5	0.8	430
312	Coniferous forest	0.4	0.045	0.5	0.8	382
313	Mixed forest	0.5	0.03926	0.5	0.8	406
321	Natural grasslands	0.15	0.1272	0.01	0.6	107
322	Moors and heathland	0.24	0.09	0.05	0.6	178
323	Sclerophyllous vegetation	0.5	0.07	0.1	0.7	432
324	Transitional woodland-shrub	0.5	0.07	0.1	0.7	432

Table C.4: Continuation of Table C.3

Code	Land cover	N $[m^{-\frac{1}{3}}s]$	Sl [mm]	Swood [-]	Kext [-]	Rooting Depth[mm]
331	Beaches dunes sands	0.15	0.04	0	0.6	107
332	Bare rocks	0.01	0	0	0.6	0
333	Sparsely vegetated areas	0.02	0.04	0.04	0.6	137
334	Burnt areas	0.02	0.04	0.04	0.6	0
335	Glaciers and perpetual snow	0.01	0	0	0.6	0
411	Inland marshes	0.15	0.1272	0.01	0.7	107
412	Peat bogs	0.15	0.1272	0	0.6	137
421	Salt marshes	0.15	0.1272	0.01	0.7	107
422	Salines	0.075	0.1272	0.005	0.7	53
423	Intertidal flats	0.075	0.1272	0.005	0.7	53
511	Water courses	0.01	0	0	0.7	0
512	Water bodies	0.01	0	0	0.7	0
521	Coastal lagoons	0.15	0.1272	0.01	0.7	107
522	Estuaries	0.15	0	0.01	0.7	107
523	Sea and ocean	0.01	0	0	0.7	0
999	NODATA	0	0	0	0	0

Table C.5: Determination of Manning's N_river based on the river Strahler order (Liu et al., 2005).

Strahler order	N_river $[m^{-\frac{1}{3}}s]$
1	0.050
2	0.040
3	0.035
≥ 4	0.030
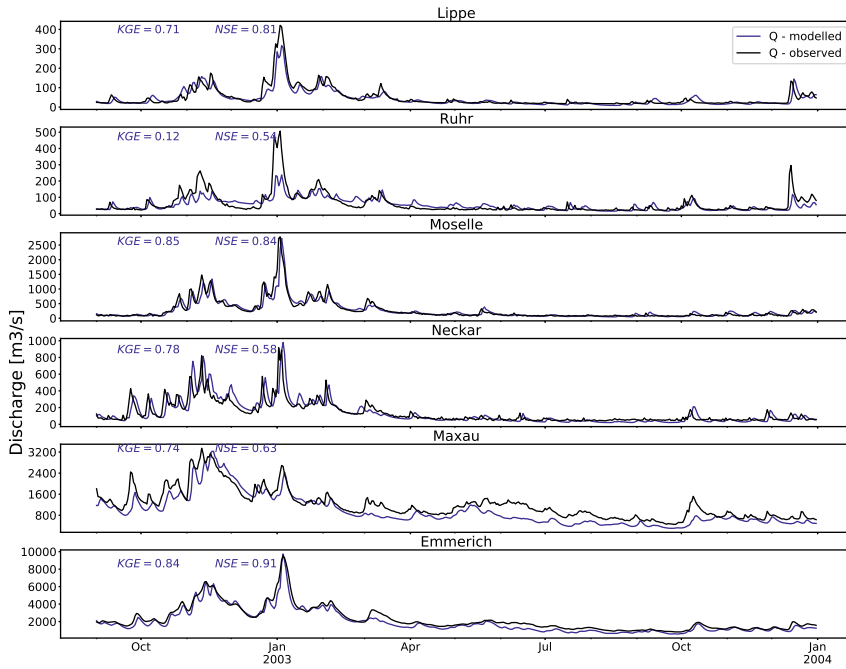
Figure C.1: Discharge simulations with wflow_sbm for four sub-catchments of the Rhine basin (Lippe, Ruhr, Moselle and Neckar) and two discharge simulations in the main river of the Rhine, one upstream (Maxau) and one downstream (Emmerich), for the hydrologically extreme year 2003. Discharge simulations are indicated by blue lines and observations in black. Reached KGE and NSE are displayed in the top left corner.
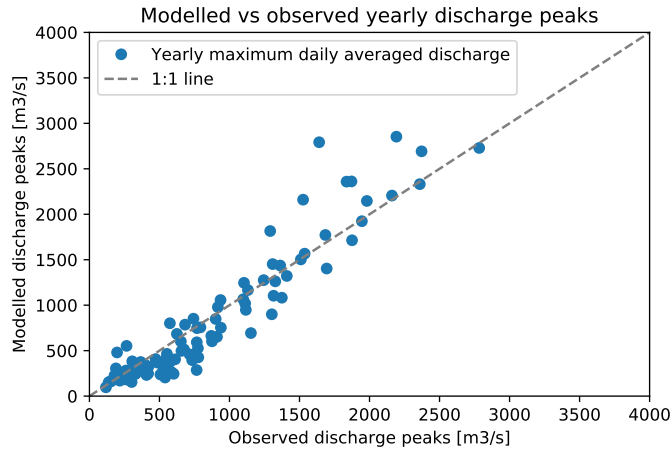
Figure C.2: Yearly maximum daily discharge peak simulations compared to observed yearly maximum daily discharge peaks. Data is based on six catchments in the Rhine basin: Moselle, Neckar, Main, Ruhr, Sieg and Lippe. Blue dots indicate modeled vs observed discharge peak and the dotted line is the 1:1 line when modeled and observed discharge peak magnitude would be the same.
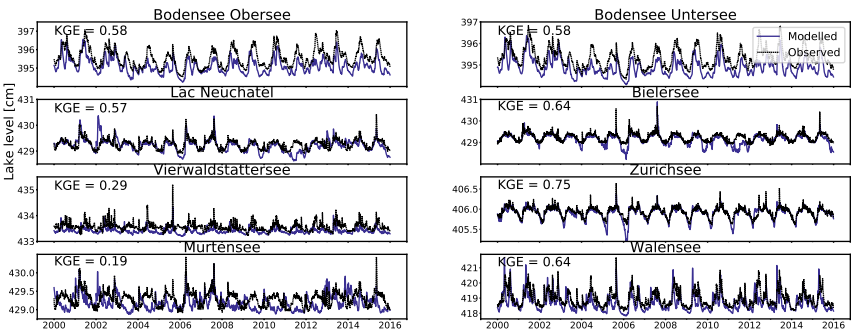
Figure C.3: Simulated lake levels of the eight largest lakes in the Rhine basin as compared with observed lake levels. All lakes are located in Switzerland or Southern Germany.
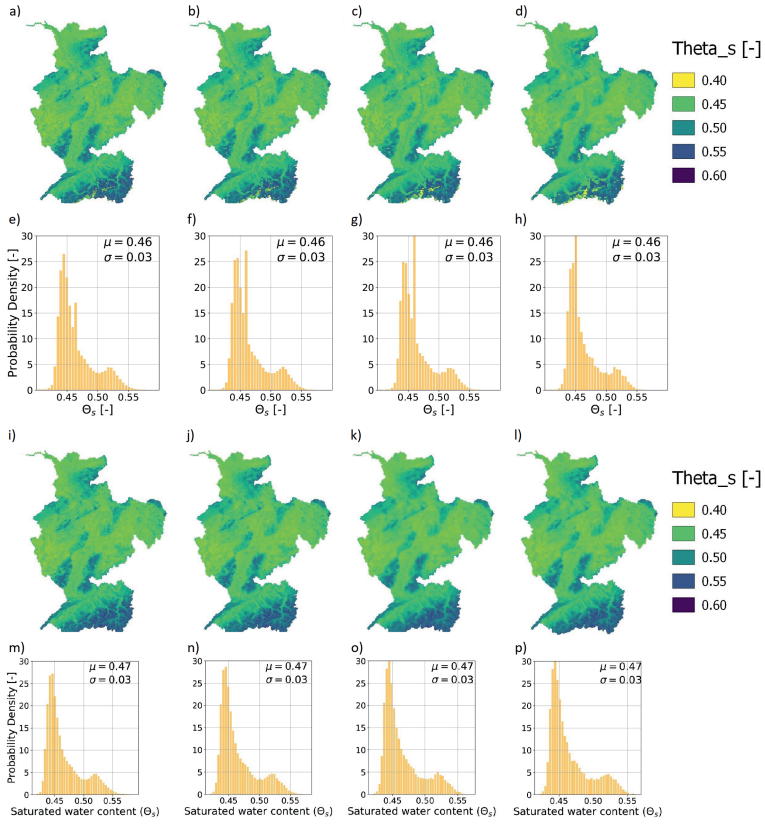
Figure C.4: Obtained parameter estimates at four resolutions for two upscaling techniques for the model parameter θ_s (saturated water content). (a)–(d), parameter estimates of θ_s from 1.2 km to 4.8 km upscaled with an arithmetic mean. (e)–(h), histograms of θ_s parameter values; histograms match with the parameter maps on top of them. (i)–(l), parameter estimates of θ_s from 1.2 km to 4.8 km as upscaled with a harmonic mean. (m)–(p), similar to the previous histograms, but for the harmonic mean upscaling.
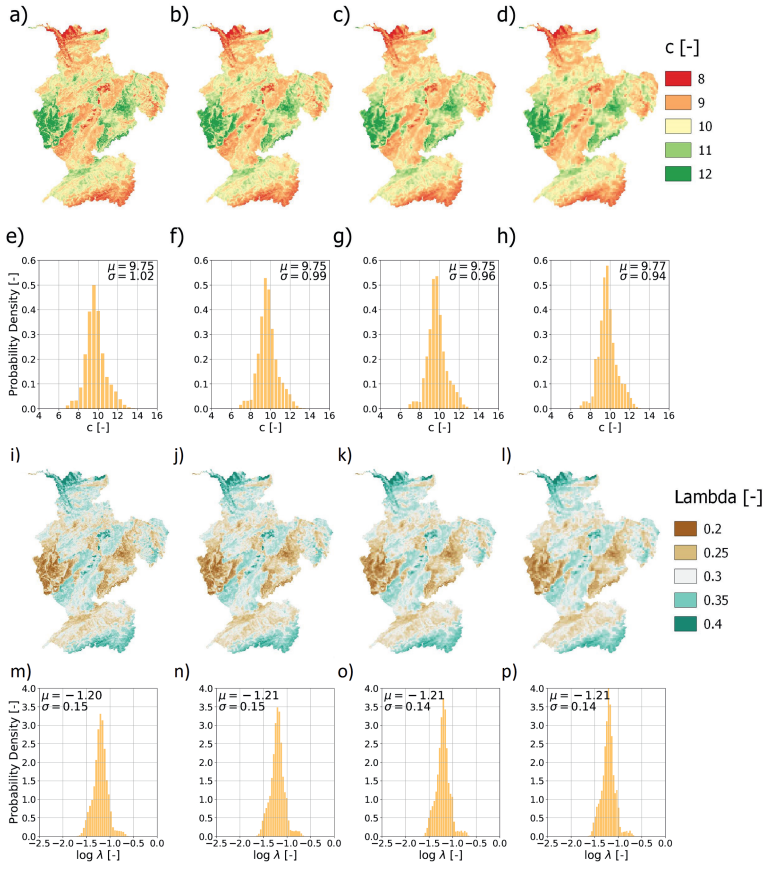
Figure C.5: Obtained parameter estimates at four resolutions for two of the model parameters: c and λ. (a)–(d), parameter estimates of c from 1.2 km to 4.8 km as upscaled with an arithmetic mean. (e)–(h), histograms of c parameter values; histograms match with the parameter maps on top of them. (i)–(l), parameter estimates of λ from 1.2 km to 4.8 km as upscaled with an arithmetic mean of the natural logarithm. (m)–(p), similar to the previous histograms, but of the natural logarithm of λ.
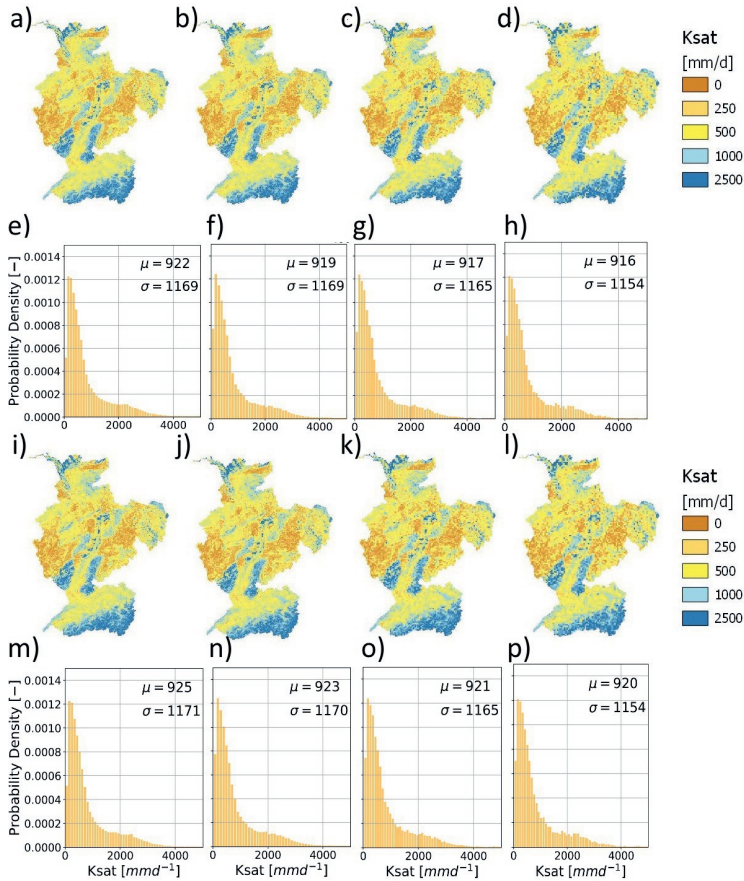
Figure C.6: *Obtained parameter estimates at four resolutions for KsatVer (saturated conductivity). (a)–(d), parameter estimates from 1.2 km to 4.8 km as upscaled with an arithmetic mean after upscaling from the natural logarithm. (e)–(h), histograms of $KsatVer$ parameter values; histograms match with the parameter maps on top of them. (i)–(l), parameter estimates from 1.2 km to 4.8 km as upscaled with an arithmetic mean of the normal parameter value. (m)–(p), similar to the previous histograms, but for (i–l).*
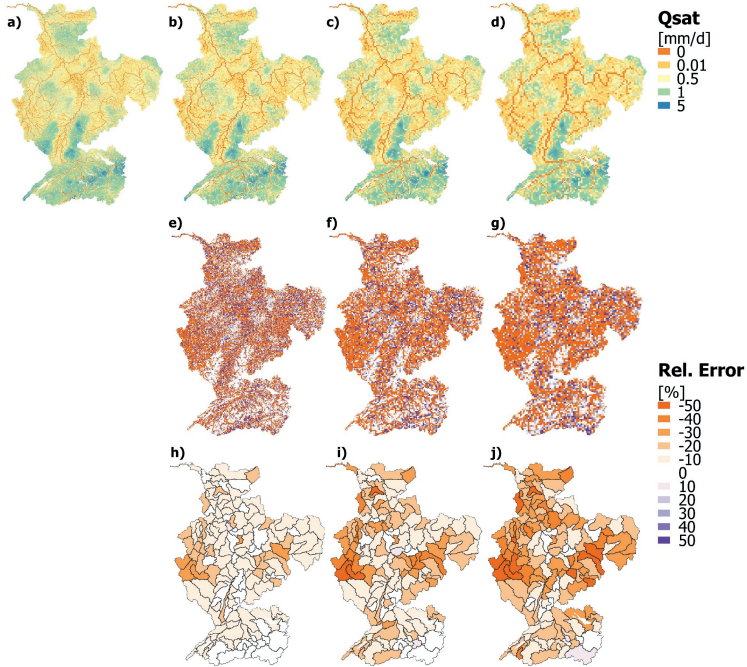
Figure C.7: Simulations of daily averaged lateral subsurface flows on four resolutions, (a) 1.2 km, (b) 2.4 km, (c) 3.6 km and (d) 4.8 km. (e) - (g) illustrate the relative error per grid cell between one of the coarser resolutions (b–d) and the simulations on 1.2 km (a) after upscaling these simulations to one of the coarser resolutions. (h) - (j), same as (e–g), but catchment averaged fluxes are used instead of fluxes per grid cell.
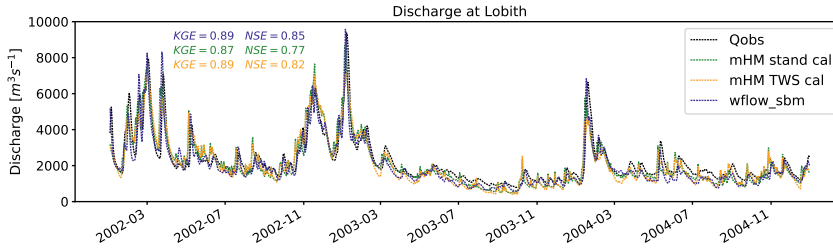
Figure C.8: Discharge simulations for three different model setups as compared with observations with Lobith (main course of the Rhine at the Dutch-German border). The illustrated model setups are wflow_sbm in blue (this study), mHM setup with transfer-functions with a calibration constrained by discharge (orange) and mHM with transfer-function with a calibration constrained by both discharge and GRACE data (green) Rakovec et al. (2016a).
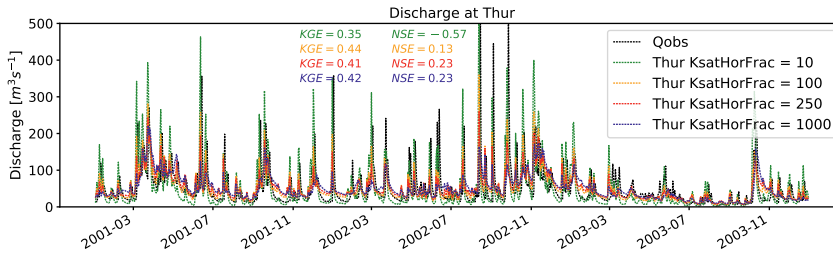
Figure C.9: Discharge simulations for the Thur for four different model setups: KsatHorFrac = 10 (green), KsatHorFrac = 100 (orange), KsatHorFrac = 250 (red) and KsatHorFrac = 1000 (blue). Discharge observations are indicated in black.
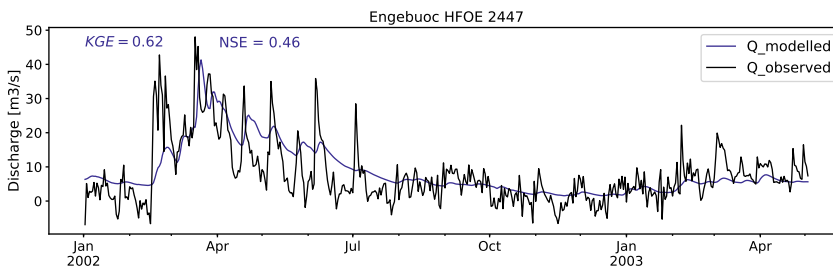
Figure C.10: Discharge simulations as compared with observations at Engebuoc (Switzerland).
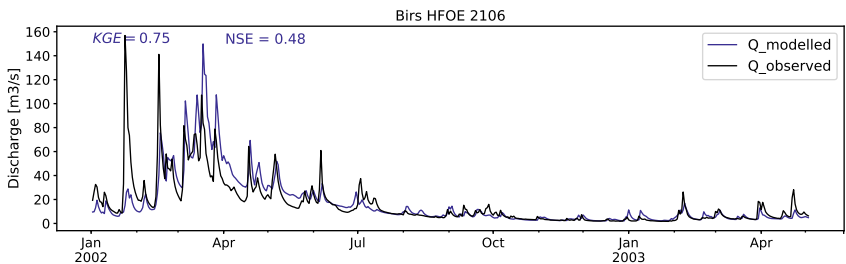
Figure C.11: Discharge simulations as compared with observations at Birs (Switzerland).
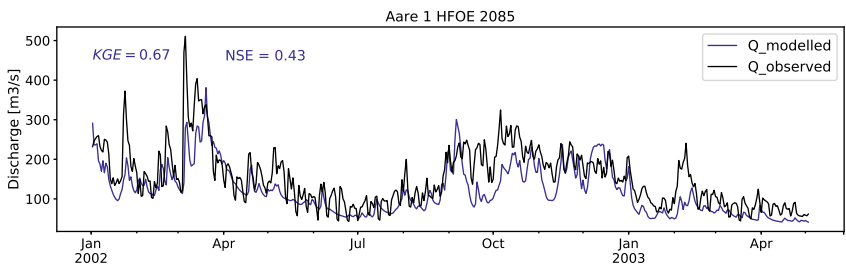
Figure C.12: Discharge simulations as compared with observations at Aare 1 (Switzerland).
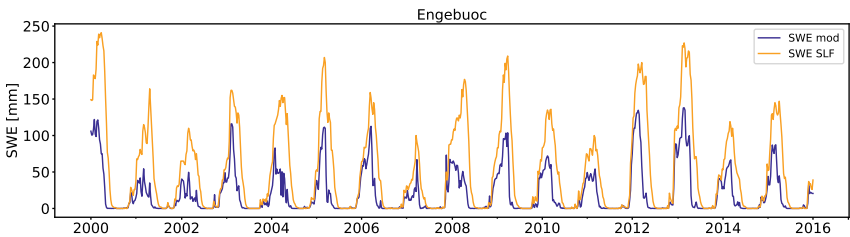
Figure C.13: SWE simulations as compared with SWE simulations from SLF Jörg-Hess et al. (2014) for Engebuoc (Switzerland).
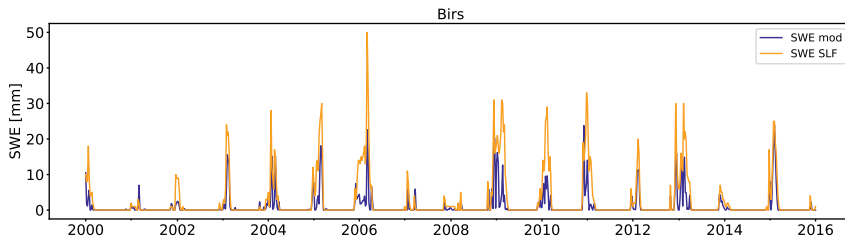
Figure C.14: SWE simulations as compared with SWE simulations from SLF Jörg-Hess et al. (2014) for Birs (Switzerland).
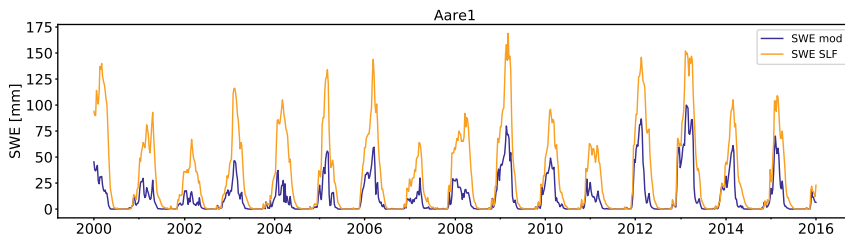
Figure C.15: SWE simulations as compared with SWE simulations from SLF Jörg-Hess et al. (2014) for Aare 1 (Switzerland).

