Improving operational flood forecasting using data assimilation

Oldřich Rakovec

Thesis committee

Promotor

Prof. Dr R. Uijlenhoet

Professor of Hydrology and Quantitative Water Management

Wageningen University

Co-promotor

Dr A.H. Weerts

Deltares, Delft, The Netherlands

Other members

Prof. Dr M.F.P. Bierkens, Utrecht University, The Netherlands

Prof. Dr C.J.F. ter Braak, Wageningen University

Dr M.P. Clark, National Center for Atmospheric Research, Boulder, USA

Prof. Dr N. Verhoest, Ghent University, Belgium

This research was conducted under the auspices of the Graduate School for Socio-Economic and Natural Sciences of the Environment (SENSE).

Improving operational flood forecasting using data assimilation

Oldřich Rakovec

Thesis

submitted in fulfilment of the requirements for the degree of doctor

at Wageningen University

by the authority of the Rector Magnificus

Prof. Dr M.J. Kropff

in the presence of the

Thesis Committee appointed by the Academic Board

on Friday 25 April 2014

at 11 a.m. in the Aula.

Oldřich Rakovec

Improving operational flood forecasting using data assimilation,

xii + 142 pages.

PhD thesis, Wageningen University, Wageningen, NL (2014)

With references, with summaries in English and Dutch

ISBN 978-94-6173-849-3

Summary

Reliable and accurate flood early warnings are very important because they can mitigate the number of casualties and reduce economic damage. Understanding the behaviour of extreme hydrological events and the ability of hydrological modellers to improve their forecast skills are distinct challenges of applied hydrology. Since models simplify the real world complexity, hydrological forecasts are prone to many sources of uncertainties, such as in initial conditions, boundary conditions, and model input, structure and parameters.

Currently, most hydrological forecast systems employ lumped hydrological models (with deterministic or manual state updating), but there is a noticeable trend to move towards distributed hydrologic models with ensemble forecasting. Contrary to deterministic forecasts, which provide only one predicted value, ensemble forecasts enable quantitative estimates of the associated uncertainty. Besides improving on the boundary conditions of the hydrologic forecast, the hydrologic forecasts can be made more reliable and less uncertain by improving initial conditions and/or the model being used (either by improved model structure or better parameters). A common way of improving the initial conditions is to make use of data assimilation (DA), a feedback algorithm or update methodology which merges model estimates with available real world observations.

This thesis contributes to improved understanding and quantification of hydrological model uncertainty especially related to the initial conditions of the model and to a lesser extent to the model structure and parameters.

A plausible rainfall ensemble generator based on time-dependent multivariate spatial conditional simulations is presented. It is defined in such a way that the rainfall fields provide proper information on the spatial error structure. Additionally, it captures the temporal coherence for each realization of a sequence of spatial rainfall fields at an hourly time step. The advantage of taking the temporal coherence in hydrological model states into account is that it avoids the necessity to smooth possible extreme state values, which can occur when neglecting temporal coherence.

To efficiently and effectively evaluate parameter sensitivity, a new computationally frugal method is proposed: Distributed Evaluation of Local Sensitivity Analysis (DELSA). DELSA distinguishes important and unimportant parameters and evaluates how model parameter importance changes as parameter values change. Two test cases are compared with the popular global, variance-based Sobol' method. The results indicate that Sobol' and DELSA identify similar important and unimportant parameters, with DELSA enabling more detailed insight at much lower computational cost. For example, in the real-world problem the time delay in runoff is considered to be the most important parameter in all models, but DELSA shows that for about 20% of the parameter sets it is not important at all and alternative mechanisms and parameters dominate. Moreover, the time delay was identified as important in regions producing poor model fits, whereas other parameters were identified as more

important in regions producing better model fits.

In the second part of this thesis, benefits of data assimilation for improved flood forecasting are presented. Several computational experiments are carried out to update model states of grid-based hydrological models by sequentially assimilating streamflow observations. For example, the configuration of the discharge observation vector and the updating frequency with which the observed data becomes available are scrutinized using the Ensemble Kalman filter (EnKF). The results show that the hydrological forecast at the catchment outlet is improved by assimilating interior gauges. This augmentation of the observation vector improves the forecast more than increasing the updating frequency.

Subsequent analysis investigates how the capabilities of the DA procedure can be improved by applying alternative Kalman-type methods, e.g., the Asynchronous Ensemble Kalman Filter (AEnKF) (comparable to 4D VAR), which calculates the analysis from several previous time steps up to the time of forecast, instead of mapping the instantaneous covariance between states and discharge. The results show that including past predictions and observations in the AEnKF improves the model forecasts as compared to the traditional EnKF. Additionally we show that elimination of the strongly non-linear relation between the soil moisture storage and assimilated observations can become beneficial for improved operational forecasting.

A related study shows the effect of noise specification on the quality of hydrological forecasts via an advanced DA procedure based on a rainfall ensemble generator and lagged particle filtering. The strength of the proposed procedure is that it requires less subjectivity in implementation of DA compared to conventional methods and therefore subjective use of additional random noise in DA is mitigated.

Samenvatting

Hoogwatervoorspellingssystemen die betrouwbaar en nauwkeurig overstromingen kunnen voorspellen zijn erg belangrijk, omdat dit het aantal slachtoffers en de economische schade van overstromingen kan beperken. Het begrijpen van het gedrag van extreme hydrologische gebeurtenissen en het vermogen van de modelleur om betere en nauwkeurigere prognoses te krijgen, zijn uitdagingen binnen de toegepaste hydrologie. Omdat modellen slechts een versimpelde weergave van de complexe werkelijkheid geven, kleven er aan hydrologische voorspellingen veel onzekerheden, zoals in de initiële condities, de randvoorwaarden, de invoerdata, de modelstructuur, de modelschematisatie en de modelparameters.

Op dit moment gebruiken de meeste hydrologische voorspellingssystemen ruimtelijk geaggregeerde hydrologische modellen (zg. lumped modellen), met deterministische of handmatige aanpassing van modeluitvoer en/of -toestanden, maar er is een trend waarbij meer en meer voorspellingssystemen gebruik gaan maken van gedistribueerde hydrologische modellen met een ensemble aan prognoses. In tegenstelling tot deterministische prognoses, die slechts één waarde voorspellen, kan met een ensemble aan prognoses een inschatting van de onzekerheid in de voorspelling gemaakt worden. Naast het verbeteren van de randvoorwaarden, kan de onzekerheid van de hydrologische voorspelling verkleind worden door het verbeteren van de initiële condities en/of het model (zowel door de modelstructuur te verbeteren als door de parameters nauwkeuriger te bepalen). Een veelgebruikte methode om de initiële condities van het model te verbeteren, is door gebruik te maken van automatische data-assimilatie (DA), een terugkoppelingsalgoritme dat modelprognoses combineert met observaties.

Dit proefschrift draagt bij aan een verbeterd begrip en kwantificatie van hydrologische modelonzekerheid, vooral gekoppeld aan de initiële condities van het model, en in mindere mate aan de modelstructuur en de parameters.

In dit proefschrift wordt een neerslagensemblegenerator gepresenteerd, gebaseerd op tijdsafhankelijke multivariate ruimtelijke processen. De generator is zo gedefiniëerd dat de gegenereerde neerslaggebieden informatie bevatten over de ruimtelijke foutenstructuur. Daarnaast beschrijft het de samenhang in de tijd van opeenvolgende neerslaghoeveelheden met een uurlijkse tijdstap. Het meenemen van deze temporele correlatie in hydrologische condities heeft als voordeel dat het niet langer nodig is om het effect van extreme condities weg te werken, wat wel voor kan komen als deze temporele correlatie niet wordt meegenomen.

Om parametergevoeligheid efficiënt en effectief te evalueren, wordt in dit proefschrift een nieuwe methode voorgesteld: Distributed Evaluation of Local Sensitivity Analysis (DELSA). DELSA onderscheidt belangrijke en minder belangrijke parameters en evalueert hoe het belang van een parameter verandert zodra de parameterwaarde zelf verandert. Er zijn twee voorbeelden gegeven waarbij DELSA vergeleken wordt met de populaire Sobol'-methode. De resultaten laten zien dat Sobol' en

DELSA beide dezelfde parameters als belangrijk en minder belangrijk aanmerken, waarbij DELSA meer inzicht geeft in de parameters en minder berekeningen nodig heeft. Vaak wordt bijvoorbeeld de parameter die de vertraging tussen neerslag en afvoer omschrijft als één van de belangrijkste parameters beschouwd, terwijl DELSA laat zien dat voor ongeveer 20% van de parametersets deze parameter helemaal niet belangrijk bleek te zijn. Daarnaast bleek deze parameter vooral belangrijk te zijn in die delen van de parameterruimte waar het model slecht presteert, terwijl andere parameters van groter belang bleken te zijn in parametergebieden waar het model wél goed presteert.

In het tweede deel van dit proefschrift worden de voordelen van data-assimilatie voor het verbeteren van overstromingsvoorspellingen gepresenteerd. Verschillende experimenten zijn uitgevoerd om de modeltoestand van een ruimtelijk gedistribueerd hydrologisch model te updaten, door continu afvoerobservaties te assimileren in het model. De configuratie van de afvoervector en de frequentie van de update zijn onderzocht met gebruikmaking van het Ensemble Kalman Filter (EnKF). De resultaten laten zien dat de hydrologische voorspelling van de afvoer bij de uitgang van het stroomgebied verbeterd wordt door afvoerobservaties van binnen het stroomgebied mee te nemen in de assimilatie. Het meenemen van meer observaties heeft een groter positief effect op de voorspellingen dan het opvoeren van de frequentie waarmee de modeltoestand wordt aangepast.

Vervolgens is onderzocht hoe de data-assimilatieprocedure verder verbeterd kan worden door het toepassen van alternatieve Kalman-methoden. Bijvoorbeeld met de Asynchronous Ensemble Kalman Filter (AEnKF), die voor de berekening meerdere tijdstappen terugkijkt. De resultaten laten zien dat het meenemen van oudere voorspellingen en observaties in de AEnKF de voorspelling verbetert in vergelijking met de voorspelling gedaan met de traditionele EnKF. Daarnaast laten we zien dat het niet-meenemen van de sterk niet-lineaire relatie tussen bodemvochtopslag en geassimileerde observaties voordelen kan hebben voor het verbeteren van operationele afvoervoorspellingen. Een gerelateerde studie laat het effect van ruis op de kwaliteit van de hydrologische voorspelling zien, via een DA procedure gebaseerd op de neerslagensemblegenerator. De kracht van de voorgestelde methode is dat hierdoor de ruisbeschrijving van het model beperkt kan blijven terwijl nauwkeurigere resultaten behaald worden voor korte voorspellingshorizons.

Contents

Chapter 1

Introduction

1.1 Background

Floods represent one of the most devastating natural hazards, endangering people's well-being and livelihoods. In the last decade, hundreds of millions of people have been affected by floods every year, and the estimated socio-economic damage is growing exponentially across the world. This trend is mainly caused by an increasing frequency of extraordinary rainfall events, undesired land use changes (agricultural intensification and deforestation) and last but not least an increased number of people living in flood prone areas (*UN*, 2013; *WMO*, 2013; *IPCC*, 2013).

To illustrate the global extend of floods, consider the 3-month summer period of 2013, which attracted attention of the mass media but also of the scientific community. Extensive flooding in central Europe in June 2013 was classified by *Blöschl et al.* (2013) as one of the largest in the past two centuries. Next to that, hundreds of thousands of people had to be evacuated because of the flooding on the Amur River (China and Russia) in August 2013 (*Deutsche Welle*, 2013). The U.S. State of Colorado was hit by a flash floods in September 2013, during which the rainfall accumulation was so high that it approached the mean annual accumulation (*HEPEX*, 2013). On top of that, the Northern India flooding left thousands of casualties in July 2013 (*BBC*, 2013).

To reduce flood risk, many initiatives are developed across the world and coordinated regionally, nationally, or even internationally as the natural hazards are not limited to country boundaries. For example, the Flood Forecasting Initiative under the international umbrella of the World Meteorological Organization (*WMO*, 2013) aims to reduce the vulnerability of society by promoting integrated flood management and improved availability of qualitative and quantitative weather forecasting products.

Although recent developments in science show that it is possible to mimic complex hydrological processes leading to extreme events by models, which simplify reality, there are many obstacles for having reliable and effective real time flood early warning systems. Often, there is a lack of data infrastructure and financial resources, besides challenges in clear and prompt communication among hydro-meteorological agencies, government decision makers, and the affected people. Prompt measures in response to potential flood warnings are crucial and when issued early in advance, they can considerably mitigate the negative impacts of floods.

The causes of flooding are typically classified as either atmospheric or geotechnical (*Sene*, 2008). The atmospheric ones originate from heavy rainfall or snowmelt. Moreover, atmospheric floods nearby the coastal line can be caused by surge, wave and wind effects from the sea. Geotechnical

factors such as landslides, debris flows and earthquakes can also trigger serious flooding in both the inland water systems and coastal regions.

This thesis is particularly focused on improved understanding of floods within inland water systems, which are often referred to as catchments, watersheds or river basins. The type of floods investigated in this thesis can be classified as atmospheric, mainly caused by heavy rainfall. For this objective, rainfall-runoff hydrological models are used to simplify and represent the complex hydrological processes. Although human control, such as reservoir operations, can influence flood forecasting, it will not be discussed here, because the emphasis of this thesis lies on the natural physical processes associated with flooding.

1.2 Hydrological modelling

1.2.1 Overview

A historical overview of rainfall-runoff models ranging from empirical to physically-based models is provided for example by *Singh* (1995); *Dooge and O'Kane* (2003); *Dingman* (2004); *Wagener et al.* (2004) and *Todini* (2007). Traditionally, hydrological models can be divided into three groups: (1) metric (empirical, black box, data driven) (2) parametric (conceptual, grey box), (3) mechanistic (physically based, process based, white box).

Metric models are purely based on the information derived from the data and no prior knowledge about catchment behaviour is needed. Examples of this type of models are Artificial Neural Networks (ANN) and Transfer Functions (TF). These models are usually spatially lumped, i.e. the catchment is assumed to be a single unit.

Parametric models aim to satisfy the water balance of the catchment, which is represented by a system of reservoirs. Storages are filled by fluxes such as precipitation, snowmelt, infiltration, and percolation, and they are emptied through evapotranspiration, discharge, drainage. The model parameters are often hardly measurable in the field and need to be derived in a calibration procedure. These models are either lumped, quasi-lumped (segmentation of catchment into smaller sub-catchments) or fully distributed. Well known and commonly used conceptual hydrological models at the catchment and the grid scale are for example the GR4J model (e.g. *Edijatno et al.*, 1999), the HBV model (e.g. *Bergström and Sandberg*, 1983), the Hymod (e.g. *Wagener et al.*, 2001), the NAM-MIKE11 (e.g. *Havno et al.*, 1995), the PCR-GLOBWB model (e.g. *van Beek et al.*, 2011), the Probability Distributed Model (PDM; e.g. *Moore*, 1985, 2007), the Precipitation Runoff Modelling System (PRMS; e.g. *Leavesley et al.*, 1983), the Sacramento model (SAC-SMA; e.g. *Burnash*, 1995), TOPMODEL (e.g. *Beven and Kirkby*, 1979; *Beven et al.*, 1995), the Variable Infiltration Capacity (VIC) model (e.g. *Liang et al.*, 1994) and the Wageningen model (e.g. *Stricker and Warmerdam*, 1982; *Brauer et al.*, 2014). Besides employing models with a fixed structure, the use of conceptual hydrological modelling toolkits gains increasing popularity (e.g. *Clark et al.*, 2011a; *Gupta et al.*, 2012), such as the Rainfall Runoff Library (RRL; e.g. *Perraud et al.*, 2003), the Rainfall Runoff Modelling Toolkit (RRMT; e.g. *Wagener et al.*, 2004), the Framework for Understanding Structural Errors (FUSE; e.g. *Clark et al.*, 2008a) or the FLEX framework (e.g. *Fenicia et al.*, 2008). These toolkits consist of reservoirs and fluxes (classical elements in conceptual hydrology) defined by the user in a very flexible way. They allow for multiple hypothesis testing of an ensemble of model equations of different complexity.

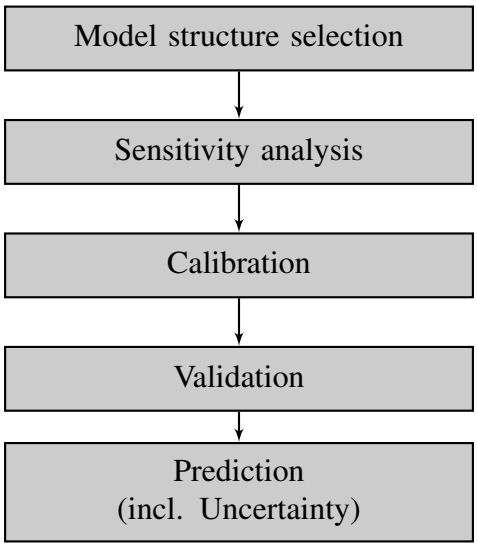


Figure 1.1: Simplified procedure of rainfall runoff modelling (after *Wagener et al.*, 2004).

Mechanistic models are described by the conservation equations of mass, momentum and energy with appropriate boundary and initial conditions. They became practically applicable in 1980s when computer power and availability increased. Nevertheless, extreme data demand, scale-related problems and overparameterization caused that in the hydrological community these models did not yet become as popular as the more conceptual parametric models. An example of a mechanistic approach is the Representative Elementary Watershed framework (REW) (*Reggiani et al.*, 1998). Additionally, the mechanistic models are particularly popular in the land-surface modelling community (e.g. the Community Land Model, *Lawrence et al.*, 2011), the agro- or ecohydrological community (e.g. the Soil Water Atmosphere Plant model, *Van Dam et al.*, 2008), and to model water flow and solute transport in variably saturated porous media, for example using the HYDRUS model (e.g., *Simunek et al.*, 2008) or the ParFlow model (e.g., *Kollet and Maxwell*, 2006).

1.2.2 Model development and uncertainty

Traditionally, the modelling procedure varies according to various circumstances but in general includes a number of steps (Figure 1.1), as they are summarized in *Wagener et al.* (2004) and *Dunn et al.* (2008). Firstly, *model structure* is selected according to the modelling purpose and required temporal resolution of the model. This can be conditioned on the available data. The catchment characteristics are important criteria to determine the modeling concept, nevertheless subjective criteria such as e.g. the modelling experience of the modeller also influence this selection step. *Sensitivity analysis* is a technique to find out how sensitive the model is to changes in parameter values. Sensitivity analysis is used to identify the most important parameters and thus the most important model components. Since the catchment averaged parameters are often hardly measurable, they need to be estimated by a *calibration* procedure. Calibration tries to optimize the fit between the modelled and observed time series for a given objective function by adjusting the values of model parameters. *Validation* repre-

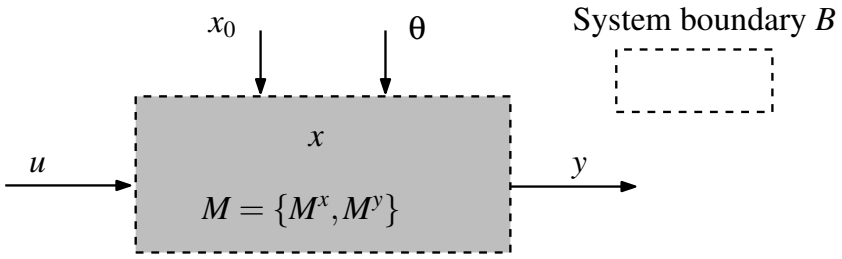


Figure 1.2: Schematic diagram of model components from a systems perspective (after *Liu and Gupta*, 2007).

sents the testing of the calibrated model on an independent period from the calibration period. It is used to verify the applicability of the optimized model. Systematic testing of hydrological simulation models including selections of calibration and validation periods is discussed a.o. in *Klemeš* (1986). *Prediction* can represent the last step of hydrological simulations. Its reliability is connected to the total model uncertainty, which originates from data uncertainty plus model uncertainty.

Since models simplify the real world complexity, hydrological predictions are prone to many sources of uncertainties. Adequate addressing of uncertainty associated with hydrological prediction is a challenging problem, which requires understanding, quantification and reduction of the uncertainty (*Liu and Gupta*, 2007). To enable identification of the uncertainties, hydrological models can be decomposed, as shown in Figure 1.2: model inputs (u), time-varying states (x) of mass and energy stored within the system boundary (B), model structure (M, which consists of the input-to-state mapping M^x and the state-to-output mapping M^y), parameters (θ), initial conditions (x_0) and model output (y). For five of these components (u, B, M, θ, x_0) the uncertainty needs to be specified beforehand and for states x and output y it is computed by running the model. To reduce the prediction uncertainty in hydrological models, data assimilation procedures are becoming popular among hydrological scientists. However, their applicability within real-time operational settings needs further investigation and validation.

1.2.3 Data assimilation (DA)

Currently, most hydrological forecasts around the world employ lumped parametric hydrological models for flood forecasting (with deterministic or manual state updating), but there is a clear tendency to move towards distributed models with hydrological ensemble forecasting (e.g. *Koren et al.*, 2004). Contrary to deterministic forecasts, which provide only one predicted value, ensemble forecasts also quantitatively estimate the probability of possible outcomes. To improve the hydrological forecasts, quantification and reduction of uncertainties within the whole flood forecasting chain need to be addressed, which can be achieved through data assimilation (DA; *Liu and Gupta*, 2007; *Liu et al.*, 2012; *Weerts et al.*, 2013).

Data assimilation is a technique to merge measurements of any type with estimates from geophysical models (*Evensen*, 2003; *Reichle*, 2008; *van Leeuwen*, 2009; *Liu et al.*, 2012). It can be

seen as an update of the model state and/or parameter with externally measured variables, which also quantifies the errors, uncertainties in input data, model structure and observations (*Pauwels and De Lannoy*, 2006; *Moradkhani*, 2008; *Clark et al.*, 2008b). Application of data assimilation arose in many different fields, such as oceanography, meteorology, hydrology, ecology, and crop science (*Karssenberg et al.*, 2009). Data assimilation systems help the data-users to organize the useful and redundant observational information into physically consistent estimates of the variables according to their relevance (*Reichle*, 2008). In other words, data assimilation is a feedback system where the forecast is conditioned on all information that is available at the time of the forecast (forecast origin). This includes information about the current state of the system and past performance of the forecast system. Quantification and, if possible, also the reduction of errors in the forecast model need to be quantified to gain (1) more accurate forecasts and (2) information regarding the uncertainty of the forecast, including its bias and skill.

DA methods used in hydrology are mostly employed for state updating of models' dynamic states to estimate the best initial conditions and/or parameter estimation/calibration problems. DA methods can be divided into two classes: (1) *sequential* and (2) *variational*. The sequential methods are employed by assimilating observations when they become available and the analysis depends on the uncertainties in observations and/or model. The variational methods, contrary to the sequential methods, rather adjust the initial value of the mathematical model, instead of changing the state directly. At the beginning, a first-guess model is constructed, which is afterwards updated by creating an adjoint model which propagates backwards in time and incorporates the mismatch between the model and observations.

Sequential DA methods

The first known sequential data assimilation technique is the *Kalman filter* (KF), which was originally developed for linear systems (*Kalman*, 1960). Since hydrological processes are rather non-linear, it further developed into the *Extended Kalman filter* (EKF) (e.g. *Georgakakos*, 1986). The major drawbacks of EKF are the high computational demand for the propagation of the background error covariance (especially for large system state vectors) and the fact that higher order derivatives for the background error covariance propagation and the mapping of the observational information to the model state variables are neglected (*Pauwels and De Lannoy*, 2009; *Salamon and Feyen*, 2009). The *Ensemble Kalman filter* (EnKF) (*Evensen*, 2003, 2009) represents another variation of the KF. The EnKF propagates an ensemble of model realizations (generated from model perturbations) through time and estimates the background error covariance matrix from the ensemble statistics. EnKF is computationally efficient, but is limited to errors with Gaussian distributions.

The particle filter is another form of a recursive Bayesian filter based on Monte Carlo simulation. Particles, associated with weights, are used to approximate the posterior probability distribution functions (*Arulampalam et al.*, 2002; *Chen*, 2003; *Moradkhani et al.*, 2005a; *Weerts and El Serafy*, 2006; *van Leeuwen*, 2009; *Noh et al.*, 2011a). The advantage of this technique is that no assumptions on the form of the prior pdf of the model states are necessary and that the full prior density is being used, in contrast to EnKF, which uses only its second-order characteristics (mean and co-variance). In theory, this would mean that particle filtering is more sensitive to the tails of the prior distribution, a property which may be of vital importance in flood forecasting, although this may be at the expense of

significantly more simulations and thus a higher computational demand (*Weerts and El Serafy*, 2006).

Variational DA methods

Variational methods have been widely used in data assimilation for numerical weather prediction. They can deal with a very large number of observations to be assimilated in a computationally efficient way. The technique depends on defining the adjoint model, which provides local gradient terms for any predicted variable that can be matched to an observable. These gradients will vary in space and time, depending on the nonlinearity of the model. Linear extrapolation is then used to adjust model-predicted variables towards the observed values, depending on an estimate of the covariance matrix. In this sense, it is similar to EKF but, unlike the EKF, does not update the covariance matrix as the data assimilation proceeds. Hydrological forecasting applications, where variational methods are investigated are presented in *Seo et al.* (2003, 2009) and *Lee et al.* (2011, 2012).

Uncertainty definitions in DA

While applying data assimilation to improve model prediction, the choice of a suitable model error is essential (e.g. *Weerts and El Serafy*, 2006; *McMillan et al.*, 2013). Uncertainties in data assimilation originate from model initialization, observation errors or an improper DA structure. The optimality of the DA procedure mainly depends on the reliability of error estimates for the model inputs (i.e., forcings) and the model itself (*Liu et al.*, 2012).

Concerning model inputs, precipitation is assumed as the most uncertain hydrological model input, because of the short temporal and spatial correlation scales. Historically, spatial precipitation information was obtained by rain gauge measurements based on point scale estimates. However, during the last decades, application of weather radar at larger scales has improved our understanding of the spatial and temporal properties of rainfall even further. Unfortunately, precipitation estimates by weather radar are prone to errors (e.g. *Seo et al.*, 1999; *Steiner et al.*, 1999; *Krajewski and Smith*, 2002; *Hazenberg et al.*, 2011). To date, this implies that operational nowcasting/forecasting systems still have to make use of rain gauge information to mitigate the errors in radar rainfall estimates (e.g. *Schuurmans et al.*, 2007; *Goudenhoofdt and Delobbe*, 2009; *Cole and Moore*, 2009). Uncertainty of precipitation fields in DA procedures is commonly quantified either by stochastic perturbations, or through conditional simulation methods (*Clark and Slater*, 2006; *Leblois and Creutin*, 2013). The latter provide statistically reliable estimates, however, they are more computationally expensive than ad hoc stochastic perturbations (*McMillan et al.*, 2011; *Liu et al.*, 2012). For recent advances in radar rainfall uncertainty estimation and its hydrological application see *Berenguer et al.* (2005); *Germann et al.* (2009); *Berenguer et al.* (2011) and *Hazenberg et al.* (2011).

The uncertainty in the model itself is associated with model conceptualizations. Mathematical implementation and its quantification is an extremely difficult task. Currently, the hydrological community has adopted four approaches to quantify model error (after *Liu et al.*, 2012): (1) stochastic perturbation of the model state variables (e.g. *Reichle et al.*, 2002; *Vrugt et al.*, 2006; *Clark et al.*, 2008b) (2) use of inverse methods to infer probability distributions for each of the model parameters (e.g. *Beven and Binley*, 1992; *Vrugt et al.*, 2003), (3) simultaneous update of model states and parameters (e.g. *Moradkhani et al.*, 2005b; *Vrugt et al.*, 2005) and (4) use of multi-model ensembles (e.g. *Georgakakos et al.*, 2004).

1.2.4 Operational flood forecasting in the Netherlands

A spatially lumped HBV-96 model (*Bergström and Sandberg*, 1983; *Lindström et al.*, 1997) is currently used operationally by the Dutch authorities for flood forecasting at the two largest rivers flowing into the Netherlands: the Rhine River and the Meuse River. The area of the Rhine River at Lobith, the entrance point of the Rhine River into the Netherlands, equals $\sim 160\,000\,\mathrm{km}^2$. The area of the Meuse River at Sint Pieter, the entrance point of the Meuse River into the Netherlands, equals $\sim 21\,000\,\mathrm{km}^2$. Both rivers are conceptualised into a number of lumped HBV-96 sub-basins. The Rhine River consists of 134 sub-basins and the Meuse River consists of 15 sub-basins. Currently, the operational forecasts derived with the lumped hourly HBV-96 models are used as lateral inflows into a 1D-hydrodynamic model. This lumped model does not employ data assimilation in terms of the sequential model state updating, as presented in Sect. 1.2.3. However, it is updated with discharge observations by means of an automated auto-regressive error correction method (*Broersen and Weerts*, 2005).

1.3 Research questions and thesis outline

The main research objective of this thesis is to *develop and test hydrological data assimilation methods for improving hydrological model forecasts used in operational flood forecasting systems*. To fulfil this challenging objective, the following research questions are formulated, discussed and answered as separate chapters of this thesis:

- How to define a plausible precipitation ensemble generator using rain gauges to capture the temporal coherence for each realisation of a sequence of spatial rainfall fields at an hourly time step? (Chapter 2, *Rakovec et al.*, 2012a)
- How to efficiently and effectively evaluate parameter sensitivity, using multiple evaluations of local parameter sensitivity distributed throughout parameter space? (Chapter 3, *Rakovec et al.*, 2014)
- What is the sensitivity of the DA procedure using the Ensemble Kalman Filter to the number and the locations of discharge gauges, which are assimilated into a grid-based distributed operational hydrological forecasting model? (Chapter 4, *Rakovec et al.*, 2012b)
- What is the effect of state augmentation using a time window of past discharge observations assimilated by the Asynchronous Ensemble Kalman Filter on forecast accuracy? Additionally, how does the choice of model states being updated affect the forecast accuracy? (Chapter 5)
- What is the effect of noise specification in a DA scheme on the quality of hydrological forecasts? (Chapter 6)

To illustrate how these research questions contribute to the application of hydrological models within a flood forecasting framework, we refer to Figure 1.3. Firstly, the model is developed and scrutinized over an extensive period. Then, a real-time operation (prediction) itself takes place, which can be further split into two modes: (1) a historical one and (2) a forecast one. These modes are separated by the current time (start of the forecast). Chapter 2 contributes to a better understanding

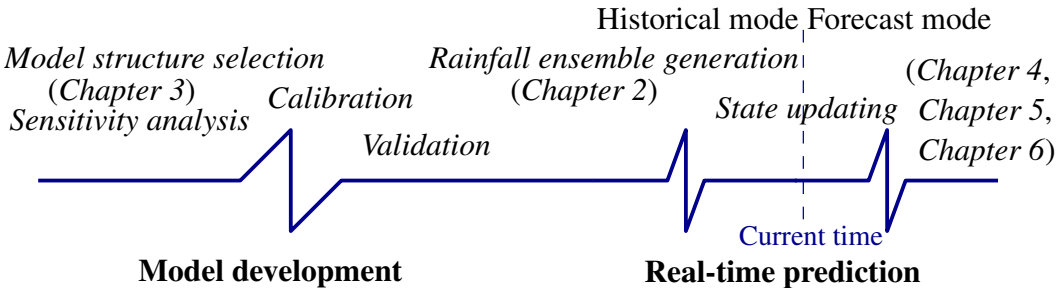


Figure 1.3: Modes of operational flood forecasting.

of spatial and temporal uncertainties of rainfall fields, either in the model development phase or in real time operation. Chapter 3 evaluates local parameter sensitivities in a distributed fashion and also partly evaluates the choice of model structure during the model development phase. Chapters 4, 5, and 6 focus on state updating at the beginning of the forecast mode. While Chapters 4 and 5 focus entirely on the uncertainty in the initial model states and the hydrological forecasts are forced with observed meteorological inputs, Chapter 6 also considers uncertainties in the meteorological forecasts.

Chapter 2

Generating spatial precipitation ensembles: impact of temporal correlation structure

Abstract

Sound spatially distributed rainfall fields including a proper spatial and temporal error structure are of key interest for hydrologists to force hydrological models and to identify uncertainties in the simulated and forecasted catchment response. The current paper presents a temporally coherent error identification method based on time-dependent multivariate spatial conditional simulations, which are conditioned on preceding simulations. A sensitivity analysis and real-world experiment are carried out within the hilly region of the Belgian Ardennes. Precipitation fields are simulated for pixels of $10\,\mathrm{km} \times 10\,\mathrm{km}$ resolution. Uncertainty analyses in the simulated fields focus on (1) the number of previous simulation hours on which the new simulation is conditioned, (2) the advection speed of the rainfall event, (3) the size of the catchment considered, and (4) the rain gauge density within the catchment. The results for a sensitivity analysis show for typical advection speeds $> 20\,\mathrm{km\,h^{-1}}$, no uncertainty is added in terms of across ensemble spread when conditioned on more than one or two previous hourly simulations. However, for the real-world experiment, additional uncertainty can still be added when conditioning on a larger number of previous simulations. This is because for actual precipitation fields, the dynamics exhibit a larger spatial and temporal variability. Moreover, by thinning the observation network with 50 %, the added uncertainty increases only slightly and the cross-validation shows that the simulations at the unobserved locations are unbiased. Finally, the first-order autocorrelation coefficients show clear temporal coherence in the time series of the areal precipitation using the time-dependent multivariate conditional simulations, which was not the case using the time-independent univariate conditional simulations. The presented work can be easily implemented within a hydrological calibration and data assimilation framework and can be used as an improvement over currently used simplistic approaches to perturb the interpolated point or spatially distributed precipitation estimates.

This chapter is largely based on the publication: Rakovec, O., P. Hazenberg, P. J. J. F. Torfs, A. H. Weerts, and R. Uijlenhoet (2012), Generating spatial precipitation ensembles: impact of temporal correlation structure, *Hydrol. Earth Syst. Sci., 16*, 3419–3434, doi:10.5194/hess-16-3419-2012.

2.1 Introduction

Precipitation is the most dominant input term determining the hydrological response at the catchment scale (*Beven*, 2001). Historically, spatial precipitation information was obtained by rain gauge measurements based on point scale estimates. However, during the last decades, application of weather radar at larger scales has improved our understanding of the spatial and temporal properties of rainfall even further. Unfortunately, precipitation estimates by weather radar are prone to errors (e.g. *Seo et al.*, 1999; *Steiner et al.*, 1999; *Krajewski and Smith*, 2002; *Seo and Breidenbach*, 2002; *Germann et al.*, 2009; *Hazenberg et al.*, 2011). To date, this implies that operational nowcasting/forecasting systems have to make use of rain gauge information to mitigate the errors in radar rainfall estimates (e.g. *Schuurmans et al.*, 2007; *Goudenhoofdt and Delobbe*, 2009; *Cole and Moore*, 2009). As such, rain gauges remain an important tool for the derivation of unbiased spatially distributed rainfall estimates.

To obtain sound spatially distributed rainfall information from rain gauge observations, these devices generally are interpolated to appropriate spatial and temporal resolutions, depending on the hydrological purpose. Among interpolation methods, geostatistical techniques (like kriging) are popular. These methods take into account information about the spatial variation within an area and provide both a mean rainfall as well as an associated error estimate (*Webster and Oliver*, 2001; *Schuurmans and Bierkens*, 2007). These errors in areal rainfall are of key interest for hydrologists because they can be used to estimate uncertainties in catchment response. An evaluation of those errors in a spatially lumped manner has been discussed, for example, by *Zawadzki* (1973) and *Willems* (2001). We refer to *Villarini et al.* (2008) and *Ciach and Krajewski* (2006) and references cited therein, for analyses of different spatial and temporal sampling errors from a rain gauge perspective.

Nevertheless, kriging is prone to smooth local variability of rainfall. Further away from observation points, high (low) values tend to be underestimated (overestimated) (*Goova erts*, 1997). These biases decrease the usability of kriging in applications sensitive to extreme values (*Goova erts*, 1997), such as within spatially distributed rainfall-runoff modelling (*Bivand et al.*, 2008).

Sound spatial and temporal estimates of precipitation and its corresponding uncertainty are of key interest for both scientific and applied hydrological studies (*Liu et al.*, 2012). The generation of an ensemble, which is a finite and discrete number of spatial realisations over time, is able to realise this goal. A common practice in hydrological data assimilation applications to obtain ensembles is to perturb the interpolated point or spatially distributed estimates by Gaussian white noise with a standard deviation ranging between 15–50 % of the observed precipitation (e.g. *Pauwels and De Lannoy*, 2006; *Weerts and El Serafy*, 2006). Although this approach leads to hydrological model simulations with wide discharge uncertainty bands, the realisations are not very realistic from a hydro-meteorological perspective, because of a lack in coherent temporal evolution of each individual precipitation realisation.

Sound spatially distributed rainfall fields including a proper spatial error structure can be obtained by conditional simulation. Unlike interpolation this technique provides both the best local estimate and ensures that realisations match the sample statistics and are conditional on neighbouring estimates. In other words, conditional simulations provide proper information about the spatial uncertainty (*Goova erts*, 1997). Several hydrological studies have applied conditional simulations at daily (e.g. *Clark and Slater*, 2006; *Schuurmans*, 2008; *Vischel et al.*, 2009; *Grimes and Pardo-Iguzquiza*, 2010) and at hourly time steps (*AghaKouchak et al.*, 2010; *Renard et al.*, 2011).

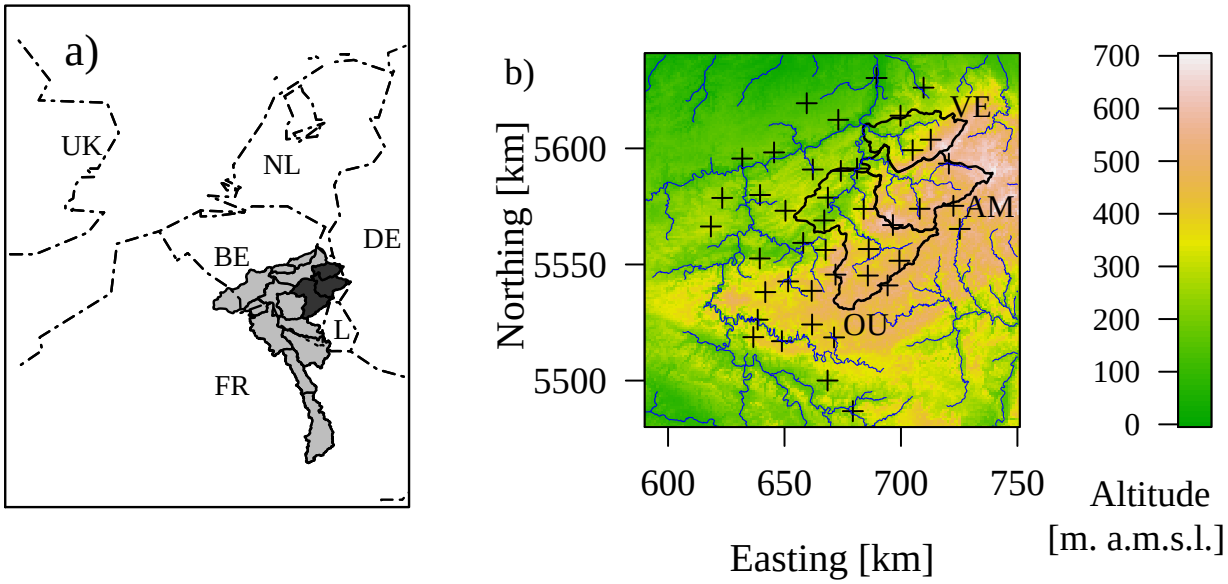


Figure 2.1: (a) The Meuse River basin upstream of Sint Pieter ($\sim 21\,000\,\mathrm{km}^2$), the entrance point into the Netherlands, and its 15 sub-catchments (grey polygons) including the Upper Ourthe, Amblève and Vesdre sub-catchments (black polygons). (b) Topographic map of the Belgian Ardennes including the Upper Ourthe (OU), Amblève (AM) and Vesdre (VE) sub-catchments, the channel network (blue lines), and rain gauges (plusses). Projection is in the Universal Transverse Mercator (UTM) 31N coordinate system. After *Hazenberg et al.* (2011).

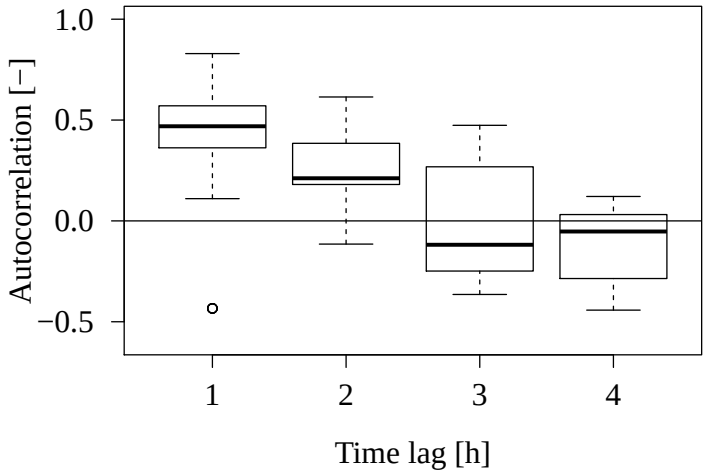


Figure 2.2: Autocorrelation coefficients of the catchment averaged precipitation of the Upper Ourthe for 30 rainy events between 1 October 2002 and 31 March 2003. The bold black line in the boxplot is the median, the body of a boxplot shows the interquantile range (Q75–Q25) and the whiskers represent the sample minima and sample maxima, unless the extreme value occurs further than 1.5 times the interquartile range from the box, in which case the outlier is shown by a dot.

Table 2.1: Characteristics of three representative rainfall events.

Starting time	Duration [h]	Speed[a] [km h^{-1}]	Mean[b] [mm]	St. dev.[c] [mm]
22 Oct 2002	10	54	12.2	7
22 Dec 2002	10	21	16.3	3.9
1 Jan 2003	10	33	17.5	5

[a] The advection speed for the three precipitation events was obtained by estimating the mean speed of the precipitation system during the event. We identified the precipitation system based on observed volumetric radar data for which the reflectivity exceeded 7 dBZ ($\sim$0.1 mm h^{-1}).

[b]Mean of the precipitation sums for all 42 rain gauges.

[c]Standard deviation of the 42 precipitation sums.

Unfortunately, conditional simulations do not primarily take the temporal evolution of the spatial field into account (*Goova erts*, 1997; *Webster and Oliver*, 2001; *Bivand et al.*, 2008). Nevertheless, for precipitation the temporal correlation structure can be an important aspect to be considered when generating spatial precipitation ensembles. Theoretically, this can be achieved using spatial conditional simulations which are made conditional on previous simulations. Neglecting this temporal aspect would lead to underestimation of the overall uncertainty in precipitation ensembles.

The objective of this study is to define a plausible precipitation ensemble generator using rain gauges to capture the temporal coherence for each realisation of a sequence of spatial rainfall fields at an hourly time step. Our analyses focus on the uncertainty in the simulated fields based on (1) the number of previously simulated hours on which the new simulation is conditioned, (2) the advection speed of the rainfall event, (3) the size of the catchment considered and (4) the rain gauge density within the catchment. Our goal is to present a technique, which can be easily implemented within a hydrological data assimilation framework to be used as an improvement over currently used simplistic approaches to perturb the interpolated point or spatially distributed estimates. The advantage of having the temporal coherence in hydrological model states is that it avoids the necessity to smooth possible extreme state values, which can occur when neglecting temporal coherence.

2.2 Material and methods

2.2.1 Data

Hourly precipitation data are available from 42 automatic rain gauges situated within the Belgian Ardennes region (Fig. 2.1). This moderately hilly terrain with maximum elevations of $\sim$700 m a.m.s.l. is predominantly drained by the Meuse River and partly by the Rhine River (*Berne et al.*, 2005; *Driessen et al.*, 2010).

In this chapter, we focus on the analysis of three representative stratiform winter rainfall events as described and analysed by *Hazenberg et al.* (2011): (1) a fast-moving stratiform system (22 October 2002), (2) a large-scale stratiform system (22 December 2002) and (3) a fast-moving frontal stratiform system (1 January 2003). For a further description of these events, the reader is referred to *Hazenberg et al.* (2011). The main characteristics of these events are given in Table 2.1. Addition-

ally, the autocorrelation coefficients of the catchment average precipitation of the Upper Ourthe (1600 km^2) for 30 rainy events with a minimum duration of 13 h are presented in Fig. 2.2. All 30 events were observed during the winter half year, from 1 October 2002 to 31 March 2003. Figure 2.2 shows that the temporal memory of the precipitation is highly significant (decorrelation time between 2 and 3 h) and that the autocorrelation roughly follows an exponential decay.

2.2.2 Geostatistical analysis

The variogram is a geostatistical measure of spatial variability in terms of the semi-variance over a lag distance h. The experimental omnidirectional semi-variogram, which is generally called the variogram, assumes stationarity and isotropy of the predicted variable. It represents half of the mean square difference between paired data values (i.e. measured precipitation) within the same binned lag distance h:

$$\hat{\gamma}(h) \;=\; \frac{1}{2\,N_h} \sum_{k=1}^{N_h} \left(z(x_k) \;-\; z(x_k \;+\; h)\right)^2, \tag{2.1}$$

where N_h is the number of data observation pairs and $z(x_k)$ and $z(x_k+h)$ are the observations separated by the lag distance h. Because the experimental variogram is derived only for several discrete lag distances, a parametric variogram model has to be fitted in order to obtain continuous estimates of the semi-variance.

The spherical model is a popular and widely employed variogram model for rainfall (*Berne et al.*, 2004; *Schuurmans et al.*, 2007; *van de Beek et al.*, 2011; *Verworn and Haberlandt*, 2011). With only three parameters it is defined as follows:

$$\gamma(h) \;=\; \begin{cases} c_0 \;+\; c_1 \left(\frac{3}{2}\frac{h}{a} \;-\; \frac{1}{2}\left(\frac{h}{a}\right)^3\right) & \text{if} \;\; h \;\leq\; a \\ c_0 \;+\; c_1 & \text{if} \;\; h \;>\; a. \end{cases} \tag{2.2}$$

In Eq. (2.2), the parameter c_0 is the nugget, representing the semi-variance at distance $h=0$. The parameter c_1 is the partial sill, while a represents the range, the distance beyond which the data are not correlated any more.

Certainly, the choice of variogram model involves some subjectivity. Since our main interest is to see how the temporal aspect (i.e., the temporal correlation structure) affects the uncertainty in simulated fields, we decided to keep the parametrisation as simple as possible. However, since we are focusing specifically on the added uncertainty that originates due to the advection of a precipitation cell, we expect these results not to change radically if another variogram model would have been chosen. Additionally, the spatial anisotropy was not considered in the variogram model. Although this is certainly a relevant issue, it is beyond the scope of the current study.

Figure 2.3 shows two examples of the experimental variogram as well as the fitted spherical model for two consecutive hours on 22 December 2002. Note that the unit of semivariance is in mm instead of mm^2. This is because the quantitative statistical measures, which are employed in this study, are particularly suitable for normally distributed data. However, rainfall by nature does not follow a Gaussian distribution at shorter time scales. Therefore, a pragmatic and popular solution to overcome this problem is to transform the rainfall data such that their distribution approaches a Gaussian distribution. As such, rainfall data are square-root transformed (*Schuurmans et al.*, 2007; *van de Beek*

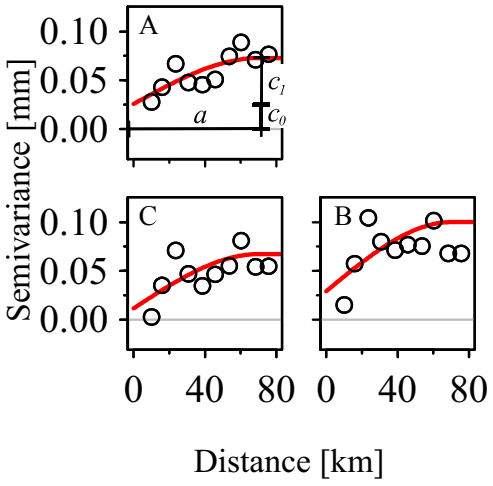


Figure 2.3: Two examples of an experimental variogram (black circles) and the fitted spherical variogram model (red curve) of square-root transformed rain gauge observations on 22 December 2002 for two consecutive hours (A and B). C shows experimental cross-variogram (circles) and fitted spherical cross-variogram model corresponding to A and B.

et al., 2011).

Additionally, Eq. (2.1) can be extended for the time-dependent multivariate case, which relates spatial dependency between two variables z_p and z_q (subindices "p" and "q" stand for the time steps) and yields the cross variogram:

$$\hat{\gamma}_{pq}(h) = \frac{1}{2\,N_h} \sum_{k=1}^{N_h} \left[\left(z_p(x_k) - z_p(x_k + h) \right) \left(z_q(x_k) - z_q(x_k + h) \right) \right]. \tag{2.3}$$

This property is defined as half the expectation of the product of the increments of two variables (*Wackernagel*, 2003). Obviously, if $z_p = z_q = z$, then Eq. (2.3) reduces to Eq. (2.1). Figure 2.3 shows an example of an experimental cross-variogram and the fitted spherical model. Interpretations of direct and cross-variograms will be discussed in Sect. 2.3.1.

The gstat R package function fit.lmc() was employed to fit the direct and cross variogram models using weighted least square fitting (*Pebesma*, 2004). The default method uses weights N_h/h^2. The initial model parameters of the spherical variogram model were set as follows: $c_1 = 0.1$ mm, $a = 70$ km and $c_0 = 0.02$ mm. Function fit.lmc() also ensures that the system is positive definite (*Pebesma*, 2004).

2.2.3 Conditional simulation

Conditional simulation is a geostatistical method, which generates multiple realisations that all reasonably match the sample statistics (variogram model) and exactly match the conditioning data (*Goova erts*, 1997). As such, conditional simulation is a useful tool to model and quantify spatial uncertainty of a variable such as precipitation (e.g. *Clark and Slater*, 2006; *Schuurmans*, 2008; *Vischel et al.*, 2009; *AghaKouchak et al.*, 2010; *Grimes and Pardo-Iguzquiza*, 2010; *Renard et al.*, 2011). Time-independent univariate conditional simulations depend on the spatial observations by

rain gauges for a given simulation hour. However, by performing time-dependent multivariate conditional simulations, it becomes possible to simulate rainfall fields conditional on both previously simulated precipitation fields as well as on the rainfall observations by rain gauges. This approach enables to introduce temporal coherence for each simulated grid point over time.

In the current study, the gstat R package (*Pebesma*, 2004; *Rossiter*, 2007; *R Development Core Team*, 2013) was used to simulate conditional precipitation fields. For a comprehensive overview of the theory behind conditional (sequential Gaussian) simulations we refer to *Goova erts* (1997). Here, only a brief summary is presented:

Initially, a normal transformation of rainfall data is carried out. Then, the simulation is performed on the transformed dataset according to following steps (*Goova erts*, 1997):

1. A random path throughout all the grid nodes is defined, where all nodes are visited only once.

2. At each grid note, given the variogram model, a random number is drawn from a Gaussian distribution with parameters equal to the kriging prediction and variance. This number is added to the dataset used to condition the subsequently simulated grid nodes.

3. After an estimate is obtained for all grid nodes, the back-transformation of the simulated normal values to the original rainfall distribution is performed.

By performing these steps, one time-independent univariate ensemble realisation per time step is generated. Other realisations can be obtained using different random paths over the simulation grid domain.

For the purpose of computational stability, we further focus on rainy periods, which are defined as a cluster of consecutive rainy hours, for which each individual hour satisfies a minimal intensity condition. More specifically, the mean of all rain gauge observations should have at least a minimum value of 0.1 mm and the maximum individual observation has to exceed 0.5 mm. Additionally, to prevent computational instability, rain gauges with zero rainfall are set to a small value of 0.05 mm.

Then, time-dependent multivariate conditional simulations are carried out for each rainy period according to the following steps:

1. Initially, N ensemble realisations conditional on the rain gauge data and the variogram model are simulated for the first hour of each rainy period (time-independent univariate conditional simulation).

2. For the following simulation hours, the multivariate simulation for realisation j and time t is conditioned on: (a) rain gauge observations at time t given the fitted direct variogram model and (b) previously simulated realisations j at times $(t-1,\ldots,t-M)$, where M is simulation memory, given the cross-variogram models. This means that the observed precipitation by rain gauges at times $(t-1,\ldots,t-M)$ is substituted by the simulated fields, which encounter all the points of the entire simulation grid.

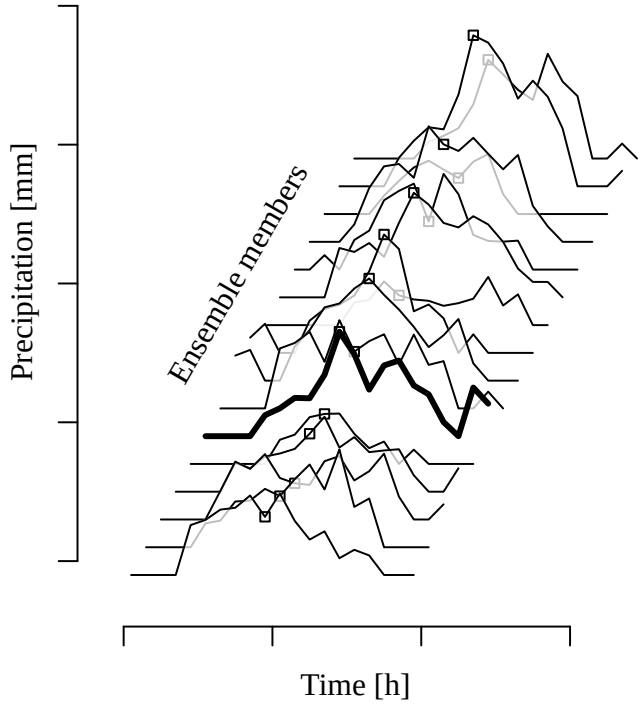


Figure 2.4: Two types of evaluating the statistics of precipitation ensembles: hydrological event-based approach (bold line for one ensemble realisation) and data assimilation across-ensemble approach (□ for one time instant).

2.2.4 Mathematical notation

Throughout this chapter, the following notations are used. A time series of rainfall realisations at the n-th pixel to be simulated is defined using the following matrix:

$$\mathbf{R}_{j,t}^{n,m} = \begin{pmatrix} R_{1,1} & R_{1,2} & \cdots & R_{1,T} \\ R_{2,1} & R_{2,2} & \cdots & R_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ R_{J,1} & R_{J,2} & \cdots & R_{J,T} \end{pmatrix} \tag{2.4}$$

where m is the simulation memory scenario, j is an index of the ensemble realisation and t stands for time. J is the ensemble size and T is the duration of the rainy period.

