

Ensemble-based uncertainty quantification and reduction in hydrological modelling and predictions

Aynom Tesfay Teweldebrhan



Thesis submitted for the degree of Philosophiae Doctor (PhD)

Department of Geosciences
Faculty of Mathematics and Natural Sciences
University of Oslo
Oslo, Norway

2019

© **Aynom Tesfay Teweldebrhan, 2019**

*Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No.2191*

ISSN 1501-7710

All rights reserved. No part of this publication may be
reproduced or transmitted, in any form or by any means, without permission.

Cover: Hanne Baadsgaard Utigard.
Print production: Representralen, University of Oslo.

Abstract

In the contemporary scene, many environmental and water resources related decisions rely upon wide range of modelling results. However, estimates of the model parameters and model predictions are generally imperfect due to the inherent uncertainties emanating from different sources. Our incomplete perception of the natural system, accompanied by the simplifications of important process representations during the development of a conceptual model as well as limitations in measurements are some of the factors that contribute to the various sources of uncertainty. This challenge is further aggravated by the lack of efficient methods to extract the available information from the noisy dataset during model calibration. Thus, uncertainty analysis should be integrated as part of the modelling process from the very beginning; and any rational decision is expected to take into account the prediction uncertainty emanating from the various sources. A comprehensive approach for addressing uncertainty in environmental modelling requires the understanding of potential sources as well as the quantification and reduction of uncertainty at various stages of the modelling process.

This work was focused on ensemble-based uncertainty quantification and reduction in hydrological modelling and predictions, while accompanied by the proper understanding of the different sources and types of uncertainty. The assessment of modelling and prediction uncertainty for a distributed conceptual hydrological model was realized using two variants of the GLUE methodology and with due consideration to the model's main application as an operational hydrological model. Model realizations were evaluated using single and multi-criteria objective functions. Uncertainty of the model parameters was assessed and the value of remote sensing snow cover data in constraining the rainfall runoff model parameters was evaluated. The main outcome of this study was the novel approach introduced to adapt the rejectionist framework for use in the identification of behavioural model realizations in continuous rainfall-runoff modelling.

In order to alleviate the heavy computational burden associated with use of the GLUE methodology in computationally expensive hydrological models, three machine learning models (MLMs) were coupled with the limits of acceptability approach. The coupled MLMs and time-relaxed limits of acceptability approach were able to effectively identify behavioural parameter sets for the hydrological model while significantly minimizing the computational time of the Monte Carlo (MC) simulation. In this study, persistency of the model realizations in reproducing the observations within the error bounds (pLoA) was introduced as a standalone objective-function and was able to effectively identify the behavioural simulations. Although two of the MLMs were not commonly used as emulators of the MC simulation in previous studies, they have yielded comparable results to the standard MLM based emulator.

Parameter sensitivity analysis was conducted using the in-built algorithms of the three MLMs, the residual-based GLUE methodology, and the regional sensitivity analysis approach. Consistent results were obtained from the different methods. The sensitivity analyses results have complemented the process of behavioural model identification by providing better insight on how the variability in the uncertain model parameters affects the variability in the likelihood values.

Reducing the prediction uncertainty of snow water equivalent (SWE), and thereby to get an improved estimate of SWE during the maximum accumulation period was among the main goals of this study. This was accomplished by introducing new fuzzy logic based efficient snow data assimilation schemes that have improved capability to extract the information content of the assimilated remote sensing snow cover data (fSCA). Two ensemble-based data assimilation (DA) schemes, i.e. particle batch smoother and another based on the limits of acceptability concept were employed. Newly introduced versions of these schemes that account for the variability in informational value of the assimilated fSCA observation were also used to reanalyze the model results. Introducing the limits of acceptability approach as a DA scheme yielded an encouraging result. Incorporating the concept of variable informational value of the remote sensing data in both DA schemes was a viable option for an improved estimates of the perturbation parameters, and thereby the reanalyzed SWE values. Results from the analyses under the premise of variable informational value of fSCA with time revealed that the observations do not carry equal information amount in constraining the perturbation parameters. Some observations were more important than others depending on their timing with respect to certain critical points in the melt season, i.e., the points where the mean snow cover changes and the start of a melt-out period. The parametric (likelihood-based) and non-parametric change point detection schemes employed in this study were able to effectively locate the critical points in each grid-cell for multiple analysis years. Although this study was focused on fSCA assimilation, the fuzzy logic based ensemble frameworks introduced in this study can be applied to assimilate other observations that display variable informational value with time.

Acknowledgment

First, I wish to thank my supervisors John Burkhart and Thomas Schuler for giving me the opportunity to be part of their research group and for their encouragement and support during the PhD period.

I would also like to thank Chong-Yu Xu and Morten Hjorth-Jensen for their support and valuable feedback on my draft manuscripts.

I would like to thank Lena Tallaksen for her encouragement and for the useful information about the summer school on uncertainty analysis at Uppsala University which was held for its last time in 2016 after ten consecutive years. That course has introduced me to wide aspects of uncertainty in environmental modelling, but most importantly to one of the biggest names in the field, Keith Beven. I am grateful to Keith Beven both for his dedication in conveying his knowledge to his students and the extended guidance and comments I received in my first paper.

The funding of this PhD work was generously provided by the research council of Norway (NFR) through the ESCYMO project. I gratefully acknowledge for their support.

I would like to express my gratitude to Gaute Lappegard, Knut Sand, Sigbjorn Helset, and all former colleagues at Statkraft AS for providing me most of the hydro-meteorological data used in this work and for the technical support when setting up their newly developed hydrological model in a windows environment. I would also like to thank Frano Cetinic (Globesar AS) for providing me the satellite snow measurement scenes.

I am grateful for all my colleagues at the Department of Geosciences for their co-operation, valuable information, and for the fruitful discussions we had during conferences and lunch times. Thanks also to my office mates and all others for creating such a nice working environment.

Finally, I would like to thank my family and friends for their understanding, patience, and encouragement.

Contents

- 1. Introduction 1**
 - 1.1 Motivation 1
 - 1.2 Goals and objectives 4
 - 1.3 Organization of the thesis 5

- 2. Scientific background 7**
 - 2.1 Brief overview of the modelling process in hydrology 7
 - 2.2 Sensitivity and uncertainty analysis in hydrological modelling 10
 - 2.2.1 Parameter sampling techniques 10
 - 2.2.2 Likelihood measures used in evaluation of hydrological models 11
 - 2.2.3 Information content of data and its implication in formulating a likelihood measure 13
 - 2.2.4 Change point detection methods 15
 - 2.2.5 Parameter sensitivity analysis 20
 - 2.2.6 Sources and nature of uncertainty 24
 - 2.2.7 Uncertainty analysis methodologies 27
 - 2.2.8 Ensemble modelling and multi-objective evaluation 28
 - 2.2.9 Philosophical aspects in testing hydrological models as hypotheses 29
 - 2.3 Surrogate modelling in hydrology 31
 - 2.3.1 Response surface surrogate models 32
 - 2.3.2 Machine learning based emulators 33
 - 2.4 Snow distribution modelling and snow data assimilation into hydrological models 41
 - 2.4.1 Main processes and controlling factors in snow distribution 41
 - 2.4.2 Snow distribution modelling approaches 45
 - 2.4.3 Ensemble-based snow data assimilation techniques 47

3. The hydrological model and data	51
3.1 The Shyft hydrological modelling toolbox.....	51
3.2 An overview of the PT_GS_K model.....	53
3.3 Study area and data.....	56
4. Summary of publications.....	59
4.1 Paper I: Parameter uncertainty analysis for an operational hydrological model using residual-based and limits of acceptability approaches.....	59
4.2 Paper II: Coupled machine learning and the limits of acceptability approach applied in parameter identification for a distributed hydrological model	61
4.3 Paper III: Improving the informational value of MODIS fractional snow cover area using fuzzy logic based ensemble smoother data assimilation frameworks.....	63
5. General discussion.....	67
5.1 Understanding uncertainty.....	67
5.2 Uncertainty quantification	68
5.3 Uncertainty reduction	71
6. Conclusions and recommendations	75
6.1 Conclusions	75
6.2 Recommendations for further work.....	77
7. References	79
8. Journal publications.....	95

Chapter 1

Introduction

1.1 Motivation

Hydrologic predictions are needed for a wide range of applications including for planning, design, and management of environmental and water resources. Streamflow estimates for water supply designs by water systems builders have been needed since the days of the ancient hydraulic societies of China and the Middle East (Schulze, 1995). The increased demand for such estimates, on the one hand, and the limited availability of sufficient streamflow measurements, on the other hand, has initiated the development of indirect methods for estimating the required streamflow characteristics. The development of hydrological models for use in synthetic streamflow simulation, including the generally recognized first hydrological model, i.e. the Rational Method proposed by Mulvaney (1850) and the unit hydrograph introduced by Sherman (1932), has been mainly motivated by such practical engineering problems (Todini, 2011). A further demand for modelling the complex interaction between the soil-surface and runoff generating mechanisms has led to the development of conceptual models following the advent of digital computers. In the contemporary scene, the role of hydrological models has extended into a multi-purpose and multi-faceted tool reflecting the environmental awareness with regard to man's impact on water related issues (e.g. Schulze, 1995).

Hydrological modelling concepts range from simple conceptual models to complex physically based ones. These models differ on their level of process description as well as their computational resource, and data requirements. The process based models tend to resolve the mass flux between the different components of the hydrological system. Therefore, they are computationally expensive and often restricted to analyzing spatially and temporally limited domains (Liston et al., 2007; Mott and Lehning, 2010). On the other hand, the oversimplified representation of important components and processes of the hydrologic system by certain lumped and semi-distributed conceptual models in order to allow multiple years and catchment scale simulation are subjected to additional sources of uncertainty and may not be applicable in other domains due to lack of transferability in space (Clark et al., 2011).

Conceptual hydrological models typically have one or more calibration parameters and commonly require some form of inverse modelling to estimate model parameters from

observations (Crawford and Linsley, 1966). During calibration, equifinality arises when different parameter sets give equally good results in terms of predefined efficiency criteria (Beven, 1993; Savenije, 2001; Wagener et al., 2003). Since all models of hydrological systems are highly simplified representations of reality (e.g., Reichert and Omlin, 1997), it is expected to have several different model structures and parameter sets that describe the system in an adequate way (Blazkova and Beven, 2002). Previous studies have also reported that the classic hydrological approach of using a single set of model parameters may lead to large predictive biases when dealing with nonlinear systems, (e.g., Mantovan and Todini, 2006). It is thus more realistic to use multiple plausible realizations from one or more models to get an ensemble of predictions rather than trying to find an optimal parameter set (Peel and Blöschl, 2011).

The estimates of model parameters through inverse modelling and the resulting predictions are generally imperfect due to the inherent uncertainties, for example in forcing data, model parameters, and model structure. Uncertainty is a situation of inadequate information (Funtowicz and Ravetz, 1990). An increased awareness of these modeling uncertainties and the need for quality control of such models requires the integration of uncertainty analysis into the modeling process from the very beginning (Beven, 1989; Saltelli et al., 2006; Refsgaard et al., 2007). The understanding (e.g. the different sources), quantification, and reduction of uncertainty are the three important aspects in hydrological modelling and prediction (Liu and Gupta, 2007). In the past, various uncertainty analysis techniques have been proposed to infer model parameter values from observations. Of these techniques, the generalized likelihood uncertainty estimation (GLUE) methodology (Beven and Binley, 1992) is the most widely used uncertainty analysis framework in hydrology (Stedinger et al., 2008; Xiong et al., 2008; Shen et al., 2012). This methodology relies on ensemble modelling to assess modelling and prediction uncertainty. Ensemble based techniques allow for the incorporation of a multi-objective evaluation within the calibration procedure to provide a number of alternative parameter sets while reducing parameter uncertainty through incorporating new knowledge from available observations (Efstratiadis and Koutsoyiannis, 2010).

The main challenge with use of ensemble-based uncertainty analysis techniques in general, and the GLUE methodology in particular, is the heavy computational burden (Yu et al., 2015). The computational cost becomes even higher and poses strong limitations in terms of practical implementation and computational requirements when such methods are applied to computationally intensive simulation models (e.g. Castelletti et al., 2012). Monte Carlo (MC) simulation is an ensemble method which is commonly used to quantify the uncertainty propagated from model parameters to predictions in most of the uncertainty analysis frameworks including the GLUE methodology. In the past, surrogate modelling methods have been coupled with the MC simulation in order to make better use of the available, but usually limited, computational budget (Razavi et al., 2012). A surrogate model is a low-order and computationally efficient model identified from the original large model and then used to replace it for computationally intensive applications (Castelletti et al., 2012). Machine learning and other surrogate models have been reported to efficiently emulate the MC simulations after trained with a small sample size (e.g. Yu et al., 2015).

Improved estimates of snow depth and its spatiotemporal distribution are important for various applications including for flood and avalanche warning as well as energy production planning (Farinotti et al., 2010; Udnæs et al., 2007; Melvold and Skaugen, 2013). Snow has important relevance for society spanning from environmental and economical to recreational and aesthetic values. Snow distribution studies are especially important in countries like Norway where approximately 30% of the annual precipitation falls as snow with strong spatial variability (Saloranta, 2012; Dyrddal et al., 2013). In snow dominated catchments, a significant part of the streamflow in late spring and summer is derived from meltwater of snow storage. As a result, the accuracy of streamflow forecasts during this period depends mainly on the accuracy of snow storage estimate at the onset of the forecast period. Water managers in hydropower and water supply sectors consider snow storage in mountain areas as a natural reservoir and they recognize the need for accurate prediction of this important resource and the subsequent impact on downstream flows (Maurer et al., 2007). However, the spatial distribution of snow, especially in alpine settings, is characterized by high heterogeneity (e.g. Winstral and Marks, 2014); and this phenomenon is expected to be more pronounced in the coming decades at the verge of increased climate change impact. According to Dyrddal et al. (2013), while more regions in Norway are expected to experience declining maximum annual snow depth, some inland and higher mountain regions may accumulate more snow in the coming decades. It is at this juncture of increased snow distribution variability as a result of climate change and anthropogenic effects, and the need for accurate prediction of this important resource, where the necessity for hydrologic models and snow re-analysis schemes capable of capturing the evolving phenomena becomes more significant. As part of the effort to address that necessity, Statkraft (2018) has recently released a hydrological modelling framework with multiple hypotheses for estimating snow storage.

The continual retrieval of different kinds of snow observations both from remote sensing and in situ measurements serves as a driving factor to continue from quantification to the reduction of uncertainty by assimilating the new observations into hydrological models (e.g. Liu and Gupta, 2007). Remote sensing data in general and optical fractional snow cover area (fSCA) in particular have been increasingly used in hydrologic modeling for constraining model parameters and states through inverse modeling and various data assimilation schemes (e.g. Clark et al., 2006). This can be attributed to the readily available observations of this data for large areas and at a relatively high temporal resolution (Parajka and Blöschl, 2006). Data assimilation (DA) is the means to obtain an estimate of the true state through use of independent observations and the prior knowledge (model state) with appropriate uncertainty modelling (Liu et al., 2012; Griessinger et al., 2016). There is a well-established understanding that the information content of the data and the efficiency with which that information is extracted are more important than the amount of the data (e.g. Gupta and Sorooshian, 1985; Vrugt et al., 2002). However, there still exists a limitation in the availability of data assimilation schemes that accommodate this concept. According to Liu and Gupta (2007), while the hydrologic science has made significant advances in the availability of data and improved hydrological models, there exists an urgent need for

schemes that can efficiently assimilate important information from the data into the models to produce improved hydrological predictions.

1.2 Goals and objectives

In an effort to address the main research problems presented in the preceding section, this work was undertaken with three main goals. The first goal was to conduct modelling and prediction uncertainty analysis for the PT_GS_K conceptual hydrological model (method stack) within the recently developed Statkraft Hydrological Forecasting Toolbox (Shyft, <https://github.com/statkraft/shyft>). This study was conducted as part and parcel of the modelling process and with due consideration to the model's main application as an operational hydrological model. The specific objectives were:

- to assess parameter uncertainty for the PT_GS_K model using the GLUE methodology and evaluate its performances in light of its intended use
- to implement and validate a new time-relaxed GLUE LoA approach for constraining model parameters
- to investigate the potential value of MODIS snow cover data as additional observation in multi-criteria model conditioning for reducing parameter uncertainty

Like other MC simulation based uncertainty analysis techniques, use of the GLUE methodology in computationally expensive hydrological models might be restricted due to heavy computational burden. Thus, the second goal of this work was to emulate the time consuming MC simulation for parameter identification through coupling of machine learning models (MLMs) with the time-relaxed limits of acceptability approach. The specific objectives were:

- to assess the possibility of using the percentage of model predictions falling within the observation error bounds, i.e. pLoA as a likelihood measure for identification of behavioural models using the coupled MLMs and the limits of acceptability approach, instead of the previously used residual-based likelihood measures.
- to assess the application of two MLMs as emulators of the MC simulation and compare their relative performances in relation to a standard machine learning based emulator and a direct estimate of the response surface from the MC simulation.
- to compare the performance of the MLMs trained using pLoA against those trained using the absolute bias based criterion, as target variables, in assessing the relative influence of the model parameters on the simulation result.

As a continuation to the uncertainty quantification and as an important aspect of uncertainty in hydrological modelling and predictions, the reduction of uncertainty by assimilating new observations was also part of this work. The third goal of this work was thus to reduce the uncertainty in selected model forcing and parameters, and thereby to get an improved estimate of snow water equivalent (SWE) during the maximum accumulation period. This goal was accomplished by introducing fuzzy logic based efficient snow data assimilation schemes for improved extraction of the information content of the assimilated

remote sensing snow cover data (fSCA). The study was conducted with the following specific objectives:

- to assess whether the assumption of variable informational value of fSCA observations depending on their location in different timing windows during the ablation period can reduce the uncertainty in SWE estimates. A novel approach was introduced that accommodates this concept by incorporating the fuzzy logic concept in two ensemble-based SWE reanalysis schemes.
- to evaluate the viability of statistical change point detection methods to locate the critical points that constitute the timing windows.
- to adopt the limits of acceptability concept as an ensemble smoother data assimilation scheme and compare effect of the assumed variability in informational value of fSCA observations on relative performance of this scheme and the particle batch smoother.

1.3 Organization of the thesis

This thesis is organized as follows: following the introductory part presented in this chapter, Chapter-2 presents a scientific review on relevant subjects including sensitivity and uncertainty analysis in hydrological modelling, the concept and application of surrogate models in hydrology as well as snow distribution modelling and snow data assimilation into hydrological models. Chapter-3 introduces the hydrological model, the study domain as well as the data used to validate the different hypotheses underlying the specific objectives outlined in the preceding section. Chapter-4 provides an overview of the methodologies and results presented in the papers published during the study period. Chapter-5 discusses the general implication of the results obtained in this work in relation to the methods and dataset as well as in relation to relevant previous studies. Chapter-6 provides the general conclusions drawn based on the work presented in the thesis and gives some insights about future work. The bibliography is provided following the general conclusion; and presentation of the entire research work is completed with the attached papers and appendices at end of this thesis.

Chapter 2

Scientific background

2.1 Brief overview of the modelling process in hydrology

Hydrological models can be described as collections of physical laws and empirical observations that are written in mathematical terms and combined in such a way that they can produce a set of results based on a set of known and/or assumed conditions (Haan, 1994). As such, hydrological models provide a quantitative expression of the observation, analysis, and prediction of the time-variant interactions of various hydrological processes (Fig. 2.1) for use in various applications (Schulze, 1995). The application of these models is highly dependent on the purpose for which they are developed (Singh and Woolhiser, 2002; Moradkhani and Sorooshian, 2009; Pechlivanidis et al., 2011). Many hydrological models are developed for research purposes aimed at improving the fundamental understanding of the processes governing the real world hydrological system and assessing the impact of change (e.g. climate and land cover change) on water resources. Other types are developed as tools for the extrapolation of hydrological variables both in space and time in order to help decision makers in taking the most effective decision with due consideration to the interaction with the other components of the real world system such as the socio-economic aspects.

Hydrological models are classified in various ways based on the modeller's perception of a hydrological system which is gained through the interaction with the system, observation and experiments (Beven, 2011). According to their description of the physical processes, hydrological models can be classified as physically-based (white-box), conceptual (grey-box) and stochastic (black-box) models; and based on their spatial description of catchment processes they can be classified as lumped or distributed (Refsgaard, 1996). In lumped models, the entire catchment is considered as a single homogenous unit, while a distributed model takes into account for the spatial variability of model forcings and parameters with explicit characterization of the processes and patterns (e.g. Moradkhani and Sorooshian, 2009).

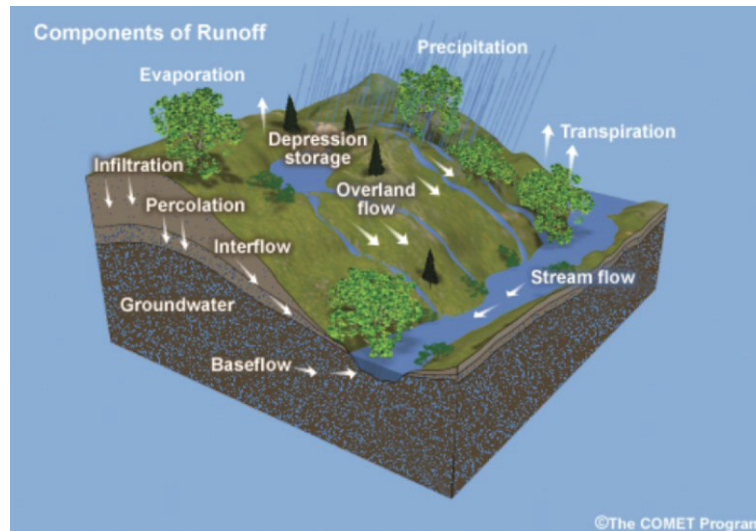


Figure 2.1: Basic components of the hydrological system (source: The COMET Program (2005))

Although, the modelling process has been formulated to varying degree of detail by different researchers, the main steps can be summarized as shown in Figure 2.2 (e.g. Refsgaard, 1997; Beven, 2011; Gupta et al., 2012). After defining the modelling purpose, the perceptual model of the hydrological processes of interest in the study domain is developed and that reflects the modeller's perception on how the catchment responds to inputs to the system, such as precipitation. As such, a perceptual model depends on the modeller's prior knowledge about the observations, components and processes in the system and it is expected to differ from that of another.

The next step in the modelling process deals with the establishment of a conceptual model and this comprises the modeller's perception of the main hydrological processes in the study domain and the corresponding simplifications which are assumed to be acceptable for the intended use of the mathematical model. A conceptual model attempts to quantitatively define the states, fluxes, and parameters of the system and the linkage between them. Similar to the perceptual model, the conceptual model depends upon the modeller's process understanding about the system (Gupta et al., 2012).

The complexity of the conceptual model may vary from a model involving simple mass balance equations for components representing storage in the system to coupled-nonlinear partial differential equations. In the latter case, an additional stage in the modelling process is required to define a procedural model in the form of a computer code (e.g. Beven, 2011). Details of the procedural model include: the selection of a numerical formulation for representing the spatial relationships (e.g. finite difference or finite element), definition of the spatial resolution for computations within the model domain, and the selection of a procedure for time integration of the governing equations (Gupta et al., 2012). It should, however, be taken into consideration that a significant interaction may exist between the choice of model equations and the choice of the spatial and temporal scales since the decision to explicitly resolve a given process will dictate the threshold spatial and temporal resolution of computation (Clark et al., 2011).

Although usually considered as part of the procedural model development phase, code verification, i.e. debugging the computer program is also an important component of the modelling process. This process is used to substantiate the correctness of the computer algorithm and its implementation (Sargent, 1991). The procedural model also needs to pass through the stage of model calibration before it is applied to make quantitative predictions (Beven, 2011). The formulations of the hydrological model involve different inputs, model parameters and state variables. Some of the model inputs such as those that define the geometric characteristics of the study domain are static that remain constant throughout the simulation period, while others such as the meteorological variables are dynamic during the simulation period. The state variables such as soil water and snow storage also tend to vary with time. The model parameters are usually assumed transferrable in time and thus remain constant during the simulation period. Since it is difficult to set the parameter value a priori for a given study domain, parameter values are conditioned through use of observational dataset (e.g. streamflow) during the calibration process. However, it may be challenging to identify an optimal parameter set from different feasible parameter sets due to the various factors discussed in the next section.

Once the model parameter values are identified, the next step is the validation of model predictions through calculation of efficiency metrics in relation to independent observations. The validation phase can be considered as the process of establishing the soundness and completeness of the model structure, i.e. a model structure based on valid reasoning that includes all the components relevant for the intended purpose (Nguyen and Kok, 2007). Furthermore, the model behaviour should not contradict the general physical laws while yielding a computed behaviour with an acceptable agreement to the observations (e.g. Parker et al., 2002; Jakeman et al., 2006). Although this section presents the modelling process in successive steps, it is more iterative rather than procedural, where the modeller can go back to a preceding step and make changes based on updated information.

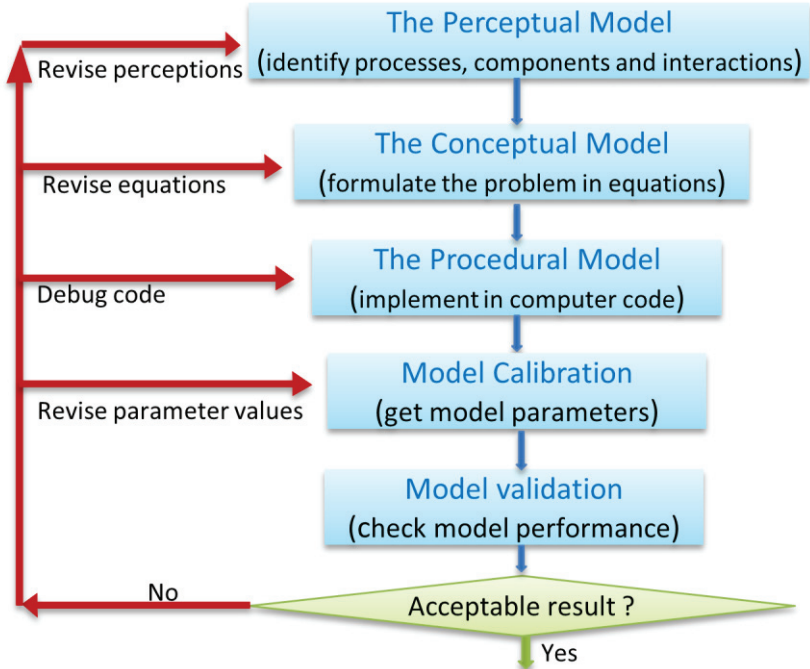


Figure 2.2: A schematic outline of the modelling process (after Beven, 2011)

2.2 Sensitivity and uncertainty analysis in hydrological modelling

Sensitivity and uncertainty analysis are important steps in the modelling process (e.g. Campolongo et al., 2007). As mentioned in the previous section, a possible use of models is to map a set of known or assumed conditions into inferences. A good modelling practice considers alternative conditions with subsequent identification of their corresponding interval of inferences, rather than mapping a single set of conditions into a single inference (Saltelli et al., 2006). This helps to present a model-based inference in the form of an empirical distribution of the predicted variable and the production of this distribution is commonly known as uncertainty analysis. The process of linking the uncertainty in the inference to the uncertainty in the conditions is known as sensitivity analysis. As such, uncertainty and sensitivity analyses are commonly run together. This section presents topics related to sensitivity and uncertainty analysis with main focus given to uncertainty quantification in hydrological modelling.

2.2.1 Parameter sampling techniques

The application of most sensitivity and uncertainty analysis techniques requires parameter sampling from a prior distribution. There are various sampling techniques that have been applied in the literature. In sensitivity analysis a distinction is usually made between One-At-a-Time (OAT) and All-At-a-Time (AAT) sampling methods based on the approach followed in selecting the parameter values from their respective dimensions (e.g. Pianosi et al, 2016). In OAT methods, only one parameter is allowed to vary, while keeping the value of all other parameters fixed, while in the AAT method the output variations are induced by simultaneously varying all parameter values. Following are some of the AAT-based sampling techniques that are commonly used to define the shape of a response surface (e.g. Beven, 2009; Beven, 2011; Yang et al, 2018).

(a) Simple Monte Carlo sampling

This method involves an independent sampling of random values of each parameter across their specified ranges. The main problem with this sampling strategy is that of taking enough samples since areas of higher likelihood on the surface might be missed. When using this method, usually very large numbers of samples are required to define the shape of local areas of high likelihood. A parameter sample as represented by the matrix, M (Eq. 2.1) is used to produce an output vector for the target variable of interest through running the model using each row of the matrix M (e.g. Saltelli et al, 2008). For example, $y^{(1)}$ is the value obtained by running the hydrological model with the parameter values given by the row vector, $M_1 = (x_1^{(1)}, x_2^{(1)}, \dots, x_r^{(1)})$.

$$\mathbf{M} = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_r^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_r^{(2)} \\ \dots & \dots & \dots & \dots \\ x_1^{(N-1)} & x_2^{(N-1)} & \dots & x_r^{(N-1)} \\ x_1^{(N)} & x_2^{(N)} & \dots & x_r^{(N)} \end{bmatrix} \longrightarrow \mathbf{Y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \dots \\ y^{(N-1)} \\ y^{(N)} \end{bmatrix} \quad (2.1)$$

(b) Monte Carlo Markov Chain sampling

The Monte Carlo Markov Chain (MCMC) sampling technique is a form of importance sampling which is commonly used in hydrological modelling. This method involves running a model using parameter sets that are randomly sampled in accordance to the prior distribution with subsequent estimation of the posterior likelihood. A new set is chosen following the prior distribution but whether the model is run with that parameter set depends on the likelihood of the previous sample and a random number. Even if the likelihood of the previous point was low, there is a probability of making a new run around it in order to avoid missing regions of the space where a new high likelihood area might be found. As such, the process involves a chain of random walks across the likelihood surface where the probability of choosing a new sample depends on the knowledge of the local likelihood values; and with the possibility of starting a new chain in areas that have not been sampled before. The Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970) and the Gibbs sampler (Geman and Geman, 1987) are the common forms of the MCMC method. Although, such methods may considerably save the time for finding the likelihood surface in the presence of a well-defined surface, they may not work well for surfaces with complex shapes (Beven, 2011).

2.2.2 Likelihood measures used in evaluation of hydrological models

Different likelihood measures can be used to evaluate the results of a hydrological model simulation depending on various factors. Some of these factors include: type of the observational data used for model evaluation (e.g. streamflow or fractional snow cover area), nature of the prediction problem (e.g. high-flow and low-flow) and a prior belief on shape of the residual distribution (e.g. Beven, 2011). The commonly used likelihood measures in hydrological modelling can be categorized as statistical and informal likelihoods.

(a) Statistical likelihood measures

Statistical likelihood measures are based on the nature of the model residuals under the assumption that all residual errors vary statistically. The most commonly used formal likelihood measure assumes the Gaussian distribution with the contribution of a single residual to the likelihood as being proportional to the square of its value. This likelihood function is based on the assumption that the model residuals (ϵ_t) are normal and independently distributed (Stedinger et al., 2008). The likelihood function of a Gaussian error with N residuals takes the form (Eq. 2.2):

$$L(\epsilon|M(\theta, I)) = (2\pi\sigma_\epsilon^2)^{-N/2} \exp\left[-\frac{1}{2\sigma_\epsilon^2}\left(\sum_{t=1}^N \epsilon_t^2\right)\right] \quad (2.2)$$

where, $L(\epsilon|M(\theta, I))$ is the likelihood (L) of the model residuals (ϵ) given predictions of the model with parameter set θ and inputs I . σ_ϵ^2 and t respectively denote variance of the residuals and a simulation time step.

(b) Informal likelihood measures

Informal likelihood measures do not assume a statistical model for model residuals and are commonly used in the GLUE methodology. The most commonly used informal likelihood measure in hydrology is the Nash-Sutcliffe efficiency (NSE) (e.g. Croke et al., 2008). For example, if the variable of interest is streamflow (Q) with n observations, NSE can be expressed as in Equation 2.3.

$$NSE = 1 - \frac{\sum_{i=1}^n (Q_{sim,i} - Q_{obs,i})^2}{\sum_{i=1}^n (Q_{obs,i} - \bar{Q}_{obs})^2} \quad (2.3)$$

Where n and Q_{sim} respectively represent the total number of simulation time steps and the simulated value for the variable of interest (here streamflow). Q_{obs} is observed streamflow and \bar{Q}_{obs} represents mean value of the observed streamflow series.

In hydrological modelling, fuzzy measures have been used as informal likelihood measures especially in situations where observational data are too scarce to use statistical measures (Pappenberger et al., 2007; Jacquin and Shamseldin, 2007; Beven, 2012). Fuzzy logic is a method to formalize the human capacity of imprecise reasoning, i.e. an ability to reason approximately and to decide under uncertainty (Ross, 2009). Since in fuzzy logic all truths are partial or approximate, this reasoning has also been called interpolative reasoning with the process of interpolating between the binary extremes of true or false being represented by the ability of a fuzzy logic to encapsulate partial truths (Ross, 2009). The emergence of new theories of uncertainty in the 1960's, including fuzzy logic, have challenged the seemingly unique connection between uncertainty and probability theory by showing that probability theory is capable of representing only one of several distinct types of uncertainty (Klir and Yuan, 1995). Zadeh (1965) introduced sets with boundaries that are not precise (fuzzy sets) where the membership in a fuzzy set is represented in terms of degree rather than as a matter of affirmation or denial.

A fuzzy measure can be defined as a membership function of the error between the observed and predicted variable of interest. The individual fuzzy measures for different time steps or variables may be combined using different methods, including: linear addition, weighted addition, multiplication, fuzzy union or fuzzy intersection. Such likelihood measures are commonly used in the limits of acceptability approach to model evaluation (e.g. Beven, 2006). Example of triangular (a) and trapezoidal (b) membership functions for standardized scores between the limits; as well as two special cases of the trapezoidal membership function, i.e. R-function (c) and L-function (d) are shown in the following figure.

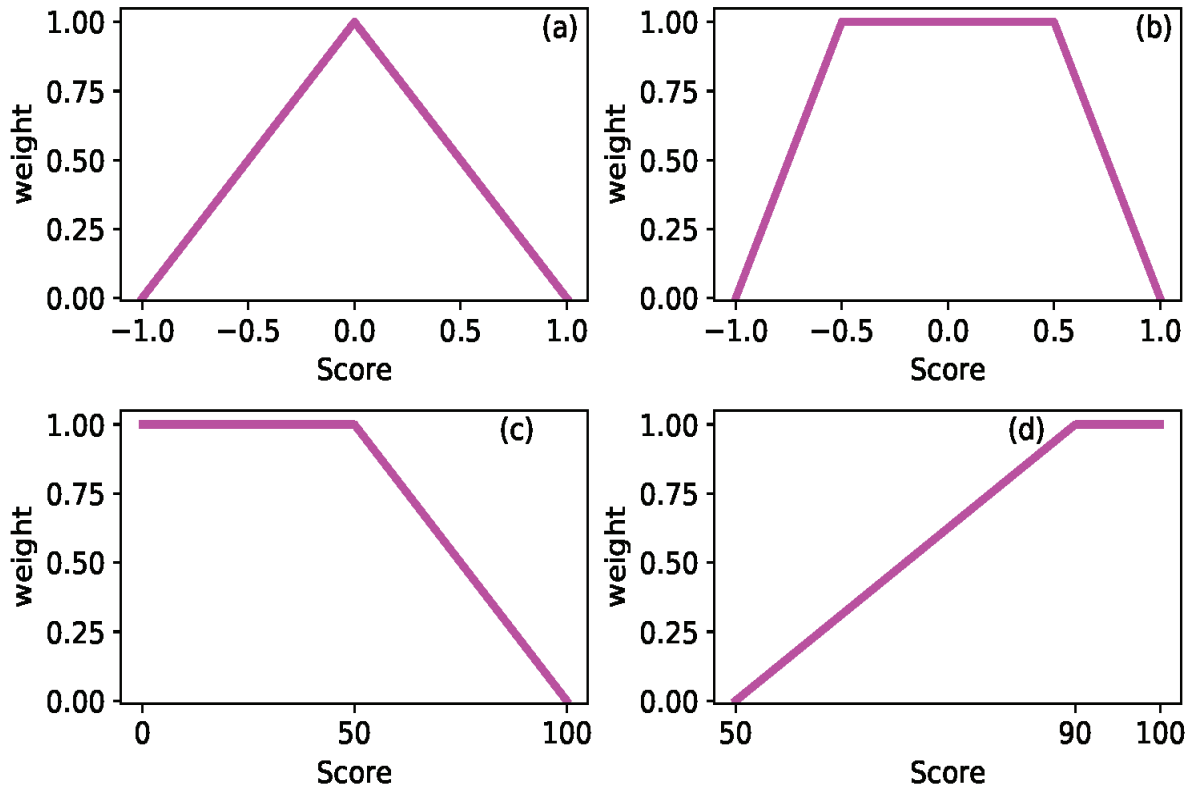


Figure 2.3: Graphical representation of a triangular (a) and trapezoidal (b) membership functions commonly used for the conversion from error scores to likelihood weights as well as the R-function (c) and L-function (d) which are special cases of the trapezoidal membership function.

2.2.3 Information content of data and its implication in formulating a likelihood measure

The amount and information content of a data series used in model evaluation require due consideration since they have a strong influence on model performance (Ritter and Munoz-Carpena, 2013). Previous studies suggest that the information content of the data and the efficiency with which that information is extracted during model identification are more important than the length of records (e.g. Gupta and Sorooshian, 1985). For example, Vrugt et al. (2002) have observed highest information content when conditioning parameters of the HYMOD model with a limited amount but less correlated streamflow measurements. Similarly, Wang et al. (2017) used isotope data from streamflow samples to assess the effect of sample size and period of sampling on the model calibration result. They observed that the performance of models calibrated with isotope data from two selected samples was comparable to simulations based on isotope data from 100 time steps. Further, the samples taken on the falling limb were more informative than samples taken from other parts of the hydrograph.

Information content of data can be assessed using different techniques. Most of the well established methods are based on the information theory which in turn rely upon the probabilistic models of communication developed by Shannon (1948). These models are

based on the basic principle that an observation with a high probability of random occurrences convey little information, while an observation that is least likely to spontaneously occur carry most information (Perone and Ham, 1985). This principle is quantitatively described by the concept of entropy which equates information and uncertainty. Shannon's formula (Eq. 2.4) defines the information content of data, i.e. entropy (I) as weighted average information content of the individual observations in the data series (Shannon and Weaver, 1964). Specific information is a function of the occurrence of a particular event and it is quantified by $-Log$ of the probability of the event (P_i).

$$I = - \sum_{i=1}^N P_i \text{Log}(P_i) \quad (2.4)$$

where N and P_i respectively represent the number of discrete intervals of the variate and the probability of the variate occurring within that interval. Maximum entropy occurs when all events are equally probable (Papoulis, 2002; Vieux and Farajalla, 1994).

Some researchers have been sceptical on the application of the information theory and other similar methods in hydrology to assess the information content of data (e.g. using entropy measures and Akaike information criteria). For example, Beven and Smith (2015) argue that the information about the error sequences is largely lost when using entropy measures since such measures are based only on the shapes of distributions. Rather, an approach for assessing the information content of data from a hydrological perspective should be event based. Croke et al. (2008) proposed a methodology that can be used to assess the information content of new data in data assimilation. It is based on the relative uncertainty before and after assimilation of the new data into the model. The uncertainty could be in the model parameters or model predictions and it can be expressed as a ratio of the standard deviation of parameter distributions before and after the data assimilation. However, they have pointed out that value of the new data for reducing the uncertainty depends on how viable a given approach is in extracting this information (for e.g. depending on type of the likelihood measure).

The main considerations when calibrating hydrological models are the amount of usable information content in the data series used to calibrate a simulation model and the identification of appropriate likelihood measure for assessing the performance of multiple models in light of the different sources of uncertainty in the modelling process (Beven and Smith, 2015). According to Mantovan and Todini (2006), the notion of coherence in learning refers to the condition of increasing amount of information extracted with increase in the number of observations. As such, the Bayesian parameter inference process with a formal likelihood function can be considered as coherent in learning. However, coherence in the Bayesian approach with formal likelihood can only be achieved if the residuals are aleatory and stationary since non-stationarity due to epistemic errors (e.g. errors in observations) might lead to disinformation (Beven and Westerberg, 2011). An ideal likelihood measure would then be one that is not biased by small amount of disinformation in the data to avoid biased inferences. In the past, different measures have been proposed to maximize information extraction from a calibration dataset. For example, Croke (2007) modified the Nash-Sutcliffe

efficiency using an optimal weighted average approach to take into account for the influence of uncertainties in the modelled and measured quantities. The modified NSE yielded a better estimate of the extreme values as compared to the original formulation.

2.2.4 Change point detection methods

As mentioned in the previous section, the amount of information in calibration data with observations that show different characteristics is higher than that of same length but with similar characteristics. Thus, methods for the identification of critical time periods in the observations (for e.g., periods which contain most of the information for parameter identification) can provide useful guidance in characterizing the evaluation data series (e.g. Bárdossy and Singh, 2008). A change point analysis helps to detect if a change has occurred, the number of change points, and the time when the changes occurred (Taylor, 2000). In hydrology, change point detection methods have been used in different areas of application that range from assessing the non-stationarity in hydrological time series data to detection of land use and land cover changes (Bayazit, 2015). The identification of temporal changes in hydrological data has particularly been gaining interest due to the potential impacts of climate change on river flow regimes (Khaliq et al., 2009). A number of change point detection schemes have been proposed for applications in hydrological and environmental problems (Sonali and Kumar, 2013). These range from the traditional approaches based on maximum likelihood estimation and Bayesian methods to newer approaches based on smoothing techniques (e.g. locally weighted regression) (Kundzewicz and Robson, 2004). The commonly used change point detection approaches can be roughly classified as parameteric and non-parameteric approaches. Table 2.1 presents summary of the commonly used change point detection methods in hydrology.

Table 2.1. Summary of the commonly used change point detection methods in hydrology (Kundzewicz and Robson, 2004; Khaliq et al., 2009; Sonali and Kumar, 2013)

Method	Assumptions		Description
	Parametric (normal distribution)	Independent observations	
Median change point test	No	Yes	rank-based test for a change in median
Mann-Kendall test	No	Yes	rank-based test
CUSUM test	No	Yes	rank-based test
Sen's test	No	Yes	slope based test
Spearman rank correlation test	Yes	Yes	rank-based test for correlation between two variables
Linear regression test	Yes	Yes	slope based test
Student's t test	Yes	Yes	parametric test on sample means

Change point problems can be continuous or discrete and in the case of continuous change point problem it is generally assumed that a continuous random variable representing a survival or failure time can be expressed by the hazard function (Ebrahimi and Ghosh, 2001). In the case of discrete change point problem, we assume n independent random sequence of data, $x_1:n = (x_1, \dots, x_n)$ with an abrupt change expected to occur within this set when there exists a time, $\tau \in \{1, \dots, n - 1\}$, such that the statistical properties, θ (e.g. mean or variance), of $\{x_1, \dots, x_\tau\}$ and $\{x_{\tau+1}, \dots, x_n\}$ are different in some way (Fig. 2.4). As such, a change point detection approach can be treated as hypothesis testing with the null (H_0) and alternative (H_1) hypotheses for a change defined as:

H_0 : No change point expected, i.e. $\theta_1 = \theta_2$

H_1 : A change point is expected to occur at time τ , i.e. $\theta_1 \neq \theta_2$



Figure 2.4: An illustration of time-instants in discrete change point problem. The required variable, τ , is the time instant when the time series data shows an abrupt change. The data during the intervals $[1, \tau]$ and $[\tau, n]$ are respectively generated from θ_1 and θ_2 which are both stationary and unknown (after Alippi et al., 2013).

This hypothesis can be tested using a likelihood-ratio based approach which requires the estimation of the maximum log-likelihood value under both the null and alternative hypotheses (e.g. Eckley et al., 2011; Killick and Eckley, 2014). For the null hypothesis, the likelihood value ($L(H_0)$) is expressed as:

$$L_{H_0} = \prod_{i=1}^n p(x_i|\theta_0) \quad (2.5)$$

Under the alternative hypothesis, we consider a model with a change point at τ ; and the log likelihood for a given τ , i.e. the likelihood given the alternative hypothesis (L_{H_1}) is expressed based on the joint probability density function as:

$$L_{H_1} = \prod_{i=1}^{\tau} p(x_i|\theta_1) \prod_{j=\tau+1}^n p(x_j|\theta_2) \quad (2.6)$$

The log likelihood ratio (R_τ) is calculated as:

$$R_\tau = \log\left(\frac{L_{H_1}}{L_{H_0}}\right) = \sum_{i=1}^{\tau} \log p(x_i|\theta_1) + \sum_{j=\tau+1}^n \log p(x_j|\theta_2) - \sum_{k=1}^n \log p(x_k|\theta_0) \quad (2.7)$$

The generalized log likelihood ratio (G), is estimated as:

$$G = \max_{1 \leq t \leq n} R_t \quad (2.8)$$

The null hypothesis is rejected if $2G > \lambda$, where λ is a predefined critical value. If the null hypothesis is rejected, the change point position is the value of τ that maximizes R_τ . The penalty factor (λ) can be defined using various criteria such as the Bayesian Information Criterion (BIC) and Akaika Information criterion (AIC).

This likelihood based change point detection approach as well as another scheme based on a nonparametric approach were applied in this work (Paper II) to identify a critical point in the time axis, i.e. where the mean fractional snow cover area (fSCA) changes. The study was conducted using fSCA data for nine sites in the study domain. The result obtained from this analysis was employed to assess the variability of this critical point spatially from one grid-cell to another and temporally from year to year in response to various physiographic and climatic factors. The identified critical time index (change point) together with start of the melt-out period were employed to define a dynamic information content of remote sensing fSCA observations for use during the assimilation of the fSCA into a hydrological model through an appropriate likelihood function. Figure 2.5 shows temporal variability of the fSCA dynamics and location of the identified change points for a grid-cell situated in site 9. This site is characterized by relatively highest average annual snow water equivalent (SWE) as compared to other sites in the catchment. The four sample years used in assessing the temporal variability of the change point location include the years with the lowest (year 2014) and highest (year 2012) annual peak snow water equivalent records in the analysis years, i.e. from year 2008 to 2016. Figure 2.6 also shows the result for similar analysis but for a grid-cell located in site 2, which is characterized by a relatively lowest average SWE value in the catchment. Some potential factors for the spatial variability of SWE in the study catchment, in particular the significances of terrain parameters, are presented in Section 2.4.1.

The result shows that both the likelihood-based parametric approach and the nonparametric approach yielded close and reasonable results. From these sub-plots it can also be noticed that, the detected change points varied spatially between the grid-cells and temporally from year to year. The year to year variability of the change points was somehow consistent with the observed amount of maximum SWE in respective years. The detected change point occurred at a latter point in the time axis for the year with highest maximum SWE, i.e. year 2012 (b) as compared to the other years. The comparison between the grid-cells with lowest (Fig. 2.5) and highest (Fig. 2.6) average SWE shows that the change points detected for the grid-cell with lowest SWE occurred at earlier point in the time axis as compared to the grid-cell with highest SWE for the corresponding years.

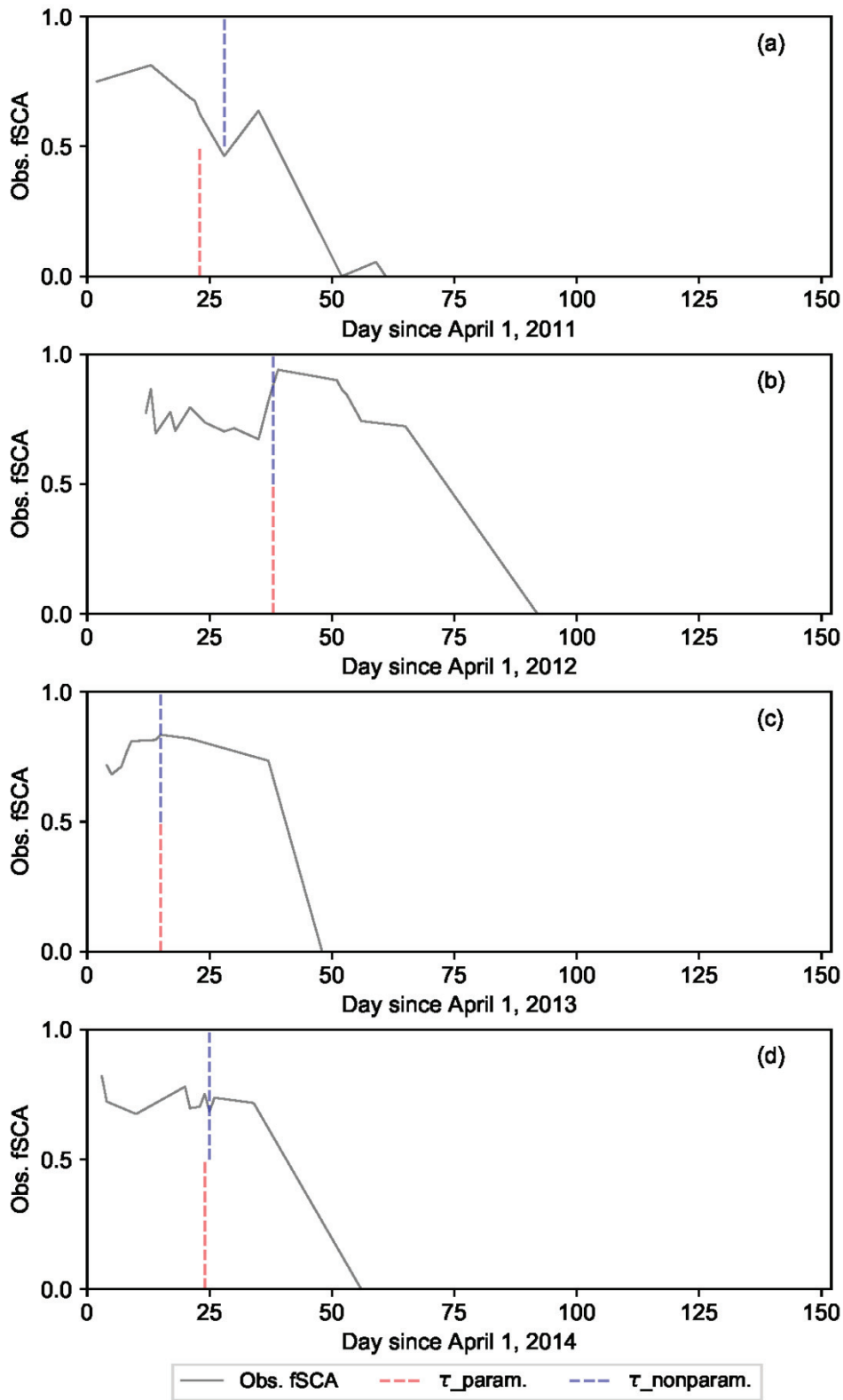


Figure 2.5: Temporal variability of the fSCA dynamics and location of the change points τ_{param} and τ_{nonparam} respectively detected using the parametric and non-parametric approaches for a sample grid-cell located in the site with lowest average SWE (site 9) and in four sample years, i.e. year 2011 (a) , 2012 (b), 2013 (c) and year 2014 (d).

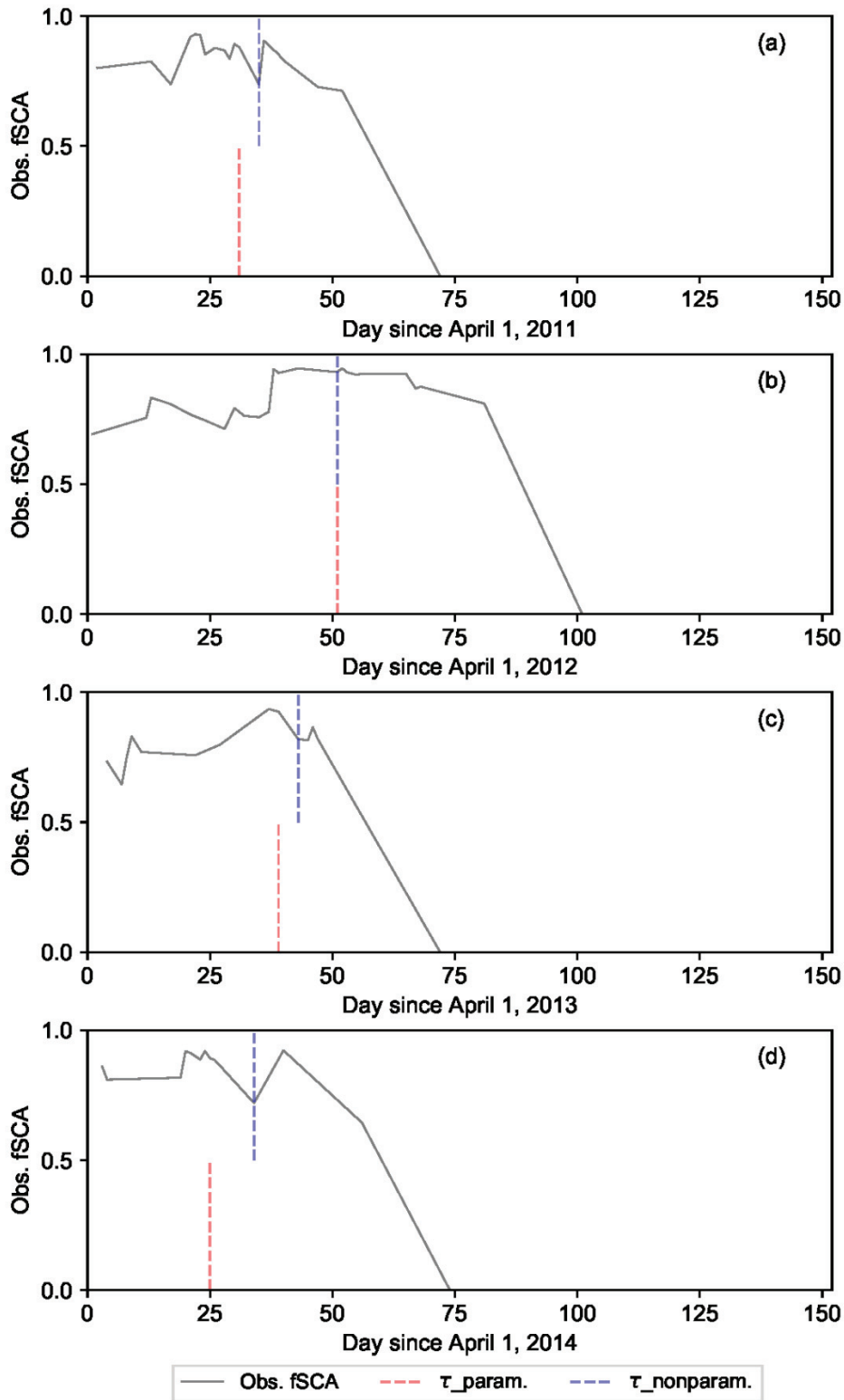


Figure 2.6: Same as the previous figure but for a sample grid-cell in the site with highest annual average SWE (site 2).

2.2.5 Parameter sensitivity analysis

Sensitivity analysis may provide guidance as to which of the parameters is a better candidate for further analysis in order to reduce the uncertainty in the simulation result (Saltelli et al., 2006). Making more focus on parameters to which the model simulation results are most sensitive gives improved efficiency of model calibration and this requires an approach to assessing parameter sensitivity with respect to predicted variables or with respect to some performance measure (Beven, 2011). Some of the possible purposes of conducting sensitivity analysis are (Saltelli et al., 2008):

- model corroboration, i.e. to test if the inference is robust or if the model is highly dependent on fragile assumptions
- research prioritization, i.e. to identify the parameter that needs further analysis or measurement
- model simplification, i.e. to assess if some parameters or model components can be fixed or simplified
- identification of critical regions in the parameter space, and
- to identify parameter interactions that may yield extreme values.

The purposes of sensitivity analysis guide the choice of the appropriate sensitivity analysis method since different methods are better suited to address different purposes (Pianosi et al., 2016). Depending on how these purposes are formulated and addressed, different types of sensitivity analysis have been distinguished in the literature (e.g. Saltelli et al., 2008; Pianosi et al., 2016) including:

- *local versus global sensitivity analysis*: local sensitivity analysis considers the output variability in response to the variations of a model parameter around a baseline point, while global sensitivity analysis considers the variations within the entire space of variability of the model parameters.
- *quantitative versus qualitative sensitivity analysis*: in quantitative sensitivity analysis each parameter is associated with a quantitative and reproducible evaluation of its relative influence, usually through use of sensitivity indices or importance measures. Whereas in qualitative sensitivity analysis, sensitivity is assessed qualitatively by visual inspection of model predictions such as scatter plots and graphical representations of the posterior distributions of the model parameters.

