UNIVERSITY OF TWENTE.

Tracking the uncertainty in streamflow prediction through a hydrological forecasting system

August 2011

Trang Van Pham

Internal supervisors:
University of Twente
Dr. M.S. Krol
Associate Professor | m.s.krol@ctw.utwente.nl

Dr.Ir. M.J. Booij
Assistant Professor | m.j.booij@ctw.utwente.nl

University of Twente
Faculty of Engineering Technology
Department of Water Engineering and Management
P.O. Box 217
7500AE Enschede
The Netherlands

External supervisor:
Cemagref
Dr. M.H. Ramos
Researcher | maria-helena.ramos@cemagref.fr

Cemagref Antony
Research Unit HBAN
1, rue Pierre-Gilles de Gennes CS 10030
92761 Antony Cedex
France

Summary

A flood forecasting system is a complex system which consists of many different components and each of these components can contain, to some extent, an uncertainty. Studying the uncertainties in flood forecasting, quantifying and propagating them through the system can help to gain more information about the different sources of uncertainty that may affect the forecasts. This information can later be added to the forecasts to improve their quality. These issues bring several challenges to the study of flow forecasting uncertainty: firstly, what is the impact of different sources of uncertainty on the quality of flood? Secondly, forecasts among all sources of uncertainty that stem from different components of the system, which sources significantly affect flood forecasts? Thirdly, which methods can be used to efficiently quantify and propagate those uncertainties through a forecasting model? Finally, which measures should be used to evaluate the uncertainty quantification and their impact on the quality of the forecasts? The aims of this research is to quantify and propagate the different kinds of uncertainty sources which play a role in flood forecasting; and to investigate methods to assess the quantified uncertainties and proper measures to evaluate the uncertainty quantification.

In this research, the GRPE forecasting system, an ensemble prediction system based on the lumped GRP hydrologic model, is applied to three catchments in France. The uncertainties from precipitation data (input precipitation which is used for flow simulation and forecast precipitation used for flow forecasting), hydrological initial conditions (discharge data) and model parameters, which are acknowledged as important sources of uncertainty in hydrological modelling and forecasting, are studied. They are individually quantified and then propagated together through the forecasting system with an experimental approach by multiplying the simulations. The model structure uncertainty is not considered in the scope of this research.

Methods for uncertainty quantification are defined and applied to each source of uncertainty. Two ensemble prediction systems from ECMWF and Météo-France are used to account for the forecast precipitation uncertainty for lead times from 1 to 9 days. For the uncertainty of input precipitation data, geo-statistical simulations of spatially averaged rainfall, conditioned on point data, and available for one study catchment, are chosen to provide the multiple statistical realizations of daily spatial rainfall fields over the study area. The hydrologic initial condition uncertainty is quantified by using an ensemble of discharges to update the state of the routing reservoir of the forecasting model. These discharges are retrieved from the analysis of uncertainties affecting the rating curves of each study catchment. Ten different periods of data, with the length of 5 years each, are selected to calibrate the model and thus to account for calibration period uncertainty. Finally, the Generalized Likelihood Uncertainty Estimator (GLUE) method is alternatively used to quantify the parameterization uncertainty. This is done by taking a large number of 125.000 sets of parameters to find the confidence intervals. To assess the results of uncertainty quantification, two probabilistic evaluation measures, the Brier (Skill) Score and the reliability diagram, are employed. In addition, confidence intervals of the forecasts are used to visualize the outcome of the research.

The results show that input precipitation uncertainty does not have noticeable impact on the forecast output. This may be due to the method used to quantify the uncertainties from this source, which may be inappropriate to correctly capture them. For the catchments studied, this source of uncertainty can, therefore, be neglected when propagating different sources of uncertainty through the system. The other sources of uncertainty show large impacts on flow forecasts. Initial condition uncertainty shows large impacts for small lead times (up to 2 days). After that, forecast precipitation uncertainty has the largest impact; this impact is more significantly pronounced at high lead times. Depending on the catchment, parameter uncertainty can have more impact if it is evaluated from the variation of the calibration period or from the GLUE method.

Based on the results of this research, and on the catchments and methods investigated, it is recommended to take into consideration the uncertainty of forecast precipitation, initial condition and model parameters in flood forecasting. There are different ways to account for parameter uncertainty, but the proposed approach of using different calibration periods proved to be a simple method but able to improve the quality of the forecast outputs.

Preface

This thesis report is the final product of my master programme on Civil Engineering and Management with the specialisation on Water Engineering and Management at University of Twente, the Netherlands. This was done partially in University of Twente and partially in Cemagref Antony, France.

I would like to express my appreciation to my supervisors Dr. Martijn Booij and Dr. Maarten Krol, who gave me very helpful advice during the research time. Special thanks go to Dr. Maria-Helena Ramos, my supervisor in Cemagref, thank you for your great help academically and personally during the time I worked in France.

I would like to also acknowledge here the work of Benjamin Renard and Etienne Leblois from Cemagref Lyon, France on rating curve and spatial average precipitation uncertainty quantification. I acknowledge Météo-France for the SAFRAN data archive and the PEARP ensemble forecasts, European Centre of Medium Range Weather Forecasting for the ensemble prediction system forecasts and the French MEEDM (Ministère de l'Ecologie, de l'Energie, du Développement durable et de la Mer) for the streamflow data.

Thanks to all of my colleagues in Cemagref, you were very friendly and helped me a lot during my four months stay there. Special thanks go to my Vietnamese dear friends in Twente, many of them are here, many of them are not here anymore, but they always support me, give me help when I need without any hesitation. Without you all, it would be impossible for me to have such a great and comfortable time here.

I would like to express my love and appreciation to my parents for their love and support and for the fact that they gave me this great opportunity in my life.

Last but not least, I want to thank my boy friend, Ali Riza Konuk, thank you so much for your incredible support and love. It was a big motivation for me to complete this research.

Enschede, August 2011

Trang Van Pham

Table of Contents

1 Introduction

1.1 Motivation

Flood is a major natural disaster in many countries all over the world. The consequences of flooding can be enormous in term of property loss and fatalities. According to the European Commission (2011), in Europe, there were more than 100 major damaging floods only in a period of three years between 1998 and 2002, which represents, on average, more than 30 floods per year. Among those, there was the catastrophic flood along the Danube and Elbe rivers in 2002; which was a 100 year-flood that caused billions of Euros of damage in many countries in Eastern Europe (BBC, 2002). Since 1998, floods have caused some 700 fatalities, the displacement of about half a million people and at least 25 billion Euros in insured economic losses in Europe (European Commission, 2011).

Flood forecasting is one of the solutions to prevent the consequences of flooding, as it can provide information on whether a potential flood might happen in the long or short term future. Based on the flood forecasts issued, flood warning, prevention and evacuation solutions can be implemented. According to the American Meteorological Society Glossary (American Meteorological Society, 2011), "flood forecasting is the use of real-time precipitation and streamflow data in rainfall-runoff and streamflow routing models to forecast flow rates and water levels for periods ranging from a few hours to days ahead, depending on the size of the watershed or river basin".

However, telling something that might happen in the future is not an easy task as the future itself is anyway something we cannot know at this moment and hence uncertain. Lack of knowledge about the physical processes involved in flooding and the data used, as well as inherent unpredictability of severe events can lead to uncertainties in the forecasts. Errors in the forecasting of flood may lead to (Smith and Ward, 1998):

- under-preparation and therefore to otherwise avoidable damage (if the forecast stage is too low and/or the forecast timing of inundation is late), or
- over-preparation, unnecessary expense and anxiety, and to a subsequent loss of credibility (if the forecast stage is too high and/or the forecast timing of inundation is premature).

A typical example of how not taking into account the errors in flood prediction can lead to severe consequences is reported by Krzysztofowicz (2001) for the flood event of Spring 1997 on the Red River in the Grand Forks, North Dakota, USA. The estimated 15 meter flood crest, with no uncertainty bounds associated to this value, led city officials and residents to prepare for the future event as if this estimate were a perfect forecast. However, the forecast actually seriously underestimated the event and the actual flood crest was of 16.5 meters, overtopping the dikes, and inundating 80% of the city which forced almost the total evacuation of its citizens.

Therefore, it can be seen that assessing uncertainties in flood forecasting is a very crucial issue; especially in a well-developed world nowadays where a great number of sophisticated infrastructure systems were built or are being built, which makes people much more vulnerable to natural disasters. Understanding the uncertainties that exist in the flood forecasting chain will help people to be aware of them and be better prepared for the future events. Furthermore, uncertainty assessment also helps scientists to find innovative solutions to quantify and reduce those uncertainties. Besides the forecast quality is affected by the lack of knowledge on uncertainty, and therefore, the quantification of uncertainty also contributes to adding more information/knowledge into the flow forecasting procedure to improve the quality of the forecasts. Forecasts can thus become more reliable in terms of probability, or/and more accurate in terms of magnitude.

In the coming section, a general overview of uncertainty analysis in streamflow prediction is presented. The focus is on the identification, quantification, propagation and evaluation of flow forecasting uncertainty.

1.2 Uncertainty analysis in flood forecasting

There are different sources of uncertainty that might affect flow forecasting. In the context of flood forecasting, Maskey et al. (2004) classified the sources of uncertainty as model uncertainty, input uncertainty, model parameter uncertainty, natural and operational uncertainty. Except the last source of uncertainty, which originates from the nature and the operation and is not related to the forecasting model, others come from the different components of the flow forecasting model as illustrated in Figure 1. The input data for a flow forecasting model have two kinds, input for flow simulation and forecasting; uncertainty of the input data can arise from both of these sources. Other source of uncertainty that is not mentioned by Maskey et al. (2004) is initial condition uncertainty which comes from the inappropriate estimation of the initial state for flow forecasting. All of these sources of uncertainty can propagate into the forecast outcome and cause the uncertainty of the flow forecasting. In addition, the forecasts are often verified against a reference which is often taken as the observed discharge. However, as the measurement of discharge is not error free, uncertainty can also stem from this component.

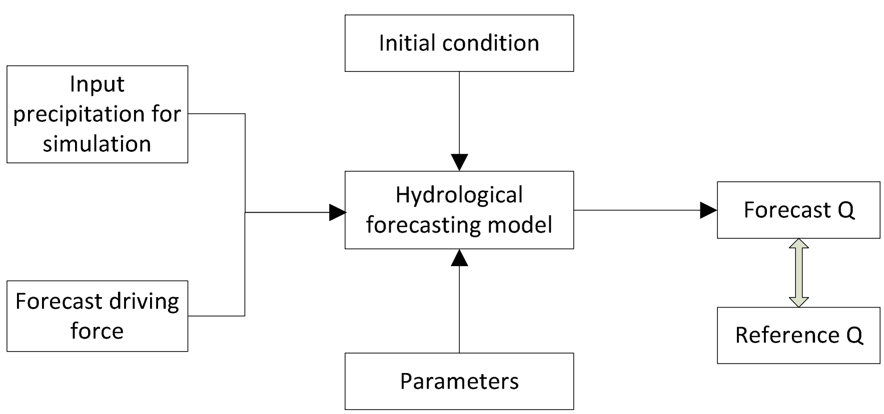


Figure 1: Main components of a typical hydrological forecasting chain

1.2.1 Quantification of uncertainties in flow forecasting

Many studies can be found in the literature that study one or some of the uncertainty sources mentioned above. Most of them show that taking into consideration the uncertainties would improve the reliability of the forecast.

Using the ensemble prediction systems (EPS) is a popular approach to assess the uncertainty in flow forecasting due to precipitation forecast uncertainty. Many published literature, which have used EPS, are listed in Cloke and Pappenberger (2009). The authors also showed the attractiveness and potential of using ensemble prediction systems (EPS) to account for uncertainty from the forecast of precipitation in flow forecasting. Ensemble forecasts of precipitation take into account the uncertainties in the atmospheric state and initial conditions, as well as the limitation of the representation of the physical process in weather forecasting. As a result, a set of possible future states of the atmosphere is provided. This uncertainty can be then propagated through flood forecasting system to produce an ensemble prediction of flow as shown in Figure 2. By reviewing published literature and based on their results, the authors demonstrated that using the precipitation ensembles in flood prediction can increase the capability of issuing the successful flood warnings. Moreover, the ensemble predictions help to add additional useful information to the deterministic forecast which is the best estimation of the future event.

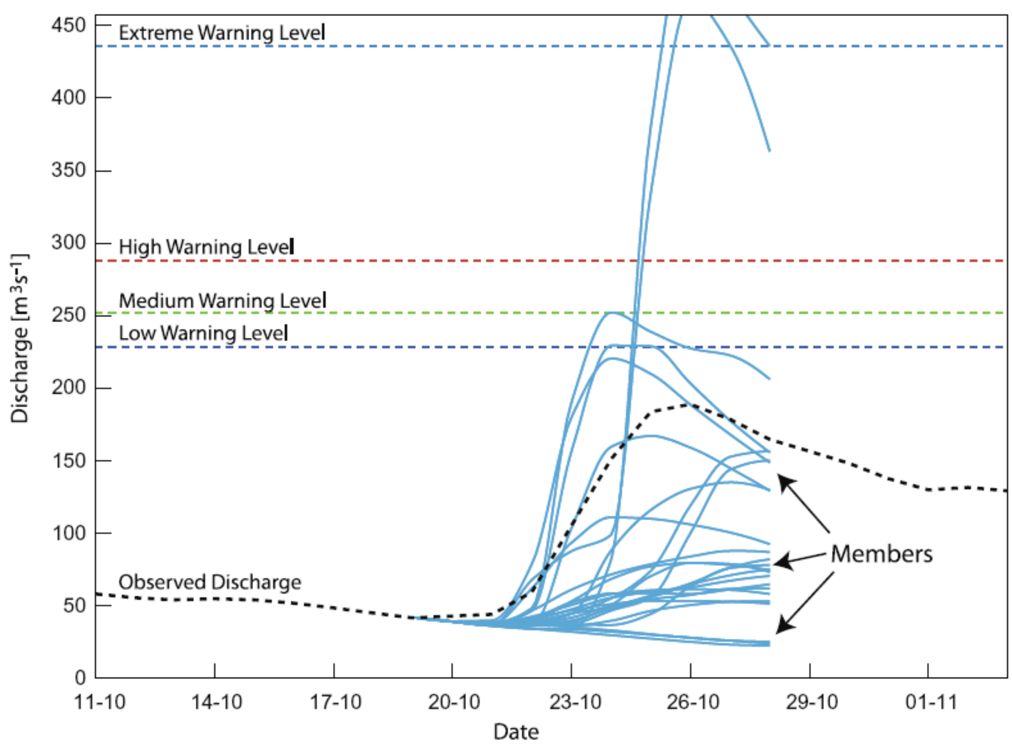


Figure 2: Ensemble hydrographs for a flood event predicted for each ensemble forecast of precipitation (Cloke and Pappenberger, 2009)

Pappenberger et al. (2005) cascaded the uncertainty of rainfall from medium range weather forecasts into flow forecasts using the ensemble prediction system from the European Centre of Medium Range Weather Forecasting (ECMWF). The authors also concluded that the errors in rainfall forecasts had a strong influence on the forecast flows.

He et al. (2009) used not only one, but multiple ensemble weather predictions from various weather centres to cascade the uncertainty of precipitation forecast through a flow forecasting model. The best six sets of model parameters were selected, to account for parameter uncertainty, and combined with 216 forecasts members to form 6*216 ensemble forecast discharges. The paper shows that forecast precipitation uncertainty dominates and propagates through the forecast chain, implying the importance of accounting for this precipitation uncertainty in flow forecasting.

Mascaro et al. (2010) also admitted the important contribution of evaluating the propagation of errors associated with ensemble precipitation forecasts into the ensemble streamflow to uncertainty reduction in flow forecasting. In their paper, the authors generated and verified three different sets of ensemble streamflow forecasts by using three ensembles of precipitation forecasts.

Concerning uncertainty from input data, in Renard et al. (2010), the authors stated that the traditional calibration methods in hydrological modelling assume all observed inputs are error free, and therefore there is no input uncertainty which is obviously not realistic. For that reason, in their papers, the Bayesian total error analysis (BATEA) framework was used to assess the uncertainty of input data. The outcome of their work proves that ignoring input uncertainty can significantly degrade the inference of flow prediction, and hence, this source of uncertainty should be considered in hydrologic prediction.

Berthet et al. (2009) focused on investigating the impact of soil moisture initial conditions on the performance of flood forecasting models and finding the level of importance of accounting for the initial condition uncertainty in flood forecasting. They applied different methods of forecast initialization (updating) and compared the prediction errors obtained from those tests to find the best approach to deal with the initial condition in a flood forecasting model. Their study shows that different methods of initialization have different impacts on the forecast outcome and there is an optimum mode for initializing. Because of the large and varied data that were used, the authors stated that the results were probably not catchment-dependent. It is also suggested in the paper that the same results achieved for the model studied could be found for many other forecasting models.

Many studies on hydrological simulation and prediction quantify parameter uncertainty using the Generalized Likelihood Uncertainty Estimator (GLUE) proposed by Beven and Binley (1992) (see Pappenberger et al. (2004), Pappenberger et al. (2005), Larsbo and Jarvis (2005), Xiong and O'Connor (2008), Jin et al. (2010), etc.). The main idea of the GLUE method is rejecting the concept of an optimum model and parameter set, and assuming that prior to input of data into a model, all model structures and parameter sets have an equal likelihood of being acceptable. By using a number of parameter sets or model structure, series of equally likely simulations can be obtained. After the data for a particular case being considered, the model structures or parameter sets can be attributed as non-behavioural or behavioural depending on the likelihood threshold that is defined by the user. Confidence intervals can be defined for the flow simulations or forecasts using the cumulative distribution function of discharge weighted by the likelihood value. Larsbo and Jarvis (2005) recommended GLUE a suitable method of parameter uncertainty estimation for prediction. Pappenberger et al. (2004) did a study in the Meuse river basin in the Netherlands and Belgium using the GLUE method. A large number of 6632 runs were performed but none of the simulations was found as behavioural. It is concluded in the paper that "uncertainty analysis should be compulsory for every model exercise even when the changes in the modelling domain are considered as controllable or minor".

Not exactly on forecasting, Carpenter and Georgakakos (2004) studied the impacts of parametric uncertainty on ensemble streamflow simulations. Ensemble flow simulations were obtained by introducing perturbations in model parameters through random sampling from prescribed probability distributions within a Monte Carlo simulation framework. The results show a significant reduction in simulation uncertainty when considering the discharge ensembles from a number of perturbed parameters.

Uhlenbrook et al. (1999) estimated the prediction uncertainty of a rainfall-runoff model caused by limitations in the identification of model parameters and structure. Different parameter sets were randomly generated with a Monte Carlo procedure to account for the parameter uncertainty and several variants of the conceptual rainfall-runoff HBV model were considered. The outcome of their study on parameter and model structure uncertainty comes to a conclusion that uncertainties of the model structure and the model parameters, and their impacts on model predictions have to be considered when applying a conceptual hydrological model. Also concerning the model structure uncertainty, Butt et al. (2009) did their research on a physically based forecasting model. The authors also came to a conclusion that model structural uncertainty should be considered in assessing model uncertainties and the use of a combination of several model structures can be a means of improving hydrological simulations.

Mc Milan et al. (2010) investigated the errors in discharge measurements, used to calibrate a rainfall runoff model, that caused by the rating curve uncertainty. By looking at the errors in individual gauging station and rating curve fits, the authors concluded that considering the uncertainty in discharge data resulted in significant improvement in flow prediction.

Considering the rainfall forecast uncertainty and the discharge observation uncertainty but more focusing on the discharge observation errors, Pappenberger et al. (2009), by using the observations for verifying the forecasts, recognized the effect of uncertainty in observations. The results show the flatten histogram and the reduced number of outliers in the hydrographs due to the effect of observation uncertainty. The authors recommended that this uncertainty coming from observation should be taken into account when evaluating the forecast skill.

In summary, quantifying different uncertainty sources is acknowledged in the literature that may improve the forecast reliability. However, it can also be seen that, normally, the studies focused on one or two uncertainty sources; and the entire flow forecasting model domain, as shown in Figure 1, is not often studied in the literature.