In general, two types of approaches can be used to disentangle the uncertainty in variability. The first way of looking at one precipitation realisation is to evaluate and quantify its corresponding statistics for the whole rainfall event. The variability of a single realisation over time during an event (bold line in Fig. 2.4), giving rise to uncertainty in the accumulated rainfall, is a typical source of uncertainty dealt with by hydrometeorologists (e.g. *Mandapaka et al.*, 2010; *Kirstetter et al.*, 2010), catchment hydrologists and rainfall-runoff modellers (e.g. *Brauer et al.*, 2011). An alternative approach is to quantify the uncertainty across the ensemble for each individual time step (□ in Fig. 2.4), which is more of interest for hydrologists employing Kalman filtering approaches in flood forecasting (e.g. *Weerts and El Serafy*, 2006; *Clark et al.*, 2008b).

Having defined the **R** matrix according to Eq. (2.4), we can derive the first two central moments of **R** from two different perspectives: (1) across-ensemble (having index J, Eqs. 2.5 and 2.6) and (2) event-based (along the time axis having index T, Eqs. 2.7 and 2.8):

$$\hat{\mu}_{J,t}[\mathbf{R}] = \frac{1}{J} \sum_{j=1}^{J} R_{j,t} \tag{2.5}$$

$$\hat{\sigma}^2_{J,t}[\mathbf{R}] = \frac{1}{J} \sum_{j=1}^{J} \left(R_{j,t} - \hat{\mu}_{J,t}[\mathbf{R}]\right)^2 \tag{2.6}$$

$$\hat{\mu}_{T,j}[\mathbf{R}] = \frac{1}{T} \sum_{t=1}^{T} R_{j,t} \tag{2.7}$$

$$\hat{\sigma}^2_{T,j}[\mathbf{R}] = \frac{1}{T} \sum_{t=1}^{T} \left(R_{j,t} - \hat{\mu}_{T,j}[\mathbf{R}]\right)^2. \tag{2.8}$$

The uncertainty within the precipitation ensemble can be expressed by the coefficient of variation (CV), the ratio of the standard deviation of the dataset to its mean, which represents a normalised dispersion and enables comparison between the generated ensembles for different scenarios (e.g. *Wackernagel*, 2003):

$$\mathrm{CV} = \frac{\hat{\sigma}}{\hat{\mu}}. \tag{2.9}$$

2.2.5 Experimental setup

To understand how the direct and cross-variograms reflect the differences in spatial variability between different precipitation fields, two examples of experimental and modelled variograms are analysed for the case of a synthetic circular-shaped rainfall cell moving over a 145 km × 145 km grid with 1 km × 1 km resolution (Fig. 2.5). As part of this case, the effect of rain gauge density on the experimental and modelled variograms is addressed and includes sampling either from all 21 025 grid pixels (a very dense synthetic observation network) or using only the 42 rain gauge pixels (actual real-world network). As such, both the impacts of using a sparse gauge network and of temporal correlation can be identified.

Next in this study, sensitivity analyses are carried out for a number of scenarios, which enable one to obtain better understanding of individual contributions of uncertainty in the synthetic simulated fields. The sensitivity analyses encompass four aspects:

1. *Time*: eight types of simulations with, given the time lag, simulation memories of 0–7 h.

2. *Advection speed*: five synthetic circular-shaped rainfall cells with an area of about 4100 km^2 (72.5 km diameter), an intensity of 2 mm h^{-1} and moving at different advection speeds of 6, 8, 11, 17, 25 km h^{-1} over the simulation domain (Fig. 2.5). The duration of the rainfall events is 18, 12, 9, 6 and 4 h, respectively. As such, the dimensions of these synthetic rainfall cells are similar to those observed within the region (Hazenberg et al., 2011).

3. *Area*: six synthetic nested sub-catchments over which the analysis is carried out (Fig. 2.6a).

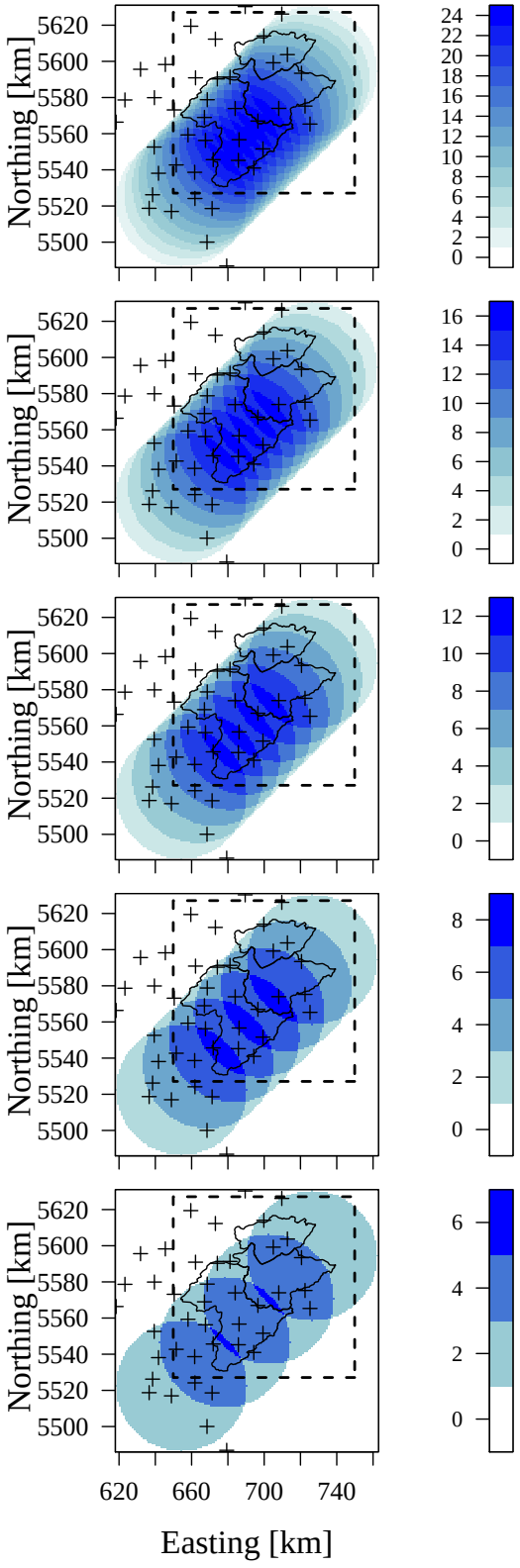


Figure 2.5: Sensitivity analysis. Five synthetic rainfall events of circular shape with an intensity of 2 mm h^{-1} and advection speeds of 6, 8, 11, 17, 25 km h^{-1} and durations of 18, 12, 9, 6 and 4 h, respectively (from top panel to bottom panel) over a 145 km × 145 km grid with 1 km × 1 km resolution. The figures show precipitation sums for the hypothetical situation, where the rain cell remains stationary within each hour and then instantaneously moves to its subsequent position. This figure is used only for illustration, the sums were not used in the analyses presented in this chapter. The plusses are rain gauges and the dashed box delineates the 100 km × 100 km simulation domain.

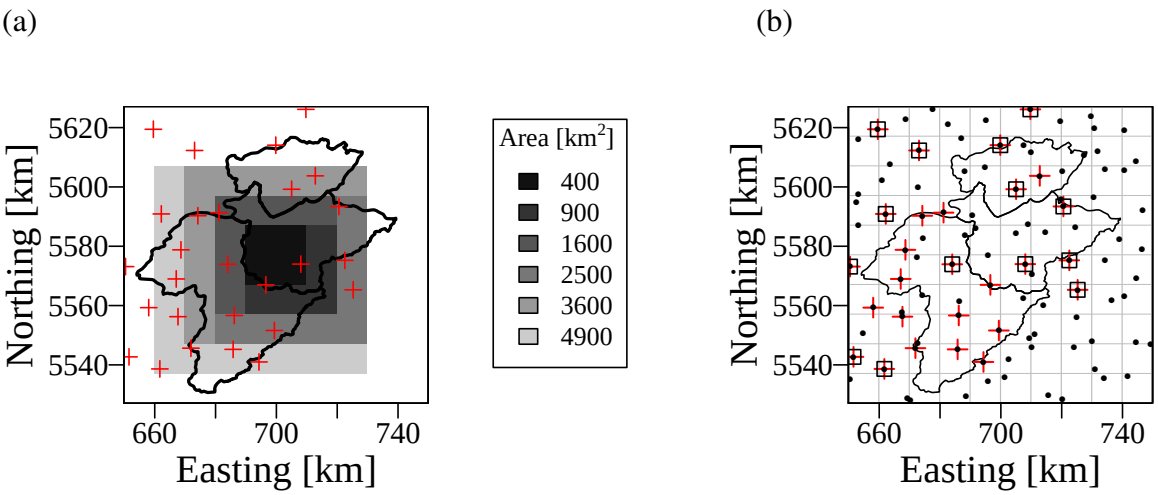


Figure 2.6: (a) Six nested "sub-catchments" with increasing catchment areas. Rain gauges are indicated by plusses. (b) Three types of rain gauge densities: dots ($N = 100$), red pluses ($N = 27$) and squares ($N = 14$). Extent of figure corresponds to the black dashed box in Fig. 2.5 and the grey lines show the 10 km × 10 km grid resolution.

4. *Observation density*: three types of rain gauge densities (Fig. 2.6b): (1) the actual observation network consists 27 rain gauges, (2) the reduced network has 14 rain gauges and (3) the complete synthetic network has 100 rain gauges. Removal of the rain gauges from the actual observation network follows a method (*Goudenhoofdt and Delobbe*, 2009), which keeps the spatial distribution of remaining gauges as uniform as possible. First, the sum of the inverse distances between the four nearest gauges is calculated for each gauge and then half of the gauges with the highest values are removed. The complete synthetic network for the catchment's pixels without any real rain gauge is obtained by randomly drawing x- and y-coordinates from a uniform distribution.

These aspects are considered as follows. Initially, the synthetic circular-shaped rainfall cell with constant intensity is advected over the catchment at five different advection speeds. The three types of rain gauge densities are used to sample the synthetic rainfall. Then, these synthetic point observations are employed in the simulations of the spatial rainfall fields. Both temporal and spatial correlations, which need to be taken into account, are expressed by the variogram models. Finally, the generated rainfall fields are analysed for different simulation memories and catchment sizes. By carrying out such a sensitivity analysis of a very simplistic rainfall cell with known spatial statistics, constant rainfall intensity and known advection speed, we are able to eliminate several sources of uncertainty which would arise with the real-world data. The purpose of this sensitivity analysis is to obtain a better understanding of the spatial and temporal scales involved in the rainfall-runoff transformation before turning to the real-world case.

Because of the higher computational costs of the time-dependent multivariate conditional simulations at high resolution grids, the simulation domain is reduced for that purpose to 100 km × 100 km with a 10 km × 10 km raster resolution (dashed box in Fig. 2.5). The analysis includes 24 ensemble realisations and the length of the rainy periods varies between 4 and 18 h, depending on the advection speed.

Finally, the real-world experiment will focus on the three events described by *Hazenberg et al.*

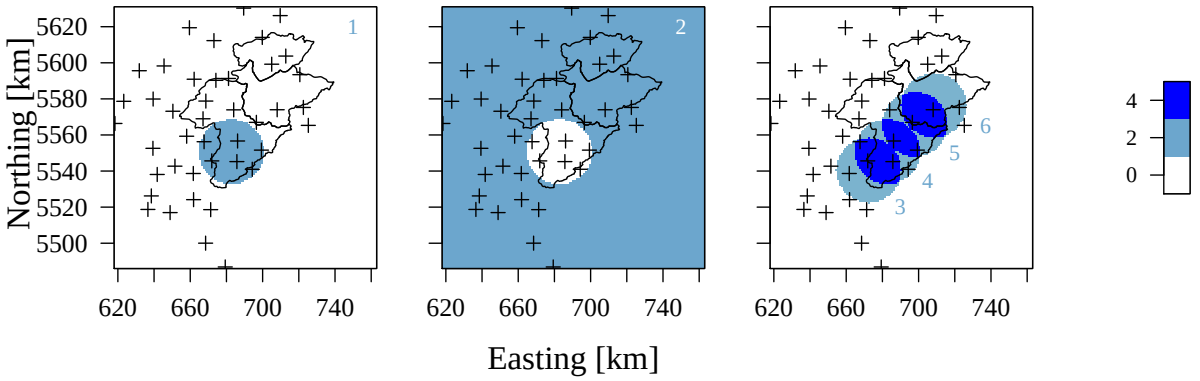


Figure 2.7: Sensitivity analysis. Circular-shaped rainfall events consisting of six rainfall fields 1–6. Raster resolution is 1 km × 1 km. Plusses show the rain gauges and black line delineates the Upper Ourthe, Amblève and Vesdre catchments. This figure is used only for illustration, the sums were not used in the analyses presented in this chapter.

Table 2.2: Sensitivity analysis. Fitted sills (mm) using the spherical model for direct and cross-variograms for rainfall fields 3–6 (see Fig. 2.7) derived for the dense synthetic observation network (Fig. 2.8c).

		Rainfall field			
Rainfall field		3	4	5	6
	3	0.31			
	4	0.16	0.32		
	5	−0.03	0.08	0.33	
	6	−0.03	−0.03	0.16	0.31

(2011) (see Sect. 2.2.1). For these events, the impact of time and area as described previously are analysed using the actual observation network (27 rain gauges). Moreover, to verify the accuracy of the method at unobserved locations, cross-validation is carried out by thinning the observation network with 50 %.

2.3 Results

2.3.1 Interpretation of direct and cross variograms

Empirical and modelled spherical direct and cross-variograms (Sect. 2.2.2) are calculated for two cases of a synthetic circular-shaped rainfall event with an intensity of 2 mm h^{-1}. The first case consists of two rain fields (see 1 and 2 in Fig. 2.7), which are complementary. The second case consists of four advected rainfall fields (Fig. 2.7, right panel). The intersection of the third and the fourth rainfall field (Fig. 2.7, right panel) yields exactly half of the rainfall area. Since the third and the fifth rainfall fields are tangent, the intersection of the fourth and the fifth rainfall field covers around 28 % of the rainfall area (*Glassner*, 1998). Moreover, the intersection of the fifth and the sixth rainfall field yields again exactly half of the rainfall area. We chose a rather small circular-shaped rainfall cell with a 36 km diameter, to minimise the impact of boundary effects.

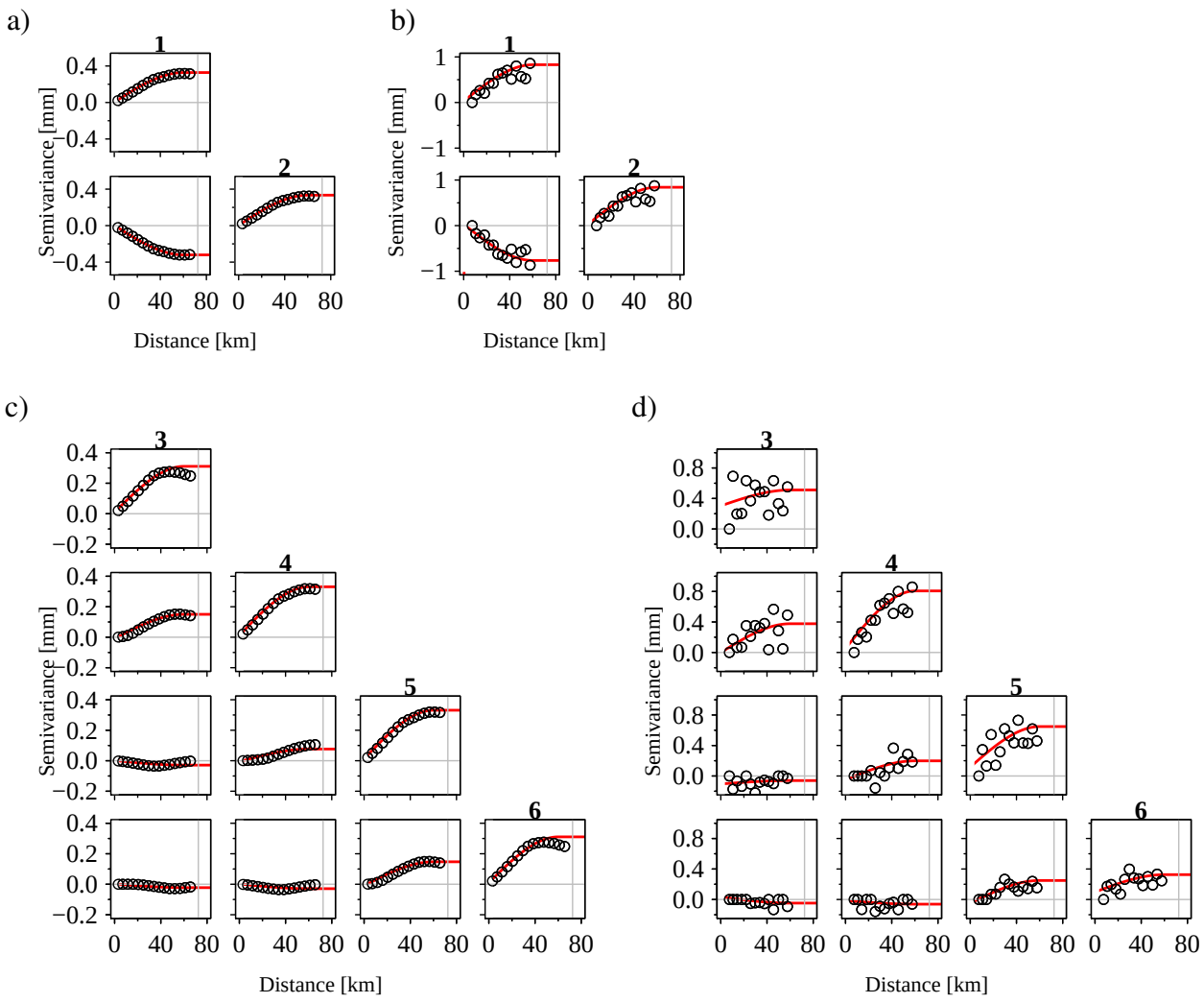


Figure 2.8: Sensitivity analysis. Direct and cross-variograms for rainfall fields 1 and 2 (Fig. 2.7) derived from all grid points (a) and only the rain gauge grid points (b). Direct and cross-variograms for rainfall fields 3–6 (Fig. 2.7) derived from all the grid points (c) and only the rain gauge grid points (d). Grey vertical line shows the range corresponding to twice the diameter of the rainfall circle.

Figure 2.8a and b show the direct and cross-variograms of the two complementary rainfall fields in Fig. 2.7 (left and middle panels). These have been square-root transformed, for the two sampling densities: a dense synthetic network (21 025 points covering all raster pixels) and the real world network (42 rain gauges). We can observe that the two direct variograms are identical and symmetrical with respect to their cross-variogram. This holds for both sampling densities. As expected, the real-world rain gauge network has a larger sill and a larger scatter in the empirical variograms than the much denser synthetic observation network.

For the advected cell of Fig. 2.7 (right panel), the direct and cross-variograms are shown in Fig. 2.8c and d. For a decreasing intersected area between the two rainfall fields, the fitted spherical sills decrease proportionally using the dense synthetic network (Table 2.2). Zero overlapping area (i.e., three combinations: rainfall fields 3 and 5, 3 and 6, 4 and 6), gives a sill of about zero. For the real-world network similar behaviour can be observed, although the scatter in the fitted models through the empirical variograms is much larger. Note that even though the circular-shaped rainfall

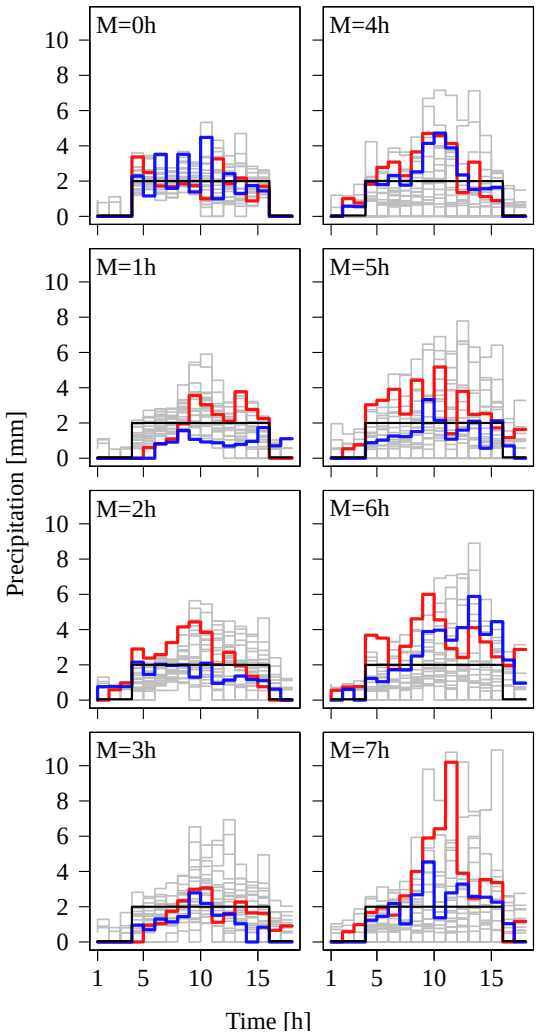


Figure 2.9: Sensitivity analysis. Precipitation ensemble (24 realisations in grey, from which two realisations are shown in red and blue) for eight types of simulation scenarios given a simulation memory M of 0–7 h, for a pixel located in the centre of the simulation domain. Black line stands for the synthetic rain gauge observation at that location. Rainfall scenario for the advection speed of 6 km h^{-1} (see Fig. 2.5).

cell is identical for all rainfall fields, the shape of the empirical variograms differs from each other, because the rain gauge configuration is not spatially uniform and effectively changes for each rainfall field.

2.3.2 Conditional simulations: sensitivity analysis

As a first example, the impact of advection speed and the number of hours used as part of the conditional simulation was identified for the setup presented in Fig. 2.5. As such, precipitation is sampled by the real-world rain gauge network ($N = 27$ in Fig. 2.6), while a total of up to eight hours are used as part of the simulation memory (0–7 h). An example of such a simulation for one pixel is shown in Fig. 2.9. We can observe that for the time-independent univariate case (i.e., conditioned on 0 h of previously simulated fields), there is no consistency for ensemble realisations over time, since

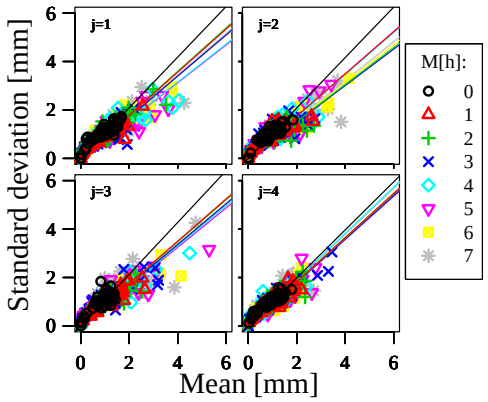


Figure 2.10: Sensitivity analysis. Scatter plot of $\hat{\mu}_{T,j}$ and $\hat{\sigma}_{T,j}$ with fitted coefficients of variation (CV_T) for four ensemble realisations, an advection speed of 6 km h^{-1} and a catchment area of 4900 km^2. Different points in each panel are for individual points (pixels) of the catchment.

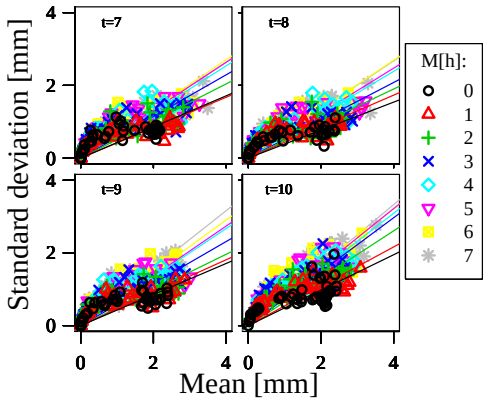


Figure 2.11: Sensitivity analysis. Scatter plot of $\hat{\mu}_{J,t}$ and $\hat{\sigma}_{J,t}$ with fitted coefficients of variation (CV_J) for four time steps, an advection speed of 6 km h^{-1} and a catchment area of 4900 km^2. Different points in each panel are for individual points (pixels) of the catchment.

no information between individual time steps is taken into account. However, for the time-dependent multivariate cases (i.e., conditioned on 1–7 h of previously simulated fields), the temporal consistency for ensemble realisations becomes more clear. Overall, the spread in simulated precipitation increases when a larger number of previously simulated fields is included as part of the simulations.

From the event-based perspective (see bold line in Fig. 2.4), the scatter plots between the simulated mean precipitation over time ($\hat{\mu}_{T,j}$) and the corresponding standard deviation ($\hat{\sigma}_{T,j}$) for the first four ensemble realisations are shown in Fig. 2.10 separately, where $\hat{\mu}_{T,j}$ and $\hat{\sigma}_{T,j}$ are plotted for all the individual pixels within the 4900 km^2 catchment (see Fig. 2.6b). The panels indicate that the spread in $\hat{\mu}_{T,j}$ gradually increases from the time-independent univariate conditional simulation to the most complex time-dependent multivariate scenario. This is in agreement with the simulation results shown in Fig. 2.9. Because we are not only interested in the mean simulated values, but also in their temporal variability, standard deviations are plotted against the corresponding means (Fig. 2.11). The slope of the fitted linear regression line represents the mean temporal coefficient of variation. The coefficients

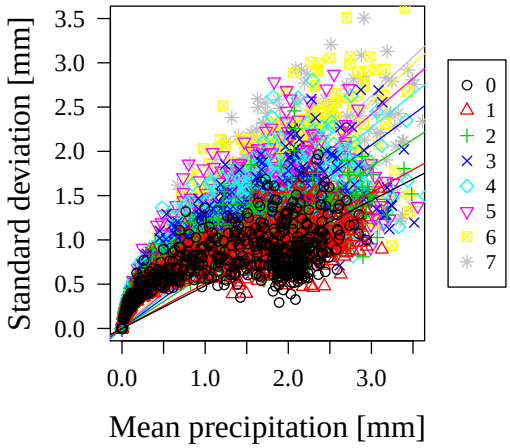


Figure 2.12: Sensitivity analysis. Scatter plot $\hat{\mu}_{J,t}$ and $\hat{\sigma}_{J,t}$ lumped over time with fitted coefficient of variations (CV_J) for the rainfall scenario with an advection speed of $6\,km\,h^{-1}$ (see Fig. 2.5) and a catchment area of $4900\,km^2$. Different points in the panel are for individual points (pixels) of the catchment.

of variation are gradually decreasing for longer simulation memories, which indicates lower temporal variability and larger temporal coherence for longer simulation memories.

For the across-ensemble perspective (see □ in Fig. 2.4), the scatter plots between simulated mean precipitation across-ensemble ($\hat{\mu}_{J,t}$) and their corresponding standard deviations ($\hat{\sigma}_{J,t}$) for the four time steps ($t = 7$, 8, 9, 10) are shown in Fig. 2.11, where $\hat{\mu}_{J,t}$ and $\hat{\sigma}_{J,t}$ are plotted for all individual pixels within the $4900\,km^2$ catchment. The figures reflect a rather constant spread in $\hat{\mu}_{J,t}$ for all simulation memories during those four time steps. Nevertheless, the fitted coefficients of variation are gradually increasing for longer simulation memories. This is in line with the higher ensemble spread for longer simulation memories (see Fig. 2.9).

Since the purpose of this chapter is to assess the impact of precipitation uncertainty estimation across the ensemble (i.e., $\hat{\mu}_{J,t}$ vs. $\hat{\sigma}_{J,t}$), which is of key interest especially for hydrological data assimilation applications, the next step is to evaluate it in a lumped manner over all time steps. This was done by overlapping the individual sub-plots (partially depicted for $t = 7$, 8, 9, 10 in Fig. 2.11) for all time steps. An example is shown in Fig. 2.12, where the coefficients of variation increase with rainfall simulations conditioned on longer simulation memories.

The behaviour of the coefficients of variation (CV_J) for individual catchment sizes, advection speeds and simulation memories is summarised in Fig. 2.13. We can observe that for a larger advection speed the coefficient of variation rises faster and approaches its upper level earlier than for slower moving systems. Note that for the faster events, with advection speeds of 25 and $17\,km\,h^{-1}$ lasting four and six hours, respectively (Fig. 2.5), the coefficients of variation can be obtained only up to the three and five previously simulated hours. Additionally, the catchment size does not have a large influence on the coefficients of variation except for very small catchment sizes consisting of only few pixels, where the estimation is significantly affected by sampling uncertainty. Because small catchments are nested, most variability is smoothed out for larger catchments sizes.

The aforementioned examples were based on the real-world rain gauge network ($N = 27$). When the number of rain gauges is decreased to the half of its original density ($N = 14$), the general shape

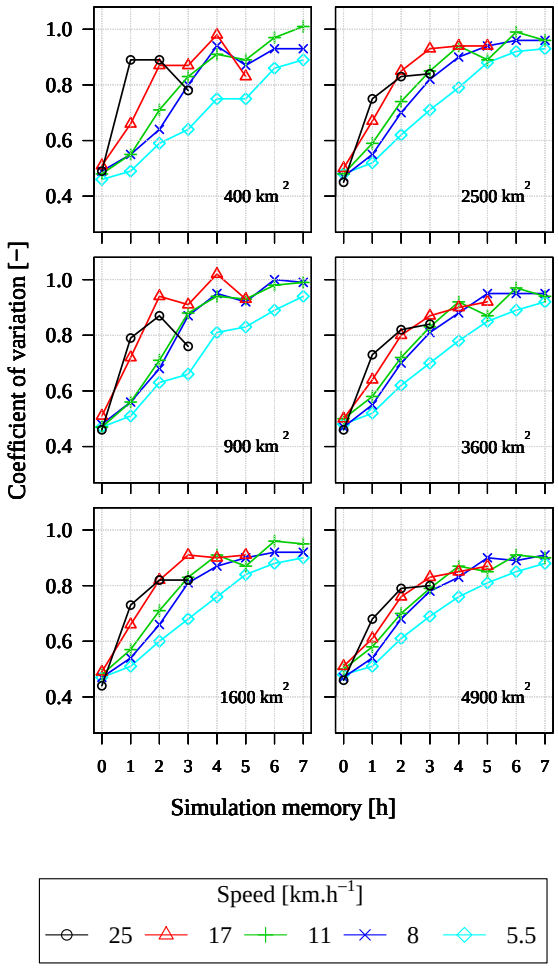


Figure 2.13: Sensitivity analysis. Coefficient of variation (CV_J) for different catchment sizes and advection speeds.

Table 2.3: Sensitivity analysis. Fitted y-intercept and horizontal asymptote values of coefficients of variation (CV_J, in Fig. 2.13) for different rain gauge networks (Fig. 2.6b).

Number of gauges	y-intercept	Horizontal asymptote
14	0.5	1
27	0.45	0.9
100	0.4	0.7

of the fitted spherical variogram through the estimated coefficients of variation remains very similar. However, both y-intercept and horizontal asymptote values become higher (Table 2.3). This indicates a slight increase in the across-ensemble variability. For a dense synthetic network ($N = 100$) the opposite occurs. Both the y-intercept and horizontal asymptote values decrease, which means a lower across-ensemble variability. These results suggest that the method is rather robust. Additionally, the effect of thinning the observation network on the mean simulated error is provided in the

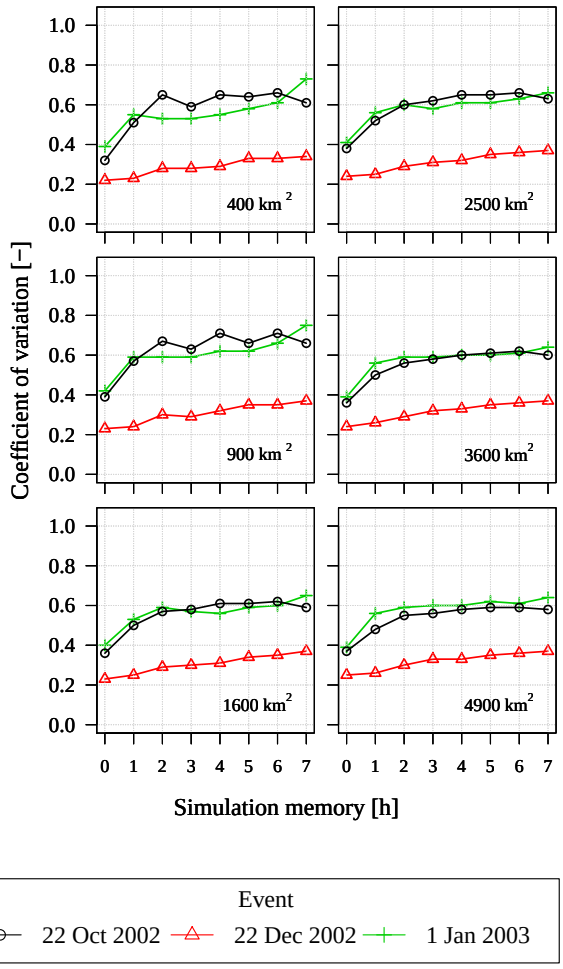


Figure 2.14: Real-world experiment. Coefficient of variation (CV_J) for different catchment sizes and three rainfall events.

following section, where the cross-validation at unobserved locations is carried out for the real-world experiment.

2.3.3 Conditional simulations: real-world experiment

The real rain gauge observations have been analysed within the same conditional simulation framework as was done within the sensitivity analysis. The resulting coefficients of variation (CV_J) for the different catchment sizes and simulation memories are shown in Fig. 2.14. They correspond well with the sensitivity analysis. For the two fast-moving systems (22 October 2002 and 1 January 2003), there is a steep increase in the across-ensemble spread (CV_J), which becomes more or less steady after simulation memories of 1–2 h. This means that no further uncertainty is added to simulated precipitation fields by conditioning on more than two hours of previous simulations. For a large-scale stratiform system (22 December 2002) moving very slowly, a gentle rise in CV_J is observed. For this event, the horizontal asymptote is reached when a simulation memory of >5 h is taken into account. Moreover, these values of CV_J are considerably smaller than for the faster systems.

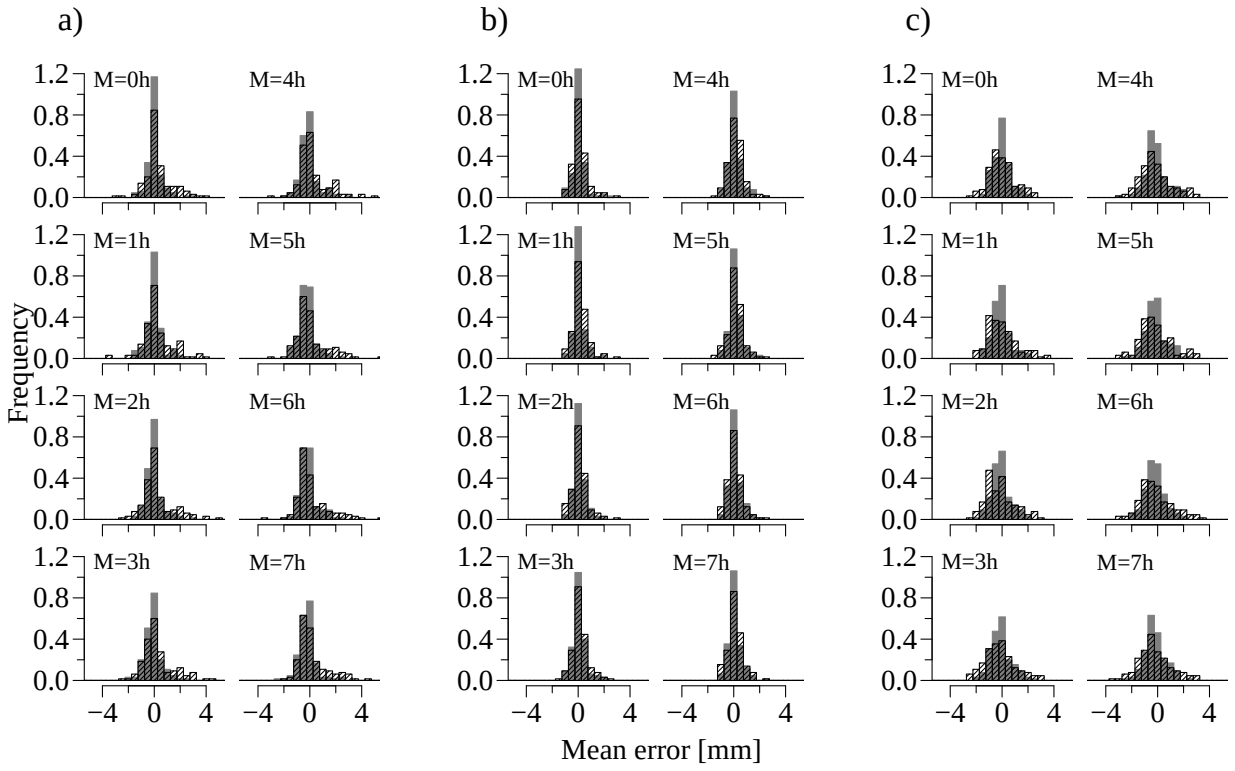


Figure 2.15: Validation for three real-world events for simulation memories ranging from 0 to 7 h: (a) 22 October 2002, (b) 22 December 2002 and (c) 1 January 2003. Histograms of the validation mean errors at 13 rain gauges (dashed histograms). Histograms of the simulation mean errors for the same 13 rain gauges (grey histograms).

To verify the accuracy of the presented method, cross-validation was carried out in the terms of the mean error. The mean error is defined as the difference between the rain gauge observation and the corresponding across ensemble mean ($\hat{\mu}_{J,t}$, Eq. 2.5). The rain gauge observations employed in the validation were independent from the data used for simulation. To simulate the precipitation fields, 14 rain gauges out of the complete observation network of 27 rain gauges were used (Fig. 2.6b). The remaining 13 rain gauges were kept for validation and their mean errors were calculated for all time steps and for all eight simulation memories (dashed histograms in Fig. 2.15). Additionally, we compared these validation mean errors with the simulation mean errors at the same 13 locations. These simulation mean errors were obtained by simulating precipitation fields using all 27 rain gauges (grey histograms in Fig. 2.15). Figure 2.15 shows that the validation mean errors at the unobserved locations are unbiased and have a consistent behaviour over all simulation memories (dashed histograms in Fig.2.15). Furthermore, the histograms of the validation mean errors have smaller peaks than the mean errors of the simulated precipitation fields. This increase in uncertainty is to be expected, because in the validation only half of the rain gauge data are used to simulate spatial precipitation fields. Note that the spread in histograms agrees well with the corresponding standard deviations shown in Table 2.1.

2.4 Synthesis and discussion

The overall temporal correlation structure of the simulated precipitation field can be quantified using the first order autocorrelation coefficient r_{1h}, which expresses the correlation of a precipitation time series for a time lag of 1 h. Figure 2.16 shows box plots of r_{1h} for time series of areal precipitation for different catchment sizes and simulation memories. For a robust investigation of the autocorrelation, it is preferable to have long time series. Therefore, we decided to use the data for the slowest synthetic event with a duration of 18 h (see Fig. 2.5). It can be observed that r_{1h} increases when moving from time-independent univariate ($M = 0$) to the time-dependent multivariate conditional simulations ($M = 1$–7 h). Nevertheless, the major difference between the univariate simulations (grey boxplots in Fig. 2.16) and the multivariate simulations (white boxplots in Fig. 2.16) decreases for larger catchments. This is a direct result of the relatively small size of the rain cell with respect to the catchment area.

The across-ensemble uncertainty was quantified using the lumped CV_J, and its shape for both the sensitivity analysis (Fig. 2.13) as well as the real-world experiment (Fig. 2.14) clearly resembles the shape of the spherical variogram (Eq. 2.2). By fitting the spherical model, the range can be obtained, which represents a simulation memory threshold of the system, after which no additional precipitation uncertainty is added by including more previous information. The fitting of the range is done using the gstat R package (*Pebesma*, 2004) and is an analogy to the method explained in Sect. 2.2.2.

For the sensitivity analysis, this leads to an inverse relation between the advection speed and its corresponding fitted range [h] (Fig. 2.17):

$$\text{Range} \times \text{Speed} \approx 50\,\text{km}. \tag{2.10}$$

This result indicates that for typical advection speeds ($> 20\,\text{km}\,\text{h}^{-1}$) no uncertainty in terms of across-ensemble spread is added to the simulated precipitation fields by conditioning it on more than two previous hourly simulations.

We need to bear in mind, however, that the synthetic case analysed here is the most simplistic example of a precipitation cell, which assumes a known and constant advection speed, has a rather small dimension, and results in a constant rainfall intensity. For real-world examples, on the other hand, a much higher spatial and temporal variability in the dynamics of precipitation systems can be expected. This means that the effective ranges of previous information are expected to increase. This corresponds well with the results obtained for the real-world experiment, which are shown in Fig. 2.17. To quantify the apparent uncertainty in the fitted ranges for the three real rainfall events, the values of the ranges for 80 %, 90 % and 100 % of the partial sills are shown. For the large-scale stratiform system (red triangles), the fitted ranges vary around 5–7 h, which is considerably longer than for the two faster systems. This can be caused by the combination of (1) the gentle increase and in general lower values of coefficient of variation, which are about half of the values for the two remaining events (recall Fig. 2.14) and (2) the size of the observed precipitation system, which for the real-world case has a larger dimension. Because of this latter property, we repeated the sensitivity analysis for rainfall cells with twice the original diameter size, which made its area four times larger. For these cases, indeed a larger coefficient of 63 km is obtained (grey line in Fig. 2.17). Finally, when we refer back to Fig. 2.2, which shows the exponential decay of the autocorrelation of the areal

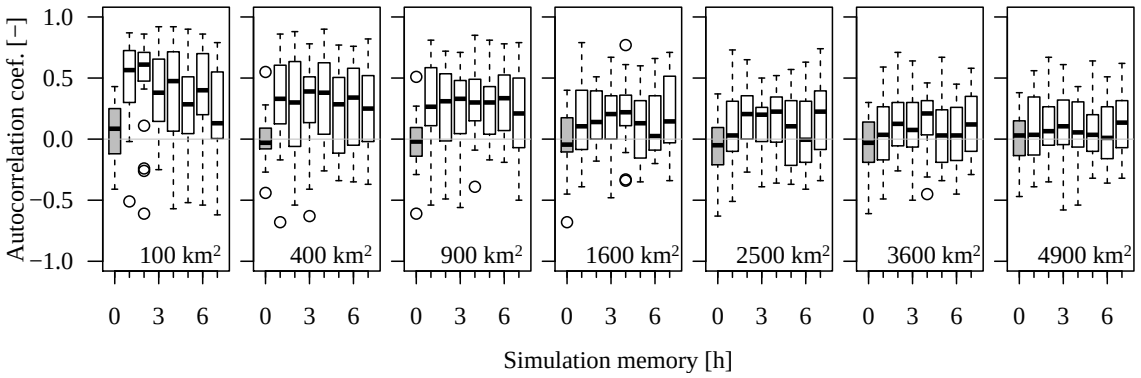


Figure 2.16: Sensitivity analysis. Boxplots of the first-order autocorrelation coefficients (24 realisations) for the time series of the difference between the catchment's averaged areal precipitation and the ensemble mean for different catchment sizes and simulation memories (time-independent univariate simulation in grey, time-dependent multivariate simulations in blank), rainfall scenario with an advection speed of 6 km h^{-1} and duration of 18 h (see Fig. 2.5).

precipitation estimates, we can conclude that considering a simulation memory of about two to three hours is relevant for the presented stratiform winter rainfall events.

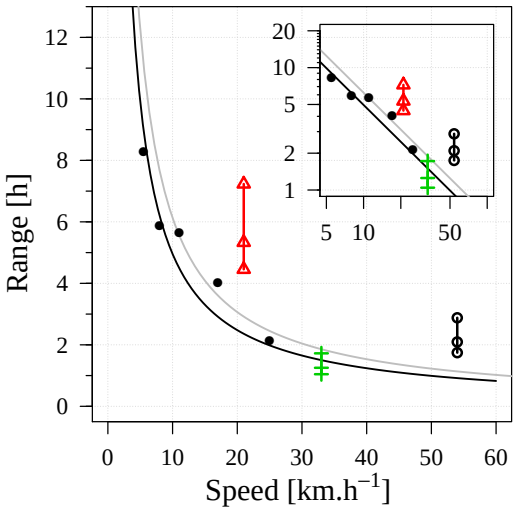


Figure 2.17: Nonlinear relation between the advection speed and the threshold range for a catchment area of 4900 km^2. Black filled circles represent five synthetic circular-shaped rainfall cells. Black line delineates Eq. (2.10). Black open circle, red triangle and green plus show three real-world rainfall events (see Fig. 2.14) with the values of the ranges for 100 %, 90 % and 80 % (from top to bottom) of the corresponding partial sills. Grey line delineates the fitted nonlinear relation for synthetic rainfall cells with double diameter. The identical figure with logarithmic axes is given in the inset.

Overall, for the time-dependent multivariate conditional simulations with longer simulation memories, we observed a larger across-ensemble spread. The commonly defined time-independent rainfall perturbations used for the hydrological data assimilation applications have noise errors with a stan-

dard deviation up to 50 % of the observed precipitation (e.g. *Pauwels and De Lannoy*, 2006; *Weerts and El Serafy*, 2006). This corresponds well with our simulations, in which the maximum slope of the fitted coefficients of variation is about 0.5 (Fig. 2.11). However, in comparison with the aforementioned references, we were able to additionally capture the temporal coherence for each realisation in space.

Moreover, conditional simulation methods increase computational costs quite dramatically in comparison with interpolation methods. Fortunately, this problem can be partly circumvented by decreasing the temporal (Δt [min]) or spatial (Δr [km]) resolution of the simulation model. From an applied hydrological point of view, an hourly time step is usually recommended for regions with an area of $\sim$10 000 km^2 (*Berne et al.*, 2004). Moreover, the choice for a rather coarse 10 km $\times$ 10 km grid resolution, as was chosen in this study, can be supported by the analysis carried out by *Berne et al.* (2004), who reported the spatial rainfall resolution to be 4.5 $\sqrt{\Delta t}$, which yields a decorrelation distance of about 35 km for $\Delta t = 60$ min.

A limiting factor of this study can be that the spatial anisotropy of the rainfall field as well as the local topography were not considered. Additionally, we did not use other sources of precipitation measurements, such as weather radar. Therefore, obtaining precipitation ensembles from radar remains an important continuation of this study. The main benefit of using radar is that one obtains much more information on the spatial characteristics of both the precipitation field and type (*Hazenberg et al.*, 2011). It is then possible to take two different approaches: (1) the rain gauge perspective, where we imagine that that the weather radar data is only used to provide information on where it rains and which gauges are specifically to be used to generate a variogram (since they belong to the same region), (2) the radar perspective, where the gauges are used to correct the radar for any remaining bias, while the uncertainty in the precipitation field is obtained from the volumetric radar information. We will present some ideas on these issues and approaches related to these issues in future publications.

2.5 Summary and conclusions

In this chapter, a rain gauge precipitation ensemble generator at hourly time step using time-dependent multivariate conditional simulations, which were made conditional on previous simulations back in time was developed. As such, a plausible way to generate temporal correlation structures for precipitation for each realisation over time was introduced. Next, we identified the uncertainty and the temporal correlation structures in the simulated fields based on (1) the number of previous simulation hours on which the new simulation is conditioned, (2) the advection speed of the rainfall event, (3) the size of the catchment considered and (4) the rain gauge density within the catchment.

The sensitivity analysis shows that for typical advection speeds of > 20 km h^{-1} no uncertainty in terms of across-ensemble spread lumped over time (expressed using the coefficient of variation) is added to simulated precipitation fields by conditioning them on more than one or two previous hourly simulations. In the real-world experiment, which exhibits a larger spatial and temporal variability, the time-dependent simulations require somewhat longer simulation memories. Additionally, by halving the observation network, i.e., using 14 rain gauges, the uncertainty in the sensitivity analysis increases only slightly. Furthermore, the cross-validation shows that the simulations at unobserved locations are unbiased and have a consistent behaviour over all simulation memories. Finally, the first-order

autocorrelation coefficient indicated the presence of temporal coherence in the time series of the areal precipitation using the time-dependent multivariate conditional simulation in comparison with the time-independent univariate conditional simulations. Nevertheless, this coherence was found to decrease with increased catchment area.

The presented technique to generate spatial precipitation ensembles can be easily implemented within a hydrological data assimilation framework to be used as an improvement over currently used simplistic approaches to perturb the interpolated point or spatially distributed estimates (as referred to in the introduction). As shown, using the time-dependent rainfall simulations with at least one hour of simulation memory, but preferably longer, we were able to reach this goal and obtain precipitation ensembles with temporal correlation structures that are plausible from a hydro-meteorological perspective. Therefore, the corresponding simulated spatially distributed model states produced by that rainfall ensemble should inherit this temporal aspect. The advantage of having the temporal coherence in model states is that it eliminates the need to smooth possible extreme state values, which can be the case when neglecting it. A hydrological application of the presented spatial precipitation ensemble generator is presented in Chapter 4.

Chapter 3

Distributed Evaluation of Local Sensitivity Analysis (DELSA), with application to hydrologic models

Abstract

This paper presents a hybrid local-global sensitivity analysis method termed the Distributed Evaluation of Local Sensitivity Analysis (DELSA), which identifies important and unimportant parameters and evaluates how model parameter importance changes as parameter values change. DELSA uses derivative-based "local" methods to obtain the distribution of parameter sensitivity across the parameter space. This promotes consideration of sensitivity analysis results in the context of simulated dynamics. This work presents DELSA, discusses how it relates to existing methods, and uses two hydrologic test cases to compare its performance with the popular global, variance-based Sobol' method. The first test case is a simple non-linear reservoir model with two parameters. The second test case involves five alternative "bucket-style" hydrologic models with up to 14 parameters applied to a medium-sized catchment (200 km^2) in the Belgian Ardennes. Results show that in both examples, Sobol' and DELSA identify similar important and unimportant parameters, with DELSA enabling more detailed insight at much lower computational cost. For example, in the real-world problem the time delay in runoff is the most important parameter in all models, but DELSA shows that for about 20% of parameter sets it is not important at all and alternative mechanisms and parameters dominate. Moreover, the time delay was identified as important in regions producing poor model fits, whereas other parameters were identified as more important in regions producing better model fits. The ability to understand how parameter importance varies through parameter space is critical to inform decisions about, for example, additional data collection and model development. The ability to perform such analyses with modest computational requirements provides exciting opportunities to evaluate complicated models as well as many alternative models.

This chapter is largely based on the manuscript: Rakovec, O., M. C. Hill, M. P. Clark, A. H. Weerts, A. J. Teuling, R. Uijlenhoet (2014), Distributed Evaluation of Local Sensitivity Analysis (DELSA), with application to hydrologic models, *Water Resour. Res., 50*, 1–18, doi:10.1002/2013WR-014063.

3.1 Introduction

The primary aim of sensitivity analysis is to identify how different components of a model affect model output. Results can affect model calibration, uncertainty evaluation and risk assessment (e.g., *Katz*, 1999; *Oakley and O'Hagan*, 2004; *Saltelli et al.*, 2008; *Kucherenko et al.*, 2009; *Göhler et al.*, 2013; *Plischke et al.*, 2013). Sensitivity analyses can be conducted to evaluate model structure and forcing (e.g., *Doherty and Welter*, 2010; *Gupta et al.*, 2012; *Foglia et al.*, 2013), yet most commonly, and in this work, the analysis is focused on evaluating how model parameters (also called factors) affect model output. Focusing on parameter sensitivity analysis is less limiting than it might seem because parameters can be defined to control many aspects of a model, and thus parameter sensitivity analysis methods can be used to understand the role played by, for example, model structure and forcing. Once parameters are defined, parameter sensitivity analysis is used to identify parameters important and unimportant to simulated processes, inverse modeling metrics, predictions, quantifying uncertainty, and evaluating risk (e.g., *Van Werkhoven et al.*, 2008; *Saltelli et al.*, 2008; *Plischke et al.*, 2013). Within these broad goals, parameter sensitivity analysis can be used to (a) detect when increasing model complexity can no longer be supported by observations and whether it is likely to affect predictions of interest (e.g., *Saltelli et al.*, 1999; *Van Werkhoven et al.*, 2008; *Rosolem et al.*, 2012; *Gupta et al.*, 2012; *Hill et al.*, 2013), (b) reduce the time of model calibration by focusing estimation efforts on parameters important to calibration metrics and predictions (e.g., *Anderman et al.*, 1996; *Hamm et al.*, 2006; *Zambrano-Bigiarini and Rojas*, 2013), (c) determine priorities for theoretical and site-specific model development (e.g., *Hill and Tiedeman*, 2007; *Saltelli et al.*, 2008; *Kavetski and Clark*, 2010), and (d) identify advantageous placement and timing of new measurements (e.g., *Musters and Bouten*, 2000; *Weerts et al.*, 2001; *Vrugt et al.*, 2001; *Tiedeman et al.*, 2003, 2004; *Tonkin et al.*, 2007; *Fienen et al.*, 2010). Of interest is also sensitivity analysis over different periods (e.g. *Wagener et al.*, 2003; *Cloke et al.*, 2007; *Herman et al.*, 2013a).

The considerable potential utility of sensitivity analysis methods is sometimes difficult to attain in practice because many sensitivity analysis methods currently popular in environmental modeling require considerable computational effort. To discuss this issue we use the common classification of sensitivity analysis methods: global and local.

Global methods calculate sensitivity measures using parameter samples obtained from a defined parameter space. For each parameter sample the model is run one or more times to obtain simulated results. Global methods provide stable results because they produce measures of parameter importance that are averaged over the range of the parameter space from which samples are obtained. However, of interest is that different global sensitivity analysis methods often lead to completely different importance rankings, as shown for example by *Pappenberger et al.* (2008a).

When considered from the perspective of uncertainty analysis, global methods can be moment-independent (e.g., *Park and Ahn*, 1994; *Chun et al.*, 2000; *Borgonovo*, 2007) or based on decomposition of the variance of the model output into variance contributors from individual parameters and parameter combinations (*Saltelli et al.*, 2008). The latter are considered in this work. Many global methods account for the effect of nonlinearity, including the effect of parameter interactions of predictions (e.g., *Box and Tiao*, 1992; *Saltelli*, 2002; *Van Werkhoven et al.*, 2008; *Rosolem et al.*, 2012).

A popular variance-based global method is the Sobol' method (e.g., *Sobol'*, 2001; *Saltelli*, 2002),

which has been applied in many environmental studies (e.g., *Hamm et al.*, 2006; *Pappenberger et al.*, 2006; *Cloke et al.*, 2007; *Tang et al.*, 2007; *Pappenberger et al.*, 2008a; *Van Werkhoven et al.*, 2008; *Saltelli et al.*, 2008; *Kavetski and Clark*, 2010; *Nossent et al.*, 2011; *Massmann and Holzmann*, 2012; *Herman et al.*, 2013a,b) and is often used in method comparisons (e.g., *Borgonovo*, 2006; *Cloke et al.*, 2007; *Pappenberger et al.*, 2008a). Thus, Sobol' can be used as a standard for comparison that allows alternative methods to be indirectly compared to each other. Many global methods, and especially the Sobol' method, require a large number of model runs (often 10 000 or more) and thus can be computationally demanding or prohibitive.