(a) Local sensitivity analysis

This method involves variation of the model parameters from their nominal values one at a time (OAT) with subsequent assessment of their impacts on the simulation results through visual inspection with the help of graphical representations of the simulated variables under the nominal and perturbed parameter values (Pianosi et al., 2016). A quantitative measure of the sensitivity of the model simulation to the i^{th} parameter can be obtained by the partial derivative of the change in model output with respect to the change in model parameter value ($\partial g/x_i$) as evaluated at the nominal value of the parameters (x_n) or by finite-difference gradient. The sensitivity measures of each parameter are rescaled using a scaling factor (c_i) for the ease of comparison of the sensitivities across the model parameters with different units

of measurement. The partial derivatives for quantifying the sensitivity index of the i^{th} parameter ($\hat{S}_i(x)$) are usually approximated by finite differences (Eq. 2.9). These methods are computationally less demanding as they require only $M+1$ number of model runs. However, the information they provide is limited to local sensitivity.

$$\hat{S}_i(x) = \frac{g(x_1, \dots, x_i + \Delta_i, \dots, x_M) - g(x_1, \dots, x_i, \dots, x_M)}{\Delta_i} c_i \quad (2.9)$$

where, g represents the relationship between the model's inputs and outputs assessed by the model's response. M and Δ_i respectively denote the number of model parameters subjected to sensitivity analysis and width of the finite variation.

(b) Global sensitivity analysis

The local sensitivity type discussed above can be extended to global sensitivity analysis by using a multi-start perturbation instead of the simple perturbation method. This can be realized by computing the output perturbations from multiple points within the feasible range of the model parameters followed by the measurement of the global sensitivity by aggregating the individual sensitivities. The methods falling under this category differ depending on whether they use finite differences directly or they involve some transformations such as their absolute or squared values; how the fixed points in the parameter space and the width of the finite variation are selected; and on the way they aggregate the individual sensitivities (Pianosi et al., 2016). Two of the most commonly used sensitivity analysis methods that fall under this category, namely, the Elementary Effect method (EE) and Regional Sensitivity analysis (RSA) are briefly discussed below. While, the EE method uses an OAT sampling approach RSA uses an AAT sampling approach.

The Elementary Effect Method (EE) (Morris, 1991) is an effective way of screening a few important parameters from several model parameters. The main aim behind this method is to determine which model parameter has an effect that is (a) negligible (b) linear and additive, or (c) nonlinear or characterized by interaction with other parameters. Different sampling strategies have been proposed to select the values in the parameter dimension and to determine width of the input variation (Saltelli et al., 2008). The original procedure by Morris (1991) involves sampling r points for M independent parameters $x^j (j=1 \dots, r)$ across user defined even number of levels (p). The width of the input variation (Δ) is determined based on p , i.e. $\Delta = p/(2(p-1))$. The elementary effect of the i^{th} model parameter is defined as (e.g. Campolongo et al., 2007; Saltelli et al., 2008; Pianosi et al., 2016):

$$EE_i = \frac{g(x_1^j, \dots, x_i^j + \Delta_i^j, \dots, x_M^j) - g(x_1^j, \dots, x_i^j, \dots, x_M^j)}{\Delta_i^j} c_i \quad (2.10)$$

For each input, two sensitivity measures are calculated. The mean and standard deviation of EE computed based on the individual EE, i.e. EE_i respectively assess the overall influence of the parameter on the model output and due to non-linear or interactions with other parameters. A high value of mean EE for a given parameter shows higher effect on model outputs and high standard deviation shows high degree of interaction with other parameters.

Results from the EE method are usually accompanied by graphical representations; and the mean EE vs standard deviation of EE graph allows for a better interpretation of the two sensitivity measures simultaneously. Figure 2.7 shows a sample graphical representation of a sensitivity analysis output using the EE method. From this figure it can be noticed that X3 is the most influential model parameter as compared to the remaining parameters. Similarly, the comparison between two less influential sample parameters, X1 and X2 shows that X2 is relatively more influential in terms of the mean effect than X1. On the other hand, the degree of interaction with other parameters is higher for X1 than for X2.

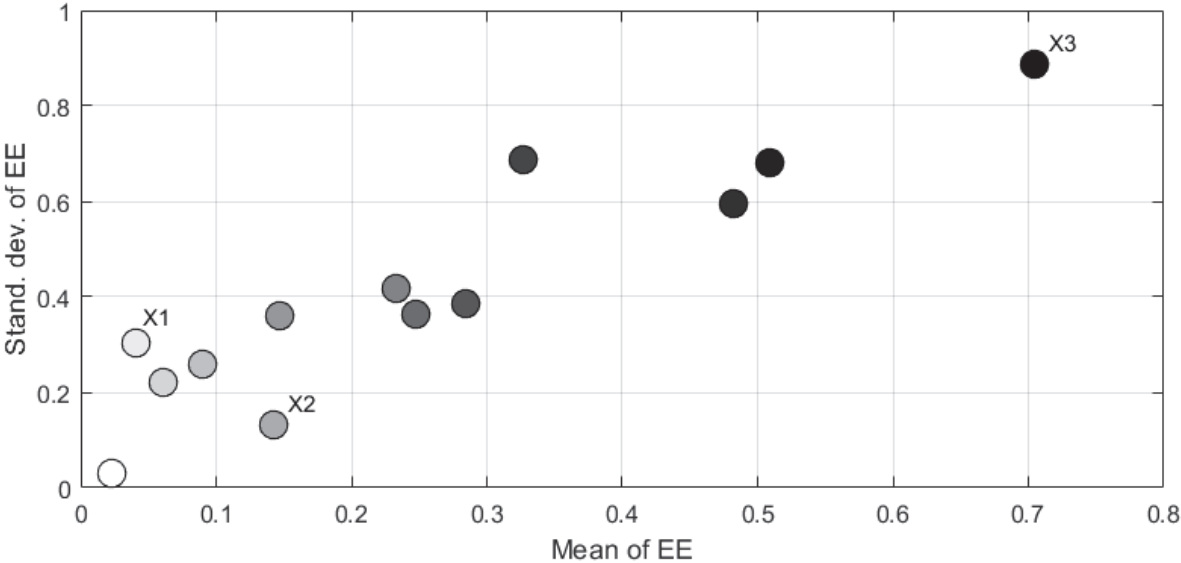


Figure 2.7: A sample graphical representation of an EE based sensitivity analysis output (Standard deviation of EE against the mean value of EE) for 13 model parameters.

The Regional sensitivity analysis (RSA) (Hornberger and Spears, 1981) was a precursor of the GLUE methodology and it is based on the Monte Carlo simulation, which makes several different model runs with each run using a randomly chosen parameter set (Beven, 2011). The simulations are classified into behavioural and non-behavioural samples with respect to the system under study. For example, behavioural (acceptable) simulations might be those with a high value of the variable of interest or a certain performance measure and non-behavioural with a low value. These two samples are compared, for example using visual inspection of the empirical cumulative distribution function (CDF) of the two samples plotted in the same graph. Figure 2.8 shows graphical representation of RSA result for selected parameters of the PT_GS_K model based on implementation of the methodology in the SAFE toolbox (Pianosi et al., 2015). The CDFs of the model parameters are associated with an efficiency metrics of streamflow simulations (NSE) above or below 0.67. From this graph it can be noticed that the catchment response parameters, i.e. $c1$, $c2$ and $c3$ are relatively more influential as compared to the remaining parameters. The maximum deviation between the CDF of behavioural and non-behavioural simulations is much wider for the more influential as compared to the less influential parameters. The model parameters sa and cv are particularly insensitive parameters that yielded very close behavioural and non-behavioural CDF curves.

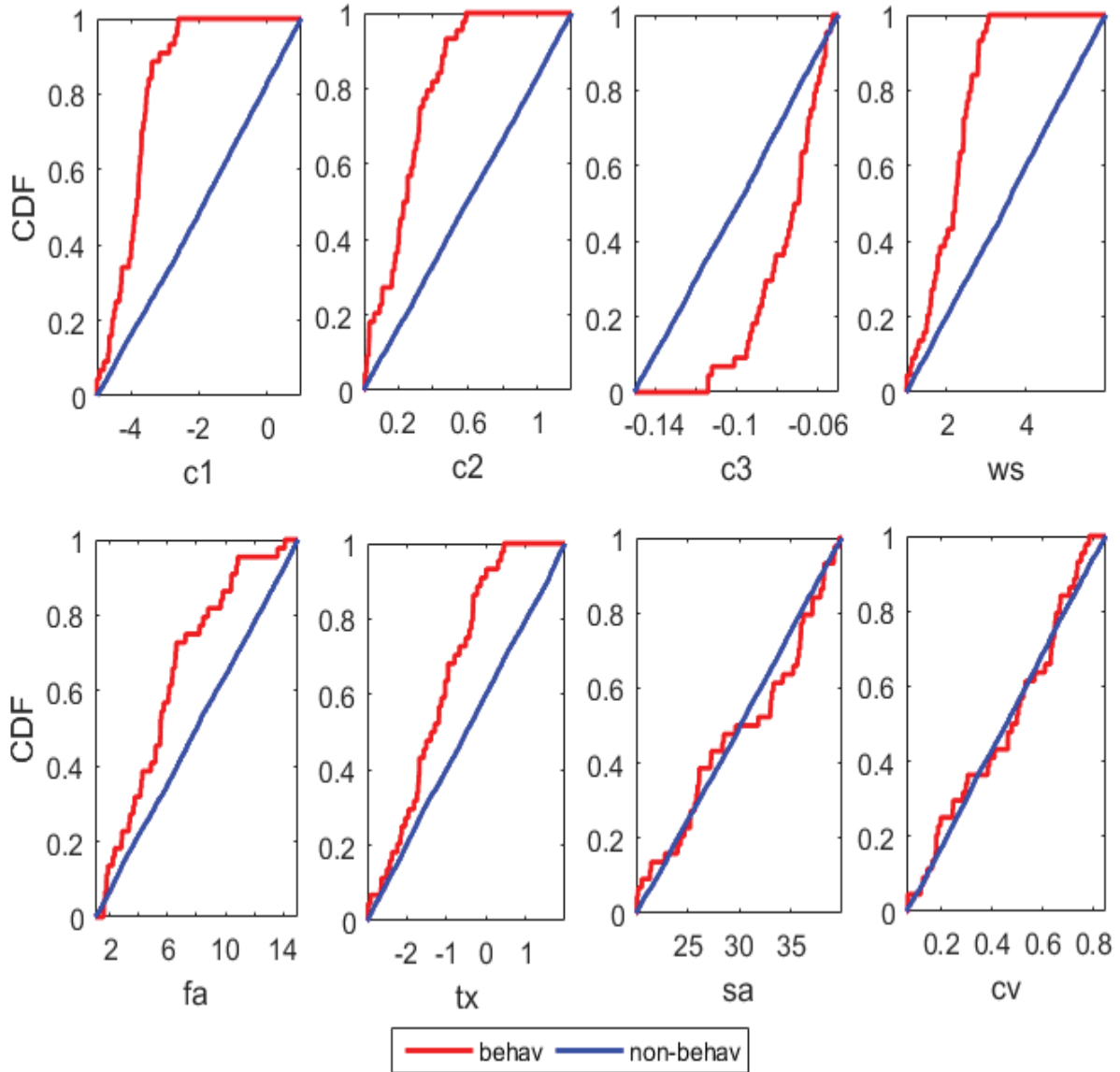


Figure 2.8: Regional sensitivity analysis result for selected parameters of the PT_GS_K model. Each plot shows cumulative distribution function (CDF) of behavioural and non-behavioural simulations against parameter values.

A quantitative measure of the divergence between the two distributions, i.e. the sensitivity index (S_i), can be calculated using the Kolmogorov-Smirnov statistic (Eq. 2.11) (Pianosi et al. 2016). Figure 2.9 shows the median and the confidence interval of the sensitivity index over bootstrap resamples of the PT_GS_K model parameters. The result from the sensitivity index was similar to the qualitative outputs displayed in the preceding figure where the catchment response parameters, i.e. $c1$, $c2$ and $c3$ have shown higher sensitivity index as compared to most of the snow and water balance related parameters, i.e. tx , fa , sa , and cv . The later parameters have also yielded slightly wider confidence interval as compared to $c1$ and $c2$. On the other hand, ws has shown higher sensitivity index with narrow confidence interval as compared to the other parameters with the exception of $c1$.

$$S_i = \max_{x_i} |F_{x_i|y_b}(x_i|y \in y_b) - F_{x_i|y_{nb}}(x_i|y \in y_{nb})| \quad (2.11)$$

where $F_{x_i|y_b}$ and $F_{x_i|y_{nb}}$ represent the empirical cumulative distribution functions of the model parameter x_i for the behavioural (y_b) and non-behavioural (y_{nb}) samples, respectively.

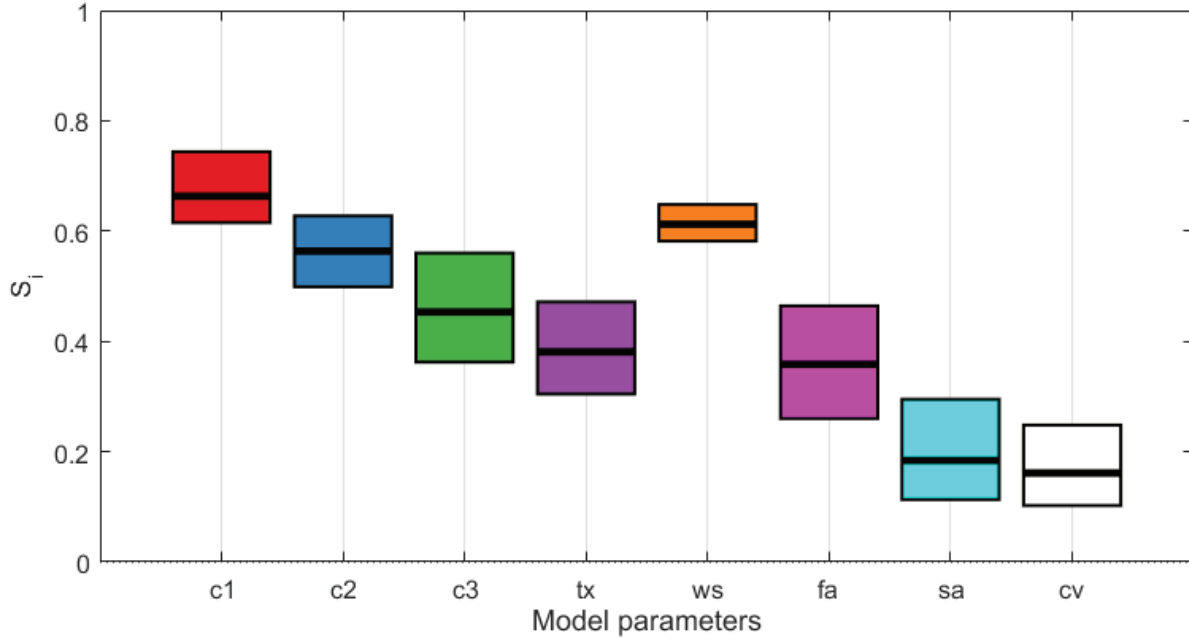


Figure 2.9: The median and 5-95 % confidence intervals of the sensitivity index (S_i) over bootstrap resamples of the PT_GS_K model parameters.

2.2.6 Sources and nature of uncertainty

As a consequence of aggregating the processes in the simplified representation of reality, conceptual hydrological models often contain one or more model parameters whose values cannot be directly obtained from physical measurements. The parameters can be inferred by using historical observations in the calibration process. During calibration, model parameters are fine tuned in such a way that the hydrological model reproduces the observed response of the system as close and consistently as possible over some defined period of time (Vrugt et al., 2005). A rapid growth in computational power coupled with increased availability of higher resolution data both in space and time as well as improved understanding of the underlying processes of the hydrological system have benefited hydrological modelling (Liu and Gupta, 2007). This has led to the development of complex semi-distributed and distributed physics-based and conceptual hydrological models with more number of calibration parameters. However, this has resulted in further difficulties in the model calibration process leading to a considerable level of uncertainty in model predictions. Efficient methods are required to deal with the increasing level of uncertainty associated with the models, forcing data, and observations. The necessity for a proper consideration of uncertainty analysis in hydrological modelling right from the beginning of the modelling process has been receiving a growing interest among the modelling community when such models are applied both for research and operational purposes (e.g. Beven, 1989 ; Wagener and Gupta, 2005; Saltelli et al., 2006).

The understanding, quantification, and reduction of uncertainty are important aspects in hydrological modelling and prediction (Liu and Gupta, 2007). Uncertainty is a situation of inadequate information that can be of three types: inexactness, unreliability, and border with ignorance (Funtowicz and Ravetz, 1990). The effort for maximizing the usefulness of a model is closely connected with the relationship among three key characteristics of all systems models, i.e. complexity, credibility, and uncertainty (Klir and Yuan, 1995). In general, allowing more uncertainty tends to reduce complexity and increase credibility of the resulting model. As such, uncertainty becomes very valuable when considered in connection to the other characteristics of a model, albeit usually undesirable when considered alone (Klir and Yuan, 1995). There are different dimensions of uncertainty related to model based decision support exercises; and Walker et al. (2003) distinguish three dimensions of uncertainty, i.e. the location (sources) of uncertainty, the nature (type) of uncertainty and the level of uncertainty (Figure 2.10). The level of uncertainty refers to the degree of knowledge about the subject matter that falls within the spectrum ranging from the ideal complete deterministic understanding at one end of the scale to total ignorance at the other end. The main dimensions, namely, the sources and nature of uncertainty are discussed in more detail in this section.

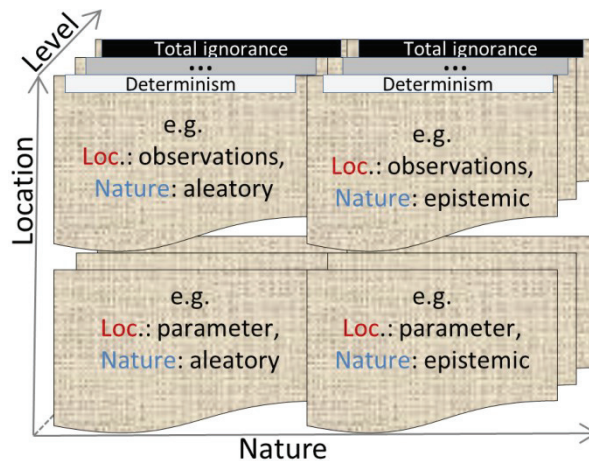


Figure 2.10: Three-dimensional conceptualization of uncertainty, i.e. the level, location (Loc) and nature of uncertainty (after Walker et al., 2003)

(a) Main sources of uncertainty

The calibration process and hydrologic predictions are highly influenced by different sources of uncertainty; and the identified behavioural parameter sets might be effective values differing from the prior estimations based on direct observations or other sources (Moradkhani and Sorooshian, 2009; Beven and Smith, 2015). Knowledge of the main sources of uncertainty (outlined below) is thus important in different aspects of the modelling process in general and model calibration in particular (e.g. Walker et al. 2003; Beven, 2016).

- *Input and validation data uncertainty*: this stems from measurement errors and scaling issues, for example, due to forcing data downscaling. Similarly errors of the rating curve may affect the streamflow estimates and thereby lead to validation data uncertainty. These errors arise independently of the hydrological model and hence their statistical properties such as mean and variance can be estimated prior to

calibration with due consideration to the data collection instruments and the procedures (Renard et al., 2010).

- *Model structure uncertainty*: As mentioned in Section 2.1, the perceptual model depends on the analyst's process understanding which in turn depends on previous experience. Alternative perceptual models can represent competing hypotheses about the structure and functioning of an observed system (Gupta et al., 2012). Further, since conceptual models are simplified representations of a perceptual model, they may introduce additional uncertainty in the model structure due to oversimplification. In general, the structural uncertainty of a hydrological model is poorly understood and it mainly depends on model formulation (e.g. number and connectivity of stores), on the specific catchment, and on the spatial and temporal scales of the analysis (Renard et al., 2010).
- *Parameter uncertainty*: this is related to the inability to specify the exact values of model parameters and it has important interaction with the input data and observations used for model conditioning as well as with the errors emanating from the analyst's process understanding (Renard et al., 2010). The most feasible scientific way of assessing parameter uncertainty is through inverse modelling (Duan et al., 1994; Hill, 2000). Section 2.2.7 presents commonly used methodologies in parameter uncertainty analysis.

(b) Nature of uncertainty

Understanding the nature of uncertainty helps to adequately address the specific uncertainty. This section presents the main types of uncertainty reported in previous studies (e.g. Walker et al, 2003; Beven, 2015).

- *Aleatory uncertainty*: this is also referred in the literature as statistical, stochastic and variability uncertainty. Generally, this type of uncertainty is characterized by stationary random distribution induced by the input data, parameters and certain model structures.
- *Epistemic uncertainty*: this type of uncertainty may stem from a lack of knowledge about how to represent the real world hydrological system in terms of model structure and parameters. This may also include processes that have not yet been included in the perceptual model, where the absence of these processes might result in reduced model performance when unexpected events occur. It also stems from the lack of knowledge about the forcing data or the validation data due to such factors as commensurability or interpolation issues, lack of knowledge on interpretation of measurement data, or non-stationarity and extrapolation in rating curves.

Other less familiar types of uncertainty include (Beven, 2015): semantic uncertainty that emanates from the ambiguity on what statements or quantities mean in the relevant domain (e.g. storm runoff vs. quick flow); and ontological uncertainty which is associated with different belief systems, for example, beliefs about whether a formal probability based framework is appropriate for the representation of model residuals. The next section discusses

some of the methodologies commonly used for parameter uncertainty analysis in hydrological modelling and their underlying philosophies.

2.2.7 Uncertainty analysis methodologies

Various calibration and uncertainty analysis methods have been proposed and applied to meet the increasing demand for such tools in hydrological modelling, including the generalized likelihood uncertainty estimation (GLUE) methodology (Beven and Binley, 1992), the Shuffled Complex Evolution algorithms (SCE-UA, SCEM) (Duan et al., 1992; Vrugt et al., 2003) and the dynamic identifiability analysis framework (DYNIA) (Wagener et al., 2003). The formal Bayesian statistical approach and the GLUE methodology are the most commonly used frameworks. A detailed description of these methodologies can be found from hydrology and statistical literatures (e.g. Beven and Binley, 1992; Mantovan and Todini, 2006) and a brief description of these approaches is presented here.

The formal statistical Bayesian approach assumes a certain probability distribution (prior) of the model parameters and a likelihood function (usually a Gaussian). For each additional data, the posterior parameter values are estimated as the product of the prior and the new likelihood. When using this method, only one set of model parameters (model realization) that optimizes the model performance is commonly selected as representative of the catchment behaviour. In a Bayesian inference, the solution to an inverse problem is given by a posterior probability distribution, $P(M|D)$, over the model space (Eq. 2.12). The posterior probability distribution is a function of the available information on the model obtained from the data (D) through the likelihood function, i.e. $P(D|M)$; and the data-independent prior information expressed by a prior probability density function, i.e. $P(M)$ (Moradkhani, 2008; Moradkhani and Sorooshian, 2009).

$$P(M|D) = \frac{P(D|M)P(M)}{P(D)} \quad (2.12)$$

where, the denominator, $P(D)$ represents a normalization factor.

The GLUE methodology employs a non-statistical formulation of uncertainty which is based on the concept of equifinality, where different model structures and parameter sets can give similarly good simulation results when compared to calibration data based on certain efficiency criteria (Beven, 2011). This methodology recognizes that most environmental models are characterized by non-identifiability mainly due to the over-parameterized structure of these models (Yang et al., 2018). The GLUE methodology has introduced a different philosophy to the venue of model calibration where the main goal had historically been the identification of an optimal single model (Liu and Gupta, 2007).

When applying the GLUE methodology the modeller defines the likelihood measure and a given threshold value which is used to assess whether a particular model realization with a randomly selected parameter set is behavioural or non-behavioural. All model realizations whose likelihood values that satisfy this threshold value are considered behavioural, i.e. representing the catchment behaviour. As such, in the case of the GLUE methodology we may have several model realizations instead of a single optimal model realization; and

prediction of future values such as streamflow is conducted using all behavioural model realizations. A cumulative distribution function (CDF) is calculated based on the predicted values and likelihood weight of each model realization (w_i , Eq. 2.13) from which the required quantiles might be extracted to estimate a crisp value (e.g. the median) or the predictive uncertainty based on width of the prediction band.

$$w_i = \frac{L(x^i)}{\sum_{k=1}^N L(x^k)} \quad (2.13)$$

where, $L(x^i)$ is the likelihood value of the model realization (i) with a parameter set x^i , and N is the total number of behavioural samples.

2.2.8 Ensemble modelling and multi-objective evaluation

Regardless of the effectiveness and consistency of the optimization methods in finding a global solution, previous studies have shown that different combinations of parameter sets may result in similar values of a likelihood measure, and thus several optimum solutions may exist due to equifinality (e.g. Moradkhani and Sorooshian, 2009). The inherent uncertainties in parameter identification due to the various sources that are both aleatory and epistemic in nature cannot be fully addressed by an automatic calibration procedure. It is, thus, more realistic to use multiple plausible realizations from one or more models to get an ensemble of predictions rather than trying to find an optimal parameter set (Peel and Blöschl, 2011). Ensemble modelling also provides a way to assess predictive uncertainty using ensemble-based frameworks such as the GLUE methodology. Previous ensemble-based hydrological studies have reported that single or multi-model ensembles provided better calibration and validation results than single realizations (e.g. Viney et al., 2009).

The use of ensemble-based techniques to optimize model parameters has allowed for the incorporation of multiple criteria within the calibration procedure resulting to a number of alternative parameter sets that are optimal on the basis of the Pareto-dominance concept (Efstratiadis and Koutsoyiannis, 2010). A multi-objective optimization involves the simultaneous optimization of several numerical measures representing multiple objective functions and such approaches have been used in different hydrological studies (e.g. Ritzel et al., 1994; Yapo et al., 1998). The fundamental concept of Pareto optimality is commonly employed to look for acceptable trade-offs in simultaneously optimizing all the objective functions (Fig. 2.11). A vector of control variables is Pareto optimal if there is no feasible vector that would improve some criterion without causing a simultaneous deterioration of at least one other criterion (Marler and Arora, 2004). A multi-objective evaluation might be formulated based on the following types of information (Madsen, 2003; Peel and Blöschl, 2011):

- *Multi-variable data*: this includes different observable variables such as streamflow and fractional snow cover area.
- *Multi-site data*: historical data of the same variable obtained from several gauges within a catchment and are reproduced through distributed mode of simulation.

- *Multi-response models*: independent criteria representing for various aspects of a single variable (e.g. streamflow) reproduced by a hydrologic model. This approach strives for a satisfactory agreement of the specific components of the time-series data (e.g. the rising and falling limbs of a hydrograph) rather than for an average good match.

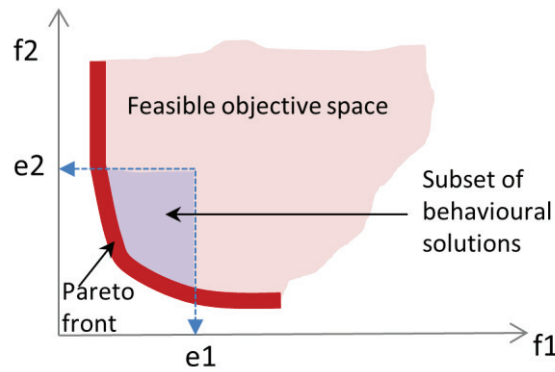


Figure 2.11: Graphical illustration of Pareto-optimal and behavioural solutions in the objective space for simultaneous minimization of two objective functions (f_1 and f_2). The axis limits (e_1 and e_2) denote the threshold for behavioural and non-behavioural solutions (after Efstratiadis and Koutsoyiannis, 2010).

2.2.9 Philosophical aspects in testing hydrological models as hypotheses

The different uncertainty analysis methods are based on some school of scientific philosophical thoughts that generally can be classified as inductive or deductive methods. The inductive method relies on the idea of the inference from singular statements (e.g. the results from observations or experiments) to universal statements, such as hypothesis or theories; whereas, in deductive reasoning predictions are derived from theories (e.g. Refsgaard and Henriksen, 2003; Wilkinson, 2013). According to the inductive school of thought, theories are proposed through inductive logic, and the proposed theories are confirmed or refuted on the basis of critical experiments designed to verify the consequences of the theories (Matalas et al., 1982). Proponents of the deductive method (e.g. Popper, 1959; Kuhn, 1962) argue that theories cannot be verified, they can only be falsified. Regardless of the number of confirmatory observations of a theory or prediction, there will always be the possibility that more than one theory can explain the observations with the false theories likely to be confronted with future observations that falsify them (Hume, 2016).

A hydrologic model can be considered as a hypothesis of catchment behaviour that encompasses a description of dominant hydrological processes and predicts how these processes interact to produce the catchment's response to external forcing such as climatic inputs (e.g. rainfall and temperature) and catchment characteristics (e.g. land cover and soil properties) (e.g. Clark et al., 2011). As such, all environmental models are representations of simplified hypotheses of the real world. These hypotheses require rigorous construction and testing before their application in solving real world hydrological problems. When the calibration and uncertainty analysis frameworks such as the formal Bayesian approach and the GLUE methodology are applied in calibration and validation of hydrological models they serve as hypothesis testing frameworks. However, since hydrology is an inexact science,

subjected to both aleatory and epistemic uncertainties, there are important issues about how to test models as hypotheses of catchment behaviour in all aspects of hydrology (Beven, 2018).

In the last years, there has been an ongoing debate in the literature between the advocates of the formal Bayesian approach and the GLUE methodology with regard to testing hydrological models as hypotheses. While proponents of the formal Bayesian approach focus on model confirmation, the latter group suggests that falsification of models is a better idea (Beven, 2018). The GLUE falsification approach is based on the Popperian philosophy that “every genuine test of a theory is an attempt to falsify it, or to refute it” (Popper, 1963). There has been an increased appreciation of the concept of rejection as positive result of modelling studies in the literature (e.g. Vache and McDonnell, 2006). In the GLUE methodology, if the hydrologic model fails to produce any behavioural model realization, the model is rejected as not fit for the intended purpose. In subsequent step of the modelling process, the model developers are expected to identify the weak points of the model and accordingly add, remove, or simply replace some of the process and/or component representations of the hydrologic system in the model. This way, the rejected model can benefit from the hypothesis test. Proponents of this approach criticise the formal Bayesian method for never rejecting the entire model as hypothesis (e.g. Beven et al., 2012).

The GLUE methodology has been criticised for using a subjectively chosen threshold value to identify behavioural models (e.g. Clark et al., 2011). Further, GLUE was criticised for losing the properties of coherence and consistency in learning due to its use of informal likelihoods that do not satisfy the Baye’s theorem. According to Mantovan and Todini (2006), a non-decreasing amount of extracted information content with the increasing number of sampling observations is one of the basic requirements that guarantee the success of a Bayesian inference process; and the GLUE likelihoods fail to guarantee these properties. The main assumption behind this argument is that each observed data is informative and leads to better estimate of parameter values. However, due to different sources of errors in the observational data, such as errors in gauging stations, as well as data reading and processing errors, not all observations carry valid information. In fact some observations might be physically inconsistent and dis-informative; and the Bayesian approach is prone to such erroneous observations (Beven et al., 2012). The Bayesian statistical method tries to represent all sources of uncertainty within a coherent statistical framework assuming epistemic uncertainties can be represented as if they were aleatory in nature (Beven, 2011).

These methodologies attempt to address the problem of parameter non-uniqueness in different ways on the basis of their underlying philosophies (e.g. Wagener and Gupta, 2005). The Bayesian approach attempts to minimize the non-identifiability problem using its Gaussian likelihood function that results to stretched likelihood surface where two models with similar root mean of squares error may have likelihoods that differ by orders of magnitude. This results to forceful differentiation of model hypotheses that might be rather similar in performance (Beven et al., 2012). On the other hand, the GLUE methodology accepts equifinality as a working paradigm for the identification of behavioural models (Choi and Beven, 2007). This is based on the philosophy that natural systems are so complex to be represented by a single model structure. Instead it is more likely to have several competing model structures (hypotheses) that can represent this complex system, i.e. an acceptable

model prediction might be achieved through use of different model structures or parameter sets (Beven, 1993; Beven, 2002). This school of thought is in line with Oreskes et al. (1994) that, since natural systems are open and model results are always non-unique, the verification and validation of environmental models is impossible and there will always be the possibility of equifinality (Refsgaard and Henriksen, 2004). In the GLUE methodology, non-uniqueness and non-identifiability are viewed as internal characteristics of the modelling process (Beven, 2006). This methodology accepts all hypotheses that satisfy a certain threshold level (based on the modeller's belief) for prediction of the future catchment behaviour. As such, the higher number of behavioural model realizations we have, the more likely are our model predictions to bracket future observations and thereby minimizing the prediction uncertainty from future surprises. Thus, for many advocates of the GLUE methodology, as Savenije (2001) has put it, equifinality is a blessing in disguise.

The implications of these different schools of thought when testing models as hypotheses coupled with the various sources of uncertainty is that, there is a possibility of accepting a poor model while it should be rejected and thereby committing Type I error (false positive) or rejecting a good model while it should be accepted, i.e. making Type II error (false negative). In statistical hypothesis testing a certain probability of rejecting the hypothesis while it is true is commonly set (e.g. at 5% level). However, it is difficult to follow such assumptions in simulation models that are commonly forced with data characterized by high spatiotemporal variability (e.g. Beven, 2011). Type I errors should be avoided since the use of falsely retained models in prediction may lead to false inferences and poor decision making. It is also important to avoid Type II errors since we do not want to reject a good model due to the errors in input or observational data which otherwise can well represent the catchment behaviour. Although, it is difficult to avoid Type I and Type II errors, these errors can be minimized if the testing is done with due consideration to some of the common causes leading to such errors. For example, a Type I error primarily occurs when there is uncertainty in the inputs to the model and the evaluation data to the extent that the performance measure cannot differentiate between good and poor performing models. Thus, only periods of good quality data should be used in model calibration and the commensurability between observed and model variables should be taken into account (Liu et al., 2009).

2.3 Surrogate modelling in hydrology

The modern simulation models tend to be computationally intensive due to their strive for increased process representations and improved capability of producing simulation results at higher spatiotemporal resolution as part of the endeavour to represent a detailed perceptual model of the real-world hydrological system (e.g. Keating et al., 2010; Zhang et al., 2009). The computational cost becomes even higher and poses strong limitations in terms of practical implementation and computational requirements when these models are coupled with other computationally intensive analyses routines involving MC simulations (e.g. Castelletti et al., 2012). In the past, surrogate modelling methods have been developed in order to make better use of the available, typically limited, computational resources (Razavi et al., 2012). A surrogate model is a low-order and computationally efficient model identified from the original large model and then used to replace it for computationally intensive applications

(Castelletti et al., 2012). Razavi et al. (2012) classify surrogate models as response surface and lower-fidelity models. Response surface surrogates attempt to empirically estimate the model response. On the other hand, lower-fidelity surrogates are simplified physically based simulation models as compared to the original simulation model while preserving the main processes. In the literature, response surface surrogates are also referred as metamodels (e.g. Blanning, 1975; Kleijnen, 2009), emulators (e.g. O'Hagan, 2006) and proxy models (e.g. Bieker et al., 2007).

2.3.1 Response surface surrogate models

Main applications of emulators in hydrological modelling have been for optimization and global sensitivity analysis (Yang et al., 2018). In optimization, the surrogate model of a computationally intensive simulation model is used to approximate the objective function and/or the constraints. Surrogate modelling has been mainly used for optimization purposes, especially in water resources modelling (Razavi et al., 2012). In the area of sensitivity analysis, emulators can be used to obtain faster evaluations of the model's response and thereby allow for applying computationally demanding sensitivity analysis methods to complex and computationally expensive simulation models (Pianosi et al., 2016). Since sensitivity analysis can be efficiently done with the means of emulators, it is often performed in tandem with emulation (Ratto et al., 2012). The use of uncertainty analysis frameworks involving the MC simulation such as the GLUE methodology might be limited due to a substantial computational burden especially when dealing with a distributed hydrological model having high number of parameter dimensions. Previous studies have reported a substantial decrease in computational cost when this framework was coupled with emulators (e.g. Shrestha et al., 2014).

Different emulators have been used in hydrological applications, including: polynomial regression (e.g. Jones, 2001); Gaussian Processes (e.g. Kennedy and O'Hagan, 2001; Yang et al., 2018) and machine learning methods (e.g. Yu et al., 2015). Polynomial emulators have been used for several years mainly due to the availability of well-established techniques for experimental designs, for estimating the coefficients as well as for the interpretation and assessment of the results (Hussain et al., 2002). However, these emulators are poor in providing global fit to smooth response functions of arbitrary shapes, which is particularly important in multi-criteria optimization; and an emulator based on neural networks is one of the promising alternatives for solving such problems (Hussain et al., 2002). In general, some of the criteria that may be considered when selecting a proper emulator include (Hussain et al., 2002; Ratto et al., 2012):

- The functional form of the emulator and its computational complexity
- The set of points used in constructing the emulator. These points are commonly generated from experimental designs such as random sampling and Latin hypercube designs.
- Assessment on the adequacy of the fitted model evaluated through different model diagnostics measures such as confidence intervals, hypothesis tests, degree of fit and cross validation especially for predictions away from observed pairs.

- Ability to gain insights on the behaviour of the simulation model using the fitted emulator.

2.3.2 Machine learning based emulators

Machine learning is a branch of computational algorithms that are intended to emulate human intelligence by learning from the available data (Naqa and Murphy, 2015). Machine learning lies at the intersection of computer science and statistics and at the core of artificial intelligence and data science. It is one of the most rapidly growing technical fields due to the development of new learning algorithms and the increased availability of data and low-cost computation (Jordan and Mitchell, 2015). Conceptually, machine learning algorithms can be viewed as searching through a large space of candidate programs, guided by training experience, to find a program that optimizes the performance metric (Jordan and Mitchell, 2015). Machine learning methods have proven particularly effective for big-data as compared to traditional multivariate statistical techniques (e.g. Li et al., 2011; Bair et al., 2018). In hydrology, machine learning techniques have been used for a number of applications including the prediction of streamflow, peak discharge, snow water equivalent, different water quality variables as well as in emulating the MC simulation in order to overcome the computational burden during uncertainty analysis. Table 2.2 presents a summary overview on some of the previous studies focused on the applications of machine learning methods in hydrology.

Table 2.2. Summary on some applications of machine learning methods in hydrology (updated after Araghinejad (2014))

Field of application	Researchers
Spatial interpolation of forcing variables	Appelhans et al. (2015), Li and Heap (2008), Gilardi and Bengio (2000)
Evapotranspiration modeling	Trajkovic et al. (2003), Kisi and Yildirim (2007)
Flood forecasting	Toth et al. (2000)
Impacts of climate change on water supply	Elgaali and Garcia (2007)
Rainfall–runoff modeling	Garbrecht (2006), Modaresi et al. (2018)
Surface reservoir modeling	Chandramouli and Raman (2001), Neelakantan and Pundarikanthan (2000)
Water quality modeling	Schmid and Koskiaho (2006), Suen and Eheart (2003)
Snow modelling	Tabari et al. (2010), Marofi et al. (2011), Bair et al. (2018)
Uncertainty analysis	Shrestha et al. (2009); Yu et al. (2015)

This section presents a brief description of three machine learning methods, namely, Artificial Neural Network (NNET), Random Forest (RF) and K-Nearest Neighbourhood (KNN). Neural networks have been effectively constructed and applied as emulators in some hydrological studies. For example, Shrestha et al. (2009) implemented a methodology to emulate the time consuming Monte Carlo simulation by using NNET for the assessment of model parameter uncertainty based on the GLUE methodology. Similarly, Yu et al. (2015)

coupled the GLUE framework with NNET models in order to improve the efficiency of uncertainty assessment in flood inundation modelling. Although, RF and KNN were not previously applied as emulators of the MC simulation in hydrological modelling, they can be considered as potential methods given their suitability in mapping non-linear relationships and other merits of these methods.

(a) Artificial Neural Network

An artificial neural network (NNET) constitutes an interconnected and weighted network of several simple processing units called neurons that are analogous to the biological neurons of the human brain (Hsieh, 1993; Tabari et al., 2010). The neurons provide the link between the predictors and the predicted variable and in the case of supervised learning the weights of the neurons, i.e. the unidirectional connection strengths, are iteratively adjusted to minimize the error (Sajikumar and Thandavesware, 1999; Bair et al. 2018). NNET has the capability to detect and self-learn complex and nonlinear relationships between variables (Yu et al., 2015).

A multilayer perceptron is the most common type of neural network used in supervised learning (Zhao et al., 2005; Marofi et al., 2011) and it consists of an input layer in which input data are fed, one or more hidden layers of neurons in which data are processed and an output layer that produces output data for the given input (e.g. Kingston et al., 2008; Senent-Aparicio et al., 2018). Information is received at the input layer in the form of vector of observed data values. The input layer contains a node (neuron) corresponding to each input variable or attribute. The information is then transmitted to each node in the next layer through weighted connections, where the weight determines the strength of the incoming signal. At each node, the weighted information is summed with a bias value and transferred using predefined activation function. The result from the activation function is transmitted to the nodes of a subsequent layer. For example, consider a NNET model with n input neurons ($x_1 \dots x_n$), h hidden neurons ($z_1 \dots z_h$) and m output neurons ($y_1 \dots y_m$). Let i, j , and k be the indices representing the input, hidden and output layers, respectively. Let τ_j and φ_k respectively be the bias for neurons z_j and y_k . Let w_{ij} be the weight of the connection from neuron x_i to neuron z_j and let β_{jk} be the weight of the connection from neuron z_j to y_k . The functions in sample hidden and output neurons can be expressed as (Sudheer and Jain, 2004):

$$y_k = g_A \left(\sum_j^h z_j \beta_{jk} + \varphi_k \right) \quad (2.14)$$

$$z_j = f_A \left(\sum_i^n x_i w_{ij} + \tau_j \right) \quad (2.15)$$

where g_A and f_A are activation functions, which are usually continuous and bounded functions such as the sigmoid function (Eq. 2.16).

$$f(s) = \frac{1}{1 + e^{-s}} \quad (2.16)$$

where s in this case represents the expression within the bracket of equation 2.14 or 2.15.

An artificial neural network model can be trained to learn the general relationship between input-output pairs using a set of training samples from a given problem. Appropriate values of the connection weights and bias terms need to be used in order for the NNET model to perform the desired function (e.g. Kingston et al., 2008). Both weight and bias are hyper-parameters of NNET that need to be estimated through calibration. During the calibration (training) phase, the model outputs are compared against the target values and the weights and biases are iteratively adjusted to minimize the error. This is done through backpropagation of the errors computed at the output layer to each neuron in the hidden layers until optimal values are obtained (e.g. Senent-Aparicio et al., 2018). The knowledge acquired about the problem domain during the training process is encoded within the NNET in the form of the network architecture (e.g. number of hidden layers and neurons) and a set of model hyper-parameters (e.g. weights) (Sudheer and Jain, 2004). As such, each NNET output results from a distributed information processing structure instead of a direct evaluation as opposed to the standard equation-based models.

Figure 2.12 shows a standard illustration of a neural network model applied on sample parameter sets (inputs) and the corresponding response surface, i.e. pLoA values, as outputs based on an MC simulation using the PT_GS_K model. The first layer represents an input layer with model parameters as independent variables (I1 to I8). Generally, the number of neurons comprising the input layer is equal to the number of the independent variables. In this example, NNET is applied for regression purpose and hence the output layer has only a single response variable (O1) which is shown in the far right layer (labelled as pLoA). The NNET model has a single hidden layer with ten neurons labelled as H1 through H10. The lines (usually accompanied by weights) link the neurons in the hidden layer with the input and output layers. B1 and B2 are the bias terms applicable to each neuron in the hidden and output layers, respectively.

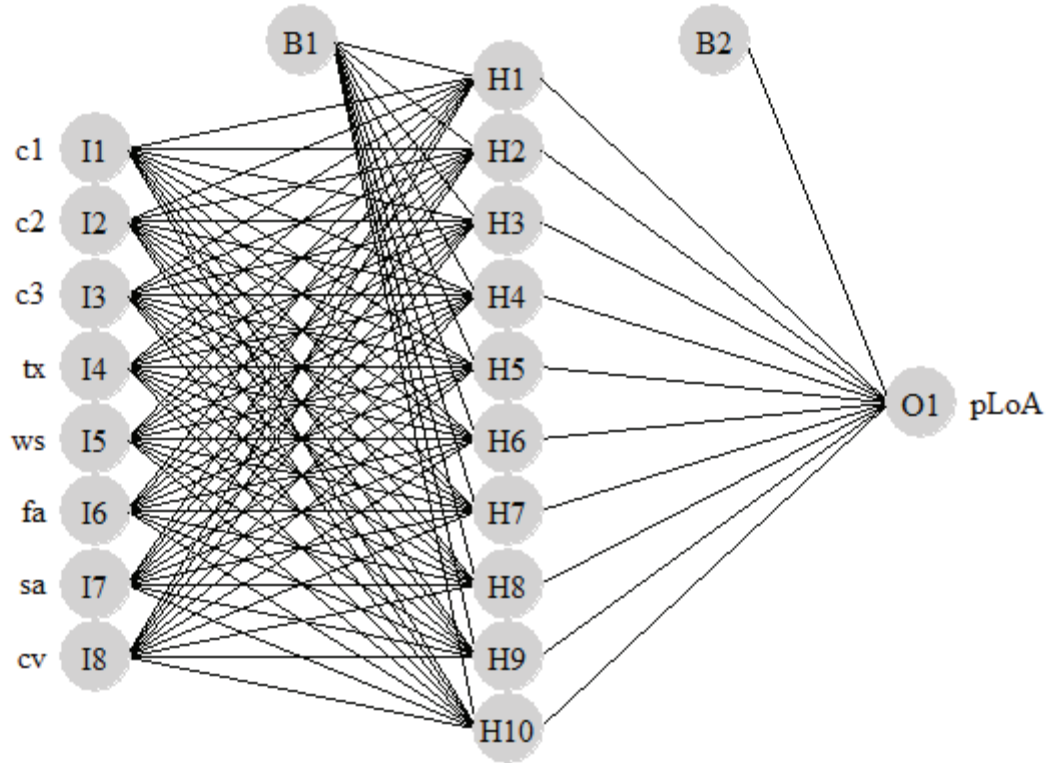


Figure 2.12: An illustration of a neural network model with an input layer containing neurons I1 to I8 and an output layer containing a single neuron (O1) as well as a hidden layer with neurons H1 to H10 applied on sample pairs of parameter sets and pLoA values.

(b) K-Nearest Neighbours

K-Nearest Neighbours (KNN) approach is one of the oldest machine learning algorithms that originated from pattern recognition work (Cover and Hart, 1967). KNN has been widely used in the field of nonlinear dynamics as a standard method in time series data estimation due to its good ability to estimate under nonlinear conditions (Hu et al, 2013). KNN regression is a nonparametric method where the information extracted from the observed datasets is used to predict the variable of interest without defining a predetermined parametric relationship between the predictors and predicted variables (Modaresi et al., 2018).

KNN uses the K-closest samples from the training dataset to predict a new sample and it is a non-linear method whose prediction solely depends on the distance of the predictor variables to the closest training dataset known to the model (Appelhans et al., 2015). Similarity measures such as the Euclidean and Manhattan distance are commonly used to compute the distance between the new dataset and each training dataset. As illustrated in the following examples, another advantage of using the KNN regression is that it follows a simple algorithm (e.g. Hu et al, 2013; Araghinejad, 2014; Modaresi et al., 2018). Suppose we have m predictors for the current condition, r , with unknown value of the target variable, i.e. $X_r = \{x_{1r}, x_{2r}, x_{3r}, \dots, x_{mr}\}$ and an $n \times m$ matrix of predictors with their corresponding already observed target variable values for the condition, t , i.e. $X_t = \{x_{1t}, x_{2t}, x_{3t}, \dots, x_{mt}\}$, with $t = 1, 2, \dots, n$. A Euclidean distance function (Eq. 2.17) can be used to calculate n distances (D_{rt}) between the current predictor and the observed predictors.

$$D_{rt} = \sqrt{\sum_{i=1}^m (x_{ir} - x_{it})^2} \quad , \quad t = 1, 2, \dots, n \quad (2.17)$$

The K-nearest neighbours, i.e. K sets of predictors and target variables (X_k, Y_k) with the lowest value of D_{rt} are selected and used to estimate the unknown target variable value (Y_r) using Equation 2.18 based on Y_k and a kernel function ($f(D_{rj})$) associated with the K-nearest neighbour (Eq. 2.19). The number of neighbours (K) is estimated through model training.

$$Y_r = \sum_{j=1}^k f(D_{rj}) \cdot Y_j \quad (2.18)$$

$$f(D_{rj}) = \frac{1/D_{rj}}{\sum_{j=1}^k 1/D_{rj}} \quad (2.19)$$

where $f(D_{rj})$ represents the Kernel function (weighted inverse distance) of the K nearest neighbours which is estimated based on the magnitude of the distances (D_{rj}).

As an alternative to direct use of D_{rj} , the kernel function can also be estimated from the order of the neighbours after sorting them in an ascending order (Lall and Sharma, 1996). As such, the neighbours with higher distance would get higher orders and thereby lower contribution to the final output. The following figure illustrates schematic overview of the KNN algorithm for a condition with K=3.

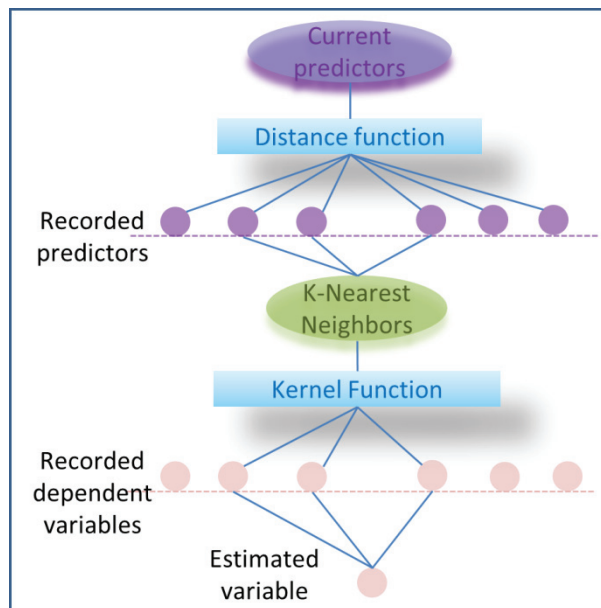


Figure 2.13: Schematic overview of the KNN algorithm for the condition K=3 (after Araghinejad (2014)).

(c) Random forest

Random forest (RF) is a version of the Bagged (bootstrap-aggregated) trees algorithm (Breiman, 2001). It is an ensemble method whereby a large number of individual trees are grown from random subsets of predictors and samples, providing a weighted ensemble of trees (Bair et al., 2018). Bagging is a successful approach for combining unstable learners;

and since RF combines bagging with a randomization of the predictor variables used at each node, it yields an ensemble of low correlation trees (Li et al., 2011, Appelhans et al., 2015). RF is also less sensitive to non-important variables since it implicitly performs variable selection (Okun and Priisalu, 2007).

The decision tree forms the basis for all regression trees which in turn are the building blocks of more complex tree-based models such as random forests and gradient boosting machines (e.g. Quinlan, 2006). A basic regression tree divides a dataset into smaller groups and fits a simple model (constant) for each subgroup. The partitioning is achieved by successive binary partitions based on the different predictors. The predicted constant value is based on the average response values for all observations that fall in that subgroup. Figure 2.14 shows a simple regression tree constructed using the regression part of the classification and regression tree (CART) method (Breiman et al., 1984). The regression tree was fitted to a sample dataset of the model parameters and response surface (pLoA) pairs generated from the MC simulations using the PT_GS_K model. This regression tree predicts a response variable (pLoA) based on the model parameters. All the observations that pass through this tree are assessed at a particular node and proceed to the left if the answer is “yes” or to the right if the answer is “no”. First all observations with c_2 value less than 0.52 go to the right branch and all other observations proceed to the left branch. The boxes contain information about the average value of the response variable and the number of observations (n) and percentage out of total number of observations that fall under this category. The first information box (1), for example, shows the average pLoA for all observations (0.081) and the number of observations (801) which is 100% of the total number of observations used for training the model. Those with c_2 value of less than 0.52 and with c_3 value of less than -0.12 proceed to the left, while those with greater than or equal to -0.12 proceed to the right. The partitioning continues until the final branches lead to the terminal nodes or leaves which contain the predicted response values.

The partitioning of the variables is done in a top-down fashion, i.e. a partition performed at earlier node in the tree will not change based on partitions at later nodes. The partitions are made by searching every value of each input variable to find a value of the predictor and the split value that partitions the data into two regions in such a way that the sum of squares error are minimized. For further details on the methodology of growing a regression tree, including deciding on splits and on finding the balance between the depth and complexity of the tree (cost complexity criterion) the reader is referred to materials focused on algorithms of the CART approach (e.g. Breiman et al., 1984).

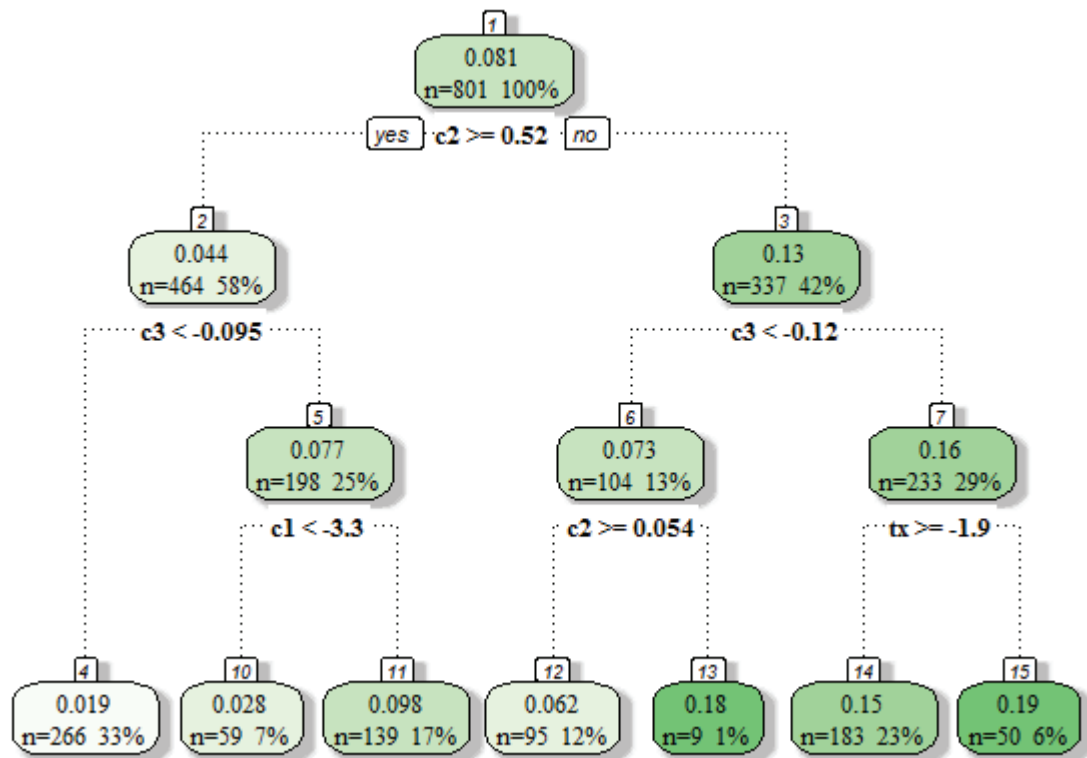


Figure 2.14: Sample regression tree for predicting a response surface (pLoA) based on the model parameters (c_1 , c_2 , c_3 , and tx).

When building the machine learning models, the dataset is commonly divided into two sets, i.e. into training and test dataset. The training dataset is used to identify the model structure and value of its hyper-parameters, while the test dataset is used to evaluate performance of the developed model through comparison of the predicted responses against new observations. Figure 2.15 shows sample cross-validation and bootstrap analyses results when estimating the hyper-parameters of the final machine learning models as used in Paper II. For the RF model (a), the optimal number of randomly selected predictors when forming each split (m_{try}) was 7; and for KNN (b), the optimal value of nearest neighbours (k) used for the final model was $k=10$. For NNET (c) two hyper-parameters were optimized using the training dataset, i.e. the weight decay and number of neurons in the hidden layer (hidden units or size). The final values used for this model were a weight decay of 0.001 and hidden units of 10.