1.2.2 Propagation of uncertainties in flow forecasting

The assessment of uncertainty in a model output requires the propagation of different sources of uncertainty through the modelling system. Different approaches to propagate the uncertainties through a flood forecasting system can be found in the literature.

Hostache et al. (2011) proposed a stochastic method to assess the uncertainty in hydro-meteorological forecasting systems. The focus of the paper is to evaluate the total uncertainty propagated into flood forecasts through a conceptual hydrologic model. But the authors recognized the difficulty in isolating the errors that stem from the individual model components. Therefore, to evaluate the predictive uncertainty of the forecasting system, instead of computing the uncertainties generated by individual model components, their approach focused on the analysis of the statistical properties of the discharge forecast errors. The results of this approach were the confidence limits computed for various lead times of prediction. These confidence limits were then compared with observations of the river discharge. The drawback of this approach is that it cannot differentiate between each individual uncertainty source that arises from each component of the forecasting system. In addition, when comparing the outcome confidence limits of the forecasts with the observations, the observed discharges were assumed to be error free, which might lead to errors in the interpretation of the results.

The Bayesian approach may overcome those limitations by directly addressing both input and output errors in hydrological modelling through the distribution model of the errors of each source. Kavetski et al. (2006) applied the BATEA methodology to two North American catchments in which the precipitation errors showed considerable effects on the forecast hydrographs and the calibrated parameters. The authors concluded that this was a promising approach to deal with uncertainty in hydrological modelling. However, the authors also recognized the shortcoming of the proposed approach, since the error models are often poorly known and it is required further work on the distribution of the input and output errors. Besides, it is also computationally challenging and, technically, an expensive method.

The combination of different uncertainties, each coming from a different component of the forecasting system, can be experimentally propagated through the model by multiplying the simulations, every time a different source of uncertainty is taken into account. For instance, Pappenberger et al. (2005) propagated the uncertainties from forecast precipitation and model parameters by taking 52 ensemble precipitation forecasts and 6 different parameter sets within the modelling framework, resulting in 52*6 simulations of runoff. This approach can explicitly quantify the uncertainties, and makes it possible to assess the impact of each uncertainty source and that of the combination of those uncertainties on the model outcome in a step-by-step propagation analysis.

The entire uncertainty analysis for a system is necessarily done through three steps: uncertainty identification, uncertainty quantification and uncertainty propagation (Magnusson, 1997). First of all, the sources of uncertainty that might play an important role in the study system should be identified, and then those recognized sources need to be quantified and propagated through the system. However, as can be seen in the reviewed literature above, the whole process of uncertainty analysis is not often done. Normally, the studies focused on some specific sources of uncertainty and propagated those uncertainties through the system of interest (Mc Milan et al. (2010), Uhlenbrook et al. (1999), Carpenter and Georgakakos (2004), Mascaro et al. (2010), etc.). This work can result in the impact of some specific sources of uncertainty in hydrological modelling, but the overall uncertainty of the model output cannot be seen. Going through the whole process is, therefore, necessary to understand the total possible uncertainty that might exist in the model output.

1.2.3 Evaluation of forecast uncertainty

The result of propagating the uncertainty through a forecasting model is a set of ensemble forecasts of the model output of interest, then the issue comes in how to evaluate those forecasts. In WMO (2011), three properties of an accurate probabilistic forecast are defined:

- Reliability: the agreement between forecast probability of an event and the mean observed frequency of that event.
- Sharpness: the tendency of forecast probabilities of an event occurring being near 0 or 1. Forecast system that are capable of predicting high probability are said to have sharpness and vice versa.
- Resolution: the ability of the forecast to resolve the set of sample events into subsets with characteristically different outcomes.

To assess these properties, over a long series of pairs of "probabilistic forecasts-observation", several statistical measures have been proposed in the literature, such as Brier (Skill) Score and the Ranked Probability (Skill) Score, as well as the reliability diagram and the rank histogram.

All of those probabilistic measures deal with the probability of forecast and observed frequency but Brier score considers the whole domain of probability; it is calculated for a specific threshold of exceedance by summing all of the square difference between the forecast probabilities and observed frequencies. Reliability Diagram looks also at the distribution of the probability while considering the event with different bins of probability. In a reliability diagram, the forecast probability and observed frequency are plotted as a pair for each probability bin; the assessment is based on the distance of the pair points to the diagonal which represents the perfection of forecast. This will be explained in more detailed later. Ranked Probability (Skill) Score also categorizes the probability to cover all possible outcomes, but it is more sensitive to the distance as the further distance between observation and forecast will be punished more. The rank histogram is used to check if the future state of the predictant is consistent with the distribution of the ensemble.

Evaluation measures have been applied in flood forecasting by several authors, see, for instance, Renner et al. (2009), Thirel et al. (2010), Olsson and Lindstrom (2008), etc. Renner et al. (2009) employed these verification tools in their papers. The results show that different forecast ensembles can be compared and improvements in the forecasting systems can be identified and measured with the help of skill scores and the reliability diagram. In addition, the verification information also provides useful information of the forecasts as it gives an expectation of uncertainty existed. This information can be used effectively in establishing the confidence in a forecast.

1.3 Problem definition

From the previous sections, it can be clearly seen that flood forecasting always contains uncertainties which can originate from many different sources. It is of importance to quantify those uncertainties as a forecast should be issued in company with its uncertainty. Accounting for the uncertainty in flood forecasting can help to improve the quality of the forecasts. However, since the uncertainty in flow forecasts comes from various sources, it is necessary to know which sources of uncertainty play a crucial role in the forecasting system and which do not. By doing this, information can be obtained about which sources of uncertainty should be primarily propagated to the forecast output, since accounting for all kinds of uncertainty might not be feasible and, if uncertainty is wrongly quantified, might lead to an overestimation of the total predictive uncertainty of the output.

The major challenges in uncertainty quantification in flood forecasting can be summarized as following:

- The entire uncertainty analysis process in flood forecasting, including uncertainty identification, quantification and propagation needs to be done.
- The uncertainties should be properly propagated into the forecast output to improve the forecast reliability.
- The impact of different uncertainty sources needs to be evaluated.
- Methods to assess different sources of uncertainty and proper evaluation measures to assess if the uncertainty quantification is correctly described the total predictive uncertainty need to be considered.

1.4 Objective and research questions

The objectives of this research are to identify the sources of uncertainty which may play a significant role in flood forecasting; to quantify and propagate the main sources of uncertainty identified through a flow forecasting system; to evaluate, individually and together, the impact of uncertainty quantification on the forecast outcome.

Based on the results of uncertainty quantification evaluation, this research will suggest the main sources of uncertainty that should be propagated into flood forecasts to improve forecast quality.

The objectives are presented in terms of four research questions:

1. Which sources of uncertainty significantly affect flood forecasts?
2. How to quantify the important uncertainty sources that affect flood forecasts?
3. How to efficiently propagate those uncertainties through a forecasting model?
4. What is the impact of different sources of uncertainty on the quality of flood forecasts?

Three catchments in France, namely Allier (code number K2330810), Ardèche (V5064010), and Arc (Y4122020), are chosen as study areas; details about the study areas are given in Section 2.1. The GRPE forecasting system (Ramos et al., 2008) is used to forecast river discharges. It is an adaptation of the hourly GRP rainfall-runoff model (Tangara, 2005; Berthet et al., 2009) to daily ensemble flow forecasting. The GRP model is a lumped hydrological model developed at Cemagref, France, which is used to forecast river flows in real time on several catchments in France.

1.5 Report outline

In the second chapter of this report, the study materials used in this research are introduced, including study areas, data and the forecasting system. The methods for uncertainty identification, quantification, propagation and evaluation are presented in chapter 3. The results on uncertainty identification from the literature review are shown after the description of uncertainty identification method. The results of uncertainty propagation and evaluation are shown and discussed in chapter 4. Finally, chapter 5 gives the conclusion and recommendation of the research.

2 Study materials

In this chapter, the study areas and their observed hydro-meteorological data are presented. Then the hydrological forecasting system used in this research is described.

2.1 Study areas

Three catchments in France, Allier (K2330810, 2269 km^2), Ardèche (V5064010, 2240 km^2), Arc (Y4122020, 728 km^2) are selected for this study. Figure 3 shows the location of three study catchments in France. These catchments are selected because they have available data necessary for uncertainty quantification performed in this study, such as spatially averaged precipitation and ensemble discharges from rating curve are available in these areas. Besides, these catchments are part of a wider study on uncertainty quantification and propagation conducted at Cemagref. Some characteristics of the catchments are summarized in Table 1. Two catchments Allier and Ardèche, which are medium sized catchments with area of more than 2000 km^2, are located in the mountainous area while Arc, the small catchment with only 728 km^2, is in a lower area.

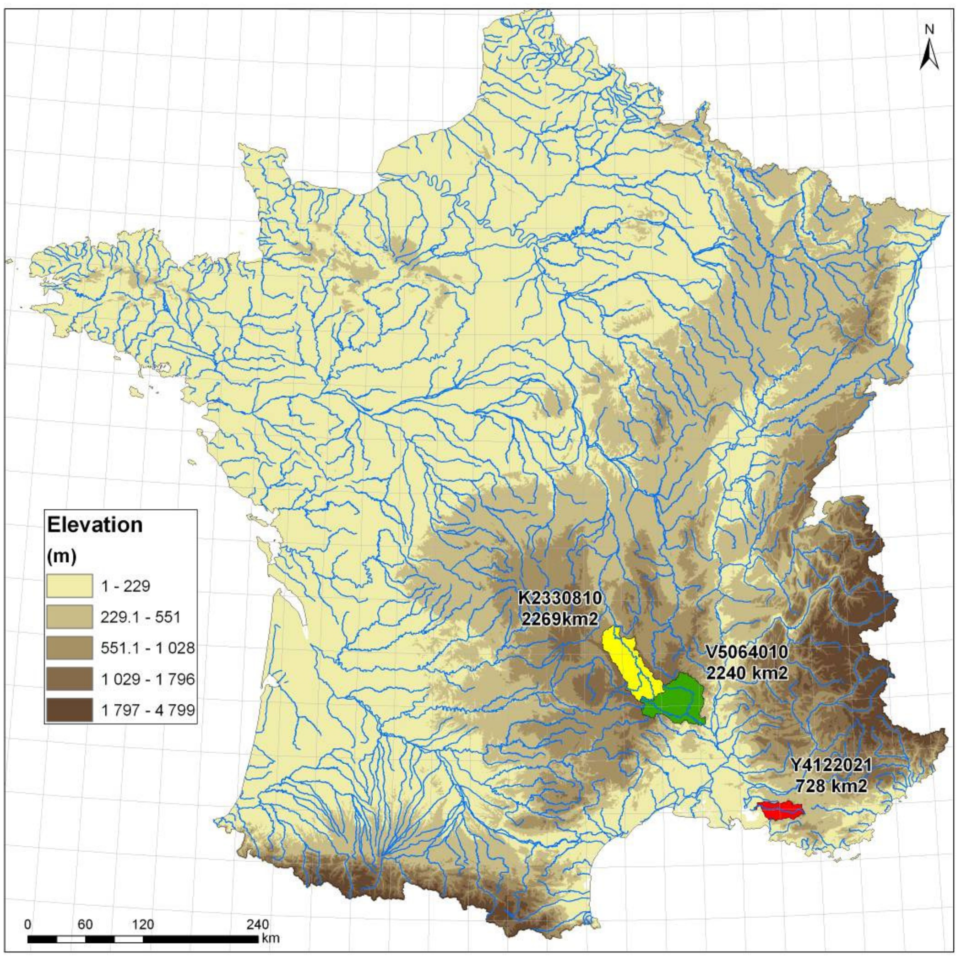


Figure 3: Location of the three study catchments with indication of their codes and surface areas

Table 1: Study areas and data availability

		K2330810	V5064010	Y4122020
Location		L'Allier at Vieille Bridge	L'Ardèche at Saint-Martin-d'Ardèche [Sauze-St-Martin]	L'Arc at Berre-l'Étang [St-Estève]
Surface area (km2)		2269	2240	728
Minimum P [mm]		0	0	0
Maximum P [mm]		86.3	195	119
Annual Mean P [mm]		895	1406	676
Minimum ETP [mm]		0	0	0
Maximum ETP [mm]		5.3	5.6	5.9
Annual Mean ETP		576	702	819
Minimum Q [m3/s]		0.965	1.5	0.167
Maximum Q [m3/s]		810	2515	224
Annual Mean Q [m3/s]		29.3	64.9	2.62
Period of observation (P,ETP, Q)		01/08/1958 to 31/07/2009	01/08/1958 to 31/07/2009	01/08/1958 to 31/07/2009
Availability of discharge data	Period with no data	04/04/1987 to 05/04/1987	29/07/1974 to 31/08/1974 09/08/1983 to 18/08/1983 24/10/1983 to 30/11/1983 01/01/1986 to 31/12/1986 15/01/1990 to 22/01/1990 30/03/1990 to 30/04/1990 06/09/1990 to 31/10/1990	01/08/1958 to 31/12/1970 01/01/1972 to 31/12/1979 01/01/1983 to 31/08/1999 21/06/2001 to 31/07/2001 18/09/2001 to 31/10/2001 22/11/2001 02/08/2002 26/08/2002 to 27/08/2002
	Total days of no data	2	543	13634
Period of rainfall prediction	PEARP	11/03/2005 to 31/07/2009	11/03/2005 to 31/07/2009	11/03/2005 to 31/07/2009
	Missing period of PEARP	29/11/2007 to 30/11/2007 02/01/2009 to 01/02/2009	29/11/2007 to 30/11/2007 02/01/2009 to 01/02/2009	29/11/2007 to 30/11/2007 02/01/2009 to 01/02/2009
	ECMWF	11/03/2005 to 30/09/2008	11/03/2005 to 30/09/2008	11/03/2005 to 30/09/2008
Q rating curve		11/03/2005 to 31/07/2009	11/03/2005 to 31/07/2009	11/03/2005 to 31/07/2009
Spatial averaged P		-	01/01/2000 to 31/12/2008	-

2.2 Observed precipitation and discharge data

The observed precipitation is used as input for the GRP model in the calibration period. In the forecast period, observed precipitation is used to simulate the discharge until the day of issuing the forecast. Observed discharges are used to calibrate and validate the model, as well as update the state of the model at each time step during the forecast period. Observed precipitation data come from the meteorological analysis system of Météo-France (SAFRAN) and observed streamflow data come from the French database Banque HYDRO. The data used in this research is daily data.

The observation period of precipitation is from 1/8/1958 to 31/7/2009 (about 51 years). During this period, availability of discharge data varies with catchments, as shown in Table 1; catchments Allier and Ardèche have long series of discharge while catchment Arc has only about 14 years. With GRP calculation algorithm, when no data of discharge are found, the model still simulates in the days with no data but the objective functions are calculated without those days. Hence, it will not affect too much the calibration and validation if the number of missing values is small and scattered in the whole period. For catchment Arc, the data are still available for about 14 years so it is still acceptable for calibrating and validating the model. Besides, it would be also interesting to look at catchment Arc because it is a small catchment (728 km^2) compared with the other two with quite similar size (2269 km^2 and 2240 km^2).

2.3 *Forecasting system and GRP model*

2.3.1 Forecasting system

A forecasting system contains many different components as shown in Figure 1. Forecasting ensemble river discharge with a hydrological model requires defining a particular model structure, with its parameters, input data and, usually, an updating procedure to start the forecasts from initial conditions as close as possible to observed conditions. This composes a forecasting model.

To implement a hydrological forecasting model, it is thus necessary to provide the input meteorological data (precipitation, temperature, etc.) for flow simulation and weather forecast (precipitation, temperature, etc.) to predict the river flow. The initial state of the system needs also to be defined so that the forecast can start from a certain, properly defined point. This initial state can be defined without a specific updating routine based only on the simulated discharge at the end of the previous time step, which already takes into account observed meteorological conditions. However, as the model flow output contains errors due to the limited representation of the real system, the simulated discharges are also subject to some errors, which might be exacerbated in time during the continuous simulation of the hydrological model. Therefore, in order to avoid such errors, other kinds of observed data, for example observed discharges can be used to update the forecasting model and bring its output closer to real time conditions before issuing a forecast. It should be kept in mind, however, that the observed discharges are not error free, as observation is still far from being perfect due to the instrumental errors and the difficulties of capturing the natural space-time variability of stream flows. As the model represents the real system through a set of parameters that characterizes that system, the parameters need to be defined before using the model to forecast. In order to define the model parameters, the model is calibrated using a long period of historic data (called calibration period); after calibration, the model can be used for forecasting. Calibration and forecasting period need to be independent.

Finally, after issuing a forecast, the output of the model, the forecast discharge, should be compared with a reference value to evaluate the forecast performance. This reference discharge can be the observed discharge. Here again, as the observed discharge is also subject to measurement errors, using it as reference might lead to a wrong interpretation of forecast outcome.

2.3.2 GRP model structure and parameters

The hydrological model used in this research is the GRP model, integrated in the GRPE forecasting system to forecast discharges from an ensemble of forecast precipitation scenarios at each forecast time step. Weeink (2010) and Berthet et al. (2009) provided a detailed description of the model structure. Here, a summarized description is presented. More detailed information about the model can be found in the above papers.

The GRP model is a daily lumped hydrological forecasting model developed in Cemagref, France (Figure 4).

Model inputs are the areal rainfall P, either observed or forecast for stream flow simulating and forecasting respectively; and the potential evapo-transpiration PE. In practice, the climatological value for the potential evapo-transpiration is used (for example, the same seasonally variable evapo-transpiration, identically repeated each year), as previous studies have shown that, for the family of GR models, there were no systematic improvements in the rainfall-runoff model efficiencies when using temporally varying evapo-transpiration (Oudin et al., 2005).

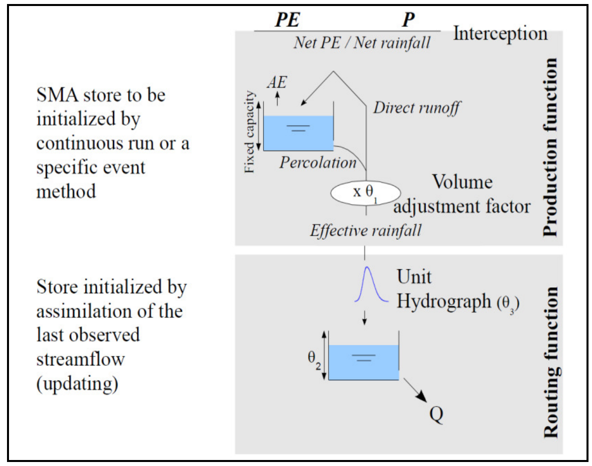


Figure 4: Model structure of the GRP hydrological forecasting model (Berthet et al., 2009)

The GRP model consists of a production function and a routing function as shown in Figure 4. The model describes the hydrologic process from rainfall to runoff as a sequence of processes in one time step. The time step used in this study is daily time step.

The GRP model consists of three calibrated parameters: the volume adjustment factor which controls the volume of effective rainfall X_1; the level of the routing store X_2 and the base time of the unit hydrograph X_3.

Besides, a number of fixed parameter is defined for the forecasting model running at daily time steps:

- Production reservoir capacity: A=350mm
- Percolation function coefficient: B=2.25
- Unit hydrograph exponent: α=2.5
- Outflow routing reservoir exponent: β=2.0

2.3.3 Calibration, validation and forecast in the GRPE forecasting system

The GRP parameters are calibrated using the historic data of discharge. The calibration and validation are done automatically. Optimization searches for the global optimum of a given objective functions. Four performance criteria are automatically computed by the calibration-validation routine:

Root Mean Square Error: $$RMSE(L) = \sqrt{\frac{\sum_t \left(\hat{Q}_{t+L|t} - Q_{t+L}\right)^2}{n}} \quad [1]$$

Persistence index: $$PI(L) = 1 - \frac{\sum_t \left(\hat{Q}_{t+L|t} - Q_{t+L} \right)^2}{\sum_t \left(Q_t - Q_{t+L} \right)^2} .100\%$$ [2]

Nash-Sutcliffe: $$NASH(L) = \left(1 - \frac{\sum_t \left(\hat{Q}_{t+L|t} - Q_{t+L} \right)^2}{\sum_t \left(\hat{Q}_{t+L|t} - \overline{Q} \right)^2} \right) .100\%$$ [3]

Transformation of Persistence Index: $$C2MP(L) = 100. \left(\frac{\frac{PI(L)}{100}}{2 - \frac{PI(L)}{100}} \right)$$ [4]

where L is the lead time, Q_t and Q_{t+L} are the observed discharges at time step t and $t+L$ respectively; $\overline{Q}$ is the average of observed discharges; while $\hat{Q}_{t+L|t}$ is the forecast issued at time step t for time step $t+L$; n is the number of time steps.