D   Supplemental tables and figures for Chapter 7
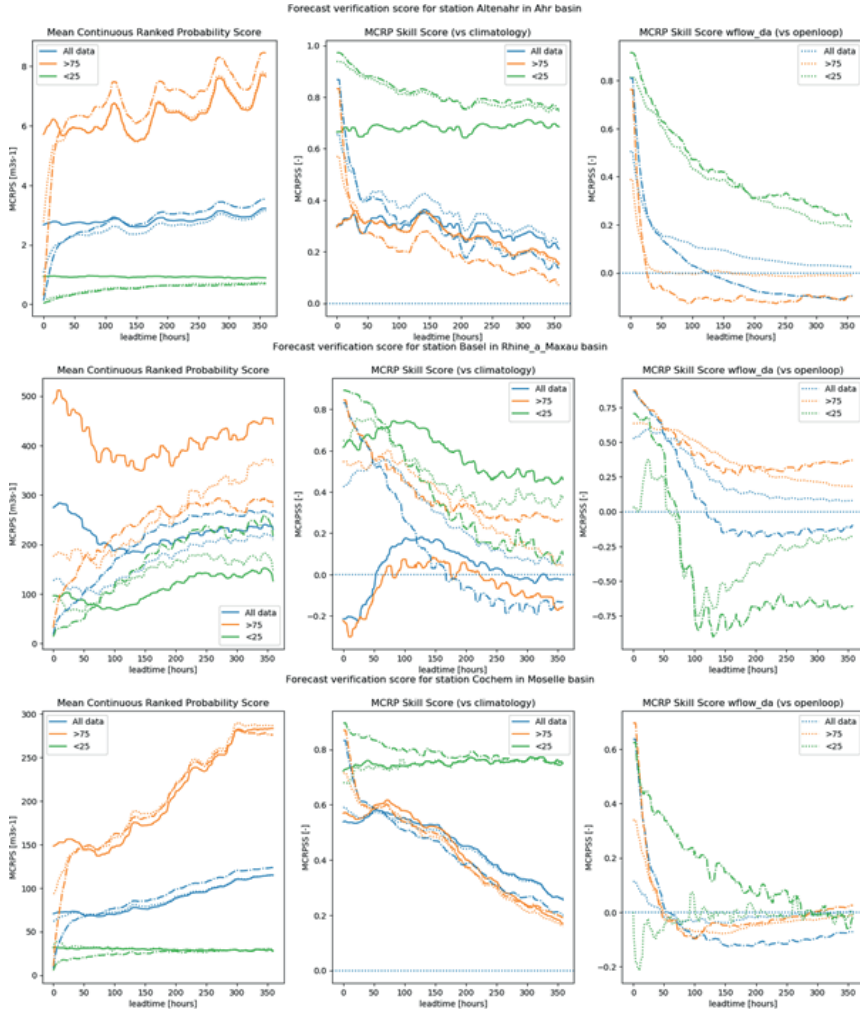
Figure D.1: *Forecast verification scores for several stations. MCRPS (left panel) and MCRPSS (middle, right). The scores are shown for three scenarios: open loop (solid line), AEnKF (dotted line) and open loop with AR correction (stripe dot line), and for three quantiles (see legend). The difference between the middle and right panel is that for the first the MCRPSS is calculated with sample climatology as benchmark forecast and for the right panel the MCRPSS is calculated with the forecast from the open loop initial state as benchmark.*
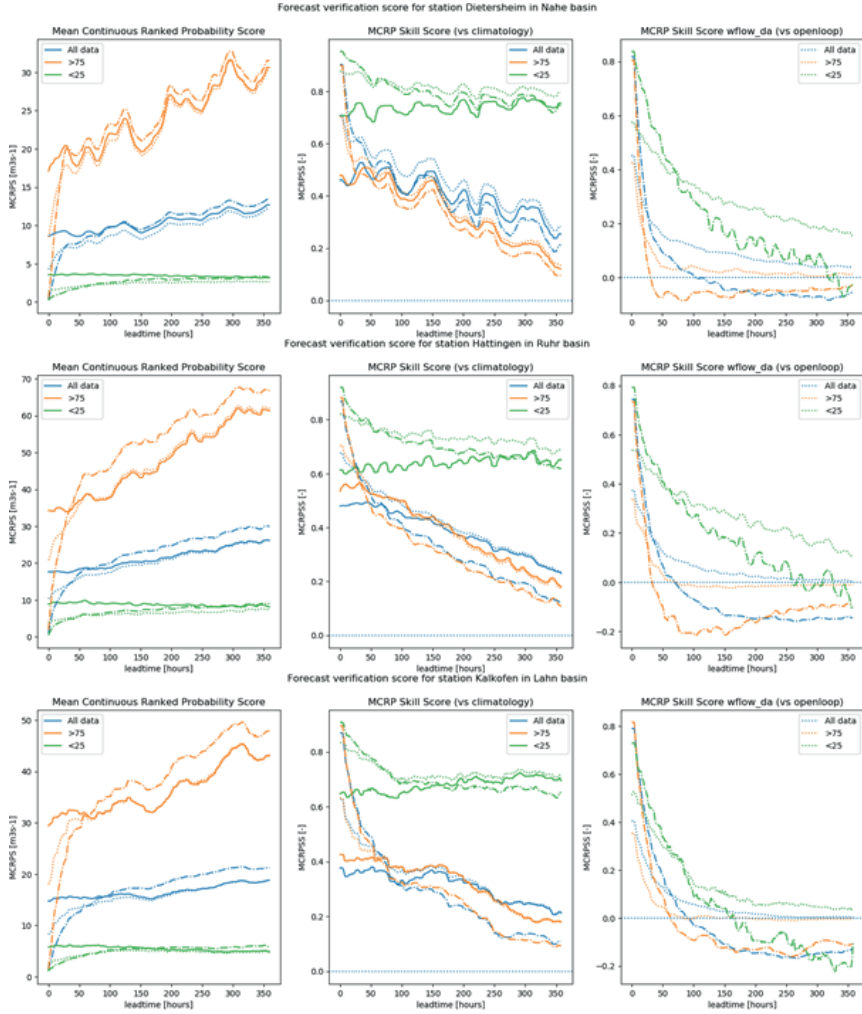
Figure D.2: Forecast verification scores for several stations. MCRPS (left panel) and MCRPSS (middle, right). The scores are shown for three scenarios: open loop (solid line), AEnKF (dotted line) and open loop with AR correction (stripe dot line), and for three quantiles (see legend). The difference between the middle and right panel is that for the first the MCRPSS is calculated with sample climatology as benchmark forecast and for the right panel the MCRPSS is calculated with the forecast from the open loop initial state as benchmark.
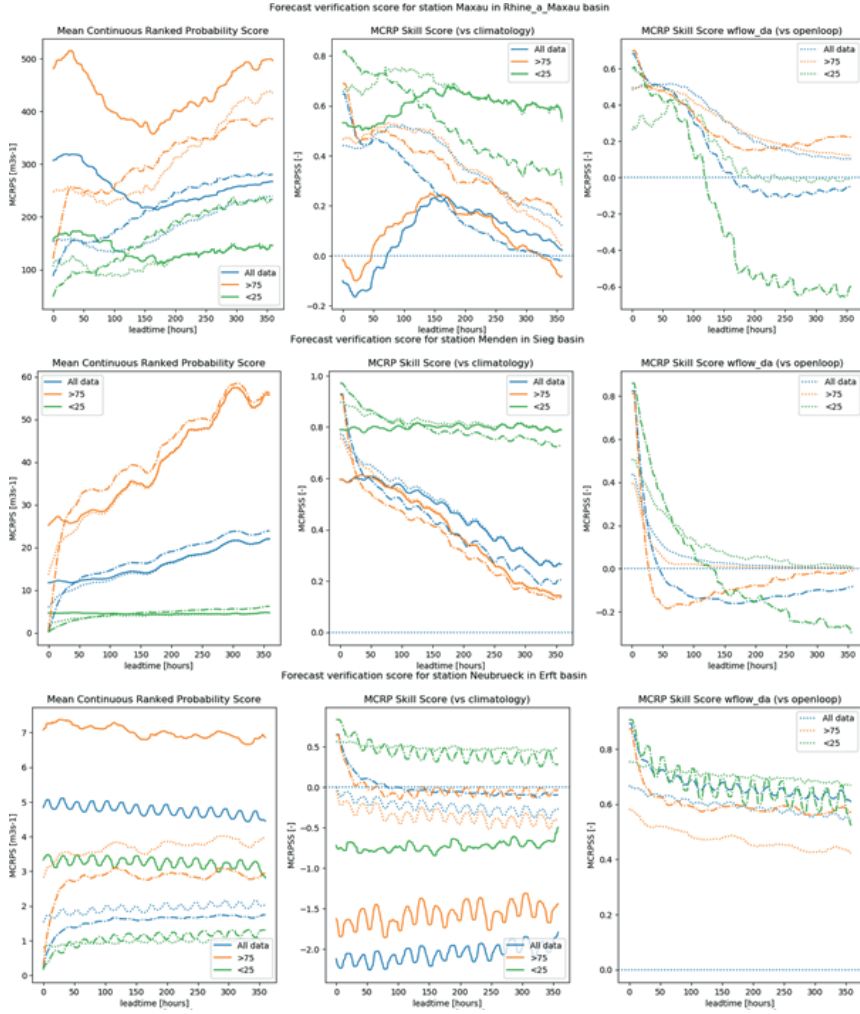
Figure D.3: Forecast verification scores for several stations. MCRPS (left panel) and MCRPSS (middle,right). The scores are shown for three scenarios: open loop (solid line), AEnKF (dotted line) and open loop with AR correction (stripe dot line), and for three quantiles (see legend). The difference between the middle and right panel is that for the first the MCRPSS is calculated with sample climatology as benchmark forecast and for the right panel the MCRPSS is calculated with the forecast from the open loop initial state as benchmark.
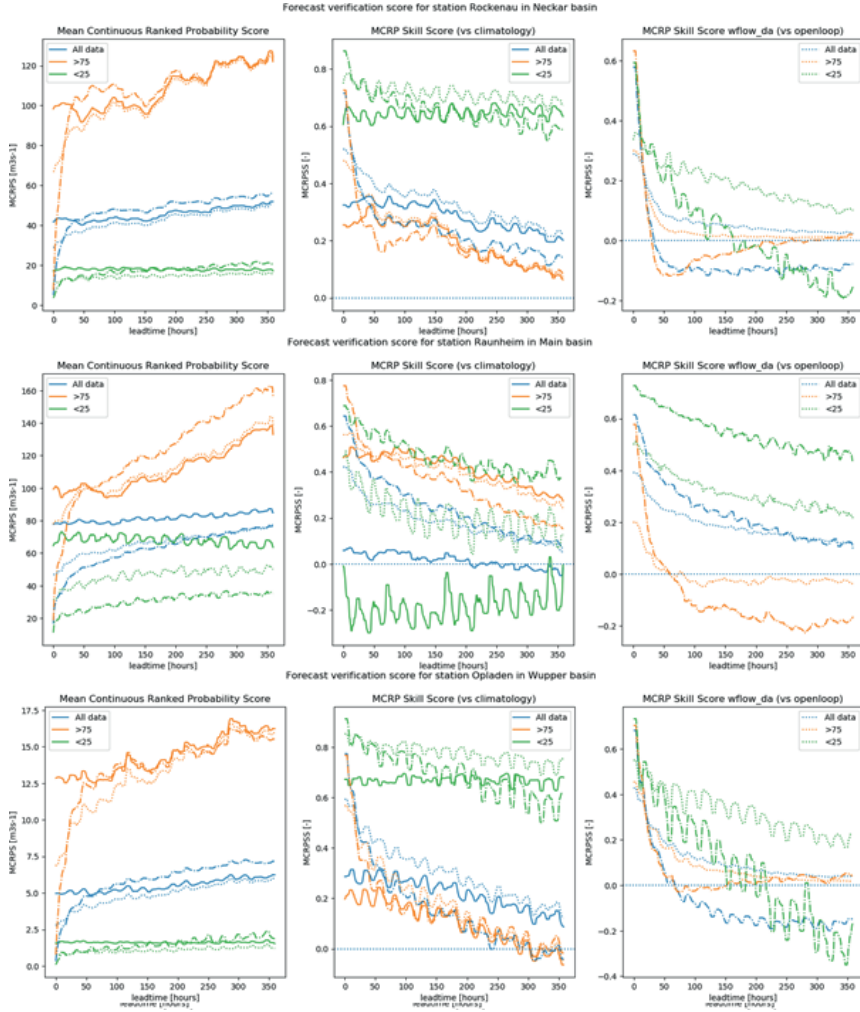
Figure D.4: Forecast verification scores for several stations. MCRPS (left panel) and MCRPSS (middle, right). The scores are shown for three scenarios: open loop (solid line), AEnKF (dotted line) and open loop with AR correction (stripe dot line), and for three quantiles (see legend). The difference between the middle and right panel is that for the first the MCRPSS is calculated with sample climatology as benchmark forecast and for the right panel the MCRPSS is calculated with the forecast from the open loop initial state as benchmark.
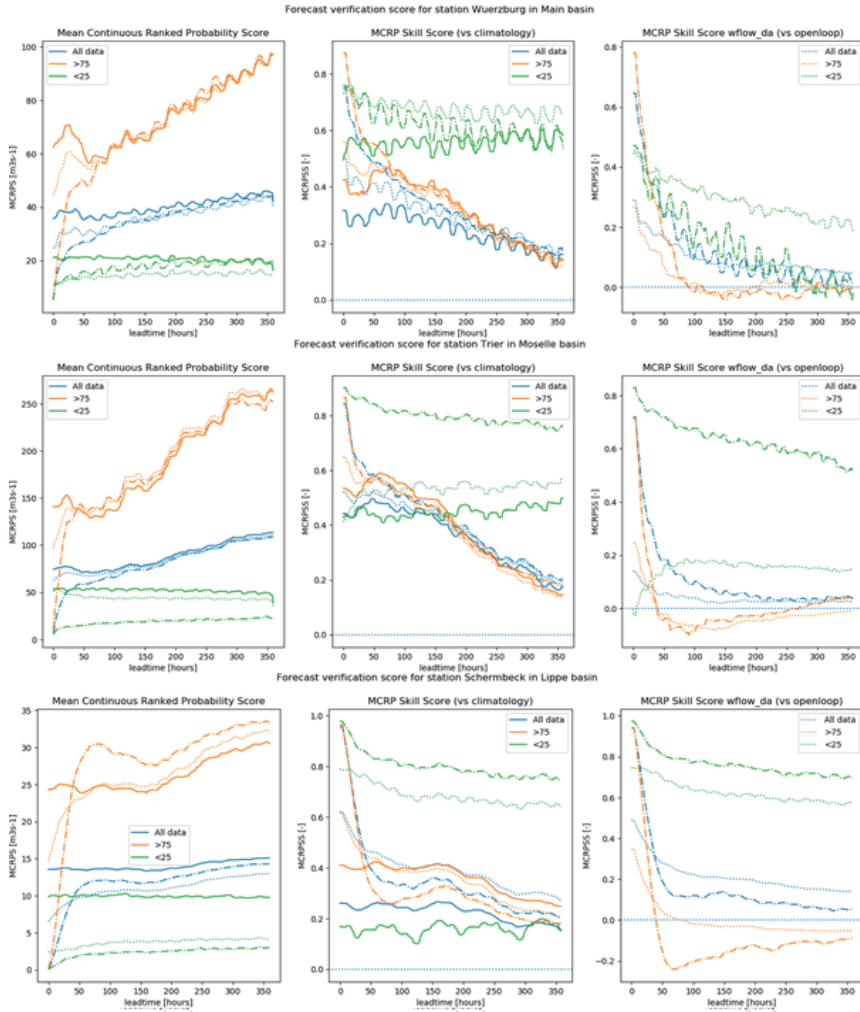
Figure D.5: Forecast verification scores for several stations. MCRPS (left panel) and MCRPSS (middle, right). The scores are shown for three scenarios: open loop (solid line), AEnKF (dotted line) and open loop with AR correction (stripe dot line), and for three quantiles (see legend). The difference between the middle and right panel is that for the first the MCRPSS is calculated with sample climatology as benchmark forecast and for the right panel the MCRPSS is calculated with the forecast from the open loop initial state as benchmark.