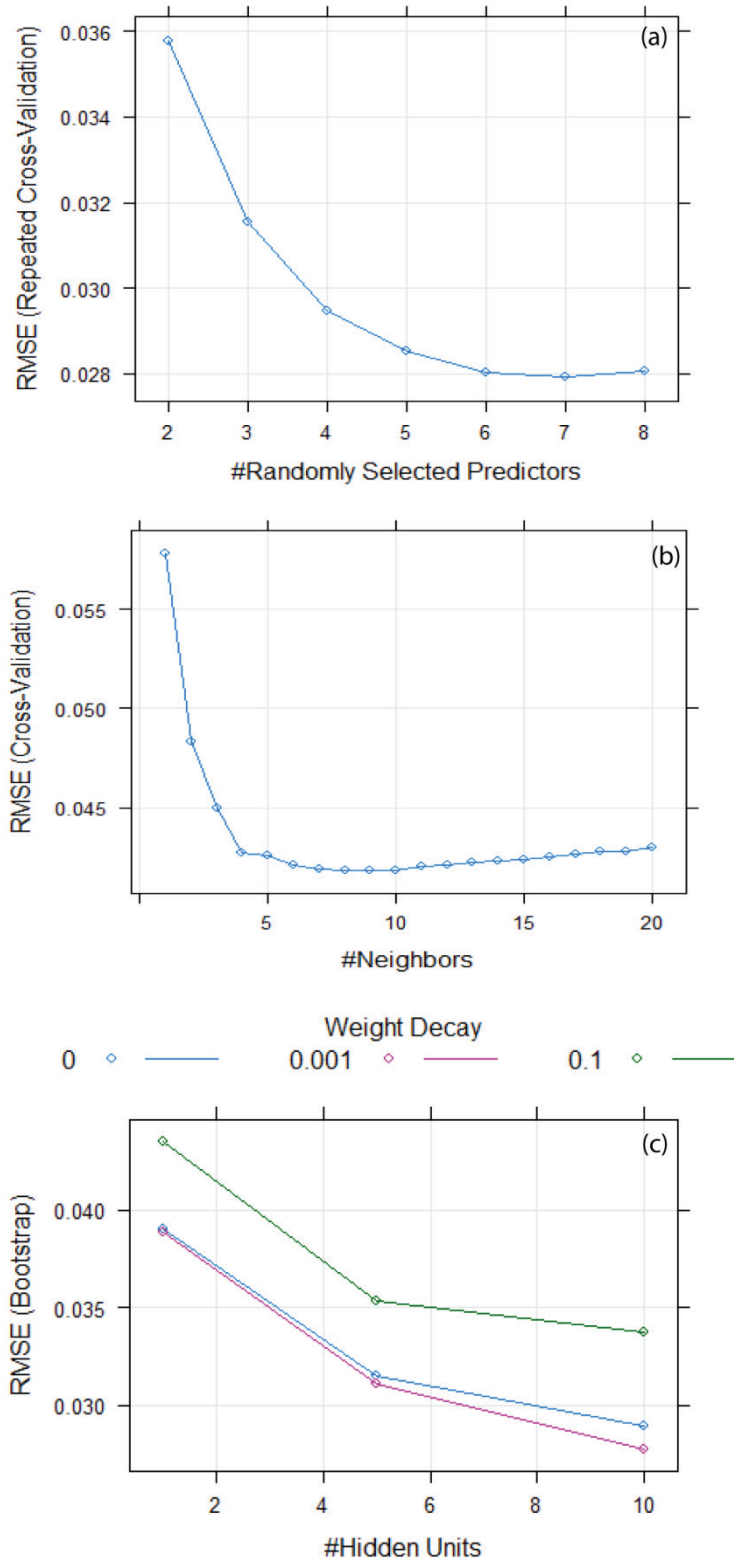


Figure 2.15: Sample bootstrap and cross-validation based estimates for hyper-parameters of the three machine learning models used as emulator of the MC simulation (Paper II), i.e. for RF (a), KNN (b), and NNET (c).

2.4 Snow distribution modelling and snow data assimilation into hydrological models

Snow plays a significant role especially in snow dominated catchments including in summer water supplies and flood event generation as well as in changing the interface between the atmosphere and the land surface through its high albedo characteristics and through thermal insulation of the soil from the atmosphere (Houser et al., 2012). Thus, improved estimate of snow state and its spatiotemporal distribution can greatly increase the accuracy in hydrological predictions (Houser and Walker, 2005). This section presents an overview on the main processes and controlling factors in snow distribution as well as some modelling and ensemble-based snow data assimilation techniques.

2.4.1 Main processes and controlling factors in snow distribution

Spatial heterogeneity of snow cover and snow depth is the result of complex interactions between topographic, climatic and land cover parameters and processes prevalent at different spatiotemporal domains. Preferential deposition, saltation, suspension as well as melting and sublimation constitute major processes contributing to heterogeneous snow distribution. The prevalence of these processes, however, varies both in space and time scales. For instance, saltation was observed to be more important transport process at higher spatial resolution (Mott and Lehning, 2010). The role of melting in snow distribution variability is evinced by the increasing standard deviations and decreasing means during the melt season (e.g. Winstral and Marks, 2014). The spatial variability of snow depth at local scale is mainly controlled by wind-induced snow redistribution and sublimation of the wind-borne snow particles. Whereas large scale snow depth variability is a function of the general climate as well as topographic influences on precipitation (Liston, 2004).

Although, the variability of meteorological input plays a significant role, it cannot always explain the variability of snow cover patterns (Frey and Holzmann, 2015). Local precipitation amounts are strongly affected by the interaction of the terrain with the local weather and climate (Beniston et al., 2003). Snow transport and sublimation rates increase with increasing wind speed and are highly sensitive to wind speed variations caused by the variability in topography and roughness (Essery et al., 1999). Generally, wind plays a dominant role in temporal and spatial distribution of snow cover during winter (Mott and Lehning, 2010). However, it is worthy to note that in regions with greater variability in wind directions, deposition and scour regions might vary and thereby cancelling out wind effect (Winstral and Marks, 2014).

Snow distribution heterogeneity is further amplified by the variable melt rate in response to the local radiation and energy balance (Mott et al., 2011; Winstral and Marks, 2014). Solar radiation has dual effect on snow distribution properties. It can increase snow loss via sublimation; and it can produce snowpack surface crusts and thereby increasing the surface density which ultimately reduces the amount of surface snow available for wind transport (e.g. Hiemstra et al., 2006). Air temperature is highly related to snow depth via its effect on the fraction of precipitation that falls as snow; snowmelt as well as on snow density. The effect of

total precipitation on snow distribution is thus conditional on temperature and temperature dependencies (Dyrrdal et al., 2013).

Topography and land cover are also important controls of snow distribution. Terrain parameters such as slope, elevation and curvature are reported to play decisive role in spatial snow distribution. On steep terrain, snow tends to slide off and accumulate on more gentle terrain. Elevation is often related to snow water equivalent (SWE), in that higher elevations have increased SWE (Kerr et al., 2013). Concave features generally show lower wind speed than convex features. As a result, they experience less erosion and higher accumulation (Dadic et al., 2010). The type and density of vegetation control snow distribution through their effects on surface roughness, wind speed, surface energy exchange, and snow interception (Erxleben et al., 2002). For instance, Rice and Bales (2010) have observed lowest snow depths being associated with canopy densities greater than 60% closure in forested areas.

Figure 2.16 presents a sample spatiotemporal variability of observed SWE (m w.e.) in response to the interaction between various physiographic and climatic factors prevalent in the Nea catchment (Norway). From this figure it can be noticed that, the amount of SWE varies from one place to another in a given observation date. The SWE distribution in some part of the catchment tends to follow the topographic pattern which signifies the prevalence of topography-induced preferential deposition. Comparison of the snow distribution between the observation dates shows a general increase in the amount of SWE with time during the accumulation period. However, the rate of accumulation varies from one place to another.

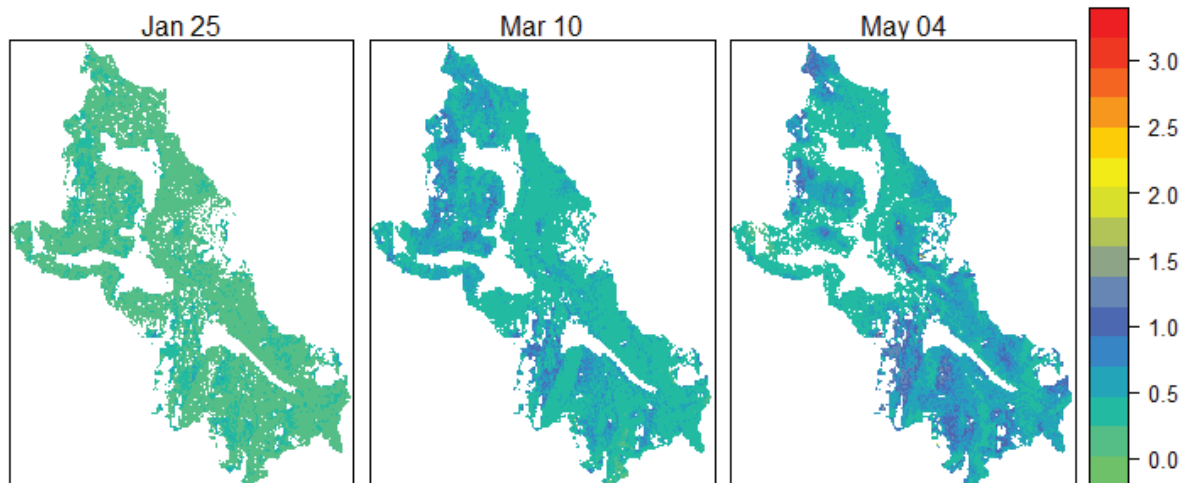


Figure 2.16: Spatiotemporal variability of SWE (m w.e.) in the Nea catchment in response to the interaction between the dominant processes and various physiographic and climatic factors. The observed satellite SWE scenes from Globesar (2018) were retrieved on January 25 (left), March 10 (middle), and May 04 (right) of year 2015. White colour denotes no data.

Studies conducted based on regression-tree models have shown that the relative importance of terrain parameters as explanatory variables of SWE varies both in space and time (e.g. Webster et al., 2015; Kerr et al., 2013). A similar result was also obtained in this work when analyzing the relative importance of different physiographic variables in explaining the observed SWE. The analysis was conducted using over 10, 000 observations

(aggregated at a spatial scale of 100m) from the Nea catchment at four observation dates distributed in different months. The physiographic variables considered include: easting, northing, elevation, slope, aspect and curvature as well as the average degree of sheltering index (SX) calculated at three potentially dominant wind directions in the study domain i.e. 135° (sx135), 180° (sx180) and 255° (sx225). Simulated SWE (simSWE) generated from the PT_GS_K model was also included as covariate in an attempt to indirectly take into account for impact of the dynamic meteorological variables.

The relative importance of these covariates at four different periods of the snow accumulation season was analyzed using a machine learning method, i.e. Random Forest (Fig. 2.17). The result shows that, although the relative importance of curvature was consistently low in all observation dates, the other variables have shown considerable variability in their relative importance between the observation dates. For example, in January 25 (a), aspect was relatively the most important variable, while in March 10 of same year (2015) this variable was downgraded to the fourth level of importance preceded by easting, elevation and northing (b). In another analysis, clear sky radiation (rad) was initially considered as a potential explanatory variable, but it was excluded from the preceding analysis based on the result obtained from a preliminary assessment using a correlation matrix (Fig. 2.18). It shows that this parameter is highly correlated to the sheltering indices (SX variables) with a Pearson correlation value of -0.7 to -0.98. A considerable degree of interaction between some of the other terrain parameters can also be noticed from this figure.

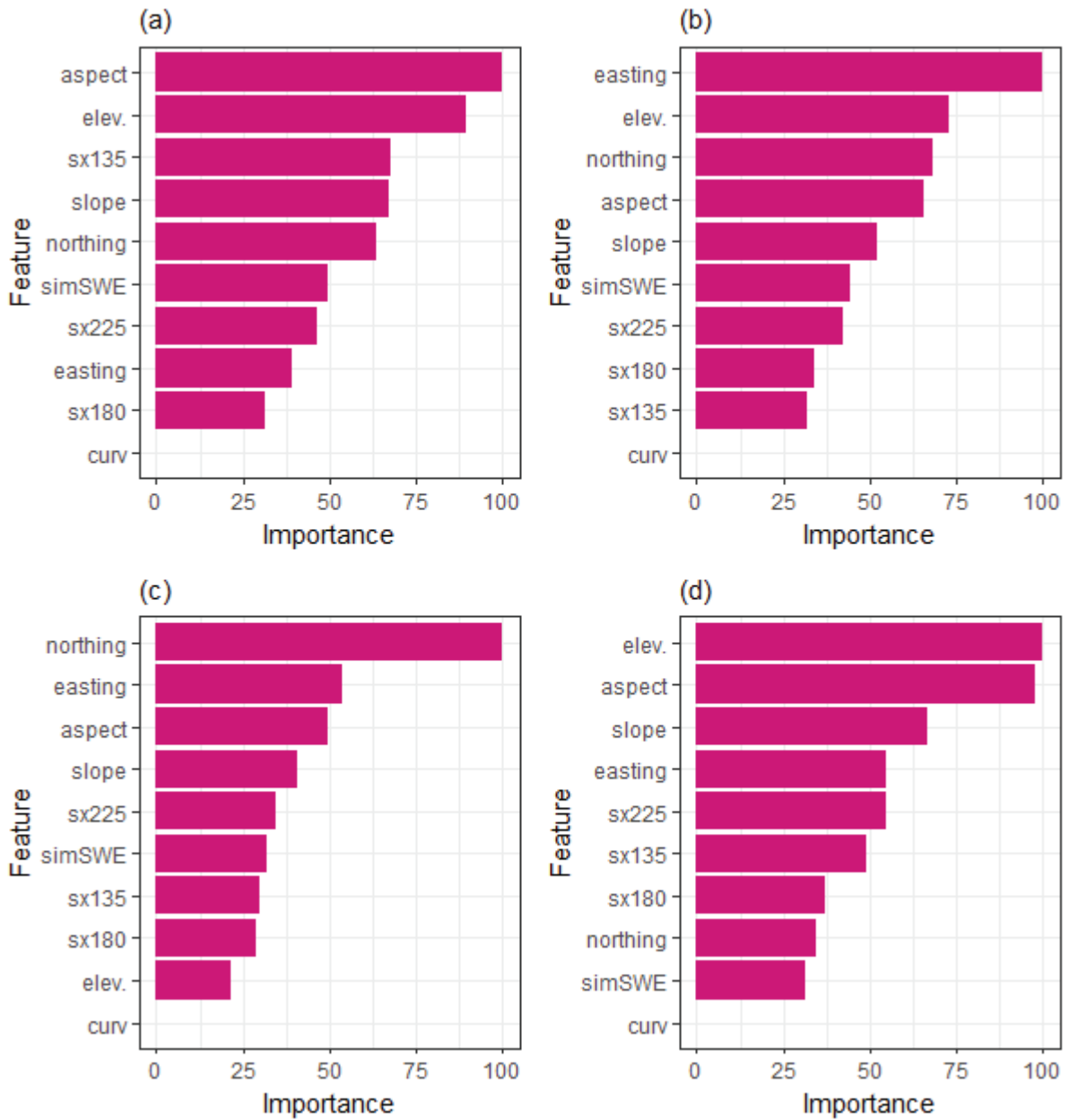


Figure 2.17: Temporal variability in relative importance of terrain parameters to explain the spatial variability in snow water equivalent (SWE) based on the analysis using the Random Forest machine learning method. The sub-plots show analysis results for January 25 (a), March 10 (b), April 12 (c), and May 04 (d) of year 2015.

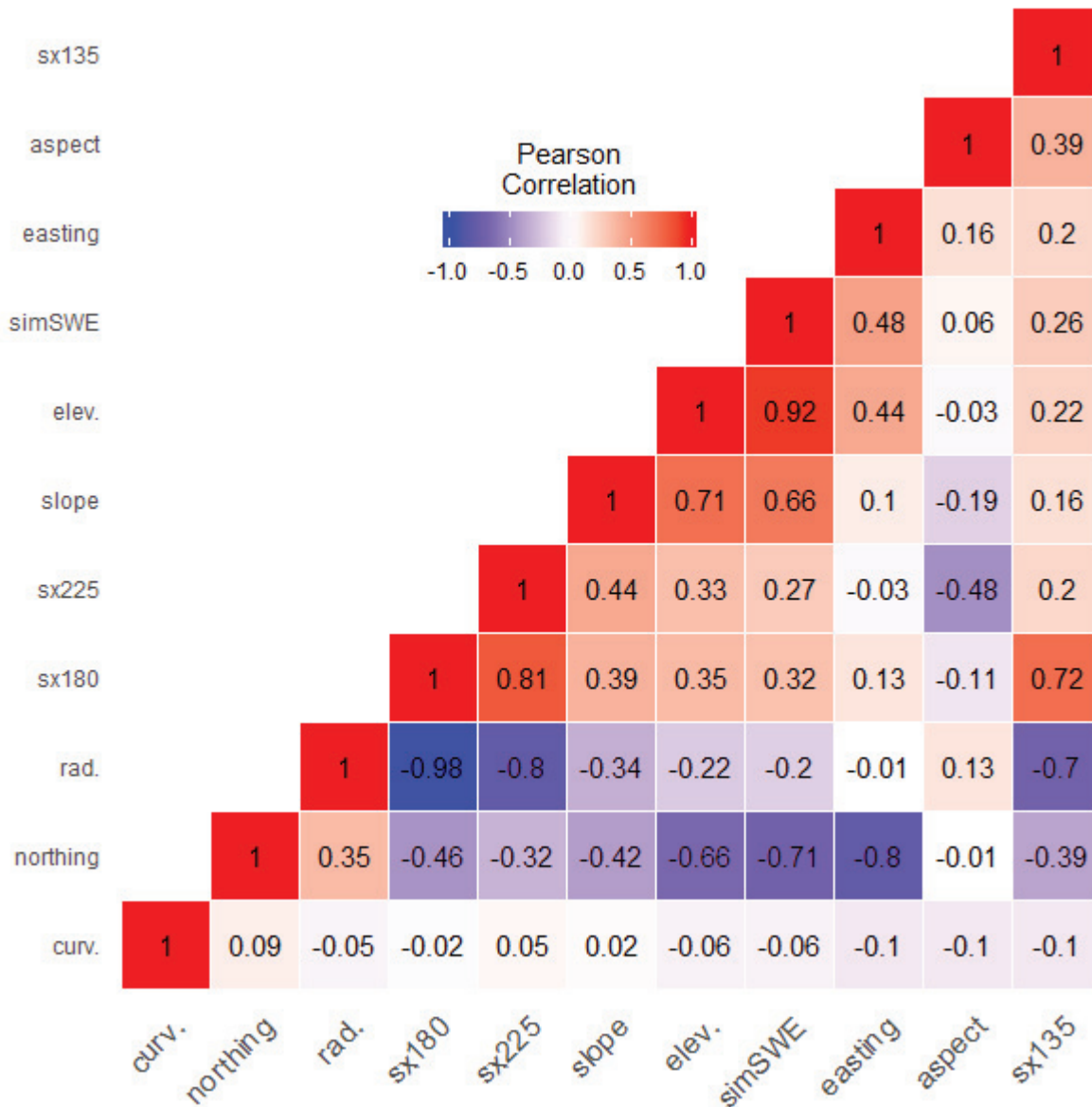


Figure 2.18: Pearson correlation matrix of potential explanatory variables for the spatial distribution of snow water equivalent.

2.4.2 Snow distribution modelling approaches

Over the past years, snow-transport models capable of simulating snow distribution processes and the resulting snow-depth patterns have been developed with wide range of complexity (Liston et al., 2007). Models for simulating snow distribution range from simple conceptual to complex, physically based ones. Physically based models attempt to consider all energy fluxes and therefore have huge meteorological data demand; and usually high spatial resolution requirement (Frey and Holzmann, 2015). As a result, they are computationally expensive. Complex snow-physics models are reported to take much longer computation time than temperature index models (e.g. Magnusson et al., 2015). Conceptual snow models, on the other hand, involve partial description of the processes, where the snow distribution effect of topography, wind, land cover, and/or other important parameters are parameterized.

SnowModel and SNOWPCK have been widely used as physically based energy balance snow models. SnowModel (Liston and Elder, 2006) is a physically based distributed snow-evolution modelling system; and it includes sub-models dealing with meteorological forcing conditions, surface energy exchanges, snow depth calculations as well as snow redistribution by wind. SnowModel is designed to run on grid increments of 1 to 200 m and temporal increments of 10 min to 1 day. SnowTran-3D is the sub-model responsible for snow redistribution by wind. Its primary components include: wind flow forcing field, wind-shear stress on the surface, transport of snow by saltation and turbulent suspension, sublimation of salting and suspended snow as well as the accumulation and erosion of snow at the snow surface. However, for spatial increments greater than 100m the redistribution components of the model become negligible and only the simulated sublimation is significant (Liston et al., 2007). SNOWPACK (Bartelt and Lehning, 2002) is a one-dimensional physical snowpack model which numerically solves the partial differential equations governing the mass, energy, and momentum conservation within the snowpack using the finite-element method. The model has process algorithms dealing with snowfall, rain, short wave and long wave radiation, sensible and latent heat exchange, wind pumping as well as snow drift (Lehning et al., 2002). Although SNOWPACK was primarily designed to handle problems related to avalanche warning, it has been applied to study ecological and hydrological problems involving snow cover. Certain processes such as those describing new snow density and albedo are statistically derived based on local data. According to Liston et al. (2007), SNOWPACK uses a mesoscale meteorological model to simulate the wind field. This approach, however, requires considerable computational energy for simulating the wind fields. The snow evolution simulation capability of the model is thus limited to not more than few days.

Although physically based energy balance models more properly account for the processes determining melt, the degree-day method has been in use in many variants for more than a century (Hock, 2003). According to Frey and Holzmann (2015), many of the conceptual models use a temperature index approach to model snow melt and snow accumulation. However, when using this modelling approach snow accumulates as long as the air temperature does not rise above a certain threshold regardless of any other processes that may lead to snow melt like radiation or turbulent fluxes of latent energy. This may lead to an intensive computational accumulation of snow in certain areas; though they barely exist in nature. In the past, various methods have been introduced to cope with this drawback. The HBV (Bergström, 1976) and COSERO (Nachtnebel et al., 1993) models are typical examples of conceptual hydrological models where a temperature index approach is used to estimate snow accumulation and snow melt. The PT_GS_K model (Section 3.2) also belongs to this family of conceptual models. In these models, snow distribution within elevation band or a sub-grid, in the case of distributed system, is described using simple two parameter statistical distributions such as the log-normal and gamma distributions. This method is reported as having the potential to indirectly consider the influence of curvature, shelter, vegetation and elevation on snow distribution (Frey and Holzmann, 2015). Temperature-index methods are most common in snow and glacier melt modeling due to their good performance, low data requirements, and simplicity (Hock, 2003).

Statistical and machine learning based snow distribution models attempt to relate snow distribution with physiographic parameters (e.g. slope, elevation, curvature and land cover), weather data (e.g. precipitation, wind speed and direction) and/or their surrogate variables. For example, Winstral et al. (2002) derived terrain-based degree of shelter/exposure parameter (SX) and assessed its effect on statistical snow distribution model. A significant improvement over the original model was observed after adding SX as a predictor. Revuelto et al. (2014) studied the relationship between eight topographic parameters and snow depth. Introducing SX into their statistical model improved the result; and this parameter was identified as one of the key variables explaining snow distribution. Similarly, Gislén et al. (2015) used such variables as the coefficient of variation of SX, grid-cell elevation, and mean maximum snow depth to characterize small scale snow distribution for application in a regional permafrost model. Machine learning methods are also increasingly used for snow distribution modelling with improved computation power and availability of spatial snow data (e.g. Tabari et al., 2010; Marofi et al., 2011).

Some hybrid snow distribution modelling approaches involve combined use of physically based model decisions and observed snow distribution or surrogate variables. Sturm and Wagner (2010) used a hybrid approach of snow distribution modelling based on a physical model, SnowModel, and annual snow depth surveys. Their study was based on the observation that snow distribution patterns and thereby relative snow depth are similar from year to year since they are largely controlled by the interaction of topography, vegetation, and consistent synoptic weather patterns. The assimilation of observed snow pattern into the physical model was realized by calibrating wind related model parameters and subsequent comparison of the simulated SWE distribution against the average observed SWE patterns for the catchment. An improvement in model output accuracy by 60% was attained through use of the hybrid approach as compared to outputs from SnowModel alone. In another similar study, Liston and Hiemstra (2008) used observed SWE data to periodically constrain SnowModel simulations to match observed SWE both in time and space. To obtain the required model constraint, they modified precipitation and melt parameters based on correction factors derived from the deviation between observed and simulated SWE values. Farinotti et al. (2010) used a simple snow accumulation model and a temperature index melt model to match the modelled melt-out pattern evolution to the pattern monitored during an ablation season through terrestrial oblique photography. A snow multiplier matrix in the accumulation model was tuned to optimize the spatial agreement between simulated and observed melt out pattern. The simulated SWE was observed to have accuracy comparable with that obtained with an inverse distance interpolation of the point measurements. Snow data assimilation techniques using different schemes such as the Kalman Filter and its variants can also be labeled under the hybrid approach.

2.4.3 Ensemble-based snow data assimilation techniques

Model predictions are generally imperfect due to the inherent uncertainties for example in model forcing data, model parameters, and model structure. The use of techniques with efficient capability to extract and assimilate the information contained in the observed data through the model identification and prediction processes is one of the main areas where

actions can be taken to reduce the uncertainty in hydrological predictions (Liu and Gupta, 2007). Hydrologic data assimilation techniques aim to improve model predictions by combining the imperfect hydrologic process knowledge embodied in a hydrologic model and the uncertain information gained from observations (e.g. Han et al., 2012). As such, data assimilation is used to not only update the hydrological model states that optimally combine model outputs with observations, but also to assess observational and model errors (Moradkhani, 2008).

Ensemble-based filtering and smoothing techniques have been widely used in solving data assimilation problems (e.g. Liu et al., 2012). The formulation of a data assimilation problem as filtering or smoothing mainly depends on the time period when the state estimation is needed (Carrasi et al., 2018). The Monte Carlo approach forms the basis for the data assimilation schemes referred to as ensemble-based methods (e.g. Evensen, 2009). This section briefly presents two commonly used ensemble-based data assimilation frameworks, one based on the filtering and another based on the smoothing techniques.

(a) The Ensemble Kalman Filter

Ensemble based filtering techniques are widely used for real-time applications (e.g. Piazzini et al., 2018). The filtering technique involves sequential processing, in which the observational data are sequentially used to update the model state as they become available (Bengtsson et al., 1981). Figure 2.19 shows a simplified graphical illustration of an ensemble filtering technique (e.g. McLaughlin, 2002; Carrasi et al., 2018). At the initial time step (t_0) an ensemble of n random replicates is generated for a given uncertain input variable that appears in the state equation. The ensemble generated based on the initial condition is used to generate a corresponding ensemble of state vector replicates (x^j) at the first measurement time (t_1). The sample statistics such as marginal probability density, mean, and covariance are calculated from the propagated state vector ensemble for use in computation of the updated state conditioned on the observation at the present time step. The sequence of propagation and update is repeated for each measurement time, yielding a propagated state $x^j(t_i|Z_{i-1})$ and an updated state $x^j(t_i|Z_i)$ at the time t_i .

Evensen (1994) introduced a Monte Carlo based Kalman filter, the Ensemble Kalman Filter (EnKF), which was extensively used in different hydrological data assimilation studies. The Kalman filter is a data assimilation algorithm that produces optimal weighting between a modeled and observed state with due consideration to errors in the model and observations (e.g. Maybeck, 1990). Variants of this data assimilation technique including the EnKF have been used for snow updating in hydrologic models resulting to considerable improvement in the efficiency of streamflow forecasts (e.g. Slater and Clark, 2006). The EnKF is especially useful for high dimensional data and it represents distribution of the system state using a random sample (an ensemble), from which the sample covariance is computed (Evensen, 1994). Unlike the Kalman filter, the EnKF does not require that the state equation be linearized during the propagation step and the model input errors generated during this step need not be additive. However, it relies on an implicit Gaussian assumption to derive the measurement update and it can be computationally demanding if the number of states is large (McLaughlin, 2002).

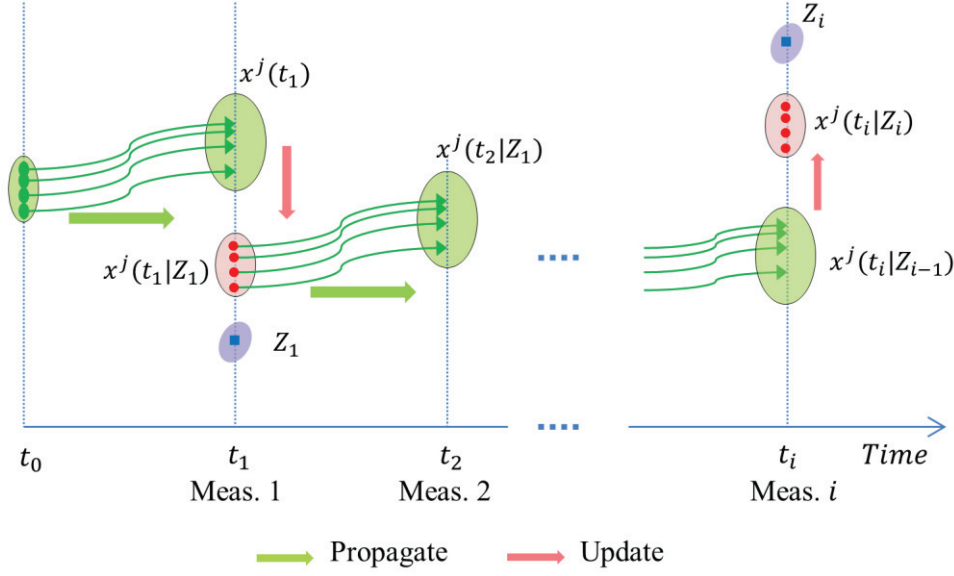


Figure 2.19: Graphical illustration of ensemble filtering. The ensembles are propagated forward to the next measurement time (t) and the variable of interest (x^j) is subsequently updated with due consideration to the new measurement (Z) (after McLaughlin, 2002; Carrassi et al., 2018).

The EnKF is based on a Monte Carlo approximation of the Kalman filter and its basic formulation can be described as follows (e.g. Clark et al., 2008; Mandel, 2009). The ensemble matrix (prior ensemble) at a given time step can be represented as:

$$\mathbf{X}^f = [\mathbf{X}_1^f, \dots, \mathbf{X}_N^f] \quad (2.20)$$

where \mathbf{X}^f is an $n \times N$ matrix of n model states whose N columns are a sample from the prior distribution. Given the ensemble matrix \mathbf{X}^f , the ensemble mean ($\bar{\mathbf{X}}^f$) and an estimate of the model error covariance (\mathbf{P}^f) are computed as:

$$\bar{\mathbf{X}}^f = \frac{1}{N} \sum_{i=1}^N \mathbf{X}_i^f \quad (2.21)$$

$$\mathbf{P}^f = \frac{\mathbf{A}\mathbf{A}^T}{N-1} \quad (2.22)$$

where, \mathbf{A} and the superscript T respectively denote the ensemble of anomalies and a matrix transpose. The ensemble anomaly for the i^{th} ensemble member is $\mathbf{A}_i^f = \mathbf{X}_i^f - \bar{\mathbf{X}}^f$, and the ensemble of anomalies is represented as:

$$\mathbf{A} = [\mathbf{A}_1^f, \dots, \mathbf{A}_N^f] \quad (2.23)$$

The model states in each ensemble member are updated (\mathbf{X}_i^a) with due consideration to the Kalman gain (\mathbf{K}) and \mathbf{y}_i which is an $m \times 1$ vector of observations (where m is the number of observations) (Eq. 2.24).

$$\mathbf{X}_i^a = \mathbf{X}_i^f + \mathbf{K}(\mathbf{y}_i - \mathbf{H}\mathbf{X}_i^f) \quad (2.24)$$

$$\text{with, } \mathbf{K} = \mathbf{P}^f \mathbf{H}^T (\mathbf{H}\mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1}$$

where, \mathbf{H} denotes to the $m \times n$ operator that is used to map the model states to observation space and \mathbf{R} is the $m \times m$ observation error covariance matrix.

(b) Ensemble smoothing using the particle batch smoother

Unlike to the filtering technique, in smoothing, all observations are assimilated in a single step (Li et al., 2017). This technique is commonly used in retrospective analysis after the observations have been collected. It is very useful in generating reanalysis and in estimating model parameters (Carrasi et al., 2018). The ensemble smoothing techniques have the advantage of low computational cost coupled with the flexibility in running independent of the forward model (Li et al., 2017; Bailey and Baù, 2010; Emerick and Reynolds, 2013). The particle batch smoother (Pbs) is one of the commonly used ensemble smoothers. Similar to the ensemble filtering technique described above, Pbs also uses an ensemble of independent randomly drawn Monte Carlo samples (particles). The procedure involves generating ensemble of model realizations by running the model over the full seasonal cycle followed by the assimilation of all observations at once. The Bayes theorem forms the basis for estimating the updated relative importance or weight (w_j) of each ensemble member (Eq. 2.25).

$$w_j = \frac{p(\mathbf{y}|\hat{\mathbf{X}}_j)p(\hat{\mathbf{X}}_j)}{\sum_{j=1}^{N_e} p(\mathbf{y}|\hat{\mathbf{X}}_j)p(\hat{\mathbf{X}}_j)} \quad (2.25)$$

where $p(\mathbf{y}|\hat{\mathbf{X}}_j)$ refers to the likelihood of the observations (\mathbf{y}) given the state value of the j^{th} ensemble member ($\hat{\mathbf{X}}_j$). $p(\hat{\mathbf{X}}_j)$ and N_e respectively denote prior values and the number of ensemble members.

In Pbs, all ensemble members are implicitly assigned equal prior weights of $1/N_e$ (Margulis et al., 2015). When using the Gaussian likelihood, this yields (Margulis et al., 2015; Aalstad et al., 2018):

$$w_j = \frac{\exp[-0.5(\mathbf{y} - \hat{\mathbf{Y}}_j)^T \mathbf{R}^{-1}(\mathbf{y} - \hat{\mathbf{Y}}_j)]}{\sum_{j=1}^{N_e} \exp[-0.5(\mathbf{y} - \hat{\mathbf{Y}}_j)^T \mathbf{R}^{-1}(\mathbf{y} - \hat{\mathbf{Y}}_j)]} \quad (2.26)$$

where \mathbf{y} and $\hat{\mathbf{Y}}_j$ respectively refer to $N_o \times 1$ vector of perturbed observations and predicted values for the j^{th} particle in the $N_o \times N_e$ matrix ($\hat{\mathbf{Y}}$). \mathbf{R} denotes $N_o \times N_o$ diagonal observation error covariance matrix. N_o refers to the number of observations.

Pbs assumes that both the prior and posterior states of the particles remain the same (Margulis et al., 2015). It updates the particle weights in such a way that particles with their predictions closer to the observations get higher weight than those with farther from the observations. The median and prediction quantiles are estimated from the cumulative distribution of the sorted values of the target variable and their associated weights.

Chapter 3

The hydrological model and data

In this PhD work different software packages and applications were used in conducting the various analyses as well as in pre- and post-processing of the relevant hydrologic, climatic, and physiographic datasets. Shyft was used as the main hydrological modelling toolbox and the R statistical software in general and its CART package in particular were used for general statistical analysis and in building the machine learning models (Appendix B). The SAFE toolbox was used as a basis for conducting sensitivity and uncertainty analysis. ArcGIS and the GDAL library in Python were mainly used for spatial data analysis and presentation. For example, ArcGIS was used to setup the grid-cells (when Shyft was run in distributed setting) and to extract the relevant physiographic and hydrological data of each grid-cell through overlay analysis. A number of Matlab and Python based algorithms were also written in this work in order to implement the different theoretical formulations including the newly introduced fuzzy logic based and the existing data assimilation schemes (Appendices A and C). Various algorithms were also implemented to carry out auxiliary tasks such as in input data preparation and post processing. Similarly, different climatic, physiographic, and hydrologic datasets were used as forcing and validation datasets. The hydrological model and the main datasets used in this study are briefly described in this section.

3.1 The Shyft hydrological modelling toolbox

The Statkraft Hydrological Forecasting Toolbox, Shyft, (<https://github.com/statkraft/shyft>) is an open-source distributed hydrological modelling framework developed by Statkraft with contributions from the University of Oslo and other institutions (e.g. Burkhart, 2016; Nyhus, 2017; Matt et al., 2018). The modelling framework has three main models (method stacks), namely PT_GS_K, PT_HS_K and PT_SS_K. These models use different methods in conducting snow mass balance calculations, although the same process representations are used for evapotranspiration and catchment response calculations. As such, these three models can be considered as alternative snow and water balance process representations within the domain of a conceptual modelling framework. The framework establishes a sequence of spatially distributed cells of arbitrary size and shape. Hence, it can provide lumped (single cell) or discretized (spatially distributed) calculations. When the framework is run in distributed setting, the selected model is applied to each of the grid-cells. Figure 3.1 shows simplified schematic representation of the Shyft models when run in distributed mode of simulation.

Information related to the model setup such as specifications of the selected Shyft model, parameter ranges, interpolation techniques, and paths to different input files are provided in ‘yaml’ files.

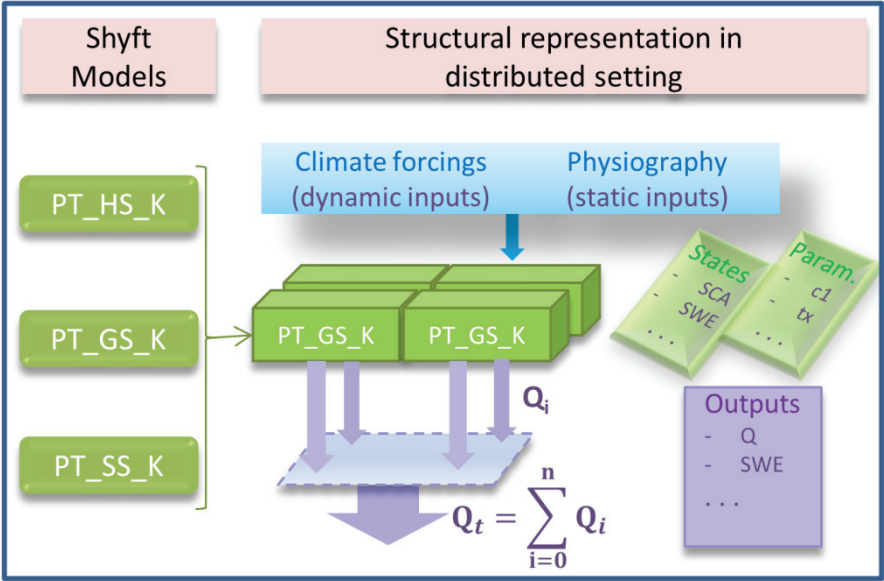


Figure 3.1: Simplified structural representation for a sample Shyft model in distributed setting

The Shyft models require both climatic and physiographic inputs. The climatic forcings include: precipitation, temperature, relative humidity, wind speed, and radiation. These forcings are dynamic in time and they can be input to the framework either as point measurements or as gridded data. In the case of point measurements, the data is distributed to the grid-cells using a suitable interpolation technique. Point temperature data is interpolated over the study domain using a Bayesian Kriging approach, while the remaining forcing variables are interpolated using the inverse distance weighting approach. Several auxiliary scripts within the framework are used to prepare the data in a format readily usable by a Shyft model and for post processing of outputs from the model. The physiographic input variables are static and they include: average elevation, total area, as well as the areal fractions of forest cover, glaciers, lakes, and reservoirs.

The Shyft models are composed of methods that are intended to carry out calculations related to the dominant processes taking place in the hydrological system of the study domain. These methods involve the conceptualization and parameterization of important processes and they can have several adjustable parameters depending on the climatic and physiographic characteristics of the study area where the model is applied. For example, the main model parameters for the PT_GS_K model are shown in Table 3.1. The optimal value for many of these parameters is determined through calibration using observed data. The major model states are snow water equivalent (SWE) and fractional snow cover area (fSCA). These variables are also generated as output (in addition to streamflow) for each grid-cell and time step in the simulation period.

3.2 An overview of the PT_GS_K model

The PT_GS_K was the most frequently used model both in this work and other studies as compared to the other models in Shyft. This model provides a relatively more physical (quasi-physical) description of the hydrological system as compared to the remaining two models. The discussion in this section is, therefore, mainly focused on the constituent methods of this model. PT_GS_K uses the Priestley-Taylor (PT) method (Priestley and Taylor, 1972) for estimating potential evaporation; a quasi-physical based method for snow melt, sub-grid snow distribution, and mass balance calculations (GS method); and a simple storage-discharge function (Lambert, 1972; Kirchner, 2009) for catchment response calculation (K). Overall, these three methods constitute the PT_GS_K model in Shyft.

The energy balance calculation in GS method follows a similar approach as used by DeWalle and Rango (2008) (Eq. 3.1). The precipitation falling in each grid-cell is classified as solid or liquid precipitation depending on a threshold temperature (tx) and on the local temperature values. The available net snow melt energy (NE) is the sum effect of different energy sources and sinks in the system. These include: incoming short wave radiation (SW), incoming (ILR) and outgoing (OLR) long wave radiation, the subsurface energy flux (G) as well as the turbulent sensible (SH) and latent (LH) energy fluxes emanating from rainfall, solar radiation, wind turbulence, and other sources.

$$NE = SW \cdot (1 - albedo) + ILR + OLR + SH + LH + G \quad (3.1)$$

Among other factors, the energy contribution from short wave radiation depends on snow albedo. For a given time step (t), the snow albedo of each grid-cell depends on the minimum (α_{min}) and maximum (α_{max}) albedo values as well as on air temperature (T_a) (Eq. 3.2). In this method the decay rates of albedo due to snow ageing as a function of temperature, i.e. the fast (α_{fdr}) and slow (α_{sdr}) albedo decay rates corresponding to temperature conditions above and below 0°C respectively, are parameterized. Turbulent heat contribution is the sum of latent and sensible heat. Wind turbulence is linearly related to wind speed using two parameters, i.e. $wind_const$ and $wind_scale$ from the intercept and slope of the linear function, respectively (Hegdahl et al., 2016).

$$\alpha_t = \begin{cases} \alpha_{min} + (\alpha_{t-1} - \alpha_{min}) \cdot \left(\frac{1}{2^{1/\alpha_{fdr}}} \right), T_a > 0 \text{ } ^\circ C \\ \alpha_{t-1} + (\alpha_{max} - \alpha_{min}) \cdot \left(\frac{1}{2^{(\alpha_{sdr})}} \right), T_a \leq 0 \text{ } ^\circ C \end{cases} \quad (3.2)$$

Sub-grid snow distribution is described by a three parameter Gamma probability distribution snow depletion curve (SDC) (Liston, 1999; Kolberg and Gottschalk, 2006). The traditional Gamma distribution is parameterized with two values, i.e. the average amount of snow at the onset of the melt season m (mm) and the shape value (k); based on the assumption that the ground is completely snow covered before the onset of melt. Since this assumption may not hold true for a number of grid-cells especially in alpine areas, a third parameter representing the bare ground fraction at the onset of snow melt season has been

introduced (Kolberg and Gottschalk, 2006). This SDC with three parameters is based on the assumption that certain proportion of a grid-cell area (y_0) remains snow-free throughout the winter season such as in steep slopes and ridges due to wind erosion or avalanches (Kolberg and Gottschalk, 2010). The two parameter Gamma distribution (Eq. 3.3) is thus applied only to the remaining portion of a grid-cell to estimate the fraction of the initially snow covered area where snow has disappeared (y'_t). Following this formulation, the bare ground fraction at each time step (y_t) is estimated using Equation 3.4. The initial bare ground fraction parameter is constant for all hydrological years. At each time step, the state parameters such as SWE and fSCA are updated using the SDC function (Fig. 3.2). In the GS method, the shape value is a direct transformation of the sub-grid snow coefficient of variation (CV_s).

$$y'_t = \int_0^{\lambda(t)} f(x; k, \theta) dx = \gamma(k, \frac{\lambda}{\theta}) \quad (3.3)$$

$$y_t = y_0 + (1 - y_0) \cdot y'_t \quad (3.4)$$

where f denotes the Gamma probability density function and γ is the incomplete Gamma function. x and $\lambda(t)$ respectively refer to point snow storage and the accumulated melt depth (mm) at time t since the onset of the melt season. θ represents the scale parameter with $m = k\theta$ and $k = CV_s^{-2}$.

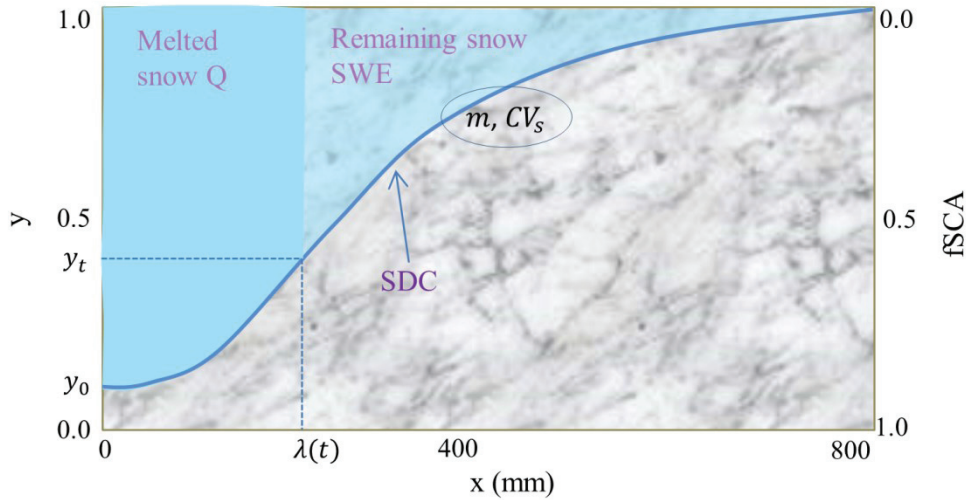


Figure 3.2: Illustration of the three-parameter snow depletion curve (SDC) parameterized by average snow storage (m), the snow coefficient of variation (CV_s), and the initial bare ground fraction (y_0) (after Kolberg et al., 2006). During the snow melt season these parameters remain constant while value of the total bare ground fraction (y) and snow covered area evolve through time following shape of the SDC and as a function of the accumulated melt depth ($\lambda(t)$) over the grid-cell.

The catchment response function is based on the storage-discharge relationship concept described in Kirchner (2009) and represents the sensitivity of discharge to changes in storage (Eq. 3.5). This method is based on the idea that catchment sensitivity to changes in storage i.e. $g(Q)$ can be estimated from the time series of discharge alone through fitting empirical

functions to the data such as the quadratic equation. Since discharge is generally non-linear and typically varies by many orders of magnitude, the recommended approach is to use log transformed discharge values in order to avoid the risk of numerical instability. In this method, the three parameters of the catchment response function, i.e. $c1$, $c2$, and $c3$ are parameterized.

$$\frac{d(\ln(Q))}{dt} = g(Q) \left(\frac{P - E}{Q} - 1 \right) \quad (3.5)$$

$$\text{with, } g(Q) = \exp(c1 + c2(\ln(Q)) + c3(\ln(Q))^2)$$

In which E and Q respectively represent actual evapotranspiration and discharge. In the original formulation P refers to precipitation, whereas in this method it refers to the liquid water supply from rainfall and snow melt.

The potential evaporation calculation in the PT method requires net radiation and the slope of saturated vapor pressure as well as the Priestley-Taylor parameter, the psychometric constant, and the latent heat of vaporization (e.g. Matt et al., 2018). The latter three variables are kept constant in the PT method. Actual evapotranspiration is assumed to take place only from snow free areas and it is estimated as a function of potential evapotranspiration and a scaling factor.

Table 3.1: Main parameters in the three methods (i.e. K, GS and PT) of the PT_GS_K model

Sr. No.	Name	Method	Description
1	$c1$	K	constant in Catchment Response Function, CRF
2	$c2$	K	linear coefficient in CRF
3	$c3$	K	quadratic coefficient in CRF
4	tx	GS	solid/liquid threshold temperature ($^{\circ}\text{C}$)
5	ws	GS	wind scale (slope in turbulent wind function)
6	fa	GS	fast albedo decay rate (days)
7	sa	GS	slow albedo decay rate (days)
8	cv	GS	spatial coefficient of variation of snowfall
9	$wind_const$	GS	intercept in turbulent wind function
10	max_water	GS	max water holding capacity of the snow pack
11	max_albedo	GS	maximum albedo for snow covered area
12	min_albedo	GS	minimum albedo for snow covered area
13	$land_albedo$	PT	albedo for snow free area
14	$alpha$	PT	dimensionless multiplier

3.3 Study area and data

Most of the case studies presented in this thesis are based on analyses conducted using climatic and physiographic data from the Nea-catchment (11.67390° - 12.46273° E, 62.77916° - 63.20405° N). The Nea-catchment constitutes the headwaters of the Nea-Nidelva water resources management area which is situated in Sør-Trøndelag County, Norway (Fig. 3.3). The hydropower generated from this area is the main source of electric supply to several places in mid-Norway including to one of the biggest cities in the country, Trondheim. As a result, this area has significance for various stakeholders responsible for the development and management of water resources in the region; and has been selected for research focused on better prediction and understanding of the snow processes and their impact on hydrology of the downstream area. The following sub-sections briefly present the climatic, physiographic, and hydrologic data for the Nea-catchment retrieved from different sources.

(a) Climatic data

As mentioned in the previous section, the PT_GS_K model requires temperature, precipitation, radiation, relative humidity, and wind speed as climatic forcing data. In this work, daily time series data of these variables for the study area that fully or partly cover the hydrological years 2008-2016 were obtained from Statkraft (2018) as point measurement, with the exception of relative humidity. Daily gridded relative humidity data was retrieved from ERA-interim (Dee et al., 2011). Mean annual precipitation in the Nea-catchment based on the nine hydrological years was 1279 mm. The highest and lowest average daily temperature values for this period were 20 and -23 °C, respectively.

(b) Physiographic data

The Nea-catchment covers a total area of 703 km² and it is characterized by a wide range of topographic and land cover characteristics. Altitude of the catchment ranges from 1783 masl on the eastern part around the mountains of Storsylen to 649 masl at its outlet on the western part of the catchment. In the case studies presented in this thesis, the PT_GS_K model was setup over each grid-cell of 1 km²; requiring average elevation, grid-cell total area, as well as the areal fraction of forest cover, glaciers, lakes, and reservoirs. Data for these physiographic variables were retrieved from two sources: the land cover data from Copernicus land monitoring service (2016) and the 10m digital elevation model (10m DEM) from the Norwegian mapping authority (Kartverket.no). The land cover data show that, the catchment is mainly dominated by moors, bogs, and some sparse vegetation; and limited part of the catchment is forest covered (3%).

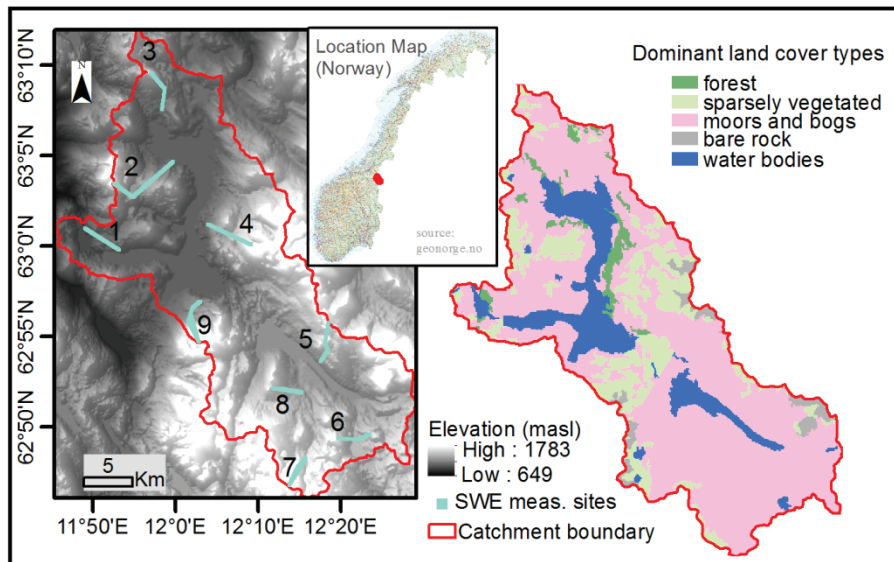


Figure 3.3: Physiography and location map of the Nea-catchment as well as the snow measurement sites.

(c) Hydrologic data

Streamflow and fractional snow cover area (fSCA) were used as validation datasets in the studies presented in Paper I and Paper II. Similarly, in Paper III snow water equivalent (SWE) observations were used as validation datasets, while fSCA observations were assimilated into the hydrological model. The streamflow and SWE datasets were provided by Statkraft (2018). The availability of long term SWE data, i.e., for nine years and in nine different sites was one of the reasons for choosing the Nea-catchment as a potential study area for validating the snow data assimilation schemes. Similar to the climatic variables, the observed SWE data covers the hydrological years 2008–2016. The SWE data were collected once a year in the month of April, where accumulated snow storage approximately attains its peak magnitude. The radar measurements roughly followed the same course each year in the nine representative sites of the catchment (Fig. 3.3). Table 3.2 presents, statistical summary of site-averaged annual peak observed SWE value of each grid-cell. While the highest average SWE value was observed in Site-2, the lowest value was recorded in Site-9. Annual mean and median catchment scale peak SWE values are also shown in Figure 3.4. From this Figure it can be noticed that the highest and lowest average catchment peak SWE values estimated from the nine sites during these hydrological years were 717 mm (in year 2012) and 331 mm (in year 2014), respectively.

Daily fSCA data was retrieved from NASA MODIS snow cover products (Hall et al., 2006). Frequent cloud cover is one of the major challenges when using MODIS and other optical remote sensing data in Norway. A composite dataset was thus formed using data retrieved from the Aqua and Terra satellites, MYD10A1 and MOD10A1 products respectively in order to minimize the effect of obstructions and misclassification errors emanating from clouds and other sources. The fSCA observations extend from the peak accumulation period until end of the ablation period, i.e. including the months April to August of each year (2008-2016) for the dates with valid satellite scenes.

Table 3.2: Summary of site-averaged annual peak observed snow water equivalent (SWE) (m w.e.) as well as average elevation of each site (masl).

Site	SWE (m w.e.)			Avg. Elev. (masl)
	Min.	Max.	Average	
1	0.269	0.558	0.4	746
2	0.411	0.931	0.606	907
3	0.159	0.764	0.487	873
4	0.299	0.659	0.454	958
5	0.398	0.818	0.603	1064
6	0.34	0.668	0.43	988
7	0.256	0.391	0.34	1251
8	0.258	0.383	0.316	976
9	0.152	0.549	0.274	993

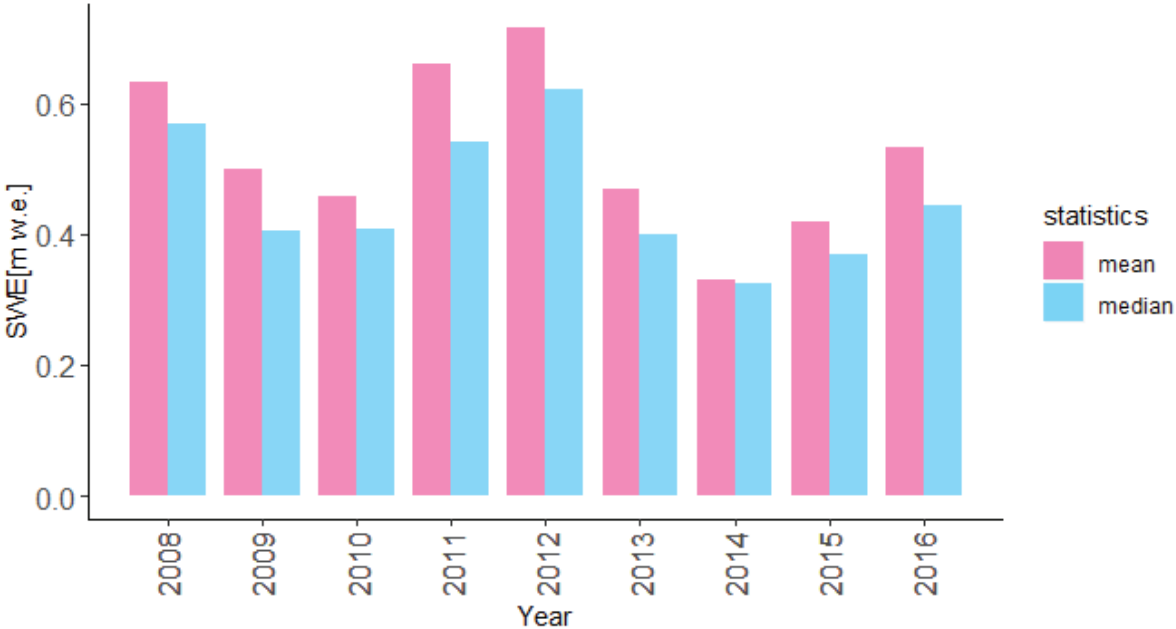


Figure 3.4: Annual statistics (mean and median) of catchment scale observed SWE values during the peak accumulation period based on radar measurements in nine sites within the Nea-catchment.

Chapter 4

Summary of publications

4.1 Paper I: Parameter uncertainty analysis for an operational hydrological model using residual-based and limits of acceptability approaches

Conceptual hydrological models are characterized by having one or more calibration parameters and parameter uncertainty estimation is one of the major challenges in using these types of models. In this study, a recently released distributed conceptual hydrological model, i.e. the PT_GS_K model of Shyft, was subjected to parameter uncertainty analysis using relevant data from the Nea catchment in Norway. Two variants of the generalized likelihood uncertainty estimation (GLUE) methodologies, one based on the residuals and the other on the limits of acceptability, were employed. Both single and multi-objective evaluation criteria were used in conditioning model parameters and in model validation. Two variables, i.e. streamflow and remote sensing snow cover data (fSCA) were used as validation datasets and model performances were evaluated using relevant objective functions. The objective functions include two forms of the Nash-Sutcliffe efficiency metrics that were used with the streamflow data, i.e. one applied on the raw data (NSE) and another on log-transformed data (LnNSE). Similarly, the root mean of squares error (RMSE) and the critical success index (CSI) were used during validation of the simulated fSCA against the remote sensing snow cover (MODIS) data.