The model uses all observed data previously to the start of forecast period to calibrate its parameters. The optimization searches to maximize the PI values. When forecasting, the model simulates stream flow until the day of issuing the forecast, and then forecast the flows for the L days after depending on the length of lead times.

3 Methods

In this chapter, literature is reviewed to identify the possible uncertainty sources that might propagate into the forecasting system of interest; as well as to determine which sources play a crucial role in this system. Based on that, methods for quantifying each of those “important” sources of uncertainty is described together with the corresponding data that are used. After that the method for experimental uncertainty propagation is introduced. At the end, the evaluation measures for assessing the uncertainty quantification are described.

3.1 Identifying uncertainty sources

3.1.1 Initial identification of uncertainty sources

A large number of researches on uncertainty of hydrological modelling or flow forecasting can be found in the literature, for example, Uhlenbrook et al. (1999), Carpenter and Georgakakos (2004), Mc Milan et al. (2010), Kuczera et al. (2010), etc. Therefore, it is useful to make use of the existing knowledge to study possible sources of uncertainty in a system. Besides, there are also different methods for doing this work; for instance, one might use the expert opinion method proposed by Warmink et al. (2010). Whatever method is applied to identify the sources of uncertainty, literature study is definitely an inevitable step. Here this method is used for identifying the uncertainty sources that play a role in hydrological forecasting in general and GRPE flow forecasting system in particular.

In the hydrological modelling context, Walker et al. (2003) categorized five uncertainty sources (or "location" as the authors used in their papers) which are context, model, inputs, model parameter, model outcome uncertainty. The model outcome uncertainty is the result of propagating the other uncertainty sources through the model; so basically there are four sources that might cause the uncertainty in hydrological modelling.

Context is an identification of the boundaries of the system to be modelled, and thus the portions of the real world that are inside the system, the portions that are outside, and the completeness of its representation. Because it is related to the limitation of the model by which the real world cannot be fully described, here the context is classified in the uncertainty source arose from the model.

In addition, the model uncertainty has two other categories, model structure uncertainty and model technical uncertainty. The model technical uncertainty is generated by the software or hardware errors. In case of GRP model, this is a lump, conceptual model with a simple code which is written in FORTRAN language; therefore, the model is expected to be not subject to the model technical uncertainty. The model structure uncertainty arises from a lack of sufficient understanding of the system, including the behaviour of the system and the interrelationships among its elements. Model structure uncertainty also involves the mathematical algorithm, equations and assumptions of the model. In GRP model structure, the uncertainty can come from the representation of rainfall-runoff process using the production and routing reservoirs, or the definition of a threshold level according to that the amount of evapo-transpiration is decided, or the hydrological routing process which is described by the unit hydrograph.

The input data uncertainty is divided by Walker et al. (2003) into uncertainty about the external driving force and uncertainty about the system data. The external driving force that produces changes within the hydrological process is the meteorological force; in the case of GRP forecasting model, it is the input precipitation data used for the flow simulation mode of the system. The uncertainty about the system data including land use map, data on water-related infrastructure, which is often the result of the lack of knowledge on the system's properties, is not applicable to the GRP model, so it is not considered in this research. Finally, it should be noted that in their paper, the authors considered the uncertainty in hydrological modelling, not forecasting; in a hydrological forecasting system, there is another uncertainty which can be also classified as uncertainty about the input data. This is the uncertainty from the forecast driving force or the forecast precipitation which is used for flow forecasting in the GRPE system.

Concerning the model parameter uncertainty, there are four different types of parameters suggested by Walker et al. (2003), divided into two sub-categories: non-calibrated parameters and calibrated parameters. Non-calibrated parameters include: (1) Exact parameters (universal constants); (2) Fixed parameters (well defined parameters like gravity acceleration g); and (3) Prior chosen parameters (which are parameters that may be difficult to identify by calibration and are chosen to be fixed to a certain value that is considered invariant). In the GRP model, as mentioned in Section 2.3.2, there are four main parameters whose values are fixed. The uncertainty of non-calibrated parameters can come from the fact that these parameters are chosen and fixed at some values. The calibrated parameters are unknown with previous experience and need to be determined by minimizing the difference between model outcomes and measured data of the same period and location. Here these are the three calibrated parameters of GRP. The choice of these parameters when forecasting river flow can affect the forecast outcome and cause uncertainties.

In Walker et al. (2003), the uncertainty coming from the initialization of the model is not mentioned. However, this source of uncertainty was investigated by other authors (for example, Berthet et al. (2009)). For the GRP model, at each forecast day, the state of the system is updated with the observed discharge at the end of the previous time step. However, this choice of initial condition might also lead to uncertainties in the discharge forecasts as the observed discharge might contain errors and might not be a good start for the simulation at the forecast time step. Another uncertainty source which is also not mentioned in Walker et al. (2003) but was studied in the literature is uncertainty of the observed discharge (See Mc Milan et al. (2010)). As the outcome of the forecasting model is compared with the observed discharges, which is an uncertain reference, this action might cause a misinterpretation of the results if the uncertainties are not considered.

The uncertainty in hydrological forecasting is the accumulated uncertainty of the above uncertainties. All sources of uncertainty in hydrological forecasting are summarized Figure 5 below:

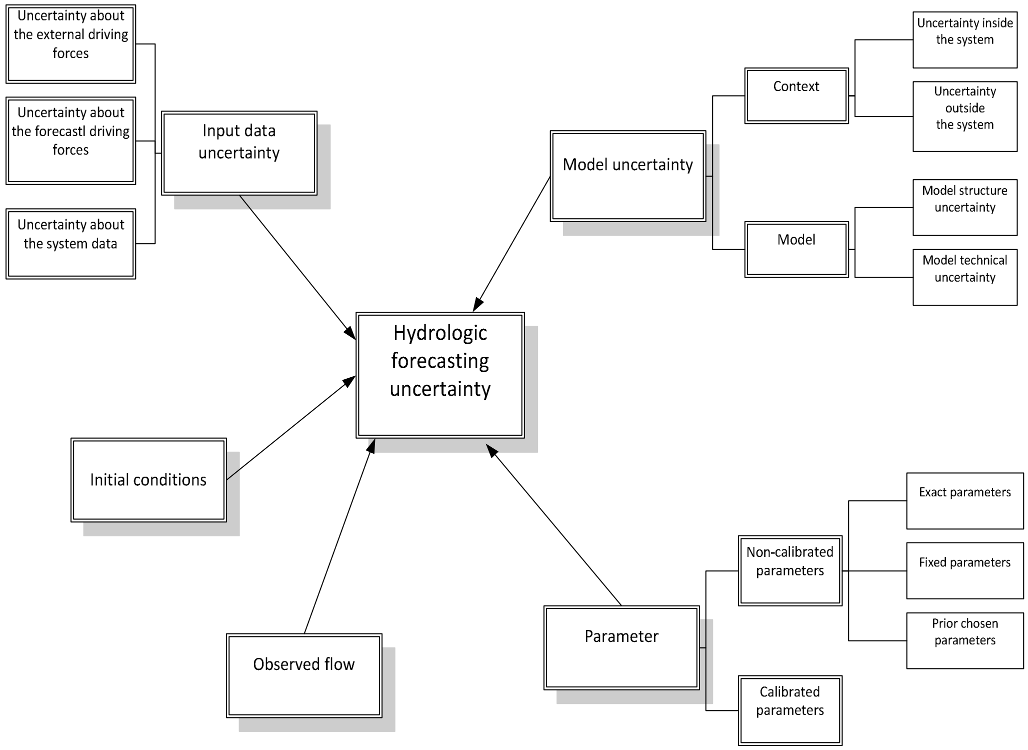


Figure 5: Different uncertainty sources that can propagate into hydrological forecasting

In the context of GRPE forecasting system, the uncertainty sources are shown in Figure 6.

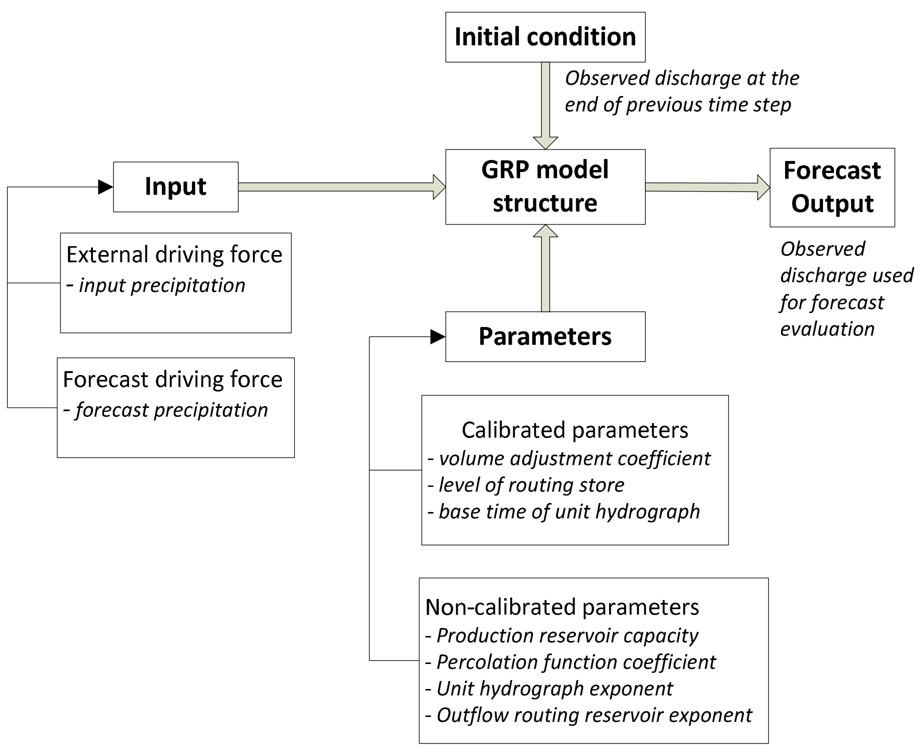


Figure 6: Identification of uncertainty sources for GRPE forecasting system

3.1.2 Determination of uncertainty source importance

As discussed above, there are different sources of uncertainty in hydrological modelling and forecasting; and the literature shows that those uncertainties do not have the same influence on the outcome flows; some might have very large influence and others might have negligible impact. For example, the uncertainty about the non-calibrated parameters is not mentioned in the literature as an important source of uncertainty that should be considered in uncertainty propagation. In this step, previous researches are considered to determine the most important uncertainty sources to set the focus for uncertainty quantification and propagation in the next steps.

Among all of those uncertainties, meteorological input uncertainty is usually assumed to be the largest source of uncertainty in the prediction of floods, at least for lead times of 2-3 days (Rossa et al., 2010). He et al. (2009) also showed that forecast precipitation uncertainties dominate and propagate through the cascade chain. The input uncertainty is significant because of the high spatial and temporal variability of precipitation (Kavetski et al., 2006). Hence, attention should be paid on this sort of uncertainty when assessing the flood forecast uncertainty.

The initial moisture conditions at the beginning of a rainfall event have a major influence on a catchment's hydrological response and therefore have a crucial impact on flood forecasting. Berthet et al. (2009) analyzed the influence of initialization on flood forecasting for 178 catchments in France. They found that different methods of initialization could result in very large differences in the flood forecasts. A persistence index was used to evaluate the discharge forecasts; and the result of their research shows that at the 1-hour lead time, the persistence differences for different (arbitrary) initial values are greater than 0.03 (which is a significant difference) on more than 75% of the catchments; for the 48-h lead time, this difference is greater than 0.14 for more than 90% of the catchments. Schaake et al. (2006) also emphasized that one of the most important source of hydrological model uncertainty is the uncertainty in initial conditions.

Beside the uncertainty caused by the selection of initial condition, the model structure and parameterization uncertainties are also significantly pronounced in flow forecasting. A study done by Butt et al. (2009) shows that the sensitivity of streamflow simulations to variations in acceptable model structure was at least as large as uncertainties arising from parameter and observation uncertainty. The authors stated that model performance was strongly dependent on model structure.

Hughes et al. (2010) also reported on relatively large uncertainties of the hydrological model they used related to parameter values even when the input data uncertainty was not taken into account. Authors suggested that the forecast uncertainty will be probably larger when the input uncertainty is considered.

The examples in their research suggest that there are differences in the degree of parameter value uncertainty for sub-basins that have different physiographic, climate and runoff response characteristics.

Uhlenbrook et al. (1999) studied the prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structure. The authors also reported the importance of accounting for model structure and parameter uncertainty when evaluating the forecasting uncertainty. They noted that the difficult identifiability of the model structure caused uncertainties in the flow predictions but they were slightly smaller than the implications caused by the parameter uncertainty. However, perhaps this conclusion comes from the rather similar HBV model variants tested in this study. The variations of simulation results would probably increase for predictions with totally different conceptual rainfall-runoff models.

Mc Milan et al. (2010) studied the errors in discharge measurements used to calibrate a rainfall runoff model, caused by the rating curve uncertainty. The authors recognized that serious considerations of uncertainties in discharge data can significantly improve the prediction ability of the model. This implies the importance of accounting for observed discharge data uncertainty in flow forecasting.

In summary, there are some uncertainty sources, which are found important in flow simulating and forecasting in the literature, namely uncertainty about the input data (external driving forces, forecasting driving force), initial conditions, model parameters, and model structure.

To be more specific to the case of the GRPE forecasting system used in this study, hereafter, the uncertainty of external driving forces is referred to as input precipitation uncertainty; and the uncertainty of forecast driving force is referred to as forecast precipitation uncertainty. These important uncertainty sources are listed in Table 2.

Table 2: Main sources of uncertainty in flood forecasting

Sources of uncertainty		Importance	References
Input data uncertainty	Uncertainty about input precipitation	High	Rossa et al. (2010)
	Uncertainty about the forecast precipitation	High	He et al. (2009)
	Uncertainty about the system data	Low	
Model uncertainty	Context	Low	
	Model structure uncertainty	High	Butt et al. (2009)
	Model technical uncertainty	High	
Model parameter	Uncertainty about non-calibrated parameters	Low	
	Uncertainty about calibrated parameters	high	Butt et al. (2009)
Initial conditions	Uncertainty about the current state of the system	high	Berthet et al. (2009)
Observed discharge	Uncertainty from the rating curve	High	Mc Milan et al. (2010)

3.2 Quantifying the impact of individual uncertainty on flood forecast

In this research, there are five main sources of uncertainty are investigated, namely (1) input precipitation uncertainty, (2) forecast precipitation uncertainty, (3) initial condition uncertainty, (4) calibration period uncertainty, (5) parameterization uncertainty. The uncertainty about the model structure is not considered here due to time constrain as it would require using several other model structures, but could be another source of uncertainty to be considered for future research. The materials those are used to account for the uncertainties are described below; the individual uncertainty tests are explained at the end of this section.

3.2.1 Input precipitation uncertainty

The observed rainfall data that are often used for catchment studies are often not areal rainfall because rainfall cannot be quantitatively measured in space with sufficient precision for catchment modelling. Usually, rainfall is only observed at some stations (point rainfall), located either inside or outside the study catchment. In order to simulate rainfall-runoff process in the whole basin area, it is necessary to spatially interpolate point data. There are many methods used for interpolating and averaging rainfall, as, for example, the Thiessen polygon method, the arithmetic mean method, the isohyetal method (Shaw, 1994). It is a fact that none of these methods can properly produce an areal rainfall; and the uncertainty from the methods combined with uncertainties from the instruments used to measure precipitation, leads to errors in observed rainfall data used in hydrological modelling.

In this research, to account for the uncertainty of observed rainfall, precisely the uncertainty of transforming point rainfall to areal rainfall (errors from rain gauge instruments are neglected), the areal rainfall generator developed at Cemagref is used. This generator is based on the geo-statistical Turning-Bands-Method (TBM) for the simulation of random field and has been applied for characterisation of rainfall intensities (Ramos, et al., 2006). The version used in this research comes from recent developments made at Cemagref (Lepioufle, 2009). It first applies the conditional simulation method to hourly rainfall observed data and then aggregates the different simulated fields to daily time steps. Conditional simulation is a geo-statistical method that generates multiple statistical realizations of the spatial rainfall field over a specific area, while preserving the information measured by the individual rain gauges. Spatial and temporal structures are also preserved in the simulations. The rainfall generator is built from the analysis of variograms and the evolution of spatial structure with time. Besides, rainfall fields are generated according to different successive rain types, defined by using a Kohonen algorithm and statistical properties of non-zero precipitation and total coverage area.

For this research, the rainfall generator is set up for one of the study catchments, catchment Ardèche. We then ran the simulator for the period from 2000 to 2008. Ten realisations (members) are generated at each hourly time step. Figure 7 Illustrates 4 generated fields of spatially distributed rainfall over catchment Ardèche for the day 12/01/2004 at 13:00.

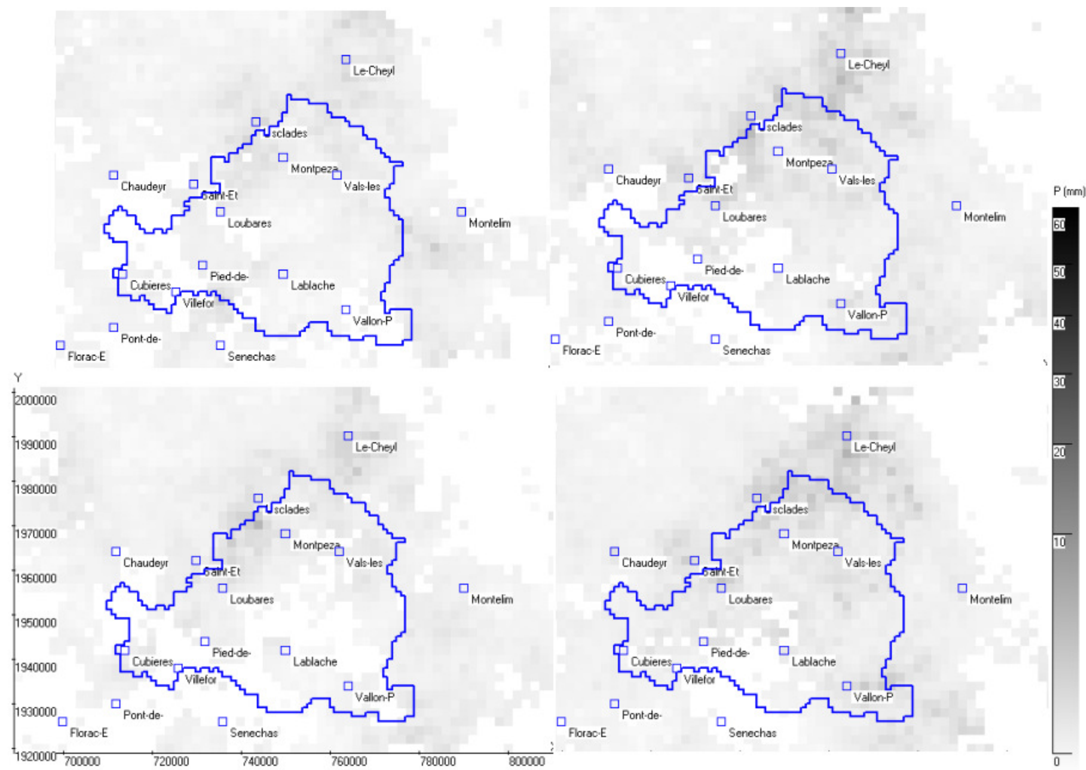


Figure 7: Four out of ten simulations of spatially distributed rainfall over catchment Ardèche on 12/01/2004 at 13:00. Rain gauges are presented by squares. Each realisation preserves the observed data at the rain gauges

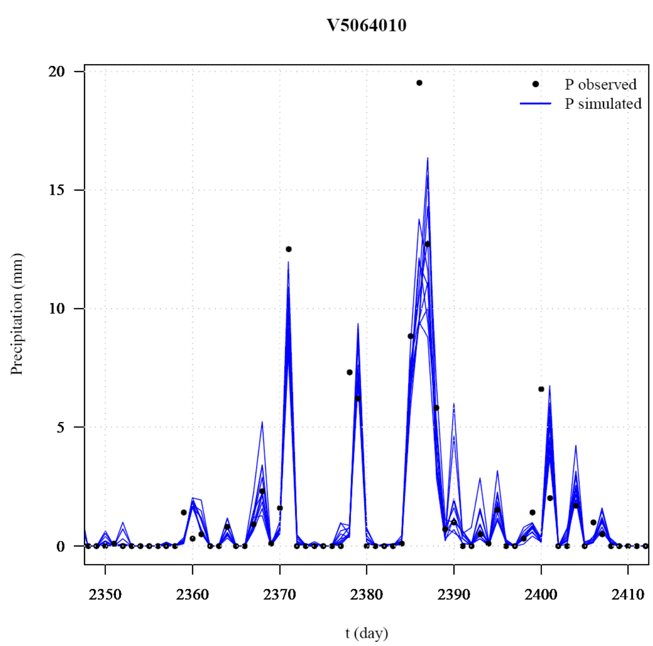


Figure 8: Ten simulated areal precipitation values for catchment Ardèche from 07/06/2006 to 06/08/2006 (blue lines). Observed precipitation given by the SAFRAN Météo-France data analysis system is also indicated (black dots).