Bibliography

Addor, N., S. Jaun, F. Fundel, and M. Zappa (2011), An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): Skill, case studies and scenarios, *Hydrology and Earth System Sciences*, *15*(7), 2327–2347, doi: $10.5194/\mathrm{hess}\text{-}15\text{-}2327\text{-}2011$.

Addor, N., A. J. Newman, N. Mizukami, and M. P. Clark (2017), The CAMELS data set: Catchment attributes and meteorology for large-sample studies, *Hydrology and Earth System Sciences*, *21*(10), 5293–5313, doi: $10.5194/\mathrm{hess}\text{-}21\text{-}5293\text{-}2017$.

Albergel, C., G. Balsamo, P. d. Rosnay, J. Muñoz-Sabater, and S. Boussetta (2012), A bare ground evaporation revision in the ECMWF land-surface scheme: evaluation of its impact using ground soil moisture and satellite microwave data, *Hydrology and Earth System Sciences*, *16*(10), 3607–3620, doi: $10.5194/\mathrm{hess}\text{-}16\text{-}3607\text{-}2012$.

Alfieri, L., P. Salamon, F. Pappenberger, F. Wetterhall, and J. Thielen (2012), Operational early warning systems for water-related hazards in Europe, *Environmental Science and Policy*, *21*, 35–49, doi: $10.1016/\mathrm{j.envsci.}$ $2012.01.008$.

Andréassian, V., C. Perrin, C. Michel, I. Usart-Sanchez, and J. Lavabre (2001), Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models, *Journal of Hydrology*, *250*(1-4), 206–223, doi: $10.1016/\mathrm{S}0022\text{-}1694(01)00437\text{-}1$.

Andréassian, V., C. Perrin, and C. Michel (2004), Impact of imperfect potential evapotranspiration knowledge on the efficiency and parameters of watershed models, *Journal of Hydrology*, *286*(1-4), 19–35, doi: $10.1016/$ $\mathrm{j.jhydrol.}2003.09.030$.

Andréassian, V., N. Le Moine, C. Perrin, M. H. Ramos, T. Oudin, T. Mathevet, J. Lerat, and L. Berthet (2012), All that glitters is not gold: The case of calibrating hydrological models, doi: $10.1002/\mathrm{hyp.}9264$.

Archfield, S. A., M. Clark, B. Arheimer, L. E. Hay, H. McMillan, J. E. Kiang, J. Seibert, K. Hakala, A. Bock, T. Wagener, W. H. Farmer, V. Andreassian, S. Attinger, A. Viglione, R. Knight, S. Markstrom, and T. Over (2015), Accelerating advances in continental domain hydrologic modeling, *Water Resources Research*, *51*, 10,078–10,091, doi: $10.1002/2013\mathrm{WR}014049.\mathrm{Received}$.

Arnal, L., M. H. Ramos, E. C. De Perez, H. L. Cloke, E. Stephens, F. Wetterhall, S. J. Van Andel, and F. Pappenberger (2016), Willingness-to-pay for a probabilistic flood forecast: A risk-based decision-making game, *Hydrology and Earth System Sciences*, *20*(8), 3109–3128, doi: $10.5194/\mathrm{hess}\text{-}20\text{-}3109\text{-}2016$.

Arsenault, R., R. Bazile, C. Ouellet Dallaire, and F. Brissette (2016), CANOPEX: A Canadian hydrometeorological watershed database, *Hydrological Processes*, *30*(15), 2734–2736, doi: $10.1002/\mathrm{hyp.}10880$.

Asrar, G. R., J. W. Hurrell, and A. J. Busalacchi (2013), A need for "actionable" climate science and information: Summary of WCRP open science conference, in *Bulletin of the American Meteorological Society*, doi: $10.1175/$ $\mathrm{BAMS}\text{-}\mathrm{D}\text{-}12\text{-}00011.1$.

Bai, P., X. Liu, T. Yang, F. Li, K. Liang, S. Hu, C. Liu, P. Bai, X. Liu, T. Yang, F. Li, K. Liang, S. Hu, and C. Liu (2016), Assessment of the Influences of Different Potential Evapotranspiration Inputs on the Performance of Monthly Hydrological Models under Different Climatic Conditions, *Journal of Hydrometeorology*, *17*(8), 2259–2274, doi: $10.1175/\mathrm{JHM}\text{-}\mathrm{D}\text{-}15\text{-}0202.1$.

Baldauf, M., A. Seifert, J. Förstner, D. Majewski, M. Raschendorfer, and T. Reinhardt (2011), Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities, *Monthly Weather Review*, *139*(12), 3887–3905, doi: $10.1175/\mathrm{MWR\text{-}D\text{-}10\text{-}05013.1}$.

Balsamo, G., A. Beljaars, K. Scipal, P. Viterbo, B. van den Hurk, M. Hirschi, and A. K. Betts (2009), A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the Integrated Forecast System, *Journal of Hydrometeorology*, *10*(3), 623–643, doi: $10.1175/2008\mathrm{JHM}1068.1$.

Balsamo, G., A. Agusti-Panareda, C. Albergel, G. Arduini, A. Beljaars, J. Bidlot, N. Bousserez, S. Boussetta, A. Brown, R. Buizza, C. Buontempo, F. Chevallier, M. Choulga, H. Cloke, M. F. Cronin, M. Dahoui, P. D. Rosnay, P. A. Dirmeyer, M. Drusch, E. Dutra, M. B. Ek, P. Gentine, H. Hewitt, S. P. Keeley, Y. Kerr, S. Kumar, C. Lupu, J. F. Mahfouf, J. McNorton, S. Mecklenburg, K. Mogensen, J. Muñoz-Sabater, R. Orth, F. Rabier, R. Reichle, B. Ruston, F. Pappenberger, I. Sandu, S. I. Seneviratne, S. Tietsche, I. F. Trigo, R. Uijlenhoet, N. Wedi, R. I. Woolway, and X. Zeng (2018), Satellite and in situ observations for advancing global earth surface modelling: A review, doi: $10.3390/\mathrm{rs}10122038$.

Bardossy, A., and J. Caspary (1990), Theoretical and Applied Climatology Detection of Climate Change in Europe by Analyzing European Atmospheric Circulation Patterns from 1881 to 1989, *Theor. Appl. Climatol*, *42*, 155–167.

Bartels, H., E. Weigl, T. Reich, P. Lang, A. Wagner, O. Kohler, and N. Gerlach (2004), Projekt RADOLAN - Routineverfahren zur Online-Aneichung der Radarniederschlagsdaten mit Hilfe von automatischen Bodenniederschlagsstationen (Ombrometer), *Tech. rep.*, MeteoSolutions GmbH.

Bartholmes, J. C., J. Thielen, M. H. Ramos, and S. Gentilini (2009), Hydrology and Earth System Sciences The european flood alert system EFAS-Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts, *Hydrol. Earth Syst. Sci*, *13*, 141–153.

Bastiaanssen, W. G. M., E. J. M. Noordman, H. Pelgrum, G. Davids, B. P. Thoreson, and R. G. Allen (2005), SEBAL Model with Remotely Sensed Data to Improve Water-Resources Management under Actual Field Conditions, *Journal of Irrigation and Drainage Engineering*, *131*(1), 85–93, doi: $10.1061/(\mathrm{ASCE})0733\text{-}9437(2005)131{:}1(85)$.

Bell, V. A., A. L. Kay, R. G. Jones, and R. J. Moore (2007), Development of a high resolution grid-based river flow model for use with regional climate model output, *Hydrology and Earth System Sciences Discussions*, *11*(1), 532–549, doi: $10.5194/\mathrm{hess\text{-}11\text{-}532\text{-}2007}$.

Benning, R. G. (1994), Towards a new lumped parameterization at catchment scale, Ph.D. thesis, Wageningen University, Wageningen, The Netherlands.

Berthet, L., V. Andréassian, C. Perrin, and P. Javelle (2009), How crucial is it to account for the antecedent moisture conditions in flood forecasting? comparison of event-based and continuous approaches on 178 catchments, *Hydrology and Earth System Sciences*, *13*(6), 819–831, doi: $10.5194/\mathrm{hess\text{-}13\text{-}819\text{-}2009}$.

Beven, K. (1989), Changing ideas in hydrology—the case of physically-based models, *Journal of hydrology*, *105*, 157–172.

Beven, K. (2006), Searching for the holy grail of scientific hydrology: Q_t=(s, r, δt)a as closure, *Hydrology and Earth System Sciences*, *10*(5), 609–618, doi: $10.5194/\mathrm{hess\text{-}10\text{-}609\text{-}2006}$.

Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrological Processes*, *6*(3), 279–298, doi: $10.1002/\mathrm{hyp}.3360060305$.

Beven, K., and A. Binley (2014), GLUE: 20 years on, *Hydrological Processes*, *28*(24), 5897–5918, doi: $10.1002/\mathrm{hyp}.10082$.

Beven, K., P. Smith, and J. Freer (2007), Comment on "Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology" by Pietro Mantovan and Ezio Todini, *Journal of Hydrology*, *338*(3-4), 315–318, doi: $10.1016/\mathrm{j.jhydrol}.2007.02.023$.

Beven, K. J., P. J. Smith, and J. E. Freer (2008), So just why would a modeller choose to be incoherent?, *Journal of Hydrology*, *354*(1-4), 15–32, doi: $10.1016/\mathrm{j.jhydrol}.2008.02.007$.

Bierkens, M. F. P., and L. P. H. Van Beek (2009), Seasonal predictability of European discharge: NAO and hydrological response time, *Journal of Hydrometeorology*, *10*(4), 953–968, doi: $10.1175/2009\mathrm{JHM}1034.1$.

Bierkens, M. F. P., V. A. Bell, P. Burek, N. Chaney, L. E. Condon, C. H. David, A. de Roo, P. Döll, N. Drost, J. S. Famiglietti, M. Flörke, D. J. Gochis, P. Houser, R. Hut, J. Keune, S. Kollet, R. M. Maxwell, J. T. Reager, L. Samaniego, E. Sudicky, E. H. Sutanudjaja, N. van de Giesen, H. Winsemius, and E. F. Wood (2015), Hyper-resolution global hydrological modelling: What is next?: "Everywhere and locally relevant", *Hydrological Processes*, *29*(2), 310–320, doi: $10.1002/\mathrm{hyp}.10391$.

Blair, G. S., K. Beven, R. Lamb, R. Bassett, K. Cauwenberghs, B. Hankin, G. Dean, N. Hunter, L. Edwards, V. Nundloll, F. Samreen, W. Simm, and R. Towe (2019), Models of everywhere revisited: A technological perspective, *Environmental Modelling & Software*, p. 104521, doi: $10.1016/\mathrm{j.envsoft}.2019.104521$.

Blenkinsop, S., E. Lewis, S. C. Chan, and H. J. Fowler (2017), Quality-control of an hourly rainfall dataset and climatology of extremes for the UK, *International Journal of Climatology*, *37*(2), 722–740, doi: $10.1002/\mathrm{joc}.4735$.

Blöschl, G., and M. Sivapalan (1995), Scale issues in hydrological modelling: A review, *Hydrological Processes*, *9*(3-4), 251–290, doi: $10.1002/\mathrm{hyp}.3360090305$.

Bourgin, F., M. Ramos, G. Thirel, and V. Andréassian (2014), Investigating the interactions between data assimilation and post-processing in hydrological ensemble forecasting, *Journal of Hydrology*, *519*, 2775–2784, doi: $10.1016/\mathrm{j.jhydrol}.2014.07.054$.

Bowman, A. L., K. J. Franz, T. S. Hogue, and A. M. Kinoshita (2016), MODIS-Based Potential Evapotranspiration Demand Curves for the Sacramento Soil Moisture Accounting Model, *Journal of Hydrologic Engineering*, *21*(1), doi: $10.1061/(\mathrm{ASCE})\mathrm{HE}.1943\text{-}5584.0001261)$.

Bowman, A. L., K. J. Franz, and T. S. Hogue (2017), Case Studies of a MODIS-Based Potential Evapotranspiration Input to the Sacramento Soil Moisture Accounting Model, *Journal of Hydrometeorology*, *18*(1), 151–158, doi: $10.1175/\mathrm{JHM\text{-}D\text{-}16\text{-}0214.1}$.

Brakensiek, D., W. Rawls, and G. Stephenson (1984), Modifying scs hydrologic soil groups and curve numbers for rangeland soils, *American Society of Agricultural Engineers*, *PNR-84-203*.

Brauer, C. (2014), Modelling rainfall-runoff processes in lowland catchments, Ph.D. thesis, Wageningen University.

Broersen, P. M., and A. H. Weerts (2005), Rainfall-Runoff models in Flood Forecasting Systems, in *IEEE Instrumentation and Measurement Technology Conference Proceedings*, pp. 17–19, doi: $10.1109/\mathrm{IMTC}.2005.1604281$.

Brooks, R., and T. Corey (1964), Hydraulic properties of porous media, *Hydrology Papers, Colorado State University*, *24*, doi: $10.13031/2013.40684$.

Brown, J. D., J. Demargne, D. J. Seo, and Y. Liu (2010), The Ensemble Verification System (EVS): A software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations, *Environmental Modelling and Software*, *25*(7), 854–872, doi: $10.1016/\mathrm{j.envsoft}.2010.01.009$.

Buitink, J., L. A. Melsen, J. W. Kirchner, and A. J. Teuling (2019), A distributed simple dynamical systems approach (dS2 v1.0) for computationally efficient hydrological modelling, *Geoscientific Model Development Discussions*, pp. 1–25, doi: $10.5194/\mathrm{gmd\text{-}2019\text{-}150}$.

Buizza, R., J.-R. Bidlot, M. Janousek, S. Keeley, K. Mogensen, and D. Richardson (2017), New IFS cycle brings sea-ice coupling and higher ocean resolution, *ECMWF Newsletter*, *150*, 14–17.

Carrer, D., S. Lafont, J.-L. Roujean, J.-C. Calvet, C. Meurey, P. Le Moigne, and I. F. Trigo (2012), Incoming Solar and Infrared Radiation Derived from METEOSAT: Impact on the Modeled Land Water and Energy Budget over France, *Journal of Hydrometeorology*, *13*(2), 504–520, doi: $10.1175/\mathrm{JHM\text{-}D\text{-}11\text{-}059.1}$.

Chen, C. T., and T. Knutson (2008), On the verification and comparison of extreme rainfall indices from climate models, *Journal of Climate*, *21*(7), 1605–1621, doi: $10.1175/2007\mathrm{JCLI}1494.1$.

Chen, L., W. Lu, J. Zhou, S. Guo, and J. Zhang (2011), Effect of streamflow forecast uncertainty on reservoir operation, *Advances in Water Resources*, *34*, 495–504, doi: $10.13243/\mathrm{j.cnki.slxb}.20150844$.

Clark, M., B. Schaefli, S. Schymanski, L. Samaniego, C. Luce, B. Jackson, and et al. (2016), Improving the theoretical underpinnings of process-based hydrologic models, *Water Resources Research*, *52*(3), 2350–2365, doi: $10.1002/2015\mathrm{WR}017910$.

Clark, M., M. Bierkens, L. Samaniego, R. Woods, R. Uijlenhoet, K. Bennett, and et al. (2017), The evolution of process-based hydrologic models: historical challenges and the collective quest for physical realism, *Hydrology and Earth System Sciences*, *21*(7), 3427–3440, doi: $10.5194/hess-21-3427-2017$.

Clark, M. P., D. E. Rupp, R. A. Woods, X. Zheng, R. P. Ibbitt, A. G. Slater, J. Schmidt, and M. J. Uddstrom (2008), Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model, *Advances in Water Resources*, *31*(10), 1309–1324, doi: $10.1016/j.advwatres.2008.06.005$.

Cloke, H., and F. Pappenberger (2009), Ensemble flood forecasting: A review, *Journal of Hydrology*, *375*(3-4), 613–626, doi: $10.1016/j.jhydrol.2009.06.005$.