When using the GLUE limits of acceptability (GLUE LoA) approach, a streamflow observation error of 25% was assumed. Neither the original limits nor relaxing the limits up to a physically meaningful value yielded a behavioral model capable of predicting streamflow within the limits in 100% of the observations based on the original formulations of this methodology. As an alternative to relaxing the limits, the requirement for the percentage of model predictions falling within the original limits (pLoA) was relaxed. A novel empirical approach was introduced to define the degree of relaxation in pLoA constrained by the 5% - 95% streamflow modelling uncertainty, reported as the containing ratio (CR) (Fig. 4.1). In this study, the CR value estimated for the calibration period using the residual-based approach was adopted as an acceptable CR value. Model realizations that fulfill this relaxed LoA criteria both in streamflow and fSCA observations were considered behavioral. A triangular membership function was used to define the weights of each criterion, where a maximum

weight of 1.0 was assigned to predictions with a perfect match to the observation and a minimum weight of 0.0 to predictions outside the acceptability limits.

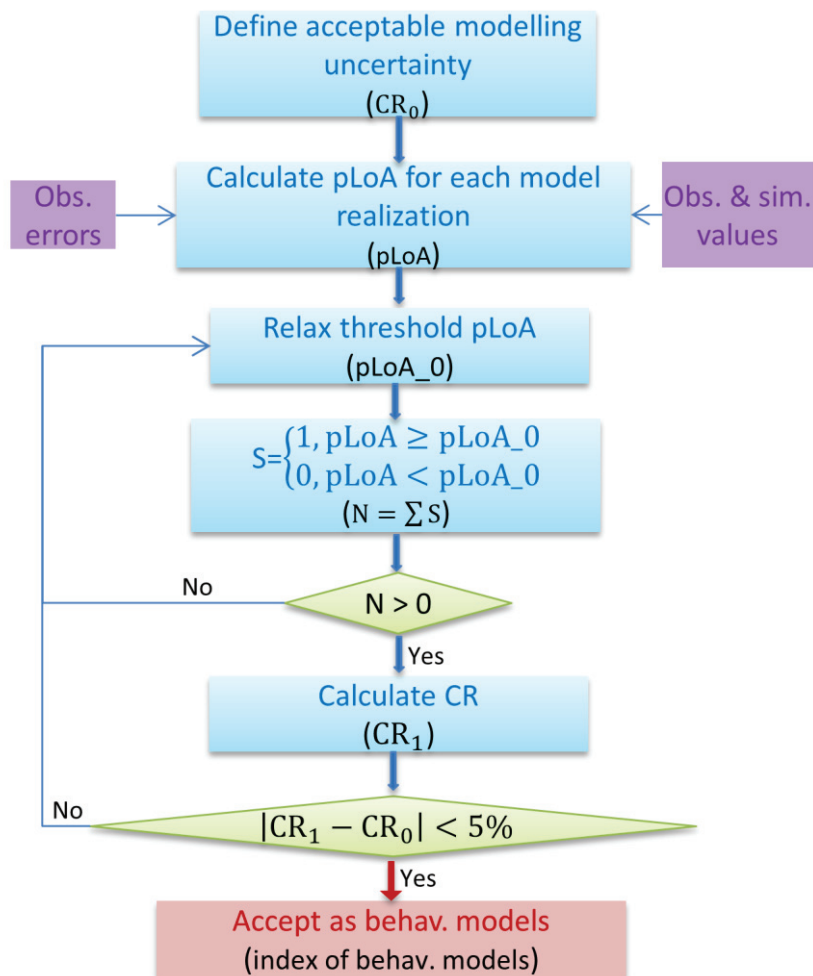


Figure 4.1: Schematic overview of the algorithm for the identification of behavioural models using the time-relaxed limits of acceptability approach.

The analysis result from the residual-based GLUE methodology shows that the catchment response parameters, i.e. $c1$, $c2$, and $c3$ as well as ws are the most sensitive model parameters. More caution is thus required when defining the value range of these parameters. On the other hand, the fast and slow albedo decay rates, i.e. fa and sa as well as the snow CV (cv) are relatively more uncertain model parameters. A multi-objective based model conditioning using combined streamflow and MODIS fSCA did not improve the median prediction of streamflow as compared to the result when model parameters are conditioned using streamflow only. The additional information from the MODIS fSCA data was thus less significant in constraining the rainfall-runoff model parameters. On the other hand, the multi-objective model conditioning using the combined objective function of NSE and LnNSE has yielded a significant improvement in model performance, especially during the low-flow condition as compared to using NSE alone.

Similar results were obtained using the residual-based GLUE and the time-relaxed GLUE LoA approaches. The median streamflow prediction of behavioral models identified using the time-relaxed GLUE LoA was able to mimic the observed values very well both under the

low- and high-flow conditions for the validation period. A slightly better capability of the 5%-95% prediction bounds in bracketing the observations was also observed as compared to predictions using the residual-based GLUE methodology when both streamflow and fSCA are used in model conditioning. The behavioural models identified using the time-relaxed GLUE LoA approach were also able to adequately reproduce the observed fSCA, although with a relatively higher prediction uncertainty during the onset of snowmelt as compared to the latter months of the ablation period. Generally, relaxing the percentage of observations required to be bracketed per simulation period by a particular model realization was found to be more effective and a viable option than relaxing the observational error bounds for balancing between type I and type II errors during behavioural model identification.

4.2 Paper II: Coupled machine learning and the limits of acceptability approach applied in parameter identification for a distributed hydrological model

Monte Carlo (MC) methods have been widely used in uncertainty analysis and parameter identification for hydrological models. The main challenge with these approaches is, however, the prohibitive number of model runs required to get an adequate sample size which may take from days to months especially when the simulations are run in distributed setting. In the past, emulators have been used to minimize the computational burden of the MC simulation by estimating the response surface associated with the residual-based GLUE methodology. In this study, emulators of the MC simulation were applied in parameter identification for the PT_GS_K distributed conceptual hydrological model using two likelihood measures, the absolute bias of model predictions (Score) and another based on the time-relaxed limits of acceptability concept (pLoA). Three machine learning models (MLMs), namely random forest (RF), K-nearest neighbours (KNN) and artificial neural network (NNET) were built using model parameter sets and response surfaces generated from limited number of model realizations (4000). The developed MLMs were applied to predict pLoA and Score for a large set of model parameters (95000). The behavioural parameter sets were identified using a time-relaxed limits of acceptability approach (GLUE pLoA) based on the predicted pLoA values and applied to estimate the quantile streamflow values using the behavioural model predictions weighted by their respective predicted Score values (Fig. 4.2).

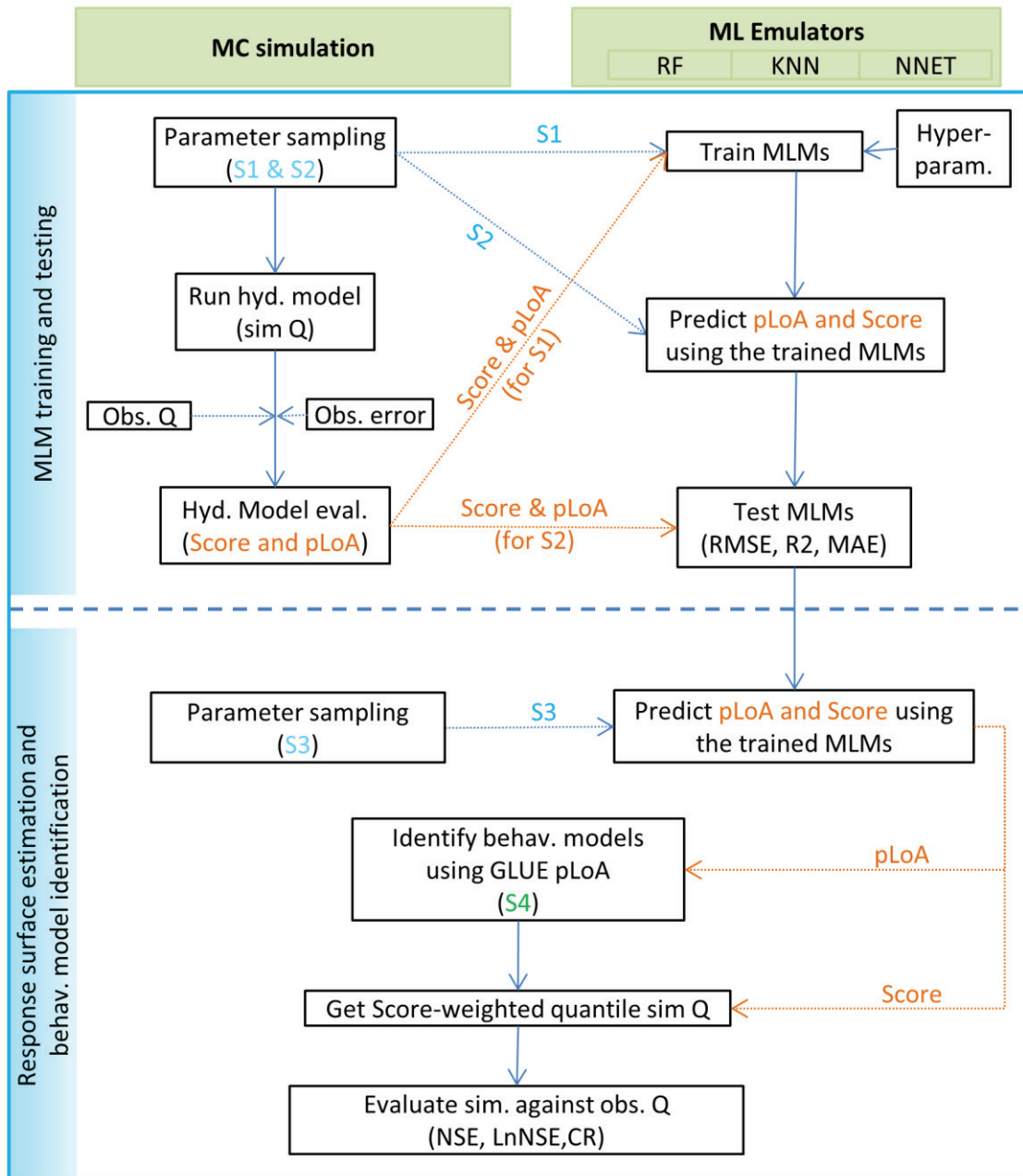


Figure 4.2: Schematic overview of the MLM training and testing as well as response surface prediction using the trained MLM with subsequent identification of behavioural models using the time-relaxed limits of acceptability approach (GLUE pLoA).

The coupled MLMs and time-relaxed limits of acceptability approach employed in this study were able to effectively identify behavioural parameter sets for the hydrological model. The MLMs were able to adequately reproduce the response surfaces for the test and validation samples. The evaluation metrics have shown variability both between the MLMs and the analysis years for the same MLM. RF and NNET yielded comparable results (especially for pLoA), while KNN has shown relatively lower result. Capability of the MLMs as emulators of the MC simulation was further evaluated through comparison of streamflow predictions using the identified behavioural model realizations against the observed streamflow values. The cross-validation result shows that the high-flow conditions as measured by average NSE were slightly better estimated both under the calibration and validation periods when KNN

was used as emulator as compared to RF and NNET. While under low-flow conditions (based on LnNSE), NNET has yielded a slightly better prediction as compared to the other two MLMS. Generally, the behavioural models identified based on KNN have yielded comparable performance to RF and NNET in terms of the efficiency measures, although characterized by a relatively higher inter-annual variability.

The sensitivity analysis conducted using the in-built algorithms of the three MLMs have yielded similar order of precedence in relative importance of the model parameters when trained using pLoA and Score as target variables. The result obtained from this analysis was also consistent with the results from previous study conducted using the residual-based GLUE methodology (Paper I) and the regional sensitivity analysis presented in Section 2.2.5. The catchment response parameters of the hydrological model, i.e. $c1$, $c2$ and $c3$ have shown higher relative importance as compared to most of the snow and water balance parameters. The highest degree of interaction was observed between the model parameters ws and tx which is also consistent with the result reported in Paper I. Thus, this study has also proven the efficiency of MLM based emulators in conducting sensitivity analysis for computationally expensive models.

4.3 Paper III: Improving the informational value of MODIS fractional snow cover area using fuzzy logic based ensemble smoother data assimilation frameworks

Remote sensing data have been increasingly used to get improved estimates of spatially distributed hydrological variables. The relative informational value of time-series observations in general and remote sensing data in particular varies in response to multiple factors. However, the commonly used ensemble-based smoothing schemes assume crisp and equal informational value for all observations of the assimilated data. The particle batch smoother (Pbs) is one of these schemes that has been gaining interest as a data assimilation (DA) scheme in hydrological models due to its distribution free likelihood and its capability to estimate the state vector directly at a relatively low computational cost. However, one of the main challenges in using Pbs and other Bayesian-based DA schemes is, that most of the weights are assigned to one or very few ensemble members and this may lead to degeneration of the statistical information in the ensembles. In this study, an alternative approach was considered that ensures fair distribution of weights among ensemble members based on the limits of acceptability concept (LoA) and certain hydrologic signatures. LoA was adopted as a DA scheme by introducing a methodology for relaxing the strict requirement of its original formulation as a rejectionist framework. New variants of these ensemble-based DA schemes (Pbs_F and LoA_F) were also introduced by incorporating the fuzzy logic concept into their likelihood measures in order to take into account for the variability in informational value of the assimilated observations.

The DA schemes were applied to a case study focused on the assimilation of the MODIS fractional snow cover area (fSCA) into the PT_GS_K model to get an improved estimate of snow water equivalent (SWE) during the maximum accumulation period. In the fuzzy logic based DA schemes, timing of the more informative period was assumed to vary both spatially

and temporally in response to various climatic and physiographic factors. Accordingly, the assimilation period was partitioned into three timing windows based on two critical points in the ablation period, i.e. the points where the mean fSCA changes (τ) and where the melt-out period starts (c). Parametric (likelihood-based) and non-parametric change point detection schemes were employed to locate τ in each grid-cell and year. The informational value of fSCA observations with time was assumed to follow an exponential-trapezoidal membership function that extends over the three timing windows (Fig. 4.3). In the first window (W-1), the fSCA measurements close to the onset of snowmelt are generally characterized by low signal to noise ratio. Similarly, during the complete melt-out period (W-3) the measurements are homogenous; and the informational value decreases with correlated data. The phenomena in W-1 and W-3 generally lead to less power in constraining model states and parameters. In contrast, the fSCA measurements in the second window (W-2) are characterized by high signal-to-noise ratio with a strong decreasing trend with time.

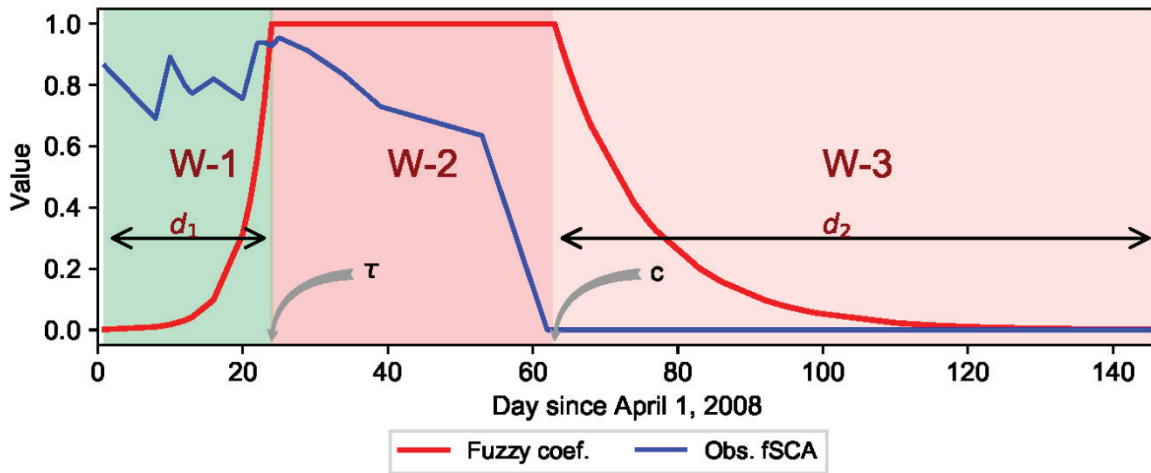


Figure 4.3: Location of the timing windows (W-1, W-2, and W-3) and value of the fuzzy coefficient in relation to a sample observed grid-cell fSCA dynamics over the ablation period as conceptualized in this study. τ and c respectively denote the point where mean snow cover area changes and start of the complete melt-out period.

The DA schemes have resulted in a posterior snow water equivalent (SWE) estimate that is better than the prior estimate in terms of accuracy as measured using one or more of the efficiency criteria used in this study. Introducing the LoA approach as a DA scheme yielded an encouraging result. An improved estimate of SWE was also obtained in most of the analysis years as a result of introducing the fuzzy coefficients in both DA schemes, although the degree of improvement in the evaluation metrics varied from year to year. Similarly, an improved estimate of fSCA was obtained for all years when using LoA_F as compared to LoA. The results were spatially analyzed both at grid-cell and site levels; and in many of the analysis years, the relative performance of the DA schemes was variable from one site to another. The most significant improvement was obtained in the correlation coefficient between the predicted and observed SWE values (site-averaged); with an increase by 8% and 16% after introducing the fuzzy coefficient in Pbs and LoA, respectively. This result reveals that, although all fSCA observations in the ablation period are important in constraining the

perturbation parameters, some observations are more important than others depending on their location in the time axis with respect to the critical points. This premise was further confirmed by the results obtained from sensitivity analysis of the DA evaluation metrics to change in location of the critical points. The detected critical points varied from year to year and spatially between the grid-cells. The parametric and non-parametric change point detection schemes employed to locate τ in each grid-cell and year have yielded close and reasonable results.

Chapter 5

General discussion

The PhD work presented in this thesis and the underlying analyses were geared towards the three important aspects for addressing uncertainty in hydrological modelling and predictions, i.e. the understanding, quantification, and reduction of uncertainty. Although distinct, these aspects of uncertainty are quite related to each other; and especially the understanding of uncertainty can be considered as an integral part of any application of uncertainty quantification or reduction (Liu and Gupta, 2007). This section will provide a general discussion on selected points from the research work presented in this thesis in relation to previous studies and organized in light of the three aspects of uncertainty.

5.1 Understanding uncertainty

Knowledge about the main sources and nature of uncertainty is the prerequisite for a successful uncertainty quantification and reduction. Although, no explicit account of the individual sources of uncertainty was done in this work, the modelling and prediction uncertainties were assumed to emanate from the different sources, i.e. from model structure, model parameters as well as from model forcings and observational datasets. Parameter uncertainty is one of the main sources of uncertainty in conceptual hydrological models and in the GLUE methodology, the likelihood weights propagated from the parameter uncertainty attempt to reflect all sources of uncertainty in the modelling process (Vrugt et al., 2009).

Understanding the philosophical background of a given uncertainty analysis framework is also very important since it may entail a considerable amount of uncertainty in the analysis results. For example, the initial attempt to find behavioural model realizations for an operational rainfall-runoff model, i.e. the PT_GS_K model, using the GLUE LoA rejectionist framework in its original formulation has yielded to the rejection of all model realizations. This result was obtained despite the presence of several behavioural model realizations that are useful for the intended purpose as identified using the residual-based GLUE methodology. The contrasting result from the two versions of the same framework suggests that there exists much uncertainty about the uncertainty quantification in hydrology, although considerable advances have been made in understanding the uncertainty quantification methodologies (Beven, 2011). Some proponents of the Proprietary falsification theory argue that the validation of numerical models of the hydrological system is impossible since natural systems are not closed (e.g. Oreskes et al., 1994). However, most of the practicing environmental scientists

follow the pragmatic realism as a working philosophy (Beven, 2001) based on the argument that models can be validated for pragmatic purposes in the sense of performance testing of whether a model is acceptable for its intended use (e.g. Rykie and Edward, 1996). Here model validity reflects the user's confidence in the model's usefulness (e.g. Nguyen and Kok, 2007). Furthermore, useful models may be discarded if the inherently imperfect models are judged by a standard they cannot achieve (e.g. Smith, 2001). A similar pragmatic realism based working philosophy was followed in Paper I when implementing the time-relaxed GLUE LoA and the subsequent application of the framework for evaluating the PT_GS_K model. This was done with due consideration to the model's inherent imperfection and its main application as a tool in operational hydrological studies.

In the presence of multiple contrasting schools of thought, it is challenging to get a universally acceptable uncertainty quantification technique. This creates difficulty in communicating the uncertainty results in hydrology. It is thus important to understand the concepts associated with model evaluation when testing hydrological models as hypotheses of catchment behaviour as viewed from different perspectives (Gupta et al., 2012). For example, from the engineering point of view model structural adequacy is mainly linked to its functional adequacy with due consideration to a decision-making perspective, while from the physical science point of view, model adequacy is purely defined in terms of the model's consistency with the physical system (e.g. Pianosi et al., 2016). The systems science, as a compromise between the two viewpoints, stresses on physical consistency to the extent that it is observable in data and it appreciates the simplicity of the engineering approach stressing on the principle of parsimony. In this work, mainly the systems science approach was followed during model evaluation (Paper I). For example, model realizations were evaluated in terms of their capability to simultaneously reproduce the low- and high-streamflow conditions in order to ensure physical consistency throughout the simulation period. On the other hand, the evaluation criteria were set with due consideration to the model's main application as an operational hydrological model.

5.2 Uncertainty quantification

In the past, several uncertainty estimation techniques have been proposed and applied to test hydrological models. These techniques can be classified as frequentist or Bayesian approaches, probabilistic or non-probabilistic approaches (e.g., Montanari et al., 2009), or as formal or informal approaches (e.g., Vrugt et al., 2009). However, these approaches are not mutually exclusive. For example, both frequentist and Bayesian are probabilistic and formal. We can also have probabilistic methods with informal (subjectively chosen) likelihoods. Though according to Nearing et al. (2016), any non-negative likelihood function with a finite total integral (including the informal ones) becomes a coherent probability distribution, and thus formal, after appropriate scaling. Non-probabilistic methods (e.g. based on fuzzy sets) can be formal (in the sense of following strictly defined rules) but will involve some subjectively chosen constraints (e.g. support for fuzzy measure). Of the different uncertainty estimation techniques in hydrology, the GLUE methodology is the most widely used. While this methodology is primarily focused on the quantification of parameter uncertainty, other

approaches have attempted to treat individual error sources with subsequent assessment of predictive uncertainty (Sadegh and Vrugt, 2013).

The parameter uncertainty analysis using the residual-based GLUE methodology (Paper I) have revealed that parameter inference based on fSCA as a conditioning observation yielded some parameter estimates that deviate significantly from those obtained when conditioned with streamflow only. However, the displacement in value of certain snow related model parameters towards the higher part of their parameter dimension, in relation to those identified using streamflow, was counteracted by an opposite effect. Other snow related parameters were shifted towards the lower part of their respective parameter dimension thereby resulting in a partial or full cancelling out of the effects of changes in the former parameters. This phenomenon may thus lead to equifinality, where different sets of model parameters give comparable fSCA responses.

Results from the analysis using the time-relaxed GLUE LoA (Paper I) provided a further insight on model performance with respect to the individual observations at different parts of the hydrograph. The number of behavioral models with their predictions falling within the observational error bounds varies from one time step to another in response to multiple factors including epistemic errors in the model structure as well as errors in the input and observational datasets. This study reveals that modelling and prediction uncertainties are relatively higher during low-streamflow than high-streamflow conditions. Figure 5.1 presents the average daily percentage of behavioral models whose predictions fall within the streamflow observational error bounds at four quantile intervals of the observed streamflow, i.e. <0.25 (q_1), $0.25-0.50$ (q_2), $0.50-0.75$ (q_3) and >0.75 (q_4). For most of the analysis years, the lowest percentages are observed at the first quantile, i.e. for the streamflow values below the 25th percentile, while highest daily percentage of behavioral models satisfying the criteria were observed for the higher percentiles. This result is consistent with the general observation that hydrologic models perform relatively well in wet conditions but break down during low-streamflow conditions (e.g. Kirchner, 2009). The uncertainty for the low-streamflow condition increases if only NSE is used as a likelihood measure during the identification of behavioural models. This highlights the challenge with transferability of catchment hydrologic models in time and space (e.g. Clark et al., 2011).

The relatively higher mismatch between simulated and observed values during the low-streamflow condition cannot always be explained by uncertainties in the hydrological model structure. It can also be partly explained by the relatively higher measurement uncertainty during low-streamflow condition as compared to high-streamflow conditions. For example, Petersen-Øverleir et al. (2009) have observed a relatively higher measurement uncertainty during low-streamflow condition as compared to high-streamflow conditions in several Norwegian streamflow gauging stations. This phenomenon was attributed to the inaccuracy of some instruments under low-streamflow condition as well as to increased cross-sectional irregularities in shallow depths leading to less precise flow area calculations.

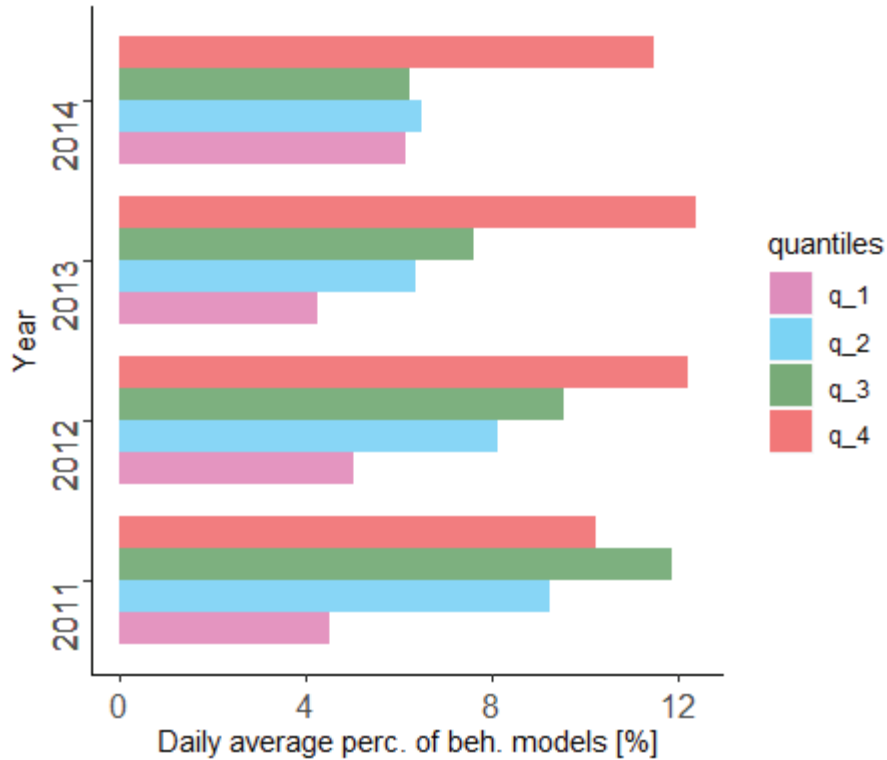


Figure 5.1: Average daily percentage of behavioral model simulations with their predictions falling within the streamflow observational error bounds at each of the streamflow quantiles, i.e. at the first quantile (q_1), second quantile (q_2), third quantile (q_3) and fourth quantile (q_4). The analysis covers four hydrological years, i.e. years 2011 to 2014.

One of the main goals of this work was to optimize the computational efficiency of the ensemble-based uncertainty quantification for use in computationally expensive hydrological applications. In the past, different approaches have been introduced in order to minimize the high computational cost associated with uncertainty analysis when using Monte Carlo (MC) simulation based frameworks. For example, the DREAM framework attempts to improve the efficiency in finding behavioural models through use of adaptive proposal updating (Vrugt et al., 2009). In Paper II, the improvement in computational efficiency was realized through coupling of three machine learning based models (MLMs), i.e. random forest (RF), K-nearest neighbours (KNN) and artificial neural networks (NNET) with the time-relaxed GLUE LoA approach. In the contemporary era of big-data, machine learning methods are playing crucial roles in many applications involving data-driven analyses (e.g. El Naqal and Murphy, 2015). In this study, the MLMs trained using a limited number of parameter sets were able to efficiently predict the response surfaces of much bigger sample size. Similar studies based on the coupling of NNET and the residual-based GLUE methodology have reported comparable performances of the emulators to GLUE in terms of the posterior parameter and prediction interval estimations at much lower computational cost (e.g. Yu et al., 2015).

The three MLMs employed in this study are non-parametric. Thus, they have an advantage over parametric models as they can be implemented to find a relationship between the predictor and target variables regardless of the error structure. However, since both the model structure and model estimates of non-parametric models are derived from the data, larger

sample sizes are required in nonparametric than parametric models (Araghinejad, 2014). Furthermore, MLMs like all data-driven models operate as black-box models, i.e. without any understanding of the modelled processes; hence, they may not behave as intended under changing conditions (Olden and Jackson, 2002). As noticed in this study, this problem can partly be solved using adequate and representative model training samples.

As demonstrated in this work and previous studies, sensitivity analysis is closely related to parameter uncertainty. It can be used to complement the process of model calibration by providing better insights on how the variability in the uncertain model parameters map onto variations of likelihood values (e.g. Pianosi et al., 2016). Since most conceptual hydrological models have several free parameters, screening is usually required before conducting parameter uncertainty analysis in order to effectively sample from the parameter dimensions and to minimize the computational cost of Monte Carlo simulations. In this work, both expert elicitation and the screening method using the Elementary Effect (EE) (Morris, 1991) were employed to identify the parameters to which the model results are more sensitive before propagation of parameter uncertainty to modelling and prediction uncertainty. The EE analysis has yielded similar result to the model parameters selected for uncertainty analysis based on the expert elicitation technique. Sensitivity analyses were also conducted following the modelling uncertainty analysis in order to identify the most influential out of the selected model parameters. The regional sensitivity analysis technique, machine learning methods as well as the GLUE methodology were employed in conducting the sensitivity analyses and have yielded similar results. As such, a relatively detailed parameter uncertainty analysis was conducted in this work as compared to previous studies that were commonly done using a single approach.

5.3 Uncertainty reduction

An adequate model structure and appropriate likelihood measures play an important role in reducing the modelling and prediction uncertainty by optimizing the amount of information extracted from the conditioning dataset. For example, as mentioned in the preceding section, a higher modelling and prediction uncertainty was observed during the low-streamflow condition than the high-streamflow condition. In the case of residual-based approach, a significant improvement was observed when using the combined streamflow likelihood of NSE and LnNSE instead of NSE alone (Paper I). However, a likelihood measure based on the combined use of remote sensing snow cover area (fSCA) and streamflow datasets for conditioning model parameters did not yield significant improvement in model performance as compared to using streamflow alone. Similar result in previous studies have led to the understanding that streamflow data contain sufficient information to ensure the identification of a suitable model structure that can closely and consistently mimic the catchment behaviour at temporal and spatial scale of measurement (e.g. Sadegh and Vrugt, 2013). In another similar study using the HBV model, Udnæs et al. (2007) have obtained an improved performance of the model in fSCA simulation without a major reduction in the accuracy of simulated runoff.

The indifference in model performance after including fSCA as additional criteria can be attributed to the limited informational value of fSCA in constraining parameters of the rainfall-runoff model. It can also be partly attributed to the lack of flexibility of the parsimonious model structure to accommodate different perceptual models based on a multi-source dataset. As reported in previous studies, such challenges are common due to the lack of process-based conceptual models that can operate on a wider scale (e.g. McMillan et al., 2011). One of the reasons for the lack of simultaneous improvement in both fSCA and streamflow in this work might be the difference in spatial scales at which the hydrological model was setup, which was 1km^2 , in comparison to the finer spatial scale at which most snow processes operate. For example, wind-driven snow redistribution processes become more important in simulations with high grid-resolution as compared to coarser resolution (e.g. Mott and Lehning, 2010). In an effort to partly overcome these challenges, previous research works focused on model structural adequacy have resulted in data assimilation and other stochastic techniques for iterative refinement of the mathematical structure of conceptual hydrological models (Sadegh and Vrugt, 2013).

The development of efficient techniques with better capability to extract and assimilate information from the available data is one of the important areas where actions can be taken towards reducing uncertainty in hydrologic prediction (Liu and Gupta, 2007). Although the GLUE methodology is a widely used uncertainty assessment framework in hydrology, the standard version of this framework has been criticized for using subjectively set likelihood measure and threshold values. The followed version of this framework which was based on the limits of acceptability concept (GLUE LoA) partly overcomes this challenge through use of observation error bounds. However, when it comes to operational hydrological models with continuous rainfall-runoff simulation, the original formulation of the GLUE LoA approach as a rejectionist framework has limited applicability since it can lead to the rejection of useful models that might adequately represent the catchment behaviour and thereby to making a type II error (false negative). For example, although useful behavioural model realizations were identified when using the residual-based GLUE, the current model structure of the PT_GS_K model was rejected when using the GLUE LoA since none of its model realizations were able to meet the strict requirement of this framework (Paper I). In the past, different methodologies have been followed to minimize the risk of making type-II errors when using this framework including relaxing the limits and use of different parameter sets for different parts of the hydrograph. In this work, the latter alternative was not considered as a potential option based on the premise that both wet and dry periods are required in order to activate all components of a hydrological model and thereby ensure its transferability in time (e.g. Sadegh and Vrugt, 2013). The first option, i.e. relaxing the limits has resulted in the inclusion of non-behavioural model realizations and thereby to making type-I errors. Thus, the time-relaxed GLUE LoA was introduced as an alternative to relaxing the limits (Paper I).

In the Nea-catchment case study involving the time-relaxed GLUE LoA approach and the PT_GS_K model, this approach was able to effectively identify behavioural model realizations based on only 30-40 % of the observations in a hydrologic year. The identified behavioural models were able to predict streamflow during the evaluation period with an acceptable degree of accuracy for the intended use based on commonly used efficiency

criteria. As discussed in the previous section, many of the relaxed time steps corresponding to the predictions that fall outside the error bounds are located during the low-streamflow condition. Visual inspection on the resulting hydrograph shows that the observations during the low-streamflow condition, especially during the winter period, are highly correlated and hence they carry low information content as compared to the observations from the remaining part of the hydrograph. Thus, the results from this analysis is consistent with the general notion that the information content of the data and the efficiency with which that information is extracted are more important than the amount of the data (e.g. Gupta and Sorooshian, 1985; Liu and Han, 2010; Sun et al., 2017).

Little attention has also been given in previous studies to reducing the prediction uncertainty through explicit account for the relative informational value of the assimilated observations. For example, the widely used schemes for reducing the uncertainty in SWE prediction through the assimilation of remote sensing or in situ measurements, such as the particle batch smoother (Pbs), have likelihood measures that assume equal informational value for all observations. The need for the development of data assimilation schemes that can extract the primary information content from observations with noisy information is thus still one of the main open areas of research in hydrological data assimilation (e.g. Houser et al., 2012). In this PhD work snow data assimilation schemes based on the fuzzy logic concept are introduced in order to take into account the variability in information content of the assimilated remote sensing snow cover data with time. The fSCA observations from the time windows characterized by low signal-to-noise ratio and high correlation were assumed to have less informational value and hence assigned less weight in conditioning the model states and parameters. The informational value of the assimilated data was assumed to be fuzzy that carry considerable amount of uncertainty (with its value ranging from 0 to 1) instead of a crisp quantity (0 or 1) which is the case in the existing ensemble smoothing schemes. An improved estimate of SWE during the peak accumulation period was obtained after introducing the fuzzy logic concept as compared to the original formulation of Pbs. This result is consistent with the general notion that many of the variables that we usually consider to be crisp quantity and deterministic are actually fuzzy that carry considerable amount of uncertainty (Ross, 2009). Further, in Pbs, most of the weights are assigned to one or very few ensemble members and this may lead to degeneration of the statistical information in the ensembles (Van Leeuwen, 2009). In an attempt to overcome this limitation, an alternative data assimilation scheme based on the limits of acceptability concept and certain hydrological signatures was thus introduced and has yielded a promising result in reducing the prediction uncertainty (Paper III).

The absence of alternative process representations during the model identification process may yield a biased result (e.g. Clark et al., 2011). In light of the different sources of uncertainty discussed herein and other hydrologic literatures, hydrological models may be non-unique in reproducing the catchment behaviour; and hence there will always be the possibility of equifinality (e.g. Beven, 2002). Thus, instead of searching for a single optimal model as in the case of traditional statistical approach, it would be more realistic to consider alternative ensemble of behavioural model representations that yield acceptable results. Embracing the limitations of a single optimal model, some studies suggest more research on

new methods for combining rather than selecting the best of imperfect models (e.g. Smith, 2001). An example of such methods is the GLUE framework that employs equifinality as part of its working paradigm. In contrast, the traditional approach attempts to find a single optimal model based on two extreme alternatives, i.e. useful or not useful. With the increasing appreciation for the ensemble-based modelling approach, the old aphorism by George Box might be further expanded and updated in the contemporary scene to: “*All models are wrong, but some are useful*”, *albeit at varying degrees of usefulness*. This expanded aphorism can be explained from a hydrological modelling viewpoint as follows:

- All models are wrong due to the omissions of certain processes from the perceptual model as well as the simplifications of important processes representations in the resulting conceptual and procedural models, but
- some are useful for the intended purpose. Models are tools designed for specified purposes, rather than as truth generators. Hence, validity of the model can be viewed in accordance to the user’s confidence in the model’s usefulness (e.g. Senge and Forrester, 1980; Nguyen and Kok, 2007).
- The relative usefulness of the ensemble members can assume any score between 0 to 1.0 instead of the binary extremes, i.e. 0 (not useful) or 1(useful). This reflects the need for an importance weighted ensemble prediction of an environmental variable.

Chapter 6

Conclusions and recommendations

6.1 Conclusions

The research work presented in this thesis was focused on ensemble-based uncertainty quantification and reduction in hydrological modelling and predictions accompanied by the proper understanding of the different sources and types of uncertainty. This section presents the conclusions drawn in light of the three main goals and specific objectives of the work.

The first goal was focused on the assessment of modelling and prediction uncertainty for a distributed conceptual hydrological model using two variants of the GLUE methodology and with due consideration to the model's main application as an operational hydrological model (Paper I). The analysis results from the residual-based GLUE methodology show that the catchment response parameters of the PT_GS_K model are the most sensitive model parameters. On the other hand, the snow and water balance related parameters induced a relatively higher streamflow uncertainty than the catchment response parameters. Results from the multi-criteria model conditioning using streamflow and MODIS fractional snow cover area (fSCA) revealed that the additional information from the fSCA data was generally less significant in constraining the rainfall-runoff model parameters, yielding no substantial improvement on the median prediction of streamflow and fSCA. The median streamflow prediction using the behavioural model realizations of the PT_GS_K model was able to mimic the validation dataset remarkably well in most of the analysis years, although the model performance was relatively low during the low-streamflow conditions.

The main outcome of the first goal was the novel approach introduced to adapt the rejectionist framework (GLUE LoA) for use in identification of behavioural model realizations for an operational hydrological model. In contrast to the result obtained from the residual-based GLUE, using the original formulation of the GLUE LoA methodology for the identification of behavioural models did not yield any behavioural simulation. A relaxation of the strict requirements of the original formulation was thus needed in order to adapt the rejectionist framework to find acceptable model realizations for an operational hydrological model. The relaxation considers the limitations of using constant observational error proportionality and that of not taking an explicit account of the other sources of uncertainty such as from input data and observational errors as well as due to the incomplete nature of model structure. This way, the risk of making type II errors, i.e. the rejection of useful models

was minimized. A time-relaxed GLUE LoA approach was introduced that allows a relaxation on the number of time steps required to achieve the LoA criteria. Similar results were obtained using both the residual-based GLUE and the time-relaxed GLUE LoA approaches. Relaxing the percentage of time steps in the simulation period where predictions by a particular model realization required to fall within the observation error bounds (pLoA) was found to be more effective than relaxing the observational error bounds. The latter approach was used in previous studies to minimize the risk of making type II errors when using GLUE LoA.

The second goal was aimed at alleviating the heavy computational burden associated with use of the time-relaxed GLUE LoA (GLUE pLoA) in computationally expensive hydrological models (Paper II). Three machine learning models (MLMs), i.e. random forest (RF), K-nearest neighbours (KNN), and an artificial neural-network (NNET) were individually coupled with the GLUE pLoA approach. The coupled MLMs and GLUE pLoA approach employed in this study were able to effectively identify behavioural parameter sets for the hydrological model, while significantly reducing the computational time. The MLMs were able to adequately reproduce the response surfaces for the test and validation samples, although the evaluation metrics have shown variability both between the MLMs and the analysis years. The cross-validation results of the simulated against observed median streamflow values have further revealed the efficiency of pLoA as a likelihood measure for the identification of behavioural models. The assessment on relative performances of the MLMs shows that, although RF and KNN were not among the favourite emulators of the Monte Carlo (MC) simulation in previous studies, they have yielded comparable results to the standard MLM based emulator, i.e. NNET. Furthermore, the cross-validation result shows that the high-streamflow conditions were slightly better estimated both under the calibration and validation periods when KNN was used as emulator as compared to RF and NNET, while NNET yielded a slightly better prediction under low-streamflow conditions.

The sensitivity analysis conducted using the in-built algorithms of the three MLMs have yielded similar order of precedence in relative variable importance when trained using pLoA and a normalized absolute bias (Score) as target variables. The result was also generally consistent with the sensitivity analysis result obtained from the previous study conducted using the residual-based GLUE methodology (Paper I) and the regional sensitivity analysis approach (Section 2.2.5).

The third goal was focused on reducing the prediction uncertainty of snow water equivalent (SWE) and thereby to get an improved estimate of SWE during the maximum accumulation period. This goal was accomplished by introducing fuzzy logic based efficient snow data assimilation (DA) schemes that have improved capability of extracting the information content of the assimilated remote sensing snow cover data. To reanalyze the model results, two ensemble-based data assimilation schemes, i.e. particle batch smoother (Pbs) and another based on the limits of acceptability concept (LoA) were used. New variants of these schemes that account for the variability in informational value of the assimilated fSCA observation, i.e., Pbs_F and LoA_F were also introduced in this work. Using the LoA approach as a DA scheme yielded an encouraging result and all DA schemes resulted in a posterior SWE estimate that is better than the prior estimate in terms of accuracy as measured

using one or more of the efficiency criteria used in this study. Incorporating the concept of variable informational value of the remote sensing data in both DA schemes was a viable option for improved estimates of the perturbation parameters, and thereby the reanalyzed SWE values.

Results from the analyses under the premise of variable informational value of fSCA with time revealed that all observations do not carry equal information amount in constraining the perturbation parameters. Although most fSCA observations in the ablation period were important in constraining the perturbation parameters, some observations were more important than others depending on their location in the time axis with respect to certain critical points in the melt season, i.e., the points where the mean snow cover changes (τ) and the start of a melt-out period. The parametric (likelihood-based) and non-parametric change point detection schemes employed to locate τ in each grid-cell and year yielded similar and reasonable results. Results from the sensitivity test conducted by moving the critical points forward and backward have also confirmed the variability in information content of the fSCA observations with time. Although the DA results were somewhat insensitive to locations of the critical points up to certain distances, moving the critical points farther from their original locations lead to deterioration of the DA results.

Generally, the new methods introduced in this work are expected to make contributions to the society at different levels. Water managers in hydropower industries and water supply sectors can benefit from the reduced level of uncertainty in model predictions. The improved estimates of hydrological variables such as snow storage and streamflow will enable them to make better water resources related decisions. The methods that mimic the human reasoning capability such as the fuzzy logic based data assimilation schemes and the machine learning based emulators will also contribute to the scientific community. The results from these methods provide further insights on significances of artificial intelligence and big-data science in hydrological modelling.

6.2 Recommendations for further work

In order to improve the relatively low performance of the PT_GS_K model during the low-flow condition, the processes affecting this part of the hydrograph need a due consideration in future studies focused on improvements in model structure. For example, a scrutiny on the underlying algorithms of the model reveal that, the precipitation falling on water bodies such as lakes and reservoirs was assumed to have a direct contribution to streamflow. However, this assumption may not hold true especially during the winter season for two main reasons. In snow dominated catchments, most of the precipitation falls as snow during the winter season and accumulates on surface of the frozen water bodies. The accumulated snow requires a sufficient amount of melt energy before it starts to contribute to downstream flows. Further, even the precipitation falling as rain may not reach the catchment outlet on the same hour or day of observation, depending on such factors as size and surface roughness of the water body. As such, this conceptualization may have substantial negative effect on simulated flows if size of the lakes and/or reservoirs constitute significant fraction of the total catchment area. The additional parameterizations for improving the model structure should be

implemented with due consideration to the value of parsimonious structure in operational models.

The time-relaxed GLUE LoA methodology requires specifying the magnitude of acceptable modelling uncertainty (CR_0) in order to determine the degree of relaxation. In this work, this value was set based on a previous assessment using the residual-based GLUE. For a wider use of the methodology, further studies need to be done for estimating CR_0 using alternative approaches. In this work, the behavioural model realizations were identified only from a single model, i.e. the PT_GS_K model. However, as an ensemble-based approach, the GLUE methodology allows the use of model realizations from a single or multiple models. Similarly, the Shyft modelling framework provides models with different process representations. Therefore, further studies can also be done to identify behavioural realizations from two or more models available in Shyft or other modelling frameworks using the time-relaxed GLUE LoA. Further studies focused on application of this methodology in other catchments with different hydrologic characteristic need also be conducted in order to get a better insight in viability of the newly introduced methodology under different conditions.

This work has demonstrated the role of individual machine learning models as emulators of the time consuming Monte Carlo simulation during parameter identification for an operational hydrological model. Further studies may assess the possibility of using the MLMs used in this work and other MLMs as ensemble emulators to get an improvement in the identification of behavioural parameter sets. Further efforts should be done to integrate machine learning methods in uncertainty analysis and parameter identification frameworks. The latter frameworks can also benefit by adopting relevant new concepts from machine learning models that are advancing at a faster rate.

Although the work presented in Paper III was focused on fractional snow cover area (fSCA) assimilation into a hydrological model, the new data assimilation (DA) schemes based on the fuzzy logic concept can be applied to assimilate other measurements that display variable informational value with time. More case studies should be conducted to assess the scope of the DA schemes under different physiographic and climatic conditions as well as different assimilated variables. Further studies can also be conducted to assimilate fSCA or other similar variables into land surface models, and thereby assess the viability of these DA schemes in different types of models.

References

- Aalstad, K., Westermann, S., Schuler, T. V., Boike, J., and Bertino, L.: Ensemble-based assimilation of fractional snow covered area satellite retrievals to estimate snow distribution at a high Arctic site, 2018.
- Alippi, C., Boracchi, G., and Roveri, M.: Ensembles of change-point methods to estimate the change point in residual sequences, *Soft Computing*, 17, 1971-1981, 2013.
- Appelhans, T., Mwangomo, E., Hardy, D. R., Hemp, A., and Nauss, T.: Evaluating machine learning approaches for the interpolation of monthly air temperature at Mt. Kilimanjaro, Tanzania, *Spatial Statistics*, 14, 91-113, 2015.
- Araghinejad, S.: Data-driven modeling: using MATLAB© in water resources and environmental engineering, *Water Science and Technology Library*, 2014.
- Bailey, R., and Baù, D.: Ensemble smoother assimilation of hydraulic head and return flow data to estimate hydraulic conductivity distribution, *Water Resources Research*, 46, 2010.
- Bair, E. H., Abreu Calfa, A., Rittger, K., and Dozier, J.: Using machine learning for real-time estimates of snow water equivalent in the watersheds of Afghanistan, *The Cryosphere*, 12, 1579-1594, 2018.
- Bárdossy, A., and Singh, S.: Robust estimation of hydrological model parameters, *Hydrology and Earth System Sciences*, 12, 1273-1283, 2008.
- Bartelt, P., and Lehning, M.: A physical SNOWPACK model for the Swiss avalanche warning: Part I: numerical model, *Cold Regions Science and Technology*, 35, 123-145, 2002.
- Bayazit, M.: Nonstationarity of hydrological records and recent trends in trend analysis: a state-of-the-art review, *Environmental Processes*, 2, 527-542, 2015.
- Bengtsson, L., Ghil, M., and Källén, E.: *Dynamic meteorology: data assimilation methods*, Springer, 1981.
- Beniston, M., Keller, F., and Goyette, S.: Snow pack in the Swiss Alps under changing climatic conditions: an empirical approach for climate impacts studies, *Theoretical and Applied Climatology*, 74, 19-31, 2003.
- Bergstrom, S.: Development and application of a conceptual runoff model for Scandinavian catchments, 1976.
- Beven, K.: Changing ideas in hydrology—the case of physically-based models, *Journal of Hydrology*, 105, 157-172, 1989.

- Beven, K., and Binley, A.: The future of distributed models: model calibration and uncertainty prediction, *Hydrological processes*, 6, 279-298, 1992.
- Beven, K.: Prophecy, reality and uncertainty in distributed hydrological modelling, *Advances in Water Resources*, 16, 41-51, 1993.
- Beven, K.: How far can we go in distributed hydrological modelling?, *Hydrology and Earth System Sciences Discussions*, 5, 1-12, 2001.
- Beven, K.: Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system, *Hydrological processes*, 16, 189-206, 2002.
- Beven, K.: A manifesto for the equifinality thesis, *Journal of hydrology*, 320, 18-36, 2006.
- Beven, K.: *Environmental modelling: An uncertain future?*, CRC Press, 2009.
- Beven, K., and Westerberg, I.: On red herrings and real herrings: disinformation and information in hydrological inference, *Hydrological Processes*, 25, 1676-1680, 2011.
- Beven, K. J.: *Rainfall-runoff modelling: the primer*, John Wiley & Sons, 2011.
- Beven, K., Smith, P., Westerberg, I., and Freer, J.: Comment on “Pursuing the method of multiple working hypotheses for hydrological modeling” by P. Clark et al, *Water Resources Research*, 48, 2012.
- Beven, K., and Smith, P.: Concepts of Information Content and Likelihood in Parameter Calibration for Hydrological Simulation Models, *Journal Of Hydrologic Engineering*, 20, 2015.
- Beven, K.: Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication, *Hydrological Sciences Journal*, 61, 1652-1665, 2016.
- Beven, K. J.: On hypothesis testing in hydrology: Why falsification of models is still a really good idea, *Wiley Interdisciplinary Reviews: Water*, 5, e1278, 2018.
- Bieker, H. P., Slupphaug, O., and Johansen, T. A.: Real-time production optimization of oil and gas production systems: A technology survey, *SPE Production & Operations*, 22, 382-391, 2007.
- Blanning, R. W.: The construction and implementation of metamodels, *simulation*, 24, 177-184, 1975.
- Blazkova, S., and Beven, K.: Flood frequency estimation by continuous simulation for a catchment treated as ungauged (with uncertainty), *Water Resources Research*, 38, 2002.
- Breiman, L.: Random forests, *Machine learning*, 45, 5-32, 2001.
- Burkhart, J. F., Helset, S., Abdella, Y. S., and Lappegard, G.: Operational Research: Evaluating Multimodel Implementations for 24/7 Runtime Environments, Abstract H51F-1541 presented at the Fall Meeting, AGU, San Francisco, California, 11–15 December 2016.
- Campolongo, F., Cariboni, J., and Saltelli, A.: An effective screening design for sensitivity analysis of large models, *Environmental modelling & software*, 22, 1509-1518, 2007.
- Carrasi, A., Bocquet, M., Bertino, L., and Evensen, G.: Data assimilation in the geosciences: An overview of methods, issues, and perspectives, *Wiley Interdisciplinary Reviews: Climate Change*, 9, e535, 2018.

- Castelletti, A., Galelli, S., Ratto, M., Soncini-Sessa, R., and Young, P. C.: A general framework for dynamic emulation modelling in environmental problems, *Environmental Modelling & Software*, 34, 5-18, 2012.
- Chandramouli, V., and Raman, H.: Multireservoir modeling with dynamic programming and neural networks, *Journal of Water Resources Planning and Management*, 127, 89-98, 2001.
- Choi, H. T., and Beven, K.: Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework, *Journal of Hydrology*, 332, 316-336, 2007.
- Clark, M. P., Slater, A. G., Barrett, A. P., Hay, L. E., McCabe, G. J., Rajagopalan, B., and Leavesley, G. H.: Assimilation of snow covered area information into hydrologic and land-surface models, *Advances in water resources*, 29, 1209-1221, 2006.
- Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J., and Uddstrom, M. J.: Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model, *Advances in water resources*, 31, 1309-1324, 2008.
- Clark, M. P., Kavetski, D., and Fenicia, F.: Pursuing the method of multiple working hypotheses for hydrological modeling, *Water Resources Research*, 47, 2011.
- Copernicus land monitoring service-CORINE land cover, available at: <https://land.copernicus.eu/pan-european/corine-land-cover>, last access: 29 August 2016.
- Cover, T. M., and Hart, P. E.: Nearest neighbor pattern classification, *IEEE transactions on information theory*, 13, 21-27, 1967.
- Crawford, N. H., and Linsley, R. K.: Digital simulation in hydrology, *Stanford Watershed Model IV*, Department of Civil Engineering, Stanford University, 1966.
- Croke, B.: The role of uncertainty in design of objective functions, *Proceedings of MODSIM 2007*, Christchurch, New Zealand, 2007.
- Croke, B., Wagener, T., Post, D., Freer, J., and Littlewood, I.: Evaluating the information content of data for uncertainty reduction in hydrological modelling, in *proceedings of the 4th international Congress on Environmental Modelling and Software*, Barcelona, Spain, 6-10 July 2008.
- Dadic, R., Mott, R., Lehning, M., and Burlando, P.: Wind influence on snow depth distribution and accumulation over glaciers, *Journal of Geophysical Research: Earth Surface*, 115, 2010.
- Dee, D. P., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., and Bauer, P.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Quarterly Journal of the royal meteorological society*, 137, 553-597, 2011.
- DeWalle, D. R., and Rango, A.: *Principles of snow hydrology*, Cambridge University Press, 2008.
- Duan, Q., Sorooshian, S., and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models, *Water resources research*, 28, 1015-1031, 1992.
- Duan, Q., Sorooshian, S., and Gupta, V. K.: Optimal use of the SCE-UA global optimization method for calibrating watershed models, *Journal of hydrology*, 158, 265-284, 1994.

- Dyrddal, A. V., Saloranta, T., Skaugen, T., and Stranden, H. B.: Changes in snow depth in Norway during the period 1961–2010, *Hydrology Research*, 44, 169-179, 2013.
- Ebrahimi, N., and Ghosh, S. K.: Bayesian and frequentist methods in change-point problems, *Handbook of statistics*, 20, 777-787, 2001.
- Eckley, I. A., Fearnhead, P., and Killick, R.: Analysis of changepoint models, *Bayesian Time Series Models*, 205-224, 2011.
- Efstratiadis, A., and Koutsoyiannis, D.: One decade of multi-objective calibration approaches in hydrological modelling: a review, *Hydrological Sciences Journal–Journal Des Sciences Hydrologiques*, 55, 58-78, 2010.
- El Naqa, I., and Murphy, M. J.: What is machine learning?, in: *Machine Learning in Radiation Oncology*, Springer, 3-11, 2015.
- El Naqa, I., and Murphy, M. J.: What is machine learning?, in: *Machine Learning in Radiation Oncology*, Springer, 3-11, 2015.
- Elgaali, E., and Garcia, L.: Using neural networks to model the impacts of climate change on water supplies, *Journal of Water Resources Planning and Management*, 133, 230-243, 2007.
- Emerick, A. A., and Reynolds, A. C.: Ensemble smoother with multiple data assimilation, *Computers & Geosciences*, 55, 3-15, 2013.
- Erxleben, J., Elder, K., and Davis, R.: Comparison of spatial interpolation methods for estimating snow distribution in the Colorado Rocky Mountains, *Hydrological Processes*, 16, 3627-3649, 2002.
- Essery, R., Li, L., and Pomeroy, J.: A distributed model of blowing snow over complex terrain, *Hydrological processes*, 13, 2423-2438, 1999.
- Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research: Oceans*, 99, 10143-10162, 1994.
- Evensen, G.: *Data assimilation: the ensemble Kalman filter*, Springer Science & Business Media, 2009.
- Farinotti, D., Magnusson, J., Huss, M., and Bauder, A.: Snow accumulation distribution inferred from time-lapse photography and simple modelling, *Hydrological processes*, 24, 2087-2097, 2010.
- Frey, S., and Holzmann, H.: A conceptual, distributed snow redistribution model, *Hydrology and Earth System Sciences*, 19, 4517-4530, 2015.
- Funtowicz, S. O., and Ravetz, J. R.: *Uncertainty and quality in science for policy*, Springer Science & Business Media, 1990.
- Garbrecht, J. D.: Comparison of three alternative ANN designs for monthly rainfall-runoff simulation, *Journal of Hydrologic Engineering*, 11, 502-505, 2006.
- Geman, S., and Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, in: *Readings in computer vision*, Elsevier, 564-584, 1987.