From the geo-statistical conditional simulation, every simulation is a possible realization of the rainfall over the catchment. Because the hydrological model applied in this research is a lumped model running at daily time steps, the distributed hourly rainfall fields were aggregated in space (average precipitation over the catchment area) and in time at daily time steps. As a result, 10 values of areal rainfall are available at each day of the forecast period. They represent observed rainfall uncertainty at the catchment scale, as illustrated on Figure 8, 10 series of spatial averaged rainfall are plotted in a period of 2 months together with the observed rainfall data from SAFRAN. These members of spatially averaged rainfall are used as input rainfall for the GRP model to assess the impact of input data (observed precipitation) uncertainty.

3.2.2 Forecast precipitation uncertainty

The weather observation is no way perfect or complete because of the chaotic atmosphere system which causes the difficulties in predicting. In addition, because of the limitation in numerical computer modelling, the weather system is inevitably an approximation of the exact solutions of the equations describing the system. Therefore, every weather forecast contains, to some extent, uncertainty. This uncertainty is unfortunately not constant but varies from day to day, depending on the stability of the atmospheric condition at the start of the forecast. The major uncertainty of weather forecasts comes from the estimation of the initial state of the atmosphere and from unavoidable simplifications in the representation of the complex nature in weather numerical models (European Centre for Medium-Range Weather Forecasts, 2011). For that reason, deterministic forecasts, which produce only one forecast based on the best estimate of the future event, are no longer suitable for many applications in practice, as uncertainty is not attached to the forecasts.

In order to deal with the problem of the weather forecast uncertainty, Ensemble Prediction Systems (EPS) were developed by several meteorological forecast centres. It is a kind of probabilistic forecast that represents the uncertainty in both initial conditions of the atmosphere and the numerical model used. Instead of using only one best estimate of the initial condition, slightly different states of the atmosphere, which are close but not identical to the best estimate, are used. Additionally, each forecast is based on a model (or on a different parameterization of a model), which is close but not identical, to the best estimate model equations. The result is not only one, but a number of forecasts spreading around the "control member", which is based on the best estimate of initial state. This technique provides an estimate of the uncertainty associated with predictions from a given set of initial conditions compatible with observation errors. If the atmosphere is in a predictable state, the spread will remain small; if the atmosphere is less predictable, the spread will be larger. In a reliable ensemble prediction system, reality will fall somewhere in the predicted range. This means that users get information on the actual predictability of the atmosphere, for example, whether a particular forecast can be expected to be certain or less certain. In addition, they also get information on the range within which they can expect reality to fall.

In this research, two EPS products are used, one from the European Centre for Medium-Range Weather Forecasts (ECMWF) and the other from the forecast ensemble system developed at Météo-France (PEARP). An example of an ECMWF EPS forecast over France is given in Figure 9.

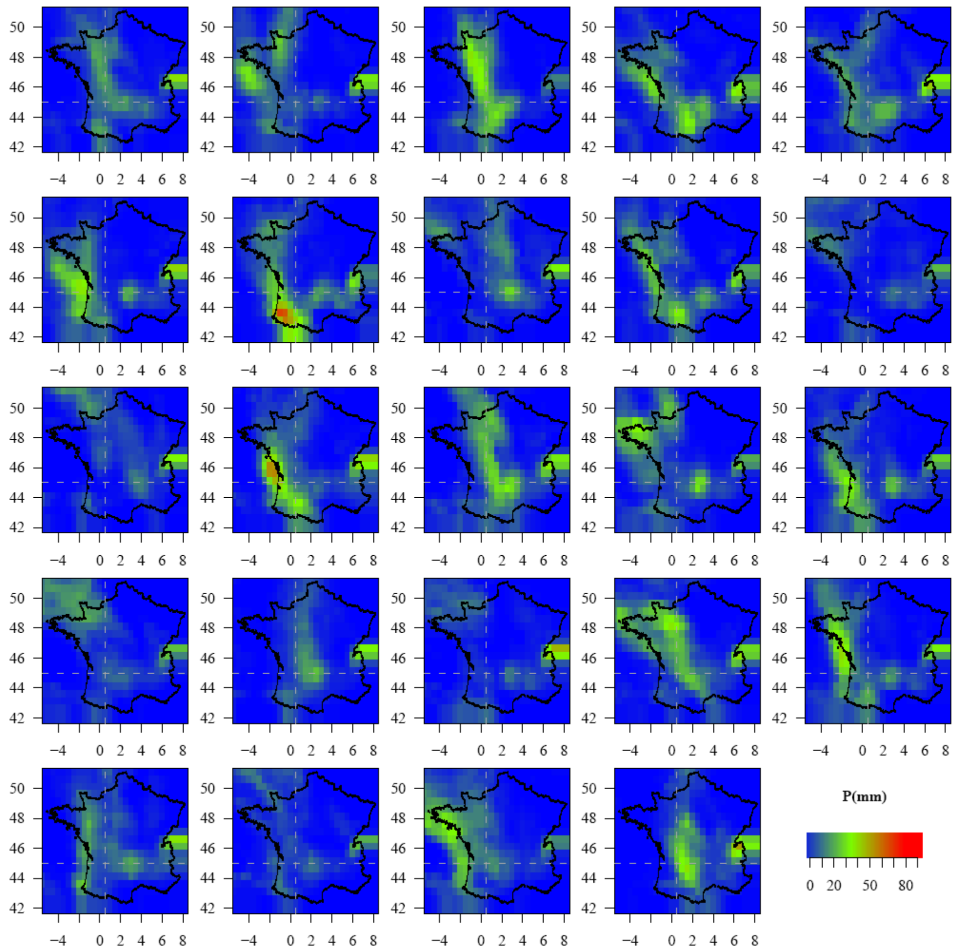


Figure 9: Maps of 24 (out of 51) members of EPS forecasts from ECMWF in France on 24 May 2008 at lead time of 4 days

ECMWF EPS data are archived forecasts and consist of one control member and 50 perturbed members, making a total of 51 forecasts at each day of the forecast period. The system focuses on medium-range forecasts (from 3 days onwards). Data was provided at a spatial resolution of 0.5° x 0.5° latitude-longitude (equivalent to spatial grid of approximately 50 km in France) and at variable time step, up to a lead time of 14 days. For this study, areal forecast precipitation was available for each catchment at daily time steps for the period from 11/03/2005 to 30/09/2008.

The Météo-France PEARP is focused on short-term forecast (less than 3-4 days of lead time) (Nicolau, 2002), and its application to hydrology has been recently reported in the literature (Thirel et al., 2008; Randrianasolo et al., 2010). Data for this study is available at spatial grid of 8 x 8 km over France and 3-hourly time steps. The number of members is limited to 11 members including one control member, and the maximum lead time is two days. For this study, areal forecast precipitation data are available for each catchment at daily time steps for the period from 11/03/2005 to 31/07/2009.

Because of its higher resolution, PEARP is used for the 1- and 2- day forecast lead times, while ECMWF EPS is used for lead times from 3 to 9 days. The predictions from ECMWF EPS for longer lead times are not used because after 9 days the resolution reduces, as well as the skill of the forecasts.

3.2.3 Initial condition uncertainty

The discharge data that are often used in hydrologic studies are not direct observed flows but they are usually retrieved from the rating curve – a relation curve between discharge and water level. Normally, the rating curve is constructed with a series of historic measured river flows and water levels. After that, water levels are measured continuously and, based on the value on the rating curve, discharge values are estimated. The values of discharge are regularly checked with the measurement but not for every time. Details about discharge measurement and rating curve construction can be reached at Shaw (1994).

Discharge data are subject to three kinds of errors: errors from discharge measurement, errors from rating curve fitting, and errors from water level measurement.

The errors in the discharge measurements can come from the operational condition during gauging or from insufficient number of verticals, from insufficient number of point velocity measurements per vertical or from the flow variation during the measurement period, etc. The rating curve fitting error is caused by the imperfection of the relationship curve, even if the true water stage and discharge were known. Water levels are usually assumed to be error free, although errors associated to the measurements device can exist. Therefore, uncertainty always exists when the rating curve is used to estimate the flow values.

In this research, initial condition uncertainty is assessed through the rating curve uncertainty, which is estimated for each catchment using a Bayesian approach (Renard et al., 2011). In the first step, the uncertainty in each individual gauging is quantified. It depends on the gauging method (e.g. velocity-area method, tracer dilution, ADCP, surface velocity measurements, etc.) and the operational characteristics of the gauging (e.g. spatial sampling of velocity and depth throughout the cross-section, unsteadiness of the flow, etc.). A set of standard deviations representing the measurement uncertainty affecting each gauge is withdrawn after this step. The output is illustrated on Figure 10.

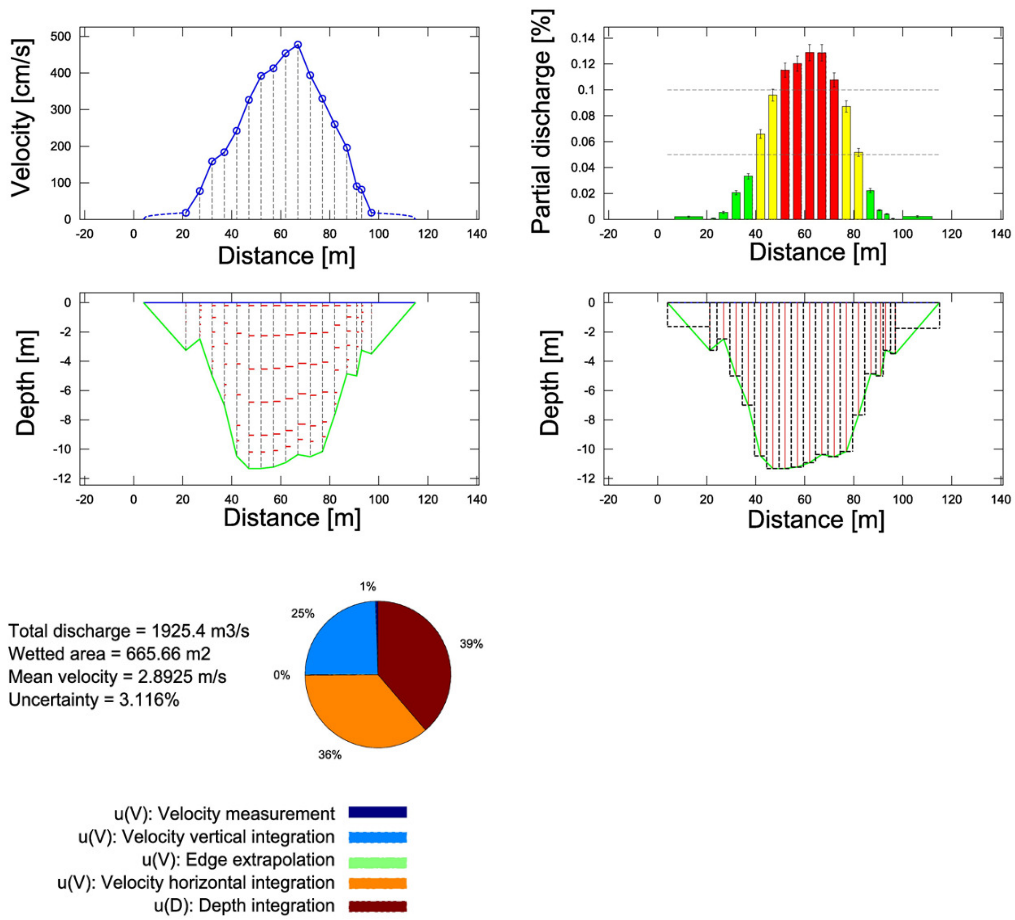


Figure 10: Example of an uncertainty analysis for one particular gauging (Renard et al., 2011)

The hydraulic configuration of the gauging cross sectional profile is then analysed to select a mathematical expression of the rating curve and specify priors for the rating curve parameters. The hydraulic configuration at a gauging station can be controlled by the section regime at low flow or channel regime at high flow.

Rating curve parameters are then estimated using a Bayesian approach. The prior information on the rating curve parameters is combined with the information brought by gauging data to perform the inference of the rating curve parameters. The outcome of this analysis is a rating curve (See Figure 11, solid black line) and its prediction intervals (dashed lines); the bars around gauging points represent measurement uncertainties which are calculated for each gauge.

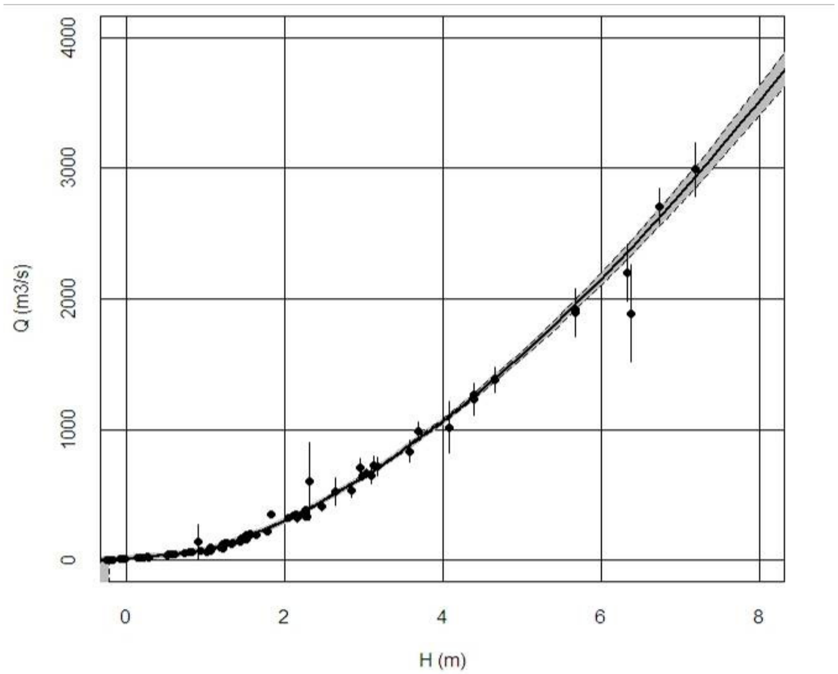


Figure 11: Example of estimated rating curve for catchment Ardèche, along with 90% prediction limits (Renard et al., 2011)

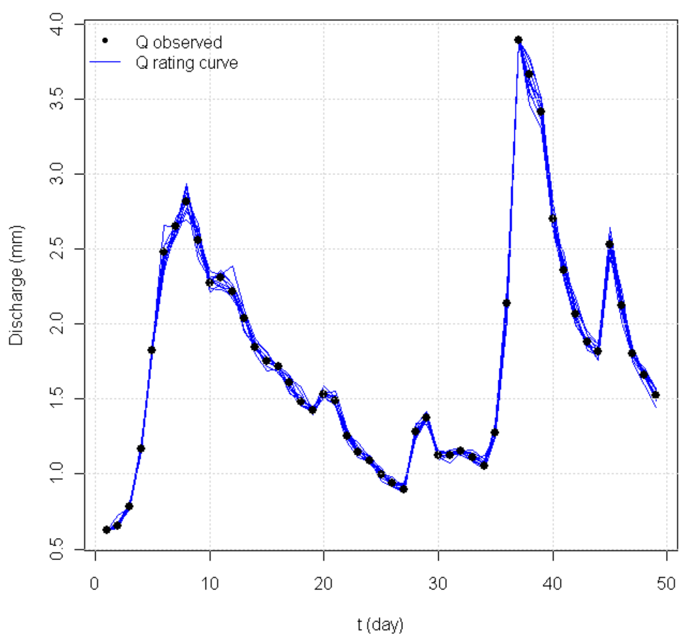


Figure 12: Ensemble discharges from rating curve for catchment Allier from 12/05/2005 to 30/04/2005. Observed discharges are also indicated in black dots

For this research, an ensemble of discharge values is built from the rating curve analysis for each day of the forecast period. For a given discharge, the corresponding stage is determined on the rating curve and then several random realizations are simulated from the predictive distribution represented by the gray interval. Here, 10 values of discharge are retrieved from each rating curve to form the ensemble on a daily basis, as shown in Figure 12.

These 10 members of ensemble discharges obtained from the rating curve are used to update the forecasting model, instead of updating it with only one value of observed discharge like it is done by default. By doing so, the impact of initial condition uncertainty on flood forecasting can be quantified.

3.2.4 Parameter uncertainty

In this research, parameter uncertainty is quantified using two alternative procedures: (a) the selection of different calibration periods for the forecasting model, and (b) the parameterization of the model using GLUE method.

a. Calibration period uncertainty

Concerning the calibration period, the uncertainty might appear due to the limited data for calibration or from the choice of calibration period which could fall into a mainly dry or wet period and lead to different optimized parameter values.

The GRP model is usually calibrated with a long series of data, which is here available from 1958 to 2005. In this research, to quantify the uncertainty, originating from the choice of the calibration period, this long period was split into different periods of five years. This is done in two catchments instead of all three due to the lack of enough observed discharge data. In total from 1958 to 2005, there are 9 periods of five years and the whole calibration period of 1958 to 2005 is also used as the 10th period to have the best estimation of parameters. The ten periods are shown in Table 3. It should be noted that normally for forecasting practice, the whole period of data is utilized for calibration to get the best possible parameters. Therefore, here the periods of data mentioned below are used all for calibration.

Table 3: Ten calibration periods for calibration period uncertainty quantification

No. of parameter sets	Calibration period (dd/mm/yyyy)
1	01/01/1959 to 31/12/1963
2	01/01/1964 to 31/12/1968
3	01/01/1969 to 31/12/1973
4	01/01/1974 to 31/12/1978
5	01/01/1979 to 31/12/1983
6	01/01/1984 to 31/12/1988
7	01/01/1989 to 31/12/1993
8	01/01/1994 to 31/12/1998
9	01/01/1999 to 31/12/2003
10	01/08/1958 to 10/03/2005

Figure 13 shows the values of the three parameters of GRP model, calibrated for 10 periods of five years, along with NASH values computed over those periods. The result does not show any clear trend (increase or decrease of the parameters with time). The performance of the model is quite stable during time; except for the period 1964-1938 for catchment Allier, for which the values are far departed from the rest. In general, there is no clear trend that the parameters are changed significantly during the time because of the change in land use or other human intervention. For that reason, it is assumed here that the difference in the choice of calibration periods here is only subject to the difference in natural variation of hydrological process; and therefore, it is reasonable to use those 10 sets of parameters to forecast discharges.