Alternative global methods such as FAST (*Cukier et al.*, 1973, 1975, 1978; *Saltelli et al.*, 1999), the method of Morris (MoM; *Morris*, 1991; *Zhan et al.*, 2013; *Herman et al.*, 2013b), regional sensitivity analysis (RSA; *Hornberger and Spear*, 1981; *Freer et al.*, 1996), and the delta method (*Borgonovo*, 2007) provide more frugal alternatives, but often remain a computational challenge: for example, FAST and the delta method require at least one solution at each sampling point and convergence properties are such that the number of sampling points is large; MoM is often conducted with multiple parameter increments for each parameter at each sampling point.

Local sensitivity analysis methods are typically based on the gradients (derivatives) of the model output with respect to parameter values evaluated at a single location in the parameter space (e.g., *Hill and Tiedeman*, 2007; *Oliver et al.*, 2008). Parameter interactions on predictions can be accounted for using, for example, the method described by *Sobol' and Kucherenko* (2010). The convenience of local methods has resulted in their considerable use (e.g., *D'Agnese et al.*, 1999; *Kunstmann et al.*, 2002). However, their applicability to nonlinear models, including models with spurious results, is of concern (*Saltelli et al.*, 2008) because single-point application of local methods can identify dramatically different important and unimportant parameters in different parts of the parameter space, and results could mislead modelers and users of model results. Comparisons of global and local methods, including *Helton* (1993), *Tang et al.* (2007), *Foglia et al.* (2007), *Kucherenko et al.* (2009), *Sobol' and Kucherenko* (2009), *Sobol' and Kucherenko* (2010), *Delenne et al.* (2012), and *Li et al.* (2013a) provide mixed results. For example, *Foglia et al.* (2007) used a hydrologic model and the non-local cross-validation method to show that local methods provided useful information on the sensitivities of 35 parameters. *Tang et al.* (2007) compared multiple sensitivity analysis methods applied to the lumped Sacramento soil moisture accounting model and found that local methods provided very different results from the global methods – however, the local methods were not scaled to be dimensionless, which potentially explains the different model sensitivities. *Delenne et al.* (2012) showed that despite the non-linearity of river flow processes simulated with a one-dimensional hydrodynamic model, a local approach yielded similar results to a global approach, even in cases of large parameter uncertainty. *Hamm et al.* (2006) compared global and local approaches to evaluate the effect of parameter values on the probability of hydrologically induced slope stability. They highlighted the robustness of the Sobol' method, and demonstrated that it accounted for parameter interactions neglected by the local method they used.

An unresolved problem in sensitivity analysis is whether the averaging over parameter space inherent in global methods can obscure important aspects of parameter sensitivity. If sensitivity analysis results vary in parameter space, the robustness of decisions and measures based on sensitivity analysis becomes an issue. For example, data collection strategies and uncertainty quantification may change significantly depending on the parameter values. Considerable variability would suggest the

importance of identifying the most relevant sensitivity analysis results using, for example, information about the system, such as measured equivalents of model-simulated quantities.

To reveal how sensitivity analysis measures vary as parameter values change, we present and evaluate the hybrid local-global Distributed Evaluation of Local Sensitivity Analysis (DELSA) method, in which multiple evaluations of local parameter sensitivity are distributed throughout parameter space. DELSA integrates methodological features from three existing sensitivity analysis methods: the Method of Morris (MoM; *Morris*, 1991; *Zhan et al.*, 2013; *Herman et al.*, 2013b), the Sobol' method (*Sobol'*, 2001; *Van Werkhoven et al.*, 2008), and regional sensitivity analysis (RSA; *Hornberger and Spear*, 1981; *Freer et al.*, 1996). We are not the first to consider the distribution of sensitivities in parameter space; *Saltelli et al.* (2008) call it "factor mapping", but expressly exclude derivative-based "local" methods as a useful tool. Our work suggests the potential utility of local methods in the context of multiscale evaluation of parameter sensitivity.

Our evaluation includes comparisons with the global variance-based Sobol' method. To enable comparison, the Sobol' method is sometimes applied at a smaller than normal scale; and sometimes local DELSA measures are averaged over large-scale parameter space. Computational costs are also compared.

We begin by briefly presenting the Sobol' method, before introducing DELSA. We compare the methods using two examples. The first example is a simple two-parameter reservoir model. The second example involves five alternative bucket-style hydrologic models applied to the Lasnenville catchment in the Belgian Ardennes (Western Europe).

3.2 Sensitivity analysis methods

3.2.1 Global variance-based approach

The Sobol' method decomposes the variance of a metric describing model output (in this work, mean simulated reservoir storage or root-mean-square-error between the simulated and observed streamflow) into contributions from individual parameters. In this study we use the global Sobol' method (*Sobol'*, 2001; *Saltelli*, 2002; *Sobol' and Kucherenko*, 2010) to estimate the first-order sensitivity. Briefly, consider a model f and a vector θ of k model parameters, which yield a metric Ψ describing model output:

$$\Psi = f(\theta) = f(\theta_1, \theta_2, \ldots, \theta_k), \tag{3.1}$$

where the range of each parameter θ_j is defined as $\theta_{j,min}$ to $\theta_{j,max}$. Equation (3.1) can be decomposed as described by *Sobol' and Kucherenko* (2010) to obtain the general variance decomposition scheme

$$V(\Psi) = \sum_j V_j + \sum_j \sum_{m>j} V_{jm} + \ldots + V_{1,2,\ldots,k}, \tag{3.2}$$

where $V(\Psi)$ is the total prior (unconditional) variance of Ψ, V_j are the first order terms, V_{jm} are the second order terms, etc. The Sobol' sensitivity indices are calculated as the ratio between partial variances and the total variance and are conveniently scaled within [0,1]. In this work we use the

first-order Sobol' sensitivity index (S_1^j) calculated as

$$S_1^j = \frac{V_j}{V(\Psi)}. \tag{3.3}$$

It represents the main effect of a parameter θ_j on the total variance $V(\Psi)$. However, it considers no parameter interactions relative to the predictions. Following *Homma and Saltelli* (1996) and *Saltelli* (2002), the first-order variance term V_j can be estimated using the following equations:

$$V_j = \widehat{U}_j - \widehat{E}^2(\Psi), \tag{3.4}$$

$$\widehat{U}_j = \frac{1}{N_S - 1} \sum_{r=1}^{N_S} (\Psi_r \times \Psi_r^j), \tag{3.5}$$

$$\widehat{E}^2(\Psi) = \left(\frac{1}{N_S} \sum_{r=1}^{N_S} \Psi_r \right)^2, \tag{3.6}$$

where $\widehat{U}_j$ is the "perturbed variance", $\widehat{E}^2(\Psi)$ is the squared mean, and N_S is the number of parameter samples used in the Sobol' procedure, which should not be confused with the number of model parameters, k. Ψ_r and Ψ_r^j are the model performance indices of the rth parameter set:

$$\Psi_r = f(\theta_{r,1}, \theta_{r,2}, \ldots, \theta_{r,k}), \tag{3.7}$$

$$\Psi_r^j = f(\theta'_{r,1}, \theta'_{r,2}, \ldots, \theta'_{r,j-1}, \theta_{r,j}, \theta'_{r,j+1}, \ldots, \theta'_{r,k}). \tag{3.8}$$

Ψ_r^j in Eq. (3.8) is constructed from Eq. (3.7) in such a way that all parameters except the jth parameter are perturbed by inserting the value of the parameters generated in another randomly selected Sobol' set of parameters. Finally, the total variance $V(\Psi)$ is estimated by $\widehat{V}(\Psi)$:

$$\widehat{V}(\Psi) = \widehat{E}(\Psi^2) - \widehat{E}^2(\Psi) = \frac{1}{N_S} \sum_{r=1}^{N_S} \Psi_r^2 - \left(\frac{1}{N_S} \sum_{r=1}^{N_S} \Psi_r \right)^2. \tag{3.9}$$

The calculation cost for the first-order Sobol' index, in terms of the number of model runs, is $N_S(k+1)$. The ranges are the only parameter information used. The parameters are sampled using the quasi-random Sobol' sequence, which draws realizations from the least visited locations relative to previous samples in the parameter space (*Bratley and Fox*, 1988) and represents an alternative to Latin hyper-cube sampling.

If Ψ is a purely additive function of the parameters, so that each additive term is a function of at most one parameter, the sum of the first-order Sobol' sensitivity indices equals one ($\sum_{i=1}^{k} S_1^i = 1$). The sampling uncertainty of S_1^j is evaluated by bootstrapping with re-sampling (e.g. *Efron and Tibshirani*, 1993; *Archer et al.*, 1997). In this study, the number of bootstrap samples is 1 000.

3.2.2 Distributed Evaluation of Local Sensitivity Analysis (DELSA)

Local sensitivity analysis is based on the sensitivity (gradient) of model output or performance index Ψ with respect to the jth parameter value

$$\left.\frac{\partial \Psi}{\partial \theta_j}\right|_l, \tag{3.10}$$

where l defines the set of parameter values at which the derivative is calculated. Equation (3.10) is estimated using a forward difference approach using 1% change in the parameter value $[\Psi(\theta_j + 0.01\,\theta_j) - \Psi(\theta_j)]/[0.01\,\theta_j]$. The choice of the 1% parameter change was also compared with 0.1% and 10% parameter change, but had a marginal effect (not shown).

The local equation for total variance comparable to Eq. (3.2) and (3.9) is the prediction variance $V_L(\Psi)$ [*Seber and Wild* (1989, p. 191); *Draper and Smith* (1998, p. 129–130); *Aster et al.* (2013, p. 224); *Lu et al.* (2012) discuss the relation between frequentist and Bayesian calculation of variance]:

$$V_L(\Psi) = \left(\left.\frac{\partial \Psi}{\partial \theta}\right|_l\right)^T (X^T \omega X)^{-1} \left(\left.\frac{\partial \Psi}{\partial \theta}\right|_l\right), \tag{3.11}$$

where X and ω are discussed below and in Appendix 3.A. Appendix 3.A also includes derivation of Eq. (3.11).

Equation (3.11) produces a linear propagation (accomplished using the derivatives $\partial\Psi/\partial\theta$) of the parameter uncertainty expressed by $(X^T\omega X)^{-1}$ to obtain the variance of Ψ. Equation (3.11) produces values identical to the Sobol' variance in Eq. (3.9) under ideal conditions, including linearity of Ψ with respect to the parameters in the range of parameters defined, and use of consistent information on parameters in the two methods. Linearity is generally more nearly approached as the range of parameter values becomes smaller.

For comparison with Sobol', X and ω of Eq. (3.11) need to include only terms associated with prior information on each parameter (see explanation in Appendix 3.A). In this circumstance X is a $k \times k$ identity matrix. In addition, for the prior information to be consistent with the parameter ranges defined for Sobol' and the even distribution of sampling points produced by a uniform or Sobol' sampling, each diagonal of ω needs to equal the reciprocal of the variance of a uniform distribution defined using $\theta_{j,min}$ and $\theta_{j,max}$ (e.g. *Mood et al.*, 1974, see also Appendix 3.B). Under these conditions, the equation for the variance for a three-parameter problem would equal:

$$V_L(\Psi) = \left(\left.\frac{\partial \Psi}{\partial \theta_1}\right|_l\right)^2 s_1^2 + \left(\left.\frac{\partial \Psi}{\partial \theta_2}\right|_l\right)^2 s_2^2 + \left(\left.\frac{\partial \Psi}{\partial \theta_3}\right|_l\right)^2 s_3^2, \tag{3.12}$$

with s_1^2, s_2^2, and s_3^2 a priori parameter variances. The first-order sensitivity measure for the jth parameter is then calculated at each sample point as (e.g., *Helton*, 1993; *Borgonovo*, 2007; *Saltelli et al.*, 2008, and others):

$$S_{L1}^j = \frac{\left(\left.\frac{\partial \Psi}{\partial \theta_j}\right|_l\right)^2 s_j^2}{V_L(\Psi)}, \tag{3.13}$$

Defining the parameter variance s_j^2 to be consistent with a uniform distribution with limits $\theta_{j,min}$ and $\theta_{j,max}$ makes Eq. (3.13) equivalent to an equation presented by *Sobol' and Kucherenko* (2010, proof

of theorem 1) if their parameter normalization to an interval of 0 to 1 is converted. Equation (3.11) does not include terms such as presented by *Sobol' and Kucherenko* (2010) that could be used to approximate parameter interactions relative to the performance criteria.

The cost for calculating S_{L1}^{j} at N_l sample points (parameter sets) is $N_l(k+1)$. $N_l = 1$ provides results at one location, and commonly local statistics are considered only to be calculated for $N_l = 1$. Here we consider opportunities provided by $N_l > 1$, where we either distribute the samples uniformly or using the quasi-random Sobol' sequence. For either sampling method, in DELSA the local sensitivity indices are calculated throughout parameter space. Generally, the full frequency distribution of the sensitivity measures are reported. However, for some comparisons to Sobol' measures the median value is reported.

3.2.3 Connection between DELSA, Method of Morris and RSA

DELSA is similar to the Method of Morris (MoM) (*Morris*, 1991) and the closely related Elementary Effects (EE_j, *Saltelli et al.*, 2008) and Latin Hypercube variant One-At-Time (LH-OAT) method of *van Griensven et al.* (2006). For all these methods model runs are conducted for a base set of parameter values and for a sequence of sets for which each has one parameter value changed at a time. The fundamental difference is that MoM and similar methods focus on providing global sensitivity measures.

MoM produces two global statistics: one measures the overall parameter importance and the other provides an overall global measure of importance variability (see *Saltelli et al.*, 2008, p. 117). One would expect the first MoM statistic to identify the same important and unimportant parameters as the mean or median of the DELSA first-order statistics. However, unlike the DELSA first-order statistic, the MoM statistic is not expected to have values that are numerically identical to Sobol' values under any condition. The MoM parameter importance variability measure would tend to be large when the standard deviation of the DELSA distribution is large, but MoM does not reveal the distribution of local importance across parameter space, for which DELSA was developed. This is consistent with the general use of large parameter value changes in MoM instead of the small values used to evaluate the derivatives of Eq. (3.10). Also, in MoM and EE_j the sequence of one-at-a-time parameter values is cumulative instead of starting each parameter value change at the same point in parameter space. That is, in MoM first one parameter changes value. From that new location, another changes value, and so on. One parameter may be changed many times before proceeding to the next parameter. In contrast, in DELSA the goal of obtaining local derivatives means that each one-at-a-time parameter value change occurs from the same point in the parameter space, the changes are intentionally small, and typically one change occurs. In this regard, DELSA is similar to LH-OAT, but the latter was used to produce global measures.

Distributions of sensitivities like those produced by DELSA have to our knowledge not been produced in applications of the other methods. *Saltelli et al.* (2008, p. 128) suggests that MoM "overcomes the limitations of a local derivative-based approach in that it attempts to explore the whole parameter space". DELSA seeks to accomplish this same goal using a different approach that provides more detail at the local scale. Experience will be needed to better understand the relative insights and utility of these two methods.

DELSA has also some similarities to regional sensitivity analysis (RSA; *Hornberger and Spear*,

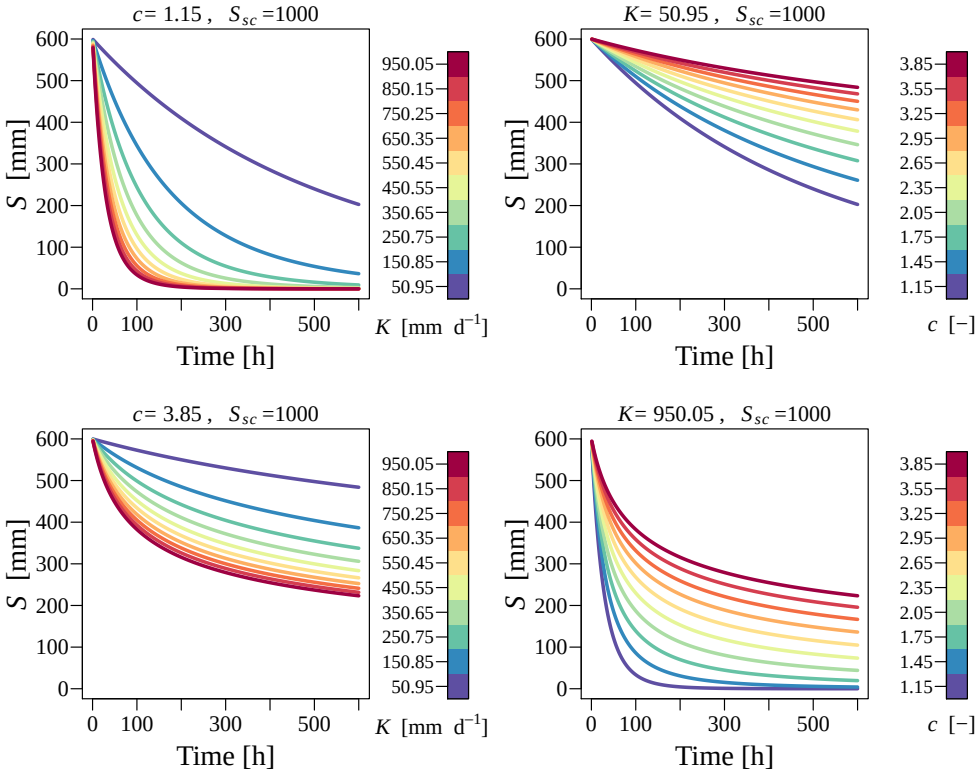


Figure 3.1: Results of drydown experiment with the simple nonlinear model. Colored lines indicate various values of parameters K and c.

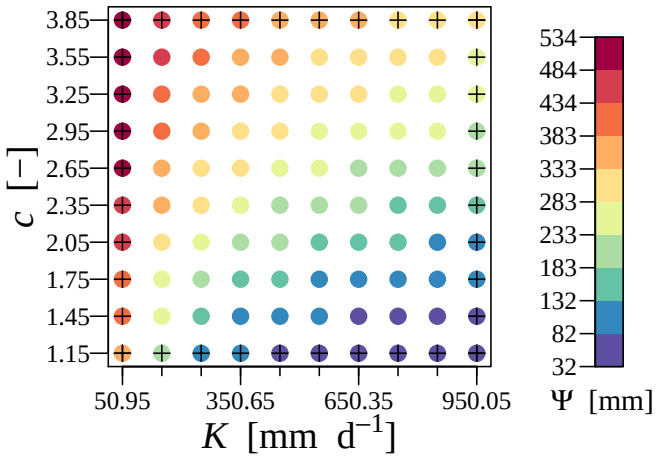


Figure 3.2: Sensitivity of performance index Ψ (mean storage S) of the simple nonlinear model to parameters K and c. Plusses indicate parameter combinations for which time series are shown in Fig. 3.1.

1981; *Freer et al.*, 1996): RSA evaluates the frequency distribution of a model prediction metric across the parameter space, whereas DELSA evaluates the frequency distribution of the gradient of a model prediction metric with respect to the parameter values.

3.3 Experiments for a synthetic test case

3.3.1 Simple non-linear reservoir model

The Sobol' and DELSA sensitivity analysis methods are first applied to the following simple non-linear reservoir model with two model parameters

$$\frac{\mathrm{d}S}{\mathrm{d}t} = p(t) - q(t), \tag{3.14}$$

where S [mm] is model storage, t [day] is time, $p(t)$ [mm day^{-1}] is precipitation (model input), which is kept zero in this case, and $q(t)$ [mm day^{-1}] is drainage (model output).

Drainage is calculated as

$$q(t) = K \left(\frac{S(t)}{S_{sc}} \right)^{c}. \tag{3.15}$$

This can be seen as the flux resulting from gravitational drainage when the Clapp-Hornberger model of unsaturated hydraulic conductivity (*Clapp and Hornberger*, 1978) is used in Richards' equation. K [mm day^{-1}] and c [$-$] are two model parameters for which sensitivity indices are analysed. The scaling parameter S_{sc} [mm] is a constant and is assigned an arbitrary value of 1 000 mm. The initial model storage $S(t = 1)$ is set to 600 mm and the model simulation is composed of 600 one-hour time steps (25 days long). The parameter range for K is defined as [1,1000] and for c as [1,4]. Finally, the performance index Ψ is defined as the mean storage S over time for most results, though results for 10-hourly averaged storages are also presented. The parameter ranges and initial conditions are defined such that the mean storage has a wide spectrum of possible values.

Equation (3.14) is solved using a fixed-time step, implicit Euler numerical scheme by implementing the Newton-Rapshon method. This numerical implementation is unconditionally stable and avoids numerical artefacts (*Clark and Kavetski*, 2010).

The time series of simulated model storage S for different combinations of parameters K and c with the scaling parameter S_{sc} kept constant are shown in Fig. 3.1. The dynamics of this simple model form part of the more complex conceptual hydrological models (further presented in Sect. 3.4): drydown from "field capacity", which is a typical saturation degree for the Lasnenville catchment (Sect. 3.4.2). Figure 3.1 shows that higher values of K yield proportionally higher outflow and model storage decreases faster. Additionally, when the parameter c has also lower values, the reservoir storage empties completely in 600 hours. In contrast, for lower values of K and higher values of c, model storage slowly decreases.

The time-averaged model storage, which is used as a model performance index Ψ (Eq. 3.1), is shown in Fig. 3.2 on a regular grid for 100 parameter combinations of K and c, in which the storage behaviour corresponds well with the simulations in Fig. 3.1. For the defined parameter ranges of K and c, Ψ values range between 32 and 534 mm.

3.3.2 Comparison of Sobol' and DELSA total variances

To compare the Sobol' and DELSA standard deviations (the square root of the total variances), values are calculated for the 100 points of the 2-D parameter space, for which Ψ values are reported in

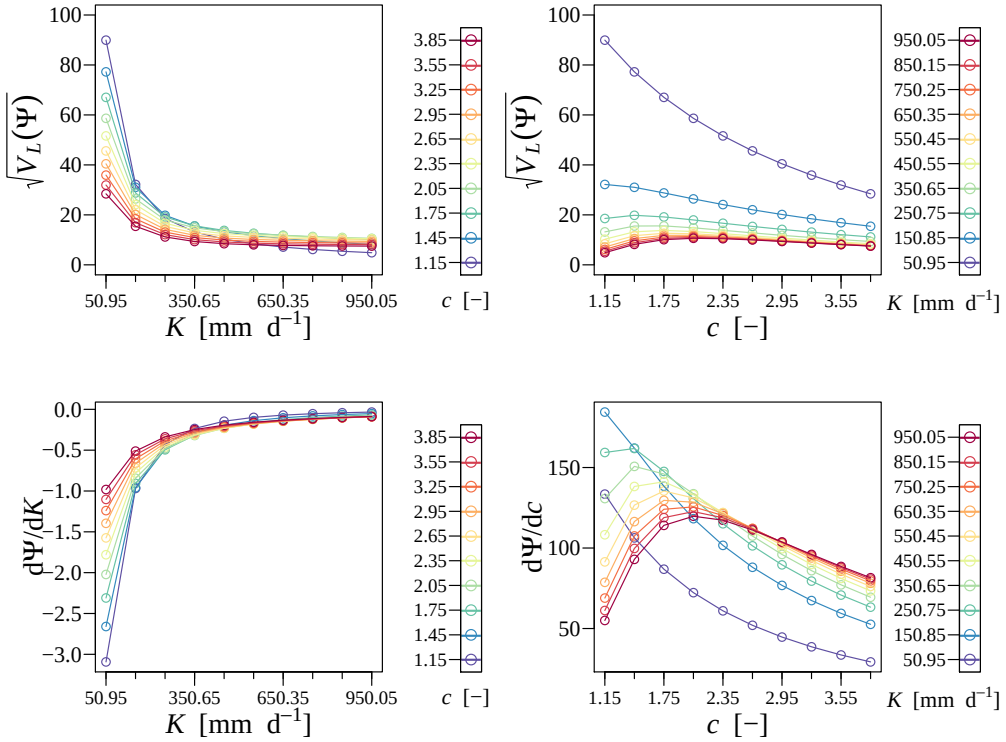


Figure 3.3: (top) Prior model prediction standard deviation ($\sqrt{V_L(\Psi)}$) calculated for values of parameters K and c at 100 points in the parameter space as indicated in Fig. 3.2. (bottom) Local gradients of the parameters K and c to model prediction Ψ calculated for values of parameters K and c at the same 100 points. These two graphs plot the same 100 values in two ways to show how the standard deviations ($\sqrt{V_L(\Psi)}$) are related to the sensitivities.

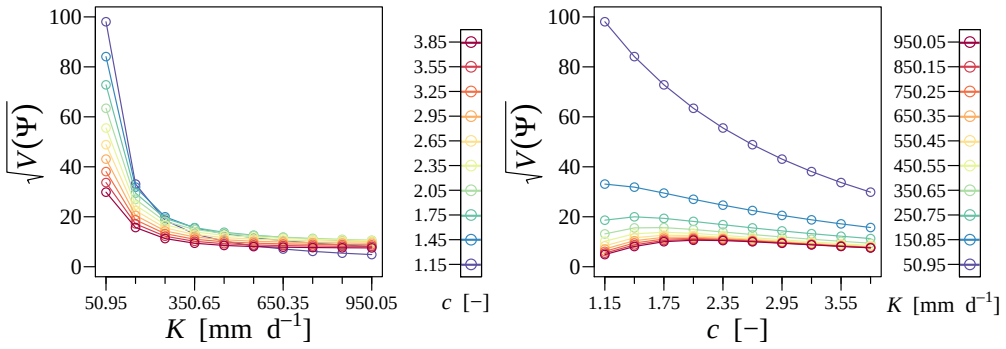


Figure 3.4: Standard deviation of the model prediction Ψ across the parameter space calculated using the Sobol' variance for 100 parameter sub-domains with 10 000 sample size each. Delineated parameter sub-domains are shown in top panel of Fig. 3.5.

Fig. 3.2. For DELSA, this is accomplished by calculating the local $\sqrt{V_L(\Psi)}$ of Eq. (3.11) at the midpoints at each location. For Sobol', $\sqrt{V(\Psi)}$ (Eq. 3.2) is calculated for the 100 parameter sub-domains that surround the 100 locations. The resulting DELSA and Sobol' standard deviations are

shown in Figs. 3.3 (top) and 3.4; results produced by the two methods are very similar. Sobol' values are slightly larger, as expected given that the local methods used here do not account for parameter interactions related to Ψ. As mentioned in the methods section, *Sobol' and Kucherenko* (2010) suggest terms for including these interactions, but they have not been included in this work. For this simple test case the small difference between Fig. 3.3 (top) and Fig. 3.4 suggests this omission has little consequence.

Using Sobol' to calculate sub-domain values is very computationally expensive: 2 000 000 model runs are needed to produce Fig. 3.4, since the computation costs for the Sobol' variance in Eq. (3.9) require $N_S k$ model runs. Such a large number of model runs is usually possible only for simplified problems such as presented here and the sub-domain Sobol' runs are only attempted to demonstrate the similarity between local and global variance estimates under close to ideal conditions.

For DELSA, the matrix ω of Eq. (3.11) was defined to be consistent with the parameter value ranges used by Sobol'. DELSA values comparable to Sobol' values calculated for 10×10 parameter sub-domains require that the prior parameter variances used in DELSA be determined using a uniform distribution, yielding $s_K^2 = 28.839$ and $s_c^2 = 0.087$ (see Appendix 3.B).

Comparison of the local standard deviations of Fig. 3.3 (top) to the local gradients (Eq. 3.10) used by DELSA (Fig. 3.3 bottom) shows how the patterns of the derivatives mirror the patterns of the standard deviations. The magnitudes of the local gradients differ, because of the defined parameter ranges. Derivatives are generally smaller for K, because the K values are relatively large (derivatives range from 0.0 to 3.0; K values range from 50.95 to 950.95). Derivatives are generally larger for c, because the c values are relatively small (derivatives range from about 25 to 175; c ranges from 1.15 to 3.85). Therefore, the local values need to be scaled to obtain comparable measures. In Eq. (3.11) the derivatives in $d\Psi/d\theta$ are scaled by the prior parameter variances in ω to obtain standard deviations plotted in Fig. 3.3 (top).

3.3.3 Comparison of Sobol' and DELSA first-order sensitivity indices

Sobol' and DELSA first-order sensitivity indices are calculated for the 100 subdomains for which the performance index and standard deviations are shown in Figs. 3.2 to 3.4. For Sobol', Eq. (3.3) is applied to each parameter sub-domain; for DELSA, Eq. (3.13) is applied at the middle point of each sub-domain. The results are shown in Fig. 3.5, and suggest that the methods provide similar sensitivity indices across the defined parameter space. Similar to the Sobol' calculation of the total variance, these simulations suggest that calculation of the Sobol' sensitivity measure distributed in parameter space is computationally expensive. Here, the calculation costs for the Sobol' first-order index are $N_S(k+1)$ times 100 sub-domains (=3 000 000), while for the local equivalent this is $N_l(k+1)$ (=300). This suggests that if the distributions produced by DELSA are of interest, obtaining them using Sobol' is likely to be impractical for most problems.

The traditional approach in the Sobol' method is to calculate one sensitivity measure for each parameter for the whole parameter space. The Sobol' S_1 indices for K and c are shown in Fig. 3.6 (top panel) and they have approximately the same sensitivity of 0.50 and 0.48, respectively. Figure 3.6 (middle panel) presents the results of Fig. 3.5 (top panel) showing the cumulative frequency distributions of the Sobol' indices for the 100 parameter sub-domains. Of interest is that the central values differ from the Sobol' global values: the median values are $S_1 = 0.33$ for K and $S_1 = 0.81$ for c.

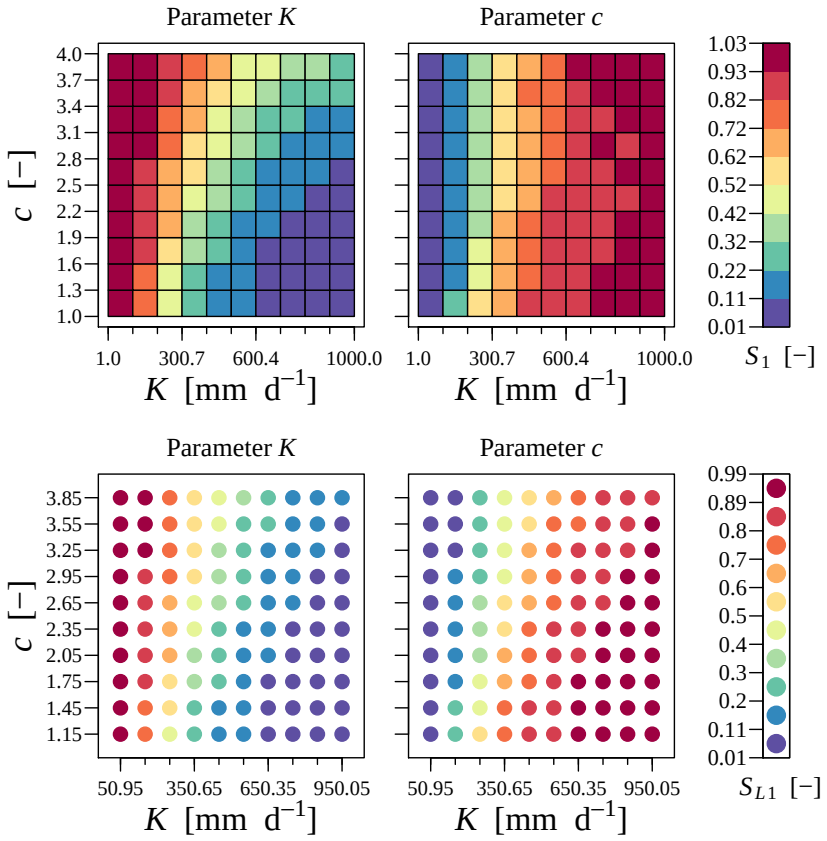


Figure 3.5: Comparison of first-order sensitivity statistics of the global method with Sobol' sampling (S_1, top panel) and the DELSA method (S_{L1}, bottom panel) for 100 parameter sub-domains with 10 000 sample size each.

To understand differences between the traditional way of calculating the Sobol' method (Fig. 3.6, top panel) and the local application of the Sobol' method (Fig. 3.6, middle panel), we decompose the total variance $\widehat{V}(\Psi)$ into its constituents for each of the 100 parameter subdomains. These are labeled $(\Psi_i - \bar{\Psi})^2$ in Fig. 3.7, where i indicates individual parameter locations and $\bar{\Psi}$ is the average over the entire parameter space. Note that calculating the mean of all $(\Psi_i - \bar{\Psi})^2$ terms approximately equals the total variance of Ψ, which is 17091 mm^2. It is clear that different parts of the parameter domain produce different contributions to $\widehat{V}(\Psi)$. For example, regions with higher c and lower K values contribute mostly to $\widehat{V}(\Psi)$ and lower c and higher K values also have significant contributions. The variance constituents over half of the parameter domain contribute little. These results indicate that while the first-order Sobol' sensitivity indices are calculated across the whole parameter domain, only certain sub-regions significantly contribute to the sensitivity indices. This explains why the results between the traditional (lumped) and the distributed Sobol' indices differ.

The sample size needed for reliable results is important because it governs the computational demands of any method. The DELSA cumulative frequency distributions of the local first-order sensitivity index (S_{L1}) for different sample sizes are shown in Fig. 3.6 (bottom). Sampling on a

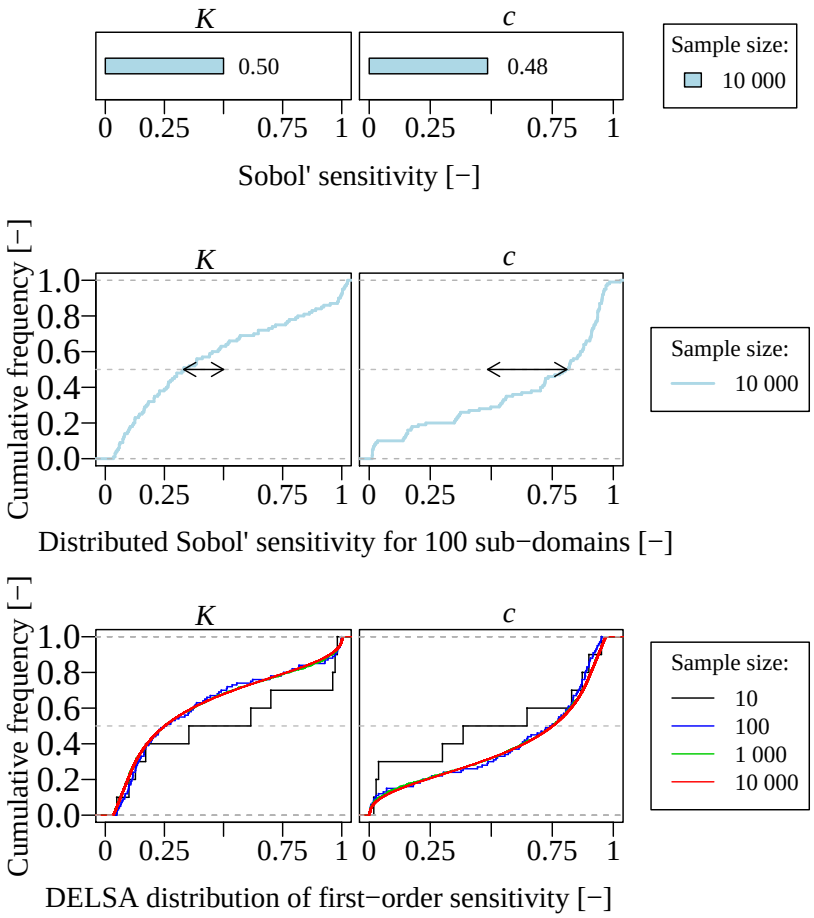


Figure 3.6: Sensitivity indices for the simple nonlinear model with two parameters K and c. (top panel) Bar chart of the first-order Sobol' global sensitivity indices (S_1) for K and c across the whole parameter space. (middle panel) Cumulative frequency distributions of the distributed Sobol' indices for 100 sub-domains. The arrows show the difference from the global Sobol' index presented in the top panel, which are thought to result from model nonlinearity. The nonlinearity is apparent in the distribution of the DELSA statistic; for linear models the same value would be expected throughout parameter space. (bottom panel) Cumulative frequency distributions of the DELSA method (S_{L1}) obtained for different sample sizes (N_l) derived using the quasi-random Sobol' sequence, showing that only 100 samples provide accurate results for this problem.

regular grid as in Fig. 3.5 does not generalize easily when considering different sample sizes. Thus, the quasi-random Sobol' sequence is used to define the samples used for Fig. 3.6 (bottom). The DELSA approach provides a full distribution of the first-order indices and we can easily obtain a number of different statistical properties. For example, the results in Fig. 3.6 (bottom) suggest that parameter c is more influential in much of the parameter space (with median of S_{L1} equal to 0.75) than the parameter K (with median of S_{L1} equal to 0.25). DELSA attains stable results with only 300 model runs ($N_l = 100$), with no significant advantage obtained from using a higher N_l.

Finally, the effect of sampling uncertainty on the first-order sensitivity indices S_{L1} and S_1 is presented in Fig. 3.8, where the sampling variability is obtained from 1 000 bootstrap samples. The median values for $N_S = 10000$ in Fig. 3.8 are the indices presented in Fig. 3.6 (top panel). When

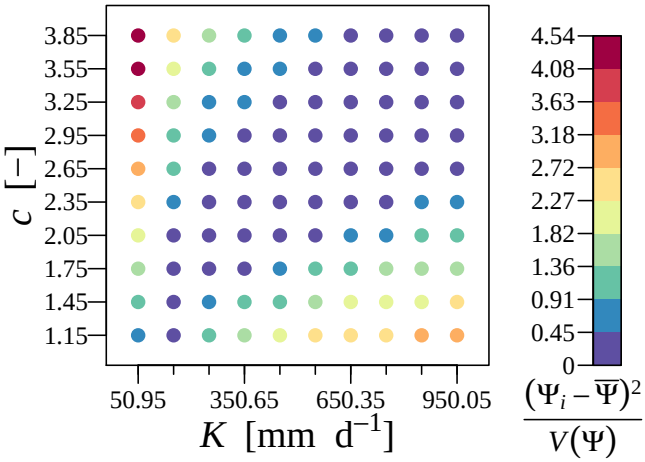


Figure 3.7: Decomposition of the total variance of the model performance index Ψ into its constituents at the 100 parameter space locations. Each constituent is divided by the global variance calculated using Sobol' for combinations of parameters K and c.

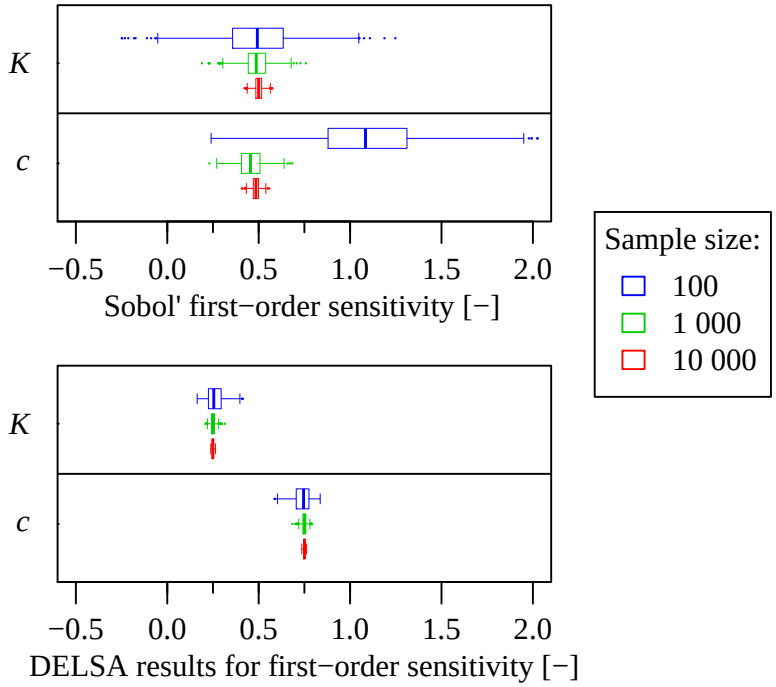


Figure 3.8: Sampling uncertainty for the non-linear model with two parameters K and c. (top panel) Effect of the sample size (N_S) on the Sobol' sensitivity index (S_1). (bottom panel) Effect of the sample size (N_l) on the median of the DELSA method (S_{L1}). The uncertainty estimates are obtained by bootstrapping (re-sampled 1 000 times). The vertical bold line in the boxplot is the median, the body of a boxplot shows the interquantile range (Q75-Q25) and the whiskers represent the sample minima and sample maxima, unless the extreme value occurs further than 1.5 times the interquartile range from the box, in which case the outlier is shown by a dot.

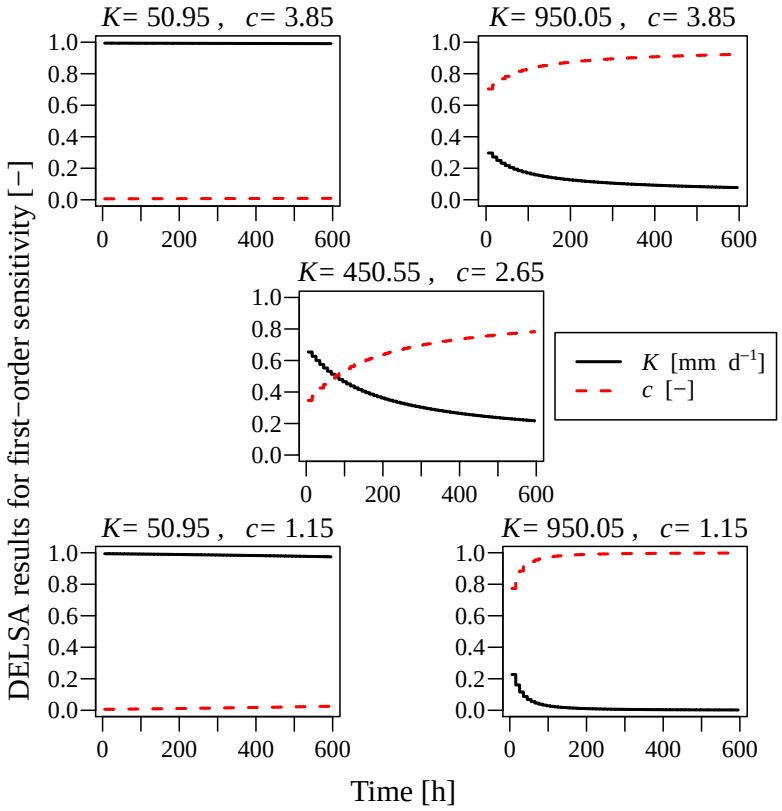


Figure 3.9: Changes in DELSA sensitivity over time for the drydown experiment. DELSA sensitivity (S_{L1}) is calculated for 60 consecutive 10-hour periods at 5 combinations of the K-c parameter space.

compared with the DELSA method, the Sobol' method required much larger N_S ($> 1\,000$) to provide meaningful results within the expected interval [0,1]. Next to that we do not expect that the sensitivity values of Sobol' and DELSA should converge to the same value because of the averaging effects in the global methods, as discussed using the total variance decomposition (Fig. 3.7).

3.3.4 Time varying DELSA

For transient problems, DELSA can be plotted over time for sampled parameter sets to understand changes in dominant model controls as suggested e.g. by *Cloke et al.* (2007); *Herman et al.* (2013a). For the synthetic test case, consider mean storage calculated over 60 consecutive 10-hour periods. We illustrate this in Fig. 3.9 for the simple model using 5 combinations of parameters K and c: the 4 corner points and center point of Fig. 3.2.

DELSA values are time-invariant for the lower K values and temporally variable for higher K values. For some parts of the parameter space (e.g., $K = 450.55$ and $c = 2.65$) the ranking of the parameter importance changes over time from one parameter to the other. In a given problem, observed transient system performance could likely be used to identify which DELSA results are most relevant, in a similar manner as dynamic identifiability analysis (DYNIA) proposed by *Wagener et al.* (2003).

3.4 Experiments for real-world case studies

3.4.1 Alternative hydrologic model structures

To understand the performance of DELSA in realistic problems, we consider a set of conceptual hydrological models within the Framework for Understanding Structural Errors (FUSE) (*Clark et al.*, 2008a). Multiple models are considered because of their utility in practice (e.g., *Ye et al.*, 2010; *Clark et al.*, 2011b; *Foglia et al.*, 2013). Alternative models allow evaluation of alternative hypotheses about system dynamics, and these alternatives can affect which parameters are important and unimportant. For one model a parameter may be insensitive over most of the parameter space, while in another the same parameter can be sensitive over much of the parameter space. The less computationally demanding DELSA makes evaluation of many models more convenient.

In this study we implement five FUSE models (FUSE-016, FUSE-014, FUSE-160, FUSE-072, and FUSE-170) analysed by *Clark et al.* (2011b). These models represent different modeling decisions, which include: (1) the choice of state variable in the unsaturated and saturated zones and (2) the choice of flux equations describing the surface runoff, vertical drainage between soil layers, baseflow and evapotranspiration. For a detailed description of the hydrological models and their model equations we refer to *Clark et al.* (2008a) and *Clark et al.* (2011b). Here we provide only a brief description.

The models FUSE-016, FUSE-014 and FUSE-160 evaluate the choice of state variables used to represent the unsaturated zone and different parametrizations of evapotranspiration. While FUSE-016 has a single reservoir, FUSE-014 and FUSE-160 use two cascading reservoirs. With respect to evapotranspiration, FUSE-016 implements a "single-layer" parametrization, FUSE-014 implements a "sequential" parametrization and FUSE-160 implements a "root weighting" parametrization. All other model components are kept constant, including a single nonlinear groundwater reservoir of unlimited size, the "ARNO/VIC" parameterization of surface runoff, and time delay routing using a gamma distribution.

The FUSE-072 model is used to evaluate different vertical drainage parametrizations. FUSE-016, FUSE-014 and FUSE-160 do not allow any vertical drainage when saturation is below field capacity. The vertical drainage in the FUSE-072 model is a nonlinear function of total storage in the unsaturated zone and for the rest FUSE-072 is identical to FUSE-016. The FUSE-170 model is used to evaluate different baseflow parametrizations. It uses two parallel linear reservoirs to represent baseflow.

The parameters used in the five models are described in Table 3.1. The effective ranges applied in this study were slightly adjusted from *Clark et al.* (2011b) based on our prior knowledge about the Lasnenville catchment (Sect. 3.4.2). The FUSE parameters PERCRTE, PERCEXP, and MAXWATR_1 can be seen as the equivalent of the simple non-linear reservoir model (Sect. 3.3.1) with parameters K, c and S_{sc} respectively (Eq. 3.15). Like the simple nonlinear model, the FUSE simulations have an hourly time step, and fixed-step implicit Euler numerical approximation.

The model performance index Ψ (Eq. 3.1) for this problem is defined as the root-mean-square-error (RMSE) between the observed streamflow (q_{obs}) and FUSE streamflow simulations (q_{sim}):

$$\Psi = \text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(q_{obs,t} - q_{sim,t})^2}. \tag{3.16}$$

Table 3.1: Parameters of the FUSE models used in this work.

Parameter name	Description	Units	Lower limit	Upper limit
MAXWATR_1	Maximum storage in the unsaturated zone	mm	50	500
MAXWATR_2	Maximum storage in the saturated zone	mm	25	250
FRACTEN	Fraction total storage as tension storage	–	0.05	0.95
FRCHZNE	Fraction of tension storage in the primary zone (unsaturated zone)	–	0.05	0.95
FPRIMQB	Fraction of free storage in the primary reservoir (saturated zone)	–	0.05	0.95
RTFRAC1	Fraction of roots in the upper soil layer	–	0.05	0.95
PERCRTE	Vertical drainage rate	$\mathrm{mm\,day^{-1}}$	0.01	1000
PERCEXP	Vertical drainage exponent	–	1	20
PERCFRAC	Fraction of drainage to tension storage in the lower layer	–	0.05	0.95
FRACLOWZ	Fraction of soil excess to lower zone	–	0.05	0.95
BASERTE	Baseflow depletion rate for the single reservoir	$\mathrm{mm\,day^{-1}}$	0.001	1000
QB_POWR	Baseflow exponent	–	1	10
QBRATE_2A	Baseflow depletion rate for the primary reservoir	$\mathrm{day^{-1}}$	0.001	0.25
QBRATE_2B	Baseflow depletion rate for the secondary reservoir	$\mathrm{day^{-1}}$	0.001	0.25
AXV_BEXP	ARNO/VIC "b" exponent for the surface runoff	–	0.001	3
LOGLAMB	Mean of the log-transformed topographic index distribution	m	5	10
TISHAPE	Shape parameter defining the topographic index distribution	–	2	5
TIMEDELAY	Routing parameter equal to the time delay in runoff	day	0.01	2

This measure tends to emphasize high flows and it is important to remember that the obtained sensitivities will change when using different performance indices.

3.4.2 Data

The sensitivity analyses using the FUSE models are carried out for a 10-year period from 1 October 1998 to 30 September 2008 for the Lasnenville catchment located in the Belgian Ardennes (Western Europe). The Lasnenville catchment is a medium-sized catchment (200 km^2) located upstream of the Lasnenville streamflow gauge, which is operated by the Hydrological Service of the Walloon Region of Belgium (MET-SETHY). The meteorological data are obtained from the Vielsalm FLUXNET site (*Papale et al.*, 2006; *Reichstein et al.*, 2005), which is located at the border of the Lasnenville catchment. The cumulative annual precipitation is about 1 000 mm and mean annual temperature 7.5°C. The land use is approximately half mixed-forest and half agricultural areas. The climate conditions can be classified as rain-fed with ephemeral snow in winter and the runoff regime is highly variable with low summer discharges and high winter discharges (e.g. *Driessen et al.*, 2010; *Hazenberg et al.*, 2011; *Rakovec et al.*, 2012b).

3.4.3 Comparison of Sobol' and DELSA

The sensitivity indices of all model parameters are shown in Fig. 3.10 for sample sizes $N_S = 1\,000$ and $N_l = 1\,000$, respectively. The same sample size was used for both Sobol' and DELSA so that use of different sample sizes would not confuse inter-comparison of results. Because the FUSE models have between 11 and 14 parameters and the calculation costs are $N_S(k+1)$ and $N_l(k+1)$, respectively, the number of model runs is between 12 000 and 15 000, depending on a model structure. For DELSA, the distribution of median values is shown.

Results suggest that, for all models for DELSA and Sobol', parameters TIMEDELAY and AXV_BEXP are the most important (see table 3.1 for a description of each parameter). This makes sense, because these two parameters significantly affect the magnitude and timing of the simulated streamflow in the RMSE. Other parameters are moderately important to one or more models. For example, FRACTEN is the third most sensitive parameter for FUSE-016 according to both methods, which is consistent with how evaporation and unsaturated zone are simulated (*Clark et al.*, 2011b). For the FUSE-072 model PERCEXP is somewhat important because vertical drainage from the unsaturated zone to the saturated zone is simulated even when the upper zone storage falls below field capacity. FUSE-160 uses two parallel linear reservoirs to represent baseflow, in contrast to the single non-linear reservoir used in all other FUSE models, and there are three other sensitive parameters: MAXWATR_1, PERCFRAC, QBRATE_2B (see table 3.1 for parameter description).

For all five models, the DELSA approach provides sensitivity indices which correspond well to the global Sobol' method.

The DELSA method yields much smaller sampling uncertainty than the global Sobol' method for the same number of model runs. This is shown in Fig. 3.10 by the error bars accounting for the 95% confidence interval: for sample size of 1 000 the DELSA error bars are much narrower. This was also analyzed for the two most sensitive parameters TIMEDELAY and AVX_BEXP using sampling sizes of 100, 1 000 and 10 000 (not shown). While 100 model runs provided consistent sensitivity indices

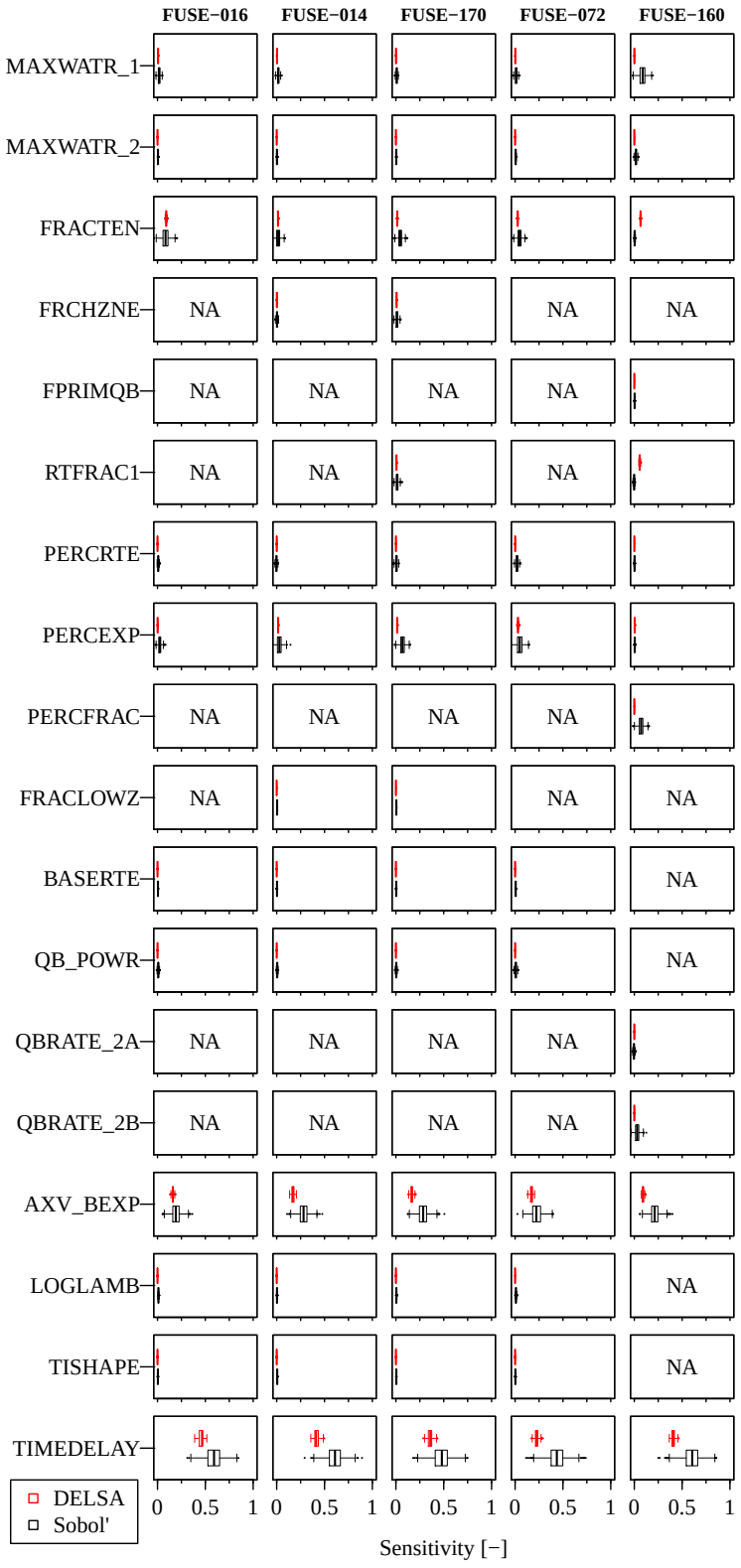


Figure 3.10: Comparison of the DELSA (S_{L1}) and Sobol' (S_1) sensitivity indices for five FUSE model structures using $N_S = 1\,000$ and $N_l = 1\,000$. The sampling uncertainty (bootstrapping with 1 000 times re-sampling) is indicated by boxplots.