- Gilardi, N., and Bengio, S.: Local machine learning models for spatial data analysis, *Journal of Geographic Information and Decision Analysis*, 4, 11-28, 2000.
- Gisnås, K., Westermann, S., Schuler, T. V., Melvold, K., and Etzelmüller, B.: Small-scale variation of snow in a regional permafrost model, *The Cryosphere*, 10, 1201-1215, 2016.
- Globesar information page, available at: <https://startupmatcher.com/s/globesaras>, last access: 09 May 2018.
- Griessinger, N., Seibert, J., Magnusson, J., and Jonas, T.: Assessing the benefit of snow data assimilation for runoff modeling in Alpine catchments, *Hydrology and Earth System Sciences*, 20, 3895-3905, 2016.
- Gupta, V. K., and Sorooshian, S.: The relationship between data and the precision of parameter estimates of hydrologic models, *Journal of Hydrology*, 81, 57-77, 1985.
- Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G., and Ye, M.: Towards a comprehensive assessment of model structural adequacy, *Water Resources Research*, 48, 2012.
- Haan, C. T., Barfield, B. J., and Hayes, J. C.: *Design hydrology and sedimentology for small catchments*, Elsevier, 1994.
- Hall, K., George, R., Vincent, S., and Grid, V.: Updated daily MODIS/Terra Snow Cover Daily L3 Global 500m Grid V005, [April 2011 to August 2014], in: National Snow and Ice Data Center, Digital media, Boulder, Colorado USA, 2006.
- Han, E., Merwade, V., and Heathman, G. C.: Implementation of surface soil moisture data assimilation with watershed scale distributed hydrological model, *Journal of hydrology*, 416, 98-117, 2012.
- Hastings, W. K.: Monte Carlo sampling methods using Markov chains and their applications, 1970.
- Hegdahl, T. J., Tallaksen, L. M., Engeland, K., Burkhart, J. F., and Xu, C. Y.: Discharge sensitivity to snowmelt parameterization: a case study for Upper Beas basin in Himachal Pradesh, India. , *Hydrology Research*, 47, 683-700, 2016.
- Hiemstra, C. A., Liston, G. E., and Reiners, W. A.: Observing, modelling, and validating snow redistribution by wind in a Wyoming upper treeline landscape, *Ecological Modelling*, 197, 35-51, 2006.
- Hill, M. C.: Methods and guidelines for effective model calibration, in: *Building Partnerships*, 1-10, 2000.
- Hock, R.: Temperature index melt modelling in mountain areas, *Journal of hydrology*, 282, 104-115, 2003.
- Hornberger, G. M., and Spear, R. C.: Approach to the preliminary analysis of environmental systems, *J. Environ. Mgmt.*, 12, 7-18, 1981.
- Houser, P. R., and WALKER, J. P.: Hydrologic data assimilation, in: *Advances in water science methodologies*, CRC Press, 45-68, 2005.
- Houser, P. R., De Lannoy, G. J., and Walker, J. P.: Hydrologic Data Assimilation, in: *Approaches to Managing Disaster-Assessing Hazards, Emergencies and Disaster Impacts*, InTech, 2012.

- Hsieh, C.-t.: Some potential applications of artificial neural systems in, *Journal of Systems Management*, 44, 12, 1993.
- Hu, J., Liu, J., Liu, Y., and Gao, C.: EMD-KNN model for annual average rainfall forecasting, *Journal of Hydrologic Engineering*, 18, 1450-1457, 2011.
- Hume, D.: An enquiry concerning human understanding, in: *Seven Masterpieces of Philosophy*, Routledge, 191-284, 2016.
- Hussain, M. F., Barton, R. R., and Joshi, S. B.: Metamodeling: radial basis functions, versus polynomials, *European Journal of Operational Research*, 138, 142-154, 2002.
- Jacquin, A. P., and Shamseldin, A. Y.: Development of a possibilistic method for the evaluation of predictive uncertainty in rainfall-runoff modeling, *Water Resources Research*, 43, 2007.
- Jakeman, A. J., Letcher, R. A., and Norton, J. P.: Ten iterative steps in development and evaluation of environmental models, *Environmental Modelling & Software*, 21, 602-614, 2006.
- Jones, D. R.: A taxonomy of global optimization methods based on response surfaces, *Journal of global optimization*, 21, 345-383, 2001.
- Jordan, M. I., and Mitchell, T. M.: Machine learning: Trends, perspectives, and prospects, *Science*, 349, 255-260, 2015.
- Keating, E. H., Doherty, J., Vrugt, J. A., and Kang, Q.: Optimization and uncertainty assessment of strongly nonlinear groundwater models with high parameter dimensionality, *Water Resources Research*, 46, 2010.
- Kennedy, M. C., and O'Hagan, A.: Bayesian calibration of computer models, *Journal of the Royal Statistical Society: Series B*, 63, 425-464, 2001.
- Kerr, T., Clark, M., Hendrikx, J., and Anderson, B.: Snow distribution in a steep mid-latitude alpine catchment, *Advances in water resources*, 55, 17-24, 2013.
- Khaliq, M., Ouarda, T. B., Gachon, P., Sushama, L., and St-Hilaire, A.: Identification of hydrological trends in the presence of serial and cross correlations: A review of selected methods and their application to annual flow regimes of Canadian rivers, *Journal of Hydrology*, 368, 117-130, 2009.
- Killick, R., and Eckley, I.: changepoint: An R package for changepoint analysis, *Journal of statistical software*, 58, 1-19, 2014.
- Kingston, G.B., Maier, H.R. and Dandy, G.C. 2018, Review of Artificial Intelligence Techniques and their Applications to Hydrological Modeling and Water Resources Management. Part 1 – Simulation, available at: https://www.researchgate.net/publication/277005048_Review_of_Artificial_Intelligence_Techniques_and_their_Applications_to_Hydrological_Modeling_and_Water_Resources_Management_Part_1_-_Simulation, last access: 15 December 2018.
- Kirchner, J. W.: Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, *Water Resources Research*, 45, 2009.
- Kişî, Ö., and Öztürk, Ö.: Adaptive neurofuzzy computing technique for evapotranspiration estimation, *Journal of Irrigation and Drainage Engineering*, 133, 368-379, 2007.

- Kleijnen, J. P.: Kriging metamodeling in simulation: A review, *European journal of operational research*, 192, 707-716, 2009.
- Klir, G. J., and Yuan, B.: *Fuzzy sets and fuzzy logic: theory and applications*, Prentice Hall PTR New Jersey, 1995.
- Kolberg, S. A., and Gottschalk, L.: Updating of snow depletion curve with remote sensing data, *Hydrological Processes*, 20, 2363-2380, 2006.
- Kolberg, S., and Gottschalk, L.: Interannual stability of grid cell snow depletion curves as estimated from MODIS images, *Water Resources Research*, 46, 2010.
- Kundzewicz, Z. W., and Robson, A. J.: Change detection in hydrological records—a review of the methodology/revue méthodologique de la détection de changements dans les chroniques hydrologiques, *Hydrological sciences journal*, 49, 7-19, 2004.
- Lall, U., and Sharma, A.: A nearest neighbor bootstrap for resampling hydrologic time series, *Water Resources Research*, 32, 679-693, 1996.
- Lambert, A.: Catchment models based on ISO-functions, *J. Instn. Water Engrs*, 26, 413-422, 1972.
- Lehning, M., Bartelt, P., Brown, B., and Fierz, C.: A physical SNOWPACK model for the Swiss avalanche warning: Part III: Meteorological forcing, thin layer formation and evaluation, *Cold Regions Science and Technology*, 35, 169-184, 2002.
- Li, J., and Heap, A. D.: A review of spatial interpolation methods for environmental scientists, 2008.
- Li, J., Heap, A. D., Potter, A., and Daniell, J. J.: Application of machine learning methods to spatial interpolation of environmental variables, *Environmental Modelling & Software*, 26, 1647-1659, 2011.
- Li, D., Durand, M., and Margulis, S. A.: Estimating snow water equivalent in a Sierra Nevada watershed via spaceborne radiance data assimilation, *Water Resources Research*, 53, 647-671, 2017.
- Liston, G. E.: Interrelationships among snow distribution, snowmelt, and snow cover depletion: Implications for atmospheric, hydrologic, and ecologic modeling, *Journal of applied meteorology*, 38, 1474-1487, 1999.
- Liston, G. E.: Representing subgrid snow cover heterogeneities in regional and global models, *Journal of climate*, 17, 1381-1397, 2004.
- Liston, G. E., and Elder, K.: A distributed snow-evolution modeling system (SnowModel), *Journal of Hydrometeorology*, 7, 1259-1276, 2006.
- Liston, G. E., Haehnel, R. B., Sturm, M., Hiemstra, C. A., Berezovskaya, S., and Tabler, R. D.: Simulating complex snow distributions in windy environments using SnowTran-3D, *Journal of Glaciology*, 53, 241-256, 2007.
- Liston, G. E., and Hiemstra, C. A.: A simple data assimilation system for complex snow distributions (SnowAssim), *Journal of Hydrometeorology*, 9, 989-1004, 2008.
- Liu, Y., and Gupta, H. V.: Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, *Water Resources Research*, 43, 2007.

- Liu, Y., Freer, J., Beven, K., and Matgen, P.: Towards a limits of acceptability approach to the calibration of hydrological models: Extending observation error, *Journal of Hydrology*, 367, 93-103, 2009.
- Liu, J., and Han, D.: Indices for calibration data selection of the rainfall-runoff model, *Water resources research*, 46, 2010.
- Liu, Y., Weerts, A., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., and Van Dijk, A.: Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, 2012.
- Madsen, H.: Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives, *Advances in water resources*, 26, 205-216, 2003.
- Magnusson, J., Wever, N., Essery, R., Helbig, N., Winstral, A., and Jonas, T.: Evaluating snow models with varying process representations for hydrological applications, *Water Resources Research*, 51, 2707-2723, 2015.
- Mandel, J.: A brief tutorial on the ensemble Kalman filter, arXiv preprint arXiv:0901.3725, 2009.
- Mantovan, P., and Todini, E.: Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology, *Journal of hydrology*, 330, 368-381, 2006.
- Margulis, S. A., Giroto, M., Cortés, G., and Durand, M.: A particle batch smoother approach to snow water equivalent estimation, *Journal of Hydrometeorology*, 16, 1752-1772, 2015.
- Marler, R. T., and Arora, J. S.: Survey of multi-objective optimization methods for engineering, *Structural and multidisciplinary optimization*, 26, 369-395, 2004.
- Marofi, S., Tabari, H., and Abyaneh, H. Z.: Predicting spatial distribution of snow water equivalent using multivariate non-linear regression and computational intelligence methods, *Water resources management*, 25, 1417-1435, 2011.
- Matalas NC, L. J., Wolman MG: Prediction in water management, in: *Scientific basis of water resource management*, National Research Council, National Academy Press, Washington, DC, 27, 1982.
- Matt, F. N., Burkhart, J. F., and Pietikäinen, J.-P.: Modelling hydrologic impacts of light absorbing aerosol deposition on snow at the catchment scale, *Hydrology & Earth System Sciences*, 22, 2018.
- Maurer, E. P., Stewart, I., Bonfils, C., Duffy, P. B., and Cayan, D.: Detection, attribution, and sensitivity of trends toward earlier streamflow in the Sierra Nevada, *Journal of Geophysical Research: Atmospheres*, 112, 2007.
- Maybeck, P. S.: The Kalman filter: An introduction to concepts, in: *Autonomous robot vehicles*, Springer, 194-204, 1990.
- McLaughlin, D.: An integrated approach to hydrologic data assimilation: interpolation, smoothing, and filtering, *Advances in Water Resources*, 25, 1275-1286, 2002.
- McMillan, H. K., Clark, M. P., Bowden, W. B., Duncan, M., and Woods, R. A.: Hydrological field data from a modeller's perspective: Part 1. Diagnostic tests for model structure, *Hydrological Processes*, 25, 511-522, 2011.

- Melvold, K., and Skaugen, T.: Multiscale spatial variability of lidar-derived and modeled snow depth on Hardangervidda, Norway, *Annals of Glaciology*, 54, 273-281, 2013.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E.: Equation of state calculations by fast computing machines, *The journal of chemical physics*, 21, 1087-1092, 1953.
- Modaresi, F., Araghinejad, S., and Ebrahimi, K.: A comparative assessment of artificial neural network, generalized regression neural network, least-square support vector regression, and K-nearest neighbor regression for monthly streamflow forecasting in linear and nonlinear conditions, *Water Resources Management*, 32, 243-258, 2018.
- Montanari, A., Shoemaker, C. A., and van de Giesen, N.: Introduction to special section on Uncertainty Assessment in Surface and Subsurface Hydrology: An overview of issues and challenges, *Water Resources Research*, 45, 2009.
- Moradkhani, H.: Hydrologic remote sensing and land surface data assimilation, *Sensors*, 8, 2986-3004, 2008.
- Moradkhani, H., and Sorooshian, S.: General review of rainfall-runoff modeling: model calibration, data assimilation, and uncertainty analysis, in: *Hydrological modelling and the water cycle*, Springer, 1-24, 2009.
- Morris, M. D.: Factorial sampling plans for preliminary computational experiments, *Technometrics*, 33, 161-174, 1991.
- Mott, R., and Lehning, M.: Meteorological modeling of very high-resolution wind fields and snow deposition for mountains, *Journal of Hydrometeorology*, 11, 934-949, 2010.
- Mott, R., Egli, L., Grünwald, T., Dawes, N., Manes, C., Bavay, M., and Lehning, M.: Micrometeorological processes driving snow ablation in an Alpine catchment, *The Cryosphere*, 5, 1083-1098, 2011.
- Mulvany, T.: On the use of self-registering rain and flood gauges, *Making Observations of the Relations of Rain Fall and Flood Discharges in a Given Catchment. Transactions and Minutes of the Proceedings of the Institute of Civil Engineers of Ireland, Dublin, Ireland, Session, 1, 1850.*
- Nachtnebel, H., Baumung, S., and Lettl, W.: Abflussprognosemodell für das Einzugsgebiet der Enns und Steyr, Report, Institute of Water Management, Hydology and Hydraulic Engineering, University of Natural Resources and Applied Life Sciences Vienna, Austria, 1993.
- Nearing, G. S., Tian, Y., Gupta, H. V., Clark, M. P., Harrison, K. W., and Weijs, S. V.: A philosophical basis for hydrological uncertainty, *Hydrological Sciences Journal*, 61, 1666-1678, 2016.
- Neelakantan, T., and Pundarikanthan, N.: Neural network-based simulation-optimization model for reservoir operation, *Journal of water resources planning and management*, 126, 57-64, 2000.
- Nguyen, T., and de Kok, J.-L.: Systematic testing of an integrated systems model for coastal zone management using sensitivity and uncertainty analyses, *Environmental Modelling & Software*, 22, 1572-1587, 2007.

- Nyhus, E.: Implementation of GARTO as an infiltration routine in a full hydrological model (Master's Thesis), NTNU, 2017.
- O'Hagan, A.: Bayesian analysis of computer code outputs: A tutorial, *Reliability Engineering & System Safety*, 91, 1290-1300, 2006.
- Okun, O., and Priisalu, H.: Random forest for gene expression based cancer classification: overlooked issues, *Iberian Conference on Pattern Recognition and Image Analysis*, Girona, Spain, June 6-8, 2007.
- Olden, J. D., and Jackson, D. A.: Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks, *Ecological modelling*, 154, 135-150, 2002.
- Oreskes, N., Shrader-Frechette, K., and Belitz, K.: Verification, validation, and confirmation of numerical models in the earth sciences, *Science*, 263, 641-646, 1994.
- Papoulis, A., and Pillai, S. U.: *Probability, random variables, and stochastic processes*, Tata McGraw-Hill Education, 2002.
- Pappenberger, F., Frodsham, K., Beven, K., Romanowicz, R., and Matgen, P.: Fuzzy set approach to calibrating distributed flood inundation models using remote sensing observations, *Hydrology and Earth System Sciences Discussions*, 11, 739-752, 2007.
- Parajka, J., and Blöschl, G.: Validation of MODIS snow cover images over Austria, *Hydrology and Earth System Sciences Discussions*, 3, 1569-1601, 2006.
- Parker, P., Letcher, R., Jakeman, A., Beck, M., Harris, G., Argent, R. M., Hare, M., Pahl-Wostl, C., Voinov, A., and Janssen, M.: Progress in integrated assessment and modelling, *Environmental modelling & software*, 17, 209-217, 2002.
- Pechlivanidis, I., Jackson, B., McIntyre, N., and Wheater, H.: Catchment scale hydrological modelling: a review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications, *Global NEST journal*, 13, 193-214, 2011.
- Peel, M. C., and Blöschl, G.: Hydrological modelling in a changing world, *Progress in Physical Geography*, 35, 249-261, 2011.
- Perone, S. P., and Ham, C. L.: of Information Content in Electrochemical Experiments, *Journal of Research of the National Bureau of Standards*, 90, 1985.
- Petersen-Øverleir, A., Soot, A., and Reitan, T.: Bayesian rating curve inference as a streamflow data quality assessment tool, *Water resources management*, 23, 1835-1842, 2009.
- Pianosi, F., Sarrazin, F., and Wagener, T.: A Matlab toolbox for global sensitivity analysis, *Environmental Modelling & Software*, 70, 80-85, 2015.
- Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., and Wagener, T.: Sensitivity analysis of environmental models: A systematic review with practical workflow, *Environmental Modelling & Software*, 79, 214-232, 2016.
- Piazzini, G., Thirel, G., Campo, L., and Gabellani, S.: A particle filter scheme for multivariate data assimilation into a point-scale snowpack model in an Alpine environment, *The Cryosphere*, 12, 2287-2306, 2018.

- Popper, K.: *The Logic of Scientific Discovery*, Hutchinson & Co, London, 1959.
- Popper, K. R.: Science as falsification, *Conjectures and refutations*, 1, 33-39, 1963.
- Priestley, C., and Taylor, R.: On the assessment of surface heat flux and evaporation using large-scale parameters, *Monthly weather review*, 100, 81-92, 1972.
- Quinlan, J. R.: Decision trees and decision-making, *IEEE Transactions on Systems, Man, and Cybernetics*, 20, 339-346, 2006.
- Ratto, M., Castelletti, A., and Pagano, A.: Emulation techniques for the reduction and sensitivity analysis of complex environmental models, in, Elsevier, 2012.
- Razavi, S., Tolson, B. A., and Burn, D. H.: Review of surrogate modeling in water resources, *Water Resources Research*, 48, 2012.
- Refsgaard, J.: Terminology, modelling protocol and classification of hydrological model codes, *Distributed Hydrological Modelling*, 22, 17, 1996.
- Refsgaard, J. C.: Parameterisation, calibration and validation of distributed hydrological models, *Journal of hydrology*, 198, 69-97, 1997.
- Refsgaard, J. C., and Henriksen, H. J.: Modelling guidelines—terminology and guiding principles, *Advances in Water Resources*, 27, 71-82, 2004.
- Reichert, P., and Omlin, M.: On the usefulness of overparameterized ecological models, *Ecological Modelling*, 95, 289-299, 1997.
- Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W.: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resources Research*, 46, 2010.
- Revuelto, J., López-Moreno, J. I., Azorin-Molina, C., and Vicente-Serrano, S. M.: Topographic control of snowpack distribution in a small catchment in the central Spanish Pyrenees: intra- and inter-annual persistence, *The Cryosphere*, 8, 1989-2006, 2014.
- Rice, R., and Bales, R. C.: Embedded-sensor network design for snow cover measurements around snow pillow and snow course sites in the Sierra Nevada of California, *Water resources research*, 46, 2010.
- Ritter, A., and Muñoz-Carpena, R.: Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments, *Journal of Hydrology*, 480, 33-45, 2013.
- Ritzel, B. J., Eheart, J. W., and Ranjithan, S.: Using genetic algorithms to solve a multiple objective groundwater pollution containment problem, *Water Resources Research*, 30, 1589-1603, 1994.
- Ross, T. J.: *Fuzzy logic with engineering applications*, John Wiley & Sons, 2009.
- Sadegh, M., and Vrugt, J. A.: Bridging the gap between GLUE and formal statistical approaches: approximate Bayesian computation, *Hydrology and Earth System Sciences*, 17, 2013.
- Sajikumar, N., and Thandaveswara, B.: A non-linear rainfall-runoff model using an artificial neural network, *Journal of hydrology*, 216, 32-55, 1999.

- Saloranta, T.: Simulating snow maps for Norway: description and statistical evaluation of the seNorge snow model, *The Cryosphere*, 6, 1323-1337, 2012.
- Saltelli, A., Ratto, M., Tarantola, S., Campolongo, F., and Commission, E.: Sensitivity analysis practices: Strategies for model-based inference, *Reliability Engineering & System Safety*, 91, 1109-1125, 2006.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S.: *Global sensitivity analysis: the primer*, John Wiley & Sons, 2008.
- Sargent, R. G.: Simulation model verification and validation, *Proceedings of the 1991 Winter Simulation Conference*, 1991.
- Savenije, H. H.: Equifinality, a blessing in disguise? , *Hydrological processes*, 15, 2835-2838, 2001.
- Schmid, B. H., and Koskiahio, J.: Artificial neural network modeling of dissolved oxygen in a wetland pond: the case of Hovi, Finland, *Journal of Hydrologic Engineering*, 11, 188-192, 2006.
- Schulze, R. E.: *Hydrology and agrohydrology: A text to accompany the ACRU 3.00 agrohydrological modelling system*, Water Research Commission, 1995.
- Senent-Aparicio, J., Jimeno-Sáez, P., Bueno-Crespo, A., Pérez-Sánchez, J., and Pulido-Velázquez, D.: Coupling machine-learning techniques with SWAT model for instantaneous peak flow prediction, *Biosystems Engineering*, 177, 67-77, 2018.
- Senge, P. M., and Forrester, J. W.: Tests for building confidence in system dynamics models, *System dynamics, TIMS studies in management sciences*, 14, 209-228, 1980.
- Copernicus land monitoring service-CORINE land cover: <https://land.copernicus.eu/pan-european/corine-land-cover>, access: 29 August 2016.
- Shannon, C. E.: A mathematical theory of communication, *Bell system technical journal*, 27, 379-423, 1948.
- Shen, Z. Y., Chen, L., and Chen, T.: Analysis of parameter uncertainty in hydrological and sediment modeling using GLUE method: a case study of SWAT model applied to Three Gorges Reservoir Region, China., *Hydrology and Earth System Sciences*, 16, 2012.
- Sherman, L. K.: Streamflow from rainfall by the unit-graph method, *Eng. News Record*, 108, 501-505, 1932.
- Shrestha, D., Kayastha, N., and Solomatine, D.: A novel approach to parameter uncertainty analysis of hydrological models using neural networks, *Hydrology Earth System Sciences*, 13, 1235-1248, 2009.
- Singh, V. P., and Woolhiser, D. A.: Mathematical modeling of watershed hydrology, *Journal of hydrologic engineering*, 7, 270-292, 2002.
- Slater, A. G., and Clark, M. P.: Snow data assimilation via an ensemble Kalman filter, *Journal of Hydrometeorology*, 7, 478-493, 2006.
- Smith, L. A.: Disentangling uncertainty and error: On the predictability of nonlinear systems, in: *Nonlinear dynamics and statistics*, Springer, 31-64, 2001.

- Sonali, P., and Kumar, D. N.: Review of trend detection methods and their application to detect temperature changes in India, *Journal of Hydrology*, 476, 212-227, 2013.
- Statkraft information page, available at: <https://www.statkraft.com/>, last access: 20 June 2018.
- Stedinger, J. R., Vogel, R. M., Lee, S. U., and Batchelder, R.: Appraisal of the generalized likelihood uncertainty estimation (GLUE) method, *Water resources research*, 44, 2008.
- Sturm, M., and Wagner, A. M.: Using repeated patterns in snow distribution modeling: An Arctic example, *Water Resources Research*, 46, 2010.
- Sudheer, K., and Jain, A.: Explaining the internal behaviour of artificial neural network river flow models, *Hydrological Processes*, 18, 833-844, 2004.
- Suen, J.-P., and Eheart, J. W.: Evaluation of neural networks for modeling nitrate concentrations in rivers, *Journal of water resources planning and management*, 129, 505-510, 2003.
- Sun, W., Wang, Y., Wang, G., Cui, X., Yu, J., Zuo, D., and Xu, Z.: Physically based distributed hydrological model calibration based on a short period of streamflow data: case studies in four Chinese basins, *Hydrology and Earth System Sciences*, 21, 251, 2017.
- Tabari, H., Marofi, S., Abyaneh, H. Z., and Sharifi, M.: Comparison of artificial neural network and combined models in estimating spatial distribution of snow depth and snow water equivalent in Samsami basin of Iran, *Neural Computing Applications*, 19, 625-635, 2010.
- Taylor, W. A.: Change-point analysis: a powerful new tool for detecting changes, 2000. Available online: <http://www.claudiobellei.com/2016/11/15/changepoint-frequentist/> (accessed on 1 June 2018)
- Todini, E.: History and perspectives of hydrological catchment modelling, *Hydrology Research*, 42, 73-85, 2011.
- Toth, E., Brath, A., and Montanari, A.: Comparison of short-term rainfall prediction models for real-time flood forecasting, *Journal of hydrology*, 239, 132-147, 2000.
- Trajkovic, S., Todorovic, B., and Stankovic, M.: Forecasting of reference evapotranspiration by artificial neural networks, *Journal of Irrigation and Drainage Engineering*, 129, 454-457, 2003.
- Udnæs, H. C., Alfnes, E., and Andreassen, L. M.: Improving runoff modelling using satellite-derived snow covered area, *Hydrology Research*, 38, 21-32, 2007.
- Vaché, K. B., and McDonnell, J. J.: A process-based rejectionist framework for evaluating catchment runoff model structure, *Water Resources Research*, 42, 2006.
- Van Leeuwen, P. J.: Particle filtering in geophysical systems, *Monthly Weather Review*, 137, 4089-4114, 2009.
- Vieux, B. E., and Farajalla, N. S.: Capturing the essential spatial variability in distributed hydrological modelling: Hydraulic roughness, *Hydrological processes*, 8, 221-236, 1994.
- Viney, N. R., Bormann, H., Breuer, L., Bronstert, A., Croke, B. F., Frede, H., Gräff, T., Hubrechts, L., Huisman, J. A., and Jakeman, A. J.: Assessing the impact of land use change on hydrology by ensemble modelling (LUCHEM) II: Ensemble combinations and predictions, *Advances in water resources*, 32, 147-158, 2009.

- Vrugt, J. A., Bouten, W., Gupta, H. V., and Sorooshian, S.: Toward improved identifiability of hydrologic model parameters: The information content of experimental data, *Water Resources Research*, 38, 48-41-48-13, 2002.
- Vrugt, J. A., Gupta, H. V., Bouten, W., and Sorooshian, S.: A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water resources research*, 39, 2003.
- Vrugt, J. A., Diks, C. G., Gupta, H. V., Bouten, W., and Verstraten, J. M.: Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water resources research*, 41, 2005.
- Vrugt, J. A., Ter Braak, C. J., Gupta, H. V., and Robinson, B. A.: Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling, *Stochastic Environmental Research and Risk Assessment*, 23, 1011-1026, 2009.
- Wagener, T., McIntyre, N., Lees, M., Wheater, H., and Gupta, H.: Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis, *Hydrological Processes*, 17, 455-476, 2003.
- Wagener, T., and Gupta, H. V.: Model identification for hydrological forecasting under uncertainty, *Stochastic Environmental Research and Risk Assessment*, 19, 378-387, 2005.
- Walker, W. E., Harremoës, P., Rotmans, J., Van Der Sluijs, J. P., Van Asselt, M. B., Janssen, P., and Kreyer von Krauss, M. P.: Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support, *Integrated assessment*, 4, 5-17, 2003.
- Wang, L., van Meerveld, H., and Seibert, J.: When should stream water be sampled to be most informative for event-based, multi-criteria model calibration?, *Hydrology Research*, 48, 1566-1584, 2017.
- Weaver, W., and Shannon, C.: *The mathematical theory of information*, Urbana, Illinois Press, 1964.
- Webster, C. S., Kingston, D. G., and Kerr, T.: Inter-annual variation in the topographic controls on catchment-scale snow distribution in a maritime alpine catchment, New Zealand, *Hydrological processes*, 29, 1096-1109, 2015.
- Wilkinson, M.: Testing the null hypothesis: The forgotten legacy of Karl Popper?, *Journal of sports sciences*, 31, 919-920, 2013.
- Winstral, A., Elder, K., and Davis, R. E.: Spatial snow modeling of wind-redistributed snow using terrain-based parameters, *Journal of hydrometeorology*, 3, 524-538, 2002.
- Winstral, A., and Marks, D.: Long-term snow distribution observations in a mountain catchment: Assessing variability, time stability, and the representativeness of an index site, *Water Resources Research*, 50, 293-305, 2014.
- Xiong, L., and O'Connor, K. M.: An empirical method to improve the prediction limits of the GLUE methodology in rainfall-runoff modeling, *Journal of Hydrology*, 349, 115-124, 2008.
- Yang, J., Jakeman, A., Fang, G., and Chen, X.: Uncertainty analysis of a semi-distributed hydrologic model based on a Gaussian Process emulator, *Environmental Modelling Software*, 101, 289-300, 2018.

- Yapo, P. O., Gupta, H. V., and Sorooshian, S.: Multi-objective global optimization for hydrologic models, *Journal of hydrology*, 204, 83-97, 1998.
- Yu, J., Qin, X., and Larsen, O.: Applying ANN emulators in uncertainty assessment of flood inundation modelling: a comparison of two surrogate schemes, *Hydrological Sciences Journal*, 60, 2117-2131, 2015.
- Zadeh, L. A.: Fuzzy sets, *Information and control*, 8, 338-353, 1965.
- Zhang, X., Srinivasan, R., and Van Liew, M.: Approximating SWAT model using artificial neural network and support vector machine 1, *JAWRA Journal of the American Water Resources Association*, 45, 460-474, 2009.
- Zhao, Y., Taylor, J. S., and Chellam, S. J.: Predicting RO/NF water quality by modified solution diffusion model and artificial neural networks, *Journal of membrane science*, 263, 38-46, 2005.

Journal publications

Paper I

Teweldebrhan, A. T., Burkhart, J. F., and Schuler, T. V.: Parameter uncertainty analysis for an operational hydrological model using residual-based and limits of acceptability approaches, *Hydrology and Earth System Sciences*, 22, 5021-5039, 2018.



Parameter uncertainty analysis for an operational hydrological model using residual-based and limits of acceptability approaches

Aynom T. Teweldebrhan¹, John F. Burkhart^{1,2}, and Thomas V. Schuler¹

¹Department of Geosciences, University of Oslo, Oslo, Norway

²Statkraft, Oslo, Norway

Correspondence: Aynom T. Teweldebrhan (aynomtt@geo.uio.no)

Received: 28 March 2018 – Discussion started: 16 April 2018

Revised: 19 July 2018 – Accepted: 10 September 2018 – Published: 28 September 2018

Abstract. Parameter uncertainty estimation is one of the major challenges in hydrological modeling. Here we present parameter uncertainty analysis of a recently released distributed conceptual hydrological model applied in the Nea catchment, Norway. Two variants of the generalized likelihood uncertainty estimation (GLUE) methodologies, one based on the residuals and the other on the limits of acceptability, were employed. Streamflow and remote sensing snow cover data were used in conditioning model parameters and in model validation. When using the GLUE limit of acceptability (GLUE LOA) approach, a streamflow observation error of 25 % was assumed. Neither the original limits nor relaxing the limits up to a physically meaningful value yielded a behavioral model capable of predicting streamflow within the limits in 100 % of the observations. As an alternative to relaxing the limits, the requirement for the percentage of model predictions falling within the original limits was relaxed. An empirical approach was introduced to define the degree of relaxation. The result shows that snow- and water-balance-related parameters induce relatively higher streamflow uncertainty than catchment response parameters. Comparable results were obtained from behavioral models selected using the two GLUE methodologies.

important role in forecasting the local inflows to the system on scales ranging from hours to years. With due recognition of the need for accurate prediction of streamflow and snow storage, Statkraft (2018) has recently released a new modeling framework mainly tailored for an operational purpose. In this study, one of the conceptual models of this framework was subjected to uncertainty analysis. Conceptual hydrological models typically have one or more calibration parameters and commonly require some form of inverse modeling to estimate model parameters from observations (Crawford and Linsley, 1966). During calibration, equifinality arises when different parameter sets give equally good results in terms of predefined efficiency criteria (Beven, 1993; Savenije, 2001; Wagener et al., 2003). The generalized likelihood uncertainty estimation (GLUE) methodology (Beven and Binley, 1992) is an extension of the generalized sensitivity analysis concept of Hornberger and Spear (1981), and it accepts equifinality as a working paradigm for parameter calibration of hydrological models (Choi and Beven, 2007). It is based on the concept that all models of hydrological systems are highly simplified representations of reality (e.g., Reichert and Omlin, 1997), and hence it is expected to have several different model structures and parameter sets that describe the system in an adequate way (Blazkova and Beven, 2002). When dealing with nonlinear systems, the classic hydrological approach of using a single set of model parameters may lead to large predictive biases (e.g., Mantovan and Todini, 2006).

Hydrological modeling is affected by four main sources of uncertainty related to input data, validation data, model structure, and model parameters (e.g., Renard et al., 2010). Input data uncertainties may arise from measurement limitations and scaling issues, for example, due to forcing data

1 Introduction

Hydrological models have numerous applications of central importance to society including for planning, design, and management of environmental and water resources. The operation of hydropower systems is mainly constrained by the availability of water resources. Hydrological models play an

downscaling. Errors of the rating curve affect streamflow estimates and thereby lead to validation data uncertainty. Structural uncertainty may result from the underlying assumptions and simplifications in the model formulation as well as from application of the model to conditions inconsistent with the model structure (Tripp and Niemann, 2008). Parametric uncertainty reflects the inability to specify exact values of model parameters (Renard et al., 2010) and it may stem from errors in input data and observations used for model conditioning as well as be due to epistemic errors in model structure. An increased awareness of these modeling uncertainties and the need for quality control of such models requires the integration of uncertainty analysis into the modeling process from the very beginning (Beven, 1989; Saltelli et al., 2006; Refsgaard et al., 2007).

Uncertainty analysis techniques can be classified as frequentist or Bayesian approaches, probabilistic or non-probabilistic approaches (e.g., Montanari et al., 2009), or as formal or informal approaches (e.g., Vrugt et al., 2009). Among the most widely used techniques in hydrological modeling are the formal Bayesian and the GLUE methods (Jin et al., 2010). The formal Bayesian approach makes strong assumptions about the statistics of observed data; with the likelihood function defined based on assumptions about the nature of the residuals (Schoups and Vrugt, 2010). However, the choice of an adequate likelihood function has been the subject of considerable debate. According to Beven and Smith (2015), a formal probabilistic likelihood function will have limited value since non-stationary epistemic uncertainties cannot be adequately represented by a statistical model. In GLUE, the likelihood measure is associated with a parameter set and should ideally reflect all the different sources of uncertainty (Beven and Smith, 2015). The original GLUE methodology has been the subject of debate for using a subjectively set threshold of behavioral models (e.g., Mantovan and Todini, 2006; Stedinger et al., 2008; Clark et al., 2011; Nearing et al., 2016). This problem is common to most residual-based model selection methods (Schaeffli, 2016). The extended concept of behavioral models in the GLUE limits of acceptability approach (GLUE LOA) (Beven, 2006) attempts to overcome this drawback through use of error bounds of the observational dataset.

The GLUE LOA methodology involves specifying limits around some observational data within which model predictions are required to lie and thereby considered acceptable for the intended model application. The acceptability limits are set prior to running a model and, among other considerations, they are expected to take into account incommensurability and uncertainty in both the input and evaluation data (Beven, 2009). However, identification of models that reproduce the observed system behavior within the limits of measurement error is not easy due to time-varying errors in the input data and model structure (e.g., Beven, 2016). This difficulty is even more pronounced when input and other sources of errors are not explicitly accounted for in defining the LOA.

Good quality time series data and associated uncertainties are not always readily available. For example, in regulated catchments the inflow hydrograph is often estimated from changes in storage volume and outflows using the water balance equation. Thus, as in the case of our study catchment, no stage–discharge relationship exists for estimating the streamflow uncertainty using the usual practice, i.e., by fitting different rating curves. In such instances the alternative is to assume an observation error proportional to the observational data. However, the identification of behavioral models without due consideration to such less precise observation error estimates may lead to the rejection of a useful model (i.e., making a type II error). Some of the measures taken to minimize the risk of making a type II error when identifying behavioral models using the GLUE LOA include extending the limits (e.g., Blazkova and Beven, 2009; Liu et al., 2009) and using different model realizations for different periods of a hydrological year (e.g., Choi and Beven, 2007). In this study, instead of relaxing the limits, the percentage of observations where model predictions are required to fall within the acceptability limits was relaxed.

The GLUE methodology has been widely used in various disciplines (Beven, 2009; Efstratiadis and Koutsoyiannis, 2010) primarily due to its conceptual simplicity and ease of implementation. Further, its suitability for parallel implementation on distributed computer systems as well as its general strategy in dealing with equifinality in model calibration make it an appealing framework (Blasone et al., 2008; Shen et al., 2012; Mirzaei et al., 2015).

In this study model parameters were constrained using streamflow and the MODIS snow cover product (Hall et al., 2006). Multi-criteria model conditioning helps to reduce prediction uncertainty through improved parameter identification (e.g., Efstratiadis and Koutsoyiannis, 2010; Finger et al., 2015), and GLUE provides a flexible approach for using multi-criteria methods through different ways of combining measures. Besides streamflow, one of the observations commonly used in multi-criteria conditioning of rainfall-runoff models in snow-dominated catchments is snow data. Remote sensing snow cover data have been used in several hydrological modeling studies for deriving and updating a snow depletion curve (SDC) (e.g., Lee et al., 2005; Kolberg and Gottschalk, 2006; Bavera et al., 2012), as well as in multi-criteria-based model calibration and simulated snow cover validation (e.g., Udnaes et al., 2007; Parajka and Blöschl, 2008; Berezowski et al., 2015). However, studies involving combined uncertainty of streamflow and snow cover predictions using the GLUE methodology are still missing in the literature.

The main objective of this study is to assess parameter uncertainty for a recently developed distributed conceptual hydrological model using the GLUE methodology with due consideration to the model's main application as an operational hydrological model. The second objective is to investigate the potential value of snow cover data as additional

observation in conditioning model parameters in the study area. The third objective is to assess the possibility of using a time-relaxed GLUE LOA approach for constraining model parameters. In doing so, we employ a novel empirical approach for implicitly accounting for the effects of input and observational data errors by relaxing the percentage of time steps in which predictions of model realizations fall within the limits.

This paper is organized as follows. First the (i) hydrological model and (ii) the study site and relevant data used in this study are briefly described in Sect. 2.1 and 2.2. The procedures followed to set up the uncertainty analyses are then outlined in Sect. 2.3. In Sect. 3, the results from parameter uncertainty as well as the uncertainty of streamflow and snow cover predictions using the residual-based GLUE approach are presented. The results from the relaxed GLUE LOA are also presented in this section. Finally, in Sects. 4 and 5, the analysis results and their implication on the hydrologic model, the data and the methodologies followed are discussed and conclusions are drawn.

2 Methods and materials

2.1 The hydrological model

The Statkraft Hydrological Forecasting Toolbox, Shyft (<https://github.com/statkraft/shyft>, last access: 1 March 2018), is an open-source distributed hydrological modeling framework developed by Statkraft (Burkhart et al., 2016). The modeling framework has three main models (method stacks) and, in this study, the PT_GS_K model was used for uncertainty analysis. PT_GS_K is a conceptual model with several adjustable parameters depending on the climatic and physiographic characteristics of the study area where the model is applied. This model requires temperature, precipitation, radiation, relative humidity, and wind speed as forcing data. PT_GS_K uses the Priestley–Taylor (PT) method (Priestley and Taylor, 1972) for estimating potential evaporation; a quasi-physical-based method for snowmelt, sub-grid snow distribution and mass balance calculations (GS method); and a simple storage–discharge function (Lambert, 1972; Kirchner, 2009) for catchment response calculation (K). Overall, these three methods constitute the PT_GS_K model in Shyft. The framework establishes a sequence of spatially distributed cells of arbitrary size and shape. As such it can provide lumped (single cell) or discretized (spatially distributed) calculations, as in this study. The model was applied to each of the grid cells and for each time step.

Within the GS method, precipitation falling in each grid cell is classified as solid or liquid precipitation depending on a threshold temperature (t_x) and on the local temperature values. The snowmelt energy is the sum effect of different energy sources in the system such as shortwave and long-wave

radiation as well as the turbulent sensible and latent energy fluxes. Among other factors, the energy contribution from shortwave radiation depends on snow albedo. For a given time step (t), the snow albedo of each grid cell depends on the minimum (α_{\min}) and maximum (α_{\max}) albedo values as well as on air temperature (T_a) (Eq. 1). In this method the decay rates of albedo due to snow ageing as a function of temperature, i.e., the fast (fast ADR, α_{fdr}) and slow (slow ADR, α_{sdr}) albedo decay rates corresponding to temperature conditions above and below 0°C , respectively, are parameterized. Turbulent heat contribution is the sum of latent and sensible heat. Wind turbulence is linearly related to wind speed using a wind constant and wind scale from the intercept and slope of the linear function, respectively (Hegdahl et al., 2016).

$$\alpha_t = \begin{cases} \alpha_{\min} + (\alpha_{t-1} - \alpha_{\min}) \cdot \left(\frac{1}{2^{1/\alpha_{\text{fdr}}}} \right) & T_a > 0^\circ\text{C} \\ \alpha_{t-1} + (\alpha_{\max} - \alpha_{\min}) \cdot \left(\frac{1}{2^{1/\alpha_{\text{sdr}}}} \right) & T_a \leq 0^\circ\text{C} \end{cases} \quad (1)$$

The sub-grid snow distribution is described by a three-parameter gamma probability distribution snow depletion curve (Liston, 1999; Kolberg and Gottschalk, 2006). The traditional gamma distribution is parameterized with two values, i.e., the average amount of snow at the onset of the melt season m (mm) and the shape value (k), based on the assumption that the ground is completely snow covered before the onset of melt. Since this assumption may not hold true for a number of grid cells especially in alpine areas, a third parameter representing the bare ground fraction at the onset of the snowmelt season has been introduced (Kolberg and Gottschalk, 2006). The two-parameter gamma distribution (Eq. 2) is thus applied only to the remaining portion of a grid cell to estimate the fraction of the initially snow-covered area where snow has disappeared (y'). The initial bare ground fraction parameter is constant for all years. At each time step, the state parameters such as snow water equivalent (SWE) and snow cover area (SCA) are updated using the SDC function. In the GS method, the shape value is a direct transformation of the sub-grid snow coefficient of variation (CV_s).

$$y' = \int_0^{\lambda(t)} f(x; k, \theta) dx = \gamma\left(k, \frac{\lambda}{\theta}\right), \quad (2)$$

where f denotes the gamma probability density function and γ is the incomplete gamma function. x and $\lambda(t)$, respectively, refer to point snow storage and the accumulated melt depth (mm) at time t since the onset of the melt season. θ represents the scale parameter with $m = k\theta$ and $k = \text{CV}_s^{-2}$.

The catchment response function (CRF) is based on the storage–discharge relationship concept described in Kirchner (2009) and represents the sensitivity of discharge to changes in storage (Eq. 3). This method is based on the idea that catchment sensitivity to changes in storage, i.e., $g(Q)$, can be estimated from the time series of discharge alone

Table 1. Range of model parameters used for the PT_GS_K model stack uncertainty analysis.

Name	Min.	Max.	Description	Method
c1	−5.0	1.0	constant in the catchment response function (CRF)	K
c2	0.0	1.2	linear coefficient in CRF	K
c3	−0.15	−0.05	quadratic coefficient in CRF	K
tx	−3.0	2.0	Solid or liquid threshold temperature (°C)	GS
Wind scale	1.0	6.0	slope in turbulent wind function	GS
Fast ADR	1.0	15.0	fast albedo decay rate (days)	GS
Slow ADR	20.0	40.0	slow albedo decay rate (days)	GS
Snow CV	0.06	0.85	spatial coefficient of variation of snowfall	GS

through fitting empirical functions to the data such as the quadratic equation. Since discharge is generally nonlinear and typically varies by many orders of magnitude, the recommended approach is to use log-transformed discharge values in order to avoid the risk of numerical instability. In this method, the three parameters of the catchment response function, i.e., c_1 , c_2 , and c_3 , are parameterized.

$$\frac{d(\ln(Q))}{dt} = g(Q) \left(\frac{P - E}{Q} - 1 \right), \quad (3)$$

with $g(Q) = e^{c_1 + c_2(\ln(Q)) + c_3(\ln(Q))^2}$,

in which E and Q , respectively, represent actual evapotranspiration and discharge. In the original formulation P refers to precipitation, whereas in this method it refers to the liquid water supply from rainfall and snowmelt.

The potential evaporation calculation in the PT method requires net radiation and the slope of saturated vapor pressure as well as the Priestley–Taylor parameter, the psychrometric constant, and the latent heat of vaporization (e.g., Matt et al., 2018). The latter three variables are kept constant in the PT method. Actual evapotranspiration is assumed to take place only from snow-free areas and it is estimated as a function of potential evapotranspiration and a scaling factor.

In the default parameter settings of the PT_GS_K model seven parameters are considered as influential and thus allowed to vary in conditioning the model. Preliminary model calibration using the BOBYQA algorithm (Powell, 2009) and the default setting gave reasonable model performance. Hence, the same setting was also followed in this study with the addition that the sub-grid snow coefficient of variation was also considered an uncertain model parameter. A similar result was also observed when this setting was later verified using the method of Morris (Morris, 1991; Saltelli et al., 2008) for screening the most influential out of the relevant model parameters. The feasible ranges of parameter values are set based on relevant literature and previous modeling studies in the Nea-Nidelva catchment. Table 1 shows a list of these parameters with their range of possible values.

2.2 Study area and data

This study was conducted using climatic and catchment data from the Nea catchment (11.67390–12.46273° E, 62.77916–63.20405° N). The Nea catchment constitutes the headwaters of the Nea-Nidelva water resources management area which is situated in Sør-Trøndelag county, Norway (Fig. 1). The hydropower generated from this area is the main source of electric supply to several places in mid-Norway including to one of the biggest cities in the country, Trondheim. As a result this area has significance for Statkraft AS and other stakeholders responsible for the development and management of water resources in the region and has been selected for research focused on better prediction and understanding of the snow processes and their impact on hydrology of the downstream area.

The Nea catchment covers a total area of 703 km² and it is characterized by a wide range of physiographic and land cover characteristics. Altitude of the catchment ranges from 1783 m a.s.l. on the eastern part around the mountains of Storsylen to 649 m a.s.l. at its outlet on the western part of the catchment. Mean annual precipitation for the hydrological years 2011–2014 was 1120 mm. The highest and lowest average daily temperature values for this period were 28 and −30 °C, respectively.

As mentioned in Sect. 2.1, the PT_GS_K model requires temperature, precipitation, radiation, relative humidity, and wind speed as forcing data. In this study, daily time series data of these variables for the study area were obtained from Statkraft (2018) as point measurements, with the exception of relative humidity. Daily gridded relative humidity data were retrieved from ERA-Interim (Dee et al., 2011). The Model uses a Bayesian kriging approach to distribute the point temperature data over the domain, while for the other forcing variables it uses an inverse distance weighting approach.

Two observational datasets, streamflow and snow cover, were used in this study. Daily observed streamflow measurements covering 4 hydrological years (1 September to 31 August) were provided for the study area. The climatic data show that these hydrological years represented periods both above and below the long-term average annual precipitation.

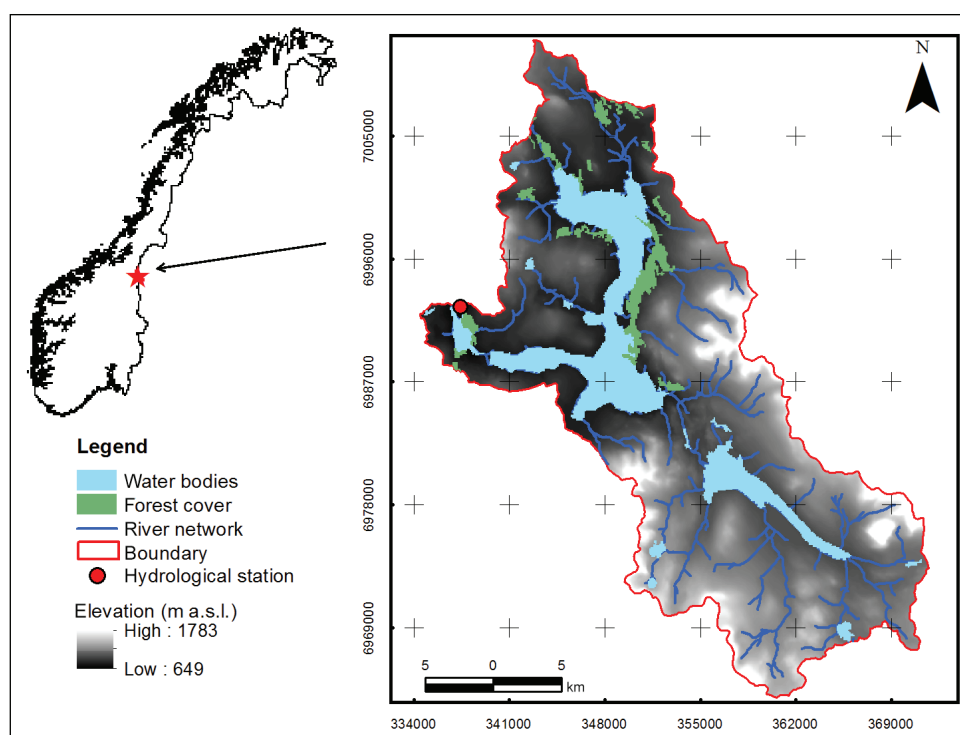


Figure 1. Physiographic and location map of the Nea catchment in Norway.

Years 2011 and 2013, respectively, were the wettest and driest years in over 10 years. Daily snow cover fraction (SCF) data were retrieved from NASA MODIS snow cover products (MODIS SCF) (Hall et al., 2006). Frequent cloud cover is one of the major challenges when using MODIS and other optical remote sensing data in Norway. In order to minimize the effect of obstructions and misclassification errors emanating from clouds and other sources, a composite dataset was formed using data retrieved from the Aqua and Terra satellites, MYD10A1 and MOD10A1 products, respectively.

In this analysis, PT_GS_K was set up in distributed mode over 812 grid cells, requiring the following physiographic data of each grid cell: average elevation and grid cell total area, as well as the areal fractions of forest, reservoir, lake, and glacier. Data for these physiographic variables were retrieved from two sources: the land cover data from Copernicus land monitoring service (2016) and the 10 m digital elevation model (10 m DEM) from the Norwegian mapping authority (2016).

2.3 The uncertainty analysis methods

In this study a modeling and parameter uncertainty analysis was conducted using two GLUE variants. First, the hydrological model and its snow sub-model were subjected to uncertainty analysis using the residual-based GLUE methodology. When using this approach, the relevant model parameters were initially conditioned using either streamflow or

MODIS SCF. In the subsequent analysis, they were conditioned using both streamflow and SCF. Following that, the uncertainty analysis was conducted using the relaxed GLUE LOA approach.

2.3.1 Sampling the parameter dimensions

The performance of all uncertainty analysis techniques depends on the efficiency of the sample in representing the entire response surface (Pappenberger et al., 2008). In this study, prior distributions of the uncertain model parameters were not known and hence a uniform distribution was assumed. The challenge in using uniform distribution is, however, to adequately sample the entire parameter dimensions. To overcome this challenge and to better identify regions of behavioral simulations, a sample size of 100 000 runs was used. Each model run is a realization of a parameter set randomly drawn from the domains of the model parameters. An all-at-a-time (AAT) sampling method (e.g., Pianosi et al., 2016) was employed. This method involves random selection of all parameter values simultaneously. The residual-based GLUE (Sect. 2.3.2) and the relaxed GLUE LOA (Sect. 2.3.3) approaches are used to identify the behavioral model runs. Matlab scripts from the SAFE toolbox (Pianosi et al., 2015) were used as a basis to characterize behavioral and non-behavioral models.

2.3.2 The residual-based GLUE approach

In this study, the performance of each model realization was evaluated by using relevant likelihood measures. Residual-based informal likelihood measures are considered suitable measures of fit when large datasets such as rainfall-runoff time series exist for model conditioning (Hassan et al., 2008). The Nash–Sutcliffe efficiency (NSE, Eq. 4) belongs to these groups of likelihood measures, and it is the most widely used likelihood measure for assessing the fitness of model parameters in hydrological modeling (Xiong and O’Connor, 2008). Further, the main end users of the model commonly use NSE both in calibration and evaluation of hydrological models. Thus, use of this performance measure as a streamflow likelihood measure makes it easier both in setting the threshold value for behavioral models (i.e., based on previous experience) and in communicating model performance outputs. However, the NSE calculated using raw values tends to overestimate model performance during peak streamflow and underestimate during low-streamflow conditions (e.g., Krause et al., 2005). To partly overcome this problem, NSE is often calculated with log-transformed observed and simulated values. In this study, both NSE and NSE with log-transformed streamflow values (LnNSE) were thus employed as likelihood measures in evaluating each model run.

$$\text{NSE} = 1 - \frac{\sum_{i=1}^n (Q_{\text{sim},i} - Q_{\text{obs},i})^2}{\sum_{i=1}^n (Q_{\text{obs},i} - \bar{Q}_{\text{obs}})^2}, \quad (4)$$

in which Q_{sim} represents simulated streamflow, Q_{obs} is observed streamflow, and \bar{Q}_{obs} represents the mean value of observed streamflow series.

Within the residual-based GLUE procedure, the definition of threshold likelihood value at which the model performance is judged reasonable is a subjective choice by the modeler. In this study, NSE and LnNSE of 0.7 and 0.6, respectively, were considered as the threshold values for behavioral models. These values were chosen with due consideration to the input and observational data quality as well as the relative importance given to high streamflow in relation to low-streamflow conditions in the hydropower industries. In the case of the combined likelihood measure, a weighted average threshold value (e.g., Hassan et al., 2008) was calculated assuming each likelihood measure to have a weight proportional to its threshold value. Accordingly, the NSE and LnNSE likelihood measures were respectively assigned weights of 0.54 and 0.46 (Eq. 5).

$$L_{\text{NS}}(O | M(\theta_i)) = 0.46(L_{\text{LnNSE}}) + 0.54(L_{\text{NSE}}), \quad (5)$$

where $L_{\text{NS}}(O | M(\theta_i))$ represents the combined likelihood measure for the i th model realization with model prediction of $M(\theta_i)$, which is a function of the set of model parameters θ_i , and corresponding to the observations (O). L_{NSE}

Table 2. Setup of the two-by-two contingency table for binary snow cover data comparison. O and S , respectively, represent observed and simulated binary snow cover and the subscripts refer to a snow-free (0) and snow-covered (1) grid cell.

	S_1	S_0	Sum
O_1	n_{11}	n_{01}	n_{x1}
O_0	n_{10}	n_{00}	n_{x0}
Sum	n_{1x}	n_{0x}	n_{xx}
$\text{CSI} = \frac{n_{11}}{n_{xx} - n_{00}}$			

and L_{LnNSE} , respectively, represent the likelihood measures based on NSE and LnNSE. Models producing likelihood measure values greater than or equal to the threshold value were labeled as behavioral models and were retained for use in further analysis.

The root mean squared error (RMSE) of simulated and MODIS fractional snow cover was used as a likelihood measure of SCF. A threshold value of 0.17 was set when using the RMSE in model conditioning. This value was fixed based on the average performance of similar conceptual hydrological models as a reference (e.g., Skaugen and Weltzien, 2016) and with due consideration to the inherent error in the MODIS SCF data. The estimated annual average error of MODIS SCF maps for the Northern Hemisphere is approximately 8 % in the absence of cloud (Pu et al., 2007), and in forest-dominated areas it may reach up to 15 % (Hall et al., 2001).

Preliminary assessment of model performance indicates that the snow yes/no-based model performance (critical success index, CSI; Table 2) is very high both before the onset of snowmelt and during the complete melt-out period. The lowest match between simulated and MODIS SCF was observed during early summer. It was thus decided to use a weighted mean likelihood measure of SCF, with maximum weight assigned to likelihoods from the middle part of the observation period. The likelihood of each SCF observation was assigned a specific weight based on the location of the observation date in a trapezoidal membership function (TMF). The start and end of the MODIS SCF observation period locate the feet of the trapezoid and the start and end of the month of June locate the shoulders (Fig. 2). For each model realization, the weighted average RMSE (wRMSE) of all SCF observations and their corresponding simulated values for the calibration period were calculated and model realizations with wRMSE below the threshold value were considered behavioral. The weight of each behavioral model was calculated as the inverse of wRMSE and was used in constructing the cumulative distribution function (CDF), based on which the predicted SCF values for different quantiles can be extracted.

When selecting behavioral models using the combined likelihoods of streamflow and SCF, the merging of these likelihoods was carried out in two steps. First the likelihoods representing low- and high-flow conditions, viz. LnNSE and

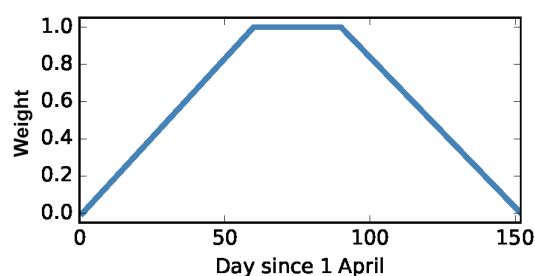


Figure 2. A trapezoidal membership function for SCF likelihoods in the observational period.

NSE, were combined following a similar procedure as described above. The likelihoods of streamflow and SCF were separately rescaled such that their respective weights would sum to unity following a similar procedure to that used in Brazier et al. (2000). The combined streamflow likelihood and the SCF likelihood were subsequently multiplied to get a combined likelihood measure of streamflow and SCF.