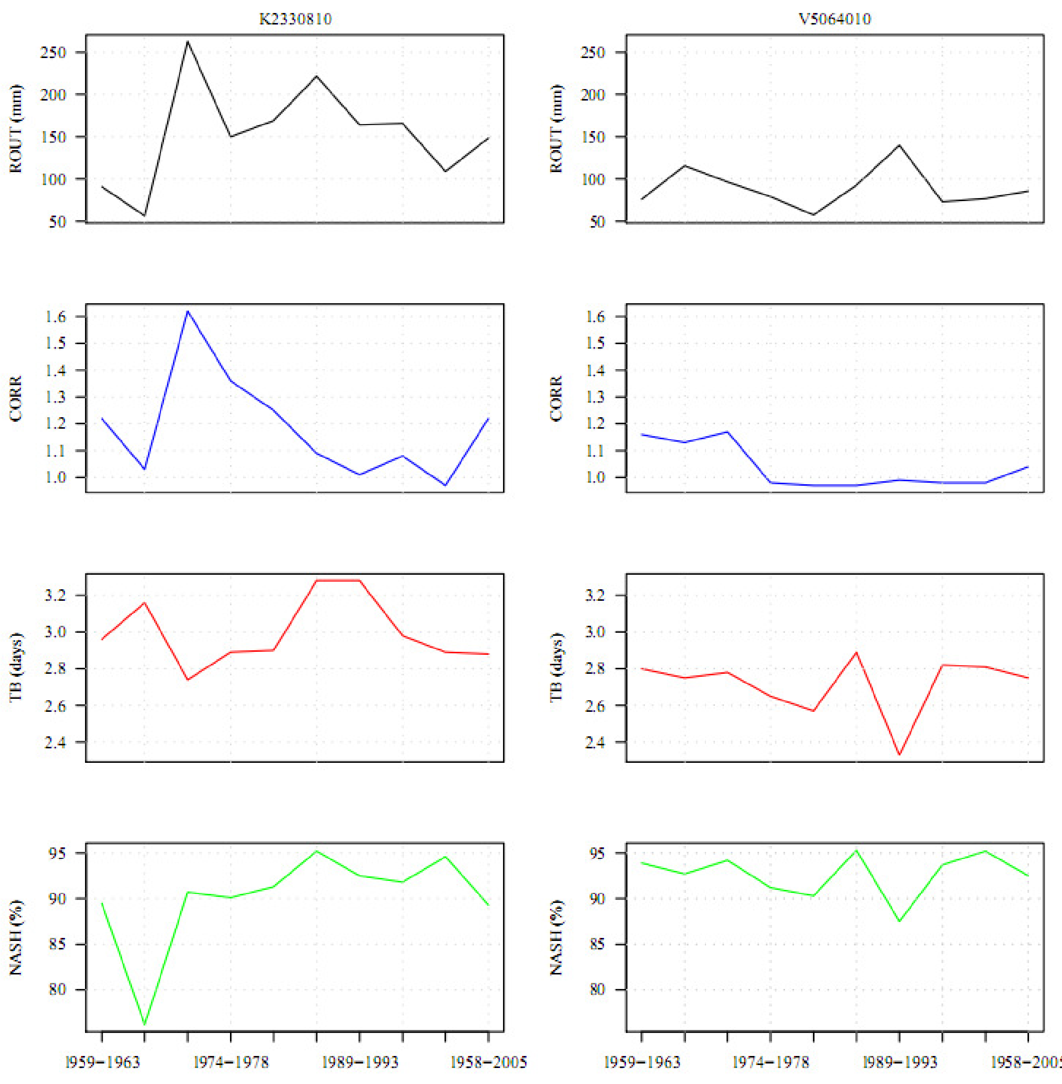


Figure 13: Changes in GRP model parameters and NASH values in 10 calibration periods for catchment Allier (K2330810, left) and Ardèche (V5064010, right)

b. Parameterization uncertainty

Parameterization uncertainty comes from the fact that different parameter sets can represent equally the behaviour of the modelled system. In this research, for the parameterization uncertainty, the Generalized Likelihood Uncertainty Estimator (GLUE) proposed by Beven and Binley (1992) is used. GLUE is performed by first identifying the parameters which most affect the output; in the case of GRP model, those are the three calibrated parameters. Then, a high number of parameter sets is generated based on the prior knowledge about the distribution of parameters; however, since this distribution is not often known, the uniform distribution is used instead. Since the three parameters of the conceptual GRP model do not have a physical interpretation; one could not say which ranges of values the parameters would fall in. For instance, the level of the routing reservoir can, theoretically, be from 0 to infinite. To obtain prior information about the ranges of the parameters of the GRP model, the model is first calibrated for 3070 catchments in France (see Figure 14). These catchments spread all over the country in different geographical and climate conditions with different quality and length of measured data. Thus, the ranges of parameters obtained are expected to be large enough to cover the possible ranges.

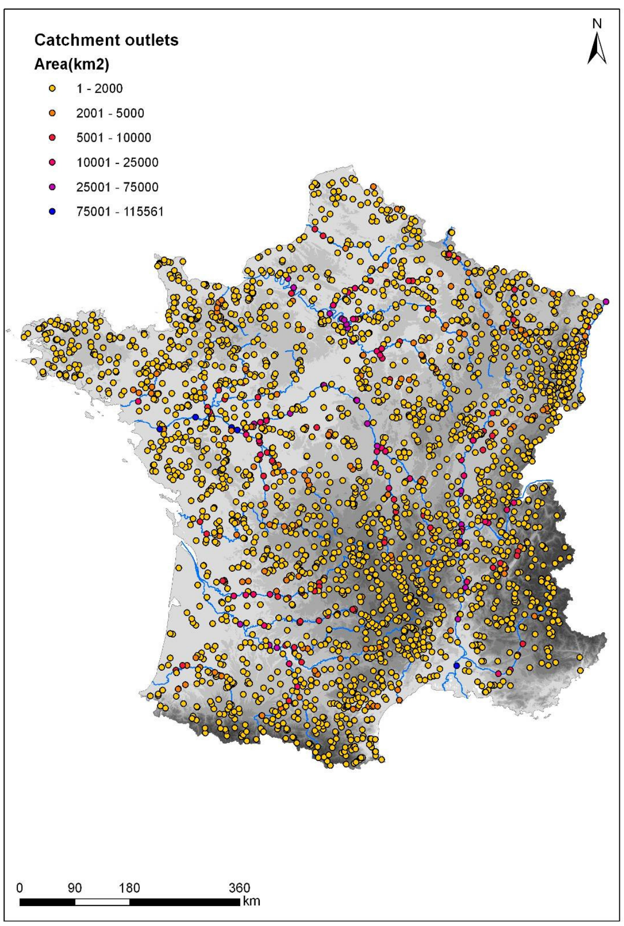


Figure 14: Map of 3070 catchments in France used for the definition of the GRP parameter ranges for the application of the GLUE method to quantify parameterization uncertainty

The scatter plots and histograms of the three parameters of the model are presented in Figure 15. The result shows the interdependence between ROUT and CORR which is reasonable because one is the level of routing reservoir and the other is the volume adjustment factor of the flow from production reservoir to routing reservoir and there might be a relation between those two. The relation between ROUT and TB are not clear. The histograms of all three parameters show that, the majority falls in the low value part of the parameter domain.

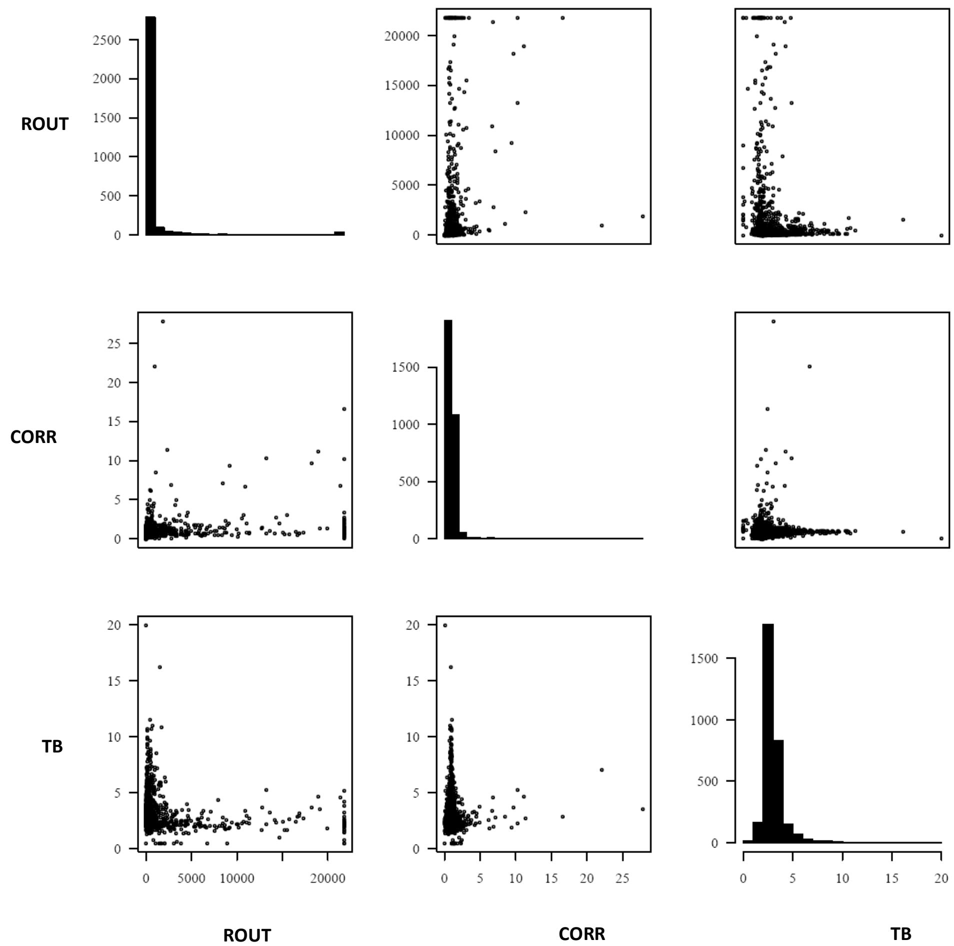


Figure 15: Scatter plots of GRP model parameters calibrated for 3070 catchments in France. Diagonal: Histograms of parameter distribution

The spatial distributions of the three parameters and the NASH values for the 3070 catchments are shown in Figure 16. No clear trend of the spatial distribution of parameters is observed. NASH values in the north (low area) tend to have very good result (>90%); but in the south (higher area) tends to have only good (>80%). The reason for that might be because the snow module of the model is not used here then the performance for the high elevation is not as good as the low area.

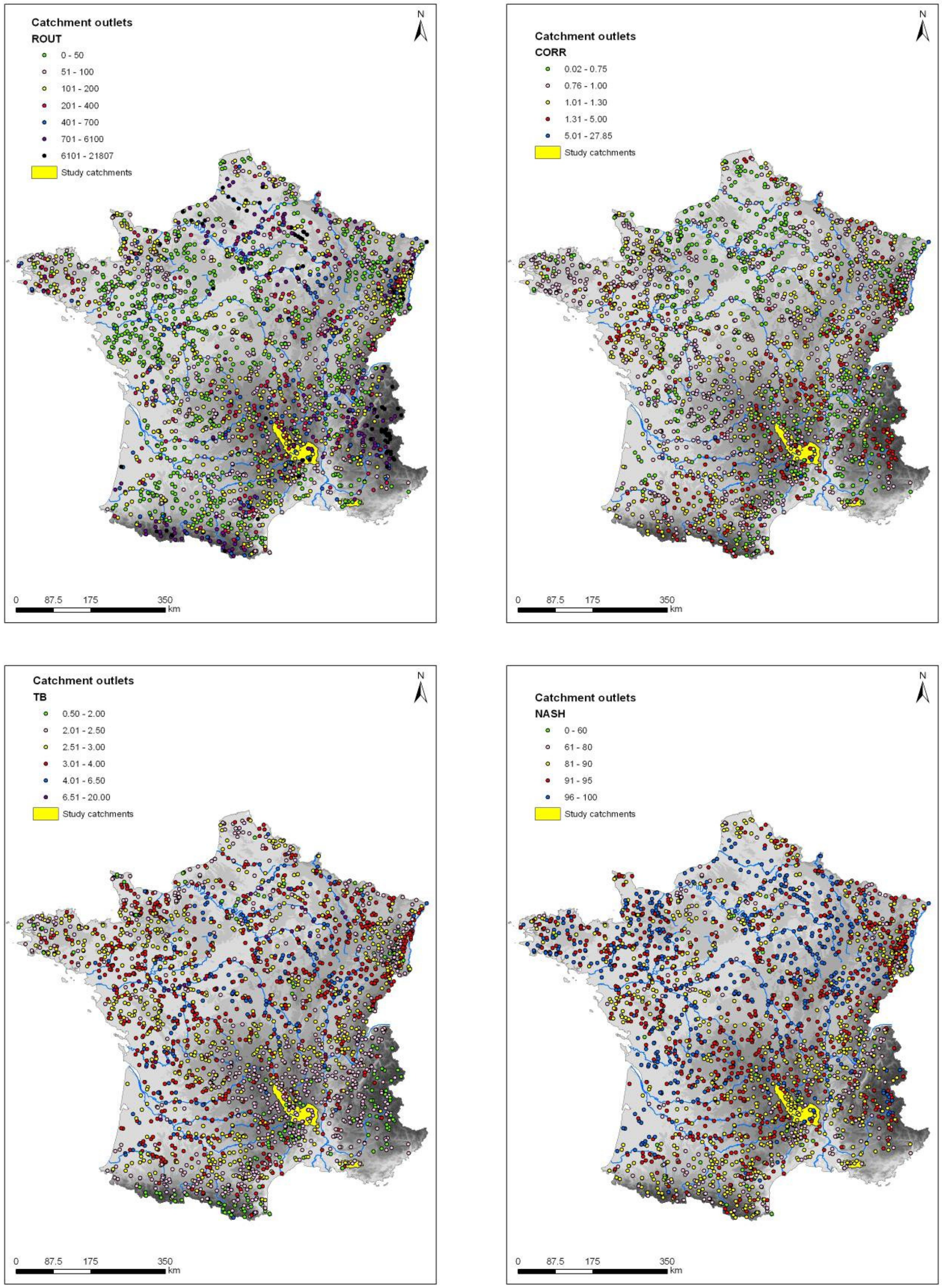


Figure 16: Distribution of GRP model parameters and NASH values in 3070 catchments in France

From the parameters calibrated for 3070 catchments, the possible ranges of parameters are defined and shown in Table 4. These ranges of parameters are much larger than those of the three study catchments; the values of parameters for the study catchments are located in the low value part of the parameter domains. Therefore it can be assumed that those ranges are sufficiently large for the catchments of interest.

Table 4: Ranges of parameter values in GRP model estimated from the calibration for 3070 catchments in France

Parameter	Minimum value	Maximum value	Ranges in 3 study catchments	
ROUT	0.22	21807.29	30 - 250	[mm]
CORR	0.02	27.85	0.59 - 1.6	[]
TB	0.5	20	2.3 - 3.2	[days]

From the ranges of parameters obtained, a uniform distribution is chosen for all parameters and 125.000 combinations of parameters are randomly taken from those distribution. The uniform distribution is taken here because the real distribution of parameters is unknown; this was also recommended by Beven and Binley (1992) and was actually implemented in Larsbo and Jarvis (2005).

These parameter sets are used to run the GRP model for the 3 catchments. The control member of forecast precipitation and lead times up to 2 days are considered (Detail explanation in Section 3.2.5). The resulted NASH values for all 125.000 runs are made visualized in Figure 17. Among the 125.000 parameter sets, a large number gives very low NASH values. Few parameter sets result in up to 90% of NASH. The performance of the model in catchment Arc seems to be the worst as highest NASH barely goes up to 70%. This is understandable because this catchment has a shortest period of available data (about 14 years compared with 51 years of the other catchments).

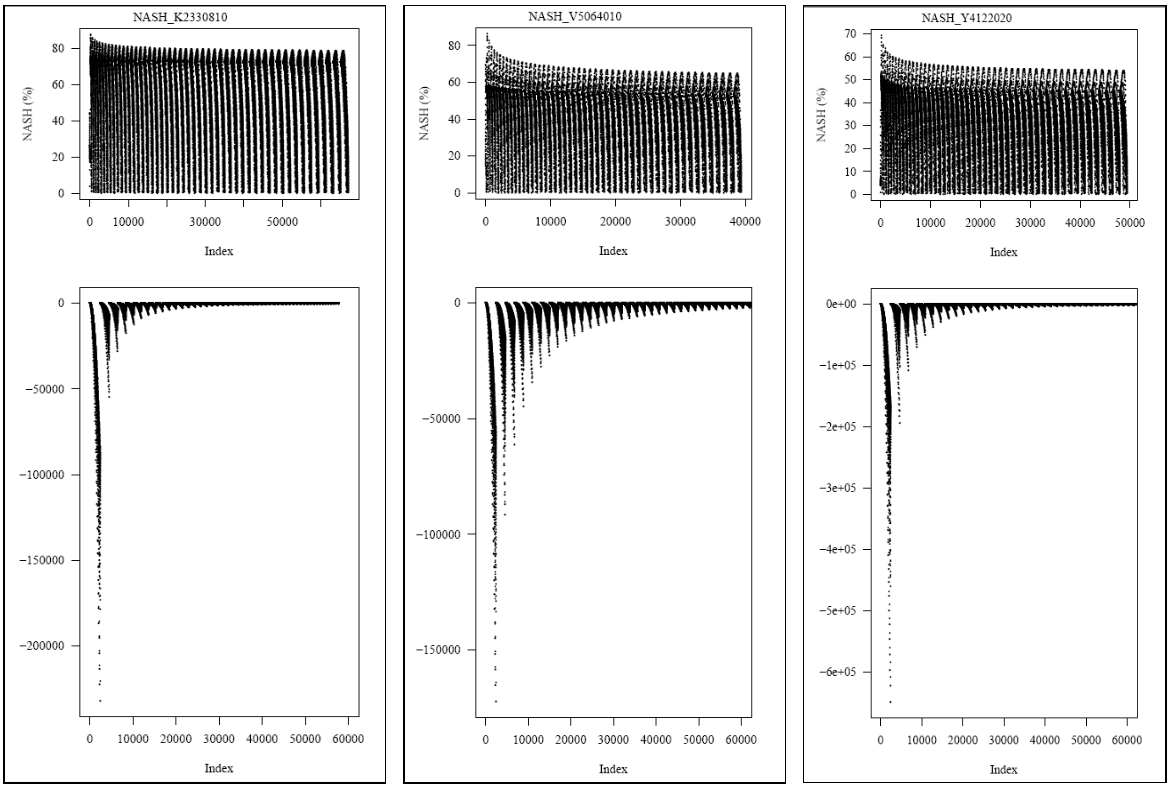


Figure 17: Result of NASH values when running 125000 sets of GRP model parameters for 3 study catchments (Lead time = 1 day)

The threshold to differentiate behavioural and non-behavioural parameters is often chosen subjectively to retain a proper number of simulations that can cover the observed discharges. Here at first the Nash-Sutcliffe threshold is chosen equal to 60% which results in the number of behavioural parameter sets shown in Table 5.

Table 5: Number of behavioural parameter sets with threshold of NASH =60% for three study catchments at lead times of 1 and 2 days

Catchment name	Number of behavioral runs (/125.000 runs)	
	Lead time = 1 day	Lead time = 2 day
Allier	32231	633
Ardèche	1645	97
Arc	31	3

Catchment Arc probably due to short data series has only 31 and 3 behavioural parameter sets for 1 and 2-day lead time, respectively. The GLUE approach therefore is only applied for catchments Allier and Ardèche. However, the number of behavioural parameter sets for 1 day lead time is too large for these catchments. A quick calculation shows that in order to run the experiments of this research for two catchments, it would take five days and 19GB of memory space of the computer. This is only for the parameterization uncertainty; if it is combined with other sources of uncertainty for example forecast precipitation uncertainty (with 11 to 51 members of forecast rainfall), and initial condition uncertainty (with 10 members of initial condition), it would take an enormous amount of time and space to store the results. On the other hand, the number of behavioural parameter sets for 2-day lead time is not too large, and, if we look at the performance of the model for lead time of 1 day, obtained by running the model with behavioural parameter sets for 2-day lead time (See Figure 18), the likelihoods locate in the high value part of the domain.