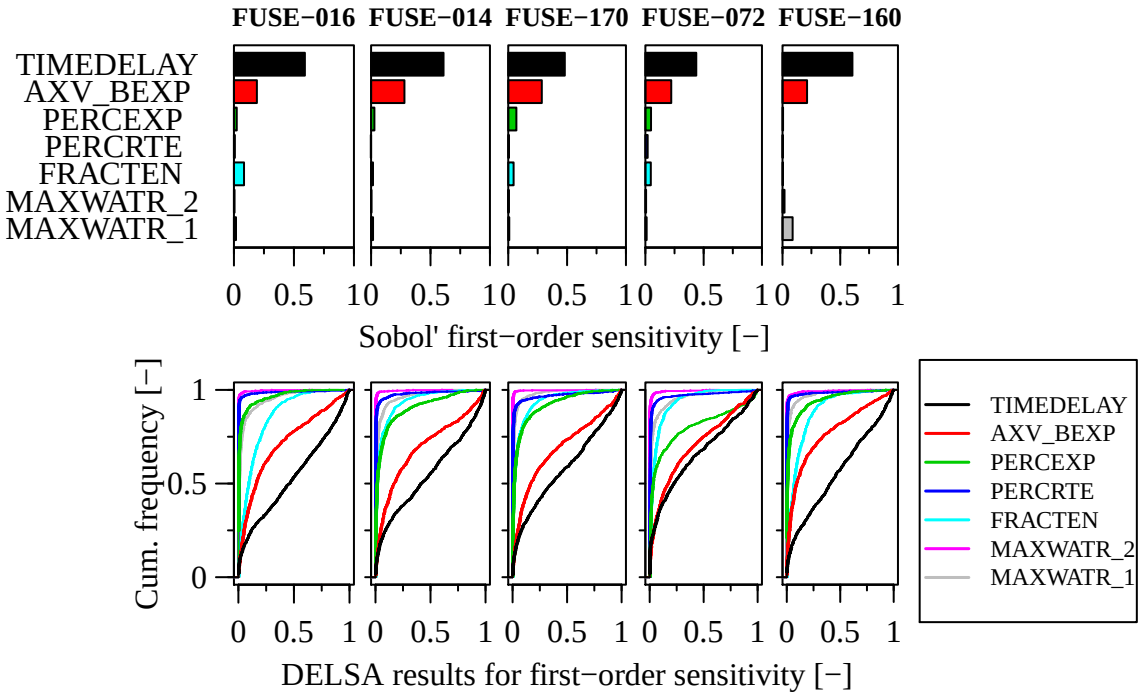


Figure 3.11: Sensitivity indices for the 7 common parameters of the five FUSE model structures using $N_S = 1\,000$ and $N_l = 1\,000$. (top panel) Global Sobol' sensitivity indices (S_1). (bottom panel) Cumulative frequency distributions of the variability across the parameter space using the DELSA method (S_{L1}).

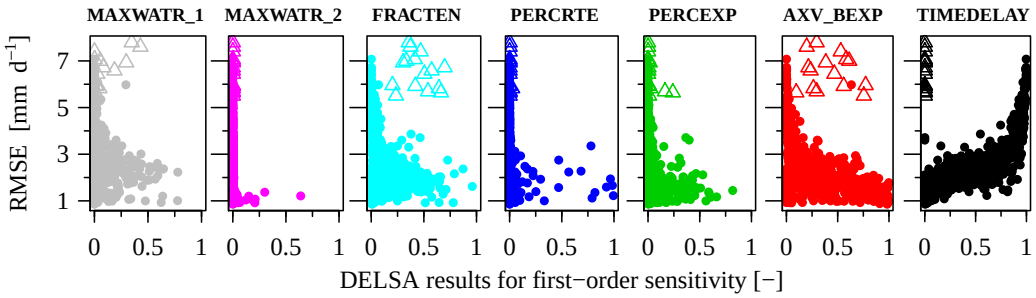


Figure 3.12: Scatterplot of the DELSA statistic (S_{L1}) related to the RMSE performance metric for the FUSE-016 model showing the same parameters as depicted in Fig. 3.11 for $N_l = 1\,000$. Triangles indicate 15 parameter sets, which are discussed in the text as outliers. Note that simulations with high RMSE are of low hydrological significance. Both AXV_BEXP and TIMEDELAY have high DELSA sensitivities, but only in the case of AXV_BEXP these correspond to low RMSE values.

using the DELSA method, Sobol' required many more model runs (preferably $N_S > 1\,000$) to provide sensitivity indices with equivalently low sampling uncertainty.

In Fig. 3.10 the Sobol' values and median DELSA values are sometimes different, though the same important and unimportant parameters are identified. Such differences have been discussed in Sect. 3.3 of this work.

Figure 3.11 compares the full distribution of parameter sensitivity provided by the DELSA method to the global Sobol' model. Results are shown for seven parameters common to all five FUSE models. While Fig. 3.10 shows only the median sensitivity values, Fig. 3.11 shows the full empirical frequency distribution. Sensitivity indices at the cumulative frequency of 0.5 correspond to the values shown in Fig. 3.10. The DELSA results allow a more detailed evaluations. For example, consider two parameters: FRACTEN in FUSE-016 and PERCEXP in FUSE-072. Both have first sensitivity index of 0.1 estimated by the global Sobol' method, but inspection of the DELSA results shows that parameter PERCEXP in FUSE-072 has a sensitivity index > 0.75 for about 20% of the sampling locations. This means that for about 20% of the parameter value samples, PERCEXP is nearly as important as the two most important parameters, TIMEDELAY and AVX_BEXP. On the other hand, the FRACTEN parameter in FUSE-016 hardly never exceeds a DELSA sensitivity of 0.5, even though the median shows a sensitivity of 0.1. Inspection of the DELSA results also suggests that the prediction is insensitive to any of the parameters for about 20% of the parameter sets. Inspection of some model runs (not shown) for which the sensitivity of the TIMEDELAY and AVX_EXP parameters equals zero showed that the FRACTEN parameter became very important. For these models the simulated response was either much more or much less flashy than the data suggested. Further evaluation is beyond the scope of the present work, but even these initial results suggest the insight available through multi-scale evaluations of sensitivity such as that provided by DELSA.

One interesting question is the relevance of global scale statistics, since global statistics often include contributions from regions of the parameter space where the model performs poorly. This issue is addressed by *Pappenberger et al.* (2008a), who, in contrast to common practice, only performed global sensitivity analysis on the model simulations with "acceptable" model performance – simulations that *Pappenberger et al.* (2008a) considered to be "behavioural". Figure 3.12 illustrates values of the DELSA statistic for each parameter value for model FUSE-016, plotted against the RMSE performance metric. Here we make two observations. First, there is a small number of outlier parameter sets, denoted as triangles in Figure 3.12, where large errors are associated with high sensitivity for FRACTEN, AXV_BEXP, and lower sensitivity for parameter TIMEDELAY. These parameter sets may provide unrealistic simulations, as indicated by the high RMSE statistic. These parameters sets all have values of TIMEDELAY that are less than 0.05; the range of TIMEDELAY was defined as 0.01 to 2, and the small values would not be expected for a basin this large (200 km^2). These results could be used to understand why model performance is so anomolous, or to omit such unrealistic results from the analysis. Second, parameter sets that have the highest values of sensitivity in the TIMEDELAY also have poor model performance. To the extent that large values of RMSE indicate an unrealistic model, this may suggest that for models of most interest the dominance of parameter TIMEDELAY may be exaggerated by the global approach. Figure 3.12 also shows the difficulties faced by using traditional $N_l = 1$ sensitivity analysis for these models. Such analyses are likely to identify MAXWATR_2, PERCRTE, and perhaps PERCEXP as insensitive, but results for other parameters are likely to be erratic. Investigation of whether spurious or realistic results account for this is an interesting question not pursued here.

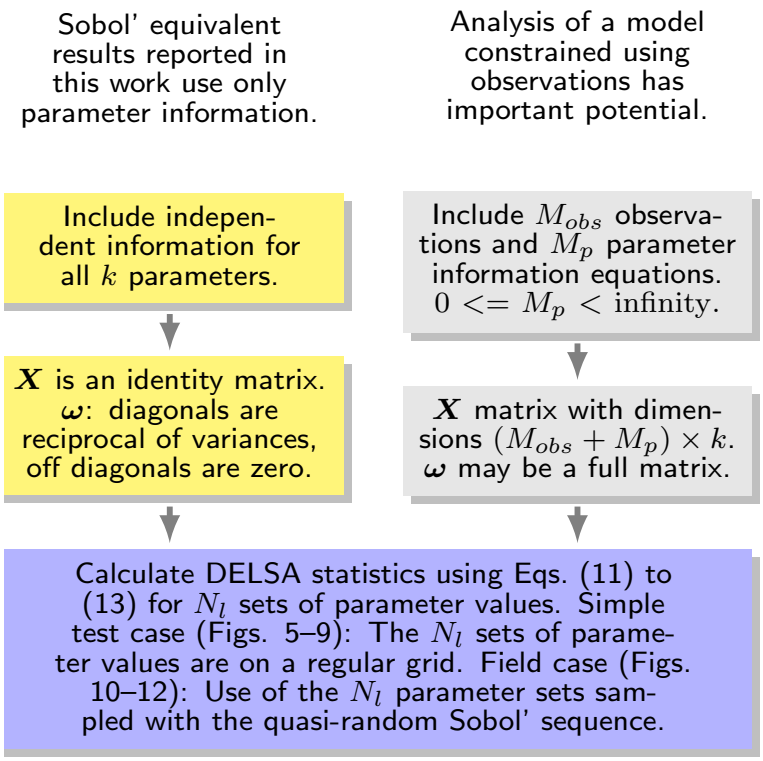


Figure 3.13: Flowchart showing definition of X and ω used to calculate DELSA for situations comparable to Sobol' variance analysis (in yellow), as considered in this work. Also shown is how X and ω are defined in the important situation of having observations with which to constrain the simulation (in gray).

3.5 Discussion

The results of the simple (synthetic) and complex (real world) problem suggest potential utility of the DELSA method. Advantages include convergence with relatively small values of N_l, and detailed characterization of variations in sensitivity over the defined range of parameter values.

The local methods considered in this work are expected to perform well when performance metrics are smooth functions of the parameters. Simulated results can be non-smooth for a number of reasons. These include spurious solutions of the type discussed by *Clark and Kavetski* (2010) and *Kavetski and Clark* (2010), and can also include small oscillations produced by otherwise valid solution schemes (for example, random-walk and method of characteristic solutions for transport often produce small oscillations). In the presence of irregularities and small oscillations, the small parameter change used to calculate derivatives in local methods can produce sensitivity measures that indicate greater parameter importance than actually occurs (also noted by, for example, *Kleijnen and Helton*, 1999; *Sobol' and Kucherenko*, 2010). This is likely the reason that in *Li et al.* (2013a) local methods indicated greater importance for some parameters than did the global methods considered.

In DELSA, such difficulties will not dominate a parameter importance measure as long as the irregularities occupy little of the parameter space. A few large values will clearly be exceptional when the DELSA distributions are plotted, as for example in the bottom panel of Fig. 3.11, and can be suitably ignored. Indeed, DELSA provides an opportunity to identify and possibly fix such model

irregularities if they are deemed to be unrealistic.

If the irregularities are pervasive and oscillate around a valid solution, DELSA will perform poorly and function smoothing or use of global methods is advised. Often, such irregularities are known to be typical of a modeling method, as is the case for random walk and the method of characteristics (MOC) transport solutions (see *Mehl and Hill*, 2001). For many reasons besides sensitivity analysis, it is advisable for modelers to be cognizant of solution characteristics such as small oscillations and other irregularities; such knowledge is needed to determine the utility of model results as well as suitable methods for model development and analysis.

The DELSA method provides a number of intriguing opportunities that we do not pursue in this first paper on DELSA. Here we briefly mention selected opportunities and significant related literature and issues.

Equation (3.11) is presented in its general form and Appendix 3.A discusses how observations can be included in X and ω to emphasize that DELSA has close ties to local sensitivity theory (*Cook and Weisberg*, 1982; *Foglia et al.*, 2009). The DELSA method appears to provide rich possibilities for using a full parameter variance-covariance matrix as required to include observations and thus allow sensitivity analysis on a posteriori distributions, as indicated in Fig. 3.13. In global methods, full matrices were considered by *Saltelli et al.* (2004) but Sobol' evaluations were found to depend on the order of the parameters (*Bedford*, 1998). *Xu and Gertner* (2011), *Zuniga et al.* (2013), and *Kucherenko et al.* (2012) present methods for accounting for parameter correlation in the context of global methods. To our knowledge numerically equivalent local methods have not been identified and development of numerically equivalent local measures is an interesting topic beyond the scope of the present work.

In this work we focus on the first-order sensitivity measure. Sobol' statistics of higher order, including total-order statistics, could perhaps be included in DELSA by using the local derivatives (for example, see *Sobol' and Kucherenko*, 2010).

3.6 Conclusions

We investigated parameter sensitivity and how it is affected by hydrologic model structure using two very different techniques: a global variance-based method (Sobol') and the proposed Distributed Evaluation of Local Sensitivity Analysis (DELSA) method. In DELSA, parameter sensitivity evaluation is based on gradients of the model performance index with respect to model parameters at multiple points throughout the parameter space and sensitivity metrics are formulated in terms of the distribution of values across the parameter space. Additionally, DELSA is constructed in a way that allows direct comparison to global variance decompositions of Sobol'. The results show that the Sobol' and DELSA methods yield similar results for the problems considered, although the DELSA method provides the full distribution of sensitivity throughout the parameter space at lower calculation costs than is incurred to obtain a single summary measure of sensitivity using Sobol'. For example, in the real-world problem the time delay in runoff is the most important parameter in all models, but DELSA shows that for about 20% of parameter sets it is not important at all and alternative mechanisms and parameters dominate. Moreover, the time delay was identified as important in regions producing poor model fits, whereas other parameters were identified as more important in regions producing better model fits. The detailed information produced by DELSA about model sensitivity provided important

insights about how parameter sensitivity varied throughout the parameter space.

This study focuses on introducing the DELSA method and applying the method to reproducible test problems. DELSA is ideally suited to analyze multiple metrics of model behaviour (e.g. *Rosolem et al.*, 2012) and environmental models with lengthy execution times. The advantage of the multiscale, hybrid local-global DELSA method is that application at multiple points across the parameter space can reveal important parameter subregions, generally undetected by global methods. We anticipate that the DELSA method will become a powerful approach to provide insight about model parameter sensitivity at very low computational cost. We look forward to working with others in the community to apply the DELSA methodology to challenging modeling problems.

3.A Appendix: Definition of X and ω and derivation of DELSA equation (3.11)

Equation (3.11) is the prediction total variance, $V_L(\Psi)$, and is composed of two parts.

The first part is the variance-covariance matrix of the parameter vector θ, or $V(\theta) = \sigma^2(X^T \omega X)^{-1}$, a standard expression stated in many texts, including *Seber and Wild* (1989, p. 191), *Menke* (1989, p. 58), *Draper and Smith* (1998, p. 129–130), *Tarantola* (2005, p. 36), and *Aster et al.* (2013, p. 224).

X is a matrix with M rows and k columns and ω is a matrix with M rows and M columns. k is the number of parameters. M depends on the type of problem considered. For model calibration problems with no prior information, M equals the number of observations. If each observation is represented by y_i, the elements of X equal $\partial y_i / \partial \theta_i$. ω would equal the weight matrix of the observations. If the model calibration problem includes prior information, $M = M_{obs} + M_p$, where M_{obs} equals the number of observations and M_p equals the number of prior information equations. The additional M_p rows of the X matrix often have zeros except for one 1 which identifies the parameter associated with the prior information. However, alternative forms are used to define, for example, first-order Tikhonov regularization. For comparison with Sobol', no observations are used and $M = N_p$. X is a $k \times k$ identify matrix, and ω is diagonal and contains the reciprocal of the variances of the parameters (see Appendix 3.B). The structure of the X and ω matrices is described and illustrated by *Hill and Tiedeman* (2007) [p. 384]. In Eq. (3.11) the assumption has been made that the sample common error variance $\sigma^2 = 1$, hence, it does not explicitly appear in Eq. (3.11). This assumption is discussed toward the end of this appendix.

For Eq. (3.11) as presented in this work, the terms for prior information are included in the matrices X and ω. Separation of X and ω into the parts for observations and prior information produces the following alternative forms of Eq. (3.11):

$$V_L(\theta) = s^2(X_{obs}^T \omega_{obs} X_{obs} + C_p^{-1})^{-1} = C_p - C_p X_{obs}(\omega_{obs}^{-1} + X_{obs} C_p X_{obs}^T)^{-1} X_{obs} C_p \tag{3.17}$$

The equivalence of the two expressions is discussed by *Oliver et al.* (2008, p. 142).

The expression $V(\theta) = \sigma^2(X^T \omega X)^{-1}$, is derived in detail by *Hill and Tiedeman* (2007, p. 396–398), and a condensed version of that derivations follows.

Start with the standard expression of variance applied to θ,

$$V(\theta) = E[(\theta - E(\theta))(\theta - E(\theta))^T]. \tag{3.18}$$

The expression for parameter values optimized using linear regression is

$$\theta = (X^T \omega X)^{-1} X^T \omega y, \tag{3.19}$$

where y is a vector of observations. Substitution of Eq. (3.19) into Eq. (3.18) for θ (not for E(θ)); expansion of terms using matrix algebra; application of the matrix property $AB^T = B^T A^T$; recognition that $(X^T \omega X)^{-1}$ and ω are symmetric so that the transpose equals the original matrix; substituting the first-order (linear) expression for y, $y = A\theta + \varepsilon$; and taking the expected value noting that only ε is stochastic yields:

$$V_L(\theta) = (X^T \omega X)^{-1} X^T \omega E[\varepsilon\varepsilon^T] \omega X (X^T \omega X)^{-1}. \tag{3.20}$$

The subscript L indicates linearity has been assumed. Consider defining the weight matrix ω such that

$$E[\varepsilon\varepsilon^T] = V(\varepsilon) = \sigma^2 \omega^{-1}. \tag{3.21}$$

Substituting Eq. (3.21) into Eq. (3.20), cancelling terms, and assuming $\sigma^2 = 1$ yields $V(\theta) = (X^T \omega X)^{-1}$, the expression used in Eq. (3.17). Assuming $\sigma^2 = 1$ is not significant for the DELSA statistics defined in this work because as a constant it divides out in the calculation of Eq. (3.13).

The second part of Eq. (3.11) is the pre and post multiplication of $V_L(\theta)$ by the derivatives of the prediction with respect to the parameters. This multiplication is a linear propagation of the parameter error as represented by $V_L(\theta)$ to the prediction. Error propagation of this type is standard in regression theory, and is presented in the references cited for Eq. (3.11) at the beginning of this appendix.

Equation (3.11) calculates the variances that appear in the denominator of Eq. (3.13) for the local first-order sensitivity indices. The numerator of Eq. (3.13) is obtained by retaining one term in the expression for $V_L(\theta)$, as described in the text.

3.B Appendix: Variance of a uniform distribution

For the Sobol' sampling used in this work, the probability for all parameter values within the range of parameters is assumed to be equal. This is consistent with a uniform probability distribution.

For the DELSA statistics (Eq. 3.13) to produce numerically equivalent results to Sobol', the parameter range and distribution implied by the prior information variance in ω needs to be consistent with that of the Sobol' method. Thus, given a parameter θ_j with ranges $\theta_{j,min}$ and $\theta_{j,max}$ defined for Sobol', the variance used in DELSA needs to equal the variance of the uniform distribution, which is $\frac{1}{12}(\theta_{j,max} - \theta_{j,min})^2$ (e.g. *Mood et al.*, 1974).

Chapter 4

State updating of a distributed hydrological model with Ensemble Kalman Filtering: effects of updating frequency and observation network density on forecast accuracy

Abstract

This paper presents a study on the optimal setup for discharge assimilation within a spatially distributed hydrological model. The Ensemble Kalman filter (EnKF) is employed to update the grid-based distributed states of such an hourly spatially distributed version of the HBV-96 model. By using a physically based model for the routing, the time delay and attenuation are modelled more realistically. The discharge and states at a given time step are assumed to be dependent on the previous time step only (Markov property).

Synthetic and real world experiments are carried out for the Upper Ourthe (1600 km^2), a relatively quickly responding catchment in the Belgian Ardennes. We assess the impact on the forecasted discharge of (1) various sets of the spatially distributed discharge gauges and (2) the filtering frequency. The results show that the hydrological forecast at the catchment outlet is improved by assimilating interior gauges. This augmentation of the observation vector improves the forecast more than increasing the updating frequency. In terms of the model states, the EnKF procedure is found to mainly change the pdfs of the two routing model storages, even when the uncertainty in the discharge simulations is smaller than the defined observation uncertainty.

This chapter is largely based on the publication: 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, *Hydrol. Earth Syst. Sci., 16*, 3435–3449, doi:hess-16-3435-2012.

4.1 Introduction

Accurate and reliable hydrological forecasts have been a challenge in applied hydrology for decades. Better forecasts can be obtained through data assimilation (DA) by merging observations with model simulations (*Reichle*, 2008). This approach basically updates the model states with externally measured variables (*Pauwels and De Lannoy*, 2006) to obtain correct initial conditions for the next time step. Currently, most operational hydrological forecasting systems employ lumped hydrological models (with deterministic or manual state updating), but there is a clear tendency to move towards distributed models combined with hydrological ensemble forecasts, (e.g. *Koren et al.*, 2004; *Cole and Moore*, 2009; *Weerts et al.*, 2013). The main advantage of spatially distributed models is the possibility to force them with spatially measured data, which nowadays become more readily available due to rapid developments in telemetry. Distributed model states also resemble the real world observations (e.g., groundwater levels, soil moisture, discharge) at the interior of the catchment more closely than lumped states over the whole catchment. Another advantage of applying distributed models is the ability to simulate and predict hydrological variables at interior locations within the catchment. Techniques on how to perform ensemble data assimilation using these models in real-time settings should be developed and tested (*Weerts et al.*, 2013; *Liu et al.*, 2012).

Data assimilation methods used in hydrology can be divided into two classes: (1) sequential and (2) variational (e.g. *Liu and Gupta*, 2007). Sequential methods are mostly employed for state updating in hydrological models by assimilating observations for each time step when they become available. Its impact depends on the uncertainties in both the observations and model states. Variational methods rather minimise a cost function over a simulation time window. At the beginning, a first-guess model is constructed, which is afterwards updated by creating an adjoint model which propagates backwards in time and incorporates the mismatch between the model and observations (*Liu and Gupta*, 2007).

A popular method often used in both meteorological and hydrological forecasting is the Ensemble Kalman Filter (EnKF) (*Evensen*, 1994, 2003, 2009). This sequential data assimilation method is an extension of the classical Kalman filter (KF), which was originally developed for linear systems (*Kalman*, 1960). The EnKF propagates an ensemble of model realisations through time and estimates the error covariance matrix from the ensemble statistics. The advantage of the EnKF to other data assimilation methods is computational efficiency, and easy and straightforward implementation within a data assimilation procedure for both lumped and distributed models (*Pauwels and De Lannoy*, 2009).

Discharge measurements are the most widely used in situ hydrological observations for model updating, since they reflect the local catchment wetness conditions and are often available at high temporal resolution (*Pauwels and De Lannoy*, 2006; *Teuling et al.*, 2010), which is necessary for operational hydrological forecasting. Comprehensive analyses of discharge data assimilation into spatially lumped hydrological models was carried out by e.g., *Pauwels and De Lannoy* (2006); *Weerts and El Serafy* (2006); *Pauwels and De Lannoy* (2009). Among others, *Aubert et al.* (2003) and *Lee et al.* (2011) assimilated, in addition to discharges, in situ soil moisture measurements. *Lee et al.* (2012) analysed the sensitivity of variational data assimilation methods for multiple spatiotemporal adjustment scales, namely assessing (1) different spatial distributions of model states and (2) temporal resolution of biases in precipitation and potential evaporation. Furthermore, the water level is an example of an in situ state variable, which can also be assimilated into a hydrodynamical model (e.g.

Madsen and Skotner, 2005; *Neal et al.*, 2007; *Weerts et al.*, 2010). Additionally, temperature observations play an important role in real-time operational forecasting systems, especially in regions with significant snow melt (e.g. *Verbunt et al.*, 2007; *Sene*, 2008).

Although the assimilation of remotely sensed data into operational hydrological models can improve model performance, this task is complicated, because remotely sensed data usually have a higher uncertainty than other in situ measurements (*Pauwels and De Lannoy*, 2009). Nevertheless, several studies focus on assimilation of remotely sensed data into hydrological models, e.g., soil moisture (*Pauwels et al.*, 2001, 2002; *Moradkhani*, 2008) and snow (*Slater and Clark*, 2006; *Thirel et al.*, 2011). Remotely sensed water levels were assimilated into a hydraulic model by *Giustarini et al.* (2011).

So far, few studies have reported on data assimilation within spatially distributed hydrological models, both from a scientific and from an operational perspective. *Clark et al.* (2008b) assimilated discharges using the adjusted EnKF for a real world case only. However, their approach was not tested and evaluated in a synthetic experiment, which would help to understand the behaviour of the simulated and updated probability density function (pdf) of the model states. Another EnKF study, by *Blöschl et al.* (2008), applied discharge assimilation into a real-time grid-based operational flash flood forecasting model. As *Blöschl et al.* (2008) did not elaborate on the importance of individual discharge gauges in the interior of the catchment, this remains an interesting question to address. However, more recently the positive effect of interior discharge gauges on hydrological forecast was described by *Lee et al.* (2011), who assimilated both discharge and/or in situ soil moisture data using the variational method.

From an operational point of view, it is also desirable to optimise the telemetry system, which delivers the observations to the forecaster. One of the interesting aspects is the frequency at which the observations become available. In most of the currently operational large-scale flood forecasting systems, it is typical to separate the forecast process into a state (i.e., carry-over or update) run, which is normally run once a day, and forecast runs which are run more frequently. The frequency of the state run is very much dictated by the frequency at which observed data becomes available (i.e., frequency of polling of the telemetry system). Hence, it is interesting to test the optimal updating frequency at which the hydrological data are assimilated into the forecasting model to obtain the most accurate forecast.

The objective of this study is to analyse the sensitivity of the DA procedure to the number and the locations of discharge gauges, which are assimilated into a grid-based distributed operational hydrological forecasting model using the EnKF. The optimal updating frequency will be addressed as well.

The focus here lies mainly on the state run part of the forecast chain and, therefore, we leave out the meteorological forecasts and error modelling of the meteorological forecast, which is a research topic on its own (e.g. *Germann et al.*, 2009). Using this approach, we disregard the large uncertainties in the meteorological forecasts and we can focus entirely on the uncertainty in the initial model states.

4.2 Material and methods

4.2.1 Catchment description and data availability

The hydrological simulations are carried out for the Upper Ourthe catchment, upstream of Tabreux (Fig. 4.1b), which drains an area of about 1600 km^2 (*Berne et al.*, 2005). This catchment forms a tributary of the Meuse River basin, originating in the hilly parts of the Belgian Ardennes. The climatic conditions can be classified as rain-fed with irregular snow in winter and the runoff regime is highly variable with low summer discharges and high winter discharges (*Leander et al.*, 2005). Relatively shallow soils in combination with significant elevation differences result in a quickly responding catchment. As such, the whole region represents a significant flood risk to the Netherlands (*de Wit*, 2008). Table 4.1 presents the catchment response time of the Upper Ourthe, which is defined as the time between the event-based catchment averaged rainfall centroid and the discharge peak for individual discharge gauges. The catchment response time is about 30 h at Tabreux and about 11 h at the two most upstream gauges (Mabompre and Ortho). This indicates that it takes about 20 h of travel time within the main channel between the two upstream gauges and the catchment outlet.

Hourly precipitation data are available from 42 automatic rain gauges situated within the Belgian Ardennes region, from which 10 are located inside the Upper Ourthe catchment (*Hazenberg et al.*, 2011). Discharge is measured at six different points at an hourly resolution. Next to that, temperature is obtained from the Saint Hubert meteorological station (Fig. 4.1b). The long-term mean monthly values of potential evapotranspiration are assumed identical to those of the operational lumped HBV-96 model, derived from the St. Mihiel station in North-Eastern France.

4.2.2 Generation of a spatially distributed precipitation ensemble

A precipitation ensemble, a finite and discrete number of spatial realisations over time, represents the uncertainty associated with temporal as well as spatial variation of precipitation. This probabilistic input enables hydrologists to evaluate more critically hydrological simulations and/or predictions. In the current chapter, we employed a time-dependent multivariate spatial conditional simulation method (*Rakovec et al.*, 2012a), which is further made conditional on preceding simulations. This method identifies, at an hourly time step, temporally coherent errors in spatial precipitation fields that are plausible from a hydro-meteorological perspective. Neglecting this temporal aspect could lead to underestimation of the overall uncertainty in the precipitation ensemble.

The theory of conditional (sequential Gaussian) simulations is fully explained by *Goova erts* (1997). For a detailed description of this time-dependent simulation method using the gstat R package (*Pebesma*, 2004; *Rossiter*, 2007; *R Development Core Team*, 2013) we refer to *Rakovec et al.* (2012a), who carried out a synthetic experiment and analysed three real rainfall events during winter 2002/2003 within the Belgian Ardennes. Altogether, 27 rain gauges were used to simulate 64 ensemble members over a 100 km $\times$ 100 km domain with a 10 km$\times$ 10 km raster resolution (see Fig. 4.1b). In the current study, we made the simulation of each realisation conditional on 3 h of the corresponding previously simulated realisation, which is a recommendation following from the results obtained by *Rakovec et al.* (2012a).

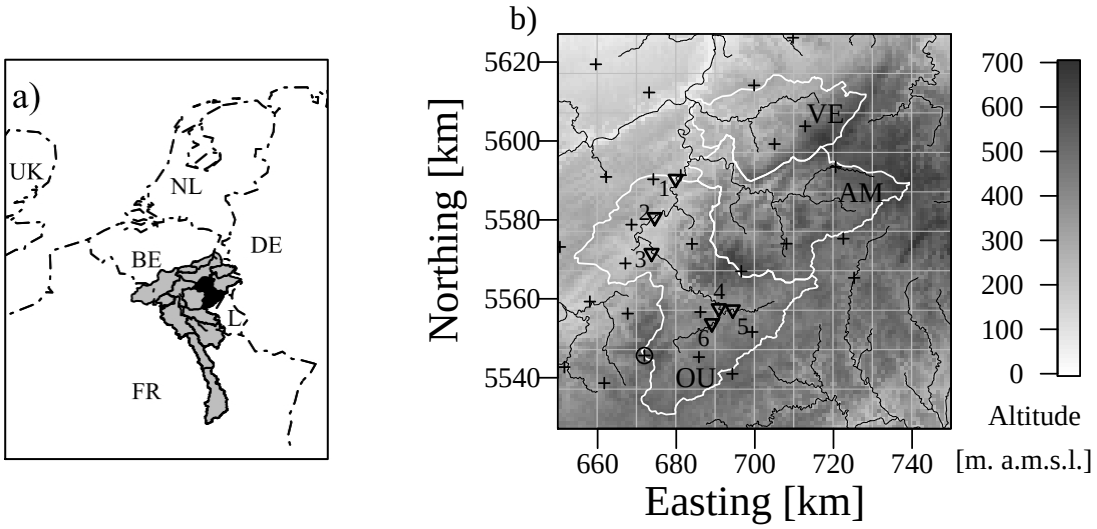


Figure 4.1: (a) The Meuse River basin upstream of Sint Pieter, the entrance point into the Netherlands, and its 15 sub-catchments (grey polygons) including the Upper Ourthe (black polygon). (b) Topographic map of the Upper Ourthe (OU, white line) including the river network (black lines), rain gauges (plusses), six river gauges (triangles and labelled with numbers: 1 = Tabreux, 2 = Durbuy, 3 = Hotton, 4 = Nisramont, 5 = Mabompre, 6 = Ortho) and the climatological station in Saint Hubert (circle). For completeness, two neighbouring sub-catchments Ambleve (AM) and Vesdre (VE) are shown, however, they were not analysed in this study. The grey grid shows the 10 km × 10 km pixel resolution of the rainfall generator, which will be further explained in Sect. 4.2.2. Projection is in the Universal Transverse Mercator (UTM) 31N coordinate system. After *Hazenberg et al.* (2011).

Table 4.1: Catchment response time between the catchment averaged rainfall centroid and the discharge peak.

Discharge gauge[a]:	1	2	3	4	5	6
Upstream area [km^2]:	1620	1225	959	743	318	387
Time to peak [h]:	31	26	21	13	11	11

[a] Location of discharge gauges is indicated in Fig. 4.1b.

4.2.3 Hydrological model

Currently, a spatially lumped HBV-96 model (*Lindström et al.*, 1997) is used operationally by the Dutch authorities for flood forecasting of the Meuse River basin at and downstream of Sint Pieter at an hourly time step (*Driessen et al.*, 2010). The Meuse River basin upstream of Sint Pieter ($\sim$21 000 km^2), the entrance point into the Netherlands, is conceptualised into 15 lumped HBV-96 sub-catchments of which the Upper Ourthe is one (Fig. 4.1a). The models for the Meuse River sub-catchments were calibrated for the period 1970–1984 and validated for the period 1985–1996 at a daily time step by *Booij* (2002). The precipitation inputs for the original HBV-96 models were derived from 39 rain gauges, of which only one station was located within the Upper Ourthe catchment. For operational purposes, the hourly HBV-96 models were derived and re-calibrated based on the work of *Booij* (2002) by *van Deursen* (2004). A comparison of daily and hourly model versions was carried out by *Weerts* (2007). Currently, the operational forecasts derived with the lumped hourly

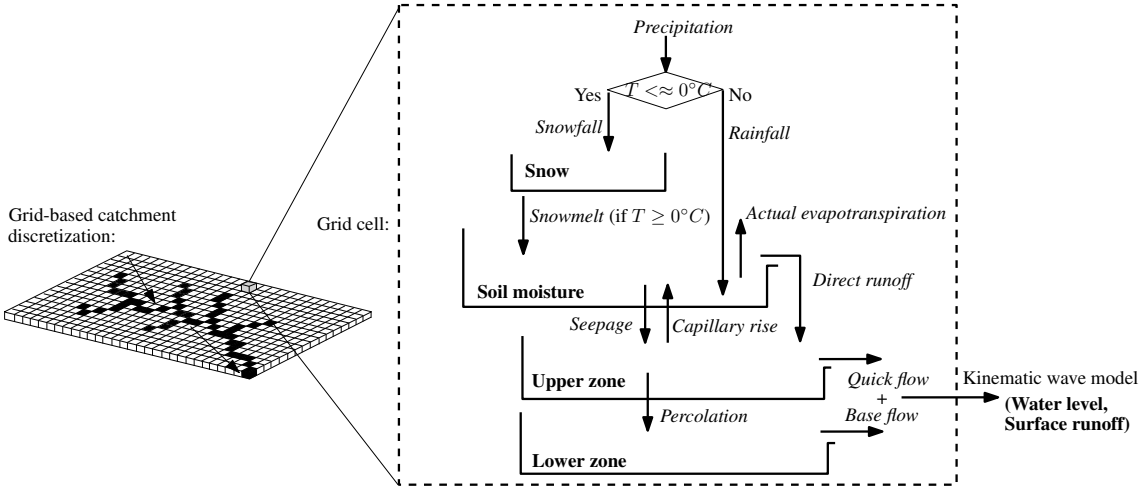


Figure 4.2: Left: Catchment discretisation using a grid-based approach including the channel delineation. Arrows indicate flow direction. Right: Schematic structure of the HBV-96 model for each grid cell. Model states are in bold and model fluxes in italics.

HBV-96 models are used as lateral inflows into a 1D-hydrodynamic model of the Meuse River. This lumped model does not employ sequential state updating, but is updated with discharge observations by means of an automated auto-regressive moving average error correction method (*Broersen and Weerts*, 2005).

For this study, we have developed a grid-based spatially distributed HBV-96 based model within PCRaster, a software environment for constructing iterative spatiotemporal environmental models (*Karssenberg et al.*, 2009; *PCRaster*, 2013). Such a grid-based approach is a popular concept in applied hydrology (e.g. *Koren et al.*, 2004; *Blöschl et al.*, 2008; *Cole and Moore*, 2009; *Thielen et al.*, 2009). For each 1 km $\times$ 1 km grid cell of the Upper Ourthe the HBV-96 model was implemented and is used as a benchmark case. Both model parameters and structure are taken identical to the lumped HBV-96 version used operationally except for the discharge routing, for which a kinematic wave model (*Chow et al.*, 1988; *PCRaster*, 2013) is used. This is only the first step forward by taking full advantage of distributed models, which would allow us to define spatially variable model parameters. However, that is beyond the scope of the current study.

The structure of the model used in this study is shown in Fig. 4.2. For each grid cell the model considers four model states: (1) snow (SN), (2) soil moisture (SM), (3) upper zone storage (UZ) and (4) lower zone storage (LZ). The dynamics of the model states are governed by the following model fluxes: rainfall, snowfall, snowmelt, actual evaporation, seepage, capillary rise, direct runoff, percolation, quick flow and base flow. The latter two fluxes force the kinematic wave model. This routing scheme calculates the overland flow using two additional model states, the water level (H) and discharge accumulation over the drainage network (Q). In this study, we use a very similar routing setup as the one applied within the distributed hydrological CQ-flow model (*Schellekens*, 2006). The main drainage network is obtained using the 8-direction steepest descent algorithm based on a digital elevation model with a grid resolution of 1 km $\times$ 1 km. Afterwards, the catchment is split between the channel and non-channel grid cells. The channel network is defined for the cells with Strahler stream order (*Strahler*, 1964) greater than 3. As such, only the major tributaries of the Upper Ourthe are identified, which corresponds well to the channel network derived from a topographic map (for

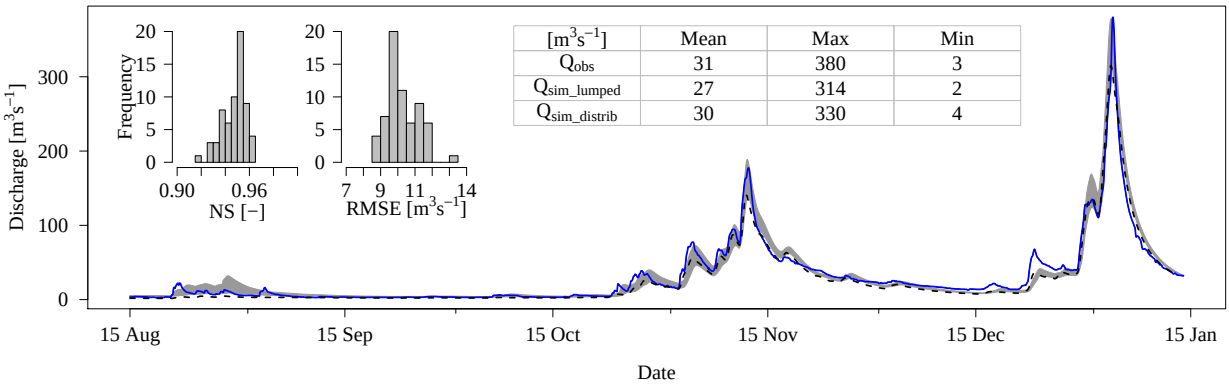


Figure 4.3: Hourly hydrograph for the Upper Ourthe at Tabreux: observed (solid blue line); HBV-96 simulation using the grid-based spatially distributed version (light grey band for the 95 % confidence bounds of 64 ensemble members); HBV-96 simulation using the operational spatially lumped version (dashed black line). Histograms of the Nash-Sutcliffe (NS) model efficiency coefficient and the root-mean-square error (RMSE) between the observed and the ensemble of simulated discharges using the grid-based version of HBV-96 are shown in the upper left corner.

Table 4.2: The channel width corresponding to Strahler stream order number

Strahler order stream number	Channel width [m]
6	20
5	15
4	10

comparison, see Figs. 4.1b and 4.4). A channel width is assigned based on a field survey and *Google maps* (2011) according to the Strahler stream order number, as shown in Table 4.2. Google orthophoto maps (satellite images) were used for rough visual estimates to obtain information on the channel widths of corresponding Strahler stream order numbers. For the non-channel cells, the channel width remains equal to the grid cell width, because the water is routed from these cells by means of sheet flow on top of the whole grid cell. By making the width of the non-channel cells very large, we are able to decrease the hydraulic gradient of the water in the "channel" of these non-channel cells and, therefore, increase the response time of these rather slow responding cells. The channel width is then used to derive the water level, which defines, together with the local topography gradient, the gravity force, which is a driver for the river flow. Channel roughness coefficients were estimated by a sensitivity analysis.

To demonstrate that the implemented spatially distributed version is able to produce reasonable hydrological discharge simulations using the aforementioned precipitation ensemble generator, Fig. 4.3 shows the observed discharge and both the spatially lumped and distributed HBV-96 discharge simulations at Tabreux. A clear consistency can be observed between both HBV-96 simulations for the 5 month period (15 August 2002–15 January 2003), which will be further used in this study. The Nash-Sutcliffe (NS) model efficiency coefficient (*Nash and Sutcliffe*, 1970) between both HBV-96 simulations is about 0.99, which gives strong evidence of model similarity. Moreover, the

NSs between the observed and the ensemble of discharges using the grid-based version of HBV-96 are between 0.92 and 0.96 and the root-mean-square error (RMSE) ranges between 9 and 12 $m^3 s^{-1}$ (Fig. 4.3). The NS between the observed and the simulated discharge using the lumped HBV-96 is about 0.96 as well. For completeness, additional statistics are summarised in Fig. 4.3. Even though the grid-based HBV-96 model was not recalibrated, the model performs very well at the scale of the Upper Ourthe catchment.

4.2.4 Ensemble Kalman filter

The Ensemble Kalman filter (*Evensen*, 2003, 2009; *Weerts and El Serafy*, 2006) is a recursive Bayesian estimation method, which estimates the true probability density function of the model states conditioned on observations. Let us denote a dynamic state space system as:

$$x_k = f(x_{k-1}, \theta, u_{k-1}) + \omega_k \quad \omega_k \sim N(\mathbf{0}, \mathbf{S}) \ , \tag{4.1}$$

where x_k is a state vector at time k, f is an operator expressing the model state transition from time step $k-1$ to k in response to the model input u_{k-1} and time-invariant model parameters θ. So f is, in fact, the hydrological model. ω_k stands for system noise, normally distributed with zero mean and covariance $\mathbf{S}$. This additive system noise incorporates the overall uncertainties in model structure, parameters and model inputs. One can expect some spatial patterns of model errors to be found in the covariance matrix $\mathbf{S}$. However, quantification of $\mathbf{S}$ for highly nonlinear hydrological systems is a complicated task and, therefore, we keep it time-invariant.

The observation process is governed by Eq. (4.2):

$$y_k = \mathbf{H}x_k + \nu_k \quad \nu_k \sim N(\mathbf{0}, \mathbf{R}_k), \tag{4.2}$$

where y_k is an observation vector derived from the model state x_k and the model parameters through the $\mathbf{H}$ operator (in our case the kinematic wave model). ν_k stands for additive observational noise, normally distributed with zero mean and covariance $\mathbf{R}_k$. For independent measurement errors between the observations in vector y_k, we can assume $\mathbf{R}_k$ to be a diagonal matrix. As such, this simplification does not consider any dependency in model simulations for observation points located close to each other.

The idea of recursive Bayesian estimation is to construct a conditional density p for an ensemble of the state x_k given all available information up to and including the step k: $p(x_k|Y_k)$, where $Y_k = (y_1, y_2, \ldots, y_k)$. This can be obtained using the Bayesian rules in two steps: forecast and update.

After the update of model states at time $k-1$, the hydrological model is used to forecast model states at time k (Eq. 4.1). The grid-based model states form a matrix, which consists of N state vectors x_k corresponding to N ensemble members:

$$\mathbf{X}_k = (x_k^1, x_k^2, \ldots, x_k^N), \tag{4.3}$$

where

$$x_k = (\mathrm{SN}_{1:m}, \mathrm{SM}_{1:m}, \mathrm{UZ}_{1:m}, \mathrm{LZ}_{1:m,1}, \mathrm{H}_{1:m}, \mathrm{Q}_{1:m})^T, \tag{4.4}$$

SN, SM, UZ, LZ, H and Q are the HBV-96 model states (Sect. 4.2.3), m gives the number of grid cells

and T is the transpose operator. The ensemble mean

$$\bar{x}_k = \frac{1}{N}\sum_{i=1}^{N} x_k^i \tag{4.5}$$

is used to derive the model error for each ensemble member:

$$\mathbf{E}_k = (x_k^1 - \bar{x}_k, x_k^2 - \bar{x}_k, \ldots, x_k^N - \bar{x}_k). \tag{4.6}$$

The ensemble estimated model covariance matrix $\mathbf{P}_k$ is defined as

$$\mathbf{P}_k = \frac{1}{N-1}\mathbf{E}_k\,\mathbf{E}_k^T. \tag{4.7}$$

When observations become available, the model states are updated as follows:

$$\mathbf{X}_k^+ = \mathbf{X}_k^- + \mathbf{K}_k(y_k - \mathbf{H}_k\mathbf{X}_k^-), \tag{4.8}$$

where $\mathbf{X}_k^+$ is the new updated (posterior) model state matrix and $\mathbf{X}_k^-$ is the forecasted (prior) model state matrix. $\mathbf{K}_k$ is the Kalman gain, a weighting factor of the errors in model and observations:

$$\mathbf{K}_k = \mathbf{P}_k\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_k\mathbf{H}_k^T + \mathbf{R}_k)^{-1}, \tag{4.9}$$

where $\mathbf{P}_k\mathbf{H}_k^T$ is approximated by the forecasted covariance between the model states and the forecasted discharge, and $\mathbf{H}_k\mathbf{P}_k\mathbf{H}_k^T$ is approximated by the variance of forecasted discharge (*Houtekamer and Mitchell*, 2001):

$$\mathbf{P}_k\mathbf{H}_k^T = \frac{1}{N-1}\sum_{i=1}^{N}(x_k^i - \overline{x_k})(\mathbf{H}_k x_k^i - \overline{\mathbf{H}_k x_k})^T, \tag{4.10}$$

$$\mathbf{H}_k\mathbf{P}_k\mathbf{H}_k^T = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{H}_k x_k^i - \overline{\mathbf{H}_k x_k})(\mathbf{H}_k x_k^i - \overline{\mathbf{H}_k x_k})^T, \tag{4.11}$$

where

$$\overline{\mathbf{H}_k x_k} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{H}_k x_k^i. \tag{4.12}$$

In previous published papers on EnKF (e.g. *Weerts and El Serafy*, 2006; *Pauwels and De Lannoy*, 2009), a time delay issue was noted due to the use of the unit-hydrograph, where the discharge at time k depends on several previous calculated discharges (for instance at $k-1$, $k-2$, $k-3$, $k-4$, etc., for hourly models often up to 24 h or more in the case of the Upper Ourthe). By using a physically based model for the routing (Sect. 4.2.3), the time delay and attenuation are modelled more realistically and the discharge and states x_k can be assumed to depend only on the states x_{k-1} (Markov property). The time delay is, thus, explicitly taken into account in the model.

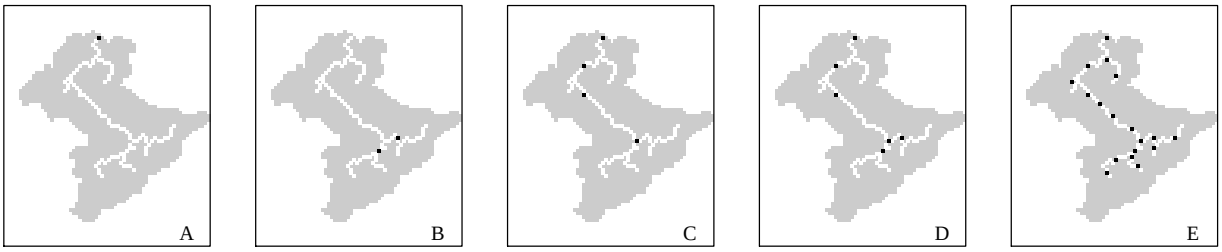


Figure 4.4: Five cases of the discharge observation vector y_k of increasing spatial extent. Discharge gauges contained in the observation vector are plotted in black. Channel delineation using Strahler stream ordering plotted in white pixels.

4.2.5 Experimental setup

Synthetic experiment

The approach of the synthetic experiment is similar to the approach used by *Weerts and El Serafy* (2006), with the main difference being that we employ a realistic stochastic representation of the spatially distributed precipitation (see *Rakovec et al.*, 2012a). We limit the analysis of the experiment to input uncertainty only. The main reason for this is that we want to fully understand and investigate the filter process using a distributed hydrological model and realistic precipitation fields. We believe this is already challenging enough without initial state and model parameter/structural uncertainty, which we leave for future work.

The synthetic experiment has been carried out to examine the ability of the EnKF to update the grid-based hydrological model states via assimilation of the spatially measured discharge. The EnKF procedure was applied to a 5-month period from 15 August 2002 to 15 January 2003, which includes a dry and a wet period. For reasons of clarity of the experimental setup, we did not update model parameters, which were kept constant. In an open loop simulation, i.e., without data assimilation, the model was initially forced with uncertain precipitation inputs with a simulation memory of 3 h (64 ensemble members) derived using time-dependent multivariate spatial conditional simulation (see Sect. 4.2.2 and *Rakovec et al.*, 2012a) and observed deterministic potential evapotranspiration and temperature. This produced an ensemble of simulated discharges from which one complete realisation was randomly selected as the true discharge ($Q_{\text{true},k}$). To introduce discharge observation uncertainty, a normally distributed error ν_k (Eq. 4.2) with heteroscedastic variance was added to $Q_{\text{true},k}$ to obtain a synthetic perturbed observation $Q_{\text{pobs},k}$:

$$Q_{\text{pobs},k} = Q_{\text{true},k} + \nu_k \quad \nu_k \sim N(0, (0.1\, Q_{\text{true},k})^2) \; . \tag{4.13}$$

The discharge measurement error was defined similar to *Thiemann et al.* (2001) and *Weerts and El Serafy* (2006). The ensemble size of 64 members corresponds to other studies as well, e.g., *Pauwels and De Lannoy* (2009). The model errors ($\mathbf{S}$, Eq. 4.1) were obtained by perturbing the model states indirectly by uncertain precipitation input. Additional direct perturbation of the model storages, which was applied by e.g., *Clark et al.* (2008b) was not considered, because this was considered to be beyond the scope of this study. That would make our example, which focuses on the input uncertainty only, even more complicated.

In the synthetic experiment, we assimilated in total 5 schemes of perturbed discharge observations,

expressed by the vectors y_k, as shown in Fig. 4.4. The first case A is identical to a lumped model for the Upper Ourthe, where only the most downstream observation is available. The second case B considers only the two most upstream discharge gauges. The third case C includes three additional observations upstream to case A. The fourth case D contains all six discharge gauges and the fifth case E includes an additional 12 imaginary gauges to the fourth case.

Moreover, the effect of the updating frequency, i.e., how often the observations become available and how often they are assimilated into the model, was analysed for updating frequencies of 24, 12 and 6 h.

The performance of the data assimilation procedure regarding discharge forecasting was then evaluated using the root-mean-square error ($\mathrm{RMSE}_{\mathrm{lt}}$):

$$\mathrm{RMSE}_{\mathrm{lt}} = \sqrt{\frac{1}{MJN}\sum_{i=1}^{M}\sum_{j=1}^{J}\sum_{n=1}^{N}\left(Q_{\mathrm{true,lt}}^{i,j,n} - Q_{\mathrm{for,lt}}^{i,j,n}\right)^2}, \tag{4.14}$$

where lt stands for lead time ($\mathrm{lt} = 1\,\mathrm{h}, 2\,\mathrm{h}, \ldots, 48\,\mathrm{h}$) and Q_{for} is a forecasted discharge vector of length J. M is the number of hydrological forecasts, which were issued over the 5-month period. To allow comparison of the different updating frequencies between each other, the hydrological forecast was issued every 6 h, i.e., 4 times a day.

Real world experiment

In the real world experiment, we applied the same model forcings as in the synthetic experiment (Sect. 4.2.5). The difference with the synthetic experiment was the assimilation of the real discharge observations ($Q_{\mathrm{obs},k}$), which were perturbed by a normally distributed observation error with a variance of $(0.1\,Q_{\mathrm{obs},k})^2$ (*Weerts and El Serafy*, 2006; *Clark et al.*, 2008b). Analogous to the synthetic experiment, Q_{obs} was used in Eq. (4.14) instead of Q_{true} to calculate the RMSE. The size of the observation vector y_k was limited to cases A, B and D (Fig. 4.4).

4.3 Results

4.3.1 Synthetic experiment

Model performance regarding discharge forecast

The long-term RMSE (Eq. 4.14) between the synthetic observed discharge and the forecasted discharge of the 64 ensemble members at Tabreux (Fig. 4.1b) for the three updating frequencies of 24, 12 and 6 h is shown in Fig. 4.5. The forecasted discharge without data assimilation gives a constant RMSE of about $5\,\mathrm{m^3\,s^{-1}}$, which corresponds to 16 % error with respect to the mean simulation over the 5-month period. For the forecasted discharge with data assimilation, there is a reduction in RMSE for all discharge observation vectors, however, the magnitudes differ as follows:

(1) for the updating frequency of 24 h, the benchmark case A performs worst of all 5 cases, although the differences between case A and B are marginal, given the small difference in their RMSE values as well as the high NS value in the base-case model simulation. Moreover, there is a gradual decrease in RMSE for the cases with a large number of assimilated gauges.