Cuo, L., T. C. Pagano, and Q. J. Wang (2011), A Review of Quantitative Precipitation Forecasts and Their Use in Short- to Medium-Range Streamflow Forecasting, *Journal of Hydrometeorology*, *12*(5), 713–728, doi: $10.1175/2011JHM1347.1$.

de Boer-Euser, T. (2016), Added value of distribution in rainfall-runoff models for the Meuse catchment, Ph.D. thesis, TU Delft, doi: $10.4233/UUID:89A78AE9-7FFB-4260-B25D-698854210FA8$.

de Bruin, H. A. R., I. F. Trigo, F. C. Bosveld, and J. F. Meirink (2016), A Thermodynamically Based Model for Actual Evapotranspiration of an Extensive Grass Field Close to FAO Reference, Suitable for Remote Sensing Application, *Journal of Hydrometeorology*, *17*(5), 1373–1382, doi: $10.1175/JHM-D-15-0006.1$.

de Vos, L. W., T. H. Raupach, H. Leijnse, A. Overeem, A. Berne, and R. Uijlenhoet (2018), High-resolution simulation study exploring the potential of radars, crowdsourced personal weather stations, and commercial microwave links to monitor small-scale urban rainfall, *Water Resources Research*, *54*(12), 10,293–10,312, doi: $10.1029/2018WR023393$.

Demargne, J., L. Wu, S. K. Regonda, J. D. Brown, H. Lee, M. He, D. J. Seo, R. Hartman, H. D. Herr, M. Fresch, J. Schaake, and Y. Zhu (2014), The science of NOAA's operational hydrologic ensemble forecast service, *Bulletin of the American Meteorological Society*, *95*(1), 79–98, doi: $10.1175/BAMS-D-12-00081.1$.

Demeritt, D., H. Cloke, F. Pappenberger, J. Thielen, J. Bartholmes, and M. H. Ramos (2007), Ensemble predictions and perceptions of risk, uncertainty, and error in flood forecasting, *Environmental Hazards*, *7*(2), 115–127, doi: $10.1016/j.envhaz.2007.05.001$.

Demirel, M., J. Mai, G. Mendiguren, J. Koch, L. Samaniego, and S. Stisen (2018), Combining satellite data and appropriate objective functions for improved spatial pattern performance of a distributed hydrologic model, *Hydrology and Earth System Sciences*, *22*(2), 1299–1315, doi: $10.5194/hess-22-1299-2018$.

Dorigo, W., W. Wagner, C. Albergel, F. Albrecht, G. Balsamo, L. Brocca, and et al. (2017), Esa cci soil moisture for improved earth system understanding: State-of-the art and future directions, *Remote Sensing of Environment*, doi: $10.1016/j.rse.2017.07.001$.

Dunn, S., and R. Colohan (1999), Developing the snow component of a distributed hydrological model: a step-wise approach based on multi-objective analysis, *Journal of Hydrology*, *223*(1-2), 1–16, doi: $10.1016/S0022-1694(99)00095-5$.

Engman, E. (1986), Roughness coefficients for routing surface runoff, *Journal of Irrigation and Drainage Engineering*, *112*(1), 39–53, doi: $10.1061/(ASCE)0733-9437(1986)112:1(39)$.

ESDAC (2004), The european soil database distribution version 2.0, european commission and the european soil bureau network, Retrieved from esdac.jrc.ec.europa.eu, accessed: 2017-11-17.

European Climate Assessment & Dataset (2017), EOBS Precipitation Station Meta Data.

European Environment Agency (2018), Corine Land Cover (CLC) 2018, Version 20.

Euser, T., M. Hrachowitz, H. C. Winsemius, and H. H. Savenije (2015), The effect of forcing and landscape distribution on performance and consistency of model structures, *Hydrological Processes*, *29*(17), 3727–3743, doi: $10.1002/hyp.10445$.

Fan, F. M., D. Schwanenberg, R. Alvarado, A. A. dos Reis, W. Collischonn, and S. Naumman (2016a), Performance of Deterministic and Probabilistic Hydrological Forecasts for the Short-Term Optimization of a Tropical Hydropower Reservoir, *Water Resources Management*, *30*(10), 3609–3625, doi: $10.1007/s11269-016-1377-8$.

Fan, J., B. McConkey, H. Wang, and H. Janzen (2016b), Root distribution by depth for temperate agricultural crops, *Field Crops Research*, *189*, 68–74, doi: 10.1016/j.fcr.2016.02.013.

Farr, T. G., P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth, D. Seal, S. Shaffer, J. Shimada, J. Umland, M. Werner, M. Oskin, D. Burbank, and D. E. Alsdorf (2007), The shuttle radar topography mission, *Reviews of Geophysics*, *45*(2), RG2004, doi: 10.1029/2005RG000183.

Fatichi, S., E. Vivoni, F. Ogden, V. Ivanov, B. Mirus, D. Gochis, and et al. (2016), An overview of current applications, challenges, and future trends in distributed process-based models in hydrology, *Journal of Hydrology*, *537*, 45–60, doi: 10.1016/j.jhydrol.2016.03.026.

Feddes, R., P. Kowalik, and H. Zaradny (1978), Water uptake by plant roots, *Simulation of field water use and crop yield*, pp. 16–30.

Ficchì, A., C. Perrin, and V. Andréassian (2016), Impact of temporal resolution of inputs on hydrological model performance: An analysis based on 2400 flood events, *Journal of Hydrology*, *538*, 454–470, doi: 10.1016/j.jhydrol.2016.04.016.

Ficchì, A., C. Perrin, and V. Andréassian (2019), Hydrological modelling at multiple sub-daily time steps: Model improvement via flux-matching, *Journal of Hydrology*, *575*, 1308–1327, doi: 10.1016/j.jhydrol.2019.05.084.

Finkl, A., U. Ulbrichl, and H. Engel (1996), Aspects of the January 1995 flood in Germany, *Weather*, *51*(2), 34–39.

Fischer, M., M. Huss, C. Barboux, and M. Hoelzle (2014), The new Swiss Glacier Inventory SGI2010: relevance of using high-resolution source data in areas dominated by very small glaciers, *Arctic, Antarctic, and Alpine Research*, *46*(4), 933–945, doi: 10.1657/1938-4246-46.4.933.

Fowler, A. (2002), Assessment of the validity of using mean potential evaporation in computations of the long-term soil water balance, *Journal of Hydrology*, *256*(3-4), 248–263, doi: 10.1016/S0022-1694(01)00542-X.

Freeze, R. A., and R. Harlan (1969), Blueprint for a physically–based, digitally–simulated hydrologic response model, *Journal of Hydrology*, *9*, 237–258.

Frei, C., R. Schöll, S. Fukutome, J. Schmidli, and P. L. Vidale (2006), Future change of precipitation extremes in Europe: Intercomparison of scenarios from regional climate models, *Journal of Geophysical Research*, *111*(D6), D06,105, doi: 10.1029/2005JD005965.

Gash, J. H. C. (1979), An analytical model of rainfall interception by forests, *Quarterly Journal of the Royal Meteorological Society*, *105*(443), 43–55, doi: 10.1002/qj.49710544304.

Georgakakos, K. P. (1986), A generalized stochastic hydrometeorological model for flood and flash-flood forecasting: 2. Case studies, *Water Resources Research*, *22*(13), 2096–2106, doi: 10.1029/WR022i013p02096.

Gijsbers, P. J. A., J. H. Baayen, and G. J. ter Maat (2017), Quick Scan Tool for Water Allocation in the Netherlands, in *Environmental Software Systems. Computer Science for Environmental Protection*, edited by P. T. Hřebíček J., Denzer R., Schimak G., pp. 97–109, Springer, Cham, doi: 10.1007/978-3-319-89935-0_9.

Görgen, K., J. Beersma, G. Brahmer, H. Buiteveld, M. Carambia, O. de Keizer, and et al. (2010), *Assessment of climate change impacts on discharge in the Rhine River Basin: results of the RheinBlick2050 project*, Lelystad, Netherlands: Internationale Kommission für die Hydrologie des Rheingebietes.

Gregersen, J. B., P. J. Gijsbers, and S. J. Westen (2007), OpenMI: Open modelling interface, *Journal of Hydroinformatics*, *9*(3), 175–191, doi: 10.2166/hydro.2007.023.

Grinsted, A. (2013), An estimate of global glacier volume, *The Cryosphere*, *7*(1), 141–151, doi: 10.5194/tc-7-141-2013.

Guo, M., J. Li, C. Sheng, J. Xu, and L. Wu (2017), A review of wetland remote sensing, doi: 10.3390/s17040777.

Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez (2009), Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *Journal of Hydrology*, *377*(1-2), 80–91, doi: 10.1016/j.jhydrol.2009.08.003.

Hamlet, A. F., D. Huppert, and D. P. Lettenmaier (2002), Economic Value of Long-Lead Streamflow Forecasts for Columbia River Hydropower, *Journal of Water Resources Planning and Management*, *128*(2), 91–101, doi: 10.

$1061/(\mathrm{asce})0733\text{-}9496(2002)128\text{:}2(91)$.

Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, and M. New (2008), A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, *Journal of Geophysical Research*, *113*(D20), D20,119, doi: $10.1029/2008\mathrm{JD}010201$.

Hazeleger, W., B. J. Van Den Hurk, E. Min, G. J. Van Oldenborgh, A. C. Petersen, D. A. Stainforth, E. Vasileiadou, and L. A. Smith (2015), Tales of future weather, *Nature Climate Change*, *5*(2), 107–113, doi: $10.1038/\mathrm{nclimate}2450$.

Hazenberg, P. (2013), Rainfall estimation for hydrology using volumetric weather radar, Ph.D. thesis, Wageningen University.

Hengl, T., J. de Jesus, G. Heuvelink, M. Gonzalez, M. Kilibarda, A. Blagotić, and et al. (2017), Soilgrids250m: Global gridded soil information based on machine learning, *PloS one*, *12*(2), e0169,748, doi: $10.1371/\mathrm{journal.pone.}0169748$.

Hiemstra, P., and R. Sluiter (2011), Interpolation of Makkink evaporation in the Netherlands, *Tech. rep.*, KNMI, De Bilt.

Hock, R. (2003), Temperature index melt modelling in mountain areas, *Journal of Hydrology*, *282*(1-4), 104–115, doi: $10.1016/\mathrm{S}0022\text{-}1694(03)00257\text{-}9$.

Houtekamer, P. L., and H. L. Mitchell (2001), A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation, *Monthly Weather Review*, *129*(1), 123–137, doi: $10.1175/1520\text{-}0493(2001)129{<}0123\text{:}\mathrm{ASEKFF}{>}2.0.\mathrm{CO}\text{;}2$.

Hrachowitz, M., H. Savenije, G. Blöschl, J. McDonnell, M. Sivapalan, J. Pomeroy, B. Arheimer, T. Blume, M. P. Clark, U. Ehret, F. Fenicia, J. E. Freer, a. Gelfan, H. Gupta, D. Hughes, R. Hut, a. Montanari, S. Pande, D. Tetzlaff, P. a. Troch, S. Uhlenbrook, T. Wagener, H. Winsemius, R. Woods, E. Zehe, and C. Cudennec (2013), A decade of Predictions in Ungauged Basins (PUB)—a review, *Hydrological Sciences Journal*, *58*(6), 1198–1255, doi: $10.1080/02626667.2013.803183$.

Hurkmans, R., H. de Moel, J. Aerts, and P. Troch (2008), Water balance versus land surface model in the simulation of rhine river discharges, *Water Resources Research*, *44*(1), W01,418, doi: $10.1029/2007\mathrm{WR}006168$.

Hurkmans, R., W. Terink, R. Uijlenhoet, P. Torfs, D. Jacob, and P. A. Troch (2009a), Changes in Streamflow Dynamics in the Rhine Basin under Three High-Resolution Regional Climate Scenarios, *Journal of Climate*, *23*, 679–699, doi: $10.1175/2009\mathrm{JCLI}3066.1$.

Hurkmans, R., W. Terink, R. Uijlenhoet, E. Moors, P. Troch, and P. Verburg (2009b), Effects of land use changes on streamflow generation in the rhine basin, *Water Resources Research*, *45*(6), W06,405, doi: $10.1029/2008\mathrm{WR}007574$.

Hutton, C., T. Wagener, J. Freer, D. Han, C. Duffy, and B. Arheimer (2016), Most computational hydrology is not reproducible, so is it really science?, *Water Resources Research*, *52*(10), 7548–7555, doi: $10.1002/2016\mathrm{WR}019285$.

Immerzeel, W. W., and P. Droogers (2008), Calibration of a distributed hydrological model based on satellite evapotranspiration, *Journal of Hydrology*, *349*(3-4), 411–424, doi: $10.1016/\mathrm{j.jhydrol.}2007.11.017$.

Ineichen, P., C. S. Barroso, B. Geiger, R. Hollmann, A. Marsouin, and R. Mueller (2009), Satellite Application Facilities irradiance products: Hourly time step comparison and validation over Europe, *International Journal of Remote Sensing*, *30*(21), 5549–5571, doi: $10.1080/01431160802680560$.

Isotta, F. A., C. Frei, V. Weilguni, M. Perčec Tadić, P. Lassègues, B. Rudolf, V. Pavan, C. Cacciamani, G. Antolini, S. M. Ratto, M. Munari, S. Micheletti, V. Bonati, C. Lussana, C. Ronchi, E. Panettieri, G. Marigo, and G. Vertačnik (2014), The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data, *International Journal of Climatology*, *34*(5), 1657–1675, doi: $10.1002/\mathrm{joc.}3794$.

Jacobs, J. M., B. Lowry, M. Choi, and C. H. Bolster (2009), GOES Solar Radiation for Evapotranspiration Estimation and Streamflow Prediction, *Journal of Hydrologic Engineering*, *14*(3), 293–300, doi: $10.1061/(\mathrm{ASCE})1084\text{-}0699(2009)14\text{:}3(293)$.

Jaun, S., and B. Ahrens (2009), Evaluation of a probabilistic hydrometeorological forecast system, *Hydrology and Earth System Sciences*, *13*(7), 1031–1043, doi: $10.5194/\mathrm{hess}\text{-}13\text{-}1031\text{-}2009$.

Jedrzej, J., S. Bojanowski, A. Vrieling, and A. K. Skidmore (2014), A comparison of data sources for creating a long-term time series of daily gridded solar radiation for Europe, *Solar Energy*, *99*, 152–171, doi: $10.1016/\mathrm{j}.$ $\mathrm{solener}.2013.11.007$.

Jörg-Hess, S., F. Fundel, T. Jonas, and M. Zappa (2014), Homogenisation of a gridded snow water equivalent climatology for Alpine terrain: methodology and applications, *The Cryosphere*, *8*(2), 471–485, doi: $10.5194/$ $\mathrm{tc}\text{-}8\text{-}471\text{-}2014$.

Journée, M., and C. Bertrand (2010), Improving the spatio-temporal distribution of surface solar radiation data by merging ground and satellite measurements, *Remote Sensing of Environment*, *114*(11), 2692–2704, doi: $10.$ $1016/\mathrm{j}.\mathrm{rse}.2010.06.010$.

Kalweit, H., W. Buck, K. Felkel, and H. Gerhard (1993), *Der Rhein unter der Einwirkung des Menschen : Ausbau, Schiffahrt, Wasserwirtschaft*, KHR, Lelystad, The Netherlands.

Karssenberg, D., O. Schmitz, P. Salamon, K. de Jong, and M. Bierkens (2010), A software framework for construction of process-based stochastic spatio-temporal models and data assimilation, *Environmental Modelling & Software*, *25*(4), 489–502, doi: $10.1016/\mathrm{j}.\mathrm{envsoft}.2009.10.004$.

Katz, R. W. (2010), Statistics of extremes in climate change, *Climatic Change*, *100*(1), 71–76, doi: $10.1007/$ $\mathrm{s}10584\text{-}010\text{-}9834\text{-}5$.

Katz, R. W., M. B. Parlange, and P. Naveau (2002), Statistics of extremes in hydrology, *Advances in Water Resources*, *25*(8), 1287 − 1304, doi: $\mathrm{https://doi.org/10.1016/S0309\text{-}1708(02)00056\text{-}8}$.

Kilgore, J. (1997), Development and evaluation of a gis-based spatially distributed unit hydrograph model, Retrieved from http://hdl.handle.net/10919/35777.

Kim, J., and T. S. Hogue (2008), Evaluation of a MODIS-Based Potential Evapotranspiration Product at the Point Scale, *Journal of Hydrometeorology*, *9*(3), 444–460, doi: $10.1175/2007\mathrm{JHM}902.1$.

Kirchner, J. W. (2006), Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, *Water Resources Research*, *42*(3), doi: $10.1029/2005\mathrm{WR}004362$.

Klemeš, V. (1983), Conceptualization and scale in hydrology, *Journal of Hydrology*, *65*(1-3), 1–23, doi: $10.1016/$ $0022\text{-}1694(83)90208\text{-}1$.

Klemeš, V. (1986), Operational testing of hydrological simulation models, *Hydrological Sciences Journal*, *31*(1), 13–24, doi: $10.1080/02626668609491024$.

Koch, J., M. Demirel, and S. Stisen (2018), The spatial efficiency metric (spaef): multiple-component evaluation of spatial patterns for optimization of hydrological models, *Geoscientific Model Development*, *11*(5), 1873–1886, doi: $10.5194/\mathrm{gmd}\text{-}11\text{-}1873\text{-}2018$.

Koster, R. D., S. P. Mahanama, B. Livneh, D. P. Lettenmaier, and R. H. Reichle (2010), Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow, *Nature Geoscience*, *3*(9), 613–616, doi: $10.1038/\mathrm{ngeo}944$.

Krysanova, V., T. Vetter, S. Eisner, S. Huang, I. Pechlivanidis, M. Strauch, and et al. (2017), Intercomparison of regional-scale hydrological models and climate change impacts projected for 12 large river basins worldwide—a synthesis, *Environmental Research Letters*, *12*(10), 105,002, doi: $10.1088/1748\text{-}9326/\mathrm{aa}8359$.

Kumar, R., L. Samaniego, and S. Attinger (2013a), Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations, *Water Resources Research*, *49*(1), 360–379, doi: $10.1029/$ $2012\mathrm{WR}012195$.

Kumar, R., B. Livneh, and L. Samaniego (2013b), Toward computationally efficient large-scale hydrologic predictions with a multiscale regionalization scheme, *Water Resources Research*, *49*(9), 5700–5714, doi: $10.1002/$ $\mathrm{wrcr}.20431$.

Kunnath-Poovakka, A., D. Ryu, L. Renzullo, and B. George (2016), The efficacy of calibrating hydrologic model using remotely sensed evapotranspiration and soil moisture for streamflow prediction, *Journal of Hydrology*,

535, 509–524, doi: $10.1016/j.jhydrol.2016.02.018$.

Kuusisto, E. (1984), *Snow accumulation and snowmelt in Finland*, chap. 5, pp. 60–105, Vesihallitus, National Board of Waters.

Lerch, S., T. L. Thorarinsdottir, F. Ravazzolo, and T. Gneiting (2015), Forecaster's Dilemma: Extreme Events and Forecast Evaluation, *Statistical Science*, *32*(1), 106–127, doi: $10.1214/16\text{-}STS588$.