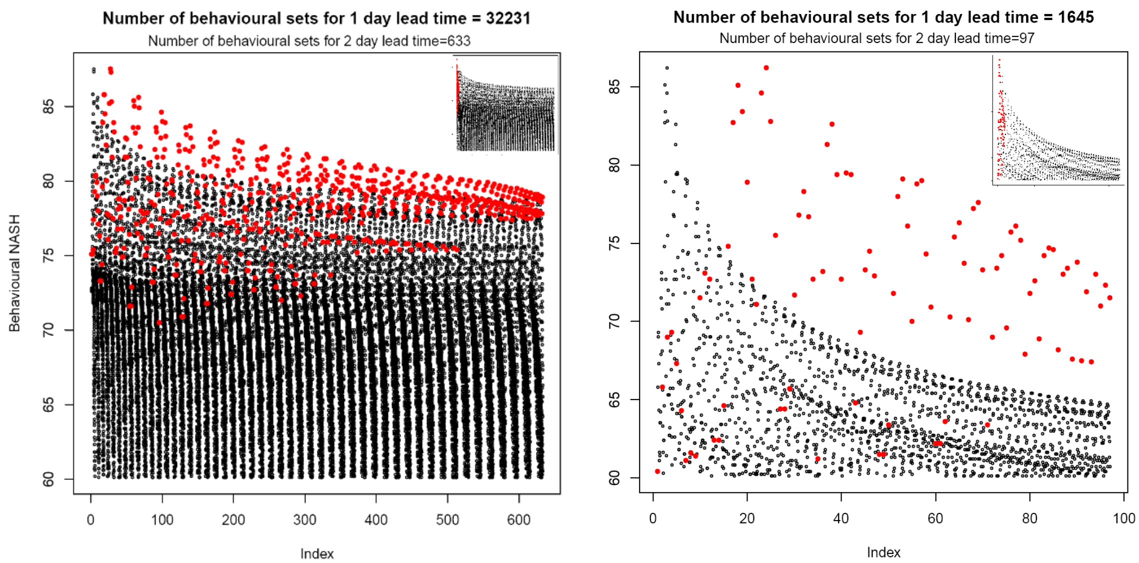


Figure 18: Likelihood values for lead time of 1 day with behavioural parameters set for lead time of 2 days (RED); and for lead time of 1 day with behavioural parameters set for lead time of 1 day (BLACK): Catchment Allier (left); Catchment Ardèche (right)

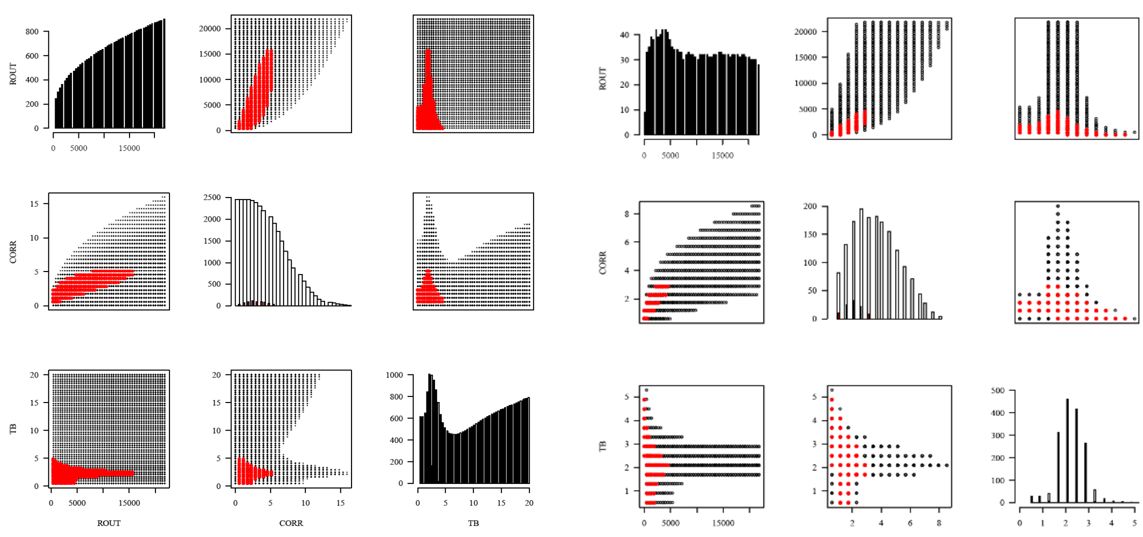


Figure 19: Parameter distributions for lead time of 1 day with behavioural parameters set for lead time of 2 days (RED); and for lead time of 1 day with behavioural parameters set for lead time of 1 day (BLACK): Catchment Allier (left); Catchment Ardèche (right)

Besides, looking at the values of the three parameters, the behavioural parameters for 2-day lead time (in red color) locate entirely in the domain of behavioural parameters for 1-day lead time (in black color) (Figure 19).

It seems then acceptable to use the parameters that highly perform for lead time of 2 days to forecast also for lead time of 1 day.

To better assess the effect on model performance when the model is optimized for a given lead time and applied for other lead times, the performance of the model with parameters optimized for each specific lead time, from 1 to 10 days, for all three catchments, and the performance at those lead times with parameters optimized for lead time of 2 days are evaluated. Figure 20 shows the results obtained when considering the four performance criteria presented in Section 2.3.3. It can be seen that values of all four objective functions do not change too much. Therefore, it is reasonable to use the behavioural parameters for 2-day lead time to forecast for other lead times; and in this research the behavioural parameters for 2-day lead time are used to forecast for all lead times and for all tests later on.

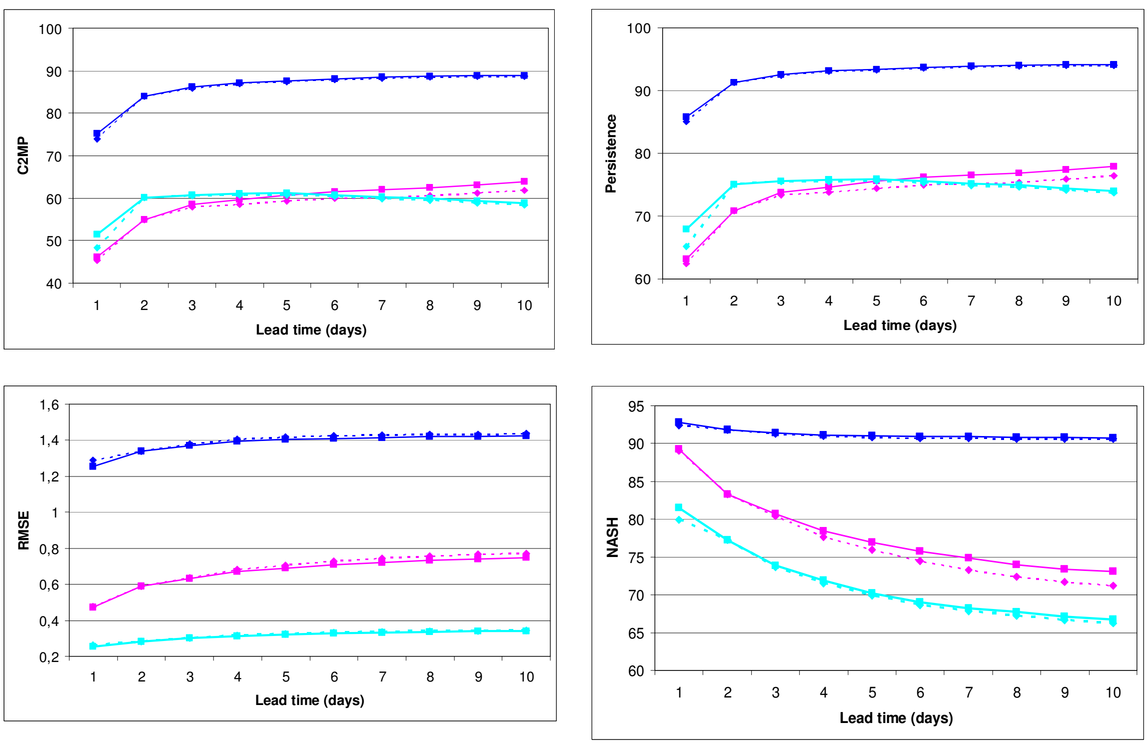


Figure 20: Performance of GRP model for lead times from 1 to 10 days with parameters calibrated for the corresponding lead time (dotted lines), or for 2 day-lead time (solid lines)

Also on Figure 19, it can be seen that ROUT and CORR parameters are clearly correlated as the scatter plots of those two parameters do not expand over the whole domain; the reverse pattern is seen for ROUT and TB. In catchment Allier, the scatter plot of CORR and TB presents a strange behaviour when values of TB tend to increase at the end of the domain, which could be an implication that after TB=20 days, there would be larger values of TB. However, this can be the result of a large number of parameter sets (32231) in which many give high likelihood values but the parameters are not physically realistic. In addition, a large number of parameter sets were implemented (125.000 sets) cover the large ranges of parameters can ensure that the result of behavioural parameters are reasonable. Moreover, shows that the typical parameter values for the study catchments are in the low domain of the parameter ranges. Therefore, the strange behaviour of TB values will not affect the result.

The likelihood weighted uncertainty bounds can be calculated using a standard procedure. First, the calculated behavioural likelihoods are rescaled to produce a cumulative sum of 1.0. A cumulative distribution function (CDF) of simulated discharges is then constructed using the rescaled weights.

For a given certainty level, here is chosen as 0.05, two quantiles of discharge are obtained by interpolating from the constructed CDF. They correspond, respectively, to the probability of 0.05 and 0.95. These two quantiles form the lower and upper prediction limit, respectively (Xiong and O'Connor, 2008). These limits form the 90% confidence bounds, meaning that 90% of the prediction is expected to fall in these bounds.

3.2.5 Individual uncertainty quantification tests

In order to assess the impact of the individual uncertainty sources, the corresponding components of the forecast system (See Figure 1) are changed to account for the considered uncertainty sources. The forecast driving force in all of these individual tests is the control forecast member of the ensemble forecast precipitations; recall that control member is the best estimation of the future precipitation (See Section 3.2.2). Since lead times up to 9 days are considered, the ensemble forecast precipitations from PEARP is used till lead time of 2 days, those from ECMWF is used for lead time from 3 days to 9 days. Five individual uncertainty assessment tests are shown in Figure 22 (from test 2 to test 6). In order to compare the individual impacts of these five sources of uncertainty, a reference test without any uncertainty is created in which the control forecast is used as forecast driving force and other components of the GRPE forecasting system is observed values (See Figure 22, Test 1).

3.3 Propagating uncertainty through the forecast system

Propagating the uncertainty through the forecast system helps to bring more information to the forecast, and therefore, may improve the forecast, making it more reliable. Looking at the forecasting process, if we consider the future being full of uncertainty where everything might happen, then, in this "naive" situation, the uncertainty in our forecast is 100% (Figure 21). But, if we consider additional information, the climatological data, which is the average of what happened in the past, then this uncertainty reduces. Then, if we consider a forecasting model to issue a forecast, then, when the forecast is performed, uncertainty is reduced considerably, as forecast is launched with the best and most updated knowledge on the future event. If the observation is assumed to be free of error, then, after observing the event, there is no uncertainty (future becomes certain). Otherwise, in the last case, if observations are not considered error free, there will be only the uncertainty of observation remaining.

From Figure 21, it can be seen that, the distance between forecast and observation, in term of remaining uncertainty, can still be large, according to the uncertainties that are not taken into account in the forecast. Propagating the uncertainties through the forecasting model can give information about the impact of those uncertainties on the forecast. If knowledge about the uncertainty is added in the forecast, it may reduce the forecast uncertainty and bring the forecast closer to the observation. In practice, however, since there are many different sources of uncertainty exist, it might happen that not all of them have a significant effect on flood forecasting. Hence, the implementation of an experimental uncertainty propagation, in which, different combinations of uncertainty sources are investigated, can be useful framework for the analysis of forecast uncertainty.

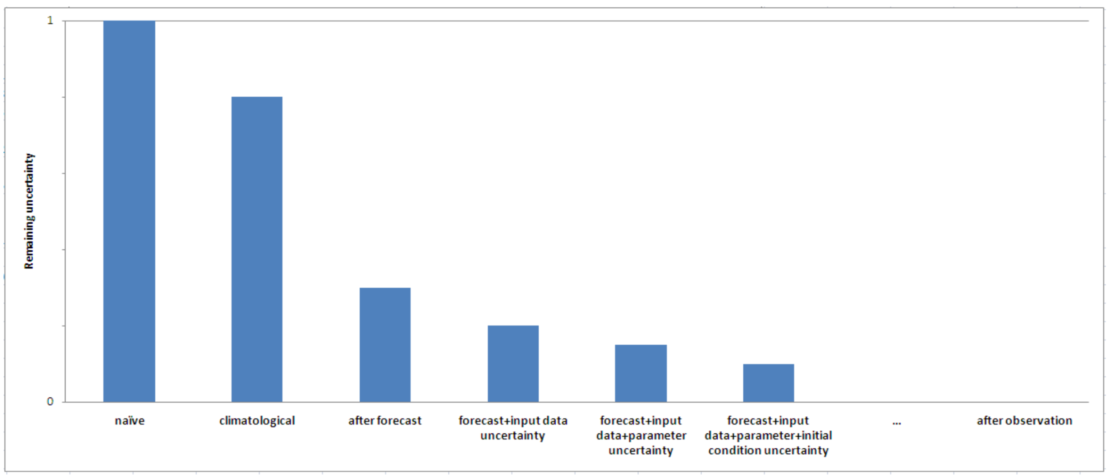


Figure 21: Remaining uncertainty of the forecast (adapted from Weijs et al., 2010)

In this research, a number of tests are performed to investigate the impact of uncertainty quantification from different sources on the quality of flood forecast. Table 6 gives an overview of all the tests performed. The data used for each test and the catchment to which the specific test is implemented are shown (More detail description of the tests is given in the Appendix). Because the availability of the data in the study catchments is not identical, not all of the tests are done for all study catchments. For example in Table 6, test 1, which employs the forecast precipitation from Meteo France and ECMWF, is done in all three catchments while test 10, using the spatially averaged precipitation to account for input precipitation uncertainty, is implemented only in catchment Ardèche.

Depending on the uncertainty test, the corresponding model components are modified. Those tests are illustrated on Figure 22. For example, one can see on this figure that in test 1 only the uncertainty of forecast precipitation is considered, hence all "basic" model components except the forecast precipitation are used, the input precipitation and initial condition in this case are the observed data, the parameters are calibrated using all of the data, the forecast precipitation from Meteo France (PEARP) is applied for lead times of 1 and 2 days and that from ECMWF is applied for lead times of 3 to 9 days.

Table 6: Uncertainty quantification and propagation tests implemented in this research

K: Catchment Allier

V: Catchment Ardèche

Y: Catchment Arc

Test	Uncertainty accounted for in the forecasts	Data used														
		Forecast precipitation			Ensemble Q from rating curve			Parameter sets from calibration periods			Parameter sets from GLUE method			Spatially averaged precipitation		
		K	V	Y	K	V	Y	K	V	Y	K	V	Y	K	V	Y
1	Control forecast															
2	Forecast precipitation															
3	Initial condition															
4	Calibration period															
5	Parameterization															
6	Input precipitation															
7	Forecast precipitation +Initial condition															
8	Forecast precipitation +Calibration period															
9	Forecast precipitation +Parameterization															
10	Forecast precipitation + Initial condition + Calibration period															
11	Forecast precipitation + Initial condition +Parameterization															

For each test, the reference discharges which are used for evaluation (comparison with the forecast) are the ten members of ensemble discharges from the rating curve analysis. These data are available for all three catchments in the same period of forecast (from 10/03/2005 to 31/07/2009). By doing so, we can avoid the error of discharge measurement if the observed discharge is used.

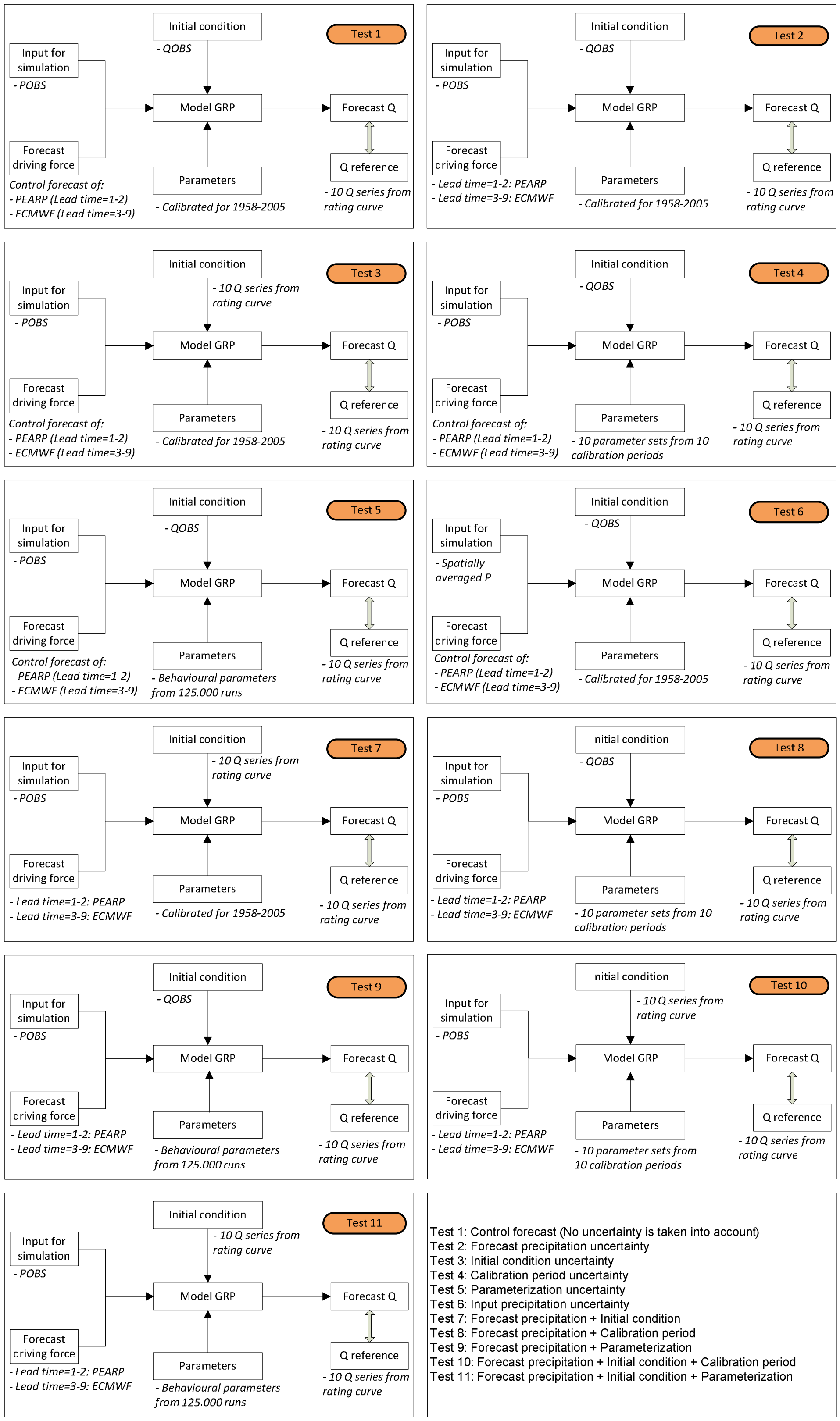


Figure 22: Illustration of changing GRP model components for uncertainty quantification and propagation tests

3.4 Evaluating the forecast outputs

It is necessary to find proper evaluation measures to assess the impact of uncertainty sources on flood forecast. In order to evaluate these impacts, these measures should be able to compare with the forecast with the reference ensemble discharge. The evaluation measures should be adapted to evaluate ensemble predictions and should consider the probability of forecast and its distribution. Based on those criteria, two probabilistic measures, reliability diagram and Brier score, are chosen to assess the quality of the forecasts produced by each test. In addition, the evaluation of confidence intervals is chosen to visualize and quantify the impact of uncertainty.

3.4.1 Reliability diagram

The reliability diagram is a probabilistic evaluation tool which compares the frequency of a reference with the probability of the forecast. Based on the analysis of the reliability diagram, one could conclude about the goodness of the forecast probabilities. A detail description of the reliability diagram is provided in Olsson and Lindstrom (2008).