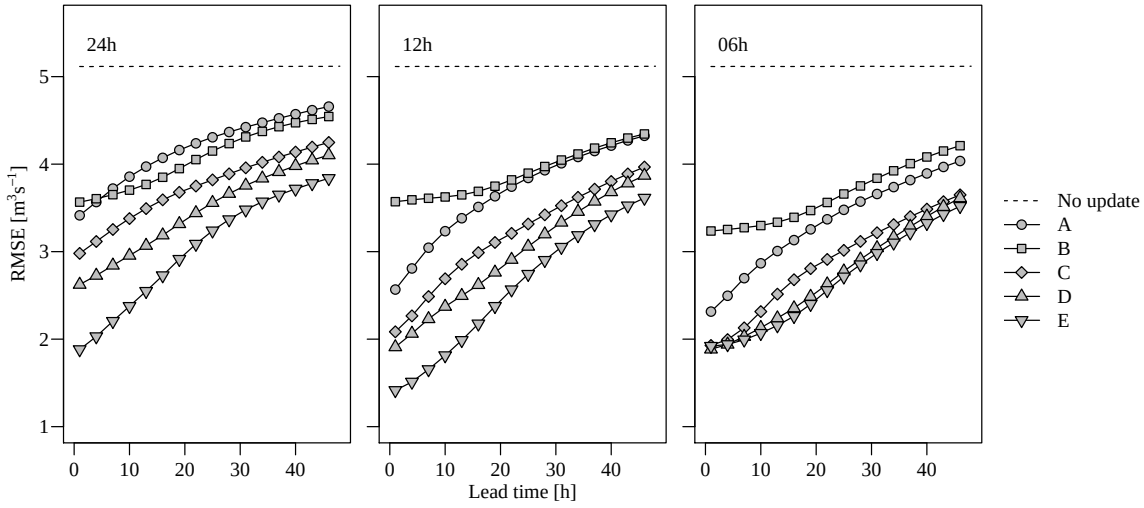


Figure 4.5: Synthetic experiment, simulation period from 15 August 2002 to 15 January 2003. Root-mean-square error at Tabreux for different discharge observation vectors. Forecast issued every 6 h. EnKF assimilation was carried out every 24h (left), 12h (centre), 6h (right).

(2) For the updating frequency of 12 h, there is even a further reduction in RMSE for all the observation vectors. The largest reduction is achieved for case E, in which the RMSE at the lead time of 1 h is 1.4 $m^3 s^{-1}$ (5 % of the mean observed discharge). Additionally, case A (one gauge at the outlet) outperforms case B (two interior gauges only) for lead times up to about 20 h, which is in line with the channel travel time from the most upstream gauges to the outlet.

(3) For the updating frequency of 6 h, there is not a pronounced improvement in RMSE. This can be expected, because within the 6 h between updating moments, hardly any rainfall is transformed into discharge, even at the most upstream gauges, as is shown in Table 4.1.

State updating

A further logical step in the analysis of the results is to have a look at the ability of the DA procedure to correctly update the model states. In other words, we wanted to check if the setup of the EnKF can identify the pdf around the true model states. However, we recall that there is not a single configuration of model states yielding one discharge value due to the fact that our system is spatially distributed.

We investigate the effect of the observation vector at the updating frequency of 24 h. We selected 2 locations (location 1 and 6 in Fig. 4.1b) within the catchment domain for which the simulated and updated model states are presented in Figs. 4.6 and 4.7. Figures 4.6 and 4.7 show in the top panel the open loop simulations for five model states with the highlighted true model states. For two time instants, on 3 November and on 2 January (dashed vertical line), we show the histograms of the 24 h lead time forecasted and the updated model states for the five discharge observation vectors (cases A, B, C, D and E, see Fig. 4.4). The true model states are indicated by asterisks. Recall that the true state for the synthetic experiment was a randomly selected sample. Note that the snow model states are not shown because there was no snow simulated during the 5-month period. We have chosen

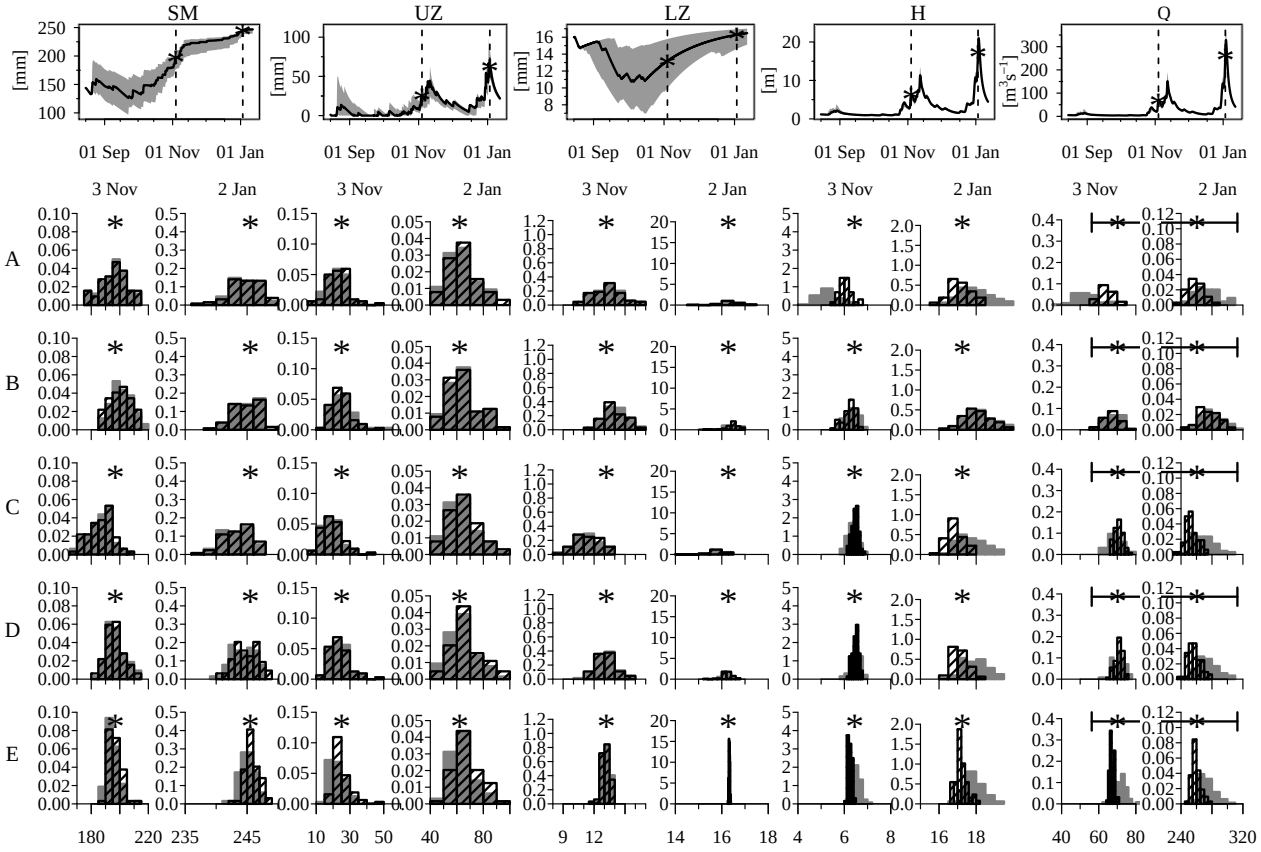


Figure 4.6: Synthetic experiment. Top: Open loop model state simulations (grey lines) including the true states (black line) at location 1 in Fig. 4.1b. Bottom: Forecasted (grey histograms) and analysed (dashed histograms) model states at location 1 for two dates (3 November and 2 January) and considering 5 discharge observation vectors (A, B, C, D, E). The true states are indicated by asterisks and the error bars represent the 95% confidence bounds of the observation errors (ν_k in Eq. 4.13).

3 November and 2 January, because both dates occur shortly before a discharge peak, although the wetness conditions of the catchment are different. The first and smaller peak is observed when the model storages are rather dry. The second and larger peak occurs after an extensive rainy period, when the model states became fully saturated.

Figure 4.6 indicates that at the catchment outlet there is hardly any difference between the forecasted and updated model states in soil moisture (SM), upper zone storage (UZ) and lower zone storage (LZ) for all discharge observation vectors and for both dry and wet conditions. However, note that both the forecasted and updated pdfs of SM, UZ and LZ tend to have more accurate peaks around the "true" values for a larger number of assimilated discharge gauges. This means that, even though there is no clear difference between the forecasted and updated pdfs at one time instant, its accumulation over time makes it visible in the higher kurtosis. Therefore, it makes sense to update those rather insensitive model states. Furthermore, the EnKF is well able to identify the "truth" in two routing storages, the water level (H) and the water storage in the channel (Q) on 3 November as well as on 2 January. There is a clear shift of the updated histogram centroid towards the true value for all discharge observation vectors, except for case B, which remains unchanged. This is caused by the fact that case B (Fig. 4.4) consists of two discharge gauges, which are located far away from the

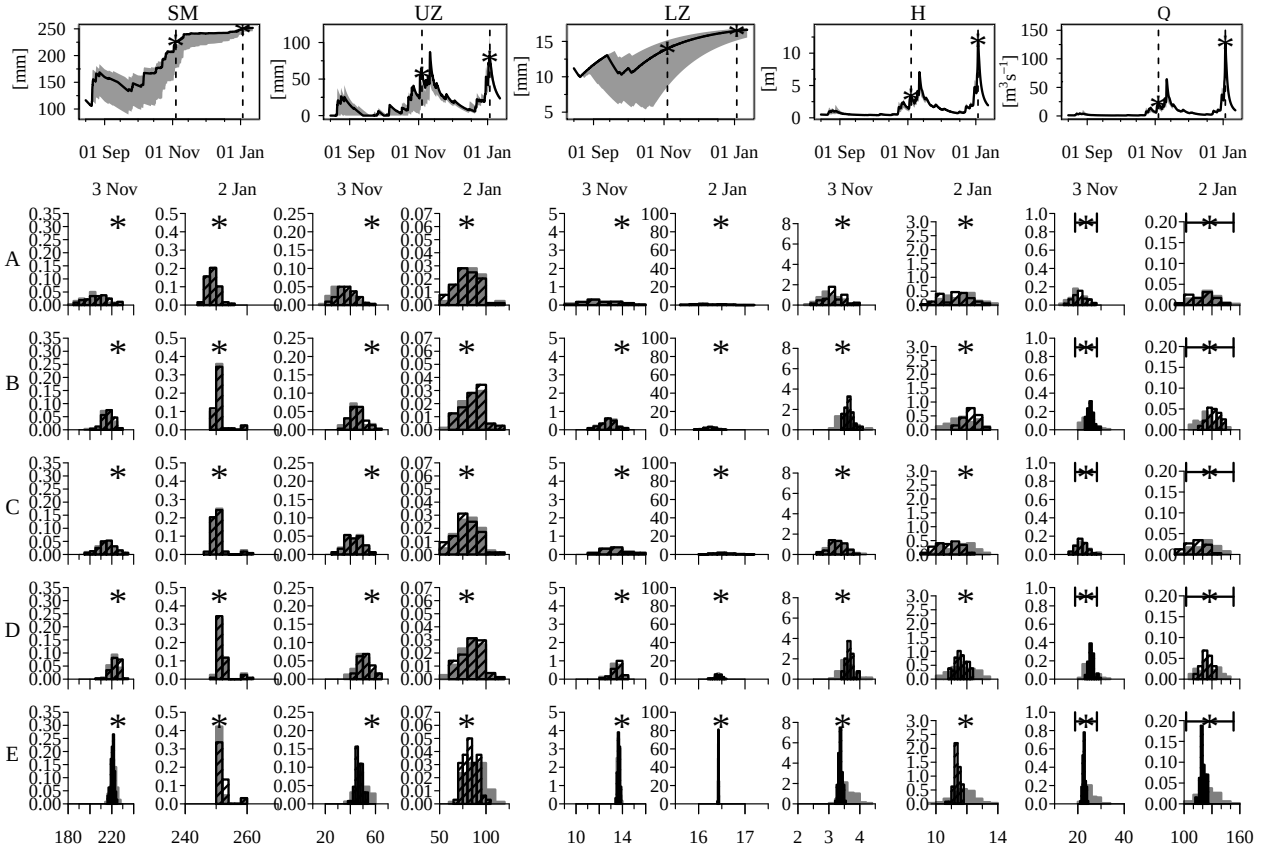


Figure 4.7: Same as Fig. 4.6, but for location 6 in Fig. 4.1b.

catchment outlet. Furthermore, it can be seen that the EnKF is well able to identify the two routing states even if the prescribed discharge observation error bands are larger than the ensemble spread of the forecasted discharge.

At the interior point (Fig. 4.7), similar to the catchment outlet (Fig. 4.6), there is no pronounced update in forecasted and updated histograms for soil moisture (SM), upper zone storage (UZ) and lower zone storage (LZ). The two routing states are again easier to identify. However, the ability of the EnKF to identify the true H and Q depends on the location of stations in the discharge observation vector. For the cases B, D, and E, which contain at least one gauge situated close to location 6, the updated histogram of H and Q approaches the true state and also its shape becomes more peaked. On the other hand, for the cases A and C, which do not include gauges close to location 6, no changes in H and Q histograms occur.

4.3.2 Real world experiment

The long-term RMSE for the real world experiment is shown in Fig. 4.8. Similar to the synthetic experiment, all three discharge observation vectors assimilated into the model improve the forecasted discharge for lead times up to 48 h, except for case B, which slightly deteriorates the forecast performance in comparison with the forecasts without discharge assimilation for lead times longer than 30 h. The best results, meaning the lowest RMSE, are achieved by assimilating all six gauges (case D) for all updating frequencies, although for longer lead times it approaches the benchmark case A (the

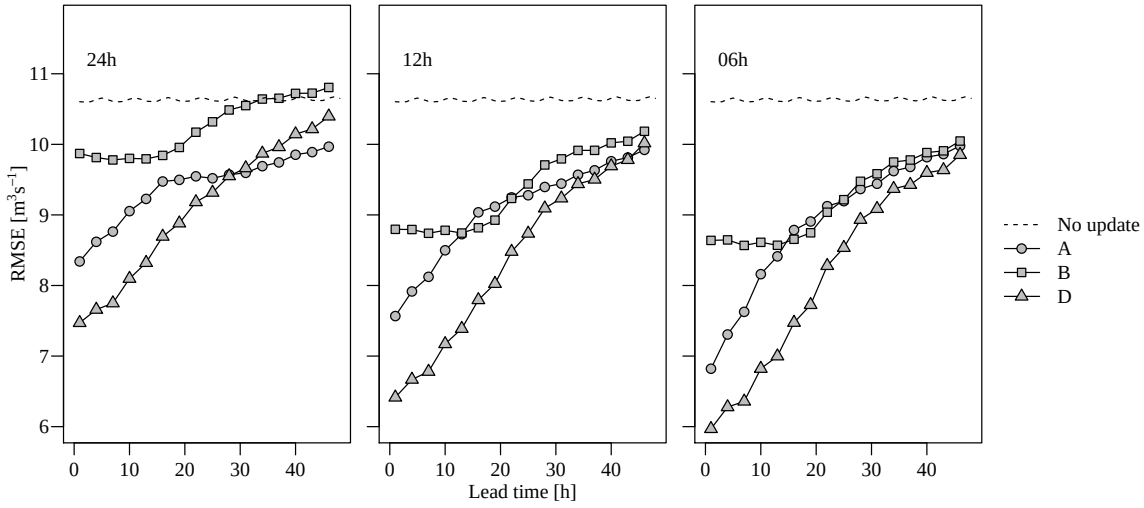


Figure 4.8: Real world experiment, simulation period from 15 August 2002 to 15 January 2003. Root-mean-square error at Tabreux for different discharge observation vectors. Forecast issued every 6 h. EnKF assimilation every 24 h (left), 12 h (centre), 6 h (right).

outlet gauge only). The largest reduction was achieved by case D, for which the RMSE at the lead time of 1 h was about 6 $m^3 s^{-1}$ (20 % of the mean observed discharge) for an updating frequency of 6 h.

Similar to the synthetic experiment, case A (one gauge at the outlet) outperforms case B (two interior gauges only) for all updating frequencies in the real world experiment. Moreover, we can observe for case B a rather constant RMSE during the first 20 h. This surprisingly steady RMSE may be explained by the assimilation effect of the most upstream gauges (locations 5 and 6), for which it takes about 20 h to reach the outlet (location 1). Although an increase in updating frequency from 24 h to 12 h improves the RMSE, a further increase in the updating frequency from 12 h to 6 h yields a more or less equal RMSE, which corresponds to the synthetic experiment.

The short-term RMSE for an individual major flood peak, which was observed at the beginning of January 2003, is shown in Fig. 4.9. Because of the rather short period used in this analysis, the shapes of the RMSE are not smoothed out and the forecasted RMSE without updating is not constant over time either. The best forecast improvements are again achieved by assimilating all six discharge gauges (case D) for all updating frequencies for lead times up to about 15–20 h. For longer lead times, case B (only two upstream gauges) gives very similar RMSE to case D, because the added value of the more downstream gauges (1–4 in Fig. 4.1b) is filtered out after about 20 h, as shown in Table 4.1. It is worth mentioning that case B outperforms case A for lead times from 5 h to 20 h, which is not observed using the long-term statistics (Fig. 4.8). This is caused by the spatial properties of this major flood peak, during which the importance of the upstream gauges is clearly shown, however, completely averaged out in the long-term statistics.

Like in the synthetic experiment, hardly any change between the forecasted and the updated histograms is observed for soil moisture, upper zone storage and lower zone storage (Figs. 4.10 and 4.11), but visible changes can be seen in the routing storages, water level and discharge. For the discharge observation vectors, which contain at least one gauge in the vicinity of the location of the state ob-

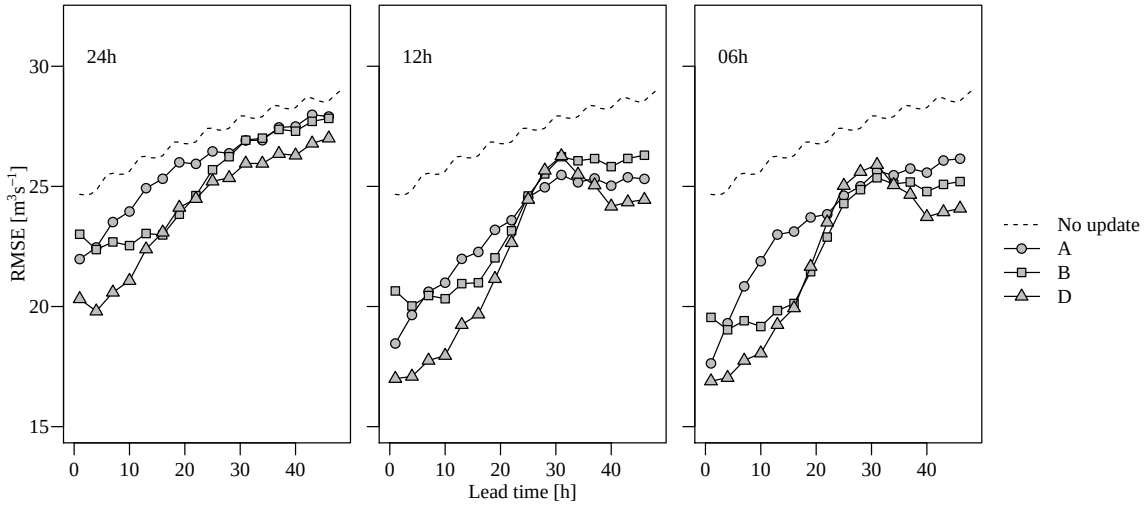


Figure 4.9: Same as Fig. 4.8, but for simulation period from 27 December 2002 to 8 January 2003.

Table 4.3: Validation at the stream gauge locations without data assimilation. The median Nash-Sutcliffe model efficiency of the 64 ensemble members for the open loop simulations and the updated simulations being assimilated using Cases A and B. Gauge numbers correspond to Fig. 4.1b.

Gauge number	Open loop	Case A	Case B
1	0.95	–	0.96
2	0.95	0.96	0.97
3	0.96	0.97	0.98
4	0.96	0.96	0.97
5	0.93	0.91	–
6	0.87	0.89	–

servation, there is a shift of the centroid of the histograms for discharge and the corresponding water level towards the uncertain discharge observation constrained by the error bars.

In order to validate the results of the real world experiments, the Nash-Sutcliffe model efficiency (NS) (*Nash and Sutcliffe*, 1970) is calculated at those stream gauge locations where the discharge data are not assimilated. The median NS of the 64 ensemble updated discharge simulations (Cases A and B) as well as the open loop simulations (without assimilation) are shown in Table 4.3 for the updating frequency of 24 h. The validation results indicate that the assimilated simulations give performance of the same order of magnitude as the open loop simulations. Small improvements can still be observed when the assimilated observations are upstream of the validation stations (Case B). However, when the assimilated observation is downstream of the validation gauges (Case A), the interior simulations can slightly deteriorate, but this is observed only at one gauge (number 5) out of five gauges. Nevertheless, this seems to be rather acceptable when we consider the large distance between the assimilation gauge (number 1) and validation gauge (number 5) and the fact that the two most upstream gauges are located at two parallel river branches (see Fig. 4.1a).

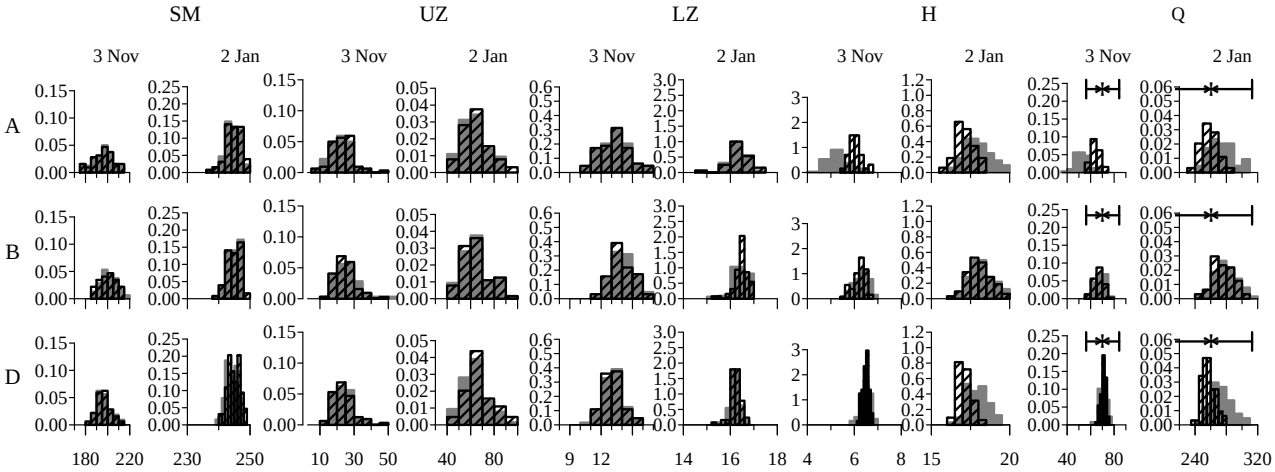


Figure 4.10: Real world experiment. Forecasted (grey histograms) and analysed (dashed histograms) model states at location 1 for two dates (3 November and 2 January) and considering 3 discharge observation vectors (A, B, D). The true states are indicated by asterisks and the error bars represent the 95% confidence bounds of the observation errors (ν_k in Eq. 4.13).

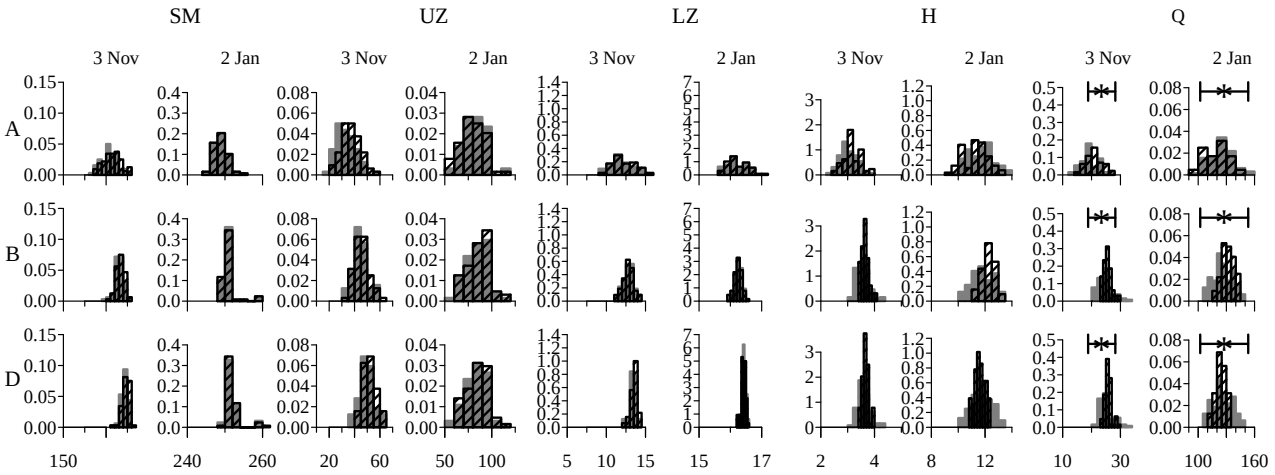


Figure 4.11: Same as Fig. 4.10, but for location 6.

4.4 Discussion

The advantage of a grid-based hydrological model with grid-based routing over a lumped model with a unit hydrograph-type of channel delay is that the modelled discharge is represented by spatially distributed model states, which quantify the volumes of water within the channel network. This means that we do not have to explicitly consider any time delay between model states and discharge, as would be needed in spatially lumped models using the retrospective EnKF (*Pauwels and De Lannoy*, 2006). Another advantage of a grid-based approach over a lumped one is that the spatially distributed

discharge observations can be easily incorporated into the model states and make the forecast more accurate for longer lead times.

A novel approach of this study was the application of time-dependent multivariate spatial conditional simulations (*Goova erts*, 1997; *Pebesma*, 2004; *Rakovec et al.*, 2012a) of hourly rain gauge observations, used to force the hydrological model in hindcasting mode. As demonstrated by *Rakovec et al.* (2012a), this multivariate approach satisfies for each precipitation realisation the requirement to have a coherent temporal evolution (required within the DA procedure), unlike the time-independent univariate simulations. Using this precipitation ensemble generator, we can achieve the goal that the corresponding simulated spatially distributed model states inherit the temporal aspect of the rainfall fields. As an alternative to rain gauge observations, the precipitation ensemble could be derived from radar rainfall estimates from the C-band radar located in the catchment (*Hazenberg et al.*, 2011), which is a possible topic of further studies.

This study also provided a closer look at the pdfs of the forecasted and updated model states during two hydrologically different situations, while the majority of hydrological DA papers on state updating focus only on the forecasted and analysed discharge and do not address the importance of individual model states. In this study, mainly the pdfs of the two routing model storages were affected by the Kalman filter update, while the other model states (SM, UZ, LZ) were found to be less sensitive to the EnKF procedure. This is because the current formulation of the EnKF (see Eq. 4.4) does not explicitly consider the strong correlation between soil moisture states in the immediate past and streamflow at the time of forecast. Therefore, it may be difficult to build a covariance matrix among the water balance model states (i.e., SM, UZ, LZ) via assimilating discharge observations. Based on our results, we can state that, given a measured discharge downstream, it is difficult to adjust (and justify) the soil moisture upstream (in a spatially distributed coherent manner) using an EnKF. Other filters like the Ensemble Kalman Smoother (EnKS), which calculate the analysis from several previous time steps (*Evensen and Van Leeuwen*, 2000), may result in better adjustment of the spatially distributed soil moisture states, which may improve forecasts for even longer lead times. However, with a larger number of assimilated discharge gauges, both the forecasted and updated pdfs of SM, UZ and LZ had more accurate peaks around their true values. Therefore, it makes sense to update those rather insensitive model states. The reason for this behaviour might be the limited model structure, which is similar to other PCRaster operational hydrological applications like the LISFLOOD model (*de Roo et al.*, 2000; *Salamon and Feyen*, 2009), where the individual neighbouring model cells are not connected by means of interflow and regional groundwater flow, but only drained by some sort of sheet flow via the routing states. This means that the SM, UZ and LZ model states are only controlled by the spatial variation of rainfall.

Finally, it is interesting to note that, unlike *Clark et al.* (2008b), we were able to improve hydrological forecasts using the standard EnKF implementation in both synthetic and real world experiments compared to open loop simulations.

4.5 Conclusions and recommendations

We analysed the sensitivity of the data assimilation procedure to the updating frequency, the number and the locations of interior discharge gauges, which were assimilated into a grid-based distributed hydrological forecasting model using the EnKF. By using a physically based model for the routing,

the time delay and attenuation are modelled more realistically than when using a unit-hydrograph approach for of the routing. As a consequence the discharge and states at time k can be assumed to depend only on the states at time $k-1$ (Markov property). The validation station of this study is Tabreux, which is the Upper Ourthe catchment outlet, Belgium.

In the synthetic experiment we showed that the hydrological forecast at the catchment outlet is improved (in terms of the forecasted root-mean-square error RMSE) by assimilating more interior gauges. This is logical, because all other discharge observations contain information from upstream to improve the posterior forecast. In addition, the EnKF procedure is mainly changing the pdfs of the two routing model storages, even when the uncertainty in the discharge simulations is smaller than the defined observation uncertainty. This is because the current formulation of the EnKF does not explicitly consider the strong correlation between soil moisture states in the immediate past and streamflow at the time of forecast. Moreover, with an increasing number of discharge observations, the centroid of the updated histograms within the observation error bounds was approaching the true value more closely and with smaller variance than for the less dense discharge observation networks.

In the real world experiment, the best results in terms of the RMSE were achieved using all observations, which includes all six discharge gauges. Given the travel time of the catchment, an updating frequency of 12 h seems to be the most appropriate. Additionally, similar to the synthetic example, only the two routing model storages showed some sensitivity to the EnKF procedure in terms of the forecasted and updated histograms. We can conclude that the hydrological forecast at Tabreux can be improved by assimilating more upstream gauges using the EnKF data assimilation procedure. This augmentation of the observation vector improves the forecast more than increasing the updating frequency.

For operational use, we recommend the implementation of additional upstream gauges into the observation system, which would enable an increase of the updating frequency and more accurate forecasts, if the polling frequency allows for it. Another recommendation for future research is to have a closer look at alternative model structures (including recalibration of spatially distributed model parameters) and their effect on the sensitivity of individual model states within the EnKF. The main limitation of the current model structure is that there is no flux between neighbouring cells except for the two routing model states. Alternatively, hydrological forecasts can be improved by applying other Kalman-type methods, e.g., the Ensemble Kalman Smoother (*Evensen and Van Leeuwen*, 2000), which calculates the analysis from several previous time steps up to the time of forecast, instead of mapping the instantaneous covariance between states and discharge (*Clark et al.*, 2008b), as shown for the standard EnKF in this study. Finally, additional in situ observations can be considered to be assimilated into the spatially distributed model states, e.g., soil moisture and/or groundwater levels. The latter are believed to resemble point-wise the actual regional water storage more closely than soil moisture observations.

Chapter 5

Operational aspects of asynchronous filtering for improved flood forecasting

This chapter is largely based on the manuscript: Rakovec, O., A. H. Weerts, J. Sumihar, and R. Uijlenhoet (2014), Operational aspects of asynchronous filtering for improved flood forecasting, *In Preparation*.

5.1 Introduction

Understanding the behaviour of extreme hydrological events and the ability of hydrological modellers to improve the forecast skills are distinct challenges of applied hydrology. Hydrological forecasts can be made more reliable and less uncertain by recursively improving initial conditions. A common way of improving the initial conditions is to make use of data assimilation (DA), a feedback mechanism or update methodology which merges model estimates with available real world observations (*Liu and Gupta*, 2007; *Reichle*, 2008; *Evensen*, 2009; *Liu et al.*, 2012).

The traditional implementation of the Ensemble Kalman Filter (EnKF; see Sect. 4.2.4) is synchronous, commonly named a three dimensional (3-D) assimilation, which means that all assimilated observations correspond to the time of update. Asynchronous DA, also called four dimensional (4-D) assimilation, refers to an updating methodology, in which observations being assimilated into the model originate from times different to the time of update (*Evensen*, 1994, 2009; *Sakov et al.*, 2010). Ensemble Kalman Smoother (EnKS; *Evensen*, 2009) is a common example of asynchronous DA method used in hydrology (e.g. *Dunne and Entekhabi*, 2006; *Crow and Ryu*, 2009; *Li et al.*, 2013b), which extends the EnKF by introducing additional information by propagating the contribution of future measurements backward in time. EnKS reduces the error variance as compared to the EnKF (*Evensen*, 2009). Traditionally, smoother is different from filter: a smoother assimilates "future observations", while filter assimilates "past observations". This implies that for operational forecasting purpose, we need a filter rather than a smoother. A smoother can help improve the model accuracy in the past (e.g. for re-analysis), but it does not help improve forecast accuracy (*Evensen*, 2009).

Therefore, *Sakov et al.* (2010) introduced the Asynchronous Ensemble Kalman Filter (AEnKF), which requires forward integration of the model to obtain simulated results necessary for the analysis and the model is updated at the analysis step using past observations over a time window. The difference among the EnKF, EnKS and AEnKF is schematized in Fig. 5.1. *Sakov et al.* (2010) showed that the formulation of the EnKS provides a method for asynchronous filtering, i.e. assimilating past data at once and that the AEnKF is a generalization of the ensemble-based data assimilation technique. Moreover, unlike the 4-D variational assimilation methods, the AEnKF does not require any adjoint model (*Sakov et al.*, 2010). The AEnKF is particularly attractive from an operational forecasting perspective as more observations can be used with hardly any extra additional computational time. Additionally, such an approach can potentially account for a more proper representation of the time-lag between the internal model states and the catchment response in terms of the discharge.

The applicability of the Kalman-type of assimilation methods was developed for an idealized modelling framework with perfect linear problems with Gaussian statistics, however, it has been proven to work well with a large number of different nonlinear dynamical models (*Evensen*, 2009). Nevertheless, it is still interesting to evaluate, whether elimination of some non-linear nature from the model updating can be beneficial. For example, *Xie and Zhang* (2013) introduced the idea of partitioned update scheme to reduce the degree of freedom of the high dimensional state-parameter estimation of a distributed hydrological model. In their study, the partitioned update scheme enabled to better capture covariances between states and parameters, which prevented spurious correlations of the non-linear relations in the catchment response. Similarly, decreasing the number of model states being perturbed and updated was suggested by *McMillan et al.* (2013) to increase the efficiency of the filtering algorithm by conserving the forecast quality. Such an approach was proposed especially

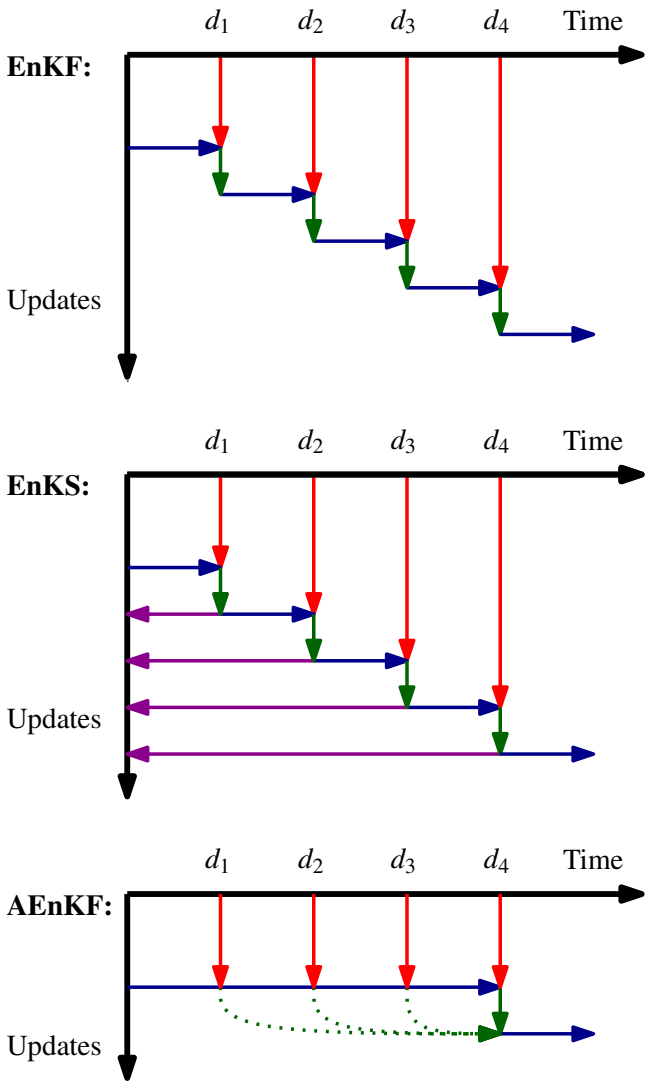


Figure 5.1: Illustration of the model updating procedure for the Ensemble Kalman Filter (EnKF), the Ensemble Kalman Smoother (EnKS), and the Asynchronous Ensemble Kalman Filter (AEnKF). The horizontal axis stands for time, observations (d_1, d_2, d_3, d_4) are given at regular intervals. The blue arrows represent forward model integration, the red arrows denote introduction of observations and green arrows indicate model update. The magenta arrows represent the model updates for the EnKS, therefore they go backward in time, as they are computed following the EnKF update every time observations become available. The green doted arrows denote past observations being assimilated using the AEnKF. The schemes for the EnKF and the EnKS are after *Evensen* (2009).

to states with small innovations, which was mainly the soil moisture storage.

In this chapter we present a follow-up on the work of *Rakovec et al.* (2012b) (Chapter 4), in which discharge observations were assimilated into a grid-based hydrological model for the Upper Ourthe catchment in Belgian Ardennes by using the AEnKF in comparison with the EnKF. To our knowledge this is the first application of AEnKF in a hydrological forecasting application. Firstly, the effect of state augmentation using a time window of past observations on forecast accuracy is analyzed. Secondly, the effect of partitioned updating scheme is scrutinized.

5.2 Methodology

5.2.1 Hydrological modelling using OpenStreams, OpenDA and Delft-FEWS

In this chapter we use a grid-based spatially distributed HBV-96 model for the Upper Ourthe, and we refer to Sect. 4.2.3 and Fig. 4.2 for the model structure description and definition of the hydrological states. In contrast to *Rakovec et al.* (2012b), in the current study we employed the HBV-96 model built within a recently developed open source modelling environment *OpenStreams* (2013), which is suitable for integrated hydrological modelling based on the Python programming language with PCRaster spatial processing engine (*Karssenberg et al.*, 2009; *PCRaster*, 2013). The advantage of using *OpenStreams* (2013) is that it enables direct communication with *OpenDA* (2013), an open source data assimilation toolbox. *OpenDA* (2013) provides a number of algorithms for model calibration and assimilation and is suitable to be connected to any kind of environmental model.

The import and export of hydrological and meteorological data to the system is done using Delft Flood Early Warning System (Delft-FEWS, *Werner et al.*, 2013), an open shell system for managing forecasting processes and/or handling time series data. Delft-FEWS is modular and highly configurable system, and it is also used by the Dutch authorities for the flood forecasting for the River Meuse basin (called RWsOS Rivers), in which the Upper Ourthe is located. The current configuration is employed in a stand-alone version of RWsOS Rivers, however, it can be easily switched into a configuration with real time data import.

5.2.2 Asynchronous Ensemble Kalman Filter (AEnKF)

The AEnKF (*Sakov et al.*, 2010) should not be considered as a new method, but rather a simple modification of the EnKF using a state augmentation approach. Therefore, the derivation of the AEnKF is based on the equations for the EnKF, which were presented in Sect. 4.2.4.

This means that the matrix of the model states ($\mathbf{X}_k$) at time k (see Eq. 4.3) is augmented with the past forecasted observations **HX** (i.e., model outputs corresponding to the observation locations) from W previous time steps, which yields

$$\tilde{\mathbf{X}}_k = \begin{pmatrix} \mathbf{X}_k \\ \mathbf{H}_{k-1}\mathbf{X}_{k-1} \\ \mathbf{H}_{k-2}\mathbf{X}_{k-2} \\ \vdots \\ \mathbf{H}_{k-W}\mathbf{X}_{k-W} \end{pmatrix}. \tag{5.1}$$

Remember that the number of rows in $\mathbf{X}_k$ and **HX** can significantly differ: $\mathbf{X}_k$ contains the complete set of model states, while **HX** contains only the forecasted observations. Additionally, with the new state definition comes a new augmented observer operator $\tilde{\mathbf{H}}_k$ (in which **I**, with the corresponding subscript, stands for identity elements on diagonal matching the dimensions in Eq. (5.1)), a new

augmented observation vector $\tilde{\mathbf{y}}_k$ and its corresponding observation covariance matrix $\tilde{\mathbf{R}}_k$:

$$\tilde{\mathbf{H}}_k = \begin{pmatrix} \mathbf{H}_k & & & \\ & \mathbf{I}_{k-1} & & 0 \\ & & \mathbf{I}_{k-2} & \\ & 0 & & \ddots \\ & & & & \mathbf{I}_{k-W} \end{pmatrix}, \tag{5.2}$$

$$\tilde{\mathbf{y}}_k = \begin{pmatrix} \mathbf{y}_k \\ \mathbf{y}_{k-1} \\ \mathbf{y}_{k-2} \\ \vdots \\ \mathbf{y}_{k-W} \end{pmatrix}, \tag{5.3}$$

$$\tilde{\mathbf{R}}_k = \begin{pmatrix} \mathbf{R}_k \\ \mathbf{R}_{k-1} \\ \mathbf{R}_{k-2} \\ \vdots \\ \mathbf{R}_{k-W} \end{pmatrix}. \tag{5.4}$$

Having these augmented equations for $\tilde{\mathbf{X}}_k$, $\tilde{\mathbf{H}}_k$, $\tilde{\mathbf{y}}_k$ and $\tilde{\mathbf{R}}_k$, it is straightforward to carry out the assimilation in a same manner as presented in Sect. 4.2.4. Note that if the time window equals the current single time step, then $W = 0$ and the AEnKF problem reduces to the traditional EnKF.

From the operational point of view, it is preferable to have a longer assimilation window, because less frequent assimilation eliminates a disruption of the ensemble integration by an update and a restart. When assimilation is done more frequently, it can cause considerably higher calculations costs, which can often be a burden for real time operational settings (*Sakov et al.*, 2010). The AEnKF uses a longer assimilation window and assimilates all observations in a single update. This makes the AEnKF attractive to be used. The added value of a longer assimilation window will be a subject for investigation in this work. Especially, it can provide an improved representation of the time-lag between the internal model states and the catchment response in terms of the discharge. Such an idea was investigated for example by *Li et al.* (2013b), who compared the effect of time-lag representation using the EnKF and EnKS.

5.2.3 Error model

The uncertainty in the hydrological model is represented by an ensemble (set of realizations) of stochastic errors, which is used to perturb the hydrological model. An ensemble of errors is a finite and discrete number of spatial realisations over time and its statistical properties can be defined using the spatial and temporal characteristics, for example a decorrelation distance. In this chapter the stochastic error fields were directly added to the soil moisture model states, unlike the procedure in Chapter 4 (*Rakovec et al.*, 2012b), where the conditionally simulated rainfall fields were directly used to obtain the ensemble of model states. Therefore, in this chapter the hydrological model was

forced using deterministic spatially distributed rainfall fields, which were obtained by inverse distance interpolation.

In this chapter, the source of model uncertainty is assumed to come from uncertain soil moisture model representation. To model this uncertainty, an AR(1) process is used to represent a temporally correlated noise process with a temporal de-correlation length of 4 hours. The noise process is further assumed spatially isotropic with a spatial de-correlation length of 30 km. The noise is assumed to have a spatially uniform standard deviation of 1 mm. The 2D noise field with such statistics were generated by using *OpenDA* (2013) toolbox. This ad-hoc modification was needed because of an incompatibility of the computer code implementations. However, the aforementioned parameterization of the current error ensures that the ensemble spread in the simulated discharge corresponds well with the control simulations as presented in Chapter 4 (not shown).

5.2.4 Experimental setup

This chapter evaluates a configuration setup of the Asynchronous Ensemble Kalman Filter (AEnKF) (Sect. 5.2.2) to assimilate discharge observations into a spatially distributed hydrological model of the Upper Ourthe catchment (Sect. 4.2.1). The objective is to improve the hydrological forecast at the catchment outlet (at Tabreux, gauge 1 in Fig. 4.1b) by assimilating up to 4 discharge gauges. However, validation at an independent location will also be performed. The discharge assimilation is done every 24 h, however, the forecasts are issued every 6 h, i.e. 4 times a day, with different independent starting points at 00:00 UTC, 06:00 UTC, 12:00 UTC, 18:00 UTC, which is the same implementation as in Chapter 4. The study analyses the 8 largest floods peaks observed within the catchment since 1998. An overview is provided in Table 5.1.

Table 5.1: Overview of the periods used in this study.

Period	Number of events	Maximum observed discharge $[m^3 s^{-1}]$
23 October 1998 – 15 November 1998	1	210
15 February 1999 – 5 March 1999	2	195
15 January 2002 – 6 March 2002	4	340
21 December 2002 – 7 January 2003	1	380

We limit the analysis of the experiment to input uncertainty only. The main reason for this is that we want to fully understand and investigate the filter process using a distributed hydrological model and realistic precipitation fields. We believe this is already challenging enough without initial state and model parameter/structural uncertainty, which we leave for future work. Additionally, since we focus on the uncertainty in the initial model states, we disregard the large uncertainty in the meteorological forecasts. Therefore, observed meteorological data are used to force the spatially distributed model. In this case, inverse distance interpolation was used to derive spatial rainfall estimates.

The ensemble of uncertain model simulations is obtained by perturbing the soil moisture state (SM) with the spatio-temporally correlated error model (Sect. 5.2.3). The ensemble size in this study was defined to be 36 realizations (for computational reasons). The error in the discharge observations

is considered to be a normally distributed observation error with a variance of $(0.1\,Q_{\mathrm{obs},k})^2$ (after e.g. *Weerts and El Serafy*, 2006; *Clark et al.*, 2008b; *Rakovec et al.*, 2012b).

The applicability of the Kalman-type of assimilation methods was developed for an idealized modelling framework with perfect linear problems with Gaussian statistics, however, it has been proven to work well with a large number of different nonlinear dynamical models (*Evensen*, 2009). Therefore, the influence of the filtering procedure on the forecast accuracy will be investigated through the effect of a partitioned state updating scheme. This means that besides updating all of the model states, we will test two other alternatives. The first alternative will leave out from the model analysis the soil moisture state (noSM), which is known to exhibit the most non-linear relation to Q. The second alternative will eliminate all the modes states except for the two routing ones (HQ). The scenarios of partitioned state updating schemes are shown in Table 5.2, including the control run without state updating (no update).

Table 5.2: Four partitioned state updating schemes (indicated in the first column) for 5 model states (indicated in the first row) being updated and thus included in the model analysis. Model states are described in Sect. 4.2.3 and Fig. 4.2 and has following acronyms: discharge (Q), water level (H), soil moisture storage (SM), upper zone storage (UZ), and lower zone storage (LZ).

name	Q	H	SM	UZ	LZ
no update					
all	√	√	√	√	√
noSM	√	√		√	√
HQ	√	√			

The experimental setup scrutinizes the problem of asynchronous filtering from two perspectives. First, we investigate the effect of state augmentation using the past asynchronous observations and assimilation of distributed observations on the state innovation. Second, the choice of which model states are included in the analysis step to be updated is analysed. The performance of the data assimilation procedure regarding discharge forecasting is evaluated using the root-mean-square error versus lead time (RMSE, Eq. 4.14).

5.3 Results

5.3.1 The effect of state augmentation and distributed observations on state innovation

To investigate and understand the effect of augmented operators (Eqs. 5.1, 5.3, and 5.2) on the innovation of spatially distributed model states, we present the following example. Figure 5.2 shows discharge simulations and corresponding discharge observations at 4 locations within the catchment on 31 December 2002, 00:00 UTC. Note that the magnitude of the discharge observations corresponds to the location within the catchment; for downstream gauges the magnitude is larger than for the more upstream gauges. The discharge observations are further distinguished according to the time window length of the state augmentation, which is set to $W = 0$ and $W = 11$. The first example represents

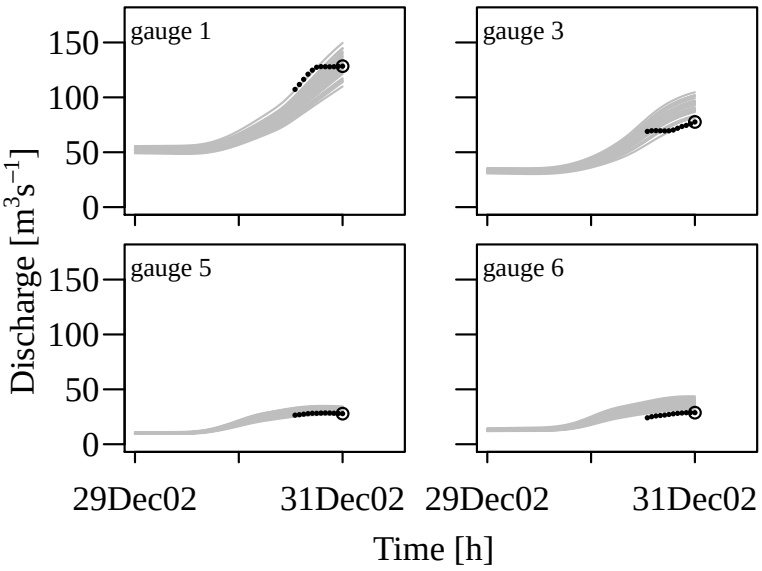


Figure 5.2: Discharge ensemble forecasts (grey lines) and observations (points) at four locations (gauges 1, 3, 5, 6; see Fig. 4.1b). Observations being assimilated using the AEnKF are schematized according to the state augmentation size for two scenarios: assimilation of the data from the current time step $W = 0$ (open circle, traditional EnKF approach) and assimilation of data including the previous 11 time steps $W = 11$ (black dots). The observations are assimilated into the model states on 31 December 2002, 00:00 UTC.

the traditional EnKF algorithm, while the latter one assimilates observation from a 12-hour time window (i.e., 1 current observation and 11 past observations), which is arbitrarily defined as a half of the 24-hour assimilation time window. For some cases alternative assimilation windows were tested, however that did not lead to noticeable differences (not shown).

The mean difference between the forecasted and updated model states for the whole ensemble is illustrated in Fig. 5.3 for four scenarios. These examples improve our understanding about the behaviour of the updated model states in relation to the information content of the observations from two perspectives: (1) the effect of the state augmentation with past observations and (2) the effect of assimilating spatially distributed observations into a grid-based hydrological model.

Let us first consider the traditional EnKF (i.e., no state augmentation with $W = 0$) to update all the grid-based model states by assimilating the observation at the catchment outlet (gauge 1). We observe that the single observation is measured approximately in the middle of the simulated ensemble (see the open circle for gauge 1 in Fig. 5.2). Therefore, there is hardly any difference between the forecasted and updated model states as we show in Fig. 5.3a. In the second scenario, we still assimilate only one gauge at the outlet, however, we use the augmented operators with $W = 11$. Because the mean of the ensemble simulations is predominantly underestimated as compared to the assimilated observations (see black dots in Fig. 5.2 for gauge 1), after the update more water is added spatially equally into the system, as shown in Fig. 5.3b. In the third scenario, we include all 4 gauges being assimilated into the model without any augmentation. Because the model simulations at the interior gauges are mostly overestimating the observations, water is removed from the catchment after the update. Moreover, since the model overestimation is largest at gauges 3 and 6, we can also observe in Fig. 5.3c how

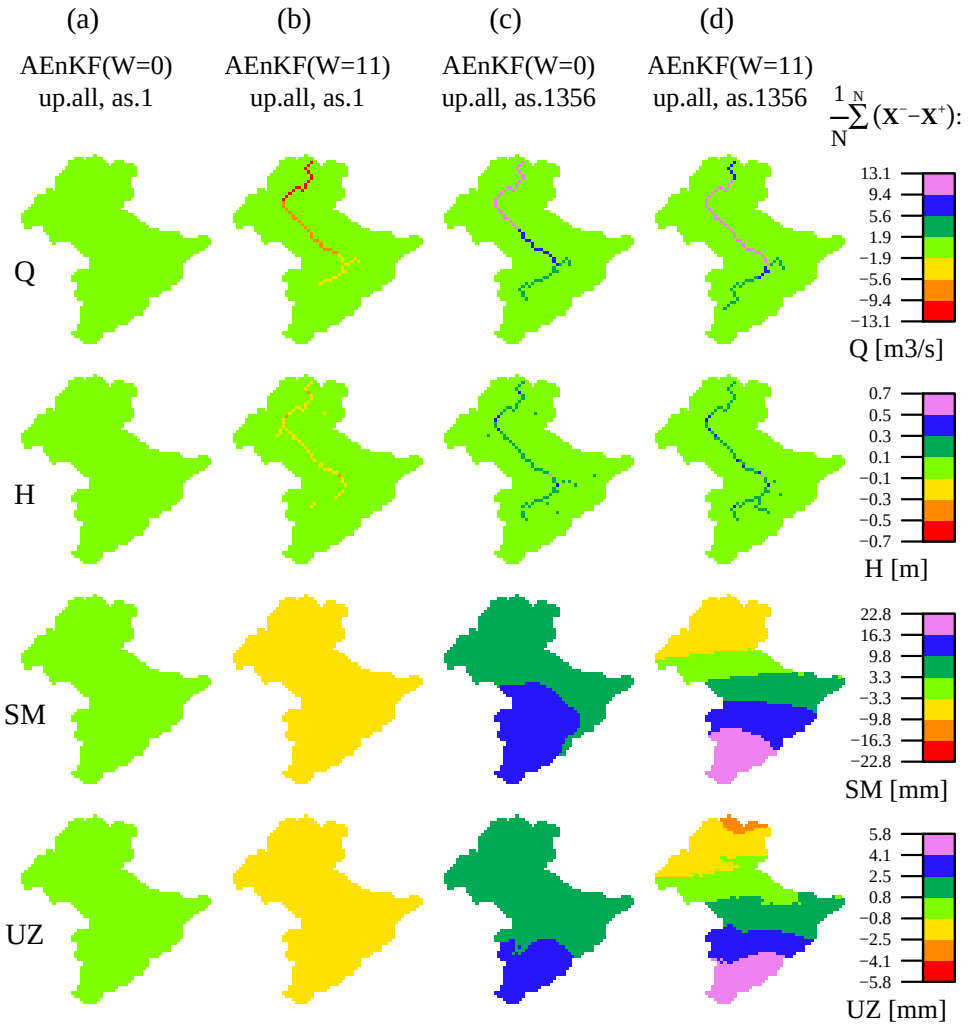


Figure 5.3: Mean difference between the forecasted ($\mathbf{X}^-$) and updated ($\mathbf{X}^+$) model states on 31 December 2002 at 00:00 UTC for different scenarios (shown in vertical panels). We show only 4 sensitive model states: discharge (Q), water level (H), soil moisture (SM) and upper zone (UZ). We dropped out the insensitive lower zone (LZ). Notation $W = 0$ and $W = 11$ indicates the size of the state augmentation. Notation up.all indicates that all of the model states are updated. Notation as."xx" indicates the gauges which are assimilated, see Fig. 4.1b for their locations. The corresponding ensemble of model forecasts and observations being assimilated are shown in Fig. 5.2.

well the EnKF is capable to identify corresponding regions in a spatial manner. In the fourth scenario (Fig. 5.3d) we still assimilate all 4 gauges, however we augment the state with $W = 11$. We can observe that the innovation of the model states gets even more spatially differentiated; the updated SM and UZ model states in the downstream part of the catchment increases the amount of water in the system, while the updated SM and UZ model states in the upstream part decreases the amount of water in the system.