2.3.3 The relaxed GLUE LOA approach

Unlike the residual-based model selection approaches, including the residual-based GLUE methodology, the GLUE LOA approach relies on an assessment of uncertainty in the observational data. The uncertainty analysis was also thus conducted in this study using the GLUE LOA approach and its results compared against those from the residual-based GLUE methodology.

In this study when using the GLUE LOA approach, both the streamflow and MODIS SCF data were considered as uncertain observations. Since no uncertainty data were available for streamflow observations in the study site, mean streamflow uncertainty of 25 % was assumed and the streamflow limits were defined using this value. Although, the maximum expected error of MODIS snow cover products under clear-sky conditions is reported to be 15 % for forest areas (Hall et al., 2001), cloud coverage coupled with a lack of contrast between clouds and snow cover may severely affect the accuracy. And in some cases this leads to misclassification of snow as land (e.g., Parajka et al., 2012). Thus, a SCF uncertainty of 25 %–50 % was assumed to represent the errors associated with the SCF observations and the input data.

An alternative approach was employed to minimize the risk of rejecting useful model realizations due to using assumed average observational error bounds and due to a lack of a viable means for explicitly accounting for the time-varying level of observational and input data uncertainties. The procedure involves relaxing the percentage of observations where model predictions fall within the acceptability limits. Model realizations whose predictions fall within the acceptable bounds in a defined percentage of the observations were considered behavioral. The minimum acceptable percentage of observations where model predictions

fall within the limits (hereafter referred as threshold pLOA) in turn was set such that the 5 %–95 % prediction limit of streamflow, reported as the containing ratio (CR, see Eq. 6), is close to the value obtained using the residual-based GLUE methodology. The procedure for relaxing the original GLUE LOA requirement during the calibration period involves the following steps.

- Step 1: define an acceptable prediction limit (CR) at a chosen certainty level (e.g., 5 %–95 %). In this study the CR value obtained for the calibration period using the residual-based GLUE methodology was adopted as an acceptable CR value.
- Step 2: relax the acceptable percentage of observations where model predictions fall within the limits. This is done by gradually lowering the requirement for bracketing the observations in 100 % of the time steps up to the acceptable pLOA.
- Step 3: run a calibration and test whether each model realization prediction falls within the limits at least for the specified percentage of the total observations. If model realizations that satisfy the relaxed acceptability criteria are found, proceed to step 4, otherwise lower the threshold pLOA further and repeat this step.
- Step 4: calculate the new CR and check if it is close to the predefined acceptable CR value. If the calculated CR is less than the predefined CR, repeat steps 2 to 4, whereas if the two CR values are close (e.g., within 5 %) then accept all model realizations that satisfy this pLOA as behavioral and store their indices for use in further analysis.

Model realizations that fulfill this relaxed LOA criteria both in streamflow and SCF observations were considered behavioral. A triangular membership function was used to define the weights of each criterion, where a maximum weight of 1.0 was assigned to predictions with a perfect match to the observation and a minimum weight of 0.0 to predictions outside the acceptability limits. For each model realization, the weights of individual time steps were added to give a generalized weight. Following the procedure by Blazkova and Beven (2009), the weights associated with streamflow and MODIS SCF were combined by taking the sum of these two criteria and rescaling them such that the sum of the weights for behavioral models is unity. The behavioral model realizations were used for prediction weighted by their overall degree of performance.

2.3.4 GLUE output analysis

A split-sample-based cross-validation of streamflow predictions was used to alternately evaluate how well the behavioral models identified at a given calibration period are able to reproduce the observed values from another period. The

hydrologic model was run for 4 years at a daily time step. The first month of each hydrological year was considered as a spin-up period and hence excluded from all uncertainty analyses. Each of the 4 years was alternately used to identify behavioral models and the remaining 3 years were individually used to assess the modeling uncertainty.

In this study the modeling uncertainty was evaluated using both qualitative and quantitative evaluation techniques. The upper and lower streamflow prediction limits as well as observed values were plotted on the same graph to visually assess the capability of the identified behavioral models in bracketing the observations. The containing ratio (CR) index was also used to analyze the prediction uncertainty following a similar procedure to that used in some studies involving the GLUE methodology (e.g., Xiong et al., 2009; He et al., 2011). CR is expressed as the ratio of the number of observations falling within respective prediction bounds to the total number of observations (Eq. 6).

$$CR = \frac{\sum_{i=1}^n I(Q_{\text{obs}, i})}{n},$$

where

$$I(Q_{\text{obs}, i}) = \begin{cases} 1, & L_{\text{lim}, i} < Q_{\text{obs}, i} < U_{\text{lim}, i} \\ 0, & \text{Otherwise} \end{cases} \quad (6)$$

$Q_{\text{obs}, i}$ represents observed streamflow at the i th time step, and $L_{\text{lim}, i}$ and $U_{\text{lim}, i}$ are the lower and upper prediction bounds, respectively.

As an alternative to a crisp prediction for an observation (e.g., Xiong and O'Connor, 2008), the median (50 %) streamflow prediction was also estimated from the behavioral model simulations and compared against observations using both NSE and LnNSE as goodness-of-fit measures. Similarly, the critical success index (Table 2) and RMSE were used as goodness-of-fit measures for median SCF prediction. When using RMSE, the fractional snow cover data of each grid cell were directly employed in validating median predictions. CSI represents the number of grid cells where the snow events are correctly predicted out of the total number of grid cells where snow is predicted in the model. It was calculated based on binary snow cover data using the two-by-two contingency table analysis (Table 2) following a similar procedure to that used in Hanzer et al. (2016). When converting the snow cover fraction to a binary measure, a grid cell was classified as snow covered if at least 50 % of its area is snow covered.

3 Results

3.1 Uncertainty analysis using the residual-based GLUE approach

3.1.1 Uncertainty of model parameters

The uncertainty of model parameters was analyzed using all years of record together as single time series data. The dot plots (Fig. 3) depict the goodness-of-fit response surface projected onto individual parameter dimensions. The parallel coordinate plots (Fig. 4) also show the distribution of model parameters within their respective parameter dimensions. The distribution of behavioral simulations across a parameter dimension varies from one parameter to another. The behavioral models are scattered nearly across the entire range of parameter dimension for fast ADR, slow ADR, and snow CV, indicating low model sensitivity to these parameters. On the other hand, the relatively localized distribution of behavioral models towards lower values when projected onto the parameter ranges of c_1 , c_2 , tx , and wind scale as well as towards higher values of c_3 reflects higher sensitivity of simulated streamflow to these calibration parameters. Furthermore, the parallel coordinate plots show an increase in likelihood measure value towards the lower (for c_1 , c_2 , tx , and wind scale) and higher (for c_3) parts of their respective parameter dimensions.

The aforementioned less sensitive model parameters can, however, have a high effect on model outputs through interaction with other parameters. Some degree of interaction between model parameters can be seen from the correlation shown in Fig. 5. For example, a general decreasing trend in model performance can be noticed with a joint increase in c_1 and c_2 . The strong influence of tx in constraining the output is also evident in these plots. A considerable level of interaction can also be observed from the correlation coefficient scores between c_1 and c_2 (0.56), c_2 and c_3 (0.53) and between tx and wind scale (0.66).

The posterior distribution histograms (Fig. 6) and the statistical summary table of posterior distribution (Table 3) illustrate variability in distribution characteristics of the model parameters. The catchment response parameters, viz. c_1 , c_2 , and c_3 , showed relatively well-defined peaks, whereas fast ADR, slow ADR, and snow CV appear less identifiable with a relatively flat distribution across their respective parameter dimensions. It should, however, be noted that, in the GLUE methodology, it is the set of parameter values that gives a behavioral model.

3.1.2 Uncertainty of streamflow predictions

Figure 7 shows a sample cross-validation of daily streamflow prediction limits against observed values. The upper and lower prediction bounds as well as the median values are generated with behavioral models identified in year 2011 us-

Table 3. Statistical summary of the posterior distribution for model parameters.

Statistics	c1	c2	c3	tx (°C)	Wind scale	Fast ADR (days)	Slow ADR (days)	Snow CV
Minimum	-5.00	0.00	-0.12	-3.00	1.01	1.00	20.07	0.06
Maximum	-2.32	0.70	-0.05	1.98	3.74	14.96	39.98	0.85
Mean	-3.90	0.22	-0.07	-1.39	2.40	7.38	30.21	0.46
Median	-3.92	0.20	-0.07	-1.57	2.48	7.01	30.71	0.47
Variance	0.33	0.02	0.00	1.15	0.48	15.46	33.15	0.05
Skewness	0.18	0.53	-0.58	0.81	-0.22	0.19	-0.05	-0.06

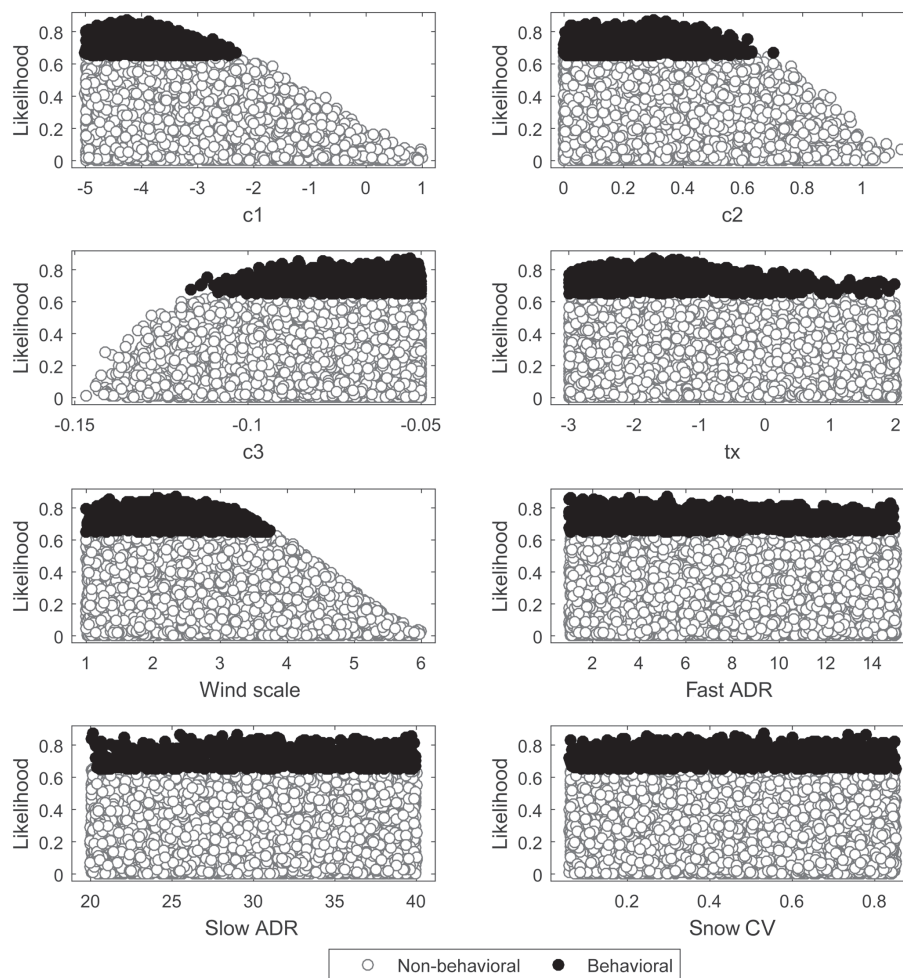


Figure 3. Dotty plots of the likelihood measure for behavioral and non-behavioral models identified using the residual-based GLUE methodology.

ing the combined NSE and LnNSE likelihood measure. The calculated uncertainty in streamflow prediction indicated by the 5–95 percentile range (shaded band) varied over time and relatively higher uncertainty was noticed during high-streamflow than low-streamflow periods.

As can be seen from the summary table of cross-validation results (Table 4), the CR values range from 0.62 to 0.91 with an overall mean value of 0.77. The mean CR values for the

calibration and validation periods are 0.78 and 0.76, respectively. The evaluation result generally shows that the median prediction of behavioral models selected using the combined likelihood was able to reproduce the observed values remarkably well with average NSE and LnNSE of 0.86 and 0.72, respectively, for the validation period. However, performance of the behavioral models identified using NSE was very low when evaluated using LnNSE in year 2014. This

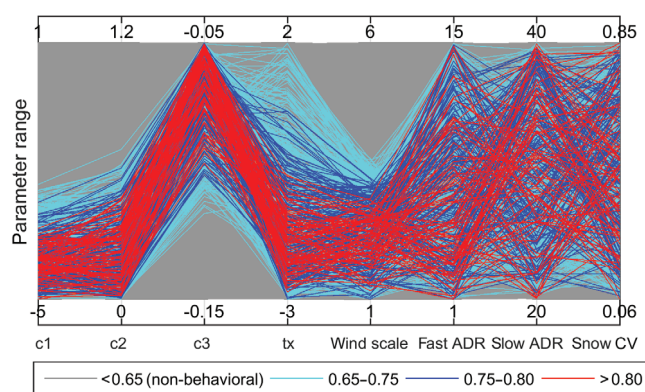


Figure 4. Distribution of model parameters within their variability ranges.

phenomenon can be attributed to the relatively low quality of streamflow observations during the low-streamflow period of this year. The validation result was also highly affected by the nature of the likelihood measure used during the identification of behavioral models. For example, a persistent low performance was observed during the early months of the hydrologic year when validating behavioral models identified using NSE alone (Fig. 7c) as compared to those identified using the combined likelihood (Fig. 7d). Similarly, excluding the first 30 observations from the validation dataset resulted in an improvement of LnNSE from -0.53 to 0.44 .

3.1.3 Uncertainty of snow cover predictions

Snow cover fractions and snow water equivalent are two main outputs of the snow sub-model (GS) of the PT_GS_K model. In this study an initial single-likelihood-based conditioning of the GS specific parameters was carried out using MODIS SCF only and RMSE as a measure of model performance.

The cross-validation result of predicted median values against MODIS SCF observations is shown in Table 5. The highest and lowest RMSE values during the calibration period were 0.15 and 0.06, respectively, with an average RMSE value of 0.11. Minimum and maximum RMSE values of 0.06 and 0.22, respectively, were observed during the validation period with an average RMSE value of 0.13. Similarly the lowest CSI values during the calibration and validation periods were 0.99 and 0.88, respectively. Comparable maximum CSI results were observed between the two periods. The 5%–95% SCF prediction interval was able to reasonably bracket the observations in most of the calibration and validation periods with mean CR values of 0.60 and 0.71, respectively, without any explicit accounting for model residuals for each parameter set. The inter-annual comparison of model performance shows that relatively lower performance was observed in years 2011 and 2012 as compared to the other periods.

3.1.4 Uncertainty of streamflow and snow cover predictions using both observations

The cross-validation result of simulated streamflow and SCF against observations is shown in Table 6. A similar model performance was observed when model parameters are conditioned using both streamflow and MODIS SCF as compared to when only streamflow was used for model conditioning. The mean NSE and LnNSE values of the median streamflow prediction in the validation periods were 0.85 and 0.71, respectively. The average streamflow prediction uncertainty (CR) in the validation period was 0.70. For SCF, average RMSE and CSI values of 0.11 and 0.99, respectively, were obtained when using the combined likelihood. The streamflow and SCF median predictions obtained in this analysis are similar to the results when model parameters are respectively conditioned with streamflow only or MODIS SCF only. This result shows that contribution from the MODIS SCF was less significant in constraining the model parameters. The relatively low quality of MODIS SCF data as compared to the streamflow data for the study site may also partly explain this phenomenon.

3.2 Uncertainty analysis using the relaxed GLUE LOA approach

The median streamflow prediction of behavioral models identified using the relaxed GLUE LOA was able to mimic the observed values very well with a mean NSE and LnNSE of 0.85 and 0.7, respectively, for the validation period (Table 7). A comparable performance was observed between models selected using the residual-based GLUE and the relaxed GLUE LOA. The similarity in median predicted streamflow by these two GLUE methodologies can also be noticed from visual comparison of the resulting hydrographs (Figs. 7 and 8). A mean streamflow CR value of 0.75 was obtained for the validation period when using the relaxed GLUE LOA. This shows slightly better capability of the 5%–95% prediction bounds in bracketing the observations as compared to predictions using the residual-based GLUE methodology when both streamflow and SCF are used in model conditioning.

The behavioral models selected using the relaxed GLUE LOA approach were also able to adequately reproduce observed SCF with a mean RMSE and CSI of 0.11 and 0.98, respectively, for the validation period. Generally, high prediction uncertainty of SCF was observed during the onset of snowmelt and low uncertainty during the summer with an average CR of 0.63. Thus, hydrological year 2011, having most of its observations coming from April, showed the lowest CR as compared to the other periods. Figure 9 shows observed and simulated average catchment SCF for the sample calibration period (2011) and validation period (2012). From this figure it can be noticed that the median prediction tends to overestimate the observed SCF values, and many



Figure 5. Model performance in response to the interaction between model parameters (upper diagonal cells) and correlation coefficient scores between the parameters (lower diagonal cells)

of the observed values from the month of April fall outside the 5%–95% prediction bounds. The overall result, however, indicates an improved capability of the 5%–95% prediction bounds in bracketing the SCF observations as compared to predictions using the residual-based GLUE methodology.

4 Discussion

The streamflow prediction uncertainty analyses results show that model performance was relatively lower during low-streamflow than high-streamflow conditions throughout most validation periods (e.g., Table 4). A similar result was reported by Choi and Beven (2007) in their multiperiod cluster-based uncertainty analysis in the Bukmoon catchment, South Korea, where a high percentage of simulation bias was observed during the drier seasons due to relatively poor model performance during these periods. The result of this study

is thus consistent with the general observation that catchment hydrologic models perform relatively well in wet conditions but break down during low-streamflow conditions (e.g., Kirchner, 2009). In the case of results from the residual-based GLUE methodology, this can also be partly attributed to the nature of the likelihood measure used to identify the behavioral models. The result reveals this observation, where model performance during low-streamflow periods (LnNSE) was improved when using the combined likelihood measures as compared to using NSE alone. This is because models identified using NSE alone strongly reflect the hydrologic characteristics of the high-streamflow periods and are expected to perform more poorly during low-streamflow conditions.

In order to assess the potential value of MODIS SCF in constraining model parameters, the snow sub-model parameters were constrained using this observation and the posterior distribution of the individual parameters were com-

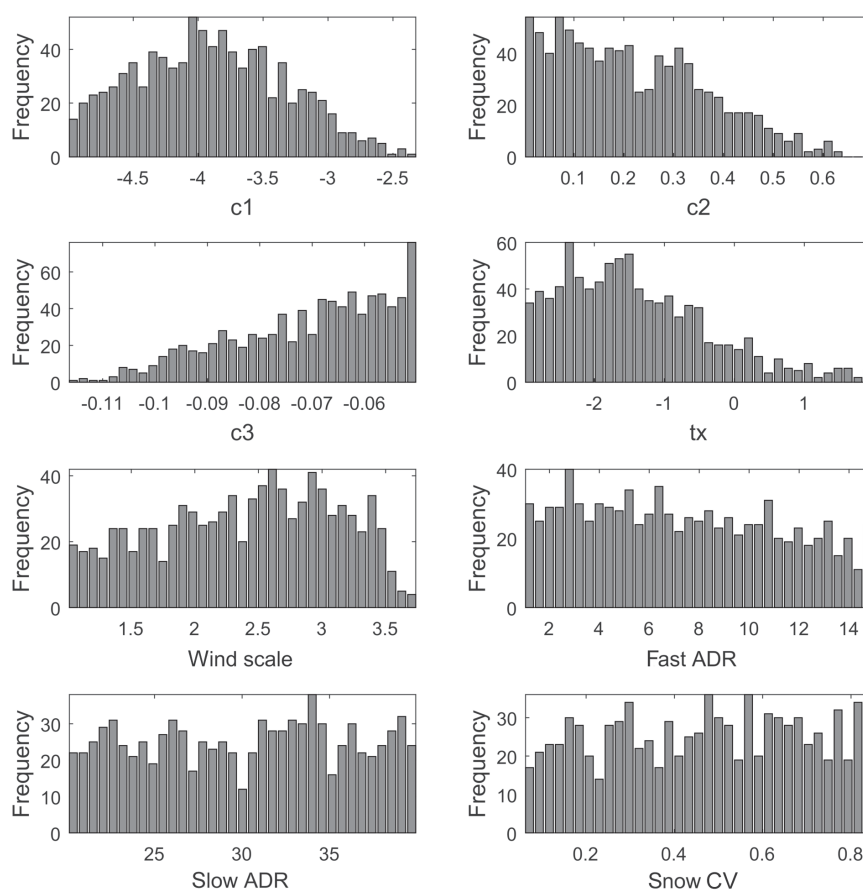


Figure 6. Posterior distribution of calibration parameters after conditioning on flow observations.

Table 4. Cross-validation of streamflow predictions against observed values. Bold numbers show the result for the calibration period.

Validation year	Likelihood (LH) measure	Calibration year							
		2011		2012		2013		2014	
		NSE	Comb. LH	NSE	Comb. LH	NSE	Comb. LH	NSE	Comb. LH
2011	NSE	0.893	0.890	0.770	0.806	0.809	0.790	0.697	0.840
	LnNSE	0.712	0.855	0.366	0.812	0.693	0.719	0.521	0.771
	CR	0.759	0.721	0.756	0.677	0.805	0.764	0.729	0.710
2012	NSE	0.842	0.869	0.920	0.930	0.818	0.787	0.910	0.874
	LnNSE	0.753	0.878	0.694	0.890	0.640	0.616	0.685	0.792
	CR	0.885	0.844	0.866	0.844	0.907	0.882	0.852	0.803
2013	NSE	0.922	0.925	0.878	0.877	0.934	0.942	0.862	0.916
	LnNSE	0.780	0.914	0.391	0.799	0.887	0.936	0.531	0.792
	CR	0.778	0.759	0.759	0.666	0.830	0.830	0.756	0.622
2014	NSE	0.828	0.884	0.860	0.892	0.826	0.810	0.901	0.924
	LnNSE	-0.346	0.566	-0.529	0.531	0.138	0.488	0.268	0.716
	CR	0.737	0.658	0.721	0.666	0.773	0.721	0.718	0.647
No. of behavioral models		1573	749	3737	1031	4725	2245	4648	604

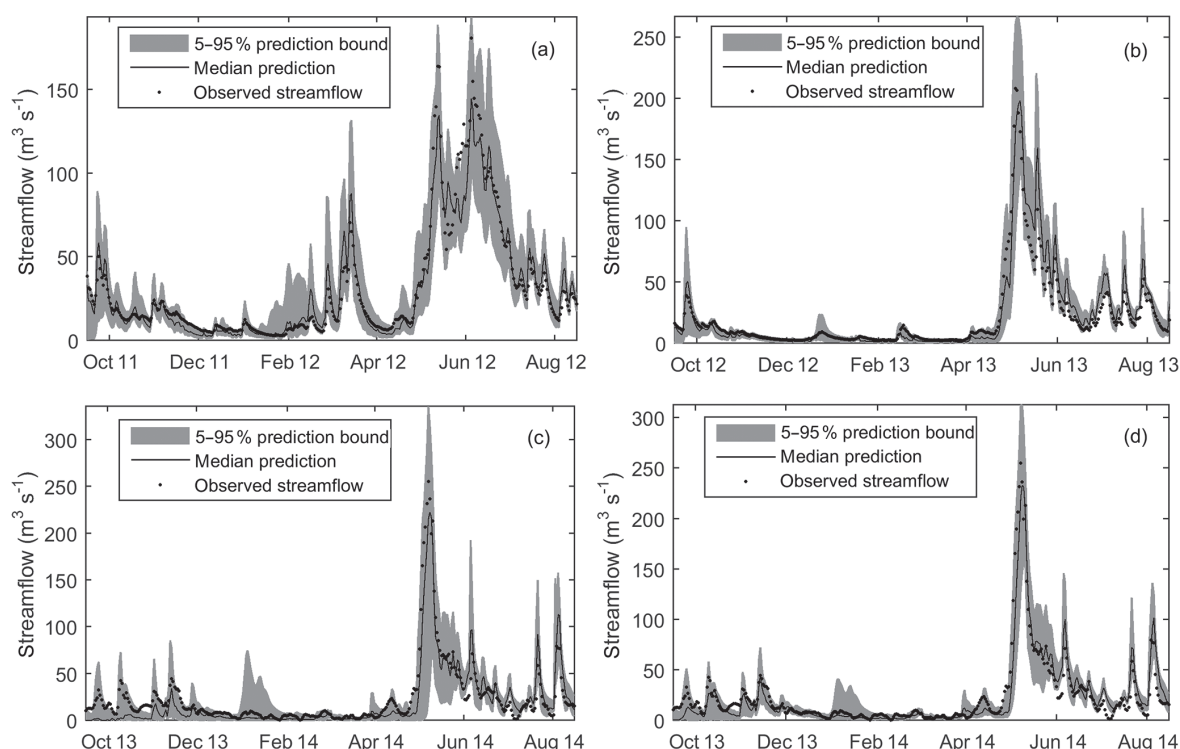


Figure 7. Median, 5–95 percentile range, and observed values of streamflow for the sample calibration period (a) and validation periods (b, c and d). The calibration result (a) and the validation results presented in (b) and (d) are based on behavioral models identified using the combined likelihood, while the result shown in (c) is based on behavioral models identified using NSE alone.

Table 5. Cross-validation of SCF predictions against MODIS SCF.

Calib. year	Validation year												No. of behav. models
	2011			2012			2013			2014			
	RMSE	CSI	CR	RMSE	CSI	CR	RMSE	CSI	CR	RMSE	CSI	CR	
2011	0.147	0.987	0.417	0.152	0.999	0.330	0.067	0.985	0.839	0.089	0.991	0.656	83922
2012	0.150	0.987	0.347	0.154	0.998	0.236	0.076	0.978	0.824	0.095	0.989	0.629	84945
2013	0.200	0.878	0.924	0.217	0.875	0.795	0.057	0.985	0.919	0.100	0.948	0.931	98400
2014	0.146	0.982	0.738	0.151	0.983	0.632	0.057	0.988	0.903	0.083	0.992	0.799	95039

pared against corresponding distributions that resulted from model conditioning using streamflow only. Parameter inference based on SCF only as a conditioning observation gave some parameter estimates that deviate significantly from those obtained when conditioned with streamflow only (Fig. 10). The box plots depict the posterior distribution of the snow-related parameters separately conditioned using streamflow and SCF. For the ease of comparison, parameter values were scaled between 0 and 1. From these plots it can be seen that tx and wind scale are the model parameters most sensitive to the conditioning data type with a significant shift in their quartiles towards the upper part of the parameter dimensions when conditioned using SCF, whereas the fast ADR, slow ADR, and snow CV did not show significant dis-

placement in their posterior distribution. These parameters were also identified as the least sensitive model parameters when the model was constrained using streamflow only.

Generally, in snow models with the sub-grid snow distribution component parameterized using the statistical probability distribution function, low snow CV results in a faster depletion rate of the snow-covered fraction (e.g., Liston, 2004). Thus, the slight displacement of snow CV posterior values towards the lower part of its parameter dimensions coupled with the increased posterior values of wind scale would give rise to lower snow cover fraction during the melting period when model parameters are constrained using SCF only. On the other hand, the increased posterior values of the rain–snow threshold (tx) would result in an increase in snow

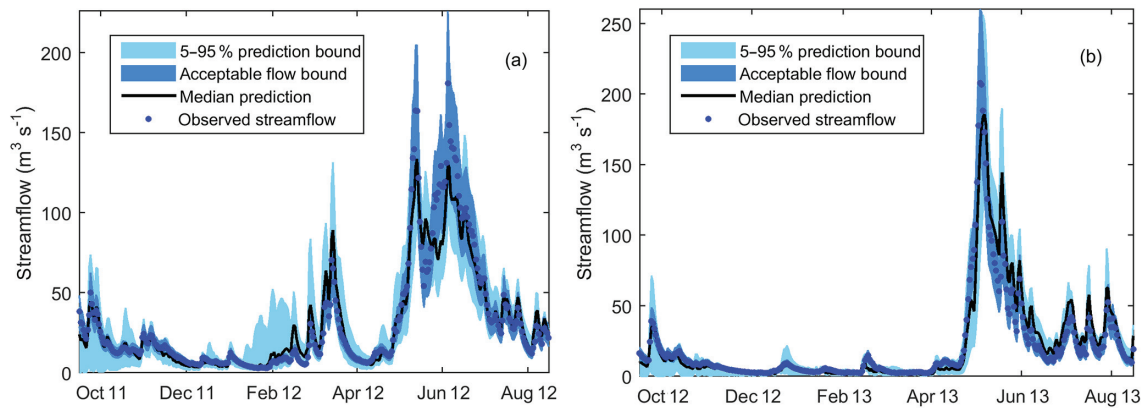


Figure 8. Prediction and acceptable flow bounds for the sample calibration period (a) and validation period (b).

Table 6. Cross-validation of streamflow and SCF predictions.

Validation year	Obs.	Likelihood measure	Calibration year			
			2011	2012	2013	2014
2011	flow	NSE	0.888	0.773	0.790	0.841
		LnNSE	0.856	0.780	0.711	0.769
		CR	0.660	0.611	0.753	0.693
	SCF	RMSE	0.142	0.146	0.155	0.143
		CSI	0.987	0.987	0.954	0.987
		CR	0.461	0.341	0.610	0.430
2012	flow	NSE	0.855	0.939	0.738	0.886
		LnNSE	0.886	0.869	0.602	0.791
		CR	0.811	0.803	0.852	0.811
	SCF	RMSE	0.158	0.150	0.165	0.150
		CSI	0.985	0.999	0.960	0.992
		CR	0.363	0.232	0.504	0.334
2013	flow	NSE	0.914	0.874	0.946	0.917
		LnNSE	0.913	0.749	0.941	0.785
		CR	0.679	0.605	0.827	0.619
	SCF	RMSE	0.053	0.063	0.049	0.055
		CSI	0.992	0.987	0.994	0.990
		CR	0.846	0.824	0.869	0.841
2014	flow	NSE	0.878	0.895	0.789	0.928
		LnNSE	0.513	0.481	0.485	0.717
		CR	0.627	0.627	0.712	0.647
	SCF	RMSE	0.079	0.087	0.078	0.078
		CSI	0.996	0.993	0.990	0.996
		CR	0.681	0.625	0.743	0.658
No. of acceptable models			726	988	2245	604

Table 7. Cross-validation of streamflow and SCF predictions after relaxing the LOA criteria.

Validation year	Obs.	Likelihood measure	Calibration year			
			2011	2012	2013	2014
2011	flow	NSE	0.881	0.861	0.769	0.854
		LnNSE	0.839	0.838	0.711	0.796
		CR	0.712	0.726	0.759	0.748
	SCF	RMSE	0.140	0.145	0.152	0.142
		CSI	0.983	0.987	0.959	0.985
		CR	0.551	0.450	0.615	0.552
2012	flow	NSE	0.808	0.914	0.758	0.837
		LnNSE	0.822	0.918	0.595	0.791
		CR	0.797	0.833	0.866	0.852
	SCF	RMSE	0.162	0.150	0.161	0.153
		CSI	0.970	0.995	0.963	0.986
		CR	0.417	0.342	0.516	0.439
2013	flow	NSE	0.947	0.896	0.940	0.941
		LnNSE	0.940	0.880	0.934	0.914
		CR	0.767	0.707	0.825	0.800
	SCF	RMSE	0.049	0.057	0.051	0.052
		CSI	0.994	0.989	0.992	0.991
		CR	0.857	0.843	0.871	0.862
2014	flow	NSE	0.872	0.859	0.787	0.898
		LnNSE	0.540	0.307	0.310	0.674
		CR	0.641	0.627	0.704	0.671
	SCF	RMSE	0.077	0.082	0.079	0.078
		CSI	0.994	0.994	0.989	0.995
		CR	0.706	0.661	0.748	0.713
No. of acceptable models			419	813	2213	1029

deposition and thereby in a partial or full canceling out of the effects of changes in snow CV and wind scale. This phenomenon may thus lead to equifinality, where different sets of model parameters give comparable SCF responses.

In the GLUE LOA approach a particular model realization is classified as acceptable if its prediction falls within the lim-

its for all observed values. In continuous rainfall-runoff modeling it is difficult for all predictions of a given model realization to lie within the observation limits in a time series. In some cases this phenomenon can be attributed to different specific processes dominating the hydrologic behavior of a catchment at different sub-periods, while in other instances

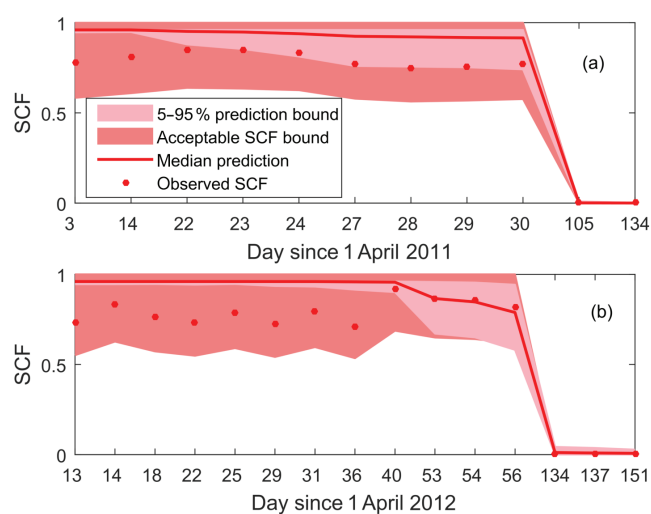


Figure 9. Prediction and acceptable bounds of average SCF for the sample calibration period (a) and validation period (b).

it may be due to a lack of a viable means for explicitly accounting for the effect of variable sources and level of uncertainties from the input data errors, which are difficult to set a priori. Thus, the time-varying likely effects of other sources of errors such as input errors on prediction uncertainty need also be implicitly taken into account when defining the limits of acceptability.

The use of GLUE LOA for testing hydrologic models as hypotheses without a due consideration of errors in input data may lead to a rejection of useful models that might adequately represent the catchment behavior and thereby to making a type II error (false negative). In the past, various attempts have been made to minimize the risk of making type II errors in model calibration studies using the GLUE and other frameworks. In some studies an improved calibration of hydrologic models was obtained through independent calibration of sub-periods of a time series (e.g., Boyle et al., 2000; Samanta and Mackay, 2003). When it comes to the GLUE LOA approach, extending the limits (e.g., Blazkova and Beven, 2009; Liu et al., 2009) and using different model realizations for different periods of a hydrological year (e.g., Choi and Beven, 2007) are some of the measures taken to minimize the risk of making type II errors. Common to all these measures is that they attempt to relax the selection criteria for behavioral models.

In this study when using the GLUE LOA approach, the streamflow bounds were set to $\pm 25\%$ and the result shows that none of the model realizations were able to satisfy the LOA criteria without one or more of their predictions falling outside the acceptable streamflow bounds. The failure rate was higher during low-flow conditions as compared to high-flow conditions. An initial attempt was made to relax the limit of acceptability by extending the streamflow bounds. Regardless, no model realization with its predictions falling

within the error bounds for all observations was found until the limits were extended to over $\pm 85\%$. This relaxed acceptability limit seems less reasonable in terms of its physical meaning as an error bound. Therefore, rather than relaxing the limits, an alternative empirical approach was followed by relaxing the number of simulation time steps which fulfilled the original LOA criterion. The procedure involves defining the acceptable percentage of observations that are required to be bracketed by model predictions (during the calibration period) based on a predefined acceptable CR value.

This empirical approach is based on the observed relationship between prediction uncertainty and number of behavioral models, which in turn is a function of the selection criterion. As the threshold value of a likelihood measure increases (in the case of residual-based GLUE) or absolute value of the limits decreases (in the case of GLUE LOA), the simulated runoff series gradually converges, though not necessarily to the observations. A similar observation was also reported in other GLUE-based uncertainty studies (e.g., Xiong et al., 2008). A further analysis in this study reveals that, as the percentage of observations required to be bracketed by each model realization (pLOA) increases, the number of behavioral models decreases and thereby the simulated runoff series converges, resulting in a low CR (Fig. 11). In this study, the threshold pLOA for each calibration period was defined in such a way that the 5%–95% prediction uncertainties of streamflow using the residual and the LOA-based GLUE methodologies are similar. Defining the threshold pLOA this way helps to set a reasonable value that minimizes the risk of making type II errors while maintaining the overall model accuracy by rejecting the inclusion of non-behavioral models. Furthermore, it helps to roughly compare the performance of behavioral models selected using the relaxed GLUE LOA against the residual-based GLUE in terms of their ability to reproduce the median streamflow and SCF predictions at a similar level of uncertainty (i.e., the CR used to set pLOA).

Although it is difficult to single out the effects of input data error from model structural error on model performance using the GLUE methodology, the error patterns may aid in assessing model performance in different periods of the hydrologic year. Generally, a good model structure coupled with good data is not expected to give a consistent bias (e.g., Liu et al., 2009). Figure 12 shows a sample daily percentage of acceptable simulations satisfying the LOA criteria during the hydrologic year 2012. The percentage of the acceptable number of model realizations in each time step was generally low during the calibration period ($< 65\%$). However, for each time step, predictions from some behavioral models are able to mimic the corresponding observation within the assumed error bound. The percentage of acceptable models was relatively higher during high than low-streamflow conditions. And this result is consistent with the general observation that most hydrological models perform relatively well during high-streamflow compared to low-streamflow periods. The

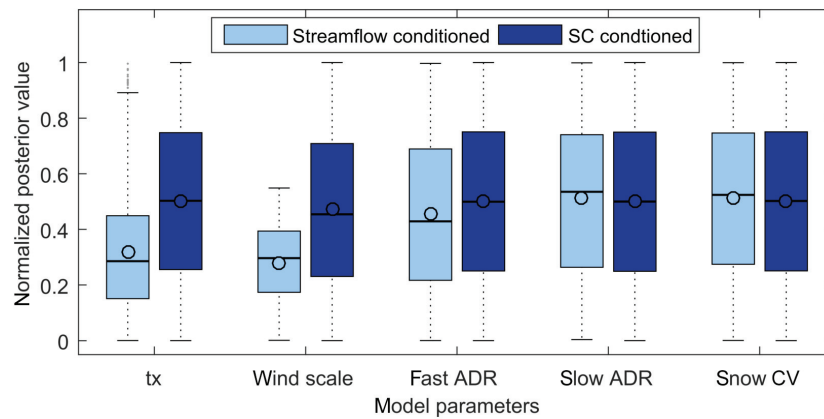


Figure 10. Box plot showing posterior distribution of model parameters when separately conditioned using streamflow and SCF. Parameter values are scaled between 0 and 1.

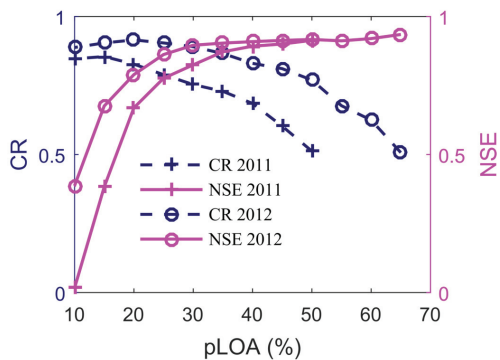


Figure 11. The effect of the percentage of observations required to be bracketed by each model realization (pLOA) on prediction uncertainty (CR) and efficiency of the median prediction (NSE) for the sample calibration periods (years 2011 and 2012).

spike in the percentage of acceptable models in the month of February 2012 when time steps around are so low, however, reveals how model performances can unexpectedly vary between time steps in response to input data errors and/or the observational error bounds. The observed spike could thus be attributed to relatively low input data errors and/or lower actual observational error bounds as compared to the assumed average values for the particular time step. The distribution of the behavioral model weights over the calibration period shows that the mean weight during the period where the spike occurred is very low. Similarly the median weight of behavioral models during this period is close to zero, implying that most of the model realizations have their predictions that barely fall within the limits.

This result reveals that the GLUE LOA with relaxation in percentage of observations where model predictions fall within observational error bounds can be used as an alternative approach for conditioning model parameters and conducting an uncertainty analysis when there is a lack of meta-

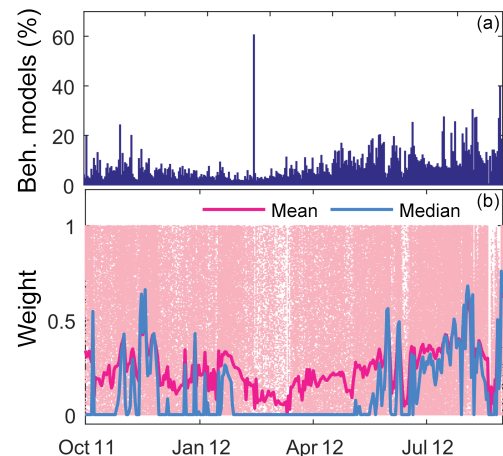


Figure 12. Daily percentage of acceptable model realizations with their predictions falling within the observation error bounds (a) and the daily weight associated with each acceptable model realization as well as daily mean and median value of the weights (b) in a sample calibration period.

data on input and observational data uncertainty coupled with a highly time-varying level of uncertainty from such sources. After relaxation, a limited sample of the total observations, i.e., 30%–40% of a hydrologic year, was able to effectively identify behavioral models, and this result is consistent with findings of other studies dealing with the effect of observation size on constraining model parameters (e.g., Seibert and Beven, 2009; Liu and Han, 2010; Sun et al., 2017). The relative accuracy of an event and other factors that affect the information content of the input and observation datasets (e.g., Beven and Smith, 2015) are more important than the length of the datasets, especially in continuous rainfall-runoff modeling.

5 Conclusions

Two GLUE methodology variants were applied for parameter uncertainty analysis of a distributed conceptual hydrological model. The analysis result from the residual-based GLUE methodology shows that the catchment response parameters, viz. c_1 , c_2 , and c_3 as well as the wind scale, are the most sensitive model parameters. More caution is thus required when defining the value range of these parameters. On the other hand, the fast and slow albedo decay rates as well as the snow CV are relatively more uncertain model parameters.

Model conditioning using combined streamflow and MODIS SCF did not improve the median prediction of streamflow as compared to the result when model parameters are conditioned using streamflow only. A similar result was also observed for SCF predictions. The additional information from the MODIS SCF data was generally less significant in constraining the rainfall-runoff model parameters.

When using the GLUE LOA approach, the model did not provide any behavioral simulation that yields predictions within the assumed observational error bound in over 90% of the time steps. A relaxation was needed in order to partly overcome the limitations of using constant observational error proportionality and not taking an explicit account of the other sources of uncertainty such as from input data errors. A relaxed GLUE LOA approach was introduced that allows a relaxation on the number of time steps required to achieve the LOA. Similar results are obtained using both the residual-based GLUE and the relaxed GLUE LOA approaches. Relaxing the percentage of observations required to be bracketed per simulation period by a particular model realization (pLOA) was found to be more effective than relaxing the observational error bounds. In this study the 5%–95% prediction uncertainty of the residual-based GLUE methodology was used as a reference to define the pLOA in the relaxed GLUE LOA analysis using forcing and observational datasets from a single catchment. More similar case studies should be conducted on catchments with different hydrologic characteristics to assess the scope of this approach under different condition.

Data availability. The underlying hydrologic observations for this analysis were provided by Statkraft AS and are proprietary within their hydrologic forecasting system. However, the data may be made available upon request. Please contact John Burkhart (john.burkhart@statkraft.com) for further information and access to the data.

Author contributions. ATT designed and performed the analysis. JFB and TSV supervised the study. ATT wrote the manuscript. JFB and TSV reviewed the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work was conducted within the Norwegian Research Council's Enhancing Snow Competency of Models and Operators (ESCYMO) project (NFR no. 244024) and in cooperation with the strategic research initiative LATICE (Faculty of Mathematics and Natural Sciences, University of Oslo <https://mn.uio.no/latice>, last access: 1 March 2018). Computational and data storage resources were provided by NOTUR/NORSTORE projects NS9333K and NN9333K. We are grateful for Keith Beven and Chong-Yu Xu for their helpful comments. Furthermore, we thank Sigbjorn Helset and Statkraft AS, in general, for helping us to set up Shyft in a windows environment and for providing us the data. We also thank Kristoffer Aalstad and Sebastian Westermann for providing us a Matlab script for retrieving the composite MODIS SCF data.

Edited by: Alberto Guadagnini

Reviewed by: two anonymous referees

References

- Bavera, D., Michele, C., Pepe, M., and Rampini, A.: Melted snow volume control in the snowmelt runoff model using a snow water equivalent statistically based model, *Hydrol. Process.*, 26, 3405–3415, 2012.
- Berezowski, T. and Batelaan, O.: Skill of remote sensing snow products for distributed runoff prediction, *J. Hydrol.*, 524, 718–732, 2015.
- Beven, K.: Changing ideas in hydrology – the case of physically-based models, *J. Hydrol.*, 105, 157–172, 1989.
- Beven, K. and Binley, A.: The future of distributed models: model calibration and uncertainty prediction, *Hydrol. Process.*, 6, 279–298, 1992.
- Beven, K.: Prophecy, reality and uncertainty in distributed hydrological modelling, *Adv. Water Resour.*, 16, 41–51, 1993.
- Beven, K.: A manifesto for the equifinality thesis, *J. Hydrol.*, 320, 18–36, 2006.
- Beven, K.: Environmental modelling: An uncertain future?, CRC Press, London, 2009.
- Beven, K.: Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication, *Hydrolog. Sci. J.*, 61, 1652–1665, 2016.
- Beven, K. and Smith, P.: Concepts of information content and likelihood in parameter calibration for hydrological simulation models, *J. Hydrol. Eng.*, 20, A4014010, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000991](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000991), 2015.
- Blasone, R.-S., Vrugt, J. A., Madsen, H., Rosbjerg, D., Robinson, B. A., and Zyvoloski, G. A.: Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov Chain Monte Carlo sampling, *Adv. Water Resour.*, 31, 630–648, 2008.
- Blazkova, S. and Beven, K.: Flood frequency estimation by continuous simulation for a catchment treated as ungauged (with uncertainty), *Water Resour. Res.*, 38, 1139, <https://doi.org/10.1029/2001WR000500>, 2002.

- Blazkova, S. and Beven, K.: A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic, *Water Resour. Res.*, 45, W00B16, <https://doi.org/10.1029/2007WR006726>, 2009.
- Boyle, D. P., Gupta, H. V., and Sorooshian, S.: Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods, *Water Resour. Res.*, 36, 3663–3674, 2000.
- Brazier, R. E., Beven, K. J., Freer, J., and Rowan, J. S.: Equifinality and uncertainty in physically based soil erosion models: application of the GLUE methodology to WEPP – the Water Erosion Prediction Project – for sites in the UK and USA, *Earth Surf. Proc. Land.*, 25, 825–845, 2000.
- Burkhart, J. F., Helset, S., Abdella, Y. S., and Lappégard, G.: Operational Research: Evaluating Multimodel Implementations for 24/7 Runtime Environments, Abstract H51F-1541 presented at the Fall Meeting, AGU, San Francisco, California, 11–15 December 2016.
- Choi, H. T. and Beven, K.: Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework, *J. Hydrol.*, 332, 316–336, 2007.
- Clark, M. P., Kavetski, D., and Fenicia, F.: Pursuing the method of multiple working hypotheses for hydrological modeling, *Water Resour. Res.*, 47, W09301, <https://doi.org/10.1029/2010WR009827>, 2011.
- Copernicus land monitoring service: CORINE land cover, available at: <https://land.copernicus.eu/pan-european/corine-land-cover>, last access: 29 August 2016.
- Crawford, N. H. and Linsley, R. K.: Digital simulation in hydrology, Stanford Watershed Model IV, Department of Civil Engineering, Stanford University, California, 1966.
- Dee, D. P., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., and Bauer, P.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. Roy. Meteor. Soc.*, 137, 553–597, 2011.
- Efstratiadis, A. and Koutsoyiannis, D.: One decade of multi-objective calibration approaches in hydrological modelling: a review, *Hydrolog. Sci. J.*, 55, 58–78, 2010.
- Finger, D., Vis, M., Huss, M., and Seibert, J.: The value of multiple data set calibration versus model complexity for improving the performance of hydrological models in mountain catchments, *Water Resour. Res.*, 51, 1939–1958, 2015.
- Hall, D. K., Riggs, G. A., Salomonson, V. V., Barton, J., Casey, K., Chien, J., DiGirolamo, N., Klein, A., Powell, H., and Tait, A.: Algorithm theoretical basis document (ATBD) for the MODIS snow and sea ice-mapping algorithms, *Nasa Gsfc*, 45, 2001.
- Hall, K., George, R., Vincent, S., and Grid, V.: Updated daily MODIS/Terra Snow Cover Daily L3 Global 500 m Grid V005, April 2011 to August 2014, in: National Snow and Ice Data Center, Digital media, Boulder, Colorado USA, 2006.
- Hanzer, F., Helfricht, K., Marke, T., and Strasser, U.: Multilevel spatiotemporal validation of snow/ice mass balance and runoff modeling in glacierized catchments, *The Cryosphere*, 10, 1859–1881, <https://doi.org/10.5194/tc-10-1859-2016>, 2016.
- Hassan, A. E., Bekhit, H. M., and Chapman, J. B.: Uncertainty assessment of a stochastic groundwater flow model using GLUE analysis, *J. Hydrol.*, 362, 89–109, 2008.
- He, M., Hogue, T. S., Franz, K. J., Margulis, S. A., and Vrugt, J. A.: Characterizing parameter sensitivity and uncertainty for a snow model across hydroclimatic regimes, *Adv. Water Resour.*, 34, 114–127, 2011.
- Hegdahl, T. J., Tallaksen, L. M., Engeland, K., Burkhart, J. F., and Xu, C. Y.: Discharge sensitivity to snowmelt parameterization: a case study for Upper Beas basin in Himachal Pradesh, India, *Hydrol. Res.*, 47, 683–700, 2016.
- Hornberger, G. M. and Spear, R. C.: Approach to the preliminary analysis of environmental systems, *J. Environ. Mgmt.*, 12, 7–18, 1981.
- Jin, X., Xu, C. Y., Zhang, Q., and Singh, V. P.: Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model, *J. Hydrol.*, 383, 147–155, 2010.
- Kirchner, J. W.: Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, *Water Resour. Res.*, 45, W02429, <https://doi.org/10.1029/2008WR006912>, 2009.
- Kolberg, S. A. and Gottschalk, L.: Updating of snow depletion curve with remote sensing data, *Hydrol. Process.*, 20, 2363–2380, 2006.
- Krause, P., Boyle, D. P., and Bäse, F.: Comparison of different efficiency criteria for hydrological model assessment, *Adv. Geosci.*, 5, 89–97, <https://doi.org/10.5194/adgeo-5-89-2005>, 2005.
- Lambert, A.: Catchment models based on ISO-functions, *J. Instn. Water Engrs.*, 26, 413–422, 1972.
- Lee, S., Klein, A. G., and Over, T. M.: A comparison of MODIS and NOHRSC snow-cover products for simulating streamflow using the Snowmelt Runoff Model, *Hydrol. Process.*, 19, 2951–2972, 2005.
- Liston, G. E.: Interrelationships among snow distribution, snowmelt, and snow cover depletion: Implications for atmospheric, hydrologic, and ecologic modeling, *J. Appl. Meteorol.*, 38, 1474–1487, 1999.
- Liston, G. E.: Representing subgrid snow cover heterogeneities in regional and global models, *J. Climate*, 17, 1381–1397, 2004.
- Liu, J. and Han, D.: Indices for calibration data selection of the rainfall-runoff model, *Water Resour. Res.*, 46, W04512, <https://doi.org/10.1029/2009WR008668>, 2010.
- Liu, Y., Freer, J., Beven, K., and Matgen, P.: Towards a limits of acceptability approach to the calibration of hydrological models: Extending observation error, *J. Hydrol.*, 367, 93–103, 2009.
- Mantovan, P. and Todini, E.: Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology, *J. Hydrol.*, 330, 368–381, 2006.
- Matt, F. N., Burkhart, J. F., and Pietikäinen, J.-P.: Modelling hydrologic impacts of light absorbing aerosol deposition on snow at the catchment scale, *Hydrol. Earth Syst. Sci.*, 22, 179–201, <https://doi.org/10.5194/hess-22-179-2018>, 2018.
- Mirzaei, M., Huang, Y. F., El-Shafie, A., and Shatirah, A.: Application of the generalized likelihood uncertainty estimation (GLUE) approach for assessing uncertainty in hydrological models: a review, *Stoch. Env. Res. Risk. A.*, 29, 1265–1273, 2015.
- Montanari, A., Shoemaker, C. A., and van de Giesen, N.: Introduction to special section on Uncertainty Assessment




- in Surface and Subsurface Hydrology: An overview of issues and challenges, *Water Resour. Res.*, 45, W00B00, <https://doi.org/10.1029/2009WR008471>, 2009.
- Morris, M. D.: Factorial sampling plans for preliminary computational experiments, *Technometrics*, 33, 161–174, 1991.
- Nearing, G. S., Tian, Y., Gupta, H. V., Clark, M. P., Harrison, K. W., and Weijs, S. V.: A philosophical basis for hydrological uncertainty, *Hydrolog. Sci. J.*, 61, 1666–1678, 2016.
- Norwegian mapping authority: Kartverket, available at: <https://www.kartverket.no/>, last access: 1 September 2016.
- Pappenberger, F., Beven, K. J., Ratto, M., and Matgen, P.: Multi-method global sensitivity analysis of flood inundation models, *Adv. Water Resour.*, 31, 1–14, 2008.
- Parajka, J. and Blöschl, G.: The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models, *J. Hydrol.*, 358, 240–258, 2008.
- Parajka, J., Holko, L., Kostka, Z., and Blöschl, G.: MODIS snow cover mapping accuracy in a small mountain catchment – comparison between open and forest sites, *Hydrol. Earth Syst. Sci.*, 16, 2365–2377, <https://doi.org/10.5194/hess-16-2365-2012>, 2012.
- Pianosi, F., Sarrazin, F., and Wagener, T.: A Matlab toolbox for global sensitivity analysis, *Environ. Modell. Softw.*, 70, 80–85, 2015.
- Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., and Wagener, T.: Sensitivity analysis of environmental models: A systematic review with practical workflow, *Environ. Modell. Softw.*, 79, 214–232, 2016.
- Powell, M. J.: The BOBYQA algorithm for bound constrained optimization without derivatives, Cambridge NA Report NA2009/06, University of Cambridge, Cambridge, 26–46, 2009.
- Priestley, C. and Taylor, R.: On the assessment of surface heat flux and evaporation using large-scale parameters, *Mon. Weather Rev.*, 100, 81–92, 1972.
- Pu, Z., Xu, L., and Salomonson, V. V.: MODIS/Terra observed seasonal variations of snow cover over the Tibetan Plateau, *Geophys. Res. Lett.*, 34, L06706, <https://doi.org/10.1029/2007GL029262>, 2007.
- Refsgaard, J. C., van der Sluijs, J. P., Højberg, A. L., and Vanrolleghem, P. A.: Uncertainty in the environmental modelling process—a framework and guidance, *Environ. Modell. Softw.*, 22, 1543–1556, 2007.
- Reichert, P. and Omlin, M.: On the usefulness of overparameterized ecological models, *Ecol. Model.*, 95, 289–299, 1997.
- Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W.: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resour. Res.*, 46, W05521, <https://doi.org/10.1029/2009WR008328>, 2010.
- Saltelli, A., Ratto, M., Tarantola, S., Campolongo, F., and Commission, E.: Sensitivity analysis practices: Strategies for model-based inference, *Reliab. Eng. Syst. Safe.*, 91, 1109–1125, 2006.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S.: *Global sensitivity analysis: the primer*, John Wiley & Sons, Chichester, 2008.
- Samanta, S. and Mackay, D. S.: Flexible automated parameterization of hydrologic models using fuzzy logic, *Water Resour. Res.*, 39, 1009, <https://doi.org/10.1029/2002WR001349>, 2003.
- Savenije, H. H.: Equifinality, a blessing in disguise?, *Hydrol. Process.*, 15, 2835–2838, 2001.
- Schaefli, B.: Snow hydrology signatures for model identification within a limits-of-acceptability approach, *Hydrol. Process.*, 30, 4019–4035, 2016.
- Schoups, G. and Vrugt, J. A.: A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors, *Water Resour. Res.*, 46, W10531, <https://doi.org/10.1029/2009WR008933>, 2010.
- Seibert, J. and Beven, K. J.: Gauging the ungauged basin: how many discharge measurements are needed?, *Hydrol. Earth Syst. Sci.*, 13, 883–892, <https://doi.org/10.5194/hess-13-883-2009>, 2009.
- Shen, Z. Y., Chen, L., and Chen, T.: Analysis of parameter uncertainty in hydrological and sediment modeling using GLUE method: a case study of SWAT model applied to Three Gorges Reservoir Region, China, *Hydrol. Earth Syst. Sci.*, 16, 121–132, <https://doi.org/10.5194/hess-16-121-2012>, 2012.
- Skaugen, T. and Weltzien, I. H.: A model for the spatial distribution of snow water equivalent parameterized from the spatial variability of precipitation, *The Cryosphere*, 10, 1947–1963, <https://doi.org/10.5194/tc-10-1947-2016>, 2016.
- Statkraft: Statkraft information page, available at: <https://www.statkraft.com/>, last access: 20 June 2018.
- Stedinger, J. R., Vogel, R. M., Lee, S. U., and Batchelder, R.: Appraisal of the generalized likelihood uncertainty estimation (GLUE) method, *Water Resour. Res.*, 44, W00B06, <https://doi.org/10.1029/2008WR006822>, 2008.
- Sun, W., Wang, Y., Wang, G., Cui, X., Yu, J., Zuo, D., and Xu, Z.: Physically based distributed hydrological model calibration based on a short period of streamflow data: case studies in four Chinese basins, *Hydrol. Earth Syst. Sci.*, 21, 251–265, <https://doi.org/10.5194/hess-21-251-2017>, 2017.
- Tripp, D. R. and Niemann, J. D.: Evaluating the parameter identifiability and structural validity of a probability-distributed model for soil moisture, *J. Hydrol.*, 353, 93–108, 2008.
- Udnæs, H. C., Alfnes, E., and Andreassen, L. M.: Improving runoff modelling using satellite-derived snow covered area, *Hydrol. Res.*, 38, 21–32, 2007.
- Vrugt, J. A., Ter Braak, C. J., Gupta, H. V., and Robinson, B. A.: Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling, *Stoch. Env. Res. Risk. A.*, 23, 1011–1026, 2009.
- Wagener, T., McIntyre, N., Lees, M., Wheeler, H., and Gupta, H.: Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis, *Hydrol. Process.*, 17, 455–476, 2003.
- Xiong, L. and O'Connor, K. M.: An empirical method to improve the prediction limits of the GLUE methodology in rainfall-runoff modeling, *J. Hydrol.*, 349, 115–124, 2008.
- Xiong, L., Wan, M., Wei, X., and O'Connor, K. M.: Indices for assessing the prediction bounds of hydrological models and application by generalised likelihood uncertainty estimation, *Hydrolog. Sci. J.*, 54, 852–871, 2009.