Lettenmaier, D. P., D. Alsdorf, J. Dozier, G. J. Huffman, M. Pan, and E. F. Wood (2015), Inroads of remote sensing into hydrologic science during the WRR era, *Water Resources Research*, *51*(9), 7309–7342, doi: $10.1002/2015WR017616$.

Li, Y., D. Ryu, A. W. Western, and Q. J. Wang (2013), Assimilation of stream discharge for flood forecasting: The benefits of accounting for routing time lags, *Water Resources Research*, *49*(4), 1887–1900, doi: $10.1002/wrcr.20169$.

Li, Y., S. Grimaldi, J. P. Walker, and V. R. Pauwels (2016), Application of remote sensing data to constrain operational rainfall-driven flood forecasting: A review, doi: $10.3390/rs8060456$.

Liguori, S., and M. A. Rico-Ramirez (2014), A review of current approaches to radar-based quantitative precipitation forecasts, *International Journal of River Basin Management*, *12*(4), 391–402, doi: $10.1080/15715124.2013.848872$.

Lindström, G., B. Johansson, M. Persson, M. Gardelin, and S. Bergström (1997a), Development and test of the distributed HBV-96 hydrological model, *Journal of Hydrology*, *201*(1-4), 272–288, doi: $10.1016/S0022\text{-}1694(97)00041\text{-}3$.

Lindström, G., B. Johansson, M. Persson, M. Gardelin, and S. Bergström (1997b), Development and test of the distributed HBV-96 hydrological model, *Journal of Hydrology*, *201*(1-4), 272–288, doi: $10.1016/S0022\text{-}1694(97)00041\text{-}3$.

Liu, S. (1998), Estimation of rainfall storage capacity in the canopies of cypress wetlands and slash pine uplands in north-central florida, *Journal of Hydrology*, *207*(1-2), 32–41, doi: $10.1016/S0022\text{-}1694(98)00115\text{-}2$.

Liu, Y., A. H. Weerts, M. Clark, H.-J. Hendricks Franssen, S. Kumar, H. Moradkhani, D.-J. Seo, D. Schwanenberg, P. Smith, A. I. J. M. van Dijk, N. van Velzen, M. He, H. Lee, S. J. Noh, O. Rakovec, and P. Restrepo (2012), Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, *Hydrology and Earth System Sciences*, *16*(10), 3863–3887, doi: $10.5194/hess\text{-}16\text{-}3863\text{-}2012$.

Liu, Z., M. L. V. Martina, and E. Todini (2005), Flood forecasting using a fully distributed model: application of the TOPKAPI model to the Upper Xixian Catchment, *Hydrology and Earth System Sciences*, *9*(4), 347–364, doi: $10.5194/hess\text{-}9\text{-}347\text{-}2005$.

Lobligeois, F., V. Andréassian, C. Perrin, P. Tabary, and C. Loumagne (2014), When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events, *Hydrol. Earth Syst. Sci*, *18*, 575–594, doi: $10.5194/hess\text{-}18\text{-}575\text{-}2014$.

Luo, Y., J. Arnold, S. Liu, X. Wang, and X. Chen (2013), Inclusion of glacier processes for distributed hydrological modeling at basin scale with application to a watershed in Tianshan Mountains, northwest China, *Journal of Hydrology*, *477*, 72–85, doi: $10.1016/j.jhydrol.2012.11.005$.

Ly, S., C. Charles, and A. Degré (2013), Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale. A review, *B S Biotechnol. Agron. Soc. Environ.*, *17*(2), 392–406.

Makkink, G. F. (1957), EKZAMENO DE LA FORMULO DE PENMAN, *Neth. J. Agri. Sci.*, *5*, 290–305.

Mantovan, P., and E. Todini (2006), Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology, *Journal of Hydrology*, *330*(1-2), 368–381, doi: $10.1016/j.jhydrol.2006.04.046$.

Mantovan, P., E. Todini, and M. L. Martina (2007), Reply to comment by Keith Beven, Paul Smith and Jim Freer on "Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology", *Journal of Hydrology*, *338*(3-4), 319–324, doi: $10.1016/j.jhydrol.2007.02.029$.

Mazzoleni, M., S. J. Noh, H. Lee, Y. Liu, D. J. Seo, A. Amaranto, L. Alfonso, and D. P. Solomatine (2018), Real-time

assimilation of streamflow observations into a hydrological routing model: effects of model structures and updating methods, *Hydrological Sciences Journal*, *63*(3), 386–407, doi: $10.1080/02626667.2018.1430898$.

McCabe, M. F., M. Rodell, D. E. Alsdorf, D. G. Miralles, R. Uijlenhoet, W. Wagner, A. Lucieer, R. Houborg, N. E. Verhoest, T. E. Franz, J. Shi, H. Gao, and E. F. Wood (2017), The future of Earth observation in hydrology, *Hydrology and Earth System Sciences*, *21*(7), 3879–3914, doi: $10.5194/hess-21-3879-2017$.

Meißner, D., and B. Klein (2016), Probabilistic Shipping Forecast, in *Handbook of hydrometeorological ensemble forecasting*, pp. 1371–1384, Springer-Verlag, doi: $10.1007/978-3-642-40457-3\_58-1$.

Meißner, D., B. Klein, and M. Ionita (2017), Development of a monthly to seasonal forecast framework tailored to inland waterway transport in central Europe, *Hydrology and Earth System Sciences*, *21*(12), 6401–6423, doi: $10.5194/hess-21-6401-2017$.

Melsen, L., A. Teuling, P. Torfs, M. Zappa, N. Mizukami, M. Clark, and R. Uijlenhoet (2016), Representation of spatial and temporal variability in large-domain hydrological models: case study for a mesoscale pre-Alpine basin, *Hydrology and Earth System Sciences*, *20*(6), 2207–2226, doi: $10.5194/hess-20-2207-2016$.

Merz, R., and G. Blöschl (2004), Regionalisation of catchment model parameters, *Journal of Hydrology*, *287*(1-4), 95–123, doi: $10.1016/j.jhydrol.2003.09.028$.

Middelkoop, H., K. Daamen, D. Gellens, W. Grabs, J. C. J. Kwadijk, H. Lang, B. W. a. H. Parmet, B. Schädler, J. Schulla, and K. Wilke (2001), Impact of climate change on hydrological regimes and water resources management in the Rhine basin, *Climatic Change*, *49*(1-2), 105–128, doi: $10.1023/A:1010784727448$.

Mizukami, N., M. Clark, A. Newman, A. Wood, E. Gutmann, B. Nijssen, and et al. (2017), Towards seamless large-domain parameter estimation for hydrologic models, *Water Resources Research*, *53*(9), 8020–8040, doi: $10.1002/2017WR020401$.

Montanari, a., G. Young, H. Savenije, D. Hughes, T. Wagener, L. Ren, D. Koutsoyiannis, C. Cudennec, E. Toth, S. Grimaldi, G. Blöschl, M. Sivapalan, K. Beven, H. Gupta, M. Hipsey, B. Schaefli, B. Arheimer, E. Boegh, S. Schymanski, G. Di Baldassarre, B. Yu, P. Hubert, Y. Huang, a. Schumann, D. Post, V. Srinivasan, C. Harman, S. Thompson, M. Rogger, a. Viglione, H. McMillan, G. Characklis, Z. Pang, and V. Belyaev (2013), "Panta Rhei—Everything Flows": Change in hydrology and society—The IAHS Scientific Decade 2013–2022, *Hydrological Sciences Journal*, *58*(6), 1256–1275, doi: $10.1080/02626667.2013.809088$.

Mu, Q., M. Zhao, and S. W. Running (2011), Improvements to a MODIS global terrestrial evapotranspiration algorithm, *Remote Sensing of Environment*, *115*(8), 1781–1800, doi: $10.1016/j.rse.2011.02.019$.

Mueller, R., C. Matsoukas, A. Gratzki, H. Behr, and R. Hollmann (2009), The CM-SAF operational scheme for the satellite based retrieval of solar surface irradiance — A LUT based eigenvector hybrid approach, *Remote Sensing of Environment*, *113*(5), 1012–1024, doi: $10.1016/J.RSE.2009.01.012$.

Myneni, R., Y. Knyazikhin, and T. Park (2015), MCD15A3H MODIS/Terra+Aqua Leaf Area Index/FPAR 4-day L4 Global 500m SIN Grid V006, *Tech. rep.*, NASA EOSDIS Land Processes DAAC.

Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part I: A discussion of principles, *Journal of Hydrology*, *10*, 282–290, doi: $10.1016/0022-1694(70)90255-6$.

Newman, A. J., M. P. Clark, J. Craig, B. Nijssen, A. Wood, E. Gutmann, N. Mizukami, L. Brekke, and J. R. Arnold (2015), Gridded ensemble precipitation and temperature estimates for the contiguous United States, *Journal of Hydrometeorology*, *16*(6), 2481–2500, doi: $10.1175/JHM-D-15-0026.1$.

Nijzink, R. C., L. Samaniego, J. Mai, R. Kumar, S. Thober, M. Zink, D. Schäfer, H. H. G. Savenije, and M. Hrachowitz (2016), The importance of topography-controlled sub-grid process heterogeneity and semi-quantitative prior constraints in distributed hydrological models, *Hydrology and Earth System Sciences*, *20*(3), 1151–1176, doi: $10.5194/hess-20-1151-2016$.

Niu, G.-Y., Z.-L. Yang, K. E. Mitchell, F. Chen, M. B. Ek, M. Barlage, A. Kumar, K. Manning, D. Niyogi, E. Rosero, et al. (2011), The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, *Journal of Geophysical Research: Atmospheres*, *116*(D12), doi: $10.1029/2010JD015140$.

Ntegeka, V., P. Salamon, G. Gomes, H. Sint, M. Lorini, M. Zambrano-Bigiarini, and J. Thielen (2013), *EFAS-Meteo: A*

European daily high-resolution gridded meteorological data set for 1990-2011, Publications Office of the European Union, Luxembourg, doi: 10.2788/51262.

OpenDA Assiociation (2016), OpenDA User Documentation.

Oudin, L., F. Hervieu, C. Michel, C. Perrin, V. Andréassian, F. Anctil, and C. Loumagne (2005a), Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2 - Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling, *Journal of Hydrology*, *303*(1-4), 290–306, doi: 10.1016/j.jhydrol.2004.08.026.

Oudin, L., C. Michel, and F. Anctil (2005b), Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 1 - Can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs?, *Journal of Hydrology*, *303*(1-4), 275–289, doi: 10.1016/j.jhydrol.2004.08.025.

Pagano, T. C., A. W. Wood, M.-H. Ramos, H. L. Cloke, F. Pappenberger, M. P. Clark, M. Cranston, D. Kavetski, T. Mathevet, S. Sorooshian, and J. S. Verkade (2014), Challenges of Operational River Forecasting, *Journal of Hydrometeorology*, *15*(4), 1692–1707, doi: 10.1175/jhm-d-13-0188.1.

Pagano, T. C., F. Pappenberger, A. W. Wood, M.-H. Ramos, A. Persson, and B. Anderson (2016), Automation and human expertise in operational river forecasting, *Wiley Interdisciplinary Reviews: Water*, *3*(5), 692–705, doi: 10.1002/wat2.1163.

Palmer, M. A. (2012), Socioenvironmental Sustainability and Actionable Science, *BioScience*, *62*(1), 5–6, doi: 10.1525/bio.2012.62.1.2.

Panagos, P., M. Van Liedekerke, A. Jones, and L. Montanarella (2012), European soil data centre: Response to european policy support and public data requirements, *Land Use Policy*, *29*(2), 329–338, doi: 10.1016/j.landusepol.2011.07.003.

Paniconi, C., and M. Putti (2015), Physically based modeling in catchment hydrology at 50: Survey and outlook, *Water Resources Research*, *51*(9), 7090–7129, doi: 10.1002/2015WR017780.

Pappenberger, F., K. J. Beven, N. M. Hunter, P. D. Bates, B. T. Gouweleeuw, J. Thielen, and A. P. J. De Roo (2005), Cascading model uncertainty from medium range weather forecasts (10 days) within the EFFS Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the European Flood F, *Hydrology and Earth System Sciences*, *9*(4), 381–393.

Pappenberger, F., M. H. Ramos, H. L. Cloke, F. Wetterhall, L. Alfieri, K. Bogner, A. Mueller, and P. Salamon (2015), How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction, *Journal of Hydrology*, *522*, 697–713, doi: 10.1016/j.jhydrol.2015.01.024.

Parmele, L. H. (1972), Errors in output of hydrologic models due to errors in input potential evapotranspiration, *Water Resources Research*, *8*(2), 348–359, doi: 10.1029/WR008i002p00348.

Peng, J., A. Loew, O. Merlin, and N. E. Verhoest (2017), A review of spatial downscaling of satellite remotely sensed soil moisture, *Reviews of Geophysics*, *55*(2), 341–366, doi: 10.1002/2016RG000543.

Peters, H. P., D. Brossard, S. De Cheveigné, S. Dunwoody, M. Kallfass, S. Miller, and S. Tsuchida (2008), Science communication: Interactions with the mass media, *Science*, *321*(5886), 204–205, doi: 10.1126/science.1157780.

Peterson, T. C., D. R. Easterling, T. R. Karl, P. Groisman, N. Nicholls, N. Plummer, S. Torok, I. Auer, R. Boehm, D. Gullett, L. Vincent, R. Heino, H. Tuomenvirta, O. Mestre, T. Szentimrey, J. Salinger, E. J. Førland, I. Hanssen-Bauer, H. Alexandersson, P. Jones, and D. Parker (1998), Homogeneity adjustments of in situ atmospheric climate data: a review, *International Journal of Climatology*, *18*(13), 1493–1517, doi: 10.1002/(SICI)1097-0088(19981115)18:13<1493::AID-JOC329>3.0.CO;2-T.

Photiadou, C., B. van den Hurk, A. van Delden, and A. Weerts (2016), Incorporating circulation statistics in bias correction of GCM ensembles: hydrological application for the Rhine basin, *Climate Dynamics*, *46*(1-2), 187–203, doi: 10.1007/s00382-015-2578-1.

Photiadou, C. S., A. H. Weerts, and B. J. J. M. van den Hurk (2011), Evaluation of two precipitation data sets for the Rhine River using streamflow simulations, *Hydrology and Earth System Sciences*, *15*(11), 3355–3366, doi: 10.5194/hess-15-3355-2011.

Pitman, J. (1989), Rainfall interception by bracken in open habitats—relations between leaf area, canopy storage and drainage rate, *Journal of Hydrology*, *105*(3-4), 317–334, doi: $10.1016/0022\text{-}1694(89)90111\text{-}X$.

Rakovec, O., A. H. Weerts, P. Hazenberg, P. J. J. F. Torfs, and R. Uijlenhoet (2012), State updating of a distributed hydrological model with Ensemble Kalman Filtering: effects of updating frequency and observation network density on forecast accuracy, *Hydrology and Earth System Sciences*, *16*(9), 3435–3449, doi: $10.5194/$ $\text{hess}\text{-}16\text{-}3435\text{-}2012$.

Rakovec, O., A. H. Weerts, J. Sumihar, and R. Uijlenhoet (2015), Operational aspects of asynchronous filtering for flood forecasting, *Hydrology and Earth System Sciences*, *19*(6), 2911–2924, doi: $10.5194/$ $\text{hess}\text{-}19\text{-}2911\text{-}2015$.

Rakovec, O., R. Kumar, S. Attinger, and L. Samaniego (2016a), Improving the realism of hydrologic model functioning through multivariate parameter estimation, *Water Resources Research*, *52*(10), 7779–7792, doi: $10.1002/2016WR019430$.

Rakovec, O., R. Kumar, J. Mai, M. Cuntz, S. Thober, M. Zink, S. Attinger, D. Schäfer, M. Schrön, and L. Samaniego (2016b), Multiscale and multivariate evaluation of water fluxes and states over european river basins, *Journal of Hydrometeorology*, *19*(2), doi: $10.1175/JHM\text{-}D\text{-}15\text{-}0054.1$.

Ramos, M. H., T. Mathevet, J. Thielen, and F. Pappenberger (2010), Communicating uncertainty in hydrometeorological forecasts: Mission impossible?, *Meteorological Applications*, *17*(2), 223–235, doi: $10.1002/$ $\text{met}.202$.

Rasmussen, J., H. Madsen, K. H. Jensen, and J. C. Refsgaard (2015), Data assimilation in integrated hydrological modeling using ensemble Kalman filtering: evaluating the effect of ensemble size and localization on filter performance, *Hydrology and Earth System Sciences*, *19*(7), 2999–3013, doi: $10.5194/\text{hess}\text{-}19\text{-}2999\text{-}2015$.

Raup, B. H., A. Racoviteanu, S. J. S. Khalsa, C. Helm, R. Armstrong, and Y. Arnaud (2007), The GLIMS Geospatial Glacier Database: a new tool for studying glacier change, *Global and Planetary Change*, *56*, 101–110, doi: $10.$ $1016/\text{j.gloplacha}.2006.07.018$.

Rauthe, M., H. Steiner, U. Riediger, A. Mazurkiewicz, and A. Gratzki (2013), A Central European precipitation climatology – Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS), *Meteorologische Zeitschrift*, *22*(3), 235–256, doi: $10.1127/0941\text{-}2948/2013/0436$.

Rawls, W. J., and D. L. Brakensiek (1989), Estimation of soil water retention and hydraulic properties, in *Unsaturated flow in hydrologic modelling - Theory and practice*, edited by H. J. Morel-Seytoux, pp. 275–300, Kluwer Academic Publishing, Dordrecht, The Netherlands, doi: $10.1007/978\text{-}94\text{-}009\text{-}2352\text{-}2\_10$.

Reggiani, P., and A. Weerts (2008), A Bayesian approach to decision-making under uncertainty: An application to real-time forecasting in the river Rhine, *Journal of Hydrology*, *356*(1-2), 56–69, doi: $10.1016/\text{j.jhydrol}.$ $2008.03.027$.

Reggiani, P., M. Renner, A. H. Weerts, and P. A. H. J. M. van Gelder (2009), Uncertainty assessment via Bayesian revision of ensemble streamflow predictions in the operational river Rhine forecasting system, *Water Resources Research*, *45*(2), n/a–n/a, doi: $10.1029/2007WR006758$.

Reichle, R. H. (2008), Data assimilation methods in the Earth sciences, *Advances in Water Resources*, *31*(11), 1411–1418, doi: $10.1016/\text{j.advwatres}.2008.01.001$.

Renner, M., M. G. F. Werner, S. Rademacher, and E. Sprokkereef (2009), Verification of ensemble flow forecasts for the River Rhine, *Journal of Hydrology*, *376*(3-4), 463–475, doi: $10.1016/\text{j.jhydrol}.2009.07.059$.