The presented educational examples showed an update for several scenarios starting from the same initial conditions. This enabled a fair comparison between scenarios, however, the sensitivity of state augmentation needs to be further scrutinized in terms of its cumulative effect over time.

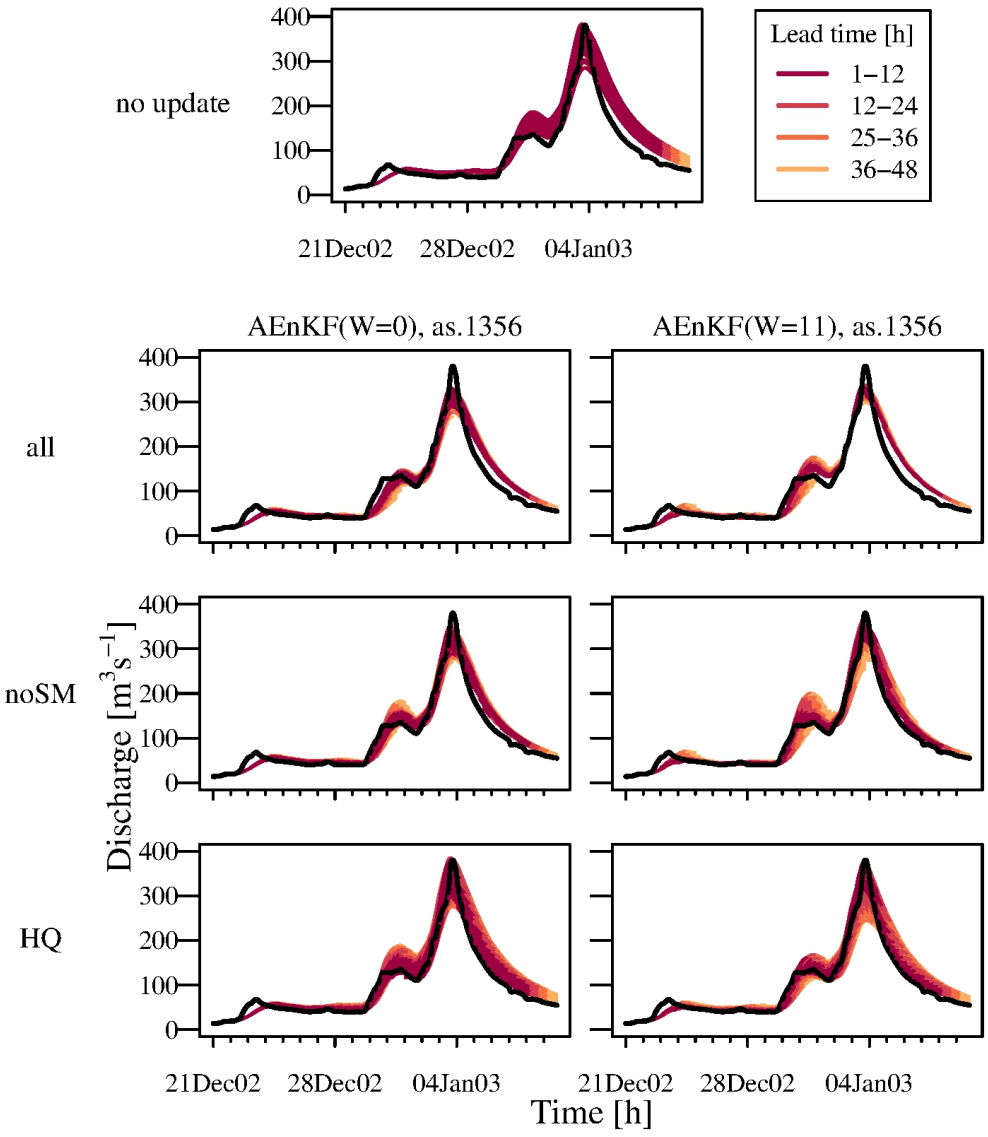


Figure 5.4: Ensemble of discharge forecasts for a typical event at the catchment outlet (Tabreux, gauge 1) for different updating scenarios. The control run (with no update) is shown in the top panel. The combined effect of the model states being updated (3 scenarios shown in rows) and the length of the state augmentation vector (W) of past observations being assimilated (2 scenarios in columns) is presented. The observations are shown in black. Gauges 1, 3, 5, and 6 are assimilated.

5.3.2 The effect of the four partitioned update schemes and asynchronous assimilation and on forecast accuracy

We present a qualitative interpretation of the hydrological forecasts with a lead time of 48 h in Fig. 5.4 for different partitioned state updating schemes as defined in Table 5.2, including both a non-augmented state ($W = 0$) and an augmented state ($W = 11$). This analysis focuses on a characteristic winter flood event (December 2002 – January 2003) being typical for a moderate temperate climate caused by a fast-moving frontal stratiform system (*Hazenberg et al.*, 2011). We observe that

the ensemble of the control runs (top panel of Fig. 5.4) simulates the major flood peak reasonably well, including the timing and the magnitude. When discharge assimilation is employed, an overall reduction of the uncertainty in the forecasted ensemble is observed. Nevertheless, the forecasted flood peak becomes underestimated and the forecasted recession remains overestimated, which is acceptable because of the defined uncertainty in the observed discharge. This happens in particular for the scenario, in which all states are updated, and there are marginal differences between the non-augmented and augmented model states. Furthermore, when we leave out SM from the state update (noSM), we can observe that the major flood peak is forecasted more accurately, including the rising limb around 31 December 2002. Moreover, for the augmented state with $W = 11$, the ensemble spread becomes somewhat wider for lead times >12 h than for the non-augmented state. Nevertheless, the observations correspond approximately with the ensemble mean. Finally, we present the effect of the scenario in which only the two routing states are updated. The results suggest that the flood peak is captured most accurately of all scenarios, however with somewhat wider uncertainty bands. Therefore, it seems more appropriate to retain the UZ storage (noSM scenario) in the model state updating, which represents water storage available for quick catchment response in the concept of the HBV model.

Besides a qualitative interpretation of the forecasted hydrographs presented in Fig. 5.4 for one particular event, we summarize these results in a more quantitative manner for the whole set of 8 flood events (see Table 5.1) in terms of the root-mean-square-error (RMSE) versus lead time. This is shown in Fig. 5.5 for different partitioned state updating schemes and for three scenarios for the state augmentation. The control model run with no update has a constant RMSE of about 32 $m^3 s^{-1}$ and an improved hydrological forecast has a RMSE lower than the control run. The results suggest that all assimilation scenarios improve the hydrological forecast, however with marked differences between the scenarios.

We can observe that updating all model states except for SM (noSM scenario) consistently leads to the most accurate forecasts across the whole range of lead times. Additionally, state augmentation using $W = 5$ and $W = 11$ indicates improvements compared to the case without augmentation ($W = 0$). However, for lead times longer than the travel time from the most upstream gauges and to the outlet (i.e. > 20 h), the difference between state augmentations $W = 5$ and $W = 11$ diminishes. Moreover, when only the two routing states (HQ scenario) are updated, the RMSE is lowered for short lead times, but the improved effect does not last too long as compared to the noSM scenario. The smallest improvement at shorter lead times is achieved when all model states are updated (all scenario), and this is due to the strongly non-linear relation between the assimilated observations and the SM storage, which is further pronounced by the time-lag between the state and the catchment response. Nevertheless, for longer lead times it seems slightly better to update all states rather than only the routing states.

Validation of the methodology is presented in Fig. 5.6 for an independent evaluation of the forecasting results at Durbuy (see gauge 2 in Fig. 4.1b), an interior location which was not used for assimilation. These results show that an improvement of discharge assimilation also occurs at the validation location and that the pattern corresponds well to the results presented in Fig. 5.5. Such an analysis indicates that there is no spurious update of the model states.

To reveal the temporal nature of the model being updated using the AEnKF, using $W = 0$ and $W = 11$, we present in Figs. 5.7 and 5.8 time series of normalized differences between the ensemble

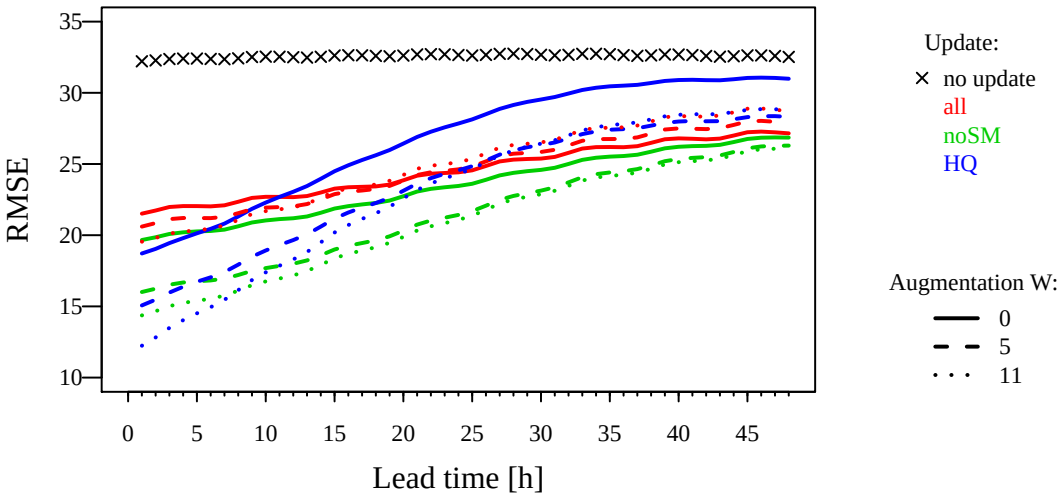


Figure 5.5: Root-mean-square-error at Tabreux (gauge 1) for different discharge observation vectors for which different model states are updated and with different lengths of the state augmentation vector (W) of past observations being assimilated. The results incorporate a set of 8 flood events shown in Table 5.1. Gauges 1, 3, 5, and 6 are assimilated.

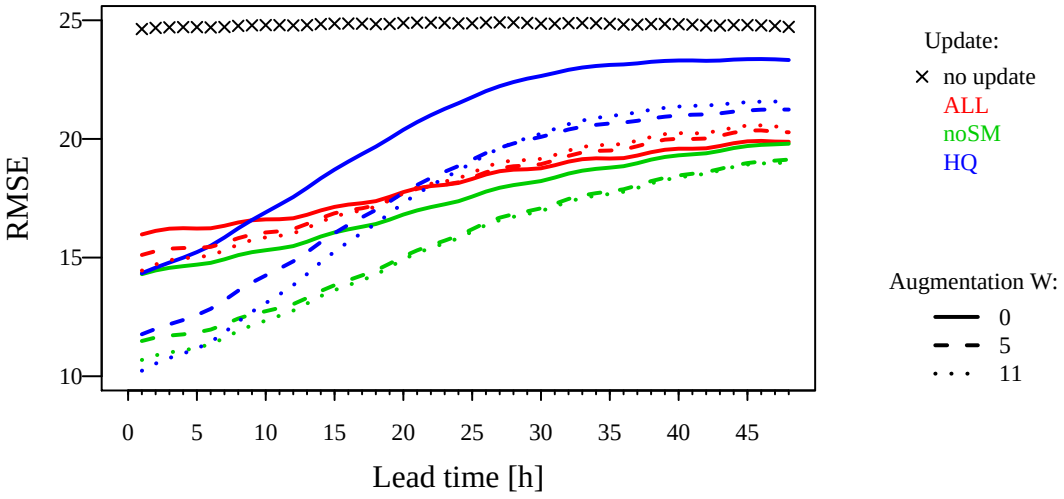


Figure 5.6: Same as figure 5.5, but the results are presented for Durbuy (gauge 2), a validation location which is not assimilated.

means for the 3 partitioned update schemes and the ensemble mean for the no update scenario. The normalization is achieved by dividing the aforementioned difference by the no update scenario mean. In such a way we obtain the relative change in each of the model states.

For the AEnKF using $W = 0$ (Fig. 5.7), we can observe that for the all scenario, which updates all the model states, the magnitude of the percentage change is approximately the same for all 4 model states and it ranges up to 25%. When all model states except for the SM are updated, no changes in the

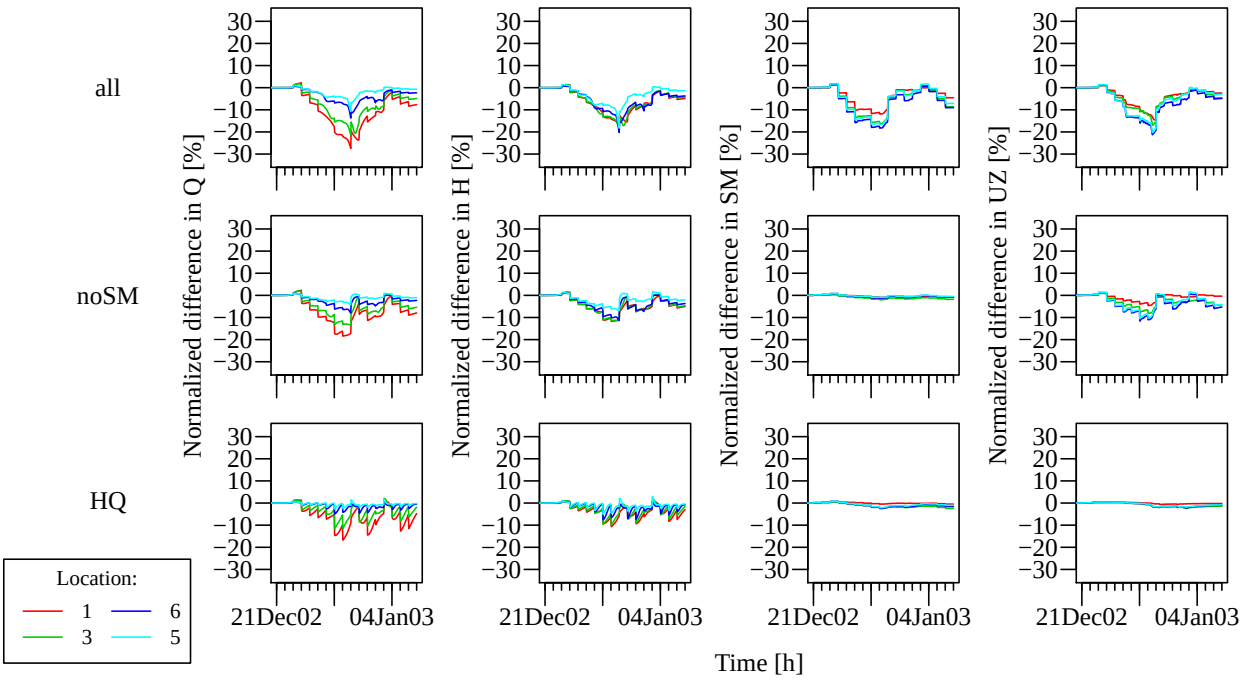


Figure 5.7: Scaled difference between the ensemble mean for the 3 partitioned update schemes and the control run without data assimilation at 4 gauging locations (shown by different colors) within the Upper Ourthe catchment using the AEnKF with $W = 0$. We dropped out the insensitive lower zone (LZ). Gauges 1, 3, 5, and 6 are assimilated. The results corresponds to the same period, which was presented in Fig. 5.4.

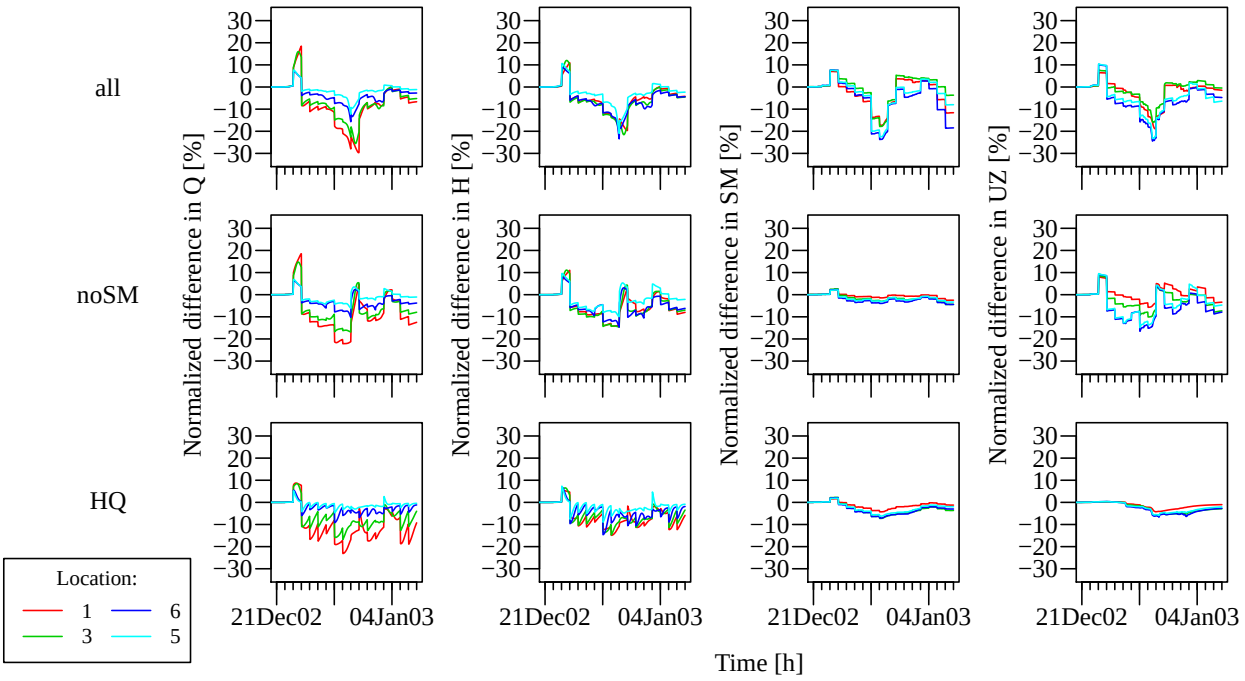


Figure 5.8: Same as Fig. 5.7 but with $W = 11$.

SM storage occur and the overall magnitude of the changes in the other states is slightly decreased and smoothed. Furthermore, when only the two routing states are updated (HQ), the SM and UZ storages remain unchanged over time and we observe a different temporal behaviour of the routing

states in comparison with the previous cases. For the HQ scenario, the updated time series have a clear zigzag shape, which indicates that the effect of update diminishes faster, because only the river channel is updated. In contrast, the routing states for the other cases show a more stable behaviour over time, illustrated by the stepwise shape. These more persistent results correspond to the updates in the UZ storage, which is used for a quick catchment response and has an impact for a longer time. The benefits of including the UZ storage in the update and leaving the SM storage out was already presented from a different point of view in Fig. 5.5 for longer lead times.

For the AEnKF using $W = 11$ (Fig. 5.7), we can observe that the overall pattern of the temporal changes in the model states is similar as for $W = 0$, but the behaviour of using $W = 11$ shows somewhat larger variability. By assimilating more observations ($W = 11$), we expect even a larger update, assuming that more observations contain more information about the unknown truth. Assuming the underlying forecast model has a significant error, by assimilating more observations the Kalman filter will pull the model even closer to the truth yielding larger not-so-smooth update.

5.4 Discussions and Conclusions

We applied the Asynchronous Ensemble Kalman Filter (AEnKF) (*Sakov et al.*, 2010) and identified the effect of augmenting the state vector with past simulations and observations. To our knowledge this is the first application of AEnKF in a hydrological forecasting application. We showed that the effect of an augmented assimilation vector improves the hydrological forecasts, but the contribution gets smaller for longer lead times. Overall, the AEnKF can be considered as an effective method for model state updating taking into account more (e.g. all) observations at hardly any additional computational burden. This makes it very suitable for operational hydrological forecasting. The AEnKF requires the choice of a certain assimilation window length which adds a degree of freedom to the data assimilation scheme. The optimal window is very likely related to the catchment size (i.e. concentration time). It was noted (not shown) that for the smaller upstream catchments the optimal window was smaller than for the complete Upper Ourthe catchment, although there was not a negative effect of longer assimilation window noted ($W = 5$ vs. $W = 11$). Note that it was not the objective of the study to determine the optimal assimilation window for the AEnKF.

We investigated the effect of a partitioned update scheme recently suggested by *Xie and Zhang* (2013). We showed that for the Ourthe catchment reducing the number of model states of a grid-based HBV model using AEnKF can lead to better forecasts of the discharge. In terms of the root-mean-square-error, the largest improvements in the forecast accuracy were observed for the scenario, when the soil moisture was left out from the analysis (similar to the PDM updating scheme presented by *Moore*, 2007). This indicates that elimination of the strongly non-linear relation between the soil moisture storage (SM) and assimilated discharge observations can become beneficial for an improved forecast. Moreover, we showed that keeping the quick catchment response storage (upper zone; UZ) in the model analysis is important especially for longer lead times, when compared to the scenario in which only two routing storages were updated. A possible alternative to exclude the SM storage from the analysis, would be to investigate the use of other algorithms, for example the Maximum Likelihood Ensemble Filter (MLEF) (*Zupanski*, 2005) that is more suited for use with highly non-linear observation operators.

Chapter 6

On noise specification in data assimilation schemes for improved flood forecasting using distributed hydrological models

Abstract

We investigate the effect of noise specification on the quality of hydrological forecasts via an advanced data assimilation (DA) procedure using a distributed hydrological model driven by numerical weather predictions. The sequential DA procedure is based on (1) a multivariate rainfall ensemble generator, which provides spatial and temporal correlation error structures of input forcing and (2) lagged particle filtering to update past and current state variables simultaneously in a lag-time window to consider the response times of internal hydrologic processes. The strength of the proposed procedure is that it requires less subjectivity to implement DA compared to conventional methods using consistent and objectively-induced error models. The procedure is evaluated for streamflow forecasting of three flood events in two Japanese medium-sized catchments (Maruyama and Katsura). The rainfall ensembles are derived from ground based rain gauge observations for the analysis step and numerical weather predictions for the forecast step. The ensemble simulation performs multi-site updating using information from the streamflow gauging network and considers the artificial effect of reservoir release. Sensitivity analysis is performed to assess the impacts of uncertainties coming from DA such as random walk state noise and different DA methods with/without objectively-induced rainfall uncertainty conditions. The results show that multivariate rainfall ensembles provide sound input perturbations and model states updated by lagged particle filtering produce improved streamflow forecasts in conjunction with fine-resolution numerical weather predictions.

This chapter is largely based on the manuscript: Noh, S. J., O. Rakovec, A. H. Weerts, Y. Tachikawa (2014), On noise specification in data assimilation schemes for improved flood forecasting using distributed hydrological models, submitted to *J. Hydrology* (in review).

6.1 Introduction

Fluvial/pluvial flooding is a natural hazard, which occurs after extensive rainfall or snowmelt events. Accurate and reliable flood early warning (systems) enable to mitigate the number of casualties and economic damage, which is causing problems especially in densely populated areas (e.g. *Sene*, 2008) . While important advances have been achieved in flood forecasting (*Biondi and De Luca*, 2013; *Cloke and Pappenberger*, 2009; *Collier*, 2007; *Hapuarachchi et al.*, 2011; *Kitanidis and Bras*, 1980, among others and references therein), due to various uncertainties that originate from simulation models, observations, and forcing data, they are still insufficient to obtain accurate prediction results with the required lead times. The quality of hydrological forecasting systems is dependent on several factors: (1) the quality of the hydrological model (in its structure and parameters estimates) and its suitability for a given catchment, (2) initial conditions of the model states at the start of the forecast, (3) external forcing during the forecasted period (weather forecasts), and (4) human control (e.g. reservoir operations, irrigation) (*Weerts et al.*, 2013). To increase the certainty of the hydrological forecast, i.e. the magnitude and timing of a flooding event, all of these sources of uncertainty need to be considered and propagated through the hydrological modelling chain embedded in a flood forecasting system. This can be done using state updating or more broadly speaking data assimilation (DA), a technique to reduce the uncertainty by combining the uncertain hydrological simulations of model states and fluxes with the actual measurements/observations of river stages/streamflow, soil water content or ground water levels.

Numerous sophisticated DA algorithms have been proposed, from ruled-based, direct-insertion methods to advanced smoothing and sequential techniques, as well as variants of these techniques (*Liu et al.*, 2012, and references therein)), which have allowed recent progress in hydrologic DA for streamflow prediction (*Aubert et al.*, 2003; *Clark et al.*, 2008b; *DeChant and Moradkhani*, 2012; *Han et al.*, 2012; *Komma et al.*, 2008; *Lee et al.*, 2011; *McMillan et al.*, 2013; *Mendoza et al.*, 2012; *Moradkhani et al.*, 2005a; *Noh et al.*, 2011a, 2013; *Rakovec et al.*, 2012b; *Salamon and Feyen*, 2009, 2010; *Seo et al.*, 2009; *Shiiba et al.*, 2000; *Smith et al.*, 2012; *Weerts and El Serafy*, 2006, among others). As discussed by *Seo et al.* (2009), the correction of model inputs, states, initial conditions, and parameters is often conducted in a rather empirical and subjective way, which may reduce credibility and transparency to operational forecasts. For example, input uncertainty such as rainfall and weather is generally assumed to be governed by different distributions: the lognormal distribution (*DeChant and Moradkhani*, 2012), the normal distribution (*Weerts and El Serafy*, 2006), or a mixture of uniform distributions (*Clark et al.*, 2008b) with different error variances using multiplicative, additive, or mixture noises. State variables such as soil moisture are usually perturbed by random normal noises with or without considering temporal correlations using hyper-parameters (e.g., *Clark et al.*, 2008b; *Noh et al.*, 2013). Observation uncertainty of streamflow is often assumed to follow a normal or lognormal distribution using different formulations (e.g. *McMillan et al.*, 2013; *Moradkhani et al.*, 2005a; *Weerts and El Serafy*, 2006). However, the impact of different noise specifications on performance has not been treated especially in operational setup. Therefore, advances in DA not only imply improvement of accuracy but also embrace establishment of an "objective DA framework" to disentangle uncertainties.

It is also worth noting that a variety of noise specifications does not imply subjectivity. An objective DA framework cannot be composed of a single, concrete procedure. We use the term "objec-

tivity" for DA procedures rendering ensembles from given information systematically and avoiding subjective random walk noises and hyper-parameters. Recently, a multivariate rainfall generator was proposed by *Rakovec et al.* (2012a) in an attempt to generate objectively-induced model forcings considering spatial and temporal correlations based on rain gauge observations and simulations having the potential to reduce subjectivity in DA. Additionally, DA algorithms can be evaluated from the viewpoint of objectivity. Objective DA algorithms may be summarized as less vulnerable to obscure noise setting and with minimized subjective decisions on DA. The lagged filtering, suggested for particle filtering (PF) by *Noh et al.* (2011a) and extended to ensemble Kalman filtering (EnKF) (*McMillan et al.*, 2013; *Noh et al.*, 2013), are DA algorithms which consider a time lag in the routing process as well as different time scales of the hydrologic processes with less subjective DA settings such as updating intervals. In this chapter, our interest is focused on particle filtering (PF) based on a Bayesian filtering theory which does not require any assumption such as linearity or a Gaussian distribution. Unlike Kalman filter-based methods, PF performs updating on particle weights instead of state variables (*Liu and Gupta*, 2007), which reduces numerical instability, especially in physically based or process-based models.

Meanwhile, due to rapid development of computing power, applications of numerical weather prediction (NWP) and distributed hydrologic modelling (DHM) have provided alternative directions in flood forecasting. Although the quality of NWP is still considered to be limited by uncertainties regarding localisation, timing, and intensity of events (*Habets et al.*, 2004)), advances in flood forecasting and reservoir operation using NWP have been reported (e.g., *Jasper et al.*, 2002; *Pappenberger et al.*, 2008b; *Saavedra Valeriano et al.*, 2010; *Schellekens et al.*, 2011; *Smiatek et al.*, 2012; *Verbunt et al.*, 2006; *Wang et al.*, 2012). In addition, despite its inherent uncertainty (*Beven*, 1993), applications of DHM have been increasing with increasing accessibility of spatial and temporal information on the earth system from various sources (e.g. remote sensing). However, there have been rarely attempts to use NWP and DHM in a DA framework to quantify uncertainties in an operational flood forecasting setting.

The main objective of this chapter is to show the effect of noise specification in a DA scheme on the quality of hydrological forecasts. Two ways of specifying noise in a DA scheme are compared: (1) random walk noise with subjective error assumptions (e.g., *Noh et al.*, 2011a) and (2) objectively induced input forcing using a multivariate rainfall generator (*Rakovec et al.*, 2012a,b). The effect of the noise specification is investigated by carrying out several DA experiments in two Japanese medium-sized catchments (Maruyama and Katsura) using a DHM and two PF methods. Finally, the operational setup of the framework is tested by combining both approaches with NWP for three flood events.

6.2 Study catchments

This study is carried out within two Japanese catchments (Fig. 6.1), the Maruyama River catchment (909 km^2) and the Katsura River catchment (887 km^2) (Fig. 6.2a,b), in which the rainfall mostly occurs in the summer season from May to September. Both catchments are characterized as fast-responding with a time of concentration of less than 10 hours. There are 10 rain gauges and 4 streamflow gauges in the Maruyama catchment and 13 rain gauges and 1 streamflow gauge in the Katsura catchment. All rain and streamflow gauges are operated by the Ministry of Land, Infrastruc-

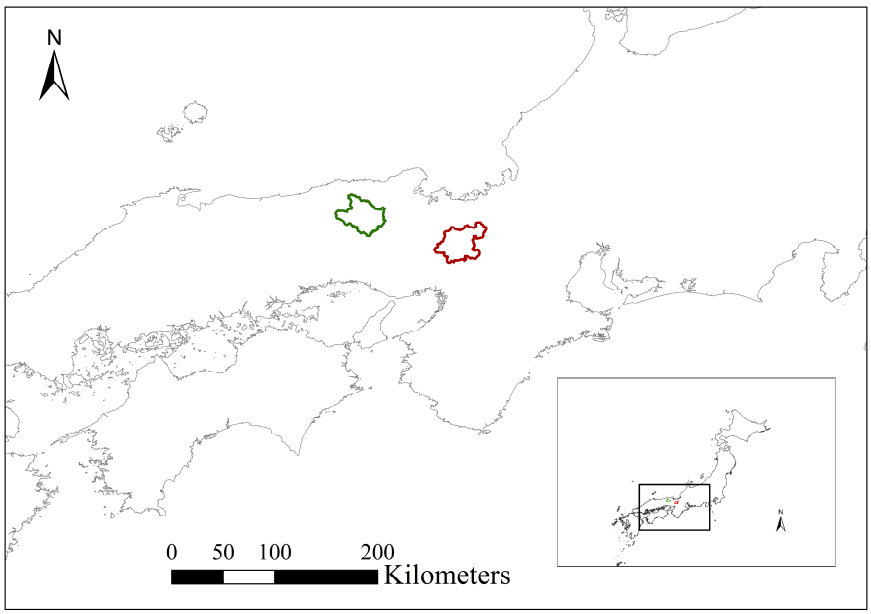


Figure 6.1: Location of two catchments: the Maruyama River catchment (greed line) and the Katsura River catchment (red line).

(a) (b)

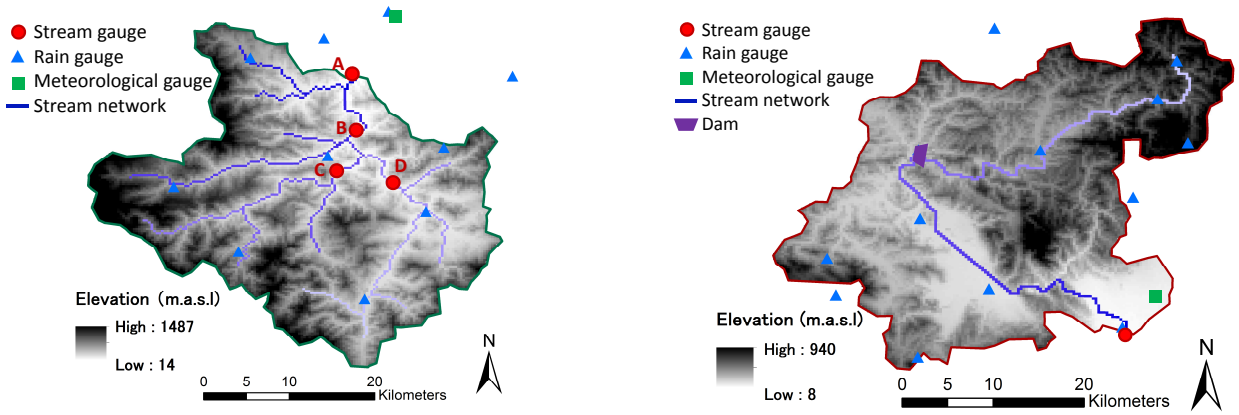


Figure 6.2: Topographic map of (a) the Maruyama River catchment and (b) the Katsura River catchment.

ture, Transport and Tourism in Japan (MLIT) (http://www1.river.go.jp/) and are available at hourly time step. The hourly time series of air temperature, relative humidity, wind speed, and duration of sunlight are obtained from two meteorological stations operated by Japan Meteorological Agency (http://www.jma.go.jp). The dominant land use of the Maruyama River catchment consists of forest (37%), savannas (10%), and crop land and natural vegetation (53%), while those of the Katsura River catchment are forest (76.7%), agricultural areas (9.3%), residential areas (7.5%) and others (6.5%). The Hiyoshi dam is located upstream of the Katsura River catchment, and the dam release from the reservoir is simulated according to its operating rule.

6.3 Hydrologic model

In this study, we applied the WEP model (Water and Energy transfer Processes), aiming for simulating spatially variable water and energy processes in catchments with complex land covers (*Jia et al.*, 2001, 2009; *Noh et al.*, 2011a). Runoff routing on hillslopes and in river beds is calculated by a one-dimensional kinematic wave method. The spatial resolution of the simulation grid in this study is 250 m $\times$ 250 m for all of the distributed input data including precipitation, topography, soil, land use, etc., in both catchments. The soil distribution was obtained from the website of the Food and Agriculture Organization of the United Nations (http://www.fao.org/nr/land/soils/en/). The basic parameters including soils, vegetation, aquifers, and roughness coefficients were selected from the ranges of previous studies and adjusted by global multipliers, which were calibrated for independent simulation periods by PEST (*Doherty*, 2003, 2009).

6.4 Particle filtering

Here, we briefly explain the basic concept of PF and two PF methods, namely sequential importance resampling (SIR) and lagged particle filtering (LPF), which will be implemented with different noise specification methods defined in Sect. 6.5.

6.4.1 Sequential importance resampling PF

In Bayesian filtering theory, it is not possible to analytically construct the posterior probabilistic density function (PDF) of the current state x_k given the measurement $y_{1:k} = \{y_j, j = 1, \ldots, k\}$, if the system and measurement models are nonlinear and non-Gaussian. In this case, suboptimal solutions can be found using simulation-based ensemble methods such as PF.

PF is a set of simulation-based methods that provide a flexible approach to computing posterior distributions without using any assumption about the nature of the distributions. The critical concept of PF is based on point mass ("particle") representations of the probability densities with associated weights (*Arulampalam et al.*, 2002):

$$p(x_k|y_{1:k}) \approx \sum_{i=1}^{n} w_k^i \delta(x_k - x_k^i), \tag{6.1}$$

where x_k and w_k^i denote the posterior state and its weight of the ith particle out of n particles (sample

size), and $\delta(.)$ denotes the Dirac delta function. According to the sequential importance sampling (SIS) algorithm, the weights are updated as follows:

$$w_k^i \propto w_{k-1}^i p(y_k|x_k^i), \tag{6.2}$$

where w_{k-1}^i is the weight of the previous time step and $p(y_k|x_k^i)$ represents the likelihood of each ensemble member. Although SIS is the basic algorithm of most PF methods, a critical flaw has been reported in this method known as the degeneracy phenomenon: most of weights are dominated by a very small number of ensemble members after a few iterations, while the remaining ensemble members have negligible weights. An alternative is to introduce a resampling step. The idea of resampling is that particles with very low weights are abandoned, while multiple copies of ensemble members are kept with uniform weights which approximate the posterior PDF, $p(x_k|y_{1:k})$ (*van Leeuwen*, 2009). Sequential importance resampling (SIR) is a PF method performing the resampling step at each analysis step, which has been widely applied in hydrologic DA studies (*Moradkhani et al.*, 2005a; *Noh et al.*, 2011b; *Salamon and Feyen*, 2010; *Weerts and El Serafy*, 2006, among others).

6.4.2 Lagged particle filtering

As mentioned by *McMillan et al.* (2013), the natural time lag between catchment state and streamflow (i.e. the time of concentration) is a significant challenge for DA in hydrology. An update of DHM states at the same time as the flow observations may not be physically realistic and can lead to under- or over-shoot in flow correction at later time steps (*Mendoza et al.*, 2012). Lagged particle filtering (LPF; *Noh et al.*, 2011a) is a PF method that considers a time lag in the routing process as well as different time scales of the hydrologic processes. The main idea of LPF is to perform updating on the past variables as well as the current ones within a fixed lag time window. LPF is less sensitive to assumptions regarding process noise (*Noh et al.*, 2011a)) and lag-time window size (*Noh et al.*, 2013), which may contribute to reduction of subjectivity in DA.

LPF may be implemented with the regularization step based on a Markov chain Monte Carlo (MCMC) method (*Gilks and Berzuini*, 2001) to preserve the diversity of the ensemble system in a long lag-time window. The main feature of the regularization step lies in a continuous approximation of the particle density, which creates a new particle system systematically with different particle locations in the regularization step (*Ristic et al.*, 2004). The MCMC move step is combined to asymptotically approximate the posterior PDF based on the Metropolis-Hastings algorithm (*Robert*, 2010). The detailed description and application of LPF can be found in *Noh et al.* (2011a, 2013).

6.5 Noise models

In this study, we consider model uncertainty by direct state perturbation of the soil moisture content with random walk noise, which provides boundary conditions of surface runoff and represents interaction from evapotranspiration and infiltration (Sect. 6.5.1) and input uncertainty by rainfall ensembles (Sect. 6.5.2).

6.5.1 Direct state perturbation with random walk noise

A single multiplier γ_s is utilized to perturb the soil moisture content θ in the entire grid:

$$\widehat{\theta} = \gamma_s \theta \tag{6.3}$$

$$\gamma_s = \frac{\widehat{S_k}}{S_{k-1}} \tag{6.4}$$

$$\widehat{S_k} = S_{k-1} + w_{soil_k} \tag{6.5}$$

where S_{k-1} represents the total water depth aggregated for the three soil layers of the entire grid within the catchment at the previous time step and w_{soil_k} represents a Gaussian noise $N(0, \sigma^2_{soil_k})$ with a heteroscedastic standard deviation:

$$\sigma_{soil_k} = \alpha_{soil} S_{k-1} + \beta_{soil} \tag{6.6}$$

where α_{soil} and β_{soil} are adaptable parameters which control the magnitude of random noise. As proper ranges of the parameters are usually obtained from sensitivity analysis, optimum values may be different from event to event, increasing additional uncertainty and subjectivity in the DA procedure. Compared to previous works, we use objectively generated rainfall ensembles to consider the uncertainty of the input forcing (Sect. 6.5.2). Therefore, the required magnitude of random noise in DA can be minimized. Impacts of objective input noise on prediction performance will be discussed by sensitivity analysis in Sect. 6.9.

The measurement error for the discharge is assumed to follow a Gaussian distribution $N(0, \sigma^2_{obs_k})$. The standard deviation of the measurement error required for the estimation of the likelihood of ensembles is defined as follows:

$$\sigma_{obs_k} = \alpha_{obs} y_k + \beta_{obs} \tag{6.7}$$

In Eq. (6.7), α_{obs} is a proportionality coefficient and β_{obs} is a constant coefficient representing uncertainty during periods of low flow, which are set at 0.1 and 5 $m^3\ s^{-1}$, respectively.

6.5.2 Rainfall ensemble generator

Spatial and temporal variation of the rainfall fields represents an important ingredient of the hydrologic hindcasting, which is used in combination with assimilation of the hydrologic observations to estimate the initial model states at the beginning of the model forecast. An ensemble of rainfall realizations is a suitable way to quantify the model input uncertainty associated with estimates of the spatial rainfall variability. This can be done using conditional simulation, a geostatistical method, which generates simulations in a way that reasonably matches the sample statistics of the rainfall fields and exactly matches the conditioning data (*Goova erts*, 1997). Conditional simulation of the ground-based rain gauge observations is a popular technique for different meteorological and hydrological applications (*Clark and Slater*, 2006; *Rakovec et al.*, 2012a; *Renard et al.*, 2011; *Vischel et al.*, 2009). In this chapter, we do not consider any other source of rainfall measurements, such as the weather radar, which is a topic on its own.

In the current study, we obtain the rainfall ensemble at hourly time step by employing multivariate

conditional simulation (*Rakovec et al.*, 2012a), which provides a proper spatial error structure given the rain gauge observations and additionally ensures that the temporal correlation structure is maintained for each ensemble realization over time. The temporal coherence of the rainfall ensembles is then propagated into the model states, which prevents averaging out possible extreme state values being of major interest for flood forecasting.

The theory of conditional simulation is explained in detail for example by *Goova erts* (1997), therefore it will not be repeated here. However, we provide a brief summary of the setup for the multivariate conditional simulation after (*Rakovec et al.*, 2012a) as follows.

First, one time-independent univariate ensemble realization i at time step k is created in the following fashion: (1) Rainfall observations are transformed into normal space; (2) A random path is defined throughout the simulation domain such that each grid node is visited once; (3) A random number is drawn from a Gaussian distribution at each grid node, given the kriging prediction, variance of the rain gauge observation and the related semi-variogram. This simulation is then included in the conditional dataset for subsequently simulated grid nodes. (4) After all grid nodes are simulated, back-transformation of the normally distributed values to the original rainfall distribution is carried out.

Secondly, if the number of consecutive rainy hours is at least two, a time-dependent multivariate conditional ensemble realization i is simulated at time $k+1$ conditioned on: (1) rain gauge observations at time $k+1$ and (2) previously simulated realizations i of the all grid nodes at times $(k,\ldots,k-M)$, where M is the simulation memory, which is set up to 1 in this case study.

To overcome problems with potential numerical instability, rainy periods were defined as a cluster of consecutive rainy hours, where each rainy hour has a minimal intensity condition. More specifically, the maximum of all rain gauge observations should be at least 3 mm and there should be more than three rain gauges with non-zero rainfall. Additionally, a small value of 0.2 mm is added to rain gauges with zero rainfall, which is finally subtracted after the simulation.

6.6 Numerical weather prediction data

The meso-scale model (MSM), which is based on the non-hydrostatic JMA model (*Saito et al.*, 2006), provides a 15 hour forecast around Japan every 3 hours: 00, 03, 06, 09, 12, 15, 18, 21UTC. MSM is one of six NWP models of the Japan Meteorological Agency (JMA) to support very short-range forecasting of severe weather disturbances. The model resolution is currently about 5 km (0.0625 and 0.05 degree in longitude and latitude) for the land surface, covering an area of 120° – 150°E, 22.4° – 47.6°N in the horizontal with 481 × 505 points. Four-dimensional variational (4D-VAR) DA method is used for analysis of the atmospheric state in the MSM.

6.7 Data assimilation experiment

The DA procedure proposed in this study is described in Fig. 6.3. Rainfall ensembles are generated as input forcing of the analysis step by the multivariate rainfall generator using rain gauge observations. Additionally, the model state (i.e. the soil moisture content) is perturbed by multiplicative noises. The model response is analyzed within a lag-time window and updated according to streamflow mea-

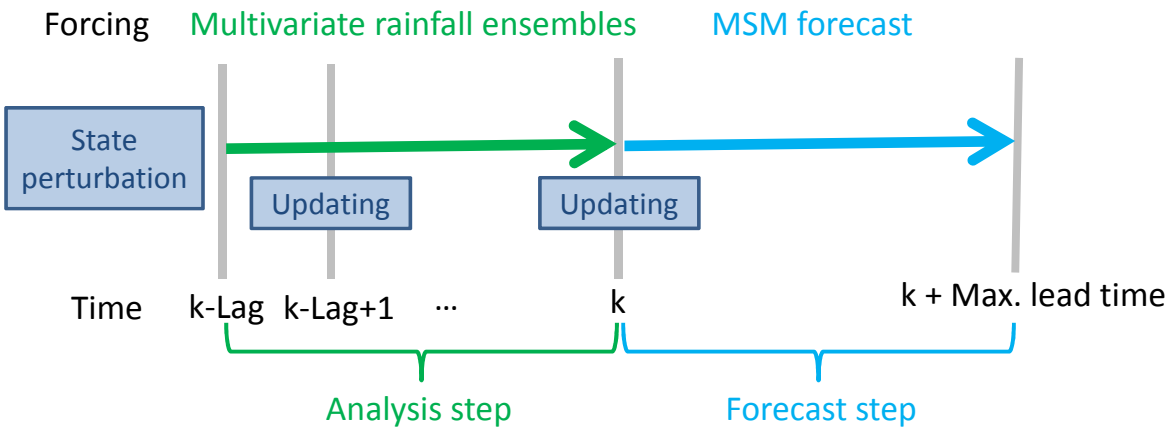


Figure 6.3: Schematic diagram of data assimilation procedure.

surements at the current time. Updated model ensembles are then forced with NWP (i.e. MSM) for given lead times. DA and forecast steps are performed at an hourly time step, which corresponds to the observation interval. The rainfall forecasts are available at an interval of 3 hours. The DA experiments are performed in three phases. First, the rainfall generator (*Rakovec et al.*, 2012a) is tuned and analyzed for the case study area (Sect. 6.2). The second experiment investigates the effect of the different noise generation methods by combining two PF methods and two noise specification methods: the state noise perturbation (Sect. 6.5.1) and the rainfall generator (Sect. 6.5.2). The Thiessen rainfall interpolation using rain gauge observations is used as a deterministic model input in the case of state noise perturbation. Forecasting results with varying lead times are evaluated using three metrics including the Nash-Sutcliffe efficiency (NSE), the root mean square error (RMSE), and the mean bias (MB) defined as:

$$NSE = 1 - \frac{\sum_{k=1}^{T}(y_k - y_{sim_k})^2}{\sum_{k=1}^{T}(y_k - \bar{y})^2} \tag{6.8}$$

$$RMSE = \sqrt{\frac{\sum_{k=1}^{T}(y_k - y_{sim_k})^2}{T}} \tag{6.9}$$

$$MB = \frac{\sum_{k=1}^{T}(y_k - y_{sim_k})}{T} \tag{6.10}$$

where y_k is the streamflow observation at time k, $\bar{y}$ is the mean of the observations, y_{sim_k} is the mean of the forecasted streamflows at the measurement site, and T is the total number of time steps. Metrics are selected to evaluate accuracy of forecasts for short lead times. Finally, in the third experiment the DA framework is applied in an operational setting (i.e. with fixed state noise) and its performance is evaluated for 3 flood events described in Table 6.1. Event 1 occurred during the monsoon season of July 2006. Events 2 and 3 were driven by consecutive typhoons in September 2011. While Event 1 moved from North to South, Event 2 and 3 followed typical directions of typhoons moving from South-West to North-East. Despite different meteorological conditions, all three events covered a large scale, which resulted in a significant amount of precipitation across both catchments.

Table 6.1: Details of selected flood events

		Katsura River catchment		Maruyama River catchment	
Flood Event	Date	Cumulated rainfall (mm)	Peak flow outlet ($m^3\ s^{-1}$)	Cumulated rainfall (mm)	Peak flow outlet ($m^3\ s^{-1}$)
Event 1	1-31 Jul. 2006	403	910	387	1530
Event 2	1-10 Sep. 2011	179	677	306	2124
Event 3	11-30 Sep. 2011	259	1151	320	1574

6.8 Rainfall ensembles: results

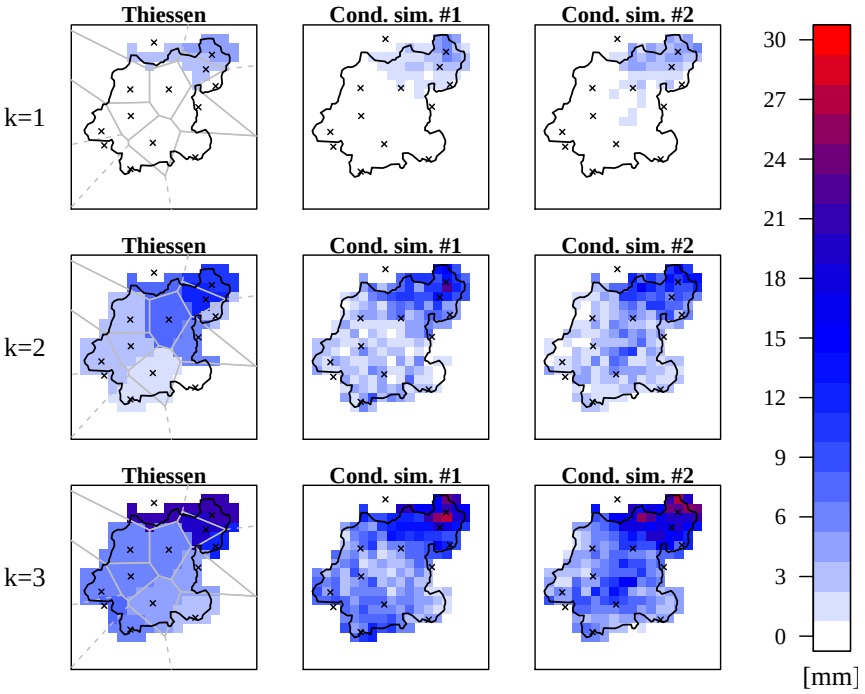


Figure 6.4: Example of spatial rainfall fields for three consecutive time steps obtained by the Thiessen interpolation (left) and two realizations of conditional simulation (center and right) for the Katsura catchment on 17 July 2006.

The ensemble of conditionally simulated rainfall fields, which is used to force the hydrological model during the analysis step (see Fig. 6.3), is generated in a way as was described in Sect. 6.5.2. The size of the simulation domain is defined to be 50 km $\times$ 50 km with a grid resolution of 2.5 km $\times$ 2.5 km for both catchments. An example of such a simulation for two ensemble realizations over the Katsura catchment is shown in Fig. 6.4 for three consecutive time steps. The Thiessen rainfall interpolation, which is used as an alternative model input, is also included in the figure for a qualitative comparison. Figure 6.4 indicates that the magnitude of the simulated and interpolated rainfall fields is about the same, but the spatial variation is much larger in the simulated ones. This suggests that the conditional simulation can take into account possible extreme values, which might be omitted by the interpola-

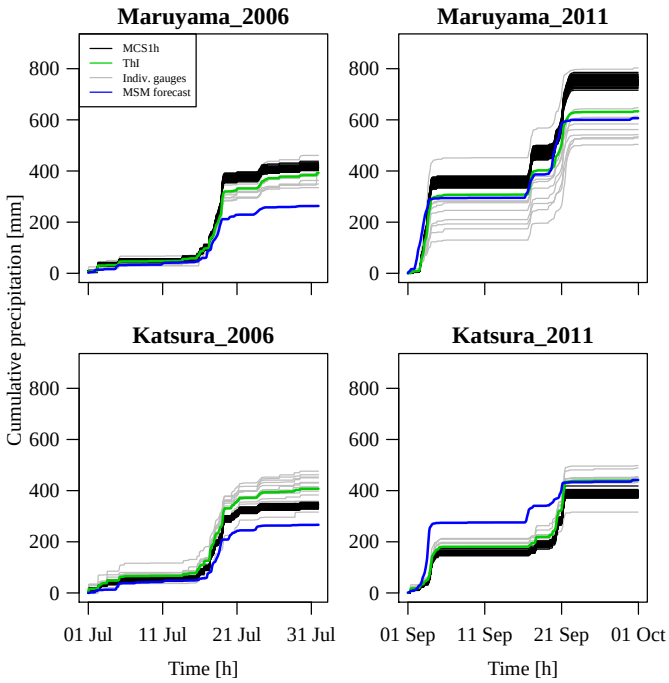


Figure 6.5: Cumulative catchment averaged rainfall obtained by 64 multivariate conditional simulations (MCS1h), Thiessen interpolation (ThI) and the MSM forecast (MSM). The cumulative rainfall of individual rain gauges is shown in grey.

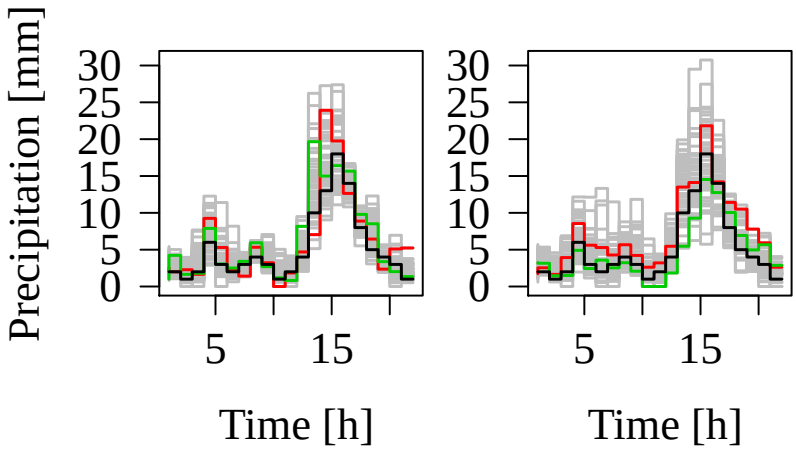


Figure 6.6: Rainfall ensemble (64 realizations in grey, from which two realizations are shown in red and green) at one grid point location within the Maruyama catchment for a 22-hour event on 18 July 2006. The black like represents the rain gauge observation within the same grid point. The left panel shows the simulation using time independent simulation memory of 0h and the right panel shows the simulations using the time dependent simulation memory of 1h.

tion. Furthermore, the cumulative catchment averaged rainfall is shown in Fig. 6.5 for 64 simulated rainfall realizations as well as the Thiessen interpolation. For guidance, we provide in the same figure the rainfall measured at individual rain gauges and the mean forecasted rainfall using the meso-scale

numerical weather prediction model (Sect. 6.6). The rain gauge observations show a strong spatial variability, which is mostly driven by the topography. Additionally, we can observe that the simulated rainfall fields are within the bounds of individual rain gauges, which indicate that the simulations are not biased. However, when we consider the rainfall forecast, we can observe some rainfall underestimation in 2006 for both catchments and rainfall overestimation in the Katsura catchment in 2011.