Paper III

Teweldebrhan, A., Burkhart, J., Schuler, T., and Xu, C.-Y.: Improving the Informational Value of MODIS Fractional Snow Cover Area Using Fuzzy Logic Based Ensemble Smoother Data Assimilation Frameworks, *Remote Sensing*, 11, 28, 2019.

Article

Improving the Informational Value of MODIS Fractional Snow Cover Area Using Fuzzy Logic Based Ensemble Smoother Data Assimilation Frameworks

Aynom T. Teweldebrhan ^{1,*}, John F. Burkhart ^{1,2}, Thomas V. Schuler ¹ and Chong-Yu Xu ¹

¹ Department of Geosciences, University of Oslo, NO-0316 Oslo, Norway; john.burkhart@geo.uio.no (J.F.B.); t.v.schuler@geo.uio.no (T.V.S.); c.y.xu@geo.uio.no (C.-Y.X.)

² Statkraft, Lilleaker, NO-0216 Oslo, Norway

* Correspondence: aynomtt@geo.uio.no

Received: 8 November 2018; Accepted: 21 December 2018; Published: 25 December 2018



Abstract: Remote sensing fractional snow cover area (fSCA) has been increasingly used to get an improved estimate of the spatiotemporal distribution of snow water equivalent (SWE) through reanalysis using different data assimilation (DA) schemes. Although the effective assimilation period of fSCA is well recognized in previous studies, little attention has been given to explicitly account for the relative significance of measurements in constraining model parameters and states. Timing of the more informative period varies both spatially and temporally in response to various climatic and physiographic factors. Here we use an automatic detection approach to locate the critical points in the time axis where the mean snow cover changes and where the melt-out period starts. The assimilation period was partitioned into three timing windows based on these critical points. A fuzzy coefficient was introduced in two ensemble-based DA schemes to take into account for the variability in informational value of fSCA observations with time. One of the DA schemes used in this study was the particle batch smoother (Pbs). The main challenge in Pbs and other Bayesian-based DA schemes is, that most of the weights are carried by few ensemble members. Thus, we also considered an alternative DA scheme based on the limits of acceptability concept (LoA) and certain hydrologic signatures and it has yielded an encouraging result. An improved estimate of SWE was also obtained in most of the analysis years as a result of introducing the fuzzy coefficients in both DA schemes. The most significant improvement was obtained in the correlation coefficient between the predicted and observed SWE values (site-averaged); with an increase by 8% and 16% after introducing the fuzzy coefficient in Pbs and LoA, respectively.

Keywords: informational value; MODIS snow cover; fuzzy logic; data assimilation (DA)

1. Introduction

Snow has important relevance for society spanning from environmental and economical to recreational and aesthetic values. Environmentally snow influences the climate and ecology of an area through its effect on solar energy absorption and by constraining land cover characteristics [1,2]. Snowmelt results in significant decrease of surface albedo and releases the water stored in seasonal snowpack, providing water for humans and the environment [3,4]. Water managers in hydropower and water supply sectors consider snow storage in mountain areas as a natural reservoir [5]. During the past decades, hydrological models have been used to assess the amount and spatiotemporal distribution of this important resource [6,7]. However, the modeling process in general and model calibration in particular is prone to various sources of uncertainty [8]. Further, the objective functions commonly used for parameter inference using inverse modeling are based on aggregate measures

of model performance and usually yield to poorly constrained model parameters [9]. A combined approach of model calibration followed by data assimilation is one of the strategies commonly followed to deal with model uncertainty [10,11]. Data assimilation (DA) is the means to obtain an estimate of the true state through use of independent observations and the prior knowledge (model state) with appropriate uncertainty modeling [9,12].

Ensemble-based filtering and smoothing techniques are commonly used in solving DA problems [9]. The filtering techniques involve updating of state variables and/or parameter values at each observation time with subsequent production of error statistics that can be cycled to future analysis times; while in smoothing all observations are assimilated in a single step [4,13]. Some of the filtering- (sequential) based DA schemes commonly used in hydrological data assimilation include the Ensemble Kalman Filter [14] and the Particle filter [15], while some examples from the smoother-based DA schemes include the ensemble batch smoother [16] as well as the particle batch smoother (Pbs) [17]. The smoother-based DA schemes have the advantage of low computational cost coupled with the flexibility in running independent of the forward model [4,18,19].

In recent years, remote sensing data in general and optical fractional snow cover area (fSCA) in particular has been increasingly used in hydrologic modeling for constraining model parameters through inverse modeling and various DA schemes [20]. This can be attributed to the readily available observations of this data for large areas and at a relatively high temporal resolution [21,22]. In DA, fSCA has been used from a simple rule-based (direct-insertion) approach [23] to advanced filtering and smoothing techniques [24,25]. The collection of fSCA measurements and their temporal evolution during the ablation period reflect the peak snow water equivalent (SWE) better than a single fSCA measurement does. Hence, the batch assimilation techniques are preferable over filtering techniques when the assimilated observation is fSCA in a snow depletion curve model [17]. The Pbs and particle filter DA schemes have been gaining interest as fSCA assimilation schemes in snow models due to their distribution free likelihood and their capability to estimate the SWE state vector directly at a relatively low computational cost [17,26]. However, these schemes assume crisp and equal informational value for all observations of the assimilated data. Further, in these schemes, most of the weights are assigned to one or very few ensemble members and this may lead to degeneration of the statistical information in the ensembles [27,28]. An alternative approach that ensures fair distribution of weights among ensemble members is the limits of acceptability (LoA) [29]. This approach provides flexibility in choosing different shapes for representing the error within specified error bounds. It further allows the incorporation of different hydrological signatures in the evaluation criteria. As such, it is more process-based as compared to the simple residual-based objective functions since it transforms the observed and simulated data into hydrologically relevant signatures [30].

In this study, the informational value of fSCA measurements is considered to be fuzzy, carrying a considerable amount of uncertainty. In previous studies different approaches have been used to increase the amount of information that can be extracted from time series data during parameter inference using inverse modeling [30]. Some of these measures include: breaking up the time series data into different periods, for example, the different components of a hydrograph [31] as well as transforming the time series data to emphasise on some aspects of the system response [32]. In several snow related DA studies all measurements are assumed to carry the same amount of information in constraining the model states and parameters. However, the amount of information that can be extracted from remote sensing data varies depending on several factors including the presence of certain physical limitations such as clouds and lack of sufficient light [33]. For example, high latitudes are characterized by polar darkness and frequent cloud cover and errors emanating from such sources can reduce the informational value of MODIS fSCA product for snow assimilation [34,35].

The informational value of MODIS fSCA was divided into three timing windows; and an automatic change point (CP) detection scheme was employed to identify the spatially and temporally variable location of the CP in the time-axis. CP detection deals with the estimation of the point at which the statistical properties of a sequence of observations changes [36]. In hydrology, CP detection

methods have been applied in various studies including the assessment of non-stationarity in time series data [37,38] as well as in the assessment of flow-solute relationship in water quality studies [39]. Nevertheless, their application has been limited when it comes to the analysis of remote sensing snow cover data. Although, snowmelt at high latitudes is a rapid process, the start date of an ablation period may vary by several weeks from year to year in response to such factors as the amount of accumulated SWE before the onset of snowmelt and the availability of melt energy [40]. Spatially, the onset of snow melt varies from one grid-cell to another depending on variability in physiographic characteristics of the area in addition to the above mentioned climatic factors. During the accumulation period, spatial variability of snow results from snow-canopy interaction in forested areas, snow redistribution by wind, and orographic effects [41]. During the ablation period, snow melts from preferential locations, yielding heterogeneous patterns [42].

The main goal of this study is to get an improved estimate of SWE during the maximum accumulation period using fuzzy logic-based ensemble fSCA assimilation schemes and by incorporating uncertainties in selected model forcing and parameters. The first objective is to assess whether the assumption of variable informational value of fSCA observations depending on their location in different timing windows can reduce the uncertainty in SWE reanalysis. We introduce a novel approach that accommodates this concept by incorporating a fuzzy coefficient in the ensemble-based SWE reanalysis schemes. The second objective is to evaluate the viability of statistical change point detection methods to locate the critical points that constitute the timing windows. The third objective is to compare effect of the assumed variability in informational value of fSCA observations on relative performance of the Pbs and the LoA-based DA schemes. As of our best knowledge, the LoA approach was not used before for DA purposes; and this study will assess the feasibility of this approach as an ensemble smoother DA scheme.

This paper is organized as follows. Section 2 presents the snow model followed by description of the study area as well as the static and time-series data used in this study. This section also presents setup of the different DA schemes and methodologies followed to detect the critical dates in the ablation period. Results from the DA schemes in comparison to the observed values as well as sensitivity of the evaluation metrics to location of the change points in the ablation period are presented in Section 3. Finally, certain points from the methodology and results sections are discussed in Section 4 in light of previous studies; and relevant conclusions are drawn in Section 5.

2. Methods and Materials

2.1. The Hydrological Model

The forward model used in this study, PT_GS_K, is a conceptual hydrological model (method stack) within the Statkraft Hydrological Forecasting Toolbox (Shyft, <https://github.com/statkraft/shyft>). This model and the modeling framework were described in previous studies [43,44] and this section presents main features of the model with focus on its snow method. PT_GS_K encompasses the Priestley–Taylor (PT) method [45] for estimating potential evaporation; a quasi-physical-based snow method (GS) for snow melt, sub-grid snow distribution and mass balance calculations; and a simple storage-discharge function (K) [46,47] for catchment response calculation.

PT_GS_K operates on a single snow layer and it can be forced by hourly or daily meteorological inputs of precipitation, temperature, radiation, relative humidity, and wind speed. The Model uses a Bayesian Kriging approach to distribute the point temperature data over the domain, while for the other forcing variables it uses an inverse distance weighting approach. The model generates streamflow, fSCA, and SWE as output variables.

The potential evaporation calculation in the PT method requires net radiation, the slope of saturated vapor pressure, the Priestley–Taylor parameter, the psychrometric constant, and the latent heat of vaporization [48]. Actual evapotranspiration is assumed to take place only from snow free areas and it is estimated as a function of potential evapotranspiration and a scaling factor.

The energy balance calculation in GS method follows a similar approach as used by DeWalle and Rango [49] (Equation (1)). The precipitation falling in each grid-cell is classified as solid or liquid precipitation depending on a threshold temperature (t_x) and on the local temperature values. The available net snow melt energy (NE) is the sum effect of different energy sources and sinks in the system. These include: incoming short wave radiation (SW), incoming (ILR) and outgoing (OLR) long wave radiation, the subsurface energy flux (G) as well as the turbulent sensible (SH) and latent (LH) energy fluxes emanating from rainfall, solar radiation, wind turbulence and other sources.

$$NE = SW * (1 - albedo) + ILR + OLR + SH + LH + G \quad (1)$$

Among other factors, the energy contribution from short wave radiation depends on snow albedo. For a given time step (t), the snow albedo of each grid cell depends on the minimum (α_{min}) and maximum (α_{max}) albedo values as well as on air temperature (T_a) (Equation (2)). In this method the decay rates of albedo due to snow aging as a function of temperature, i.e., the fast (α_{fdr}) and slow (α_{sdr}) albedo decay rates corresponding to temperature conditions above and below 0 °C respectively, are parameterized [50].

$$\alpha_t = \begin{cases} \alpha_{min} + (\alpha_{t-1} - \alpha_{min}) * \left(\frac{1}{2^{1/\alpha_{fdr}}}\right), T_a > 0 \text{ } ^\circ\text{C} \\ \alpha_{t-1} + (\alpha_{max} - \alpha_{min}) * \left(\frac{1}{2^{(\alpha_{sdr})}}\right), T_a \leq 0 \text{ } ^\circ\text{C} \end{cases} \quad (2)$$

The incoming and outgoing longwave radiations are estimated based on the Stephan–Boltzmann law. Turbulent heat contribution is the sum of latent and sensible heat. Wind turbulence is linearly related to wind speed using a wind constant and wind scale from the intercept and slope of the linear function, respectively [50]. The subsurface energy flux is a function of snow surface layer and the snow surface temperature.

The sub-grid snow distribution in each grid cell is described by a three parameter Gamma probability distribution snow depletion curve (SDC) [51,52]. This SDC is based on the assumption that certain proportion of a grid cell area (y_0) remains snow-free throughout the winter season such as in steep slopes and ridges due to wind erosion or avalanches [53]. As such, the two parameter Gamma distribution function, characterized by the average amount of snow m (mm) and the sub-grid snow coefficient of variation (CV_s) at the onset of the melt season (Equation (3)), is applied only to the remaining portion of a grid cell area to estimate the fraction of the initially snow covered area where snow has disappeared (y'_t). Following this formulation, the bare ground fraction at each time step (y_t) is estimated using Equation (4).

$$y'_t = \int_0^{\lambda(t)} f(x; m, CV_s) dx = \gamma\left(CV_s^{-2}, \frac{\lambda(t)}{m \cdot CV_s^2}\right) \quad (3)$$

$$y_t = y_0 + (1 - y_0) * y'_t \quad (4)$$

where, f denotes the Gamma probability density function and γ is the incomplete Gamma function with shape CV_s^{-2} and scale $m \cdot CV_s^2$. The variables x and $\lambda(t)$ respectively refer to point snow storage and the accumulated melt depth (mm) at time t since the onset of the melt season.

Although, the catchment response function (K) was used for estimating the regional model parameter values through calibration against observed streamflow data, this method was not directly employed in this study.

2.2. Study Area and Data

The study sites are located in Nea-catchment, Sør-Trøndelag County, Norway (Figure 1). The Nea-catchment covers a total area of 703 km² and its geographical location ranges from 11.67390° to 12.46273° E and from 62.77916° to 63.20405° N. Altitude of the study sites ranges from around

700 masl at site-1 to above 1200 masl at site-7. The fact that the model was previously setup for the study area plus the availability of long term SWE data, i.e., for nine years and in nine different sites, make the Nea-catchment an appealing study area to apply the DA schemes. The observed SWE data cover the hydrological years 2008–2016; and the datasets in these years were used in validating the DA schemes. The highest and lowest average catchment peak SWE values estimated from the nine sites during these hydrological years were 717 and 331 mm respectively in years 2012 and 2014. Mean annual precipitation for the nine hydrological years was 1279 mm. The highest and lowest average daily temperature values for this period were 20 and -23 °C, respectively.

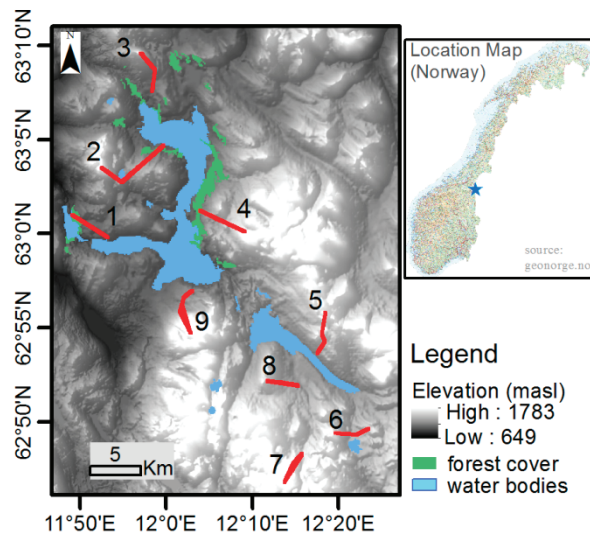


Figure 1. Physiography and location map of the study area in Norway. The red lines numbered 1–9 represent snow measurement sites.

The SWE reanalysis was conducted on 33 grid cells located at nine sites (Table 1). The location of these grid-cells was constrained by the availability of snow course measurements in the catchment. Daily time series data of most meteorological forcing variables, i.e., temperature, precipitation, radiation, and wind speed were obtained from Statkraft [54] as point measurement, while daily gridded relative humidity data were retrieved from ERA-interim [55]. Different physiographic and observational datasets were also used to generate the simulation results and in subsequent reanalysis using the DA schemes as well as during evaluation of the results. The model was setup over each grid cells of 1 km²; requiring average elevation, grid cell total area, as well as the areal fraction of forest cover. Data for these physiographic variables were retrieved from two sources: the land cover data from Copernicus land monitoring service [56] and the 10m digital elevation model (10m DEM) from the Norwegian mapping authority (Kartverket.no). The land cover data show that, the catchment is mainly dominated by moors, heathland and some sparse vegetation; and only limited part of the catchment is forest covered (3%). The grid cells considered in the DA study are located on the open areas of the catchment. The median SWE data of each grid cell were derived from radar measurements provided by Statkraft [54]. The data were collected once a year in the month of April, where accumulated snow storage approximately attains its peak magnitude. The radar measurements roughly follow the same course each year in the nine representative sites of the catchment. Daily fraction of snow cover area (fSCA) was retrieved from NASA MODIS snow cover products [57]. Frequent cloud cover is one of the major challenges when using MODIS and other optical remote sensing data in Norway. A composite dataset was thus formed using data retrieved from the Aqua and Terra satellites, MYD10A1 and MOD10A1 products respectively in order to minimize the effect of obstructions and misclassification errors emanating from clouds and other sources. The days with available snow cover data were used without filtering based on further criteria such as the remaining clouds. A direct assessment of the

bias associated with this product was not done in this study as independent snow cover observations were not available for the study domain. The estimated annual average bias of these MODIS fSCA products for Northern Hemisphere is approximately 8 % in the absence of cloud [58] and in forest dominated areas it may reach up to 15% [59]. This MODIS product was reported to have a relatively lower accuracy as compared to other more recent products, e.g., the MODSCAG algorithm [1] mainly due to its limited use of the available information in a given scene [60].

Table 1. Summary of site-averaged annual peak observed snow water equivalent (SWE) (m w.e.) as well as average site elevation (masl) and number of model grid-cells in each site of the study area.

Site	SWE (m w.e.)			No. Of Grid-Cells	Avg. Elev. (masl)
	Min.	Max.	Average		
1	0.269	0.558	0.400	4	746
2	0.411	0.931	0.606	9	907
3	0.159	0.764	0.487	2	873
4	0.299	0.659	0.454	5	958
5	0.398	0.818	0.603	3	1064
6	0.34	0.668	0.430	2	988
7	0.256	0.391	0.340	3	1251
8	0.258	0.383	0.316	1	976
9	0.152	0.549	0.274	4	993

2.3. The Data Assimilation Strategy

This section presents the perturbed forcing and model parameters followed by description of the two main DA schemes employed in this study, i.e., Pbs and LoA as well as other two similar versions of these schemes that take into account for variability in informational value of fSCA. The ensemble-based DA schemes involve two steps, namely prediction and updating. The prediction step provides the prior estimates of SWE using the dynamic (meteorological) and static (physiographic) forcings. In the updating step, the DA schemes are applied to constrain the prior SWE estimates using the fSCA observations and thereby yielding a posterior SWE estimates. Each grid cell is updated independently.

2.3.1. Perturbed Forcing and Model Parameters

In this study, one meteorological forcing and two model parameters are randomly perturbed to produce an ensemble of 100 model realizations. Precipitation was perturbed since meteorological forcing, in general and precipitation in particular, are the major sources of uncertainty in simulating snow storage and depletion processes [61,62]. Considering uncertainty in precipitation provides the means to control the amount of accumulated SWE in the absence of a viable spatial snow redistribution mechanism [17]. The perturbed model parameters are sub-grid snow coefficient of variation (CV_s) and initial bare ground fraction (y_0). Snow coefficient of variation affects the SDC through its impact on the rate of snow depletion. In snow models with the sub-grid snow distribution component parameterized using statistical probability distribution function, the rate of snow depletion is inversely related to CV_s [41]. Thus, when the sub-grid snow cover is more variable, the ground gets exposed earlier than when it is uniformly distributed [20]. The initial bare ground fraction plays an important role in the SDC especially during the onset of snow melt when the fraction of initial snow covered area where snow has disappeared (y'_t , Equation(3)) is at its lowest [52].

Both log-normal and logit-normal probability distributions are used to perturb the model forcing and parameters. A log-normal distribution was assumed for precipitation, while for CV_s and y_0 a logit-normal distribution was assumed. A similar approach to previous SWE reanalysis studies [17,28] was followed in perturbing precipitation and the model parameters. Instead of directly perturbing the precipitation input at each time step (P_t), a perturbation parameter (b_j) was randomly generated from a log-normal distribution and the generated, b_j values are used as multiplicative biases for their respective ensemble member, j (Equation (5)). Since the correction for wind induced precipitation

under-catch was previously done, it was assumed that the average ensemble, b , value is unbiased with a value of unity. The snow coefficient of variation (CV_s) estimated during calibration of the model against observed streamflow was adopted as the prior mean value. The mean value of y_0 was set to 0.04 based on previous studies in Norway [52]; and its variability was kept to a narrow range, since this parameter can easily compensate for differences between the observed and simulated fSCA values [52].

$$P_{j,t} = b_j P_t \quad (5)$$

A standard deviation (σ_{fSCA}) of 15% was used to represent the measurement error in fSCA. This value is set based on previous studies focused on MODIS measurement errors [59]. Table 2 presents the perturbation parameters and their assumed prior values.

Table 2. Main input variables used in the SWE reanalysis using the different data assimilation (DA) schemes. fSCA: fractional snow cover area.

Variable	Value	Description
b_{mean}	1.0	mean precipitation bias
b_{CV}	1.0	coefficient of variation of precipitation bias
$CV_{s,\text{mean}}$	0.68	mean snow coefficient of variation (calibrated value)
$CV_{s,\text{min}}$	0.0	minimum snow coefficient of variation
$CV_{s,\text{max}}$	1.5	maximum snow coefficient of variation
$CV_{s,\text{std}}$	0.55	standard deviation of the snow coefficient of variation
$y_{0,\text{mean}}$	0.04	mean initial bare ground fraction
$y_{0,\text{min}}$	0.0	minimum initial bare ground fraction
$y_{0,\text{max}}$	0.1	maximum initial bare ground fraction
$y_{0,\text{std}}$	0.05	standard deviation of initial bare ground fraction
σ_{fSCA}	15%	measurement error standard deviation of fSCA
N_e	100	ensemble size
N_o	-	number of fSCA observations

2.3.2. fSCA Assimilation using Particle Batch Smoother

Previous studies indicate that the batch smoother is better suited than the filtering approach when dealing with SWE reanalysis through assimilation of fSCA observations [17,63]. A particle batch smoother (Pbs) uses an ensemble of independent randomly drawn Monte Carlo samples (particles) and estimates the posterior weight of each particle. The procedure involves generating ensemble of model realizations by running the model over the full seasonal cycle followed by the assimilation of all fSCA measurements at once. The Bayes theorem forms the basis for estimating the updated relative importance (weight, w_j) of each ensemble member (Equation (6)).

$$w_j = \frac{p(y|\hat{X}_j)p(\hat{X}_j)}{\sum_{j=1}^{N_e} p(y|\hat{X}_j)p(\hat{X}_j)} \quad (6)$$

where $p(y|\hat{X}_j)$ refers to the likelihood of the observations (y) given the state value of the j th ensemble member (\hat{X}_j); and $p(\hat{X}_j)$ denotes prior values.

In Pbs, all ensemble members are implicitly assigned equal prior weights of $1/N_e$ [17]. When using the Gaussian likelihood, this yields [17,28]:

$$w_j = \frac{\exp\left[-0.5(y - \hat{Y}_j)^T \mathbf{R}^{-1}(y - \hat{Y}_j)\right]}{\sum_{j=1}^{N_e} \exp\left[-0.5(y - \hat{Y}_j)^T \mathbf{R}^{-1}(y - \hat{Y}_j)\right]} \quad (7)$$

where y and \hat{Y}_j respectively refer to $N_o \times 1$ vector of perturbed fSCA observations and predicted fSCA for the j th particle in the $N_o \times N_e$ matrix ($\hat{\mathbf{Y}}$). \mathbf{R} denotes $N_o \times N_o$ diagonal observation error covariance matrix.

Pbs assumes that both the prior and posterior states of the particles remain the same [17]. It updates the particle weights in such a way that particles with their predictions closer to the observations get higher weight than those with farther from the observations. The median and prediction quantiles are estimated from the cumulative distribution of the sorted SWE values and their associated weights.

2.3.3. fSCA Assimilation Using the LoA

We use two hydrologic signatures to constrain the perturbation parameters by the LoA scheme through weighting the relative importance of the particles with a fuzzy rule. In a fuzzy rule, the information contained in a fuzzy set is described by its membership function [64]. The first signature is based on the ability of each ensemble member to reproduce the observations within pre-defined error bounds of fSCA observations. The same value of measurement error as used in Pbs was also adopted here, i.e., 15% (Table 2). The deviation between the observed and simulated values was converted into a normalized criterion using a fuzzy rule-based scoring function. A triangular membership function was assumed with its support representing the uncertainty in MODIS fSCA observations and the pointed core representing a perfect match between the observed and predicted fSCA (Equation (8)). The weight of each ensemble member with respect to this signature (w_{e_j}) is calculated as the membership grade of the prediction error, summed over all observations (Equation (9)).

$$\mu_{fSCA}(e) = \begin{cases} 0, e \leq l \\ \frac{e-l}{m-l}, l < e \leq m \\ \frac{u-e}{u-m}, m < e < u \\ 0, e \geq u \end{cases} \quad (8)$$

$$w_{e_j} = \sum_{i=1}^{N_o} \mu_{fSCA}(e) \quad (9)$$

where $\mu_{fSCA}(e)$ is the membership grade of each prediction error (e) corresponding to the observed fSCA value i ; m is the point in the support with perfect match between the observed and predicted fSCA values. The variables l and u respectively refer to the lower and upper bounds of the fSCA observation error.

The second hydrologic signature is based on the degree of persistency of an ensemble member in reproducing the observations. This is estimated by the percentage of observations (p) where the predicted values fall within the MODIS fSCA error bounds. Ensemble members whose prediction failed to reproduce at least 50% of the observations within the error bounds are considered as non-persistent and assigned a membership grade of zero. The degree of membership linearly increases for those that satisfy the condition from 50% to 95% of the observations; and attains its maximum value after 95%. Thus, ensemble members that are able to reproduce 95% to 100% of the observations within the error bounds are considered as the most persistent and accordingly assigned a membership grade of 1.0. The weight of each ensemble member with respect to its persistence in reproducing the observations (w_{p_j}) is equal to the degree of membership ($\mu_N(p)$, Equation 10). The combined weight (w_j) resulting from the two membership functions is determined by calculating the product of w_{e_j} and w_{p_j} followed by normalizing the result in such a way that all particle weights would sum up to 1.

$$\mu_N(p) = \begin{cases} 0, p < l \\ \frac{p-l}{m-l}, l \leq p \leq m \\ 1, p \geq m \end{cases} \quad (10)$$

where $\mu_N(p)$ is the membership grade of each ensemble member with respect to its persistency (p) in reproducing the observations; l represents the minimum bound of the support ($p = 50\%$) and m refers to the point in the support where the membership grade starts to attain its maximum value ($p = 95\%$).

2.3.4. DA Schemes with an Account for Fuzzy Informational Value of fSCA

Here, the ablation period is divided into three timing windows with respect to the assumed relative informational value of fSCA observations (Figure 2). This assumption is based on the concept that information is gained from an observation only if there is uncertainty about it; and an event that occurs with a high probability conveys less information and vice versa [65]. The plot shows temporal distribution of fSCA for a sample grid-cell in the study area. The first timing window (W-1) is characterized by some snow melt with a subsequent decrease in SWE while the fSCA remains close to its maximum value. The fSCA observations from this window are characterized by low signal-to-noise ratio. In the second window (W-2), fSCA starts to drop at a faster rate and this window is characterized by high signal-to-noise ratio with a strong decreasing trend with time. The third window (W-3) corresponds to the complete melt-out period; and with the exception of few fSCA observations from sporadic snow events, fSCA is expected to remain null throughout this period. The fSCA observations in W-1 and W-3 are thus relatively certain, conveying less information as compared to observations in W-2.

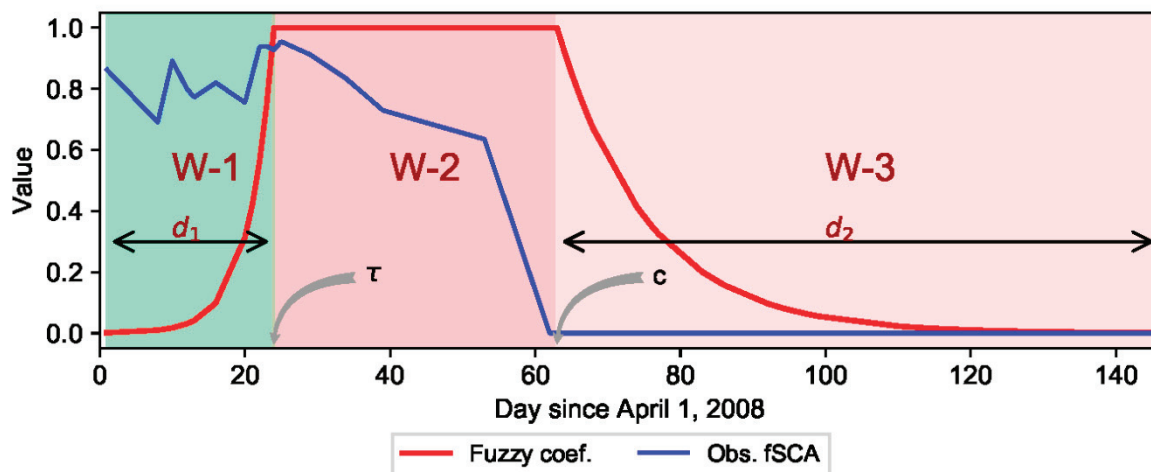


Figure 2. Location of the timing windows (W-1, W-2, and W-3) and value of the fuzzy coefficient in relation to a sample observed grid-cell fSCA dynamics over the ablation period as conceptualized in this study. τ and c respectively denote the change point and start of the complete melt-out period.

The informational value of MODIS fSCA observation is considered to be fuzzy and varying with location of each observation in the time axis and with respect to the aforementioned conceptualization of the informational value of fSCA in the three timing windows. A fuzzy coefficient (α) was introduced in the original formulations of Pbs and the LoA schemes in order to account for the variability in informational value of fSCA on the DA results. This coefficient was described by a fuzzy degree of membership function whose value is greater than zero based on the assumption that all observations contribute certain information in constraining the perturbation parameters. An exponential trapezoidal membership function was adopted to describe α (Equation (11)). The informational value of the MODIS fSCA observations is assumed to exponentially increase in W-1 until it reaches end of this timing window at τ . After start of the complete melt-out period (c), the informational value is assumed to exponentially decay with time in W-3. The fSCA observation is assumed to be most informative in the period between these two timing windows, i.e., in W-2. Hence, all observations in this period are assigned a maximum α value of unity.

$$\mu_{fSCA}(t) = \begin{cases} \exp\left(-\frac{\tau-t}{d_1}\right), t < \tau \\ 1, \tau \leq t \leq c \\ \exp\left(-\frac{t-c}{d_2}\right), t \geq c \end{cases} \quad (11)$$

where $\mu_{fSCA}(t)$ is the membership grade of each fSCA observation date index (t); τ and c respectively refer to the change point and start of the melt-out period in the time-axis. d_1 and d_2 respectively denote the widths of W-1 and W-3 in the time-axis.

The α value of each fSCA observation is equal to the membership grade corresponding to that observation ($\mu_{fSCA}(t)$); and this fuzzy coefficient is incorporated in the likelihood measure of Pbs in a way that it rescales the level of uncertainty of the residuals corresponding to each observation (Equation (12)).

$$w_j = \frac{\exp\left\{-0.5[(y - \hat{Y}_j)\alpha]^T R^{-1}[(y - \hat{Y}_j)\alpha]\right\}}{\sum_{j=1}^{N_e} \exp\left\{-0.5[(y - \hat{Y}_j)\alpha]^T R^{-1}[(y - \hat{Y}_j)\alpha]\right\}} \quad (12)$$

For the LoA-based DA scheme, the membership grades of the fSCA prediction errors ($\mu_{fSCA}(e)$, Equation (8)) were rescaled through multiplication by their respective α value before taking sum of $\mu_{fSCA}(e)$. The remaining steps are similar to the procedure followed in the LoA scheme described earlier, i.e. to get the product of w_{-e_j} and w_{-p_j} with subsequent rescaling of the result in such a way that the weights of all ensemble members would sum up to unity.

2.4. Detection of Critical Periods with Respect to the Informational Values of fSCA

The critical dates that form the timing windows and the fuzzy coefficient curve (Figure 2) were identified using a change point detection algorithm. The two critical dates, i.e., the point in the time axis where the change in mean snow cover of a given grid-cell occurs and start of the melt-out period in that grid cell, vary from year to year; and spatially from one grid-cell to another in response to various climatic and physiographic characteristics of an area. The use of a flexible change point detection scheme is a viable option to address the impact of such spatially and temporally variable factors on critical dates and thereby on spatiotemporal variability of SWE. The timing for start of the melt-out period in each grid cell was identified as the first incident with null fSCA observation. Detection of the CP location was realized by performing an offline change point analysis using the likelihood-based parametric approach. For the sake of comparison, change points were also estimated using the non-parametric approach. In order to minimize effect of the noisy data, the change point detection methods were applied over the transformed fSCA data (cumulative sum of fSCA). This section describes the methodology followed in detecting the change points using these two approaches.

The CP detection schemes are based on the idea that given n ordered sequence of data, $x_1 : n = (x_1, \dots, x_n)$, a change point is expected to occur within this set when there exists a time, $\tau \{1, \dots, n-1\}$, such that the statistical properties (e.g., mean or variance) of $\{x_1, \dots, x_\tau\}$ and $\{x_{\tau+1}, \dots, x_n\}$ are different in some way. The parametric change point analysis scheme employed in this study was adopted from previous studies focused on implementation of the methodology based on the likelihood ratio [36,66,67]. Change point detection was presented as a hypothesis test where the null hypothesis corresponds to no change in mean; while under the alternative hypothesis, a change point is expected to occur at time τ . The parametric method based on the likelihood ratio requires the calculation of the maximum log-likelihood under both null and alternative hypotheses [36]. Assuming the transformed fSCA data (x) as independent and normally distributed random variable, the likelihood ratio under the null (L_{H0}) and alternative (L_{H1}) hypotheses as well as the resulting log-likelihood ratio (R_τ) can be described as follows.

$$L_{H0} = \frac{1}{\sqrt{2\pi\sigma^2n}} \prod_{i=1}^n \exp\left[-\frac{(x_i - \mu_0)^2}{2\sigma^2}\right] \quad (13)$$

$$L_{H1} = \frac{1}{\sqrt{2\pi\sigma^2n}} \prod_{i=1}^{\tau} \exp\left[-\frac{(x_i - \mu_1)^2}{2\sigma^2}\right] \prod_{j=\tau+1}^n \exp\left[-\frac{(x_j - \mu_2)^2}{2\sigma^2}\right] \tag{14}$$

$$R_{\tau} = \log\left(\frac{L_{H1}}{L_{H0}}\right) = -\frac{1}{2\sigma^2} \left[\sum_{i=1}^{\tau} (x_i - \mu_1)^2 + \sum_{j=\tau+1}^n (x_j - \mu_2)^2 - \sum_{k=1}^n (x_k - \mu_0)^2 \right] \tag{15}$$

where μ_0, μ_1 and μ_2 respectively denote mean value of the transformed fSCA for the whole observation period, before the change point and after the change point. σ^2 refers to variance of the transformed fSCA data.

The generalized log-likelihood ratio (G) is estimated as:

$$G = \max_{1 \leq t \leq n} R_{\tau} \tag{16}$$

The null hypothesis is rejected if $2G > \lambda$, where λ is a predefined critical value; and the value of τ that maximizes R_{τ} is taken as the change point position. In this study the Bayesian Information Criterion (BIC) was employed to define this critical value, i.e., $\lambda = k \cdot \log(n)$, where k is the number of extra parameters used to define the change point (here $k = 1$).

When performing the change point analysis using a nonparametric approach, the method of Taylor [68] was followed. The procedure involves calculation of the cumulative sums (S_i) of the differences between the observations and their mean value (Equation (17)). The index in the time axis corresponding to the maximum cumulative sum (Equation (18)) is adopted as location of the change point τ .

$$S_i = S_{i-1} + (x_i - \bar{x}) \tag{17}$$

$$|S_m| = \max_{i=0, \dots, n} |S_i| \tag{18}$$

where x_i and \bar{x} respectively refer to the transformed fSCA observations and their mean value. For the first observation $S_0 = 0$ and the cumulative sum ends at $S_n = 0$.

The confidence level for the change point was determined by performing bootstrap analysis. First, the magnitude of the change is estimated as the difference between the maximum and minimum S_i (S_{diff}). Then, $N = 1000$ bootstraps of n observations are sampled through random reshuffling of the original data. For each bootstrap sample, the S_i values corresponding to each observation as well as the difference between the maximum and minimum S_i values (S_{diff}^0) are calculated. Accordingly, the percentage confidence level (CI) is estimated using Equation (19).

$$CI = 100 \frac{\sum_{i=1}^t score}{N} \tag{19}$$

where, $score = \begin{cases} 1, S_{diff}^0 < S_{diff} \\ 0, S_{diff}^0 \geq S_{diff} \end{cases}$

2.5. Evaluation Metrics

The performance of each DA scheme was evaluated through comparison of the median value of the ensemble predictions (*sim*) against the observed (*obs*) values. The following three evaluation metrics, i.e., mean absolute bias (MAB), root mean of squared errors (RMSE) as well as correlation coefficient (R) are employed during evaluation of estimated values against fSCA and SWE observations.

$$MAB = \frac{1}{N} \sum_{k=1}^N |obs_k - sim_k| \tag{20}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{k=1}^N (obs_k - sim_k)^2} \tag{21}$$

$$R = \frac{cov(obs, sim)}{\sigma_{obs} \cdot \sigma_{sim}} \quad (22)$$

where σ , cov and N respectively refer to standard deviation, covariance, and the number of observations.

3. Results

3.1. Comparison of Change Points Detected Using the Parametric and Non-Parametric Approaches

In this study, location of the change point detected using the parametric approach was also compared against those identified by the nonparametric approach. The parametric approach was based on the maximum likelihood estimate assuming that the observations are drawn from independent normal random distribution. The test for normality using chi-square on the raw and transformed fSCA data shows that although the percentage of grid cells with data that satisfy this criterion is generally low for both datasets, the result is higher for the transformed than the raw data. For example, the highest percentage of grid cells with normally distributed data at a significance level of 1% was 85% when using the transformed data as compared to 21% using the raw data. On the other hand, the non-parametric approach followed in this study is distribution free, i.e., it can be applied to any time series data regardless of the underlying probability distribution the data follows.

The change points detected using the parametric and non-parametric approaches yielded close results. Figure 3 shows the change points detected using both approaches for all grid-cells in two sample years (2012 and 2014). As mentioned in Section 2.2, these years are characterized by contrasting mean catchment SWE values during the peak accumulation period. The effect of the relatively high SWE value observed in year 2012 was reflected in the snow depletion curve of that year with the earliest grid-cells melt out day occurring after day 75 and in most of the grid-cells around day 100 (a). In year 2014, on the other hand, an early melt out date (around day 60) was observed in most of the grid-cells (b). The change point detection schemes were able to reveal this phenomenon. For example, in most of the grid-cells the τ value detected using the parametric approach was around day 25 in year 2014 as compared to day 50 in year 2012. Within a given year, the fSCA dynamics over time varied spatially from one grid cell to another. As a result, the critical dates in the ablation period, i.e., τ and start of the melt-out period for many of the grid-cells occurred at different points in the time axis. This variability in the critical dates in turn affects width of the timing windows (d_1 and d_2 , Equation (11)) and thereby the relative informational value of an fSCA observation from a particular date.

Tables 3 and 4 shows statistical summary of τ and melt-out dates of each analysis year based on spatial domain (i.e., grid-cell values). The statistical summary was computed based on the occurrence of the critical points in days since April 1 of each year (days) and time indices that show relative position of the critical points in the time axis (index). It can be noticed that the τ detected using non-parametric approach occurred at a relatively latter point in the time axis as compared to those identified using parametric approach with an average difference of about three days (one time-step). The spatial variability is relatively higher for τ detected using the non-parametric approach with an overall average value of 30 days as compared to 26 days for τ detected using the parametric approach. Based on the bootstrap test, the non-parametric approach was able to detect τ with over 99% confidence level for all grid cells and analysis years. From this table it can also be observed that start of the complete melt-out period varied spatially from one grid-cell to another within a given year and temporally between the different years. For example, highest negative skew of -0.8 and positive skew of 0.4 were respectively observed in years 2011 and 2015 (based on the time indices). Similarly, the average number of fSCA observations per grid cell varied from year to year ranging from 32 to 47.

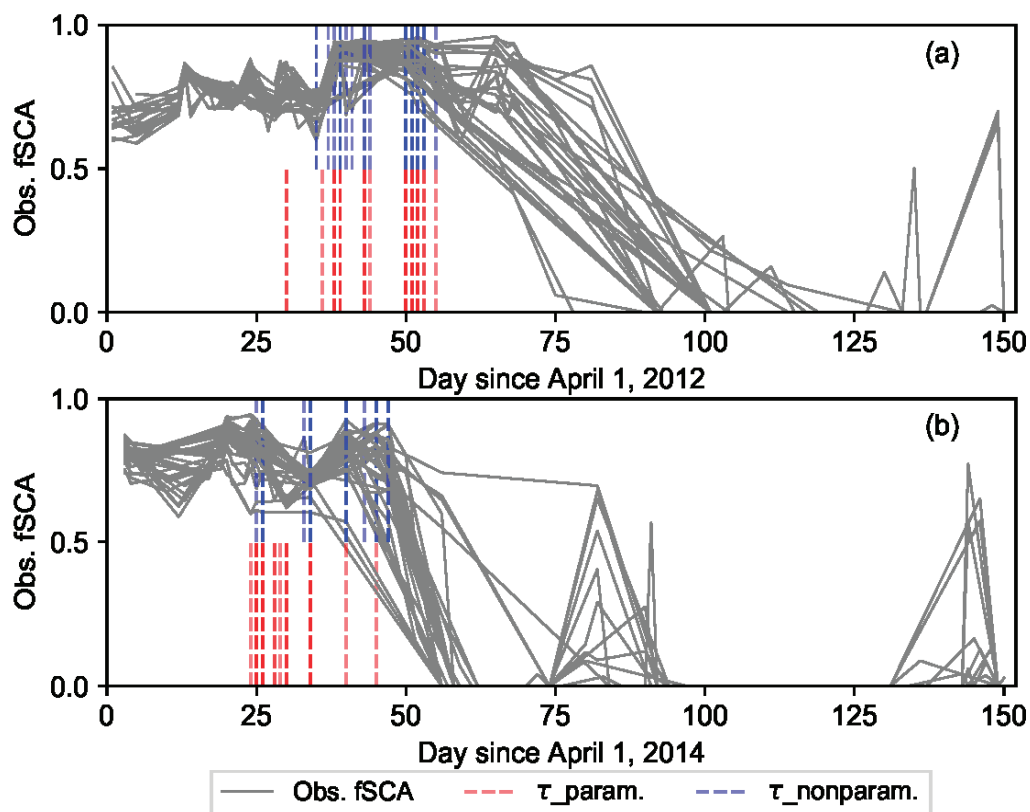


Figure 3. Spatiotemporal variability of the fSCA dynamics and location of the change points (τ_{param} and τ_{nonparam}) detected using parametric and non-parametric approaches for 33 grid-cells in two sample years, year 2012 (a) and year 2014 (b). Many of the change points are overlapping and the color intensity shows the degree of overlap in τ values of different grid-cells.

3.2. Evaluation of the DA Schemes Against Observed Data

We assess the effect of the different DA schemes through comparison of the median simulated against observed SWE values. For each hydrologic year, the median values resulting from each of the 100 iterations are averaged and compared against the observed values using the evaluation metrics. Figure 4 shows the evaluation result of the DA schemes in their original formulation (Pbs and LoA) as well as with fuzzy coefficients (Pbs_F and LoA_F). The DA schemes yielded an improved estimate of SWE as compared to the prior simulation in all years with the exception of years 2010 and 2015. The relative performance of Pbs and LoA as well as Pbs_F and LoA_F varied from year to year. A similar result was also observed between LoA and LoA_F. On the other hand, the incorporation of a fuzzy coefficient in Pbs resulted to an improved estimate of SWE in all years as compared to the original formulation. The fSCA-based evaluation using similar metrics also revealed that the DA schemes resulted to an improved estimate of fSCA as compared to the prior values. While an improved performance was observed when using LoA_F as compared to LoA in terms of the three evaluation metrics, a slight deterioration in MAB and RMSE was observed when using Pbs_F as compared to Pbs. In most of the analysis years, LoA has shown significant improvement both in SWE and fSCA estimation after the persistency in reproducing the observations was included as a hydrologic signature.

Table 3. Annual statistical summary of change points (τ) in observed fSCA of the grid-cells detected using the parametric (Par) and non-parametric (Npar) approaches as well as the melt-out points (c) in the time axis: Critical point statistics conducted on days since April 1.

Year	Variable	Statistics														Avg. No. Obs.
		Min		Max		Mean		Median		Variance		Skewness				
		Par	Npar	Par	Npar	Par	Npar	Par	Npar	Par	Npar	Par	Npar			
2008	τ	22.0	23.0	41.0	49.0	28.1	33.0	25.0	34.0	26.1	56.8	0.8	0.4	42.5		
	c	57.0		94.0		65.6		64.0		67.6		2.6				
2009	τ	22.0	25.0	35.0	46.0	27.9	31.2	29.0	31.0	6.1	16.0	-0.1	1.5	42.7		
	c	51.0		85.0		68.4		60.0		165.6		0.3				
2010	τ	33.0	36.0	40.0	41.0	36.2	37.8	36.0	38.0	2.1	2.9	0.3	0.4	35.4		
	c	69.0		92.0		76.6		70.0		102.7		0.9				
2011	τ	22.0	23.0	35.0	38.0	27.5	29.6	28.0	29.0	8.4	14.0	0.0	0.3	33.7		
	c	50.0		96.0		68.4		69.0		116.9		0.5				
2012	τ	30.0	35.0	55.0	55.0	45.2	46.0	50.0	50.0	49.9	40.6	-0.5	-0.3	32.2		
	c	78.0		133.0		99.8		101.0		96.2		1.2				
2013	τ	15.0	15.0	39.0	43.0	30.8	34.3	34.0	37.0	65.8	57.3	-0.8	-1.1	36.0		
	c	47.0		96.0		56.5		55.0		103.0		2.0				
2014	τ	24.0	25.0	45.0	47.0	28.2	37.6	26.0	40.0	22.1	54.5	1.9	-0.3	47.1		
	c	56.0		94.0		60.6		58.0		70.7		2.7				
2015	τ	27.0	34.0	47.0	50.0	37.2	43.0	38.0	41.0	41.5	21.6	-0.1	-0.2	35.1		
	c	80.0		122.0		90.3		88.0		78.0		1.9				
2016	τ	24.0	33.0	40.0	45.0	34.7	37.0	37.0	37.0	12.4	6.0	-1.2	0.5	34.6		
	c	60.0		93.0		66.6		61.0		108.5		1.6				
Avg.	τ	24.3	27.7	41.9	46.0	32.9	36.6	33.7	37.4	26.0	30.0	0.0	0.1	37.7		
	c	60.9		100.6		72.5		69.6		101.0		1.5				

Table 4. Annual statistical summary of change points (τ) in observed fSCA of the grid-cells detected using the parametric (Par) and non-parametric (Npar) approaches as well as the melt-out points (c) in the time axis: Critical point statistics conducted on time indices.

Year	Variable	Statistics												Avg. No. Obs.
		Min		Max		Mean		Median		Variance		Skewness		
		Par	Npar	Par	Npar	Par	Npar	Par	Npar	Par	Npar	Par	Npar	
2008	τ	7.0	8.0	14.0	16.0	10.0	11.6	10.0	12.0	2.8	4.1	0.2	0.0	42.5
	c	11.0		24.0		16.5		17.0		10.2		0.1		
2009	τ	8.0	9.0	12.0	15.0	10.5	12.4	11.0	13.0	1.6	2.3	-0.7	-0.6	42.7
	c	13.0		21.0		17.7		18.0		5.2		-0.7		
2010	τ	6.0	7.0	12.0	14.0	9.0	10.1	9.0	10.0	3.2	3.9	0.0	0.1	35.4
	c	10.0		19.0		14.5		14.0		7.9		0.0		
2011	τ	3.0	4.0	13.0	14.0	8.8	10.0	9.0	11.0	6.1	6.8	-0.8	-0.9	33.7
	c	7.0		21.0		15.1		16.0		14.5		-0.8		
2012	τ	8.0	9.0	15.0	15.0	12.1	12.5	12.0	13.0	3.4	3.0	-0.5	-0.4	32.2
	c	13.0		25.0		19.8		20.0		9.2		-0.4		
2013	τ	4.0	5.0	11.0	12.0	7.4	8.3	7.0	8.0	3.0	3.8	-0.1	-0.2	36.0
	c	7.0		17.0		11.7		12.0		6.7		0.1		
2014	τ	5.0	6.0	9.0	12.0	7.5	9.1	7.0	9.0	1.2	2.0	-0.3	0.0	47.1
	c	9.0		15.0		12.1		12.0		3.0		0.0		
2015	τ	5.0	7.0	12.0	13.0	8.7	9.8	8.0	10.0	2.8	3.0	0.2	0.3	35.1
	c	10.0		19.0		14.1		14.0		6.4		0.4		
2016	τ	7.0	7.0	12.0	13.0	9.4	10.2	9.0	10.0	1.9	1.9	0.0	-0.2	34.6
	c	11.0		20.0		15.3		16.0		5.5		-0.1		
Avg.	τ	5.9	6.9	12.2	13.8	9.3	10.4	9.1	10.7	2.9	3.4	-0.2	-0.2	37.7
	c	10.1		20.1		15.2		15.4		7.6		-0.2		

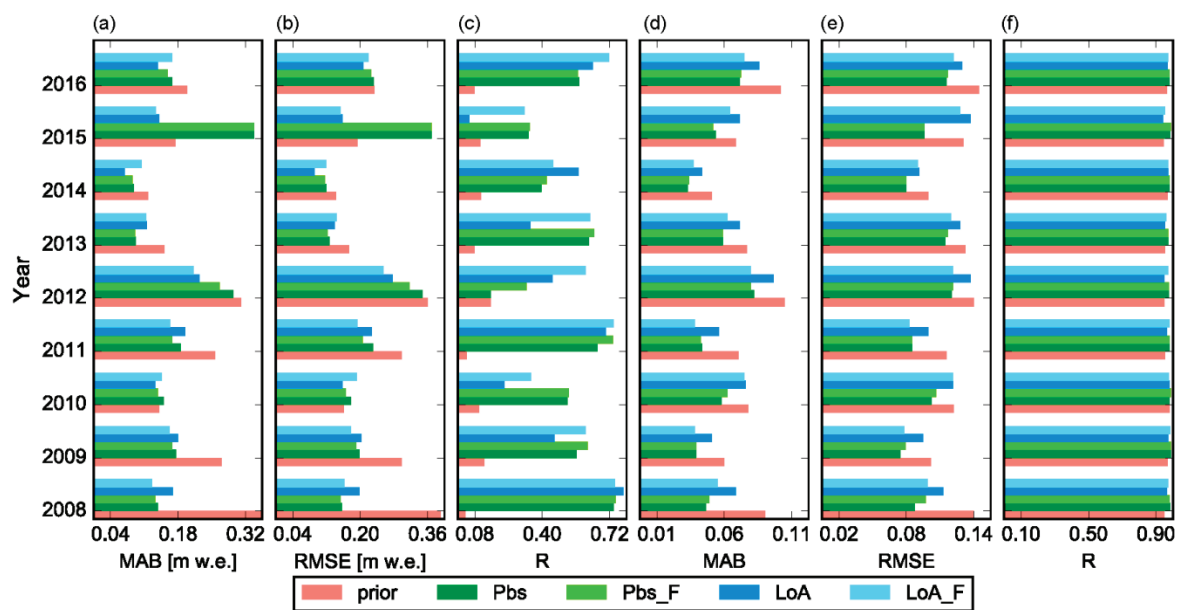


Figure 4. Comparison of simulated against observed SWE using MAB (a), RMSE (b) and R (c) as well as simulated against observed fSCA using MAB (d), RMSE (e) and R (f) under the scenarios without DA (prior) as well as with DA using Pbs, Pbs_F, LoA and LoA_F. The evaluation metrics are calculated using spatial (grid-cell) values.

Figure 5 compares observed and median simulated SWE for all years under the four DA schemes and at four elevation ranges. It shows that the prior predictions tend to overestimate the low SWE values especially for grid-cells located at the highest elevation range, i.e., above 1200 masl, and underestimate the high values (a). This phenomenon was improved in the posterior SWE estimates using the DA schemes. The line fitted to the observed versus reanalyzed SWE values (red line) yielded a closer slope to the identity line (broken grey line) as compared to the prior (dark full line) (b, c, d and e). The reanalyzed values are also closer to the identity line under Pbs with a slope of 0.57 and R value of 0.59 (b) than the LoA with a slope of 0.34 and R value of 0.56 (c). However, the number of outliers in the reanalyzed values is generally lower under LoA as compared to Pbs. This is due to the fairly distributed weight between ensemble members in LoA as compared to Pbs resulting to less surprises under variable conditions. A better fit was also observed when using the DA schemes with fuzzy coefficient (Pbs_F and LoA_F) as compared to the schemes with the original formulation (Pbs and LoA). The improvement was more significant for the DA schemes based on LoA where the slope of the line fitted to the reanalyzed versus observed values has increased from 0.34 to 0.53 and the R value has improved from 0.56 to 0.65 (e). It can also be noticed that most of the simulated SWE values for grid-cells that lie in the two extreme elevation ranges, i.e., <800 m and >1200 m lie above the identity line implying that they are overestimated. This can be attributed to the errors emanating from interpolation of the point forcings in these elevation ranges. The observed low SWE values are overestimated while high values are underestimated both before and after reanalysis using all DA schemes. This phenomenon is relatively more pronounced when using LoA than Pbs.

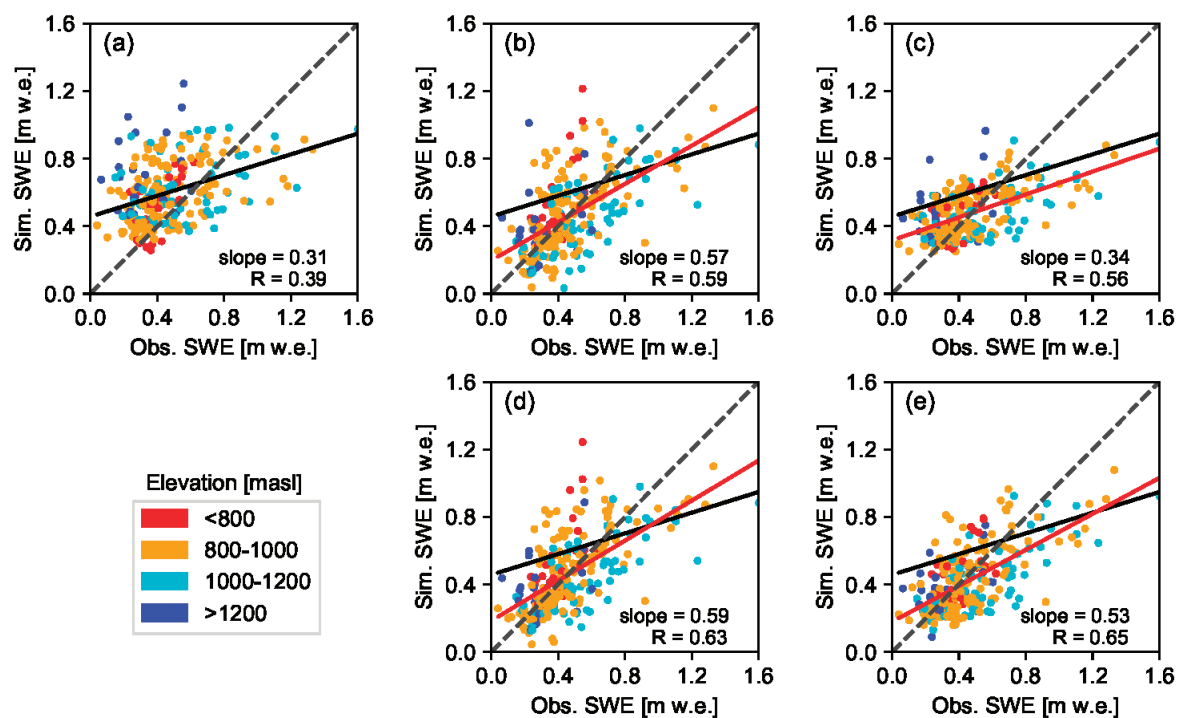


Figure 5. Correlation plot between observed and reanalyzed SWE using the prior estimate (a) and Pbs (b), LoA (c), Pbs_F (d), and LoA_F (e) depicting the effect of DA on estimated values from grid cells located at four elevation ranges. The broken grey line represents the 1:1 identity line, while the full dark line and red line are respectively fitted to the prior and reanalyzed values. The slope and R values in plots (b), (c), (d), and (e) refer to the reanalyzed versus observed SWE fitted lines.