RGI Consortium (2017), Randolph Glacier Inventory – A Dataset of Global Glacier Outlines: Version 6.0, *Tech. rep.*, Global Land Ice Measurements from Space, CO, USA, doi: $10.7265/N5\text{-}RGI\text{-}60$.

Ridler, M. E., N. van Velzen, S. Hummel, I. Sandholt, A. K. Falk, A. W. Heemink, and H. Madsen (2014), Data assimilation framework: Linking an open data assimilation library (OpenDA) to a widely adopted model interface (OpenMI), *Environmental Modelling and Software*, *57*, 76–89, doi: $10.1016/\text{j.envsoft}.2014.02.008$.

Rijkswaterstaat (n.d.), Watermanagementcentrum nederland, `https://en.wikipedia.org/w/index.php?title=LaTeX&oldid=413720397`, [Online; accessed 28-November-2019].

Rios Gaona, M. F. (2017), Rainfall over the Netherlands & beyond: a remote sensing perspective, Ph.D. thesis,

Wageningen University, doi: $10.18174/414112$.

Rios Gaona, M. F., A. Overeem, H. Leijnse, and R. Uijlenhoet (2015), Measurement and interpolation uncertainties in rainfall maps from cellular communication networks, *Hydrology and Earth System Sciences*, *19*(8), 3571–3584, doi: $10.5194/\text{hess-}19\text{-}3571\text{-}2015$.

Rios Gaona, M. F., A. Overeem, H. Leijnse, and R. Uijlenhoet (2016), First-year evaluation of GPM rainfall over the Netherlands: IMERG day 1 final run (V03D), *Journal of Hydrometeorology*, *17*(11), 2799–2814, doi: $10.1175/\text{JHM-D-}16\text{-}0087.1$.

Rutter, A. J., K. A. Kershaw, P. C. Robins, and A. J. Morton (1971), A predictive model of rainfall interception in forests, 1. derivation of the model from observations in a plantation of corsican pine, *Agricultural Meteorology*, *9*, 367–384, doi: $10.1016/0002\text{-}1571(71)90034\text{-}3$.

Rutter, A. J., A. J. Morton, and P. C. Robins (1975), A predictive model of rainfall interception in forests. ii. generalization of the model and comparison with observations in some coniferous and hardwood stands, *Journal of Applied Ecology*, pp. 367–380, doi: $10.2307/2401739$.

Sakov, P., G. Evensen, and L. Bertino (2010), Asynchronous data assimilation with the EnKF, *Tellus A: Dynamic Meteorology and Oceanography*, *62*(1), 24–29, doi: $10.1111/\text{j.}1600\text{-}0870.2009.00417.\text{x}$.

Samaniego, L., R. Kumar, and S. Attinger (2010), Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, *Water Resources Research*, *46*(5), n/a–n/a, doi: $10.1029/2008WR007327$.

Samaniego, L., R. Kumar, S. Thober, O. Rakovec, M. Zink, N. Wanders, and et al. (2017), Toward seamless hydrologic predictions across spatial scales, *Hydrology and Earth System Sciences*, *21*(9), 4323–4346, doi: $10.5194/\text{hess-}21\text{-}4323\text{-}2017$.

Schaake, J., S. Cong, and Q. Duan (2006), The US MOPEX Data Set, in *Large Sample Basin Experiments for Hydrological Model Parameterization: Results of the Model Parameter Experiment–MOPEX*, pp. 9–28, IAHS Publ., doi: $10.1080/13241583.2007.11465316$.

Schellekens, J. (2016), OpenStreams wflow documentation.

Schellekens, J., W. van Verseveld, T. de Boer-Euser, H. Winsemius, C. Thiange, L. Bouaziz, D. Tollenaar, S. de Vries, and A. Weerts (2017), Openstreams wflow documentation.

Schellekens, J., W. van Verseveld, T. de Boer-Euser, H. Winsemius, C. Thiange, L. Bouaziz, D. Tollenaar, S. de Vries, and A. Weerts (2018), Openstreams wflow documentation.

Schellekens, J., W. van Verseveld, T. de Boer-Euser, H. Winsemius, C. Thiange, L. Bouaziz, D. Tollenaar, S. de Vries, and A. Weerts (2019a), openstreams/wflow: unstable-master.

Schellekens, J., W. Verseveld van, T. Euser, H. Winsemius, C. Thiange, L. Bouaziz, and et al. (2019b), Openstreams/wflow, Retrieved from https://github.com/openstreams/wflow.

Schenk, H. J., and R. B. Jackson (2002), The global biogeography of roots, *Ecological monographs*, *72*(3), 311–328, doi: $10.1890/0012\text{-}9615(2002)072[0311\text{:TGBOR}]2.0.\text{CO};2$.

Seibert, J. (1999), Regionalisation of parameters for a conceptual rainfall-runoff model, *Agricultural and forest meteorology*, *98*, 279–293, doi: $10.1016/\text{S}0168\text{-}1923(99)00105\text{-}7$.

Semádeni-Davies, A. (1997), Monthly snowmelt modelling for large-scale climate change studies using the degree day approach, *Ecological Modelling*, *101*(2-3), 303–323, doi: $10.1016/\text{S}0304\text{-}3800(97)00054\text{-}9$.

Seo, D. J., L. Cajina, R. Corby, and T. Howieson (2009), Automatic state updating for operational streamflow forecasting via variational data assimilation, *Journal of Hydrology*, *367*(3-4), 255–275, doi: $10.1016/\text{j.jhydrol.}2009.01.019$.

Seuffert, G., P. Gross, C. Simmer, and E. F. Wood (2002), The Influence of Hydrologic Modeling on the Predicted Local Weather: Two-Way Coupling of a Mesoscale Weather Prediction Model and a Land Surface Hydrologic Model, *Journal of Hydrometeorology*, *3*(5), 505–523, doi: $10.1175/1525\text{-}7541(2002)003<0505\text{:TIOHMO}>2.0.\text{CO};2$.

Sivapalan, M., G. Blöschl, L. Zhang, and R. Vertessy (2003), Downward approach to hydrological prediction, *Hydrological processes*, *17*(11), 2101–2111, doi: $10.1002/\text{hyp.}1425$.

Spies, R. R., K. J. Franz, T. S. Hogue, and A. L. Bowman (2015), Distributed Hydrologic Modeling Using Satellite-Derived Potential Evapotranspiration, *Journal of Hydrometeorology*, *16*(1), 129–146, doi: $10.1175/$ $\mathrm{JHM\text{-}D\text{-}14\text{-}0047.1}$.

Steiner, M., J. A. Smith, S. J. Burges, C. V. Alonso, and R. W. Darden (1999), Effect of bias adjustment and rain gauge data quality control on radar rainfall estimation, *Water Resources Research*, *35*(8), 2487–2503, doi: $10.$ $1029/1999\mathrm{WR}900142$.

Stisen, S., M. McCabe, J. Refsgaard, S. Lerer, and M. Butts (2011), Model parameter analysis using remotely sensed pattern information in a multi-constraint framework, *Journal of Hydrology*, *409*(1), 337–349, doi: $10.1016/\mathrm{j.jhydrol.2011.08.030}$.

Su, Z. (2002), The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes, *Hydrology and Earth System Sciences*, *6*(1), 85–100, doi: $10.5194/\mathrm{hess\text{-}6\text{-}85\text{-}2002}$.

Sun, L., O. Seidou, I. Nistor, and K. Liu (2016), Review of the Kalman-type hydrological data assimilation, *Hydrological Sciences Journal*, *61*(13), 2348–2366, doi: $10.1080/02626667.2015.1127376$.

Tan, J., W. A. Petersen, and P.-E. Kirstetter (2017), Performance of IMERG as a Function of Spatiotemporal Scale, *Journal of Hydrometeorology*, *18*(2), 307–319, doi: $10.1175/\mathrm{JHM\text{-}D\text{-}16\text{-}0174.1}$.

Tao, J., D. Wu, J. Gourley, S. Q. Zhang, W. Crow, C. Peters-Lidard, and A. P. Barros (2016), Operational hydrological forecasting during the IPHEx-IOP campaign — Meet the challenge, *Journal of Hydrology*, *541*, 434–456, doi: $10.1016/\mathrm{j.jhydrol.2016.02.019}$.

Te Linde, A., J. Aerts, A. Bakker, and J. Kwadijk (2010), Simulating low-probability peak discharges for the rhine basin using resampled climate modeling data, *Water Resources Research*, *46*(3), doi: $10.1029/$ $2009\mathrm{WR}007707$.

Terink, W., R. T. W. L. Hurkmans, P. J. J. F. Torfs, and R. Uijlenhoet (2010), Evaluation of a bias correction method applied to downscaled precipitation and temperature reanalysis data for the Rhine basin, *Hydrology and Earth System Sciences*, *14*(4), 687–703, doi: $10.5194/\mathrm{hess\text{-}14\text{-}687\text{-}2010}$.

Terti, G., I. Ruin, S. Anquetin, and J. J. Gourley (2015), Dynamic vulnerability factors for impact-based flash flood prediction, *Natural Hazards*, *79*(3), 1481–1497, doi: $10.1007/\mathrm{s11069\text{-}015\text{-}1910\text{-}8}$.

Thakur, J. K., S. K. Singh, and V. S. Ekanthalu (2017), Integrating remote sensing, geographic information systems and global positioning system techniques with hydrological modeling, *Applied Water Science*, *7*(4), 1595–1608, doi: $10.1007/\mathrm{s13201\text{-}016\text{-}0384\text{-}5}$.

Thiboult, A., and F. Anctil (2015), On the difficulty to optimally implement the Ensemble Kalman filter: An experiment based on many hydrological models and catchments, *Journal of Hydrology*, *529*, 1147–1160, doi: $10.1016/\mathrm{j.jhydrol.2015.09.036}$.

Tobin, C., L. Nicotina, M. B. Parlange, A. Berne, and A. Rinaldo (2011), Improved interpolation of meteorological forcings for hydrologic applications in a Swiss Alpine region, *Journal of Hydrology*, *401*(1-2), 77–89, doi: $10.$ $1016/\mathrm{j.jhydrol.2011.02.010}$.

Tobin, C., B. Schaefli, L. Nicótina, S. Simoni, G. Barrenetxea, R. Smith, M. Parlange, and A. Rinaldo (2013), Improving the degree-day method for sub-daily melt simulations with physically-based diurnal variations, *Advances in Water Resources*, *55*, 149–164, doi: $10.1016/\mathrm{j.advwatres.2012.08.008}$.

Todini, E., and L. Ciarapica (2002), *Mathematical models of large watershed hydrology*, chap. The TOPKAPI model, pp. 471–506, Water Resources Publications, Littleton, Colorado.

Tóth, B., M. Weynants, A. Nemes, A. Makó, G. Bilas, and G. Tóth (2015), New generation of hydraulic pedotransfer functions for europe, *European journal of soil science*, *66*, 226–238, doi: $10.1111/\mathrm{ejss.12192}$.

Trigo, I. F., C. C. Dacamara, P. Viterbo, J.-L. Roujean, F. Olesen, C. Barroso, F. Camacho-de Coca, D. Carrer, S. C. Freitas, J. García-Haro, B. Geiger, F. Gellens-Meulenberghs, N. Ghilain, J. Meliá, L. Pessanha, N. Siljamo, and A. Arboleda (2011), The Satellite Application Facility for Land Surface Analysis, *International Journal of Remote Sensing*, *32*(10), 2725–2744, doi: $10.1080/01431161003743199$.

Urraca, R., T. Huld, A. Gracia-Amillo, F. J. Martinez-de Pison, F. Kaspar, and A. Sanz-Garcia (2018), Evaluation of global horizontal irradiance estimates from ERA5 and COSMO-REA6 reanalyses using ground and satellite-

based data, *Solar Energy*, *164*, 339–354, doi: $10.1016/j.solener.2018.02.059$.

van Andel, S. J., A. Weerts, J. Schaake, and K. Bogner (2013), Post-processing hydrological ensemble predictions intercomparison experiment, *Hydrological Processes*, *27*(1), 158–161, doi: $10.1002/hyp.9595$.

van Beek, L. P. H., Y. Wada, and M. F. P. Bierkens (2011), Global monthly water stress: 1. Water balance and water availability, *Water Resources Research*, *47*(7), doi: $10.1029/2010WR009792$.

van Bemmelen, J. M. (1890), Über die Bestimmung des Wassers, des Humus, des Schwefels, der in den colloïdalen Silikaten gebundenen Kieselsäure, des Mangans u. s. w. im Ackerboden, *Die Landwirthschaftlichen Versuchs-Stationen*, *37*, 279–290.

van de Beek, C., H. Leijnse, P. Torfs, and R. Uijlenhoet (2012), Seasonal semi-variance of Dutch rainfall at hourly to daily scales, *Advances in Water Resources*, *45*, 76–85, doi: $10.1016/j.advwatres.2012.03.023$.

van den Bergh, J., and E. Roulin (2010), Hydrological ensemble prediction and verification for the Meuse and Scheldt basins, *Atmospheric Science Letters*, *11*(2), 64–71, doi: $10.1002/asl.250$.

van den Hurk, B. J., L. M. Bouwer, C. Buontempo, R. Döscher, E. Ercin, C. Hananel, J. E. Hunink, E. Kjellström, B. Klein, M. Manez, F. Pappenberger, L. Pouget, M.-H. Ramos, P. J. Ward, A. H. Weerts, and J. B. Wijngaard (2016), Improving predictions and management of hydrological extremes through climate services: www.imprex.eu, *Climate Services*, *1*, 6–11, doi: $10.1016/j.cliser.2016.01.001$.

van den Hurk, B. J. J. M., P. Viterbo, A. C. M. Beljaars, and A. K. Betts (2000), *ECMWF Technical Memoranda*, chap. Offline validation of the ERA40 surface scheme, 295, European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom, doi: $10.21957/9aoaspz8$.

van Dijk, A., and L. Bruijnzeel (2001), Modelling rainfall interception by vegetation of variable density using an adapted analytical model. part 2. model validation for a tropical upland mixed cropping system, *Journal of Hydrology*, *247*(3-4), 239–262.

Van Griensven, A., T. Meixner, S. Grunwald, T. Bishop, M. Diluzio, and R. Srinivasan (2006), A global sensitivity analysis tool for the parameters of multi-variable catchment models, *Journal of Hydrology*, *324*(1), 10–23, doi: $10.1016/j.jhydrol.2005.09.008$.

van Looy, K., J. Bouma, M. Herbst, J. Koestel, B. Minasny, U. Mishra, C. Montzka, A. Nemes, Y. A. Pachepsky, J. Padarian, et al. (2017), Pedotransfer functions in earth system science: challenges and perspectives, *Reviews of Geophysics*, *55*(4), 1199–1256, doi: $10.1002/2017RG000581$.

van Osnabrugge, B. (2017), Gridded precipitation dataset for the Rhine basin made with the genRE interpolation method, doi: $10.4121/uuid:c875b385-ef6d-45a5-a6d3-d5fe5e3f525d$.

van Osnabrugge, B. (2018), Gridded Hourly Temperature, Radiation and Makkink Potential Evaporation forcing for hydrological modelling in the Rhine basin, doi: $10.4121/uuid:e036030f-c73b-4e7b-9bd4-eebc899b5a13$.

van Osnabrugge, B., A. H. Weerts, and R. Uijlenhoet (2017), genRE: A Method to Extend Gridded Precipitation Climatology Data Sets in Near Real-Time for Hydrological Forecasting Purposes, *Water Resources Research*, *53*(November), doi: $10.1002/2017WR021201$.

van Osnabrugge, B., R. Uijlenhoet, and A. Weerts (2019), Contribution of potential evaporation forecasts to 10-day streamflow forecast skill for the Rhine River, *Hydrology and Earth System Sciences*, *23*(3), doi: $10.5194/hess-23-1453-2019$.

Verkade, J., J. Brown, P. Reggiani, and A. Weerts (2013), Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales, *Journal of Hydrology*, *501*, 73–91, doi: $10.1016/j.jhydrol.2013.07.039$.

Verkade, J., J. Brown, F. Davids, P. Reggiani, and A. Weerts (2017), Estimating predictive hydrological uncertainty by dressing deterministic and ensemble forecasts; a comparison, with application to meuse and rhine, *Journal of Hydrology*, *555*, 257–277, doi: $10.1016/j.jhydrol.2017.10.024$.

Verkade, J. S. (2015), Estimating Real-Time Predictive Hydrological Uncertainty, Ph.D. thesis, TU Delft.

Vertessy, R., and H. Elsenbeer (1999), Distributed modeling of storm flow generation in an amazonian rain forest

catchment: Effects of model parameterization, *Water Resources Research*, *35*(7), 2173–2187, doi: $10.1029/1999WR900051$.

Viterbo, P., and A. C. M. Beljaars (1995), An improved land surface parameterization scheme in the ECMWF model and its validation, *Journal of Climate*, *8*(11), 2716–2748, doi: $10.1175/1520\text{-}0442(1995)008<2716:AILSPS>2.0.CO;2$.

Vivoni, E. R., G. Mascaro, S. Mniszewski, P. Fasel, E. P. Springer, V. Y. Ivanov, and R. L. Bras (2011), Real-world hydrologic assessment of a fully-distributed hydrological model in a parallel computing environment, *Journal of Hydrology*, *409*(1-2), 483–496, doi: $10.1016/j.jhydrol.2011.08.053$.

Vorogushyn, S., and B. Merz (2013), Flood trends along the Rhine: The role of river training, *Hydrology and Earth System Sciences*, *17*(10), 3871–3884, doi: $10.5194/hess\text{-}17\text{-}3871\text{-}2013$.

Vrugt, J. a., C. G. H. Diks, H. V. Gupta, W. Bouten, and J. M. Verstraten (2005), Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resources Research*, *41*, 1–17, doi: $10.1029/2004WR003059$.

Vrugt, J. a., C. J. F. ter Braak, M. P. Clark, J. M. Hyman, and B. a. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resources Research*, *44*, 1–52, doi: $10.1029/2007WR006720$.

Wagener, T., H. Wheater, and H. V. Gupta (2004), *Rainfall-runoff modelling in gauged and ungauged catchments*, Imperial College Press.

Wagener, T., M. Sivapalan, P. A. Troch, B. L. McGlynn, C. J. Harman, H. V. Gupta, P. Kumar, P. S. C. Rao, N. B. Basu, and J. S. Wilson (2010), The future of hydrology: An evolving science for a changing world, *Water Resources Research*, *46*(5), doi: $10.1029/2009WR008906$.