The approach:

- Choose the event to consider ($Q > Q_{threshold}$), here, for each catchment, the discharge thresholds are chosen with the probability of 50%, 70%, 80%, 90%, 95%, 99. These threshold discharges with relevant to those probabilities are shown in Table 7;
- Calculate the forecast probability for each time step (forecast day) as the number of ensemble forecasts exceeding a given $Q_{threshold}$ divided by the total number of ensemble forecasts in that time step;
- Divide the forecasts into bins of probability categories; here, five bins (categories) are chosen 0-20%, 20%-40%, 40%-60%, 60%-80%, 80%-100%;
- Calculate the reference frequency: for each day, the reference frequency is either 1, if the reference discharge exceeds the threshold, or 0, if not.
- Plot the forecast probability and reference frequency on the x and y axis, each point of the graph representing for a category of probability.

Table 7: Threshold discharges for the 3 study catchments

Catchment name	Q thresholds [mm]					
	50%	**70%**	**80%**	**90%**	**95%**	**99%**
Allier	0.5920	0.9627	1.3442	1.8784	2.4097	4.2505
Ardèche	0.9655	1.8127	2.5272	4.2221	7.3328	19.8952
Arc	0.1230	0.1900	0.3188	0.6218	0.9272	2.8607

Reference discharges in flood forecasting are usually chosen as observed values (measured discharges or discharges obtained from measured water levels and a rating curve) and proxy-observed values (simulated discharges using observed precipitation to drive the simulation model). An example of a reliability diagram is illustrated on Figure 23. The diagonal represents the perfect match between the forecast probability and the observed frequency. The closer the diagram is to the diagonal, the more similar the forecast probability is to the observed frequency.

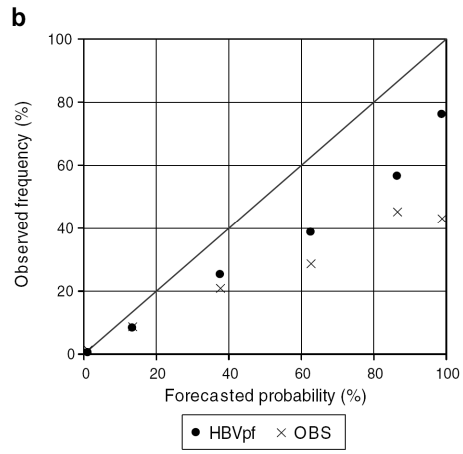


Figure 23: Example of reliability diagram where references are simulated (HBVpf) and observed (OBS) discharges (Olsson and Lindstrom, 2008)

Because the forecast is always uncertain, when adding the uncertainty from each test, it is expected that the forecast flow will be closer to the observed, with the members spread correctly around it. Hence, it is expected that the reliability diagram would be closer to the diagonal. The distances between the reliability diagrams and the diagonal indicate the relative importance of different uncertainties.

However, observed discharge is subject to the uncertainty of measure and of the rating curve, using an error-free reference might lead to incorrect comparison. Therefore, here the ensemble discharges retrieved from the rating curve are used as reference for all tests. Moreover, since in this research different tests will be implemented to evaluate different sources of uncertainty, the use of this reference is fixed for all tests.

Concerning the forecast probability, normally, in many studies using reliability diagram, authors often plot the forecast probability at the center of each bin (See Renner et al. (2009); Olsson and Lindstrom (2008)). However, that way of plotting would lead to an unwanted deviation from the diagonal like an example on Figure 24. Brocker et al. (2007) showed that it would increase the reliability of the reliability diagram if the forecast probability of one bin is calculated as the average probability of all forecasts that fall within that bin. This approach is be used here; for each probability category, the average forecast probability is then computed by taking the average of the values (1 or 0) over the days belonging to the probability category

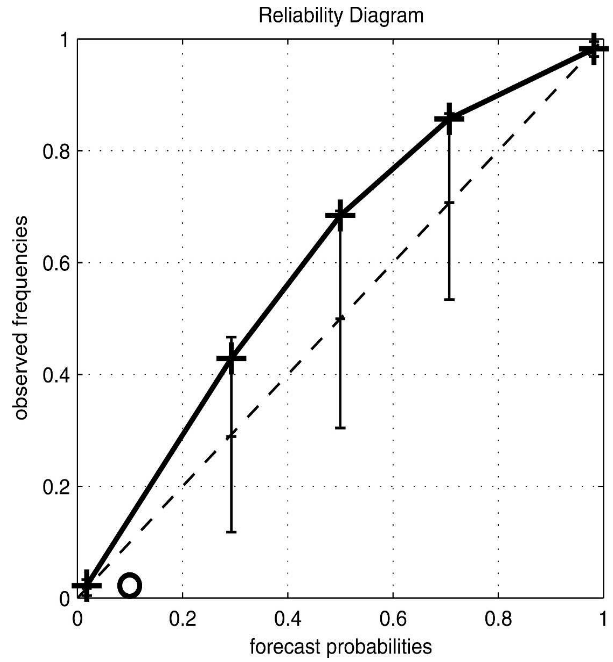


Figure 24: Plotting versus the bin centres would have caused substantial deviations from the diagonal (circle) (Brocker et al., 2007)

3.4.2 Brier Score

The Brier score (BS) is a measure that is often used to evaluate the quality of probabilistic forecasts. It evaluates the mean square difference between the forecast probability and the observed frequency. Franz and Houge (2011) recommended using this tool for evaluating the uncertainty in forecasting. The BS was proposed by Brier (1950):

$$BS = \frac{1}{N}\sum_{t=1}^{N}\sum_{i=1}^{R}(f_{ii} - o_{ii})^2 \qquad [5]$$

In which: t is time step; N is number of time steps; f is the forecast probability; o is the observed frequency; R is the number of possible classes in which the event can fall, for example: Rain / No rain (for rain) or Cold / Normal / Warm (for temperature). Here R is used as the number of bins to aggregate all of the forecast probability for all bins.

However, another version of BS which is more popular in application today is applied in this research (adapted from Thirel et al., 2010):

$$BS = \frac{1}{N}\sum_{t=1}^{N}(f_{ii} - o_{ii})^2 \qquad [6]$$

The Brier score can be decomposed into 3 additive components: Uncertainty, Reliability and Resolution (Murphy, 1973):

$BS = REL - RES + UNC$ [7]

$$BS = \frac{1}{N}\sum_{k=1}^{K} n_k (f_k - \overline{o}_k)^2 - \frac{1}{N}\sum_{k=1}^{K} n_k (\overline{o}_k - \overline{o})^2 + \overline{o}(1-\overline{o}) \quad [8]$$

Where: N is the number of time steps; K the number of probability bins.

$\overline{o} = \sum_{t=1}^{N} o / N$ is the observed climatological base rate for the event to occur

n_k is the number of forecasts with the probability category k

$\overline{o}_k$ is the observed frequency of the occurrence of events in bin k, given forecasts of probability f_k .

Uncertainty

The uncertainty term measures the inherent uncertainty in the event; this component is only dependent on the observed data not on the forecast. It should be noted that the "uncertainty" here has a different meaning with the uncertainty that is studied in this research. For binary events, it is at a maximum when the event occurs 50% of the time and the uncertainty is zero if the event always occurs (see). This uncertainty component of the BS is the average of events that are observed in the past, and forms the climatological based forecast (See Section 3.3)

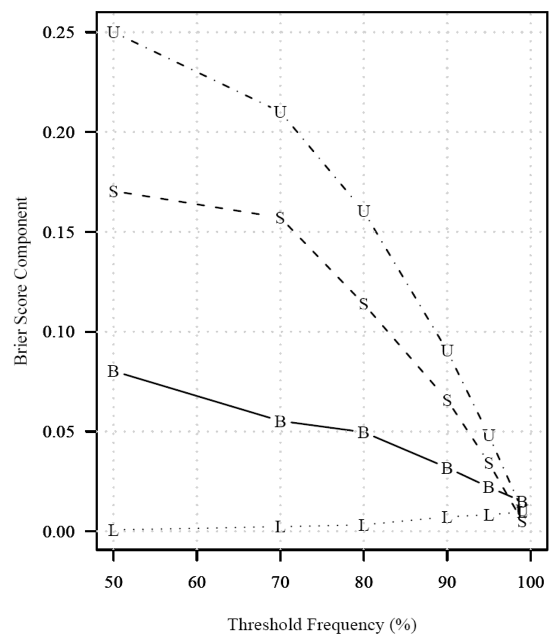


Figure 25: Illustration of the Brier score (B) and its three components: Uncertainty (U), Reliability (L), Resolution (S)

Reliability

The reliability term measures how close the forecast probabilities are to the true probabilities, given that forecast. It should be emphasized here that the reliability is defined in the contrary direction compared to English language. If the reliability is 0, the forecast is perfectly reliable. For example, if we group all forecast instances where 80% chance of rain was forecast, we get a perfect reliability only if it rained 4 out of 5 times after such a forecast was issued.

Resolution

The resolution term measures how much the conditional probabilities given the different forecasts differ from the climatic average. The higher this term is the better. In the worst case, when the climatic probability is always forecast, the resolution is zero. In the best case, when the conditional probabilities are zero and one, the resolution is equal to the uncertainty.

An example of the Brier score and its three components is given in Figure 25.

In order to evaluate the uncertainty, forecast flow should be compared with a reference flow; hereby, for all of the tests, the ensemble discharge retrieved from the rating curve is used as reference to avoid the uncertainty of observed discharge.

3.4.3 Confidence intervals

Confidence intervals are presented by bounds with a certain confidence level. They are a visualization measure which can help to evaluate the relative position between the forecast and the observed flow. If correctly evaluated, a predictive confidence interval should correspond to the proportion of observation falling within the forecast limits given by the confidence intervals. For example, on Figure 26, the 95% confidence intervals of the forecasts are shown , it means that 95% of the time, the observed flow is expected to fall between this interval.

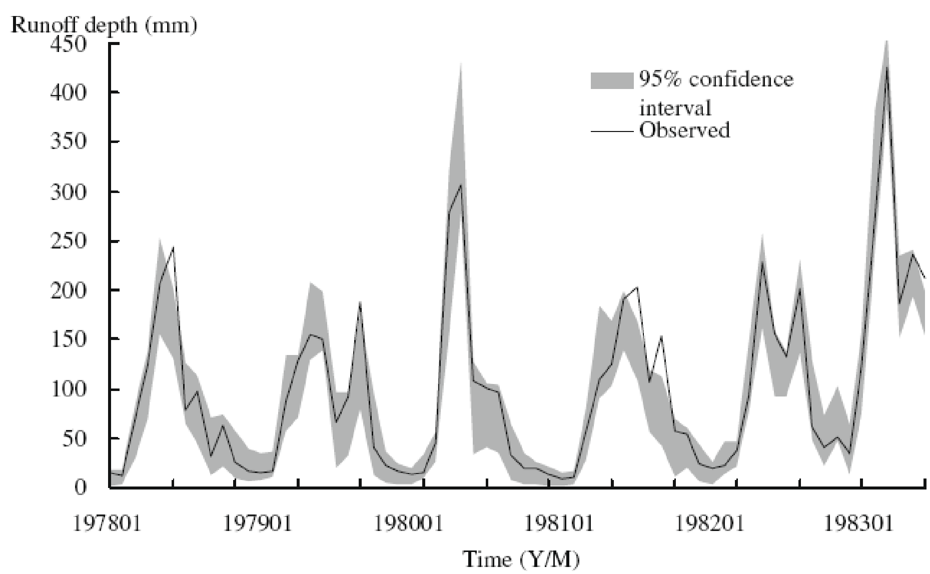


Figure 26: Example of the 95% confidence intervals of monthly runoff due to parameter uncertainty calculated by GLUE method (Jin et al., 2010)

4 Results and discussions

4.1 Individual impact of the uncertainty sources

The individual impacts of the uncertainty sources are quantified by accounting for that uncertainty in the corresponding component of the GRPE forecasting system. The control member of the ensemble prediction of precipitation is used as the forecast driving force. There are five uncertainty sources are taken into consideration which are forecast precipitation, initial condition, calibration period, parameterization, and input precipitation (Test 2 to 6). Due to the different availability of data in three study catchments, not all of the tests are implemented for all three catchments as explained in Table 6. In order to assess the relative impacts of the uncertainty sources on forecast discharge, the tests are compared with Test 1 which no uncertainty is taken into account. The evaluation tool used to assess these individual uncertainty sources here is the total Brier Scores (BS) aggregated for all threshold discharges; results are calculated for 9 lead times. The results are shown in Figure 27 to Figure 30, each diagram is the total BS calculated based on the forecast discharges and ensemble discharges from rating curve (See 3.4.2) when no uncertainty (Control forecast) or an individual uncertainty source is taken into account. The control forecasts which do not account for any uncertainty are shown in the first bar of the diagrams, when accounting for the individual uncertainty sources, it is expected that the forecasts would be more reliable, and therefore the total BS would be smaller.

In overall, for three catchments, the results show that forecast precipitation has a biggest impact on forecast discharge as the total BS of this test is the smallest. The impact of forecast precipitation is more pronounced at longer lead times due to the large influence of initial condition at small lead times (1 or 2 days). Initial condition uncertainty only and strongly affects the forecasts at small lead times (Figure 30); after that, it does not have impact on the predict forecast, as the total BS does not change compared with the Control forecast test.

Depending on the catchment (Allier or Ardèche), the uncertainty due to the calibration period is larger or smaller than the uncertainty due to the parameterization. The impact of these uncertainties is also stronger for higher lead times. Uncertainty of input precipitation is only quantified for catchment Ardèche due to the data availability but does not show impact on the forecast output (Figure 30). The total BS does not change significantly when accounting or not accounting for this uncertainty while other uncertainty sources show the impacts on the forecasts.

In conclusion, the intial condition uncertainty shows large impact at small lead times up to 2 days, after that forecast precipitation has most significant impact on forecast discharge. The relative impacts of uncertainty due to calibration period and parameterization depend on the catchment. Finally, accounting for the input precipitation uncertainty does not change the forecast reliability, therefore, this source of uncertainty is not taken into account when propagating different sources of uncertainty through the forecasting system.

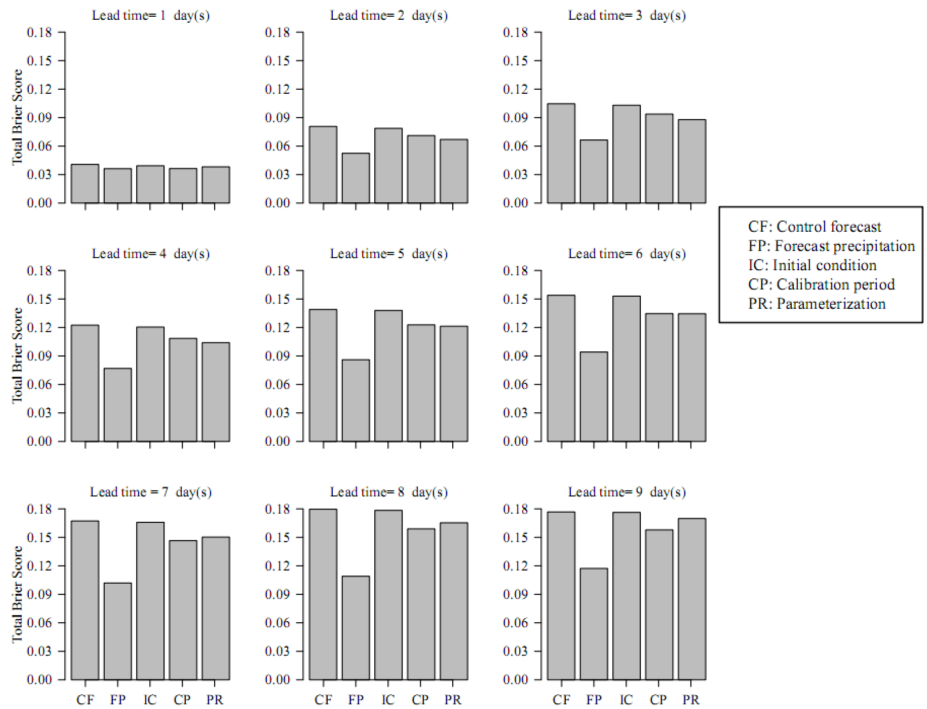


Figure 27: Total Brier Scores of the individual uncertainty tests for catchment Allier with lead times up to 9 days. Test with no uncertainty taken into account (Control forecast) and four sources of uncertainty are shown.

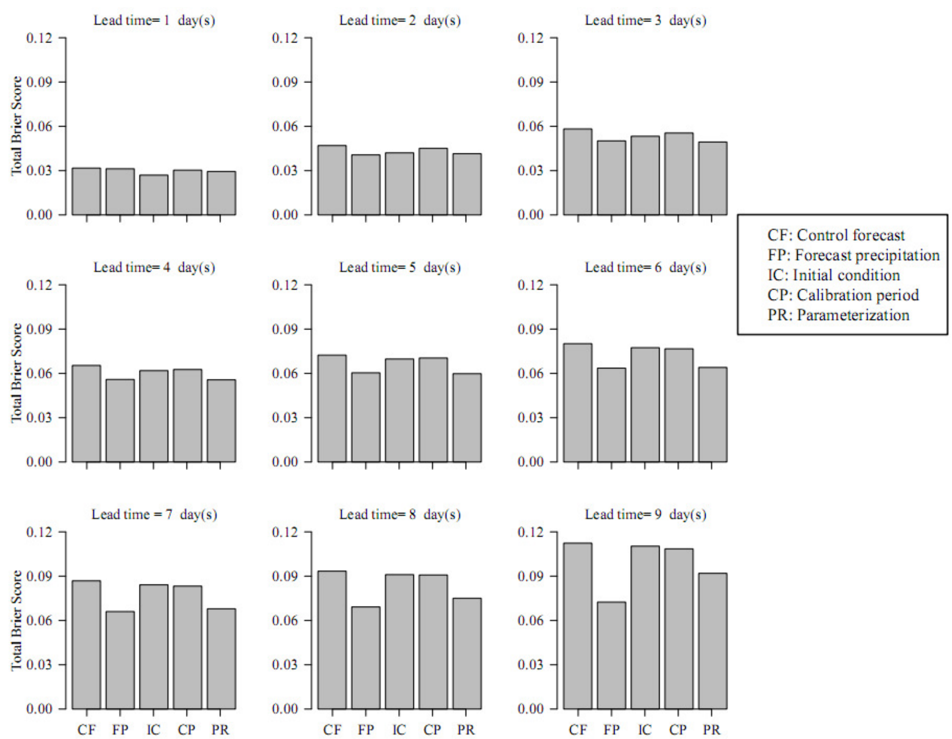


Figure 28: Total Brier Scores of the individual uncertainty tests for catchment Ardèche with lead times up to 9 days. Test with no uncertainty taken into account (Control forecast) and four sources of uncertainty are shown.