The DA result and prior values were also analyzed by aggregating the grid cells into their respective sites (Figure 6). The average SWE value of a given site was estimated from the individual SWE values of the grid-cells in the site. Generally, a better fit between observed and simulated SWE values can be noticed at site level as compared to the analysis result at grid-cell level. For example, when using Pbs_F, the slope and R values of the line fitted to the reanalyzed versus observed SWE values have respectively increased from 0.59 to 0.80 and from 0.63 to 0.69 (i.e., an 8% increase in R). This can be attributed to the lower number of outliers at site level due to the smoothing effect of the aggregation. Similar to the analysis at grid-cell level, a better fit was observed when using Pbs as compared to LoA; and the effect of accounting for the informational value of fSCA observations was more significant in LoA_F than in Pbs_F. The value of R has increased by 16% after introducing the fuzzy coefficient in LoA. No significant spatial pattern in effect of the DA was observed with the exception that the SWE values of site 6 were less constrained by the DA schemes and the estimated SWE values of sites 1 and 4 were relatively underestimated when using the LoA approach.

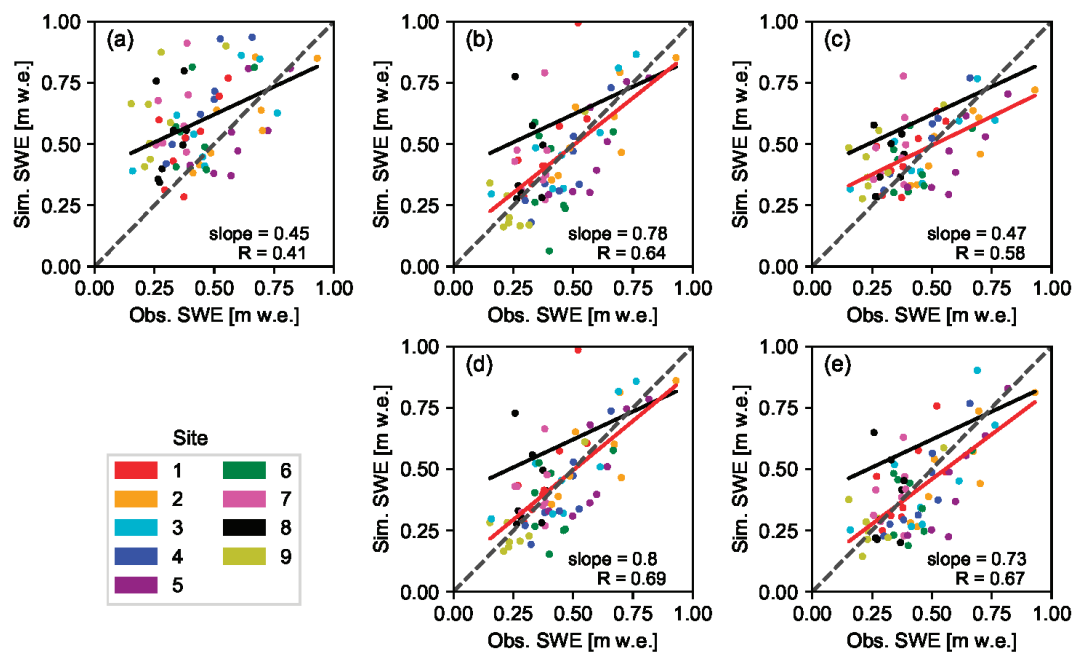


Figure 6. Correlation graph of site-averaged observed and predicted SWE values based on the prior estimates (a) and the four DA schemes, i.e. Pbs (b), LoA (c), Pbs_F (d), and LoA_F (e). The nine sites are represented by different colors and the lines in each sub-plot are as explained in Figure 5.

Figure 7 presents a sample annual inter-comparison of the prior and reanalyzed SWE estimates against observed values. The comparison was based on site-averaged observed and simulated SWE values. The relative performance of the DA schemes varied both spatially from one site to another and temporally from year to year. For example, while the SWE values reanalyzed using Pbs were closer to the observed values at sites 7 and 9 than those from LoA in year 2008 (a), the latter was better at reproducing the observed SWE values at sites 4 and 5 for the same year. A similar example can be mentioned from a temporal perspective; Pbs showed better performance in years 2009 (b) and 2010 (c) at site 1, while LoA yielded better result in years 2011 (d) and 2012 (e) for the same site. An improvement was also observed in uncertainty of the SWE estimates after reanalysis using the DA schemes, although the level of uncertainty for a given DA scheme varied spatially from one site to another and temporally from year to year. Relatively low level of uncertainty was observed on SWE values reanalyzed using Pbs and Pbs_F as compared to LoA and LoA_F. The inclusion of fuzzy coefficient in LoA yielded to an improved level of uncertainty. On the other hand, the prior and LoA resulted in SWE values with high level of uncertainty. In Figure 7, the upper quantile (95%) values from the prior and LoA are scaled down to 10% of their respective values in order to maintain proportional look with results from the other DA schemes displayed in the same plot.

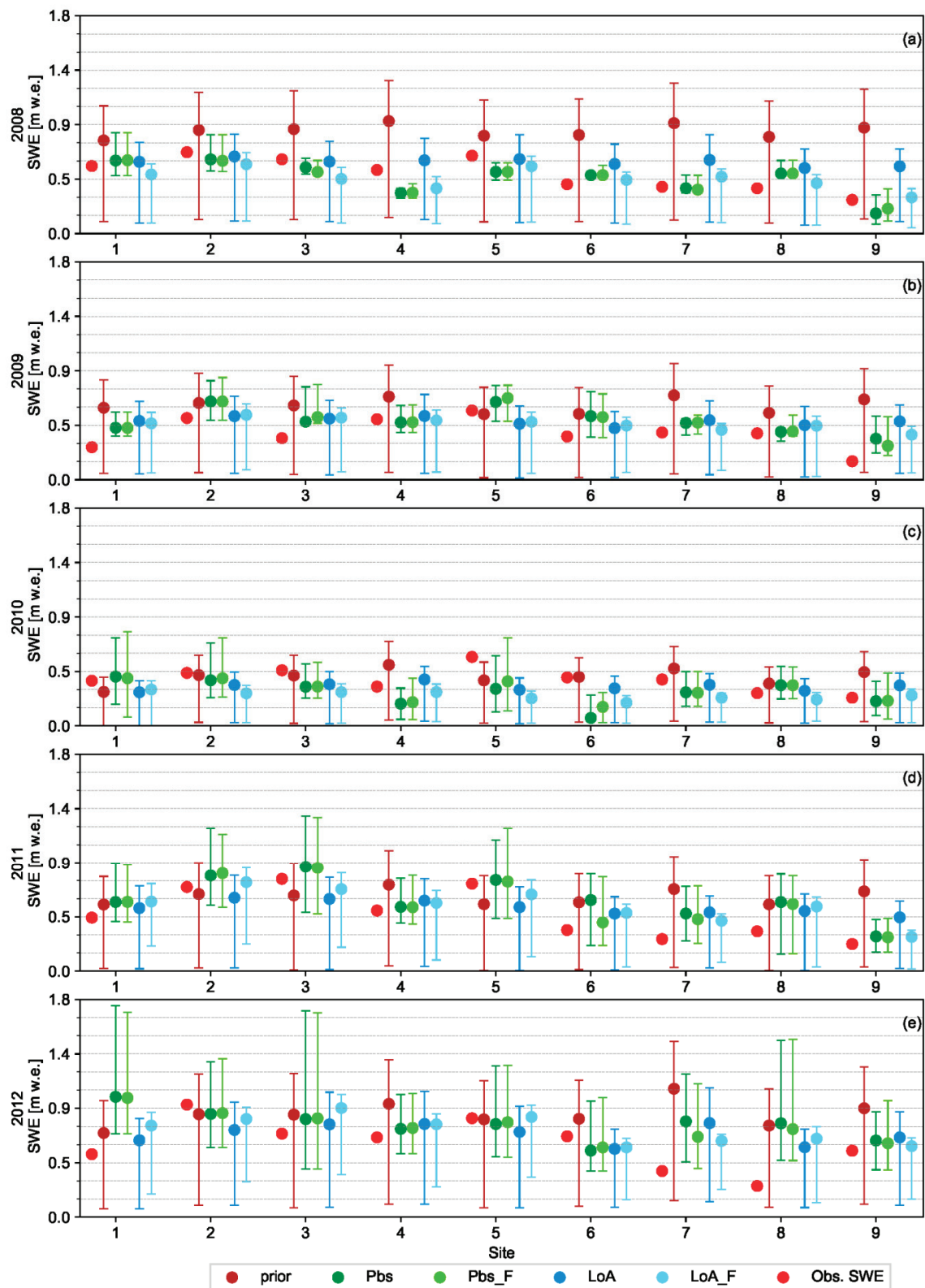


Figure 7. Comparison of site-averaged observed against the prior and reanalyzed SWE values using the original DA schemes (Pbs and LoA) and the DA schemes with fuzzy coefficient (Pbs_F and LoA_F) for years 2008 (a)–2012 (e). Error bars denote the 90% prediction interval. For the prior and LoA schemes the upper quantile (95%) values are scaled down to 10% of their respective values.

The difference in the allocation of weight between the ensemble members under the four DA schemes and the prior partly explains the difference in the width of the resulting uncertainty bound of their SWE predictions. Some ensemble members may yield very high or low SWE prediction depending on the combination of the perturbed parameter values. The effect of these predictions on the resulting quantile values, however, depends on the weight assigned to such ensemble members. In prior estimate, all ensemble members are assigned equal weight of $1/N_e$. LoA also tends to maintain fair distribution of the total weight among the ensemble members as compared to Pbs, while most of the weight is carried by few ensemble members when using the latter. For example, the maximum weight carried by a single ensemble member when using Pbs was 54.6% as compared to only 1.6% in LoA-based on analysis carried out in year 2008 and a sample grid-cell (result not shown). As such, when using LoA the quantile values fairly represent the SWE predictions of several ensemble members; and this yields to wide uncertainty bound. Whereas, Pbs yields to narrow uncertainty bounds of predicted SWE, although not necessarily bracketing the observed SWE values. For example, the uncertainty bounds fail to bracket the observed SWE values in sites 4, 5, and 8 when using Pbs in year 2008 (Figure 7a). However, this phenomenon was not consistent throughout all years since this DA scheme was able to bracket the observed SWE of the same sites in other years, for example in years 2011 and 2012 (Figure 7d,e).

Performance of the different DA schemes was also analyzed in terms of their capability in reproducing the snow depletion curve (Figure 8). The fSCA values of all grid-cells are aggregated to estimate a catchment scale daily mean value. The result shows that, the complete melt-out period occurred at a later time step when using the prior estimates as compared to the reanalyzed values using the DA schemes. As such, the median prior values tend to overestimate the fSCA observations during this period. This phenomenon was observed in all years with the exception of year 2011; where the prior yielded relatively better estimate than Pbs. Neither the prior nor the posterior median estimates were able to adequately reproduce the fSCA observations in the month of April. Visual inspection of these plots also shows that the uncertainty associated with fSCA estimates was highest under LoA as compared to the other DA schemes. On the other hand, the proportion of fSCA observations bracketed by the 90% uncertainty bound was highest under this scheme. This implies that fSCA observations during anomaly years might be better reproduced by the LoA as compared to the other DA schemes. The uncertainty bound of each DA scheme also varied from year to year.

3.3. Sensitivity of the DA Evaluation Metrics to Change in Location of the Critical Points

In this study the fSCA observations that fall in different timing windows are assumed to have varying degree of informational value in constraining model states and parameters; with observations between the critical points, i.e., the detected τ and the first incidence of null fSCA, carrying the highest informational value. The validity of this assumption was examined by moving the window with highest value of fuzzy coefficient (W-2) forward and backward in the time axis, followed by subsequent evaluation of its effect on the DA result using Pbs and LoA schemes (Figure 9). The percentage change in τ was calculated using the total number of observation time indices between April 1 and the detected τ as a reference. Since, the number of observations is generally low during this period, different percentage changes in τ might point to the same nearest time index; yielding similar value in the evaluation metrics. The result shows that the evaluation metrics started to deteriorate when the location of τ was shifted forward by about 20% (i.e., -20%) and backward by about 50% in the time axis. The sensitivity was also higher when the location of τ was moved forward as compared to backward in the time axis. This can be attributed to the relatively low signal to noise ratio of the observations in W-1 as compared to those in W-3. Further, the sensitivity to the change in location of τ was higher under Pbs than the LoA. The sensitivity curves of the three DA performance metrics have shown similar trend when applied on SWE (a, c and e) and fSCA (b, d and f) in response to the displacement of τ in the time axis.

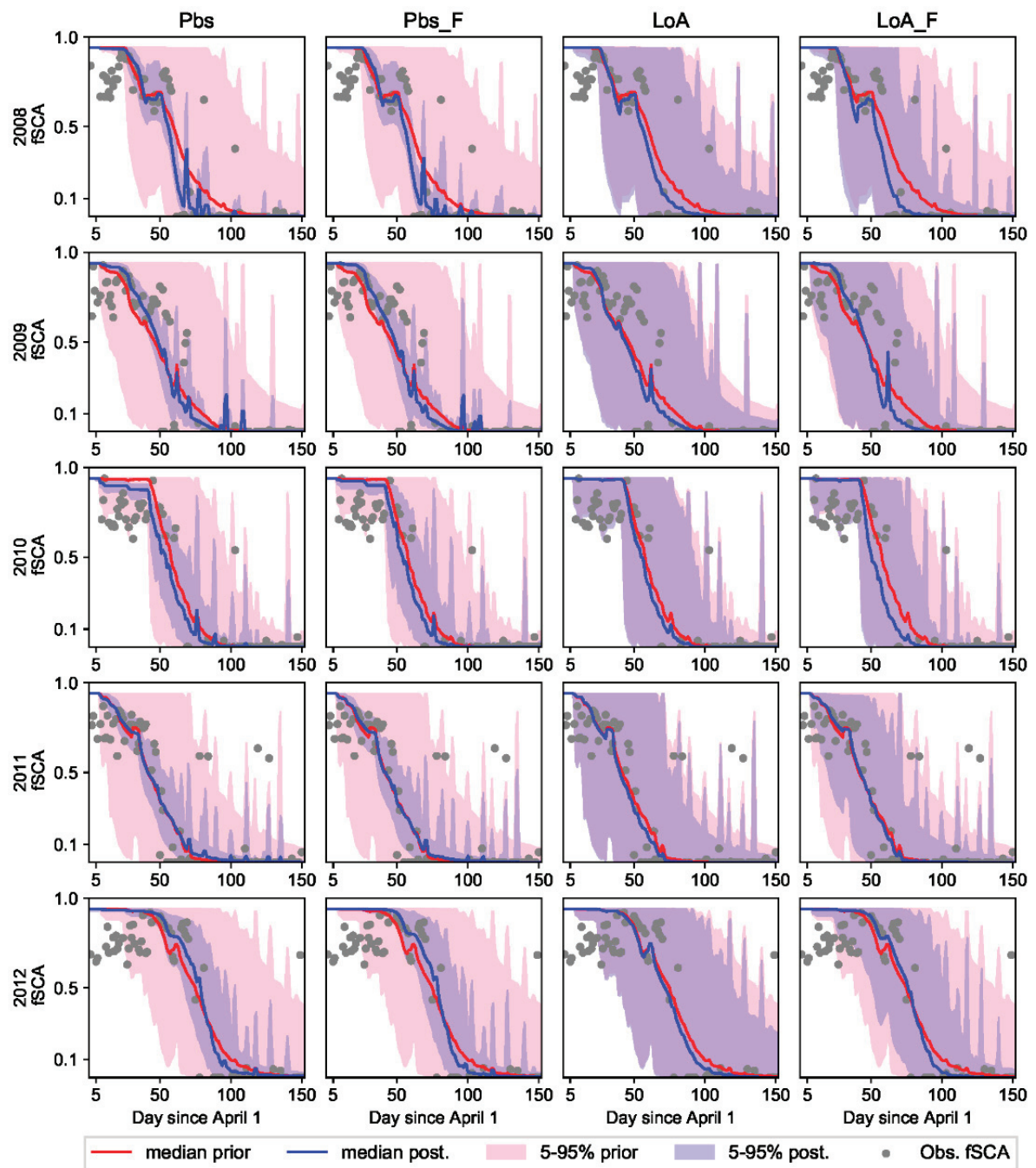


Figure 8. Median and prediction uncertainty of fSCA for the prior and posterior estimates reanalyzed using the different DA schemes (Pbs, LoA, Pbs_F, and LoA_F).

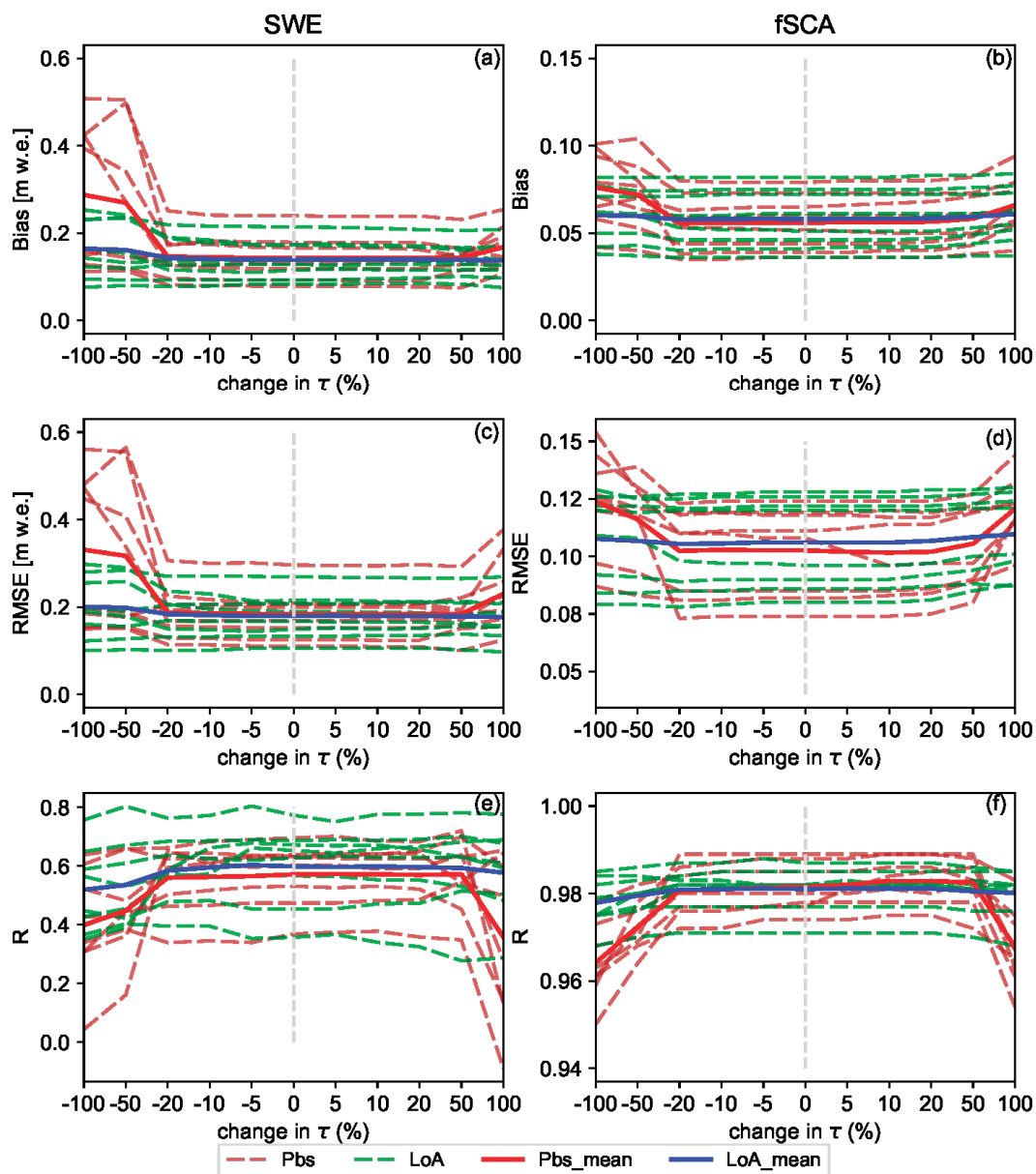


Figure 9. Sensitivity of DA performance metrics of SWE (a,c,e) and fSCA (b,d,f) to change in location of the critical points (τ and complete melt-out) in relation to the points detected using the automatic detection scheme. Broken lines represent sensitivity result of the evaluation metrics in each year and bold lines represent average value of all years.

4. Discussion

In this study four ensemble-based batch smoother schemes were used to retrospectively estimate SWE during the peak accumulation period through assimilation of fSCA into a snow hydrological model. Effect of the DA schemes was analyzed for nine years and with due consideration to different elevation zones as well as two spatial scales, i.e., at grid-cell and site levels. One of these schemes was based on a Bayesian approach (Pbs), while the other was based on the limits of acceptability concept (LoA). In both approaches, model predictions were scaled proportional to weight of each ensemble member. Due to nature of the likelihood function used, most of the weights in Pbs were carried by few ensemble members, while in the LoA framework, the weights were fairly shared by all ensemble members.

Although reanalysis using these DA schemes has yielded a significant improvement of the correlation coefficient in all analysis years, only a slight improvement was observed in most of these years when accuracy of the reanalyzed SWE was compared against the prior values based on the MAB and RMSE statistical criteria. The result was generally encouraging given the high level of uncertainty associated with remote sensing data in high latitude areas due to frequent clouds and polar darkness. Other similar studies focused on assimilation of MODIS fSCA into hydrological models have also reported only a modest improvement in SWE and streamflow estimates after assimilation; and it was mainly attributed to the very short transition period between the full snow cover and the complete melt-out dates [20,24]. On the other hand, studies based on the assimilation of multiple-source remote sensing data have reported significant improvement in these evaluation metrics [28].

The relative performance of Pbs and LoA varied in response to various factors. LoA was generally more resistant to uncertainties in the assimilated data. When using low quality fSCA measurements with several missing observations, the LoA may yield better result than Pbs. For example, LoA was able to yield relatively better result as compared to the prior estimates in years 2010 and 2015 where there was an overall low performance of the DA schemes in general and Pbs in particular (Figure 4). The sensitivity test shown in Figure 9 also reveals that the Pbs performance gets easily deteriorated as compared to the LoA due to the displacement of the timing window with highest coefficient ($W-2$) following the changes in location of τ . A further advantage of the LoA is that different intuitive evaluation criteria including hydrologic signatures can easily be accommodated in constraining model states and parameters. For example, the overall grade and persistency of an ensemble member in reproducing the observed fSCA within the assumed measurement error bounds are included as evaluation criteria in this study. However, in the presence of several correlated observations such as the fSCA observations from the complete melt-out period, LoA may suffer from identifiability problem to distinguish the best performing ensemble members and this leads to low performance as compared to Pbs. On the other hand, the likelihood with the product of squared residuals renders Pbs a better capability in identifying the optimal ensemble members under such circumstances.

The overall low performance of the DA schemes in years 2010 and 2015 can be attributed to different factors. Following year 2014, these years have lowest observed mean catchment SWE value of 419 and 458 mm respectively as compared to the other years. However, these values are not too far from the average annual value of 525 mm to explain the relatively poor performance of the DA schemes in these years. The total number of fSCA observations in these years and distribution of the observations in the time axis were also assessed since they are expected to have an effect on the DA result. Similarly, no significant difference on the number and distribution of the observations was noticed as compared to the other years. Future studies may thus assess quality of the assimilated MODIS fSCA observations and climatic inputs in relation to the other years since such factors are also expected to play an important role in the DA result.

A fuzzy coefficient was introduced into the original formulations of the two DA schemes in order to take into account for the difference in informational value of fSCA measurements from different timing windows in the ablation period. An fSCA observation in a given time index can assume different informational value depending on its location in the time axis in relation to the critical points. In Pbs and other schemes with square of errors-based likelihood function, value of the efficiency criteria is biased towards data with highest uncertainty, with the bias proportional to the uncertainty squared [30]. The fuzzy coefficient was thus introduced to rescale the level of this uncertainty with due consideration to the assumed informational value of each observation. The fSCA observations from the more informative period were promoted by allocating higher importance weight, while observations from the other periods were punished in accordance to their location within their respective timing windows. A further assumption in this study was that all observations carry some informational value; albeit to varying degree.

The informational value of fSCA observations with time was assumed to follow an exponential-trapezoidal membership function that broadly divides the ablation period into three

timing windows. In W-1, the fSCA measurements close to the onset of snowmelt are generally characterized by low signal to noise ratio. Similarly, during the complete melt-out period (W-3) the measurements are homogenous; and the informational value generally decreases with correlated data that leads to less power in constraining model states and parameters. The critical points of the timing windows, i.e., τ and start of the melt-out period, vary both spatially and temporally in response to various climatic and physiographic factors. This has highlighted the need for a flexible and automatic detection of the critical points. Both the parametric and nonparametric change point detection schemes used in this study yielded similar results with well characterization of the changes. A change point analysis conducted on the transformed data (running total) was found to be more powerful and robust to outliers than the one conducted on the raw data. This can be attributed to the smoothing effect of the noises in the former case. Studies dealing with critical point detection such as the timing of snowmelt onset are also crucial to a better estimate of the surface energy balance in high latitudes, since the fraction of available solar radiation that is actually absorbed almost quadruple during the onset of snowmelt [3].

The results obtained in this study through assimilation of fSCA into a hydrologic model could also be extended to land surface models (LSMs). Inaccuracies in snow estimates within LSMs can lead to substantial errors in simulation results since snow cover has a significant impact on the surface radiation budget as well as turbulent energy fluxes to the atmosphere [69]. An assimilation of fSCA into LSMs was reported to improve the simulation result as compared to an open-loop model without assimilation [69,70]. Thus, the fuzzy logic-based DA concept presented in this study can also be applied in the assimilation of fSCA, or other variables that display fuzzy informational value with time, into LSMs in order to get an improved result. Alternatively, the use of these schemes could indirectly yield an improvement in the performances of LSMs through improvement in the hydrologic component when LSMs are coupled with the latter. In recent years, there has been growing interest in incorporating a hydrologic component into land surface models (LSMs) in order to improve weather and extreme hydrologic condition forecasts [71,72].

Generally an improved estimate of SWE was observed as a result of reanalysis using the DA schemes with a fuzzy coefficient, although, a slight deterioration in some of the evaluation metrics of fSCA was observed when using Pbs_F. This can be attributed to the relatively lower weight assigned to observations from start of the melting period coupled with the fact that most of the weights are carried only by few ensemble members. During the onset of snowmelt, sensitivity of the SDC to the initial bare ground fraction is highest and updates during this period are more likely to be dominated by this parameter than the snow coefficient of variation or precipitation [52]. The increase in fSCA error was, however, very small as compared to the improvement attained in SWE estimate. Results from this study suggest that, the concept of variable informational value of fSCA observations is consistent with the notion that many of the variables that we usually consider to be crisp quantity and deterministic are actually fuzzy that carry considerable amount of uncertainty [64].

5. Conclusions

Two ensemble-based data assimilation schemes, i.e. particle batch smoother (Pbs) and the limits of acceptability (LoA) as well as newly introduced versions of these schemes that account for the variability in informational value of the assimilated observation (fraction of snow cover area, fSCA), i.e., Pbs_F and LoA_F, were used to reanalyze the model results. Using the LoA approach as a DA scheme yielded a promising result. All DA schemes resulted in a posterior snow water equivalent (SWE) estimate that is better than the prior estimate in terms of accuracy as measured using one or more of the efficiency criteria used in this study. The result was generally encouraging given the high level of uncertainty associated with optical remote sensing data in general, and MODIS fSCA product in particular in high latitude areas.

Although all fSCA observations in the ablation period were important in constraining the perturbation parameters, some observations were more important than others depending on their

location in the time axis with respect to certain critical points in the melt season, i.e., the points where the mean snow cover changes (τ) and the start of a melt-out period. These critical points varied from year to year and spatially between the grid cells. The parametric (likelihood-based) and non-parametric change point detection schemes employed to locate τ in each grid cell and year yielded reasonable results. A fuzzy coefficient was introduced in the original formulations of Pbs and LoA that renders highest importance weight to the assumed most informative timing window, i.e., to observations between these critical points. A sensitivity test conducted by moving the fuzzy coefficient value assigned to this timing window forward and backward in the time axis revealed that, although somewhat insensitive up to certain distance from its original location, moving this window farther lead to deterioration of the DA results.

The reanalyzed SWE using Pbs_F was superior to Pbs throughout all years, although the degree of improvement in the evaluation metrics varied from year to year. Similarly, an improved estimate of fSCA was obtained for all years when using LoA_F as compared to LoA. On the other hand, the improvement in SWE estimate when using LoA_F in comparison to LoA as measured using a given efficiency criterion was not consistent throughout the analyses years. In many of the analysis years, the relative performance of the DA schemes was also variable from one site to another. Future studies can examine the physiographic and climatic factors affecting the relative performance of these DA schemes in order to get a better insight into the relationship between the underlying concept of the DA schemes and these factors.

Generally, incorporating the variable informational value of the remote sensing data in the DA schemes was a viable option for an improved estimates of the perturbation parameters and thereby the reanalyzed SWE values. The assumed informational values of the fSCA observations with respect to their location in the time axis was based on the general concept that observations with high uncertainty are more likely to have more informational value than certain observations. Future studies can be focused on assessment of the informational values using the well-established techniques based on the information theory and statistical measures. Although this study was focused on fSCA assimilation, the fuzzy logic-based ensemble frameworks presented in this study can be applied to assimilation of other measurements that display variable informational value with time.

Author Contributions: A.T. designed and performed the analysis and wrote the manuscript; J.B. and T.S. supervised the study and reviewed the manuscript; C.Y. reviewed the manuscript.

Funding: This research was conducted within the Norwegian Research Council's—Enhancing Snow Competency of Models and Operators (ESCYMO) project (NFR no. 244024) and in cooperation with the strategic research initiative LATICE (Faculty of Mathematics and Natural Sciences, University of Oslo <https://mn.uio.no/latice>).

Acknowledgments: We thank Statkraft AS, in particular Gaute Lappegard for making the hydro-meteorological data of the Nea-catchment available and Knut Sand for providing us the radar SWE measurements. We also thank Kristoffer Aalstad for providing us a pseudocode for particle batch smoother.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Painter, T.H.; Rittger, K.; McKenzie, C.; Slaughter, P.; Davis, R.E.; Dozier, J. Retrieval of subpixel snow covered area, grain size, and albedo from MODIS. *Remote Sens. Environ.* **2009**, *113*, 868–879. [[CrossRef](#)]
2. Trujillo, E.; Molotch, N.P.; Goulden, M.L.; Kelly, A.E.; Bales, R.C. Elevation-dependent influence of snow accumulation on forest greening. *Nat. Geosci.* **2012**, *5*, 705. [[CrossRef](#)]
3. Zhang, T.; Stamnes, K.; Bowling, S. Impact of the atmospheric thickness on the atmospheric downwelling longwave radiation and snowmelt under clear-sky conditions in the Arctic and Subarctic. *J. Clim.* **2001**, *14*, 920–939. [[CrossRef](#)]
4. Li, D.; Durand, M.; Margulis, S.A. Estimating snow water equivalent in a Sierra Nevada watershed via spaceborne radiance data assimilation. *Water Resour. Res.* **2017**, *53*, 647–671. [[CrossRef](#)]
5. Maurer, E.P.; Stewart, I.; Bonfils, C.; Duffy, P.B.; Cayan, D. Detection, attribution, and sensitivity of trends toward earlier streamflow in the Sierra Nevada. *J. Geophys. Res. Atmos.* **2007**, *112*. [[CrossRef](#)]

6. Kirnbauer, R.; Blöschl, G.; Gutknecht, D. Entering the Era of Distributed Snow Models: Paper presented at EGS XVIII General Assembly (Wiesbaden, Germany—May 1993). *Hydrol. Res.* **1994**, *25*, 1–24. [[CrossRef](#)]
7. Clark, M.P.; Hendrikx, J.; Slater, A.G.; Kavetski, D.; Anderson, B.; Cullen, N.J.; Kerr, T.; Hreinsson, E.Ö.; Woods, R.A. Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review. *Water Resour. Res.* **2011**, *47*. [[CrossRef](#)]
8. Beven, K.; Smith, P. Concepts of information content and likelihood in parameter calibration for hydrological simulation models. *J. Hydrol. Eng.* **2014**, *20*, A4014010. [[CrossRef](#)]
9. Liu, Y.; Weerts, A.; Clark, M.; Hendricks Franssen, H.-J.; Kumar, S.; Moradkhani, H.; Seo, D.-J.; Schwanenberg, D.; Smith, P.; Van Dijk, A. Advancing data assimilation in operational hydrologic forecasting: Progresses, challenges, and emerging opportunities. *Hydrol. Earth Syst. Sci.* **2012**, *16*, 3863–3887. [[CrossRef](#)]
10. Griessinger, N.; Seibert, J.; Magnusson, J.; Jonas, T. Assessing the benefit of snow data assimilation for runoff modeling in Alpine catchments. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 3895–3905. [[CrossRef](#)]
11. Leach, J.M.; Kornelsen, K.C.; Coulibaly, P. Assimilation of near-real time data products into models of an urban basin. *J. Hydrol.* **2018**, *563*, 51–64. [[CrossRef](#)]
12. Ghil, M.; Malanotte-Rizzoli, P. Data assimilation in meteorology and oceanography. *Adv. Geophys.* **1991**, *33*, 141–266.
13. Bannister, R. A review of operational methods of variational and ensemble-variational data assimilation. *Q. J. R. Meteorol. Soc.* **2017**, *143*, 607–633. [[CrossRef](#)]
14. Evensen, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res. Oceans* **1994**, *99*, 10143–10162. [[CrossRef](#)]
15. Gordon, N.J.; Salmond, D.J.; Smith, A.F. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *Proc. IEE Proc. F Radar Signal Process.* **1993**, *140*, 107–113. [[CrossRef](#)]
16. Van Leeuwen, P.J.; Evensen, G. Data assimilation and inverse methods in terms of a probabilistic formulation. *Mon. Weather Rev.* **1996**, *124*, 2898–2913. [[CrossRef](#)]
17. Margulis, S.A.; Giroto, M.; Cortés, G.; Durand, M. A particle batch smoother approach to snow water equivalent estimation. *J. Hydrometeorol.* **2015**, *16*, 1752–1772. [[CrossRef](#)]
18. Bailey, R.; Baù, D. Ensemble smoother assimilation of hydraulic head and return flow data to estimate hydraulic conductivity distribution. *Water Resour. Res.* **2010**, *46*. [[CrossRef](#)]
19. Emerick, A.A.; Reynolds, A.C. Ensemble smoother with multiple data assimilation. *Comput. Geosci.* **2013**, *55*, 3–15. [[CrossRef](#)]
20. Clark, M.P.; Slater, A.G.; Barrett, A.P.; Hay, L.E.; McCabe, G.J.; Rajagopalan, B.; Leavesley, G.H. Assimilation of snow covered area information into hydrologic and land-surface models. *Adv. Water Resour.* **2006**, *29*, 1209–1221. [[CrossRef](#)]
21. Parajka, J.; Blöschl, G. Validation of MODIS snow cover images over Austria. *Hydrol. Earth Syst. Sci. Discuss.* **2006**, *3*, 1569–1601. [[CrossRef](#)]
22. Thirel, G.; Salamon, P.; Burek, P.; Kalas, M. Assimilation of MODIS snow cover area data in a distributed hydrological model using the particle filter. *Remote Sens.* **2013**, *5*, 5825–5850. [[CrossRef](#)]
23. Rodell, M.; Houser, P. Updating a land surface model with MODIS-derived snow cover. *J. Hydrometeorol.* **2004**, *5*, 1064–1075. [[CrossRef](#)]
24. Andreadis, K.M.; Lettenmaier, D.P. Assimilating remotely sensed snow observations into a macroscale hydrology model. *Adv. Water Resour.* **2006**, *29*, 872–886. [[CrossRef](#)]
25. Su, H.; Yang, Z.L.; Niu, G.Y.; Dickinson, R.E. Enhancing the estimation of continental-scale snow water equivalent by assimilating MODIS snow cover with the ensemble Kalman filter. *J. Geophys. Res. Atmos.* **2008**, *113*. [[CrossRef](#)]
26. Piazzzi, G.; Thirel, G.; Campo, L.; Gabellani, S. A particle filter scheme for multivariate data assimilation into a point-scale snowpack model in an Alpine environment. *Cryosphere* **2018**, *12*, 2287–2306. [[CrossRef](#)]
27. Van Leeuwen, P.J. Particle filtering in geophysical systems. *Mon. Weather Rev.* **2009**, *137*, 4089–4114. [[CrossRef](#)]
28. Aalstad, K.; Westermann, S.; Schuler, T.V.; Boike, J.; Bertino, L. Ensemble-based assimilation of fractional snow covered area satellite retrievals to estimate snow distribution at a high Arctic site. *Cryosphere* **2018**, *12*, 247–270. [[CrossRef](#)]
29. Beven, K. A manifesto for the equifinality thesis. *J. Hydrol.* **2006**, *320*, 18–36. [[CrossRef](#)]

30. Croke, B.; Wagener, T.; Post, D.; Freer, J.; Littlewood, I. Evaluating the information content of data for uncertainty reduction in hydrological modelling. In Proceedings of the 4th International Congress on Environmental Modelling and Software, Barcelona, Spain, 6–10 July 2008.
31. Wagener, T.; McIntyre, N.; Lees, M.; Wheeler, H.; Gupta, H. Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. *Hydrol. Process.* **2003**, *17*, 455–476. [[CrossRef](#)]
32. Gupta, H.V.; Sorooshian, S.; Yapo, P.O. Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resour. Res.* **1998**, *34*, 751–763. [[CrossRef](#)]
33. Narayanan, R.M.; Ponnappan, S.K.; Reichenbach, S.E. Effects of noise on the information content of remote sensing images. *Geocarto Int.* **2003**, *18*, 15–26. [[CrossRef](#)]
34. Gao, Y.; Xie, H.; Lu, N.; Yao, T.; Liang, T. Toward advanced daily cloud-free snow cover and snow water equivalent products from Terra–Aqua MODIS and Aqua AMSR-E measurements. *J. Hydrol.* **2010**, *385*, 23–35. [[CrossRef](#)]
35. Liu, Y.; Peters-Lidard, C.D.; Kumar, S.; Foster, J.L.; Shaw, M.; Tian, Y.; Fall, G.M. Assimilating satellite-based snow depth and snow cover products for improving snow predictions in Alaska. *Adv. Water Resour.* **2013**, *54*, 208–227. [[CrossRef](#)]
36. Killick, R.; Eckley, I. changepoint: An R package for changepoint analysis. *J. Stat. Softw.* **2014**, *58*, 1–19. [[CrossRef](#)]
37. Kendall, M. *Rank Correlation Measures*; Charles Griffin: London, UK, 1975; 202p.
38. Xiong, L.; Jiang, C.; Xu, C.Y.; Yu, K.X.; Guo, S. A framework of change-point detection for multivariate hydrological series. *Water Resour. Res.* **2015**, *51*, 8198–8217. [[CrossRef](#)]
39. Alameddine, I.; Qian, S.S.; Reckhow, K.H. A Bayesian changepoint–threshold model to examine the effect of TMDL implementation on the flow–nitrogen concentration relationship in the Neuse River basin. *Water Res.* **2011**, *45*, 51–62. [[CrossRef](#)]
40. Kane, D.L.; Hinzman, L.D.; Zarling, J.P. Thermal response of the active layer to climatic warming in a permafrost environment. *Cold Reg. Sci. Technol.* **1991**, *19*, 111–122. [[CrossRef](#)]
41. Liston, G.E. Representing subgrid snow cover heterogeneities in regional and global models. *J. Clim.* **2004**, *17*, 1381–1397. [[CrossRef](#)]
42. Magand, C.; Ducharne, A.; Le Moine, N.; Gascoin, S. Introducing hysteresis in snow depletion curves to improve the water budget of a land surface model in an Alpine catchment. *J. Hydrometeorol.* **2014**, *15*, 631–649. [[CrossRef](#)]
43. Burkhart, J.; Helset, S.; Abdella, Y.; Lappegard, G. Operational Research: Evaluating Multimodel Implementations for 24/7 Runtime Environments. In Proceedings of the AGU Fall Meeting, San Francisco, CA, USA, 12–16 December 2016.
44. Teweldebrhan, A.T.; Burkhart, J.F.; Schuler, T.V. Parameter uncertainty analysis for an operational hydrological model using residual-based and limits of acceptability approaches. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 5021–5039. [[CrossRef](#)]
45. Priestley, C.; Taylor, R. On the assessment of surface heat flux and evaporation using large-scale parameters. *Mon. Weather Rev.* **1972**, *100*, 81–92. [[CrossRef](#)]
46. Lambert, A. Catchment models based on ISO-functions. *J. Inst. Water Eng.* **1972**, *26*, 413–422.
47. Kirchner, J.W. Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward. *Water Resour. Res.* **2009**, *45*. [[CrossRef](#)]
48. Matt, F.N.; Burkhart, J.F.; Pietikäinen, J.-P. Modelling hydrologic impacts of light absorbing aerosol deposition on snow at the catchment scale. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 179–201. [[CrossRef](#)]
49. DeWalle, D.R.; Rango, A. *Principles of Snow Hydrology*; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2008.
50. Hegdahl, T.J.; Tallaksen, L.M.; Engeland, K.; Burkhart, J.F.; Xu, C.-Y. Discharge sensitivity to snowmelt parameterization: a case study for Upper Beas basin in Himachal Pradesh, India. *Hydrol. Res.* **2016**, *47*, 683–700. [[CrossRef](#)]
51. Liston, G.E. Interrelationships among snow distribution, snowmelt, and snow cover depletion: Implications for atmospheric, hydrologic, and ecologic modeling. *J. Appl. Meteorol.* **1999**, *38*, 1474–1487. [[CrossRef](#)]
52. Kolberg, S.A.; Gottschalk, L. Updating of snow depletion curve with remote sensing data. *Hydrol. Process.* **2006**, *20*, 2363–2380. [[CrossRef](#)]

53. Kolberg, S.; Gottschalk, L. Interannual stability of grid cell snow depletion curves as estimated from MODIS images. *Water Resour. Res.* **2010**, *46*. [[CrossRef](#)]
54. Statkraft Information Page. Available online: <https://www.statkraft.com/> (accessed on 20 June 2018).
55. Dee, D.P.; Uppala, S.M.; Simmons, A.; Berrisford, P.; Poli, P.; Kobayashi, S.; Andrae, U.; Balmaseda, M.; Balsamo, G.; Bauer, D.P. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **2011**, *137*, 553–597. [[CrossRef](#)]
56. Copernicus Land Monitoring Service. Available online: <https://land.copernicus.eu/pan-european/corine-land-cover> (accessed on 29 August 2016).
57. Hall, D.K.; George, A.; Vincent, V. *Updated Daily. MODIS/Terra Snow Cover Daily L3 Global 500 m Grid V005*; National Snow and Ice Data Center: Boulder, CO, USA, 2006.
58. Pu, Z.; Xu, L.; Salomonson, V.V. MODIS/Terra observed seasonal variations of snow cover over the Tibetan Plateau. *Geophys. Res. Lett.* **2007**, *34*. [[CrossRef](#)]
59. Hall, D.K.; Riggs, G.A.; Salomonson, V.V.; Barton, J.; Casey, K.; Chien, J.; DiGirolamo, N.; Klein, A.; Powell, H.; Tait, A. *Algorithm Theoretical Basis Document (ATBD) for the MODIS Snow and Sea Ice-Mapping Algorithms*; NASA/Goddard Space Flight Center: Greenbelt, MD, USA, 2001.
60. Masson, T.; Dumont, M.; Mura, M.D.; Sirguey, P.; Gascoin, S.; Dedieu, J.-P.; Chanussot, J. An Assessment of Existing Methodologies to Retrieve Snow Cover Fraction from MODIS Data. *Remote Sens.* **2018**, *10*, 619.
61. Raleigh, M.; Lundquist, J.; Clark, M. Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework. *Hydrol. Earth Syst. Sci.* **2015**, *19*, 3153–3179. [[CrossRef](#)]
62. Durand, M.; Molotch, N.P.; Margulis, S.A. A Bayesian approach to snow water equivalent reconstruction. *J. Geophys. Res. Atmos.* **2008**, *113*. [[CrossRef](#)]
63. Giroto, M.; Cortés, G.; Margulis, S.A.; Durand, M. Examining spatial and temporal variability in snow water equivalent using a 27 year reanalysis: Kern River watershed, Sierra Nevada. *Water Resour. Res.* **2014**, *50*, 6713–6734. [[CrossRef](#)]
64. Ross, T.J. *Fuzzy Logic with Engineering Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2009.
65. Singh, V.P. Hydrologic synthesis using entropy theory. *J. Hydrol. Eng.* **2011**, *16*, 421–433. [[CrossRef](#)]
66. Alippi, C.; Boracchi, G.; Roveri, M. Ensembles of change-point methods to estimate the change point in residual sequences. *Soft Comput.* **2013**, *17*, 1971–1981. [[CrossRef](#)]
67. Bellei, C.; Change-point Detection. Part I—A Frequentist Approach. Available online: <http://www.claudiobellei.com/2016/11/15/changepoint-frequentist/> (accessed on 1 June 2018).
68. Taylor, W.A. Change-Point Analysis: A Powerful New Tool for Detecting Changes. 2000. Available online: <http://www.variation.com/cpa/tech/changepoint.html> (accessed on 7 May 2018).
69. Zaitchik, B.F.; Rodell, M. Forward-looking assimilation of MODIS-derived snow-covered area into a land surface model. *J. Hydrometeorol.* **2009**, *10*, 130–148. [[CrossRef](#)]
70. Toure, A.M.; Reichle, R.H.; Forman, B.A.; Getirana, A.; De Lannoy, G.J. Assimilation of MODIS Snow Cover Fraction Observations into the NASA Catchment Land Surface Model. *Remote Sens.* **2018**, *10*, 316. [[CrossRef](#)] [[PubMed](#)]
71. Bouilloud, L.; Chancibault, K.; Vincendon, B.; Ducrocq, V.; Habets, F.; Saulnier, G.-M.; Anquetin, S.; Martin, E.; Noilhan, J. Coupling the ISBA land surface model and the TOPMODEL hydrological model for Mediterranean flash-flood forecasting: Description, calibration, and validation. *J. Hydrometeorol.* **2010**, *11*, 315–333. [[CrossRef](#)]
72. Shi, Y.; Davis, K.J.; Duffy, C.J.; Yu, X. Development of a coupled land surface hydrologic model and evaluation at a critical zone observatory. *J. Hydrometeorol.* **2013**, *14*, 1401–1420. [[CrossRef](#)]



Appendices

Summary of algorithms implemented in this work

Appendix A: Implementation of the algorithms employed in Paper I

Although an existing uncertainty analysis framework, i.e. the SAFE toolbox was employed to conduct the core analyses in this subproject, a number of algorithms were also implemented based on this framework or as independent scripts. Unless specified, the algorithms employed in this particular work are implemented using Matlab and they are available at: <https://github.uio.no/aynomtt/UA>

Parameter sampling and Monte Carlo simulation

GLUE_shyft_param_sampling: this algorithm is used to coordinate parameter sampling. Details about setup of the sampling technique such as the number of samples, parameters that are allowed to vary, and definition of their minimum and maximum range is specified in this script. It prepares the inputs and calls an existing algorithm that samples parameter values using the All-At-a-Time (AAT) sampling technique. The implemented algorithm retrieves the sampled parameter values and exports the result as a netcdf file.

run_MC: this python algorithm is used to run a Monte Carlo simulation using the PT_GS_K model and the sampled parameter sets. The simulation results, i.e. streamflow and fractional snow cover area are exported to netcdf files for use in further analysis.

Uncertainty analysis

The main scripts written to carry out the different uncertainty analysis related tasks are outlined below.

GLUE_shyft_QandSC_param_unc: this script is mainly intended to conduct uncertainty analysis based on streamflow as observational dataset. It generates different plots and summary statistics such as mean, median, variance, and skewness of the behavioural parameter sets.

GLUE_shyft_Q: used to conduct the residual based uncertainty analysis based on streamflow only. Outputs from this script include the cross-validation result of streamflow predictions against observed values both for the calibration and validation years. It also generates different plots including the observed and simulated hydrographs for each year of analysis.

GLUE_shyft_snow: used to conduct the residual based uncertainty analysis based on fractional snow cover area as observational dataset. Outputs from this script include the cross-validation result of fractional snow cover area (fSCA) predictions against observed values (MODIS fSCA) both for the calibration and validation years.

GLUE_shyft_Q_and_snow: carries out the residual based uncertainty analysis using a combined likelihood of streamflow and fSCA as observational datasets. Outputs from this script include the cross-validation plots and tabular result of the evaluation metrics for streamflow and fSCA both for the calibration and validation years.

GLUE_LOA_shyft_Q_and_snow: used to conduct uncertainty analysis based on the limits of acceptability (LoA) concept and a combined likelihood of streamflow and fSCA. This algorithm modifies the original formulation of GLUE LoA. Here model realizations that satisfy the LoA criteria above certain percentage of the observation time steps (pLoA) are considered behavioural. Outputs from this script include the cross-validation plots and tabular result of the evaluation metrics for streamflow and fSCA both for the calibration and validation years.

GLUE_LOA_shyft_Q_and_snow_CR_pLoA_analysis: this script is used to assess the effect of pLoA on modelling uncertainty (CR) and NSE of the median streamflow prediction. Plots displaying CR and NSE against pLoA values are generated for the sample analysis years.

Appendix B: Implementation of the algorithms employed in Paper II

In this subproject, the R statistical software and its CART package were used for general statistical analysis and in building the machine learning models. This section presents summary of the main algorithms implemented in this subproject. Unless specified, the algorithms are implemented using R and they are available at:

<https://github.uio.no/aynomtt/ML>

ML_idx_score_pLoA_nea: this matlab script is used to map the model realizations to their corresponding efficiency metrics, i.e. pLoA and Score through comparison of the simulated against observed streamflow values. The index of model realizations, pLoA, and Score are stored for use in training and testing phases of the machine learning methods implemented in the following algorithm.

ML_emulators_rf_knn_nnet: used to build the three machine learning models, i.e. random forest, k-nearest neighbours, and artificial neural network as emulators of the Monte Carlo simulation. The machine learning models (MLMs) are deployed for use in prediction and subsequent analyses.

ML_emulator_hyd_model_valid: this algorithm is used to identify behavioural model realizations based on the time relaxed GLUE LoA concept and the pLoA and score values predicted using the MLMs. The deployed MLMs are also tested and validated in this algorithm. The evaluation metrics both for the MLMs and the hydrological model are computed and exported to a text file. This algorithm also generates plots displaying the evaluation result.

ML_emulator_posterior_param_statDistn: main task of this algorithm is to compute the statistics (e.g. mean, median, variance and skewness) of the posterior distribution for model parameters identified using the coupled MLMs and MC simulation. It also calculates similar statistics for those directly identified from the MC simulation. The summary statistics is exported to a text file.

ML_emulator_Variable_Imp_and_correlation: used to analyse variable importance, i.e. streamflow sensitivity to model parameters. It is also used to compute the correlation between the model parameters. Outputs from this algorithm include graphs that display relative variable importance and Pearson correlation matrix.

ML_emulator_percentileQ_plots_of_calib_vs_validn_periods: this algorithm is used to analyse quantiles of observed streamflow values for the calibration and validation periods.

Appendix C: Implementation of the algorithms employed in Paper III

Unless specified, the algorithms employed in this particular work are implemented using python. The scripts are available at: <https://github.uio.no/aynomtt/DA>

Estimating global parameter values

The global parameter values for the PT_GS_K model are estimated through inverse modelling using streamflow as an observational dataset. The directory paths to the meteorological and physiographic dataset files as well as information related to setup of the model are specified in relevant ‘yaml’ files accompanying the model. The generated calibration file containing the global parameter values is saved for use in subsequent model runs to estimate the prior values. An existing template script accompanying the PT_GS_K model is used for running the model.

Snow water equivalent (SWE) reanalysis

The estimated prior SWE values are reanalysed using four data assimilation schemes. The perturbation parameters are constrained using fractional snow cover area (fSCA) as observational dataset. The following scripts are implemented to accomplish different tasks including data preparation, reanalysis, and post-processing.

DA_main: This algorithm is responsible to organize the main tasks. Basic setup of the SWE reanalysis schemes such as value of the perturbation parameters, the reanalysis period, as well as fSCA and SWE output file names are specified in this script. The reanalysis routine is repeated for a user specified number of random iterations and number of reanalysis years. The perturbation parameters are generated for the specified number of ensemble members. The model iteratively runs using these values. After each model run, the simulated prior fSCA and SWE values are retrieved and stored. Following that, the methods within the DA_analysis algorithm such as the one responsible to prepare the fSCA observations as well as the data assimilation methods with and without the fuzzy-logic concept are invoked in this algorithm. The reanalysed values are retrieved and stored for use in further analysis.

DA_analysis: This algorithm contains methods that carry out specific tasks. These methods include:

get_obs_scf_swe: used to extract the observed fSCA values from a netcdf file for selected number of grid cells (33), reanalysis period, and year of analysis. The netcdf files contain fSCA data for the whole Nea-catchment, i.e. 812 grid cells.

pbs_analysis: this method is used to carry out SWE reanalysis tasks based on the different data assimilation schemes, i.e. LoA, LoA_F, PBS and PBS_F. The data assimilation schemes with the fuzzy coefficient, i.e. LoA_F and PBS_F require the identification of critical periods in the time axis. Hence, an algorithm responsible to detect the change points (*scf_cpt_detector_param_and_nonParam*) is also called by this method. It also invokes other methods implemented in the same class that are written to accomplish specific tasks such as the *get_wt_trimf*, *get_type_2_mf*, *get_perc_days_mf*, and the *weighted_quantile* method.

eval_assimilation: this method is used to evaluate the simulated prior and reanalysed SWE and fSCA values through comparison against their respective observed values. The evaluation metrics calculated in this method include: the mean absolute bias (MAB), Pearson correlation (r), and root mean squared error (RMSE).

Other algorithms implemented in this subproject include:

scf_cpt_detector_param_and_nonParam: This algorithm estimates the change point positions using parametric and non-parametric approaches. The parametric approach was adopted from the script by Bellei (2018), while the non-parametric approach, i.e. CUMSUM was newly implemented based on the paper by Taylor (2000). This algorithm has methods that are meant to carry out the change point detection and calculation of the confidence interval associated with the detected change points.

DA_sensitivity_and_fSCA_distn: this algorithm is used to conduct sensitivity analysis of the efficiency metrics, i.e. RMSE, r, and MAB in response to change in position of the critical points. The critical points are the point where the mean snow cover fraction changes and the point where the melt-out period starts. The algorithm shifts position of the critical points both forward and backward in the time axis with subsequent analysis of its impact on the evaluation metrics as compared to the originally detected change points. Outputs from this algorithm include the sensitivity analysis graphs as well as the conceptualization graph displaying value of the fuzzy coefficient in relation to dynamics of fSCA in the ablation period.

DA_TS_analysis_and_plot: used to analyse the prior and reanalysed SWE estimates and fSCA dynamics both in space and time. It displays the output using relevant plots.

DA_CP_param_vs_nonparam_plots_and_statcal_summary: used to analyse and display the change points for all grid cells as detected using the parametric and non-parametric approaches. This algorithm also generates annual statistical summary (e.g. mean, median, variance, and skewness) of the detected change points based on grid cell values.

DA_plots: used to display different plots based on the analysis on the prior and reanalysed SWE and fSCA values.

extract_transect_sd_swe: this matlab script is used to extract SWE values, their geographical location, and observation date from the radar measurement files. It masks the data points based on geographical location of the model grid cells. It also displays location of the observed values on map of the Nea-catchment (mainly as a check on whether the data points are correctly extracted), and exports the extracted data to a mat file.