Wagner, B., V. Tarnawski, V. Hennings, U. Müller, G. Wessolek, and R. Plagge (2001), Evaluation of pedo-transfer functions for unsaturated soil hydraulic conductivity using an independent data set, *Geoderma*, *102*(3), 275–297, doi: $10.1016/S0016\text{-}7061(01)00037\text{-}4$.

Wanders, N., D. Karssenberg, a. De Roo, S. M. De Jong, and M. F. P. Bierkens (2014), The suitability of remotely sensed soil moisture for improving operational flood forecasting, *Hydrology and Earth System Sciences*, *18*, 2343–2357, doi: $10.5194/hess\text{-}18\text{-}2343\text{-}2014$.

Wanders, N., S. Thober, R. Kumar, M. Pan, J. Sheffield, L. Samaniego, and E. F. Wood (2019), Development and evaluation of a pan-european multimodel seasonal hydrological forecasting system, *Journal of Hydrometeorology*, *20*(1), 99–115, doi: $10.1175/JHM\text{-}D\text{-}18\text{-}0040.1$.

Weerts, A., D. Meissner, and S. Rademacher (2008), Input data rainfall-runoff model operational system FEWS-NL & FEWS-DE, *Tech. rep.*, Deltares, Delft.

Weerts, A. H., and G. Y. H. El Serafy (2006), Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, *Water Resources Research*, *42*(9), 1–17, doi: $10.1029/2005WR004093$.

Weijs, S. V., G. Schoups, and N. Van De Giesen (2010), Why hydrological predictions should be evaluated using information theory, *Hydrology and Earth System Sciences*, *14*(12), 2545–2558, doi: $10.5194/hess\text{-}14\text{-}2545\text{-}2010$.

Werner, M., J. Schellekens, P. Gijsbers, M. van Dijk, O. van den Akker, and K. Heynert (2013), The Delft-FEWS flow forecasting system, *Environmental Modelling & Software*, *40*, 65–77, doi: $10.1016/j.envsoft.2012.07.010$.

Winsemius, H. M., W. van Verseveld, A. H. Weerts, and M. Hegnauer (2013a), Generalised Likelihood Uncertainty Estimation for the daily HBV model in the Rhine Basin, Part A: Germany, *Tech. rep.*, Deltares, Delft.

Winsemius, H. M., W. van Verseveld, A. H. Weerts, and M. Hegnauer (2013b), Generalised Likelihood Uncertainty Estimation for the daily HBV model in the Rhine Basin, Part B: Switzerland, *Tech. rep.*, Deltares, Delft.

Wit, M., and T. Buishand (2007), *Generator of rainfall and discharge extremes (GRADE) for the Rhine and Meuse basins*, Rijkswaterstaat/RIZA.

WMO (2009), Guide to Hydrological Practices, Volume II: Management of Water Resources and Application of

Hydrological Practices, 6th edition.

Wood, E. F., J. K. Roundy, T. J. Troy, L. P. H. van Beek, M. F. P. Bierkens, E. Blyth, A. de Roo, P. Döll, M. Ek, J. Famiglietti, D. Gochis, N. van de Giesen, P. Houser, P. R. Jaffé, S. Kollet, B. Lehner, D. P. Lettenmaier, C. Peters-Lidard, M. Sivapalan, J. Sheffield, A. Wade, and P. Whitehead (2011), Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, *Water Resources Research*, *47*(5), W05,301, doi: $10.1029/2010WR010090$.

Wösten, J. H. M. (1997), Pedotransfer functions to evaluate soil quality, *Developments in Soil Sciences*, *25*, 221–245, doi: $10.1016/S0166\text{-}2481(97)80037\text{-}2$.

Xie, X., S. Meng, S. Liang, and Y. Yao (2014), Improving streamflow predictions at ungauged locations with real-time updating: Application of an EnKF-based state-parameter estimation strategy, *Hydrology and Earth System Sciences*, *18*(10), 3923–3936, doi: $10.5194/hess\text{-}18\text{-}3923\text{-}2014$.

Xu, C.-Y., and V. P. Singh (2002), Cross Comparison of Empirical Equations for Calculating Potential Evapotranspiration with Data from Switzerland, *Water Resources Management*, *16*(August), 197–219, doi: $10.1023/A: 1020282515975$.

Xystrakis, F., and A. Matzarakis (2011), Evaluation of 13 Empirical Reference Potential Evapotranspiration Equations on the Island of Crete in Southern Greece, *Journal of Irrigation and Drainage Engineering*, *137*(4), 211–222, doi: $10.1061/(ASCE)IR.1943\text{-}4774.0000283$.

Yalew, S., A. van Griensven, N. Ray, L. Kokoszkiewicz, and G. D. Betrie (2013), Distributed computation of large scale SWAT models on the Grid, *Environmental Modelling and Software*, *41*, 223–230, doi: $10.1016/j.envsoft. 2012.08.002$.

Yuan, X., E. F. Wood, and Z. Ma (2015), A review on climate-model-based seasonal hydrologic forecasting: physical understanding and system development, *Wiley Interdisciplinary Reviews: Water*, *2*(5), 523–536, doi: $10.1002/wat2.1088$.

Zhang, H., H.-J. Hendricks Franssen, X. Han, J. A. Vrugt, and H. Vereecken (2017), State and parameter estimation of two land surface models using the ensemble kalman filter and the particle filter, *Hydrology and Earth System Sciences*, *21*(9), 4927–4958, doi: $10.5194/hess\text{-}21\text{-}4927\text{-}2017$.

Zhang, Y., and D. S. Oliver (2011), Evaluation and error analysis: Kalman gain regularization versus covariance regularization, *Computational Geosciences*, *15*(3), 489–508, doi: $10.1007/s10596\text{-}010\text{-}9218\text{-}y$.

Zhu, J., and B. Mohanty (2002), Spatial averaging of van genuchten hydraulic parameters for steady-state flow in heterogeneous soils, *Vadose Zone Journal*, *1*(2), 261–272, doi: $10.2136/vzj2002.2610$.

List of publications

Peer-reviewed publications

- van Osnabrugge, B., Weerts, A.H., Uijlenhoet, R., genRE: A Method to Extend Gridded Precipitation Climatology Data Sets in Near Real-Time for Hydrological Forecasting Purposes (2017), *Water Resources Research*, 53(11), 9284-9303, doi: 10.1002/2017WR021201

- van Osnabrugge, B., R. Uijlenhoet, and A. H. Weerts (2019), Contribution of potential evaporation forecasts to 10-day streamflow forecast skill for the Rhine River, *Hydrology and Earth System Sciences*, 23(3), doi: 10.5194/hess-23-1453-2019

- Imhoff, R. O., W. J. van Verseveld, B. van Osnabrugge, and A. H. Weerts (2019), Scaling point-scale (pedo)transfer functions to seamless large-domain parameter estimates for high-resolution distributed hydrologic modelling: An example for the Rhine river, *submitted to Water Resources Research*

- van Osnabrugge, B., R. Uijlenhoet, and A. H. Weerts (2019), Assimilation of multiple lake levels in an operational integrated catchment model of the Swiss Rhine basin, *submitted to Journal of Hydrology*

Conference abstracts

- van Osnabrugge, B., Weerts, A. H., and Uijlenhoet, R., Gridded precipitation fields at high temporal and spatial resolution for operational flood forecasting in the Rhine basin (2017), EGU General Assembly, 24–28 Apr., Vienna, Austria

- Davids, F., van Osnabrugge, B., den Toom, M., and Verkade, J. S., Verification of operational ECMWF-EPS and COSMO-LEPS precipitation forecasts for the Rhine catchment area (2017), EGU General Assembly, 24–28 Apr., Vienna, Austria

- Weerts, A. H., van Osnabrugge, B., and van Verseveld, W. J., Improving hydrological modelling/predictions for the Rhine River in the context of the IMPREX project (2018), EGU General Assembly, 09–13 Apr., Vienna, Austria

- van Osnabrugge, B., Uijlenhoet, R. , and Weerts, A. H., A comprehensive high resolution hydrometeorological forcing dataset for the Rhine river basin for hydrologic modelling experiments (2018), EGU General Assembly, 09–13 Apr., Vienna, Austria

- van Osnabrugge, B., Weerts, A. H., and Uijlenhoet, R., Contribution of potential evaporation forecasts to 10-day hydrological streamflow forecast skill, EGU General Assembly, 09–13 Apr., Vienna, Austria

- van Osnabrugge, B., Smoorenburg, M., Uijlenhoet, R. , and Weerts, A. H., Assimilating water levels of 8 Swiss lakes into a distributed hydrological model to improve low flow forecasts for the river Rhine at Basel with the AEnKF and wflow-OpenDA (2019), EGU General Assembly, 08–12 Apr., Vienna, Austria

- van Verseveld, W. J., Imhoff, R. O., van Osnabrugge, B., and Weerts, A. H., Scaling point-scale pedotransfer functions to seamless large-domain parameter estimates for high-resolution distributed hydrological modelling: An example for the Rhine river (2019), EGU General Assembly, 08–12 Apr., Vienna, Austria

A digital version of this thesis is available at edepot.wur.nl

Statement of authorship contribution

The introduction and synthesis chapter have been completely written by Bart van Osnabrugge. Bart van Osnabrugge was the main author of all other chapters, with the exception of Chapter 5. The work in Chapter 5 is mainly the result of the extended MSc thesis work of Ruben Imhoff, who is also the first author of that chapter. Other contributions can be found below:

AW	=	Albrecht Weerts (WUR,Deltares)	RI	=	Ruben Imhoff (WUR,Deltares)
BO	=	Bart van Osnabrugge (WUR,Deltares)	MS	=	Maarten Smoorenburg (Deltares)
ES	=	Erik Sprokkereef (RWS)	WV	=	Willem van Verseveld (Deltares)
RU	=	Remko Uijlenhoet (WUR)			

For Chapter 3, ES aided in collecting gauge data by contacting data providers. For Chapter 4, AW contributed by preprocessing the satellite data used in that study. For both Chapters 3 and 4 AW and RU provided critical reviews of the article.

Chapter 5 is the work of multiple persons. BO and RI started the research based on initial ideas from the research proposal for this thesis. RI then proceeded setting up the model and performing the analysis, with aid from AW, WV and BO. BO performed a critical review of the first draft of the paper and model results, which led to a revision of the model code of the hydrological model. After the paper was first rejected, RI performed the analysis requested by the reviewers and RI and BO wrote the version that is currently under review, with aid from AW and WV.

For Chapter 6, AW supported BO in setting up the openDA software. For Chapter 7, MS aided in the generation of the data by adjusting the set-up of the software to run for multiple basins and in parallel. As with the earlier chapters, AW and RU provided critical feedback and guidance.

Statement of code and data availability

Input and evaluation data

The data used in this study has been gratefully received from a large number of organisations:

- Meteorological data for this research has been gratefully received from the Deutscher Wetterdienst Climate Data Center; KNMI Data Centrum; Météo France; Federal Office of Meteorology and Climatology MeteoSwiss; Administration de la gestion de l'eau du Grand-Duché de Luxembourg; and Service Publique de Wallonie Département des Etudes et de l'Appui à la Gestion.
- Discharge data has been gratefully received from SCHAPI (Service Central d'Hydro météorologie et d'Appui à la Prévision des Inondations) through 'Banque HYDRO'; Bundesamt für Umwelt BAFU; Bundesanstalt für Gewässerkunde (BfG); Administration de la gestion de l'eau du Grand-Duché de Luxembourg; Bavarian Environment Agency, www.lfu.bayern.de; Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg, LUBW; Landesamtes für Umwelt, Wasserwirtschaft und Gewerbeaufsicht Rheinland-Pfalz; Landesamt für Umwelt- und Arbeitsschutz Saarland; and Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen.
- Data for the set-up of the wflow_sbm model in Chapter 5 is described in that Chapter and references therein.

Software and models

- the wflow framework and wflow_sbm model code can be downloaded from: https://github.com/openstreams/wflow
- openDA can be downloaded from: https://github.com/OpenDA-Association/OpenDA
- Delft-FEWS is available on: https://oss.deltares.nl/web/delft-fews/home

Output data

Generated gridded hourly forcing data is available through the 4TU Data Centre:
- van Osnabrugge, B. (2017), Gridded precipitation dataset for the Rhine basin made with the genRE interpolation method, doi: 10.4121/uuid:c875b385-ef6d-45a5-a6d3-d5fe5e3f525d.
- van Osnabrugge, B. (2018), Gridded Hourly Temperature, Radiation and Makkink Potential Evaporation forcing for hydrological modelling in the Rhine basin, doi: 10.4121/uuid:e036030f-c73b-4e7b-9bd4-eebc899b5a13.

Summary

Hydrological forecasts are a useful and cost-effective tool to aid decision making. Hydrological forecasts are based on a set-up consisting of several model and data components, which need to be integrated for an effective forecast. Part of this model train is the hydrological model. Regarding hydrological models, there is a current trend to move towards high-resolution spatially distributed gridded models, both in the wider literature and in the specific case of the forecasting system RWSosRivers of Rijkswaterstaat. Concurrently, the forecasting models are used for ensemble forecasts and are further improved with data assimilation techniques.

A number of constraints particular to the operational forecasting chain are identified:

- Near real-time availability of data;
- Continuous operation;
- Processing time;
- Desicion making under uncertainty.

The combination of the movement towards gridded models, ensemble forecasts and the inclusion of data assimilation leads to new challenges, conceptually, but even more so when intersected with operational practice. The challenges that arise on the intersection of hydrological science disciplines and the constraints above are called in this thesis *operational aspects*.

The thesis consists of five research chapters that each contribute original research on operational aspects of hydrological forecasts, with the Rhine basin as case study.

In **Chapter 3**, the operational aspect studied was precipitation interpolation under the constraint of limited data availability. Specifically, limited network density of hourly real-time reporting gauges and the constraint of reliability in the sense that any interpolation method should also work in case of missing data. Collected rain gauge data were spatially interpolated using the genRE method, which uses climatological grids to inform the interpolation about the spatial distribution of precipitation. It was shown that improved hourly interpolation results could be obtained with the operational interpolation method by using climatological grids which were derived from non-operationally available daily data with a higher quality and quantity, thereby successfully using non real-time data to enhance the information available in real-time.

The subject of **Chapter 4** was the treatment of potential evaporation (PET) in hydrological forecasting. The operational aspect that was studied was the choice between using average PET climatology based on a long time series of offline data, or the use of near real-time available data to calculate PET in near real-time, including calculating PET from forecast data for use in hydrological forecasts. A 20-year reforecast was done and the resulting skill scores with online and offline PET data were compared. It was shown that for our case there was a negligible difference in discharge forecast skill between using the offline or online PET data for forecasts up to 10 days.

Chapter 5 reported the set-up of the wflow_sbm model concept without calibration. Parameters for the model were derived based on pedotransfer functions found in literature

and open access spatial data, such as soil properties. The model parameters were subsequently scaled from the highest data resolution to several coarser model resolutions. It was shown that this resulted in adequate modelling results of discharge throughout the Rhine basin, as well as flux conservation between modelling on different spatial scales. The scalability and uncalibrated properties of the model fit the use in an operational context very well.

Last, Chapters 6 and 7 showed the results of two experiments with state updating. **Chapter 6** employed the Ensemble Kalman Filter (EnKF) to investigate if measured lake water levels could be used to improve downstream discharge forecasts. The results showed that indeed state updating with lake level measurements can aid in making discharge forecasts, but that the robustness of this state updating is highly dependent on the quality of the lake modelling.

Chapter 7 applied the Asynchronous Ensemble Kalman Filter (AEnKF) to assimilate discharge into the gridded hydrological model at the subbasin level. The results of the assimilation were compared with ARMA postprocessing. It was shown that the ARMA correction was very strong for the first two days, but that the AEnKF provided more consistent improvements for longer lead times.

Based on these results, specific advice is given to improve the forecasting system for the Rhine. It is recommended to:

· Apply the methods used in this thesis for the spatial preprocessing of the forcing data and to keep an up-to-date database containing all the data that is needed for a reforecast analysis;

· Implement the data assimilation set-up as used in Chapter 7, with the annotation that this method can be further optimized for individual subbasins;

· Make the decision to base further developments in the hydrological model on the wflow model framework, under the vision of 'choose a modelling framework first, and hydrological model concept later';

· Invest in the set-up of an automated benchmarking system.

This advice fits in a broader picture on the future of state-of-the-art forecasting systems. Such forecasting systems are likely to:

· Be modular in set-up with interchangeable components;

· Have integrated uncertainty estimation, with coupled models of hydrology, meteorology and other environmental models;

· Include the user as integral part of the forecasting chain and include social elements in the forecast;

· Apply improved integration of commensurable data sources throughout and up-and-down the forecasting chain;

· Be multi-purpose and cater to a wide range of clients with different needs;

· Be automated learning systems that utilize standard benchmarks to continuously improve modelling;

· Be operated by consortia of parties that have the necessary expertise and resources.

Samenvatting

Prognoses (forecasts) van hydrologische variabelen zijn een bruikbare en kosteffectieve manier om betere beslissingen te nemen omtrent watermanagement. Zulke hydrologische verwachtingen worden gemaakt aan de hand van een serie van verschillende componenten die samen moeten werken om een voorspelling te creëren. Onderdeel van deze trein aan modellen is het hydrologische model. Op het gebied van hydrologisch modelleren wordt er steeds meer gewerkt met ruimtelijk gedistribueerde modellen. Ook voor het operationele systeem voor afvoervoorspellingen voor de Rijn en de Maas, RWsOS Rivieren, worden zulke modellen op grid basis overwogen. Tegelijkertijd worden dergelijke hoge resolutie modellen ook gebruikt voor ensemble verwachtingen en gecombineerd met data-assimilatie methodes.

Daarnaast is er nog een aantal specifieke randvoorwaarden waar rekening mee gehouden moet worden wanneer het gaat om operationele hydrologische voorspellingen:

- · Onmiddelijke beschikbaarheid van de benodigde datastromen;
- · De systemen zijn continu in gebruik;
- · De beschikbare verwerkingstijd is beperkt;
- · Beslissingen moeten worden genomen op basis van onzekere verwachtingen.

De combinatie van ruimtelijk explicite modellen en het gebruik van ensembles en data-assimilatie leidt tot nieuwe uitdagingen, vooral in combinatie met de specifieke randvoorwaarden behorende bij operationele voorspellingen. In dit proefschrift wordt een aantal van deze *operationele aspecten*, ofwel de uitdagingen die ontstaan op het raakvlak tussen bestaande theorie en de genoemde randvoorwaarden, onderzocht.