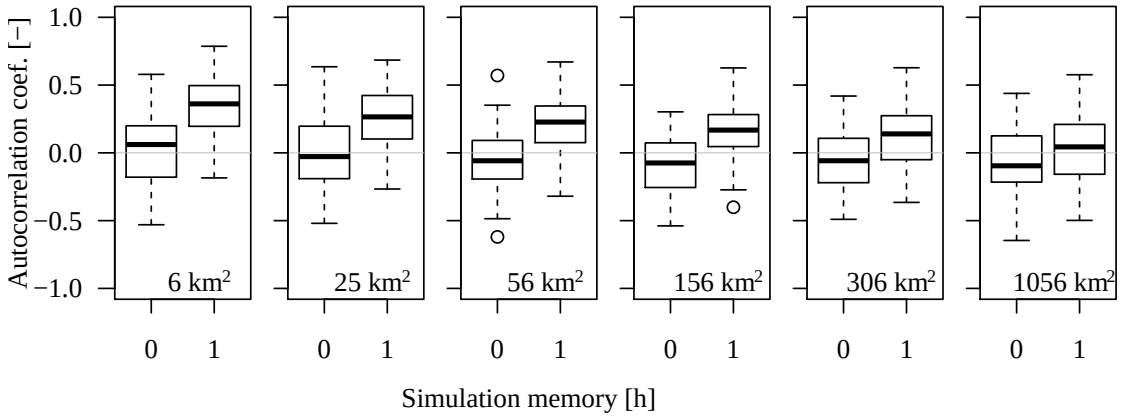


Figure 6.7: Boxplots of the first-order autocorrelation coefficients (64 ensemble realizations) for the time series of the difference between the catchment's averaged areal precipitation and the ensemble mean for the time independent simulation memory of 0 h and time dependent simulation memory of 1 h. Catchment averaged areal precipitation is derived for increasing size of sub-catchments within the Maruyama catchment (rainfall event on 18 July 2006 with duration of 22 hours).

Further analyses of the difference between (1) univariate time-independent (unconditioned on previously simulated realizations) and (2) multivariate time-dependent simulations (conditioned on 1h of previously simulated fields) are shown Figs. 6.6 and 6.7. Let us first consider a single grid cell within the Katsura catchment for which we plot the time series of the 64 ensemble realizations for both types of simulation memory (Fig. 6.6). We can observe that the overall spread is slightly higher for the multivariate simulation ($M = 1$ h) than for the univariate simulation ($M = 0$ h). Additionally, the temporal consistency of the realization is evident for the multivariate approach, while it is less present for the univariate approach. This can be seen when we have a closer look at the two depicted realizations and the frequency at which they intersect the rain gauge observation. More quantitative evidence about the temporal coherence of the multivariate simulations over the univariate simulation is provided in Fig. 6.7, in which we show the first order autocorrelation coefficients for different simulation memories and increasing catchment sizes. The boxplots are composed of 64 realizations for a 22-hour rainfall event, which occurred in the Maruyama catchment. It is clear that for all the univariate simulations there is no temporal correlation, because the simulations at each hour are carried out independently from the previous hour. Therefore, the univariate conditional simulations tend to average out extreme rainfall accumulations over time. On the other hand, the temporal correlation for the multivariate simulations is visible up to the catchment size of >300 km^2, which corresponds well with the predefined spatial properties of the simulations.

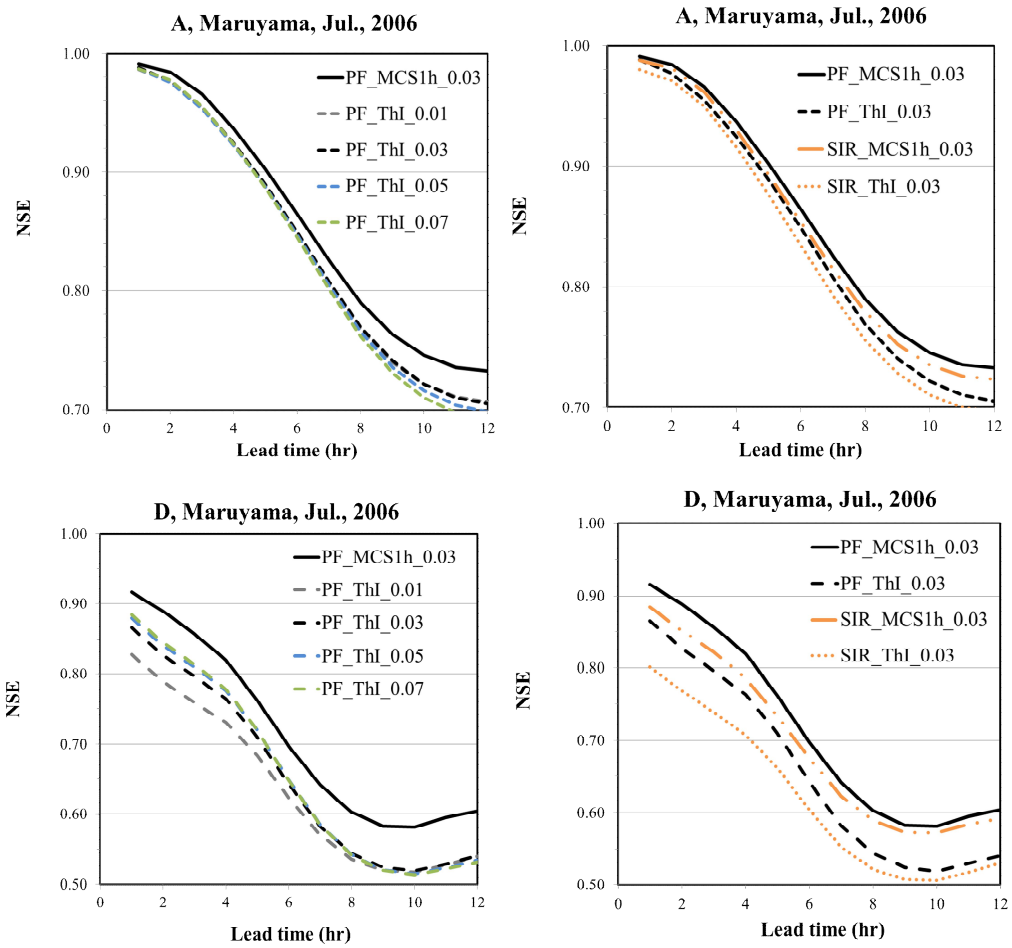


Figure 6.8: Comparison of NSE with different error specification and filtering methods at the outlet (A) and interior points (B, C, and D) (Maruyama, Event1).

6.9 Impact of noise generation methods on forecast quality

Although DA is usually applied to reduce uncertainties in hydrologic predictions, there are other sorts of uncertainties coming from DA such as error models of input forcing and state variables. As discussed by *Weerts and El Serafy* (2006), obtaining good estimates of the error (or noise) models is the most difficult part when using DA. The error models and the magnitude of the noise are usually selected by heuristics and are prone to be affected by subjectivity, which may reduce transparency and credibility in hydrologic predictions.

In this section, sensitivity analysis is carried out to assess impacts of different error models and combinations of DA methods in the forecasting mode. The DA methods with two different rainfall inputs in the analysis step are implemented for Event 1 in both catchments while varying the magnitude of state noise α_{soil} from 0.01 to 0.07, by increment of 0.01.

The performance evaluated by NSE for Event 1 is illustrated in Figs. 6.8 and 6.9. For clarity, sensitivity of state noise is compared using selected cases: one LPF+MCS1h case with a state noise

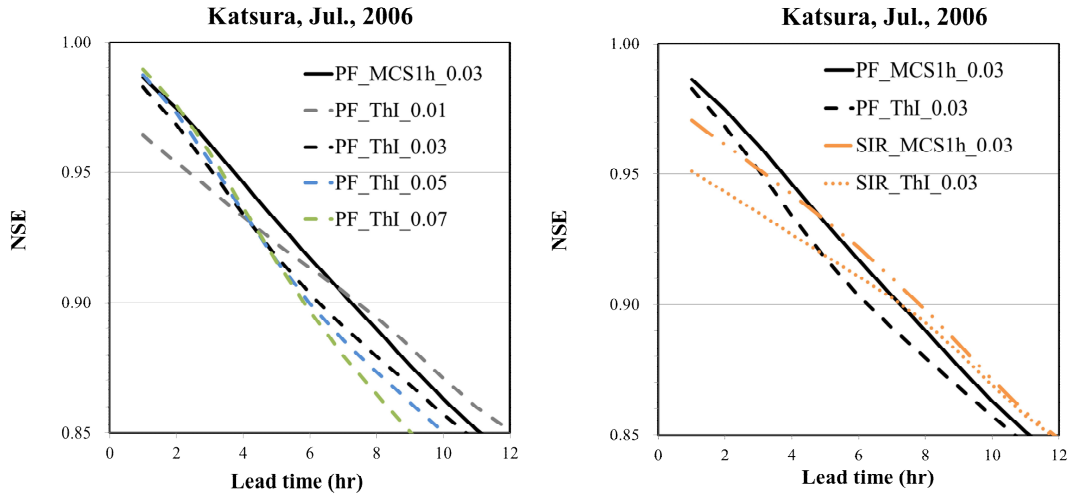
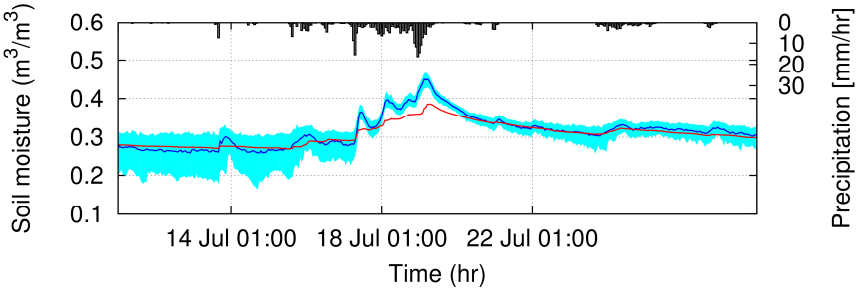


Figure 6.9: Comparison of NSE with different error specification and filtering methods at the outlet (Katsura, Event1).

(a) Maruyama, Jul., 2006, PF+MCS1h with state noise of 0.03



(b) Maruyama, Jul., 2006, PF+ThI with state noise of 0.06

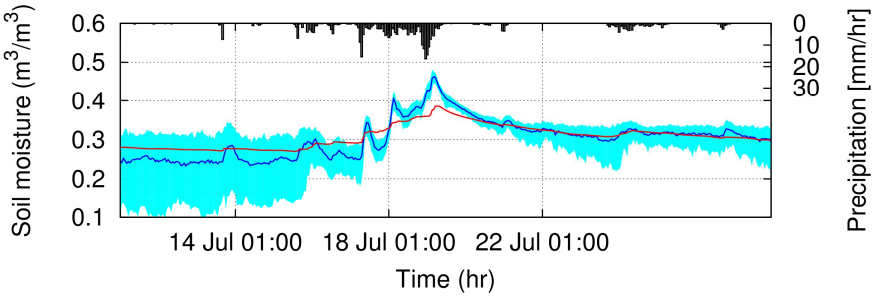


Figure 6.10: Traces of the total soil water storage (Maruyama, Event 1). Blue lines and area represent the median and the 95% confidence interval of the total soil water storage ensembles. Red lines represent the open loop simulation.

α_{soil} of 0.03 and four LPF+ThI cases with a state noise α_{soil} ranging from 0.01 to 0.07 with a 0.02 increment. Here, MCS1h and ThI represent multivariate conditional simulations with 1 hour simulation memory and Thiessen interpolation of rainfall input, respectively. In the case of LPF+MCS1h, small noise magnitudes such as 0.02 or 0.03 provide the best performance consistently in both catchments, with negligible difference (not shown here). However, LPF+ThI shows different optimal values of the state noise α_{soil}, ranging from 0.03 to 0.07, with high sensitivity for both events and lead times.

In the case of the Maruyama River catchment (Fig.. 6.8), LPF+MCS1h outperforms all the cases of LPF+ThI. LPF+ThI shows a higher sensitivity in the head water area than at the outlet. Effects of MCS and LPF are found consistent at both observing stations, while Thiessen interpolation and SIR show a larger degeneracy of performance. For Event 1 in the Katsura River catchment (Fig. 6.9), a high state noise (e.g. 0.07) is required by LPF+ThI to get an equivalent performance for short lead times with PF+MCS1h, while deterioration of the forecast performance can be observed for longer lead times. Rainfall ensembles using MCS1h show improvement compared to Thiessen interpolation. Statistics of different metrics in this experiment are shown in Table 6.2 and confirm the results described above and presented in Figure. 6.8 and 6.9. It is worth to note that the conventional filters such as SIR require appropriate updating time intervals, which vary according to the location and lead to an additional source of uncertainty in DA. A description of impacts of the updating interval in DA can be found in *Rakovec et al.* (2012b).

To illustrate impacts of random noise on state variables, the soil moisture averaged across the entire grid in the updating stage is shown in Fig. 6.10. Despite a similar performance at the outlet using LPF+MCS1h and LPF+ThI with different state noises, patterns and uncertainty spreads of the soil moisture are different. LPF+ThI with a 0.06 state noise shows abrupt changes and wide uncertainty spreads of the soil moisture, while LPF+MCS1h with a 0.03 state noise shows more gradual changes.

6.10 Evaluation of the DA framework in an operational setting

We carried out a flood forecasting experiment for three rainfall events with the distributed hydrologic model described in the previous section. The number of rainfall and model ensembles used for DA is 63 and 126, respectively. PF is applied with an 8-hour lag-time window according to previous sensitivity analyses (*Noh et al.*, 2013). The magnitude of random state noise is controlled by values of $\alpha_{soil} = 0.03$ and $\beta_{soil} = 20$ for all simulation cases. The ensemble of model states is updated every hour by observed streamflows. In the Maruyama River catchment, observed streamflows are assimilated using four streamflow gauges including one at the outlet and the three interior gauges, while one observing gauge at the outlet is used in the Katsura River catchment.

Figure 6.11 shows forecast results of Event 1 (typhoon events in July 2006) in the Katsura catchment with different lead times: 3, 6, 9, and 12 hours. Obvious improvement is found until a 9-hour-lead forecast (Fig. 6.11(a)-(c)) as compared to the open loop, whereas the gains of updating become smaller with delayed and attenuated flood peaks as the lead time increases. Effects of reservoir releases characterized by a sharp increase and decrease of hydrographs are represented properly for short lead times. No instability of forecasted streamflows coming from the short updating time interval of 1 hour is found because of the smoothing procedure within the lag-time window. Similar patterns are found for both catchments for other data periods, while predictions deteriorate more quickly in Event 2 and 3 due to overestimated rainfall for several forecast lead times (not shown).

Table 6.2: Statistics of flood forecasting of Event 1 in the Katsura River catchment. See Eqs. (6.8) – (6.10) and Table 6.1 for abbreviations.

Method	State Noise	Lead time (hour)	1			3			6			12		
			NSE	RMSE	MB	NSE	RMSE	MB	NSE	RMSE	MB	NSE	RMSE	MB
Open loop	0		0.85	60.1	-35.7	0.85	60.5	-35.9	0.83	63.7	-36.6	0.75	77.3	-40.2
LPF+MCS1h	0.01		0.97	25.9	-4.2	0.95	33.1	-5.4	0.92	43.1	-7.7	0.85	60	-14.1
	0.03		0.99	18	-3.4	0.96	30.4	-4.3	0.92	44.2	-6.5	0.84	60.9	-13.4
	0.05		0.99	15.6	-4.1	0.96	29.7	-5	0.91	45.6	-7.7	0.83	63.4	-15.2
	0.07		0.99	14.5	-4.6	0.96	29.3	-5.7	0.91	46.3	-8.7	0.82	65.6	-16.7
SIR+MCS1h	0.03		0.97	26.3	-5.5	0.95	33.8	-6.4	0.92	43	-8.3	0.85	59.5	-14
LPF+ThI	0.01		0.96	29.1	-3.3	0.94	36.6	-4.1	0.91	45.2	-5.8	0.85	59.1	-12
	0.03		0.98	20.2	-2.6	0.95	33.9	-3.3	0.9	47.8	-5.1	0.84	61.6	-11.9
	0.05		0.99	17.3	-3.5	0.95	32.8	-4.3	0.9	48.6	-6.6	0.83	63	-13.9
	0.07		0.99	15.8	-4.5	0.96	31.7	-5.7	0.9	49.4	-8.5	0.81	65.9	-16.7
SIR+ThI	0.03		0.95	34.1	-5.8	0.94	39.2	-6.2	0.91	45.8	-7.2	0.85	59.7	-12.3

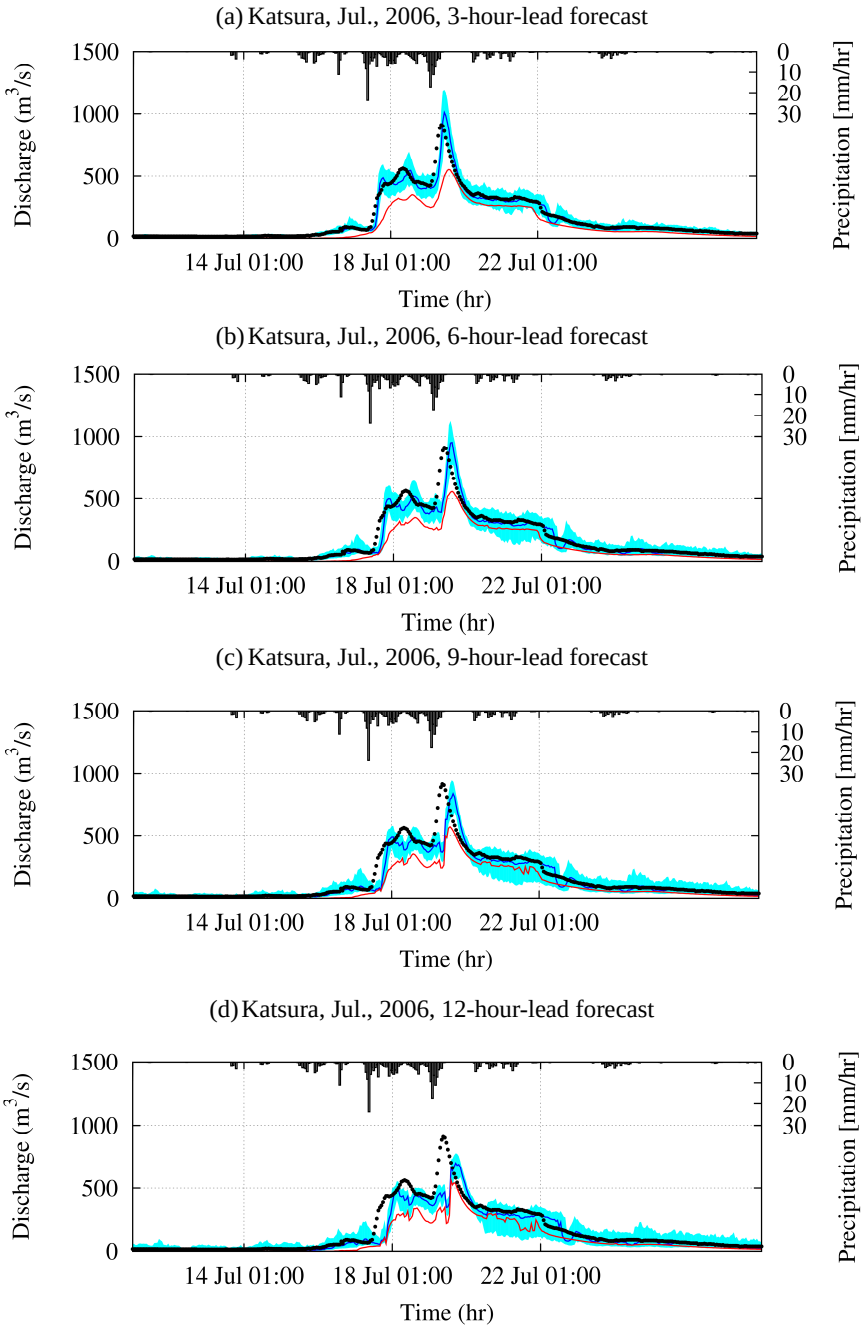


Figure 6.11: Forecasts of the July 2006 event in the Katsura River catchment: (a) 3-hour-lead forecasts (b) 6-hour-lead forecasts (c) 9-hour-lead forecasts and (6) 12-hour-lead time. Dots represent observed streamflow discharge at the catchment outlet. Blue lines and area represent the median and the 95% confidence interval of the streamflow ensemble by LPF. Red lines represent the open loop simulation.

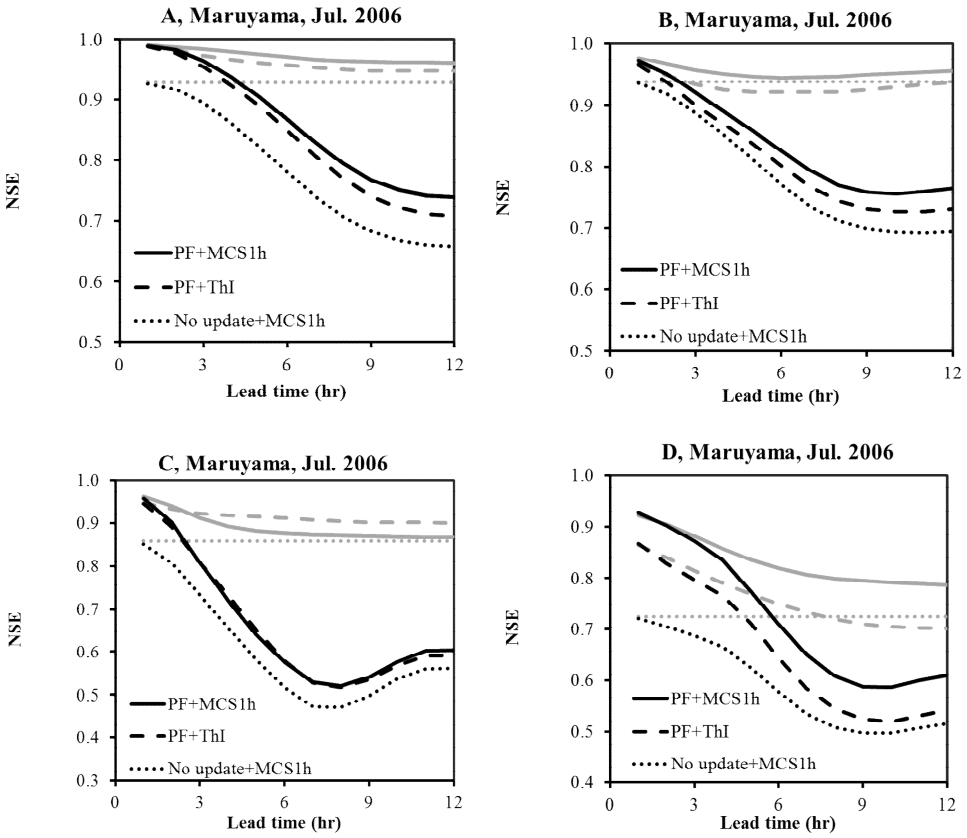


Figure 6.12: NSE model performance versus the lead time for the 4 streamflow gauges within the Maruayma catchment in July 2006. Hindcasting is shown in grey, forecasting in black.

Comparison of the NSE performance results is shown in Figs. 6.12–6.13. Black lines and dashed lines show the forecast results with PF using rainfall ensembles by multivariate conditional simulation with a 1 hour memory (LPF+MCS1h) and PF using rainfall by Thiessen interpolation (LPF+ThI), respectively. Two DA cases are compared with open loop simulations using rainfall by Thiessen interpolation (No update+ThI). Grey lines and dashed lines represent the hindcasting results using the rainfall fields of the analysis step for the forecast step with the same DA methods.

In both catchments, the two DA methods show improved performance over all forecast lead times and simulation periods in both hindcasting and forecasting modes (Figs. 6.12–6.13). For the results of Event 1 in the Maruyama River catchment (Fig. 6.12), PF+MCS1h provides improved performance compared to LPF+ThI at each observing station except for station C. Impacts of generated rainfall ensembles are high in the head water area (D, Maruyama), while the difference at the catchment outlet (A, Maruyama) is smaller. Overall, the difference induced by rainfall increases for longer lead times. Unlike other stations, station C shows an equivalent performance in the forecasting mode and a slight deterioration of LPF+MCS1h over LPF+ThI even in the hindcasting mode. Considerable missing data of the streamflow observations for station C may be a possible reason of this deterioration in the performance.

For the results of Event 2 and 3 in the Maruyama River catchment, there is no significant differ-

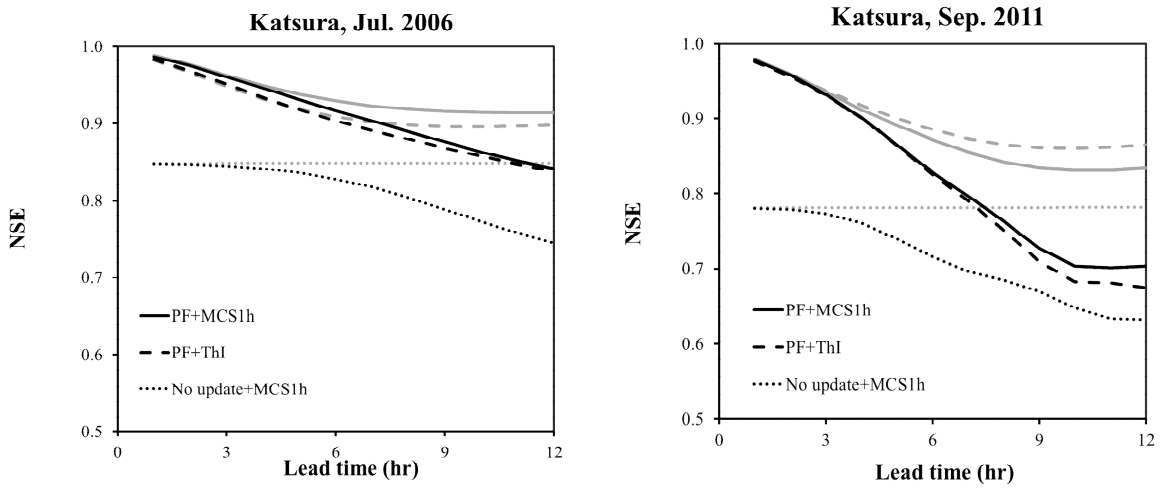


Figure 6.13: NSE model performance versus the lead time for at the catchment outlet of the Katsura catchment for July 2006 and September 2011. Hindcasting is shown in grey, forecasting in black.

ence between the two DA methods in the forecasting mode, while clear improvement is found in the hindcasting mode. In the forecasting mode, prediction accuracy decreases sharply with increasing lead times because MSM provides significantly overestimated rainfall forecasts for several lead times (not shown). Figure 6.13 illustrates results of both simulation periods in the Katsura River catchment. As expected, the two DA methods show improved predictions over open loops in both forecasting and hindcasting modes. LPF+MCS1h shows a better performance compared to LPF+ThI for both events in the forecasting mode. However, the difference is not significant, which is similar to the results at the outlet of the Maruyama catchment.

6.11 Conclusions

This study investigated the effect of noise specification on the quality of hydrological forecasts via an advanced data assimilation (DA) procedure based on a rainfall ensemble generator and lagged particle filtering. The former provides spatial and temporal correlation structures of the input forcing and the latter updates past and current state variables in a lag-time window without loss of information due to updating time intervals. Two ways of specifying noise in a DA scheme were compared with a random walk noise having subjective error assumptions and objectively induced input forcing using the multivariate rainfall generator. The procedure was evaluated for three flood events in two Japanese medium-sized catchments (Maruyama and Katsura) with fine-resolution numerical weather prediction output.

Rainfall ensembles based on different assumptions, such as univariate (time-independent) and multivariate (time-dependent) conditional simulations (MCS), were compared. The MCS showed a slightly larger spread and improved temporal consistency and coherence compared to the univariate simulation.

Sensitivity analysis was carried out with a different setup of random state noise and input forcing

to assess impacts on the forecast quality. Lagged particle filtering (LPF) forced with MCS provided good performance with small and consistent random state noise, while LPF forced with Thiessen rainfall interpolation required larger random state noise to yield comparable performance as LPF+MCS for short lead times.

In the forecasting mode, the results showed improved performance via DA compared to the open loop simulations for all lead times. The forecasts using MCS showed improved or at least equivalent performance compared to those by Thiessen interpolation. Gains by MCS were higher in the headwater area, however, significant differences were not found at the outlet, especially when the accuracy of NWP deteriorated sharply. Overall, MCS provided improvement compared to Thiessen interpolation in the hindcasting mode.

The strength of the proposed procedure is that it requires less subjectivity in implementation of DA compared to conventional methods. The multivariate rainfall generator provides an appropriate spatial and temporal error structure given the rain gauge observations and simulations. Therefore, subjective use of additional random noise in DA is mitigated. The study demonstrates that advances in different DA methods can be combined properly to reduce the subjectivity in DA procedures. A future study is expected to expand the proposed DA framework to large scale flood forecasting with more realistic artificial regulations, merging with new techniques and input forcing such as radar rainfall estimates.

Chapter 7

Synthesis

This thesis has contributed to improved understanding and quantification of hydrological model uncertainty especially related to the initial conditions of the model and to a lesser extent to the model structure and parameters. This final chapter synthesizes the most important findings presented in detail in Chapters 2 – 6. First, the research questions posed in the introduction (Sect. 1.3) are answered and second, the overall results of the whole thesis are put into a broader perspective.

7.1 Answers to the research questions

7.1.1 Quantification of model input uncertainty

How to define a plausible precipitation ensemble generator using rain gauges to capture the temporal coherence for each realisation of a sequence of spatial rainfall fields at an hourly time step?

Unlike interpolation, the conditional simulation technique provides both the best local estimate and ensures that realisations match the sample statistics and are conditional on neighbouring estimates (e.g. *Goova erts*, 1997). Chapter 2 introduced a temporally coherent error identification method for rainfall, which was based on time-dependent multivariate spatial conditional simulations, which were conditioned on preceding simulations. The experiment was carried out using 27 rain gauges within the hilly region of the Belgian Ardennes. The sensitivity analysis using synthetic data showed that for typical advection speeds of $> 20\,\mathrm{km\,h^{-1}}$ no uncertainty in terms of across-ensemble spread lumped over time (expressed using the coefficient of variation) was added to simulated precipitation fields by conditioning them on more than one or two previous hourly simulations. In the real-world experiment, which exhibited a larger spatial and temporal variability, the time-dependent simulations required somewhat longer simulation memories. Additionally, by halving the observation network, i.e., using 14 rain gauges, the uncertainty in the sensitivity analysis increased only slightly. Furthermore, the cross-validation showed that the simulations at unobserved locations are unbiased and have a consistent behaviour over all simulation memories. Finally, the first-order autocorrelation coefficient indicated the presence of temporal coherence in the time series of the areal precipitation using the time-dependent multivariate conditional simulation in comparison with the time-independent univariate conditional simulations. This coherence was found to decrease with increased catchment area. The presented technique to generate spatial precipitation ensembles can be easily implemented within

a hydrological data assimilation framework to be used as an improvement over currently used simplistic approaches to perturb the interpolated point estimates or spatially distributed estimates.

7.1.2 Distributed evaluation of parameter sensitivity

How to efficiently and effectively evaluate parameter sensitivity, using multiple evaluations of local parameter sensitivity distributed throughout parameter space?

Sensitivity analysis identifies how different components of a model affect model output and consequences for model calibration, uncertainty evaluation and risk assessment (e.g., *Saltelli et al.*, 2008). To reveal how sensitivity analysis measures vary as parameter values change, Chapter 3 presented and evaluated the hybrid local-global Distributed Evaluation of Local Sensitivity Analysis (DELSA) method, in which multiple evaluations of local parameter sensitivity were distributed throughout parameter space. DELSA is a new computationally frugal method, which integrates methodological features from three existing sensitivity analysis methods: the Method of Morris (MoM; *Morris*, 1991), the Sobol' method (*Sobol'*, 2001), and regional sensitivity analysis (RSA; *Hornberger and Spear*, 1981). Chapter 3 suggested the potential utility of local methods in the context of multiscale evaluation of parameter sensitivity. Evaluation included comparisons with the global variance-based Sobol' method and was applied to Lasnenville catchment (200 km^2) in the Belgian Ardennes. To enable comparison, the Sobol' method was sometimes applied at a smaller than normal scale; and sometimes local DELSA measures were averaged over large-scale parameter space. Results showed that in both examples, Sobol' and DELSA identified similar important and unimportant parameters, with DELSA enabling more detailed insight at much lower computational cost. For example, in the real-world problem the time delay in runoff was the most important parameter in all models, but DELSA showed that for about 20% of parameter sets it was not important at all and alternative mechanisms and parameters dominated. Moreover, the time delay was identified as important in regions producing poor model fits, whereas other parameters were identified as more important in regions producing better model fits. The ability to understand how parameter importance varies through parameter space is critical to inform decisions about, for example, additional data collection and model development.

7.1.3 Improving initial conditions of hydrological model using discharge data assimilation

What is the sensitivity of the data assimilation procedure using the Ensemble Kalman Filter to the number and the locations of discharge gauges, which are assimilated into a grid-based distributed operational hydrological forecasting model?

The results presented in Chapter 4 for the Upper Ourthe catchment showed that the hydrological forecast at the catchment outlet was improved by assimilating interior gauges. By using a physically based model for the routing, the time delay and attenuation were modelled more realistically. The discharge and states at a given time step were assumed to be dependent on the previous time step only (Markov property). The EnKF procedure was mainly changing the pdfs of the two routing model storages, even when the uncertainty in the discharge simulations was smaller than the defined observation uncertainty. This was because the current formulation of the EnKF did not explicitly

consider the strong correlation between soil moisture states in the immediate past and streamflow at the time of forecast. Moreover, with an increasing number of discharge observations, the centroid of the updated histograms within the observation error bounds was approaching the true value more closely and with smaller variance than for less dense discharge observation networks.

What is the effect of state augmentation using a time window of past discharge observations assimilated by the Asynchronous Ensemble Kalman Filter (AEnKF; Sakov et al., 2010) on forecast accuracy? Additionally, how does the choice of model states being updated affect the forecast accuracy?

Chapter 5 showed that an augmented assimilation vector improved hydrological forecasts, but that contribution became smaller for longer lead times. Additionally, Chapter 5 showed that for the Upper Ourthe catchment reducing the number of model states of a grid-based HBV model using AEnKF could lead to better forecasts of the discharge. In terms of the root-mean-square-error, the largest improvements in the forecast accuracy were observed for the scenario where the soil moisture was left out from the analysis (similar to the PDM updating scheme presented by *Moore*, 2007). This indicates that elimination of the strongly non-linear relation between soil moisture storage (SM) and assimilated discharge observations can become beneficial for improved forecasts. Additionally, it was also shown that keeping the quick catchment response storage (upper zone; UZ) in the model analysis was important, especially for longer lead times, when compared to the scenario in which only two routing storages were updated.

What is the effect of noise specification in a DA scheme on the quality of hydrological forecasts?

In Chapter 6, an advanced sequential DA procedure for two Japanese catchments employed two methods: (1) lagged particle filtering (*Noh et al.*, 2011a) and (2) the multivariate rainfall ensemble generator (Chapter 2). The first method was used to update past and current state variables simultaneously in a lag-time window to consider the response times of internal hydrologic processes. The second method was used to obtain spatial and temporal correlation error structures of input forcing at hourly time step. The results showed that objectively generated input forcing using the multivariate rainfall generator reduced the need for larger and subjective error noise model assumptions. The strength of the proposed procedure is that it requires less subjectivity in the implementation of DA compared to conventional methods. Therefore subjective use of additional random noise in DA is mitigated. Finally, the gains of objective noise and state updating deteriorated sharply when used in combination with a meso-scale numerical weather prediction model.

7.2 Outlook

This thesis provided insight in discharge assimilation to update grid-based hydrological models. Chapters 4 and 5 showed the potential of model updating for improved flood forecasting. The best results obtained in Chapter 5, in terms of the root-mean-square-error (RMSE), using model updating are presented by the dashed line in Fig. 7.1 (see also the original Fig. 5.5). This represents a considerable improvement as compared to control runs without any model updating (dotted line). However, there

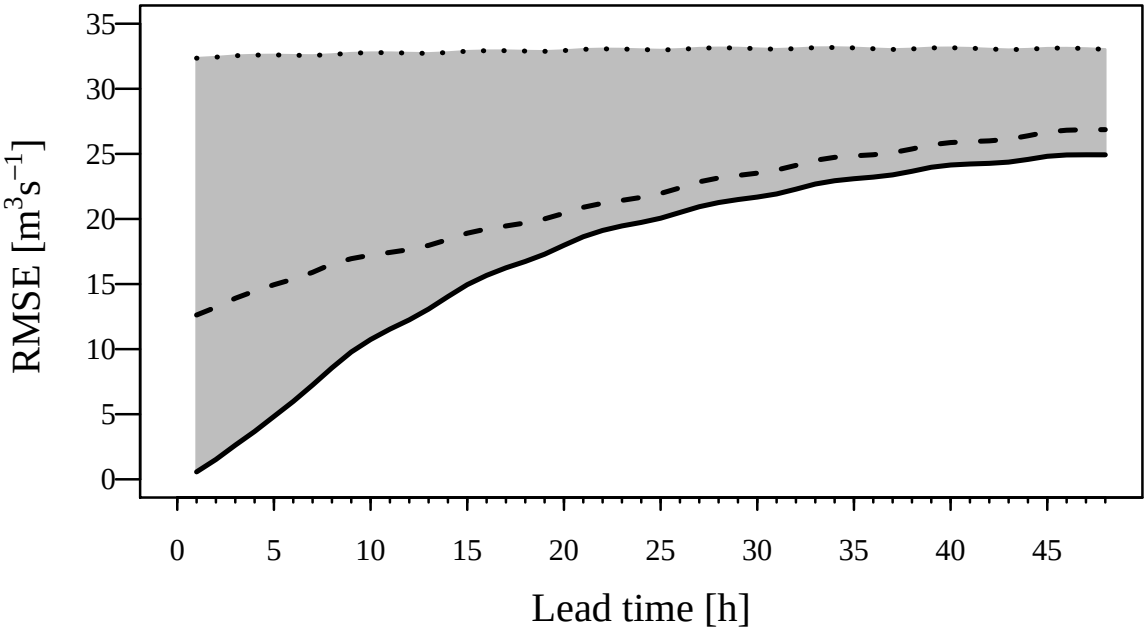


Figure 7.1: Root-mean-square-error (RMSE, Eq. 4.14) as a measure of forecast skill at Tabreux using the grid-based HBV model for the 8 largest flood peaks observed since 1998 (see Chapter 5, Table 5.1). The grey region indicates feasible RMSE values using different modelling approaches, varying between the upper bound (dotted line) representing control simulations without an update, and the lower bound (solid line) representing the output error updating scenario. The best achievement using discharge assimilation for model state updating (see Chapter 5, Fig. 5.5) is shown by the dashed line.

remains room for further improvement, because the benefits of model updating are not as skillful as when employing a simple error updating procedure (Sect. 1.2.4, *Broersen and Weerts*, 2005). This is presented by the solid line in Fig. 7.1. It is important to note that error updating is based purely on the statistics of the errors/residuals between the past model forecasts and observations, which is fitted using a 3-parameter auto-regressive model (minimizing the RMSE for each update). Additionally, error updating does not incorporate any knowledge about hydrology or the associated uncertainties. This approach is currently used operationally.

The gap between the RMSE for model updating (dashed line) and error updating (solid line) in Fig. 7.1 could possibly be eliminated and ideally even reversed by including model structural uncertainty in the data assimilation framework. Note that one of the limiting factors in Chapters 4 and 5 was that only one model with a fixed structure was used. Possibly, the DELSA methodology developed in Chapter 3 could be used to evaluate model structure more appropriately. Nevertheless, the choice of an appropriate model structure to simulate and predict hydrological processes is an outstanding research challenge (e.g. *Gupta et al.*, 2012). This could possibly be achieved when the hydrological analysis of the modeled processes would further evolve from focusing only on simulated discharge, to also including other observations such as soil moisture and groundwater levels as well. In this sense, the DELSA method is ideally suited to analyze multiple metrics of model behaviour (e.g. *Rosolem et al.*, 2012). The use of a hierarchical system of metrics was also underlined by *Gupta*

et al. (2012). Moreover, such a model evaluation should be conducted in a way that isolates different processes within models – this would help avoid compensatory errors that provide "good" simulations for the wrong reasons. This is important, especially when the conditions shift beyond the range of prior experience, for example due to extreme precipitation events, climate change, or shifts in land use (*Kirchner*, 2006).

Probably, the current use of simple conceptual models in catchment hydrology needs to be reconsidered, because of their limited ability to represent the dynamics of a vast variety of hydrological processes occurring in natural catchments (*Clark et al.*, 2011a; *Gupta et al.*, 2012). Therefore, the future of operational catchment hydrological modeling can be seen in employing more complex, more physically/process-based models, which can better represent the movement of water through the catchment (*Clark et al.*, 2011b). Improved physical representation of hydrological processes in hydrological models was ranked high by *Wetterhall et al.* (2013) among the top priorities of the professional flood forecasters. This goal, however, represents one of the most difficult and most expensive actions for improved flood forecasts. *Wetterhall et al.* (2013) also speculate that implementing multi-model approaches could lead to more robust decisions applicable over a wider range of flow situations.

In addition to a focus on the aforementioned objectives regarding model structure adequacy, there exists a vast number of other research opportunities within data assimilation for operational hydrological forecasting. For example, *Liu et al.* (2012) suggest that developments in spaceborne and ground-based remote sensing retrieval algorithms (e.g. weather radar) are important for operational forecasting, as new measurement techniques emerge (see *Liu et al.*, 2012, for a comprehensive overview). Furthermore, as the spatial and temporal characteristics of the measured variables are rarely identical, there is a need for aggregation and/or disaggregation and associated bias corrections and error estimation methods. On top of that, the observation matrices from remote sensing might become very large and current engineering infrastructure often prohibits real-time DA application. Therefore, the computational efficiency and optimization of DA algorithms should be enhanced.

Within the hydrological community, the application of DA experiments is often missing a comprehensive evaluation, which would enable a fair inter-comparison of algorithms among various research groups using the same datasets. The lack of data sharing often prohibits scientific collaboration, in particular in Europe, where it still represents a sensitive issue. Hopefully, these inhibitory effects will not persist too long into the 21st century. Another issue represents the incompatibility of numerical code implementation among various research groups, which is a real challenge for practical and robust code inter-comparisons. Therefore, development of user-oriented generic toolboxes, such as *OpenDA* (2013), presented in Chapter 5, could facilitate the implementation of various DA algorithms (*Liu et al.*, 2012). Another option could be making use of e.g. the Data Assimilation Research Testbed (*DART*, 2013), an open source community framework originating from the atmospheric sciences.

The current configuration employed to produce the results presented in Chapter 5 is implemented in a stand-alone version of the RWsOS Rivers, an operational forecasting system used by the Dutch authorities. However, it can be easily switched into a configuration with real-time data import. Although the current test case involves only one medium-sized sub-catchment of the Meuse basin, it can in principle be set up for the complete set of sub-catchments within the whole Meuse basin, and also for the Rhine basin (Sect. 1.2.4). Nevertheless, human control, such as reservoir operations, needs to be considered for an appropriate implementation. Additionally, on-going applications of the methodology are currently being developed and tested for several cases outside Europe, such as the

Taiwanese forecasting system. First results suggest that the methodology is robust and applicable to different climatic and hydrological conditions.

Bibliography

AghaKouchak, A., A. Bardossy, and E. Habib (2010), Conditional simulation of remotely sensed rainfall fields using a non-Gaussian v-transformed copula, *Adv. Water Resour.*, *33*(6), 624–634, doi:10.1016/j.advwatres.2010.02.010.

Anderman, E. R., M. C. Hill, and E. P. Poeter (1996), Two-dimensional advective transport in groundwater flow parameter estimation, *Groundwater*, *34*(6), 1001–1009.

Archer, G. E. B., A. Saltelli, and I. M. Sobol (1997), Sensitivity measures, ANOVA-like techniques and the use of bootstrap, *J. Statist. Comput. Simulation*, *58*(2), 99–120, doi: 10.1080/00949659708811825.

Arulampalam, M. S., S. Maskell, N. Gordon, and T. Clapp (2002), A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, *IEEE Transactions On Signal Processing*, *50*(2), 174–188.

Aster, R. C., B. Borchers, and C. H. Thurber (2013), *Parameter estimation and inverse problems*, Amsterdam, Academic Press.

Aubert, D., C. Loumagne, and L. Oudin (2003), Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall - runoff model, *J. Hydrol.*, *280*(1–4), 145–161.

BBC (2013), India floods: More than 5,700 people 'presumed dead', http://www.bbc.co.uk/news/world-asia-india-23282347, (last access: 1 November 2013).

Bedford, T. (1998), Sensitivity indices for (tree)-dependent variables, in *Proceedings of the Second International Symposium on Sensitivity Analysis of Model Output, Venice, Italy (pp. 17–20).*

Berenguer, M., C. Corral, R. Sanchez-Diezma, and D. Sempere-Torres (2005), Hydrological validation of a radar-based nowcasting technique, *J. Hydrometeorol.*, *6*(4), 532–549.

Berenguer, M., D. Sempere-Torres, and G. G. S. Pegram (2011), SBMcast – An ensemble nowcasting technique to assess the uncertainty in rainfall forecasts by Lagrangian extrapolation, *J. Hydrol.*, *404*(3–4), 226 –240, doi:10.1016/j.jhydrol.2011.04.033.

Bergström, S., and G. Sandberg (1983), Simulation of groundwater response by conceptual models, *Nord. Hydrol.*, pp. 71–84.

Berne, A., G. Delrieu, J.-D. Creutin, and C. Obled (2004), Temporal and spatial resolution of rainfall measurements required for urban hydrology, *J. Hydrol.*, *299*(3–4), 166–179, doi: 10.1016/j.jhydrol.2004.08.002.

Berne, A., M. ten Heggeler, R. Uijlenhoet, L. Delobbe, P. Dierickx, and M. de Wit (2005), A preliminary investigation of radar rainfall estimation in the Ardennes region and a first hydrological application for the Ourthe catchment, *Nat. Hazards Earth Syst. Sci.*, *5*, 267–274.

Beven, K. (1993), Prophecy, reality and uncertainty in distributed hydrological modelling, *Adv. Water Resour.*, *16*, 41–51.

Beven, K., and A. Binley (1992), The future of distributed models: model calibration and uncertainty prediction, *Hydrol. Processes*, *6*(3), 279–298.

Beven, K., and M. Kirkby (1979), A physically based, variable contributing area model of basin hydrology, *Hydrological Sciences-Bulletin-des Sciences Hydrologiques*, *24*(1), 43–69.

Beven, K., P. Lamb, P. Quinn, R. Romanowicz, and J. Freer (1995), TOPMODEL, in *Computer Models of Watershed Hydrology*, edited by V. P. Singh, chap. TOPMODEL, Water Resources Publications.

Beven, K. J. (2001), *Rainfall-runoff modelling: the primer*, John Wiley & Sons.

Biondi, D., and D. L. De Luca (2013), Performance assessment of a Bayesian Forecasting System (BFS) for real-time flood forecasting, *J. Hydrol.*, *479*, 51–63.

Bivand, R. S., E. J. Pebesma, and V. Gómez-Rubio (2008), *Applied spatial data analysis with R*, Springer.

Blöschl, G., C. R. Reszler, and J. Komma (2008), A spatially distributed flash flood forecasting model, *Environ. Mod. & Soft.*, *23*, 464–478.

Blöschl, G., T. Nester, J. Komma, J. Parajka, and R. A. P. Perdigão (2013), The June 2013 flood in the Upper Danube basin, and comparisons with the 2002, 1954 and 1899 floods, *Hydrol. Earth Syst. Sci. Discuss.*, *10*(7), 9533–9573, doi:10.5194/hessd-10-9533-2013.

Booij, M. (2002), *Appropriate modelling of climate change impacts on river flooding*, Ph.D. thesis, University of Twente, Enschede, Netherlands.

Borgonovo, E. (2006), Measuring uncertainty importance: investigation and comparison of alternative approaches, *Risk Anal.*, *26*(5), 1349–61, doi:10.1111/j.1539-6924.2006.00806.x.

Borgonovo, E. (2007), A new uncertainty importance measure, *Reliab. Eng. Syst. Safe.*, *92*(6), 771–784, doi:10.1016/j.ress.2006.04.015.

Box, G. E. P., and G. C. Tiao (1992), *Bayesian inference in statistical analysis*, New York, Wiley.

Bratley, P., and B. l. Fox (1988), Implementing Sobol's quasirandom sequence generator, *ACM T. Math. Software*, *14*(1), 88–100, doi:10.1145/42288.214372.

Brauer, C. C., A. J. Teuling, A. Overeem, Y. van der Velde, P. Hazenberg, P. M. M. Warmerdam, and R. Uijlenhoet (2011), Anatomy of extraordinary rainfall and flash flood in a Dutch lowland catchment, *Hydrol. Earth Syst. Sci.*, *15*(6), 1991–2005, doi:10.5194/hess-15-1991-2011.

Brauer, C. C., A. J. Teuling, P. J. J. F. Torfs, and R. Uijlenhoet (2014), The Wageningen Lowland Runoff Simulator (WALRUS): a lumped rainfall-runoff model for catchments with shallow groundwater, *Geosci. Model Dev. Discuss.*, *7*(1), 1357–1411, doi:10.5194/gmdd-7-1357-2014.

Broersen, P. M. T., and A. H. Weerts (2005), Automatic Error Correction of Rainfall-Runoff models in Flood Forecasting Systems, in *Instrumentation and Measurement Technology Conference Proceedings, Ottawa, Canada, 17-19 May 2005*.

Burnash, R. J. C. (1995), The NWS river forecast system – catchment modeling, in *Computer Models of Watershed Hydrology*, edited by V. P. Singh, Water Resources Publications.

Chen, Z. (2003), Bayesian Filtering: From Kalman Filters to Particle Filters and Beyond, Adaptive Syst. Lab., McMaster University, Hamilton, ON, Canada.

Chow, V. T., D. Maidment, and L. Mays (1988), *Applied Hydrology*, McGraw-Hill.

Chun, M.-H., S.-J. Han, and N.-I. Tak (2000), An uncertainty importance measure using a distance metric for the change in a cumulative distribution function, *Reliab. Eng. Syst. Safe.*, *70*(3), 313–321, doi:10.1016/S0951-8320(00)00068-5.

Ciach, G., and W. Krajewski (2006), Analysis and modeling of spatial correlation structure in small-scale rainfall in Central Oklahoma, *Adv. Water Resour.*, *29*(10), 1450–1463, doi:10.1016/j.advwatres.2005.11.003.

Clapp, R. B., and G. M. Hornberger (1978), Empirical equations for some soil hydraulic properties, *Water Resour. Res.*, *14*(4), 601–604.

Clark, M., and D. Kavetski (2010), Ancient numerical daemons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes, *Water Resour. Res.*, *46*(10), 1–23, doi:10.1029/2009WR008894.

Clark, M. P., and A. G. Slater (2006), Probabilistic quantitative precipitation estimation in complex terrain, *J. Hydrometeorol.*, *7*, 3–22, doi:10.1175/JHM474.1.

Clark, M. P., A. G. Slater, D. E. Rupp, R. A. Woods, J. A. Vrugt, H. V. Gupta, T. Wagener, and L. E. Hay (2008a), Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, *Water Resour. Res.*, *44*, 1–14, doi:10.1029/2007WR006735.

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

Clark, M. P., D. Kavetski, and F. Fenicia (2011a), Pursuing the method of multiple working hypotheses for hydrological modeling, *Water Resour. Res.*, *47*(9), 1–16, doi:10.1029/2010WR009827.

Clark, M. P., H. K. McMillan, D. B. G. Collins, D. Kavetski, and R. A. Woods (2011b), Hydrological field data from a modeller's perspective: Part 2: Process-based evaluation of model hypotheses, *Hydrol. Processes*, *25*(4), 523–543, doi:10.1002/hyp.7902.

Cloke, H. L., and F. Pappenberger (2009), Ensemble flood forecasting: A review, *J. Hydrol.*, *375*, 613–626.

Cloke, H. L., F. Pappenberger, and J.-P. Renaud (2007), Multi-Method Global Sensitivity Analysis (MMGSA) for modelling floodplain hydrological processes, *Hydrol. Processes*, *22*(11), 1660–1674, doi:10.1002/hyp.6734.

Cole, S. J., and R. J. Moore (2009), Distributed hydrological modelling using weather radar in gauged and ungauged basins, *Adv. Water Resour.*, *32*, 1107–1120.

Collier, C. G. (2007), Flash flood forecasting: What are the limits of predictability?, *Q. J. R. Meteorol. Soc.*, *133*, 3–23.

Cook, R. D., and S. Weisberg (1982), *Residuals and Influence in Regression*, New York: Chapman and Hall.

Crow, W. T., and D. Ryu (2009), A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals, *Hydrol. Earth Syst. Sci.*, *13*(1), 1–16, doi:10.5194/hess-13-1-2009.

Cukier, R. I., C. M. Fortuin, K. E. Shuler, A. G. Petschek, and J. H. Schaibly (1973), Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I. Theory, *J. Chem. Phys.*, *59*(8), 3873–3878, doi:10.1063/1.1680571.

Cukier, R. I., J. H. Schaibly, and K. E. Shuler (1975), Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. III. Analysis of the approximations, *J. Chem. Phys.*, *63*(3), 1140–1149, doi:10.1063/1.431440.

Cukier, R. I., H. B. Levine, and K. E. Shuler (1978), Nonlinear sensitivity analysis of multiparameter model systems, *J. Comput. Phys.*, *26*(1), 1–42, doi:10.1016/0021-9991(78)90097-9.

D'Agnese, F. A., C. C. Faunt, M. C. Hill, and A. Keith Turner (1999), Death valley regional groundwater flow model calibration using optimal parameter estimation methods and geoscientific information systems, *Adv. Water Resour.*, *22*(8), 777–790.

DART (2013), Data Assimilation Research Testbed, http://www.image.ucar.edu/DAReS/DART, (last access: 1 November 2013).

de Roo, A. P. J., C. G. Wesseling, and W. P. A. Van Deursen (2000), Physically based river basin modelling within a GIS: The LISFLOOD model, *Hydrol. Processes*, *14*, 1981–1992, doi:10.1002/1099-1085(20000815/30)14:11/12¡1981::AID-HYP49¿3.0.CO;2-F.

de Wit, M. (2008), *Van regen tot Maas*, Veen magazines, Diemen.

DeChant, C. M., and H. Moradkhani (2012), Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting, *Water Resour. Res.*, *48*, doi:10.1029/2011WR011011.

Delenne, C., B. Cappelaere, and V. Guinot (2012), Uncertainty analysis of river flooding and dam failure risks using local sensitivity computations, *Reliab. Eng. Syst. Safe.*, *107*, 171–183, doi: 10.1016/j.ress.2012.04.007.

Deutsche Welle (2013), China, Russia hit by severe floods, http://dw.de/p/19Rf0, (last access: 1 November 2013).

Dingman, S. (2004), *Physical Hydrology*, Prentice-Hall Inc.

Doherty, J. (2003), Ground water model calibration using pilot points and regularization, *Groundwater*, *41*, 170–177.

Doherty, J. (2009), *PEST: Model-Independent Parameter Estimation*, Watermark Numerical Computing, Australia.

Doherty, J., and D. Welter (2010), A short exploration of structural noise, *Water Resour. Res.*, *46*(5), 1–14, doi:10.1029/2009WR008377.

Dooge, J. C. I., and J. P. O'Kane (2003), *Deterministic Methods in Systems Hydrology*, A. A. Balkema Publishers.

Draper, N. R., and H. Smith (1998), *Applied regression analysis*, New York, Wiley.

Driessen, T. L. A., R. T. W. L. Hurkmans, W. Terink, P. Hazenberg, P. J. J. F. Torfs, and R. Uijlenhoet (2010), The hydrological response of the Ourthe catchment to climate change as modelled by the HBV model, *Hydrol. Earth Syst. Sci.*, *14*(4), 651–665, doi:10.5194/hess-14-651-2010.

Dunn, S. M., J. Freer, M. Weiler, M. Kirkby, D. Seibert, J., Quinn, P. F., Lischeid, G., Tetzlaff, and C. Soulsby (2008), Conceptualization in catchment modelling: Simply learning?, *Hydrol. Processes*, *22*(13), 2389–2393.

Dunne, S., and D. Entekhabi (2006), Land surface state and flux estimation using the ensemble Kalman smoother during the Southern Great Plains 1997 field experiment, *Water Resour. Res.*, *42*(1), 1–15, doi:10.1029/2005WR004334.

Edijatno, N. D. De Oliveira Nascimento, X. Yang, Z. Makhlouf, and C. Michel (1999), GR3J: A daily watershed model with three free parameters, *Hydrological Sciences Journal*, *44*(2), 263–277.

Efron, B., and R. J. Tibshirani (1993), *An Introduction to the Bootstrap*, Chapman & Hall, New York.

Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*(C5), 10,143–10,162.

Evensen, G. (2003), The Ensemble Kalman Filter: theoretical formulation and practical implementation, *Ocean Dyn.*, *53*, 343–367.

Evensen, G. (2009), *Data Assimilation: The Ensemble Kalman Filter*, Springer, doi:10.1007/978-3-642-03711-5.

Evensen, G., and P. Van Leeuwen (2000), An ensemble Kalman smoother for nonlinear dynamics, *Mon. Wea. Rev.,*, *128*(6), 1852–1867.

Fenicia, F., H. Savenije, P. Matgen, and L. Pfister (2008), Understanding catchment behavior through stepwise model concept improvement, *Water Resour. Res.*, *44*(1), 1–13, doi: 10.1029/2006WR005563.

Fienen, M. N., J. E. Doherty, R. J. Hunt, and H. W. Reeves (2010), *Using Prediction Uncertainty Analysis to Design Hydrologic Monitoring Networks: Example Applications from the Great Lakes Water Availability Pilot Project.*, U.S. Geological Survey Scientific Investigations Report 2010-5159: 44.

Foglia, L., S. W. Mehl, M. C. Hill, P. Perona, and P. Burlando (2007), Testing alternative ground water models using cross-validation and other methods, *Groundwater*, *45*(5), 627–641, doi: 10.1111/j.1745-6584.2007.00341.x.

Foglia, L., M. C. Hill, S. W. Mehl, and P. Burlando (2009), Sensitivity analysis, calibration, and testing of a distributed hydrological model using error-based weighting and one objective function, *Water Resour. Res.*, *45*(6), 1–18, doi:10.1029/2008WR007255.

Foglia, L., M. C. Hill, S. W. Mehl, and P. Burlando (2013), Evaluating model structure adequacy: The case of the Maggia Valley groundwater system, Southern Switzerland, *Water Resour. Res.*, *49*, 1–13, doi:10.1029/2011WR011779.

Freer, J., K. Beven, and B. Ambroise (1996), Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach, *Water Resour. Res.*, *32*(7), 2161–2173, doi:10.1029/95WR03723.

Georgakakos, K. P. (1986), A generalized stochastic hydrometeorological model for flood and flash-flood forecasting, *Water Resour. Res.*, *22*(13), 2083–2095.

Georgakakos, K. P., D.-J. Seo, H. Gupta, J. Schaake, and M. B. Butts (2004), Towards the characterization of streamflow simulation uncertainty through multimodel ensembles, *J. Hydrol.*, *298*, 222–241.

Germann, U., M. Berenguer, D. Sempere-Torres, and M. Zappa (2009), REAL - Ensemble radar precipitation estimation for hydrology in a mountainous region, *Q. J. R. Meteorol. Soc.*, *135*(639, Part B), 445–456, doi:10.1002/qj.375.

Gilks, W., and C. Berzuini (2001), Following a moving target – Monte Carlo inference for dynamic Bayesian models, *J. R. Statist. Soc. B*, pp. 127–146.

Giustarini, L., P. Matgen, R. Hostache, M. Montanari, D. Plaza, V. R. N. Pauwels, G. J. M. De Lannoy, R. De Keyser, L. Pfister, L. Hoffmann, and H. H. G. Savenije (2011), Assimilating SAR-derived water level data into a hydraulic model: a case study, *Hydrol. Earth Syst. Sci.*, *15*(7), 2349–2365, doi:10.5194/hess-15-2349-2011.

Glassner, A. S. (Ed.) (1998), *Graphics gems*, Elsevier.

Göhler, M., J. Mai, and M. Cuntz (2013), Use of eigendecomposition in a parameter sensitivity analysis of the Community Land Model, *J. Geophys. Res-Biogeo.*, *118*(2), 904–921, doi: 10.1002/jgrg.20072.

Goova erts, P. (1997), *Geostatistics for Natural Resources Evaluation*, Applied geostatistics series, 483 pp., Oxford University Press.

Google maps (2011), http://maps.google.co.uk/.

Goudenhoofdt, E., and L. Delobbe (2009), Evaluation of radar-gauge merging methods for quantitative precipitation estimates, *Hydrol. Earth Syst. Sci.*, *13*(2), 195–203, doi:10.5194/hess-13-195-2009.

Grimes, D. I. F., and E. Pardo-Iguzquiza (2010), Geostatistical analysis of rainfall, *Geogr. Anal.*, *42*(2), 136–160, doi:10.1111/j.1538-4632.2010.00787.x.

Gupta, H. V., M. P. Clark, J. A. Vrugt, G. Abramowitz, and M. Ye (2012), Towards a comprehensive assessment of model structural adequacy, *Water Resour. Res.*, *48*(8), 1–16, doi: 10.1029/2011WR011044.

Habets, F., P. LeMoigne, and J. Noilhan (2004), On the utility of operational precipitation forecasts to served as input for streamflow forecasting, *J. Hydrol.*, *293*, 270–288.

Hamm, N. A. S., J. W. Hall, and M. G. Anderson (2006), Variance-based sensitivity analysis of the probability of hydrologically induced slope instability, *Computers & Geosciences*, *32*(6), 803–817, doi:10.1016/j.cageo.2005.10.007.

Han, E., V. Merwade, and G. Heathman (2012), Implementation of surface soil moisture data assimilation with watershed scale distributed hydrological model, *J. Hydrol.*, *416–417*, 98–117, doi: 10.1016/j.jhydrol.2011.11.039.

Hapuarachchi, H. A. P., Q. J. Wang, and T. C. Pagano (2011), A review of advances in flash flood forecasting, *Hydrol. Processes*, *25*, 2771–2784.

Havno, K., M. N. Madsen, and J. Dorge (1995), MIKE 11, in *Computer Models of Watershed Hydrology*, edited by V. P. Singh, Water Resources Publications.

Hazenberg, P., H. Leijnse, and R. Uijlenhoet (2011), Radar rainfall estimation of stratiform winter precipitation in the Belgian Ardennes, *Water Resour. Res.*, *47*(2), 1–15, doi:10.1029/2010WR009068.

Helton, J. C. (1993), Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal, *Reliab. Eng. Syst. Safe.*, *42*(2–3), 327–367, doi:10.1016/0951-8320(93)90097-I.

HEPEX (2013), Expect the unexpected: Flash floods in Northern Colorado September 2013, http://hepex.irstea.fr/boulder-flooding-september-2013, (last access: 1 November 2013).

Herman, J. D., P. M. Reed, and T. Wagener (2013a), Time-varying sensitivity analysis clarifies the effects of watershed model formulation on model behavior, *Water Resour. Res.*, pp. 1–15, doi: 10.1002/wrcr.20124.

Herman, J. D., J. B. Kollat, P. M. Reed, and T. Wagener (2013b), Technical note: Method of Morris effectively reduces the computational demands of global sensitivity analysis for distributed watershed models, *Hydrol. Earth Syst. Sci..*, *17*, 2893–2903, doi:10.5194/hessd-10-4275-2013.

Hill, M. C., and C. R. Tiedeman (2007), *Effective groundwater model calibration: with analysis of data, sensitivities, prediction and uncertainty*, 455 pp., Wiley.

Hill, M. C., D. Kavetski, M. Clark, M. Ye, M. Arabi, D. Lu, L. Foglia, and S. Mehl (2013), A strategy to enhance transparency and refutability of environmental models using local to global methods, *Submitted to Water Resources Research (in review)*.

Homma, T., and A. Saltelli (1996), Importance measures in global sensitivity analysis of nonlinear models, *Reliab. Eng. Syst. Safe.*, *52*(1), 1–17, doi:10.1016/0951-8320(96)00002-6.

Hornberger, G., and R. Spear (1981), An approach to the preliminary analysis of environmental systems, *J. Environ. Management*, *12*(1), 7–18.

Houtekamer, P. L., and H. L. Mitchell (2001), A sequential ensemble Kalman filter for atmospheric data assimilation, *Mon. Wea. Rev.,*, *129*(1), 123–137.

IPCC (2013), Intergovernmental Panel on Climate Change, http://www.ipcc.ch, (last access: 1 November 2013).

Jasper, K., J. Gurtz, and H. Lang (2002), Advanced flood forecasting in Alpine watersheds by coupling meteorological observations and forecasts with a distributed hydrological model, *J. Hydrol.*, *267*, 40–52.

Jia, Y., G. Ni, Y. Kawahara, and T. Suetsugi (2001), Development of WEP model and its application to an urban watershed, *Hydrol. Processes*, *15*, 2175–2194.

Jia, Y., X. Ding, C. Qin, and H. Wang (2009), Distributed modeling of landsurface water and energy budgets in the inland Heihe river basin of China, *Hydrol. Earth Syst. Sci.*, *13*, 1849–1866.

Kalman, R. E. (1960), A New Approach to Linear Filtering and Prediction Problems, *ASME - Journal of Basic Engineering*, *82*, 35–45.

Karssenberg, D., O. Schmitz, P. Salamon, C. De Jong, and M. F. P. Bierkens (2009), A software framework for construction of process-based stochastic spatio-temporal models and data assimilation, *Environ. Mod. & Soft.*, *25*, 1–14, doi:10.1016/j.envsoft.2009.10.004.

Katz, R. W. (1999), Extreme value theory for precipitation: Sensitivity analysis for climate change, *Adv. Water Resour.*, *23*(2), 133–139, doi:10.1016/S0309-1708(99)00017-2.

Kavetski, D., and M. Clark (2010), Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction, *Water Resour. Res.*, *46*(10), 1–27, doi:10.1029/2009WR008896.

Kirchner, J. W. (2006), Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, *Water Resour. Res.*, *42*(3), 1–16, doi:10.1029/2005WR004362.

Kirstetter, P.-E., G. Delrieu, B. Boudevillain, and C. Obled (2010), Toward an error model for radar quantitative precipitation estimation in the Cévennes-Vivarais region, France, *J. Hydrol.*, *394*(1–2), 28–41, doi:10.1016/j.jhydrol.2010.01.009.

Kitanidis, P. K., and R. L. Bras (1980), Real time forecasting with a conceptual hydrological model 1: Analysis of uncertainty, *Water Resour. Res.*, *16*(6), 1025–1033.

Kleijnen, J. P. C., and J. C. Helton (1999), Statistical analysis of scatterplots to identify important factors in large-scale simulations, 1: Review and comparison techniquest, *Reliab. Eng. Syst. Safe.*, *65*(2), 147–185.

Klemeš, V. (1986), Operational testing of hydrological simulation models, *Hydrological Sciences Journal*, *31*(1), 13–24.

Kollet, S. J., and R. M. Maxwell (2006), Integrated surface–groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model, *Adv. Water Resour.*, *29*(7), 945–958, doi:10.1016/j.advwatres.2005.08.006.

Komma, J., G. Blöschl, and C. Reszler (2008), Soil moisture updating by Ensemble Kalman Filtering in real-time flood forecasting, *J. Hydrol.*, *357*(3–4), 228–242, doi:10.1016/j.jhydrol.2008.05.020.

Koren, V., S. Reed, M. Smith, Z. Zhang, and D.-J. Seo (2004), Hydrology laboratory research modeling system (HL-RMS) of the US National Weather Service, *J. Hydrol.*, *291*(3-4), 297–318.

Krajewski, W. F., and J. A. Smith (2002), Radar hydrology: rainfall estimation, *Adv. Water Resour.*, *25*(8–12), 1387–1394, doi:10.1016/S0309-1708(02)00062-3.

Kucherenko, S., M. Rodriguez-Fernandez, C. Pantelides, and N. Shah (2009), Monte Carlo evaluation of derivative-based global sensitivity measures, *Reliab. Eng. Syst. Safe.*, *94*(7), 1135–1148, doi:10.1016/j.ress.2008.05.006.

Kucherenko, S., S. Tarantola, and P. Annoni (2012), Estimation of global sensitivity indices for models with dependent variables, *Comput. Phys. Commun.*, *183*(4), 937 –946, doi:10.1016/j.cpc.2011.12.020.

Kunstmann, H., W. Kinzelbach, and T. Siegfried (2002), Conditional first-order second-moment method and its application to the quantification of uncertainty in groundwater modeling, *Water Resour. Res.*, *38*(4), 1–14, doi:10.1029/2000WR000022.

Lawrence, D. M., K. W. Oleson, M. G. Flanner, P. E. Thornton, S. C. Swenson, P. J. Lawrence, X. Zeng, Z.-L. Yang, S. Levis, K. Sakaguchi, G. B. Bonan, and A. G. Slater (2011), Parameterization improvements and functional and structural advances in Version 4 of the Community Land Model, *J. Adv. Model. Earth Syst.*, *3*(1), 1–27, doi:10.1029/2011MS00045.

Leander, R., A. T. Buishand, and M. J. M. De Wit (2005), Estimation of extreme floods of the River Meuse using a stochastic weather generator and a rainfall-runoff model, *Hydrological Sciences Journal*, *50*(6), 1089–1103, doi:10.1623/hysj.2005.50.6.1089.

Leavesley, G. H., R. W. Lichty, B. M. Troutman, and L. G. Saindon (1983), *Precipitation-runoff modeling system: Users manual*, vol. 83-4238, U.S. Geol. Surv. Water Invest. Rep.

Leblois, E., and J.-D. Creutin (2013), Space-time simulation of intermittent rainfall with prescribed advection field: Adaptation of the turning band method, *Water Resour. Res.*, *49*(6), 3375–3387, doi:10.1002/wrcr.20190.

Lee, H., D.-J. Seo, and V. Koren (2011), Assimilation of streamflow and in situ soil moisture data into operational distributed hydrologic models: Effects of uncertainties in the data and initial model soil moisture states, *Adv. Water Resour.*, *34*, 1597–1615, doi:10.1016/j.advwatres.2011.08.012.

Lee, H., D.-J. Seo, Y. Liu, V. Koren, P. McKee, and R. Corby (2012), Variational assimilation of streamflow into operational distributed hydrologic models: effect of spatiotemporal scale of adjustment, *Hydrol. Earth Syst. Sci.*, *16*, 2233–2251, doi:10.5194/hess-16-2233-2012.

Li, J., Q. Y. Duan, W. Gong, A. Ye, Y. Dai, C. Miao, Z. Di, C. Tong, and Y. Sun (2013a), Assessing parameter importance of the Common Land Model based on qualitative and quantitative sensitivity analysis, *Hydrol. Earth Syst. Sci.*, *17*(8), 3279–3293, doi:10.5194/hess-17-3279-2013.

Li, Y., D. Ryu, A. W. Western, and Q. J. Wang (2013b), Assimilation of stream discharge for flood forecasting: The benefits of accounting for routing time lags, *Water Resour. Res.*, *49*, 1887–1900, doi:10.1002/wrcr.20169.

Liang, D. P. Lettenmaier, E. F. Wood, and S. J. Burges (1994), A simple hydrologically based model of land surface water and energy fluxes for general circulation models, *J. Geophys. Res. Atmos.*, *99*(D7), 14,415–14,428.

Lindström, G., B. Johansson, M. Persson, M. Gardelin, and S. Bergström (1997), Development and test of the distributed HBV-96 hydrological model, *J. Hydrol.*, *201*(1–4), 272–288.

Liu, Y., and H. V. Gupta (2007), Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, *Water Resour. Res.*, *43*(7), 1–19, doi:10.1029/2006WR005756.

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, *Hydrol. Earth Syst. Sci.*, *16*, 3863–3887, doi:10.5194/hess-16-3863-2012.

Lu, D., M. Ye, and M. C. Hill (2012), Analysis of regression confidence intervals and Bayesian credible intervals for uncertainty quantification, *Water Resour. Res.*, *48*, 1–20, doi:10.1029/2011WR011289.

Madsen, H., and C. Skotner (2005), Adaptive state updating in real-time river flow forecasting - A combined filtering and error forecasting procedure, *J. Hydrol.*, *308*(1–4), 302–312, doi:10.1016/j.jhydrol.2004.10.030.

Mandapaka, P., G. Villarini, B.-C. Seo, and W. Krajewski (2010), Effect of radar-rainfall uncertainties on the spatial characterization of rainfall events, *J. Geophys. Res. Atmos.*, *115*(17), D17,110, doi:10.1029/2009JD013366.

Massmann, C., and H. Holzmann (2012), Analysis of the behavior of a rainfall-runoff model using three global sensitivity analysis methods evaluated at different temporal scales, *J. Hydrol.*, *475*, 97–110, doi:10.1016/j.jhydrol.2012.09.026.

McMillan, H., B. Jackson, M. P. Clark, D. Kavetski, and R. Woods (2011), Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models, *J. Hydrol.*, *400*(1–2), 83–94, doi:10.1016/j.jhydrol.2011.01.026.

McMillan, H. K., E. O. Hreinsson, M. P. Clark, S. K. Singh, C. Zammit, and M. J. Uddstrom (2013), Operational hydrological data assimilation with the recursive ensemble Kalman filter, *Hydrol. Earth Syst. Sci.*, *17*(1), 21–38, doi:10.5194/hess-17-21-2013.

Mehl, S. W., and M. C. Hill (2001), A comparison of solute-transport solution techniques and their effect on sensitivity analysis and inverse modeling results, *Groundwater*, *39*(2), 300–307.

Mendoza, P., J. McPhee, and X. Vargas (2012), Uncertainty in flood forecasting: A distributed modeling approach in a sparse data catchment, *Water Resour. Res.*, *48*(9), doi:10.1029/2011WR011089.

Menke, W. (1989), *Geophysical data analysis, Discrete inverse theory*, New York, Academic Press, International Geophysics Series, vol. 45.

Mood, A. M., F. A. Graybill, and D. C. Boes (1974), *Introduction to the theory of statistics*, McGraw-Hill International Editions (Third Edition).

Moore, R. J. (1985), The probability-distributed principle and runoff production at point and basin scales., *Hydrological Sciences Journal/Journal des Sciences Hydrologiques*, *30*(2), 273–297.

Moore, R. J. (2007), The PDM rainfall-runoff model, *Hydrol. Earth Syst. Sci.*, *11*(1), 483–499.

Moradkhani, H. (2008), Hydrologic remote sensing and land surface data assimilation, *Sensors*, *8*(5), 2986–3004.

Moradkhani, H., K. L. Hsu, H. Gupta, and S. Sorooshian (2005a), Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, *Water Resour. Res.*, *41*(5), 1–17.

Moradkhani, H., S. Sorooshian, H. Gupta, and P. R. Houser (2005b), Dual state-parameter estimation of hydrological models using ensemble Kalman filter, *Adv. Water Resour.*, *28*(2), 135–147.

Morris, M. D. (1991), Factorial sampling plans for preliminary computational experiments, *Technometrics*, *33*(2), 161–174.

Musters, P. A. D., and W. Bouten (2000), A method for identifying optimum strategies of measuring soil water contents for calibrating a root water uptake model, *J. Hydrol.*, *227*, 273–286.

Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part I - A discussion of principles, *J. Hydrol.*, *10*(3), 282–290.

Neal, J. C., P. M. Atkinson, and C. W. Hutton (2007), Flood inundation model updating using an ensemble kalman filter and spatially distributed measurements, *J. Hydrol.*, *336*(3–4), 401–415, doi: DOI: 10.1016/j.jhydrol.2007.01.012.

Noh, S., Y. Tachikawa, M. Shiiba, and S. Kim (2011a), Applying sequential Monte Carlo methods into a distributed hydrologic model: Lagged particle filtering approach with regularization, *Hydrol. Earth Syst. Sci.*, *15*(10), 3237–3251, doi:10.5194/hess-15-3237-2011.

Noh, S. J., Y. Tachikawa, M. Shiiba, and S. Kim (2011b), Dual state-parameter updating scheme on a conceptual hydrologic model using sequential Monte Carlo filters, *Journal of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering)*, *67*(4), I1–I6.

Noh, S. J., Y. Tachikawa, M. Shiiba, and S. Kim (2013), Ensemble Kalman filtering and particle filtering in a lag-time window for short-term streamflow forecasting with a distributed hydrologic model, *J. Hydraul. Eng. ASCE*, *18*(12), 1684–1696.

Nossent, J., P. Elsen, and W. Bauwens (2011), Sobol' sensitivity analysis of a complex environmental model, *Environ. Mod. & Soft.*, *26*(12), 1515–1525, doi:10.1016/j.envsoft.2011.08.010.

Oakley, J. E., and A. O'Hagan (2004), Probabilistic sensitivity analysis of complex models: A Bayesian approach, *J. R. Statist. Soc. B*, *66*(3), 751–769, doi:10.1111/j.1467-9868.2004.05304.x.

Oliver, D. S., A. C. Reynolds, and N. Liu (2008), *Inverse theory for petroleum reservoir characterization and history matching*, Cambridge University Press.

OpenDA (2013), The OpenDA data-assimilation toolbox, www.openda.org, (last access: 1 November 2013).

OpenStreams (2013), OpenStreams, www.openstreams.nl, (last access: 1 November 2013).

Papale, D., M. Reichstein, M. Aubinet, E. Canfora, C. Bernhofer, W. Kutsch, B. Longdoz, S. Rambal, R. Valentini, T. Vesala, and D. Yakir (2006), Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: Algorithms and uncertainty estimation, *Biogeosciences*, *3*(4), 571–583, doi:10.5194/bg-3-571-2006.

Pappenberger, F., P. Matgen, K. J. Beven, J. B. Henry, L. Pfister, and P. Fraipont (2006), Influence of uncertain boundary conditions and model structure on flood inundation predictions, *Adv. Water Resour.*, *29*(10), 1430–1449.

Pappenberger, F., K. J. Beven, M. Ratto, and P. Matgen (2008a), Multi-method global sensitivity analysis of flood inundation models, *Adv. Water Resour.*, *31*(1), 1–14, doi: 10.1016/j.advwatres.2007.04.009.

Pappenberger, F., J. Bartholmes, J. Thielen, H. L. Cloke, R. Buizza, and d. A. (2008b), New dimensions in early flood warning across the globe using grand-ensemble weather predictions, *Geophys. Res. Lett*, *35*, doi:10.1029/2008GL033837.

Park, C. K., and K.-I. Ahn (1994), A new approach for measuring uncertainty importance and distributional sensitivity in probabilistic safety assessment, *Reliab. Eng. Syst. Safe.*, *46*(3), 253–261, doi:10.1016/0951-8320(94)90119-8.

Pauwels, V. R. N., and G. J. M. De Lannoy (2006), Improvement of modeled soil wetness conditions and turbulent fluxes through the assimilation of observed discharge, *J. Hydrometeorol.*, *7*(3), 458–477, doi:10.1175/JHM490.1.

Pauwels, V. R. N., and G. J. M. De Lannoy (2009), Ensemble-based assimilation of discharge into rainfall-runoff models: A comparison of approaches to mapping observational information to state space, *Water Resour. Res.*, *45*, 1–17, doi:10.1029/2008WR007590.

Pauwels, V. R. N., R. Hoeben, N. E. C. Verhoest, and F. P. De Troch (2001), The importance of the spatial patterns of remotely sensed soil moisture in the improvement of discharge predictions for small-scale basins through data assimilation, *J. Hydrol.*, *251*(1–2), 88–102, doi:10.1016/S0022-1694(01)00440-1.

Pauwels, V. R. N., R. Hoeben, N. E. Verhoest, F. P. De Troch, and P. A. Troch (2002), Improvement of TOPLATS-based discharge predictions through assimilation of ERS-based remotely sensed soil moisture values, *Hydrol. Processes*, *16*(5), 995–1013, doi:10.1002/hyp.315.

PCRaster (2013), Environmental Modelling language, (last access: 1 November 2013).

Pebesma, E. J. (2004), Multivariable geostatistics in S: the gstat package, *Computers & Geosciences*, *30*, 683–691, doi:10.1016/j.cageo.2004.03.012.

Perraud, J. M., G. M. Podger, J. M. Rahman, and R. A. Vertessy (2003), A new rainfall runoff software library, *Proceedings of MODSIM*, pp. 1733–1738.

Plischke, E., E. Borgonovo, and C. L. Smith (2013), Global sensitivity measures from given data, *Eur. J. Oper. Res.*, *226*(3), 536–550, doi:10.1016/j.ejor.2012.11.047.

R Development Core Team (2013), R: A language and environment for statistical computing, (last access: 1 November 2013).

Rakovec, O., P. Hazenberg, P. J. J. F. Torfs, A. H. Weerts, and R. Uijlenhoet (2012a), Generating spatial precipitation ensembles: impact of temporal correlation structure, *Hydrol. Earth Syst. Sci.*, *16*, 3419–3434, doi:10.5194/hess-16-3419-2012.

Rakovec, O., A. H. Weerts, P. Hazenberg, P. J. J. F. Torfs, and R. Uijlenhoet (2012b), State updating of a distributed hydrological model with Ensemble Kalman Filtering: effects of updating frequency and observation network density on forecast accuracy, *Hydrol. Earth Syst. Sci.*, *16*, 3435–3449, doi:hess-16-3435-2012.

Rakovec, O., M. C. Hill, M. P. Clark, A. H. Weerts, A. J. Teuling, and R. Uijlenhoet (2014), Distributed Evaluation of Local Sensitivity Analysis (DELSA), with application to hydrologic models, *Water Resour. Res.*, *50*, 1–18, doi:10.1002/2013WR014063.

Reggiani, P., M. Sivapalan, and S. M. Hassanizadeh (1998), A unifying framework for watershed thermodynamics: balance equations for mass, momentum, energy and entropy, and the second law of thermodynamics, *Adv. Water Resour.*, *22*(4), 367–398, doi:10.1016/S0309-1708(98)00012-8.

Reichle, R. H. (2008), Data assimilation methods in the Earth sciences, *Adv. Water Resour.*, *31*(11, Sp. Iss. SI), 1411–1418, doi:10.1016/j.advwatres.2008.01.001.

Reichle, R. H., D. B. McLaughlin, and D. Entekhabi (2002), Hydrologic data assimilation with the ensemble Kalman filter, *Mon. Wea. Rev.,*, *130*(1), 103–114.

Reichstein, M., E. Falge, D. Baldocchi, D. Papale, M. Aubinet, P. Berbigier, C. Bernhofer, N. Buchmann, T. Gilmanov, A. Granier, T. Grünwald, K. Havránková, H. Ilvesniemi, D. Janous, A. Knohl, T. Laurila, A. Lohila, D. Loustau, G. Matteucci, T. Meyers, F. Miglietta, J.-M. Ourcival, J. Pumpanen, S. Rambal, E. Rotenberg, M. Sanz, J. Tenhunen, G. Seufert, F. Vaccari, T. Vesala, D. Yakir, and R. Valentini (2005), On the separation of net ecosystem exchange into assimilation and ecosystem respiration: Review and improved algorithm, *Glob. Change Biol.*, *11*(9), 1424–1439, doi:10.1111/j.1365-2486.2005.001002.x.

Renard, B., D. Kavetski, E. Leblois, M. Thyer, G. Kuczera, and S. W. Franks (2011), Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation, *Water Resour. Res.*, *47*(W11516), doi:10.1029/2011WR010643.

Ristic, B., S. Arulampalam, and N. Gordon (2004), *Beyond the Kalman filter: particle filters for tracking applications*, Artech Print on Demand.

Robert, C. P. (2010), *Monte Carlo statistical methods*, Springer.

Rosolem, R., H. V. Gupta, W. J. Shuttleworth, X. Zeng, and L. G. G. De Gonçalves (2012), A fully multiple-criteria implementation of the Sobol' method for parameter sensitivity analysis, *J. Geophys. Res. Atmos.*, *117*(7), D07,103, doi:10.1029/2011JD016355.

Rossiter, D. G. (2007), Technical Note: Co-kriging with the gstat package of the R environment for statistical computing, Version 2.1, Self-published online, (last access: 1 February 2012).

Saavedra Valeriano, O. C., T. Koike, K. Yang, T. Graf, X. Li, L. Wang, and X. Han (2010), Decision support for dam release during floods using a distributed biosphere hydrological model driven by quantitative precipitation forecasts, *Water Resour. Res.*, *46*, doi:10.1029/2010WR009502.

Saito, K., T. Fujita, Y. Yamada, J. Ishida, Y. Kumagai, K. Aranami, S. Ohmori, R. Nagasawa, S. Kumagai, C. Muroi, T. Kato, H. Eito, and Y. Yamazaki (2006), The Operational JMA Nonhydrostatic Mesoscale Model, *Mon. Wea. Rev.*, *134*, 1266–1298.

Sakov, P., G. Evensen, and L. Bertino (2010), Asynchronous data assimilation with the EnKF, *Tellus, Series A: Dynamic Meteorology and Oceanography*, *62*(1), 24–29, doi:10.1111/j.1600-0870.2009.00417.x.

Salamon, P., and L. Feyen (2009), Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter, *J. Hydrol.*, *376*, 428–442.

Salamon, P., and L. Feyen (2010), Disentangling uncertainties in distributed hydrological modeling using multiplicative error models and sequential data assimilation, *Water Resour. Res.*, *46*, doi: 10.1029/2009WR009022.

Saltelli, A. (2002), Making best use of model evaluations to compute sensitivity indices, *Comput. Phys. Commun.*, *145*(2), 280–297, doi:10.1016/S0010-4655(02)00280-1.

Saltelli, A., S. Tarantola, and K. P.-S. Chan (1999), A quantitative model-independent method for global sensitivity analysis of model output, *Technometrics*, *41*(1), 39–56.

Saltelli, A., S. Tarantola, F. Campolongo, and M. Ratto (2004), *Sensitivity analysis in practice*, New York, Wiley.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola (2008), *Global Sensitivity Analysis: The Primer*, Wiley.

Schellekens, J. (2006), CQ-flow: a distributed hydrological model for the prediction of impacts of land-cover change, with special reference to the Rio Chiquito catchment, Northwest Costa Rica, Annex 4, in: Final Technical Report DFID-FRP Project no. R7991 Hydrological impacts of converting tropical montane cloud forest to pasture, with initial reference to northern Costa Rica, edited by Bruijnzeel, L.A.

Schellekens, J., A. H. Weerts, R. J. Moore, C. E. Pierce, and S. Hildon (2011), The use of MOGREPS ensemble rainfall forecasts in operational flood forecasting systems across England and Wales, *Adv. Geosci*, *29*, 77–84.

Schuurmans, J. M. (2008), *Hydrological now- and forecasting*, Ph.D. thesis, Utrecht University, Netherlands.

Schuurmans, J. M., and M. F. P. Bierkens (2007), Effect of spatial distribution of daily rainfall on interior catchment response of a distributed hydrological model, *Hydrol. Earth Syst. Sci.*, *11*(2), 677–693, doi:10.5194/hess-11-677-2007.

Schuurmans, J. M., M. F. P. Bierkens, E. J. Pebesma, and R. Uijlenhoet (2007), Automatic prediction of high-resolution daily rainfall fields for multiple extents: The potential of operational radar, *J. Hydrometeorol.*, *8*(6), 1204–1224, doi:10.1175/2007JHM792.1.

Seber, G. A. F., and C. J. Wild (1989), *Nonlinear regression*, Wiley, New York.

Sene, K. (2008), *Flood Warning, Forecasting and Emergency Response*, Springer.

Seo, D.-J., and J. Breidenbach (2002), Real-time correction of spatially nonuniform bias in radar rainfall data using rain gauge measurements, *J. Hydrometeorol.*, *3*(2), 93–111, doi:10.1175/1525-7541(2002)003¡0093:RTCOSN¿2.0.CO;2.

Seo, D.-J., J. Breidenbach, and E. Johnson (1999), Real-time estimation of mean field bias in radar rainfall data, *J. Hydrol.*, *223*(3-4), 131–147, doi:10.1016/S0022-1694(99)00106-7.

Seo, D.-J., V. Koren, and N. Cajina (2003), Real-time variational assimilation of hydrologic and hydrometeorological data into operational hydrologic forecasting, *J. Hydrometeorol.*, *4*(3), 627–641.

Seo, D.-J., L. Cajina, R. Corby, and T. Howieson (2009), Automatic state updating for operational streamflow forecasting via variational data assimilation, *J. Hydrol.*, *367*, 255–275.

Shiiba, M., X. Laurenson, and Y. Tachikawa (2000), Real-time stage and discharge estimation by a stochastic-dynamic flood routing model, *Hydrol. Processes*, *14*, 481–495.

Simunek, J., M. T. van Genuchten, and M. Sejna (2008), Development and applications of the HYDRUS and STANMOD software packages and related codes, *Vadose Zone J.*, *7*(2), 587–600, doi:10.2136/vzj2007.0077.

Singh, V. P. (1995), *Computer models of watershed hydrology*, Highlands Ranch, CO : Water Resources Publications.

Slater, A. G., and M. P. Clark (2006), Snow data assimilation via an ensemble Kalman filter, *J. Hydrometeorol.*, *7*(3), 478–493, doi:10.1175/JHM505.1.

Smiatek, G., H. Kunstmann, and J. Werhahn (2012), Implementation and performance analysis of a high resolution coupled numerical weather and river runoff prediction model system for an Alpine catchment, *Environ. Mod. & Soft.*, *38*, 231–243.

Smith, P. J., K. J. Beven, A. H. Weerts, and D. Leedal (2012), Adaptive correction of deterministic models to produce probabilistic forecasts, *Hydrol. Earth Syst. Sci.*, *16*, 2783–2799.

Sobol', I. M. (2001), Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Math. Comput. Simulation*, *55*(1–3), 271–280, doi:10.1016/S0378-4754(00)00270-6.

Sobol', I. M., and S. Kucherenko (2009), Derivative based global sensitivity measures and their link with global sensitivity indices, *Math. Comput. Simulat.*, *79*(10), 3009–3017, doi:10.1016/j.matcom.2009.01.023.

Sobol', I. M., and S. Kucherenko (2010), A new derivative based importance criterion for groups of variables and its link with the global sensitivity indices, *Comput. Phys. Commun.*, *181*, 1212–1217, doi:10.1016/j.cpc.2010.03.006.

Steiner, M., J. Smith, S. Burges, C. Alonso, and R. Darden (1999), Effect of bias adjustment and rain gauge data quality control on radar rainfall estimation, *Water Resour. Res.*, *35*(8), 2487–2503.

Strahler, A. N. (1964), *Quantitative geomorphology of drainage basins and channel networks, section 4-II.*, in: Handbook of Applied Hydrology, edited by Chow, V.T., Maidment, D.R. and Mays, L.W. (1988). McGraw-Hill.

Stricker, J. N. M., and P. M. M. Warmerdam (1982), Estimation of the waterbalance in the Hupselse Beek basin over a period of three years and a first effort to simulate the rainfall-runoff process for a complete year., in *Proc. Symp. Hydrolog. Research Basins*, vol. 2, pp. 379–388.

Tang, Y., P. Reed, T. Wagener, and K. van Werkhoven (2007), Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation, *Hydrol. Earth Syst. Sci.*, *11*(2), 793–817, doi:10.5194/hess-11-793-2007.

Tarantola, A. (2005), *Inverse problem theory and methods for model parameter estimation*, Philadelphia, Society for Industrial and Applied Mathematics.

Teuling, A., I. Lehner, J. Kirchner, and S. Seneviratne (2010), Catchments as simple dynamical systems: Experience from a Swiss prealpine catchment, *Water Resour. Res.*, *46*(10), 1–15, doi: 10.1029/2009WR008777.

Thielen, J., J. Bartholmes, M.-H. Ramos, and A. De Roo (2009), The European Flood Alert System - Part 1: Concept and development, *Hydrol. Earth Syst. Sci.*, *13*(2), 125–140, doi:10.5194/hess-13-125-2009.

Thiemann, M., M. Trosset, H. Gupta, and S. Sorooshian (2001), Bayesian recursive parameter estimation for hydrologic models, *Water Resour. Res.*, *37*(10), 2521–2535.

Thirel, G., P. Salamon, P. Burek, and M. Kalas (2011), Assimilation of MODIS snow cover area data in a distributed hydrological model, *Hydrol. Earth Syst. Sci. Discuss.*, *8*(1), 1329–1364, doi: 10.5194/hessd-8-1329-2011.

Tiedeman, C. R., D. M. Ely, M. C. Hill, and G. M. O'Brien (2004), A method for evaluating the importance of system state observations to model predictions, with application to the Death Valley regional groundwater flow system, *Water Resour. Res.*, *40*(12), 1–14, doi:10.1029/2004WR003313.

Tiedeman, C. R., M. C. Hill, F. A. D'Agnese, and C. C. Faunt (2003), Methods for using groundwater model predictions to guide hydrogeologic data collection, with application to the Death Valley regional groundwater flow system, *Water Resour. Res.*, *39*(1), 1–17, doi:10.1029/2001WR001255.

Todini, E. (2007), Hydrological catchment modelling: Past, present and future, *Hydrol. Earth Syst. Sci.*, *11*(1), 468–482.

Tonkin, M. J., C. R. Tiedeman, D. M. Ely, and M. C. Hill (2007), *OPR-PPR, a Computer Program for Assessing Data Importance to Model Predictions Using Linear Statistics*, Reston Virginia, U.S. Geological Survey Techniques and Methods Report TM-6E2.

UN (2013), Extreme flooding must be turning point on disaster response UN official, http://www.un.org/apps/news/story.asp?NewsID=45261#.UnqTgHA3uxp, (last access: 1 November 2013).

van Beek, L. P. H., Y. Wada, and M. F. P. Bierkens (2011), Global monthly water stress: 1. Water balance and water availability, *Water Resour. Res.*, *47*(7), 1–25, doi:10.1029/2010WR009791.

Van Dam, J. C., P. Groenendijk, R. F. A. Hendriks, and J. G. Kroes (2008), Advances of modeling water flow in variably saturated soils with SWAP, *Vadose Zone J.*, *7*(2), 640–653, doi: 10.2136/vzj2007.0060.

van de Beek, C. Z., H. Leijnse, P. J. J. F. Torfs, and R. Uijlenhoet (2011), Climatology of daily rainfall semivariance in The Netherlands, *Hydrol. Earth Syst. Sci.*, *15*(2), 171–183.

van Deursen, W. (2004), Afregelen HBV model Maasstroomgebied, *Tech. rep.*, Rapportage aan RIZA. Carthago Consultancy. Rotterdam, The Netherlands.

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, *J. Hydrol.*, *324*(1–4), 10 –23, doi:10.1016/j.jhydrol.2005.09.008.

van Leeuwen, P. J. (2009), Particle filtering in geophysical systems, *Mon. Wea. Rev.*, *137*, 4089–4114, doi:10.1175/2009MWR2835.1.

Van Werkhoven, K., T. Wagener, P. Reed, and Y. Tang (2008), Characterization of watershed model behavior across a hydroclimatic gradient, *Water Resour. Res.*, *44*(1), 1–16, doi: 10.1029/2007WR006271.

Verbunt, M., M. Zappa, J. Gurtz, and P. Kaufmann (2006), Verification of a coupled hydrometeorological modelling approach for alpine tributaries in the Rhine basin, *J. Hydrol.*, *324*, 224–238.

Verbunt, M., A. Walser, J. Gurtz, A. Montani, and C. Schär (2007), Probabilistic flood forecasting with a limited-area ensemble prediction system: Selected case studies, *J. Hydrometeorol.*, *8*(4), 897–909, doi:10.1175/JHM594.1.

Verworn, A., and U. Haberlandt (2011), Spatial interpolation of hourly rainfall – effect of additional information, variogram inference and storm properties, *Hydrol. Earth Syst. Sci.*, *15*(2), 569–584, doi:10.5194/hess-15-569-2011.

Villarini, G., P. V. Mandapaka, W. F. Krajewski, and R. J. Moore (2008), Rainfall and sampling uncertainties: A rain gauge perspective, *J. Geophys. Res. Atmos.*, *113*(11), doi:10.1029/2007JD009214.

Vischel, T., T. Lebel, S. Massuel, and B. Cappelaere (2009), Conditional simulation schemes of rain fields and their application to rainfall–runoff modeling studies in the Sahel, *J. Hydrol.*, *375*, 273–286.

Vrugt, J., W. Bouten, and A. Weerts (2001), Information content of data for identifying soil hydraulic parameters from outflow experiments, *Soil Sci. Soc. Am. J.*, *65*(1), 19–27.

Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian (2003), A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, *39*(8), 1–16, doi:10.1029/2002WR001642.

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 Resour. Res.*, *41*(1), 1–17, doi:10.1029/2004WR003059.

Vrugt, J. A., H. V. Gupta, B. Ó Nualláin, and W. Bouten (2006), Real-time data assimilation for operational ensemble streamflow forecasting, *J. Hydrometeorol.*, *7*(3), 548–565.

Wackernagel, H. (2003), *Multivariate geostatistics: an introduction with applications*, Springer.

Wagener, T., D. P. Boyle, M. J. Lees, H. S. Wheater, H. V. Gupta, and S. Sorooshian (2001), A framework for development and application of hydrological models, *Hydrol. Earth Syst. Sci.*, *5*(1), 13–26.

Wagener, T., N. McIntyre, M. J. Lees, H. S. Wheater, and H. V. Gupta (2003), Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis, *Hydrol. Processes*, *17*(2), 455–476, doi:10.1002/hyp.1135.

Wagener, T., H. S. Wheater, and H. V. Gupta (2004), *Rainfall-Runoff modelling in Gauged and Ungauged Catchments*, Imperial College Press.

Wang, F., L. Wang, H. Zhou, O. C. Saavedra Valeriano, T. Koike, and W. Li (2012), Ensemble hydrological prediction-based real-time optimization of a multiobjective reservoir during flood season in a semiarid basin with global numerical weather predictions, *Water Resour. Res.*, *48*, doi: 10.1029/2011WR011366.

Webster, R., and M. Oliver (2001), *Geostatistics for environmental scientists*, Statistics in practice, John Wiley & Sons.

Weerts, A. H. (2007), Validation HBV for FewsNL Meuse, *Tech. rep.*, WL—Delft Hydraulics.

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 Resour. Res.*, *42*, 1–17, doi:10.1029/2005WR004093.

Weerts, A. H., J. A. Huisman, and W. Bouten (2001), Information content of time domain reflectometry waveforms, *Water Resour. Res.*, *37*(5), 1291–1299, doi:10.1029/2000WR900381.

Weerts, A. H., G. Y. El Serafy, S. Hummel, J. Dhondia, and H. Gerritsen (2010), Application of generic data assimilation tools (DATools) for flood forecasting purposes, *Computers & Geosciences*, *36*(4), 453–463, doi:10.1016/j.cageo.2009.07.009.

Weerts, A. H., D.-J. Seo, M. Werner, and J. Schaake (2013), Operational Hydrologic Ensemble Forecasting, in *Applied uncertainty analysis for flood risk management*, edited by K. Beven and J. Hall, chap. Operational Hydrologic Ensemble Forecasting, in press.

Werner, M., J. Schellekens, P. Gijsbers, M. van Dijk, O. van den Akker, and K. Heynert (2013), The Delft-FEWS flow forecasting system, *Environ. Mod. & Soft.*, *40*(0), 65–77, doi: http://dx.doi.org/10.1016/j.envsoft.2012.07.010.

Wetterhall, F., F. Pappenberger, L. Alfieri, H. L. Cloke, J. Thielen-del Pozo, S. Balabanova, J. Daňhelka, A. Vogelbacher, P. Salamon, I. Carrasco, A. J. Cabrera-Tordera, M. Corzo-Toscano, M. Garcia-Padilla, R. J. Garcia-Sanchez, C. Ardilouze, S. Jurela, B. Terek, A. Csik, J. Casey, G. Stankunvičius, V. Ceres, E. Sprokkereef, J. Stam, E. Anghel, D. Vladikovic, C. Alionte Eklund, N. Hjerdt, H. Djerv, F. Holmberg, J. Nilsson, K. Nyström, M. Sušnik, M. Hazlinger, and M. Holubecka (2013), HESS Opinions "Forecaster priorities for improving probabilistic flood forecasts", *Hydrol. Earth Syst. Sci.*, *17*(11), 4389–4399, doi:10.5194/hess-17-4389-2013.

Willems, P. (2001), Stochastic description of the rainfall input errors in lumped hydrological models, *Stoch. Environ. Res. Risk Assess.*, *15*(2), 132–152.

WMO (2013), Flood Forecasting Initiative, http://www.wmo.int/pages/prog/hwrp/FloodForecastingInitiative.php, (last access: 1 November 2013).

Xie, X., and D. Zhang (2013), A partitioned update scheme for state-parameter estimation of distributed hydrologic models based on the ensemble Kalman filter, *Water Resour. Res. in press*, doi: 10.1002/2012WR012853.

Xu, C., and G. Z. Gertner (2011), Reliability of global sensitivity indices, *J. Statist. Comput. Simulation*, *81*(12), 1939–1969, doi:10.1080/00949655.2010.509317.

Ye, M., K. F. Pohlmann, J. B. Chapman, G. M. Pohll, and D. M. Reeves (2010), A model-averaging method for assessing groundwater conceptual model uncertainty, *Groundwater*, *48*(5), 716–728, doi:10.1111/j.1745-6584.2009.00633.x.

Zambrano-Bigiarini, M., and R. Rojas (2013), A model-independent Particle Swarm Optimisation software for model calibration, *Environ. Mod. & Soft.*, *43*, 5–25, doi: 10.1016/j.envsoft.2013.01.004.

Zawadzki, I. I. (1973), Errors and fluctuations of raingauge estimates of areal rainfall, *J. Hydrol.*, *18*(3–4), 243–255.

Zhan, C.-S., X.-M. Song, J. Xia, and C. Tong (2013), An efficient integrated approach for global sensitivity analysis of hydrological model parameters, *Environ. Mod. & Soft.*, *41*, 39–52, doi: 10.1016/j.envsoft.2012.10.009.

Zuniga, M. M., S. Kucherenko, and N. Shah (2013), Metamodelling with independent and dependent inputs, *Comput. Phys. Commun.*, *184*(6), 1570–1580, doi:10.1016/j.cpc.2013.02.005.

Zupanski, M. (2005), Maximum Likelihood Ensemble Filter: Theoretical Aspects, *Mon. Wea. Rev.*, *133*, 1710–1726, doi:10.1175/MWR2946.1.

List of publications

Peer-reviewed articles

- Rakovec, O., Hill, M. C., Clark, M. P., Weerts, A. H., Teuling, A. J., Uijlenhoet, R. (2014): Distributed Evaluation of Local Sensitivity Analysis (DELSA), with application to hydrologic models, *Water Resour. Res., 50*, 1-18, doi:10.1002/2013WR014063

- Rakovec, O., Hazenberg, P., Torfs, P. J. J. F., Weerts, A. H., and Uijlenhoet, R. (2012): Generating spatial precipitation ensembles: impact of temporal correlation structure, *Hydrol. Earth Syst. Sci., 16*, 3419-3434, doi:10.5194/hess-16-3419-2012

- Rakovec, O., Weerts, A. H., Hazenberg, P., Torfs, P. J. J. F., and Uijlenhoet, R. (2012): State updating of a distributed hydrological model with Ensemble Kalman Filtering: effects of updating frequency and observation network density on forecast accuracy, *Hydrol. Earth Syst. Sci., 16*, 3435-3449, doi:10.5194/hess-16-3435-2012

- Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A. I. J. M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O., and Restrepo, P. (2012): Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, *Hydrol. Earth Syst. Sci., 16*, 3863-3887 doi:10.5194/hess-16-3863-2012

- Van Huijgevoort, M. H. J., Van Loon, A. F., Rakovec, O., Haddeland, I., Horacek, S., and Van Lanen, H. A. J. (2010) Drought assessment using local and large-scale forcing data in small catchments, in: Global Change: Facing Risks and Threats to Water Resources, edited by: Servat, E., Demuth, S., Dezetter, A., Daniell, T., Ferrari, E., Ijjaali, M., Jabrane, R., Van Lanen, H., and Huang, Y., *IAHS Publ. No. 340*, 77-85.

Acknowledgements

Completion of this thesis would not have been possible without endless encouraging support of Albrecht Weerts and Remko Uijlenhoet. Dear Albrecht and Remko, thank you for sharing your ideas with me and being so patient all the time. I am very delightful for having your trust. It was wonderful four years to have you as supervisors. Next to that I would like to thank to Pieter Hazenberg and Paul Torfs, who were always available for immediate help with any programming or conceptual problem I encountered on daily basis – I really appreciate your advice. My big thanks also belong to other HMW colleagues: Anne, Claudia, Marjolein, Sonja, Lieke, Joris, Hidayat, Bart, Maxi, Manuel, Karl, Henk, Aart, Gerald, Geraldine, Ryan, Henny, Roel and Ton – your presence made my stay in Atlas, Aqua and Lumen very comfortable. Matthijs, díky for the coffee/tea breaks in Czech. Trees, Tamara and Annelies, thank you for your great administrative support! I also partly worked at Deltares in Delft and I am very thankful in particular to Julius, Arno and Jaap for their support with my never ending questions regarding Deltares software configurations. Additionally, I was lucky to meet Seong Jin Noh, whom I want to thank for friendly hydrological discussions and being a great guide on our Korean expedition.

I am also very thankful for the opportunity to spend a beautiful half year at NCAR in Boulder, Colorado, where I was visiting Martyn Clark who introduced me to Mary Hill. Dear Martyn and Mary, I really appreciate your hospitability and numerous intriguing discussions, which often took more time than originally planned, and from which I learned a lot. Thanks to Pablo, Naoki and Nic for everyday life in Boulder, joint lunches, afternoon runs, dinners and snowshoeing trips.

In my research I had the luxury of all data being collected before I even started. I would like to express my thanks to the Hydrological Service of the Walloon Region of Belgium (MET-SETHY), the Royal Meteorological Institute of Belgium (KMI) and the University of Liège for providing the hydrological and meteorological data.

I really can't imagine living in Wageningen without my second family, my best friends whom I got to know mostly during my Erasmus student time: Klára, Alba, Majo, Chris, Indira, Luděk, Guillaume, Victoria, Surya, Pabitra, Jelmer, Suzanne, Durk, Nick, Frederic, Mark, ... Thanks for the true friendship, inspiration, frequent dinners (not only with halušky and becherovka) and your never ending visits to Wageningen.

Nakonec chci poděkovat mojí celé rozvětvené rodině. Především děkuji mým rodičům, bratrovi Tomášovi a sestře Janě za veškerou podporu během celého studia. Daně děkuji za vše a za Neklana.

Netherlands Research School for the
Socio-Economic and Natural Sciences of the Environment

C E R T I F I C A T E

The Netherlands Research School for the
Socio-Economic and Natural Sciences of the Environment
(SENSE), declares that

Oldřich Rakovec

born on 1 February 1984 in Prague, Czechoslovakia

has successfully fulfilled all requirements of the
Educational Programme of SENSE.

Wageningen, 25 April 2014

the Chairman of the SENSE board	the SENSE Director of Education
Prof.dr.ir. Huub Rijnaarts	Dr. Ad van Dommelen

The SENSE Research School has been accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW)

KONINKLIJKE NEDERLANDSE
AKADEMIE VAN WETENSCHAPPEN

The SENSE Research School declares that **Mr. Oldřich Rakovec** has successfully fulfilled all requirements of the Educational PhD Programme of SENSE with a work load of 47 ECTS, including the following activities:

SENSE PhD Courses

- Environmental Research in Context
- Research Context Activity: Co-organizing WIMEK/SENSE Symposium on Water & Energy Cycles at Multiple Scales, Wageningen, 1 March 2012
- Uncertainty Modelling and Analysis
- Bayesian Statistics

Other PhD Courses

- Summer school on data assimilation and its applications
- Techniques for Writing and Presenting a Scientific Paper

Management and Didactic Skills Training

- Supervision BSc thesis Exploratory analysis of observed groundwater time series from the Belgian Ardennes
- Website management of the Hydrology and Quantitative Water Management Group

Oral Presentations

- *Quantification of input uncertainty within a spatially distributed conceptual grid-based model using streamflow data assimilation.* EGU Leonardo, Topical Conference, 10-12 November 2010, Luxembourg, Luxembourg
- *Quantification of input uncertainty within a spatially distributed conceptual grid-based model using streamflow data assimilation.* International Workshop on Data Assimilation for Operational Hydrologic Forecasting , 1-3 November 2010, Delft, The Netherlands
- *Ensemble Kalman Filter-based discharge observation sensitivity within a spatially distributed hydrological model.* EGU General Assembly, 4-8 April 2011, Vienna, Austria
- *On the generation of spatial precipitation ensemble: the impact of including temporal correlation structures.* EGU General Assembly, 23-27 April 2012, Vienna, Austria
- *State updating of a distributed hydrological model with Ensemble Kalman Filtering: Effects of updating frequency and observation network density on forecast accuracy.* International Workshop on Data Assimilation for Operational Hydrologic Forecasting, 10-12 September 2012, Seoul, Korea

SENSE Coordinator PhD Education

Dr. ing. Monique Gulickx