In **Hoofdstuk 3** wordt neerslaginterpolatie onder de loep genomen in het geval van beperkte near-real time beschikbaarheid van neerslaggegevens. Specifiek wordt onderzocht hoe een betrouwbare ruimtelijke verdeling van de neerslag kan worden bepaald op basis van een beperkte dataset die in near-real time beschikbaar is. Dit wordt gedaan op basis van klimatologische neerslagverdelingen die zijn afgeleid van een grotere, niet in real-time beschikbare, dataset. De resultaten laten zien dat de geteste genRE methode succesvol informatie gebruikt over de ruimtelijk verdeling van de neerslag uit de klimatologische neerslagkaarten en zo de ruimtelijke interpolatie verbetert.

Hoofdstuk 4 kijkt naar het effect dat verschillende potentiëleverdampingsproducten (PET) hebben op de hydrogische voorspelling. Het operationele aspect dat hiermee wordt onderzocht is de keuze tussen het gebruik van een PET klimatologie, gebaseerd op een lange tijdreeks van historische PET, en het uitrekenen van PET in near-real time op basis van actuele gegevens en weersverwachtingen. Op basis van een twintigjarige reforecast zijn skill scores uitgerekend voor beide gevallen. Hieruit blijkt dat er voor een tiendaagse voorspelling nauwelijks verschil is tussen de twee methodes.

Hoofdstuk 5 beschrijft een methode om het hydrologische wflow_sbm model te parameterizen op basis van pedotransferfuncties uit de literatuur en publiekelijk beschikbare ruimtelijke gegevens zoals bodemeigenschappen. De modelparameters worden geschaald van de hoogste dataresolutie naar verschillende lagere resolutie waarop de modelberekeningen worden uitgevoerd. De resultaten laten zien dat de voorgestelde methode zorgt voor redelijk tot goede prestaties op het gebied van het voorspellen van

afvoeren. Ook blijven de hydrologische fluxen, zoals verdamping, behouden wanneer er wordt gewisseld tussen verschillende ruimtelijke modelresoluties. Deze schaalbaarheid en het feit dat bij deze methode niet gekalibreerd wordt, zijn eigenschappen die voordelig zijn in een operationele context.

In **Hoofdstuk 6** wordt het Ensemble Kalman Filter (EnKF) gebruikt om te onderzoeken of het assimileren van waterstandsmetingen van de Zwitserse meren bijdraagt aan betere afvoervoorspellingen. De resultaten laten zien dat waterstandsmetingen van meren op deze manier kunnen bijdragen aan betere voorspellingen, maar dat het resultaat sterk afhangt van hoe goed de afvoeren van de meren worden gemodelleerd.

Hoofdstuk 7 gebruikt het Asynchrone Ensemble Kalman Filter (AEnKF) om afvoermetingen te assimileren in het gedistribueerde model voor de Rijn op het niveau van deelstroomgebieden. De data-assimilatie wordt vergeleken met de ARMA-nabewerkingsmethode. Resultaten laten zien dat ARMA zeer goede resultaten geeft voor de eerste twee dagen van de verwachting, maar dat data-assimilatie meer consistente verbeteringen geeft voor langere tijd vooruit.

Naast de wetenschappelijke relevantie op het gebied van operationeel voorspellen van afvoeren, geven de resultaten uit deze hoofdstukken ook aanleiding tot specifieke adviezen over het verbeteren van hydrologische prognoses voor het Rijnstroomgebied. Het wordt aangeraden om:

· de methodes die gebruikt zijn in dit proefschrift over te nemen en om daarvan een continu geactualiseerde database bij te houden samen met (een selectie van) alle andere data die binnenkomt in het operationele systeem;

· de data-assimilatie toe te passen zoals gebruikt in Hoofdstuk 7, met de kanttekening dat de gebruikte implementatie nog verder geoptimaliseerd kan worden voor de verschillende deelstroomgebieden;

· verdere ontwikkelingen aan het hydrologische model te baseren op de hydrologische modelleeromgeving wflow, op basis van de visie 'kies eerst een framework, en dan pas een concept;'

· te investeren in een workflow voor automatische benchmarking van resultaten voor validatie en om verdere ontwikkelingen te bevorderen.

Dit advies past in een bredere visie op hoe state-of-the-art systemen voor hydrologische voorspellingen er uit zullen zien in de toekomst. Toekomstige systemen:

· zijn modulair qua opzet, met uitwisselbare componenten;

· hanteren geïntegreerde methodes om de onzekerheid van voorspellingen te bepalen, in combinatie met gekoppelde hydrologische, meteorologische en ecologische modellen;

· houden rekening met de gebruiker als integraal onderdeel van het systeem en worden ook sociale aspecten meegenomen in het voorspellingsproces;

· kenmerken zich door vergevorderde integratie van verschillende bronnen van informatie die elkaar aanvullen;

· zijn multi-inzetbaar en leveren gegevens aan een breed spectrum van eindgebruikers;

- zijn zelfverbeterend door middel van standaard benchmarking en continue verbeteringsprocessen;
- worden beheerd en geëxploiteerd door consortia die gezamenlijk de nodige expertise en middelen hebben voor zulke geavanceerde systemen.

Poetry slam entries

My baby

Did I show you yet, a picture of my baby?
No?
Here, a picture! A picture! A figure!
Here its processing it's first (it's first!) time steps!
Here you see it making it's first steps!
Is anything more wonderful?
Here, the poor child, is throwing errors.
Is it not adorable?
Doesn't it do this adorably?
Don't you think it is so cute?
Only Yesterday it made it's first appearance on a conference.
Such a great model, just like it's modeller.
Such a great toddler, just like it's modeller
Somebody said we should train it better,
but I think it is just perfect as it is.
It is after all such a great model.

Of course, some times are hard.
The errors, the little accidents, the long and broken nights.

Sigh.

But, enough about my model.
Do you have any plans to start making models?

Het regent

Er valt regen, dat valt tegen!
Geen festival, zo'n valpartij aan regenval!

Valt het in een regenmeter, per toeval,
duizenden druppels in een regen-val?
Of op asfalt als een waterval?

Hoeveel regen er valt, is niet te valideren.
Maar na de regen-zonde-val, volgt noodgeval.

O thesis my thesis

O Thesis, my Thesis, our lonely trip is done,
Our theories weather'd every rack the prize we sought is won.
The bell comes near, 'hore est,' I hear, the people all exculting,
While follow ears with answers filled, the questions grim and daring;
   But O heart, heart, heart!
   O the fleeting drops of dread,
   Where on the desk my Thesis lies,
   It's pages cold un-read.

O thesis, my thesis! Endure this lonely spell;
Endure! for you my insight lives - for you the wine now spills,
For you the title and ribbon'd hats - for you the room a-crowding,
For you they came, t'interested mass, their eager int'lects churning,
   Here thesis! Dear thesis!
   A place upon the shelf,
   It is some dream that on the desk,
   You've fallen cold and dead.

My thesis has some answers, more daring questions still,
My child awaits now readers, afraid they never will,
Uncover theories safe and sound, its voyage closed and done,
From fearfull trip now membership of science circles won.
   Exult O Chairs! And ring! O Bell!
   But I with muddled head,
   Walk from the desk my thesis lies,
   It's pages cold un-read.

after 'O Captain, my Captain' by Walt Whitman

Dankwoord

Mag je nog Interessante Tijden meemaken! (eng: May You Live in Interesting Times) is een ironisch gezegde waarbij je iemand rampspoed en onzekerheid toewenst. De onderliggende gedachte is dat het vele malen beter is om een rustig en voorspelbaar leven te hebben. Een quote in dezelfde trant is 'Geluk is nooit groots' (eng: Happiness is never grand), waarbij wordt gesteld dat alle emoties die komen kijken bij voor- en tegenspoed niet opwegen tegen een stabiele gelukzaligheid.

We leven in interessante tijden. Je kan er van alles van vinden, maar dat verandert er niet zoveel aan. Een van de voordelen van onzekere tijden, is dat het makkelijker is om te zien waar je dankbaar voor bent. Zeker op een moment waarop je eigen baan stabiel is, je wederzijds verloofd bent en, toch ook nog noemenswaardig, je proefschrift hebt afgerond. Een proefschrift is een persoonlijke beproeving. Echter, een ander populair metafoor over de wetenschap is evenzeer waar: je staat op de schouders van reuzen.

In mijn opsomming van reuzen, is er maar een plek waar ik kan beginnen: mijn ouders, Henk en Ellen, en mijn broer Jelle. Wanneer je geboren wordt, zijn dat natuurlijk letterlijk reuzen, maar dat is hier te simpel. Lieve pappa, jij bent voor mij de koning van het relativeren. Als ik een ding mag noemen dat ik van jou geleerd heb, is het wel om dingen in perspectief te plaatsen. Lieve mamma, van jou heb ik meegekregen hoe belangrijk het is om dingen te ondernemen met volledige passie. Beste broer, het is lastig te beschrijven hoeveel makkelijker je het mij hebt gemaakt door dingen voor lange tijd altijd als eerste uit te vinden. Je blijft toch altijd mijn voorbeeld.

Dan spring ik vooruit in de tijd en wil ik graag Vera Liem bedanken. We kennen elkaar op dit moment langer wel dan niet. Tijdens mijn studietijd is er niemand waarop ik zo heb kunnen terugvallen als op jou. Als ik benoem wat ik van jou heb meegekregen is het 'noblesse oblige' of 'vrijwillig is niet vrijblijvend'.

Dan is er een reeks aan docenten en professoren waarbij de relatie tot het volbrengen van mijn proefschrift een stuk duidelijker is. Daarbij wil ik beginnen bij Prof. Savenije en Prof. van de Giesen die mij geïnspireerd hebben op de TU Delft voor de hydrologie. Ook mijn thesisbegeleider, Thom Bogaard, wil ik niet vergeten. Zijn advies: je hebt wel de capaciteiten om een promotie te gaan doen, maar misschien moet je eerst een paar jaar bij Deltares gaan werken.

In dit rijtje van leermeesters horen ook mijn promotoren thuis. Albrecht, van jou neem ik in het bijzonder mee hoe je complexe problemen en vraagstukken kan omzetten naar praktische problemen die aan te pakken zijn. Remko, van jou neem ik mee dat 'je nog geen reden ziet voor paniek.'

Ik begon dit dankwoord met een verwijzing naar de 'interesting times' waarin we leven. Er is één iemand die mij onveranderlijk hoopvol laat zijn over een gelukzalige toekomst. Marloes liefste schat, ik ben ontelbaar dankbaar voor je bestaan.

Vrienden, familie, collega's, bestuursgenootjes, huisgenoten, studiegroepjes, IMPREX-ers, thesis-klankborden, Delftsche Zwervers en altviooldocenten, voor iedereen geld dat ik op de een of andere manier zonder jullie dit proefschrift niet zou hebben kunnen maken. Reuze bedankt.

Netherlands Research School for the
Socio-Economic and Natural Sciences of the Environment

D I P L O M A

for specialised PhD training

The Netherlands research school for the
Socio-Economic and Natural Sciences of the Environment
(SENSE) declares that

Bart van Osnabrugge

born on 28 August 1990 in The Hague, The Netherlands

has successfully fulfilled all requirements of the
educational PhD programme of SENSE.

Wageningen, 25 April 2020

Chair of the SENSE board

Prof. dr. Martin Wassen

The SENSE Director

Prof. Philipp Pattberg

The SENSE Research School has been accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW)

K O N I N K L I J K E   N E D E R L A N D S E
A K A D E M I E   V A N   W E T E N S C H A P P E N

The SENSE Research School declares that **Bart van Osnabrugge** has successfully fulfilled all requirements of the educational PhD programme of SENSE with a work load of 34.4 EC, including the following activities:

SENSE PhD Courses

o   Environmental research in context (2016)
o   Research in context activity: 'Exploring poetic approaches to the communication on the challenges and pleasures of doing PhD research' (2018)
o   Time and project management (2016)

Other PhD, Advanced MSc and broader skill Courses

o   Delft FEWS training course, Deltares (2016)
o   The Choice, Wageningen University (2018)
o   Ideating and building a business case, Deltares (2019)

Management and Didactic Skills Training

o   Supervising four MSc students with thesis, entitled 'Continental scale data assimilation of discharge and its effect on flow predictions across the CONUS', 'Spatial parameter regionalization for the SBM model in the French part of the Rhine Catchment', 'Improving snow model impact on hydrological prediction and simulation', and 'A calibration-free hydrological modelling approach with wflow_sbm for the Rhine basin' (2016-2019)

Oral Presentations

o   *Pre-processing of forcing data for usage in a gridded model of the Rhine*. Delft Software Days, 24 October 2017, Delft, The Netherlands
o   *Contribution of potential evaporation forecasts to 10-day hydrological streamflow forecast skill*. EGU General Assembly, 09-13 April 2018, Vienna, Austria
o   *Assimilating water levels of 8 Swiss lakes into a distributed hydrological model to improve low flow forecasts for the river Rhine at Basel with the AEnKF and wflow-OpenDA*, EGU General Assembly, 08-12 April 2019, Vienna, Austria
o   *IMPREX: Naar een nieuw hydrologisch model voor het Rijnstroomgebied*, 8ste Kennisdag Zoetwater, 24 September 2019, Utrecht, The Netherlands
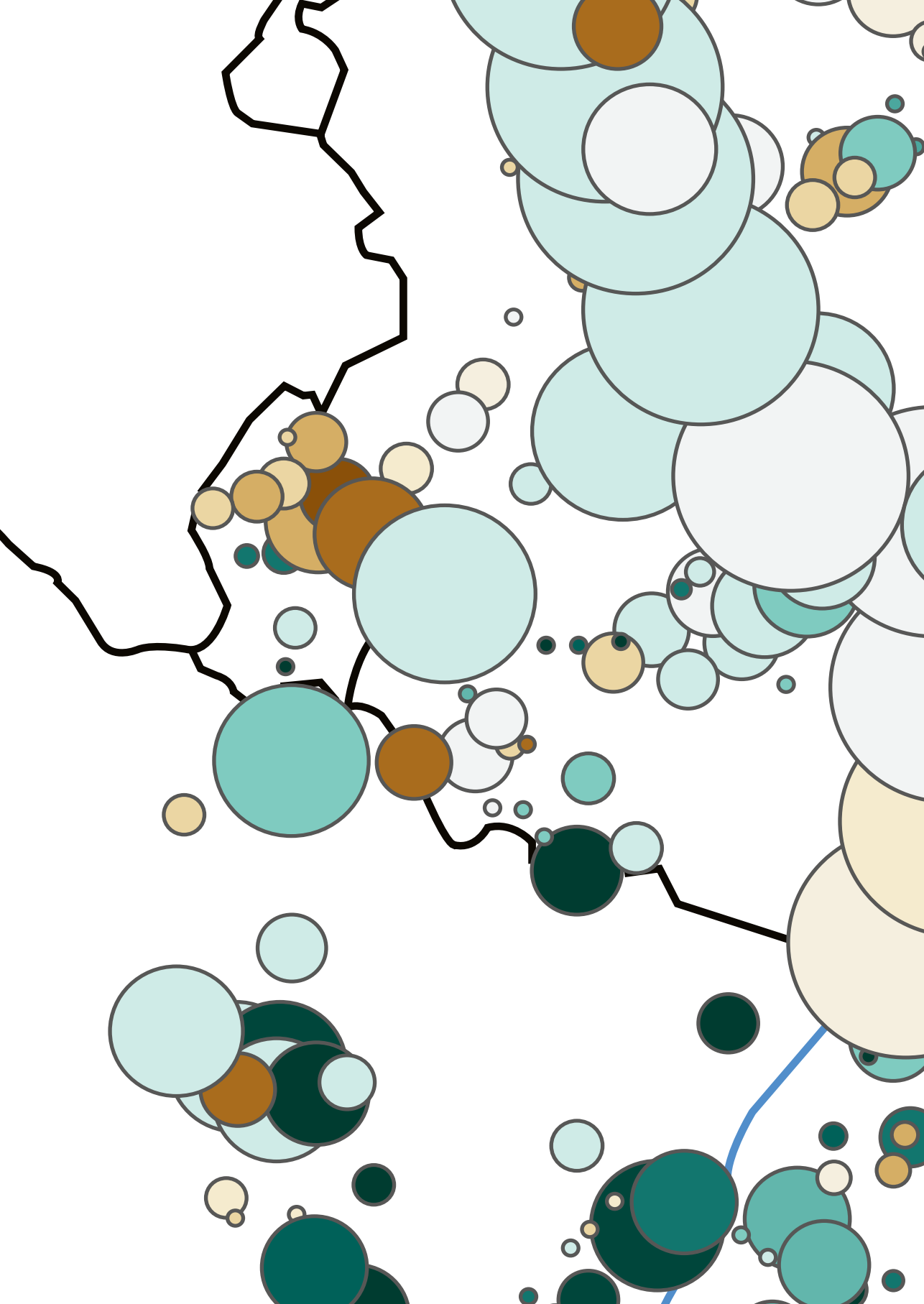
SENSE coordinator PhD education

Dr. ir. Peter Vermeulen

Propositions

1. The study of the operational aspects of hydrological forecasting is a distinctive specialization in hydrologic science.
   (this thesis)

2. Lake water level measurements can inform about upstream hydrological system states, but are less reliable for updating these modelled states than discharge measurements.
   (this thesis)

3. Since many computational science studies are not reproducible (Peng, 2011), dropping the distinction between 'hard' and 'soft' science in favour of 'reproducible' and 'non-reproducible' science would deflate the ego of the 'hard' sciences.

   Peng, R. D. (2011). Reproducible research in computational science. Science, 334(6060), 1226-1227.

4. Replacing the unwritten rule of four papers in a PhD thesis in favour of communicating the minimum requirement of 'one or two publishable chapters' (Doctoral degree regulations; Appendix 6) will increase originality of PhD research while reducing work load related stress.

   Doctoral degree regulations Wageningen University (Nov. 2019).

5. The high number of PhD students with psychiatric disorders (Levecque et al., 2017) is depressing.

   Levecque, K., Anseel, F., De Beuckelaer, A., Van der Heyden, J., & Gisle, L. (2017). Work organization and mental health problems in PhD students. Research Policy, 46(4), 868-879.

6. The added value of PhD peer groups to formal mentoring (Lewinski et al., 2017) is evidence that PhD candidates can play roles in science that are traditionally reserved for senior scientists.

   Lewinski, A. A., Mann, T., Flores, D., Vance, A., Bettger, J. P., & Hirschey, R. (2017). Partnership for development: A peer mentorship model for PhD students. Journal of Professional Nursing, 33(5), 363-369.

7. Global commercial tech companies cannot be trusted to provide free hydrological forecasts as public service, as their objective, by their nature, is to make profit.

8. Efficiency is for machines, while the essence of human life is to waste time meaningfully.

Propositions belonging to the thesis, entitled

Interpolate, Simulate, Assimilate: operational aspects of improving hydrological forecasts in the Rhine basin

Bart van Osnabrugge
Wageningen, 12 May 2020