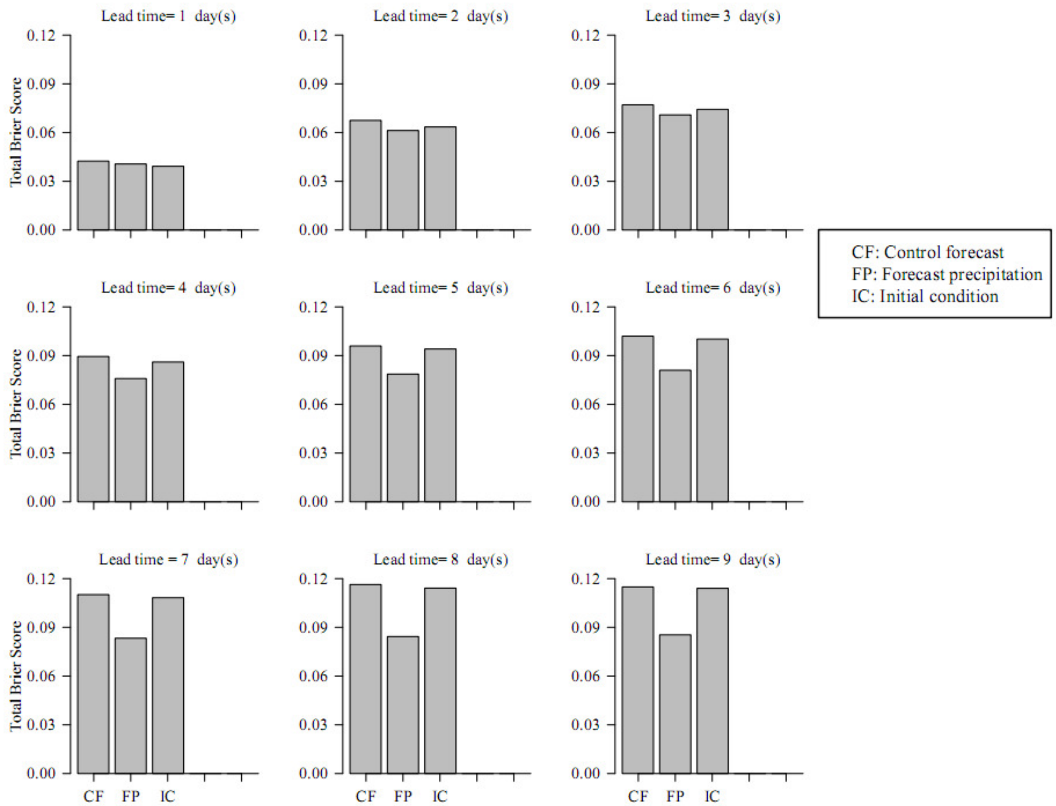


Figure 29: Total Brier Scores of the individual uncertainty tests for catchment Arc with lead times up to 9 days. Test with no uncertainty taken into account (Control forecast) and two sources of uncertainty are shown.

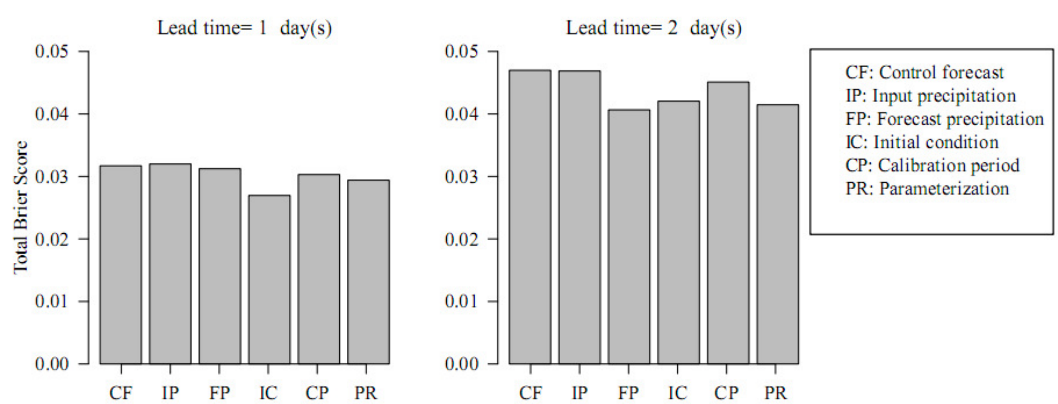


Figure 30: Total Brier Scores of the individual uncertainty tests for catchment Ardèche with lead times of 1 and 2 days. Test with no uncertainty taken into account (Control forecast) and five sources of uncertainty including the uncertainty from input precipitation are shown.

4.2 Combined impacts of the uncertainty sources

In overall, for all 3 catchments, for all tests, higher lead times result in lower quality of flood forecast showing through the increase of BS and its reliability component while the resolution decreases with lead times. However, the scores change quickly till lead time of 5 days and after that till 9 day lead time, it does not change too much (can be seen in Figure 31 and

Figure 38 in this Section). It means that the quality of forecast starts very good for small lead time and get worse by lead times but reach a sill after 5 days ahead; after that forecast quality stays the same. It should be noted that the BS tends to get better with higher flow (when threshold frequency increases). It is understandable because for higher flow thresholds, there is less non-zero probability of exceeding this threshold for both forecast and reference discharge so that the score of match between the forecast and reference is better.

4.2.1 Uncertainty of forecast precipitation and initial condition

The availability of data allows the uncertainty of forecast precipitation and initial condition can be quantified for all three study catchments up to lead time of 9 days. The uncertainty of forecast precipitation is quantified (test 2), using the ensemble forecast PEARP from Meteo France for lead times of 1 and 2 days, and the ensemble forecast from ECMWF for lead times from 3 to 9 days (See Section 3.2b). The forecast precipitation uncertainty is then combined with initial condition uncertainty in test 7, where the members of forecast precipitation (11 or 51 depending on the type of ensemble forecast used) combined with 10 members of discharges from the rating curve make 110 or 510 members of forecast discharge respectively. These members of forecast discharge are then evaluated against the references which are the ensemble discharges from the rating curve.

The Uncertainty component of BS is dependent on the climatological base, therefore, is does not change for all lead times, for all tests as the reference discharge used here is always the ensemble discharge retrieved from rating curve. Because this component does not tell any information about the forecast, it will not be shown here in all figures. There is a general increasing trend in BS with lead times; therefore the results are shown at some lead times not at all 9 lead times here.

Figure 31 shows the BS and its reliability and resolution components for catchment Ardèche from lead time of 1, 3, 5, 7 and 9 days; two results are shown where the uncertainty of forecast precipitation is accounted individually and together with initial condition uncertainty. It is seen that there is a difference between the BS values obtained between two tests for all three study catchments. After adding the 10 ensemble initial conditions to flood forecasting system, the BS becomes smaller than considering only ensemble forecast precipitation; visually it can be seen on Figure 31 that red line (forecast precipitation and initial condition) is always lower than black line (forecast precipitation).

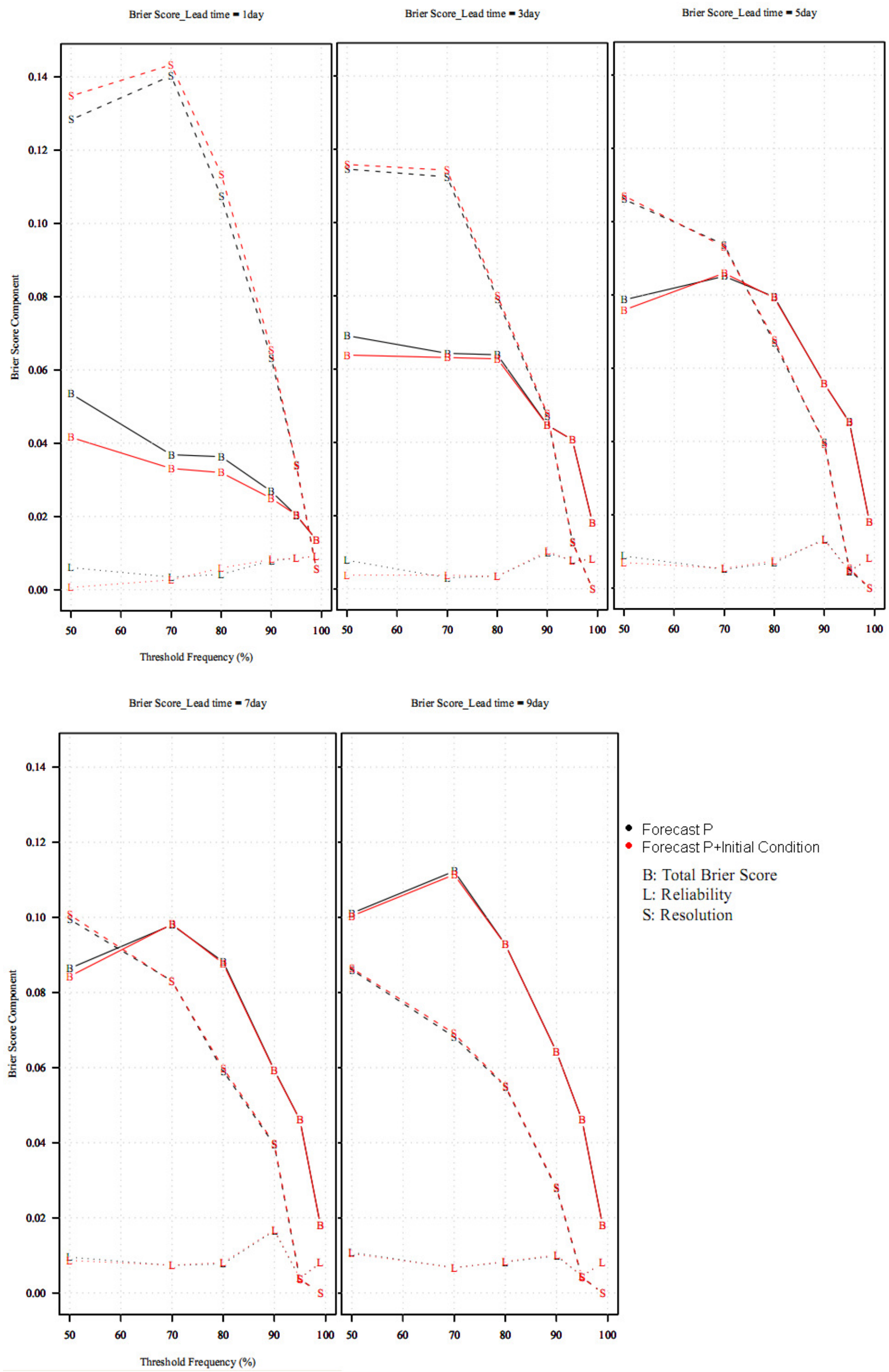


Figure 31: Brier scores at lead times 1, 3, 5, 7 and 9 day(s) when accounting for forecast precipitation uncertainty (black lines), forecast precipitation and initial condition uncertainty (red line) for catchment Ardèche

It shows that the quality of flow forecast is improved when accounting for both uncertainty sources than for only forecast precipitation uncertainty since the BS is reduced. Another point worth to comment is that the difference of BS values for the two sources of uncertainty tends to decrease with lead times. This means that relative difference brought by accounting for the uncertainty of forecast precipitation and initial condition is dependent on lead time; when lead time increases the difference between the two tests becomes insignificant.

The decrease in forecast performance with lead times that can be observed on the BS values is also acknowledged in the literature. He et al. (2009) studied the uncertainties with lead times ranged from 1 to 9 days; and the results from 7 weather forecast centres in Australia, Europe, UK, Canada, USA, China, Japan all show the clear decrease of skill scores when increasing lead times.

Although similar behaviour is found in all three study catchments, the significance is different. For catchment Allier and Arc, the difference between those two tests is not significant; example is given in Figure 32 while the red line (forecast precipitation + initial condition uncertainty) does not depart too much from the black one (only forecast precipitation uncertainty) even at small lead time (2 days).

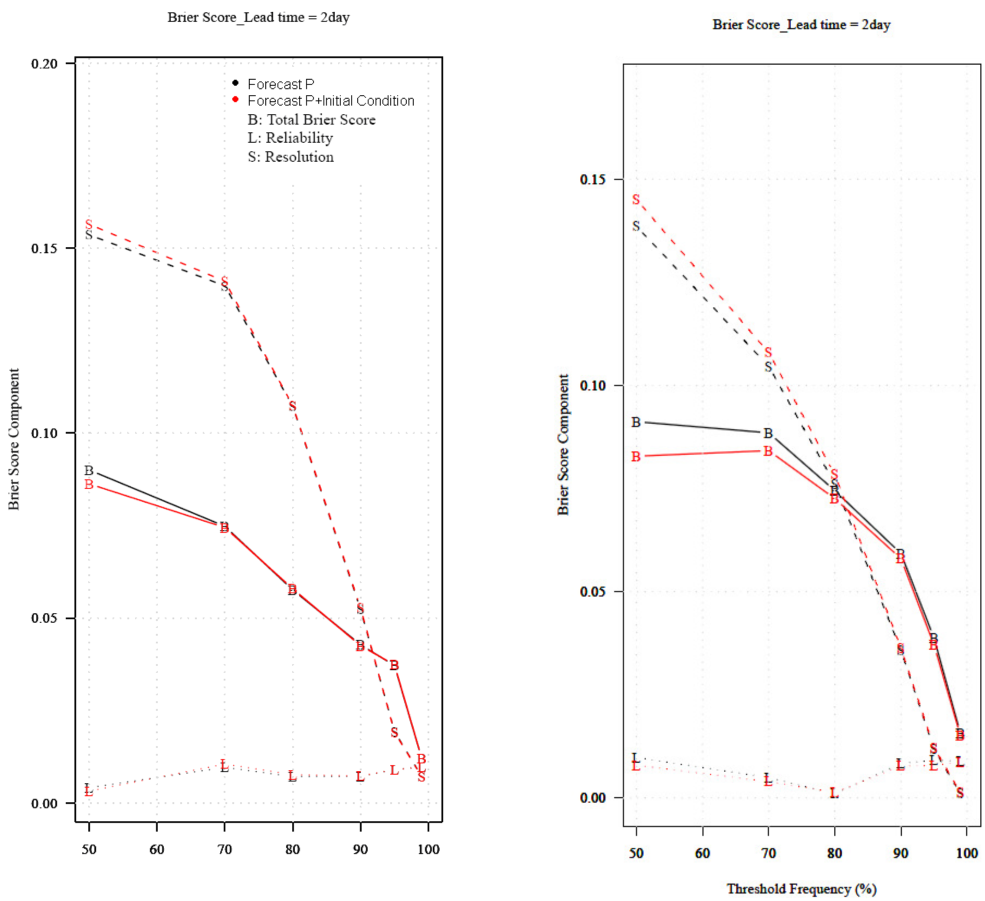


Figure 32: Brier scores at lead time of 2 days for catchment Allier and Arc

It seems that for catchment Ardèche, the uncertainty of initial condition shows the strongest effect among the three study catchments. It can be explained by looking at Figure 33 where the standard deviation of ensemble initial discharges around its mean values is shown for all 3 catchments; for catchment Ardèche, the initial discharge are largely varied with very large standard deviation. That leads to the larger difference when accounting for the uncertainty of initial condition for this catchment. Recall the flow characteristic of the catchments in , catchment Allier and Ardèche have almost the same surface area (2269 and 2240 km^2 respectively) but Ardèche has annual mean discharge twice of that of Allier. The higher values of discharge and probably of water level cause the larger uncertainty of the rating curve. For Arc, this is a small catchment compared with the other two (only 728 km^2), this catchment must respond very fast with and strongly be influenced by the external forcing, which is precipitation, so that the system is not sensitive to initial condition like for big catchments.

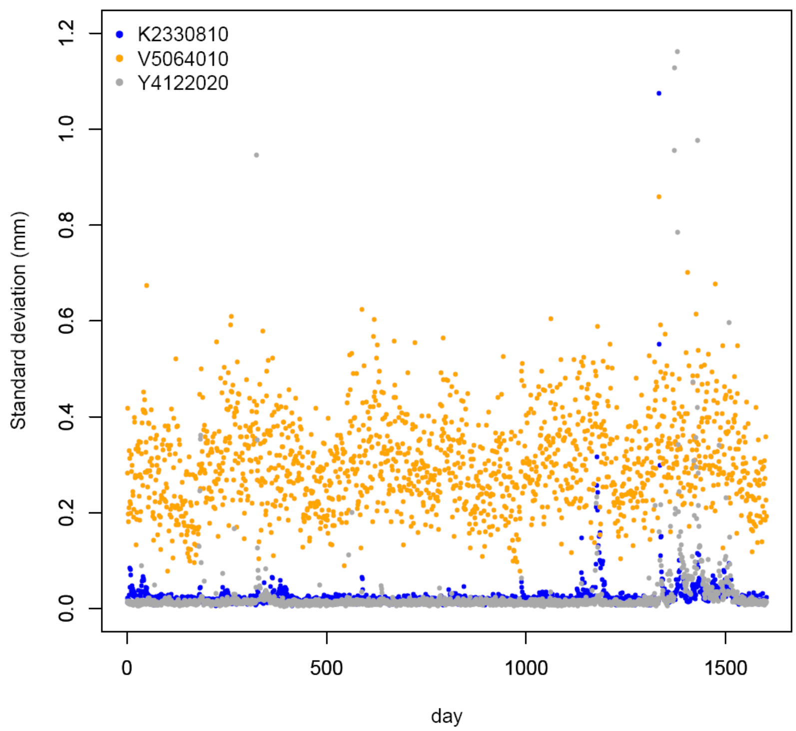


Figure 33: Standard deviation of ensemble discharges from rating curve for three study catchments: Allier (K2330101), Ardèche (V5064010), Arc (Y4122020)

The results agree with previous researches. In Rossa et al. (2010), the authors stated that the relevance of uncertainty in the precipitation field used as input into a hydrological model depends strongly on the size of the catchment. The larger the catchment, the stronger it acts to effectively filter the variations and uncertainties in the precipitation input. It might be the reason why in the small catchment like Arc, the impact of forecast precipitation uncertainty is more significantly pronounced than for big catchment like Ardèche.

The results of reliability diagram show the same trend, accounting for initial condition uncertainty together with forecast precipitation uncertainty makes the forecasts more reliable than accounting for only forecast precipitation uncertainty. On Figure 34, the reliability diagrams at lead times 1, 3, 5, 7 and 9 days for catchment Ardèche are shown, the forecast probabilities accounting for forecast precipitation individually and together with initial condition are plotted against the reference discharge frequency. It can be seen that the reliability diagram taking into account the two uncertainty sources approaches very close to the diagonal; and the largest effect is also observed in catchment Ardèche. The difference between the two tests is getting smaller as the lead time increases due to the broader ensemble forecast at high lead times. As can be seen on Figure 35, where an example of the flow forecasts taking into account the forecast precipitation uncertainty at lead times of 1, 4 and 9 days are shown, the range of the flow forecast becomes larger with higher lead times. In addition, the effect of initial condition is also degraded with time, so for high lead times, the uncertainty of initial condition does not affect significantly the forecasts making the difference between the two tests insignificant.

There is no obvious trend with flow thresholds that can be detected from the reliability diagram.

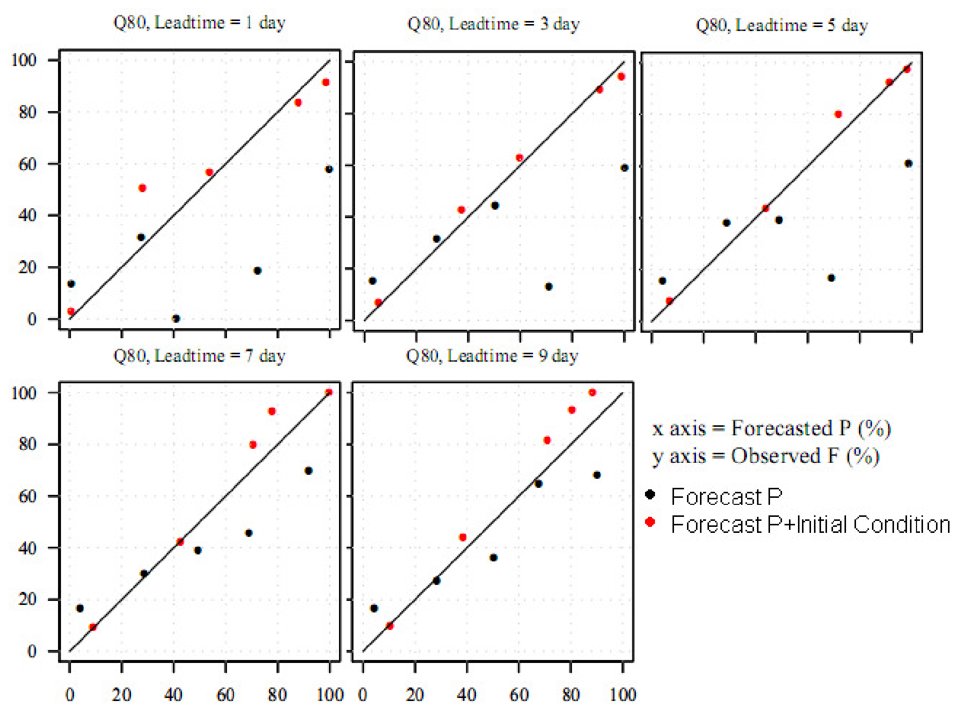


Figure 34: Reliability diagram for catchment Ardèche at lead time from 1 to 9 days exceeding Q80%

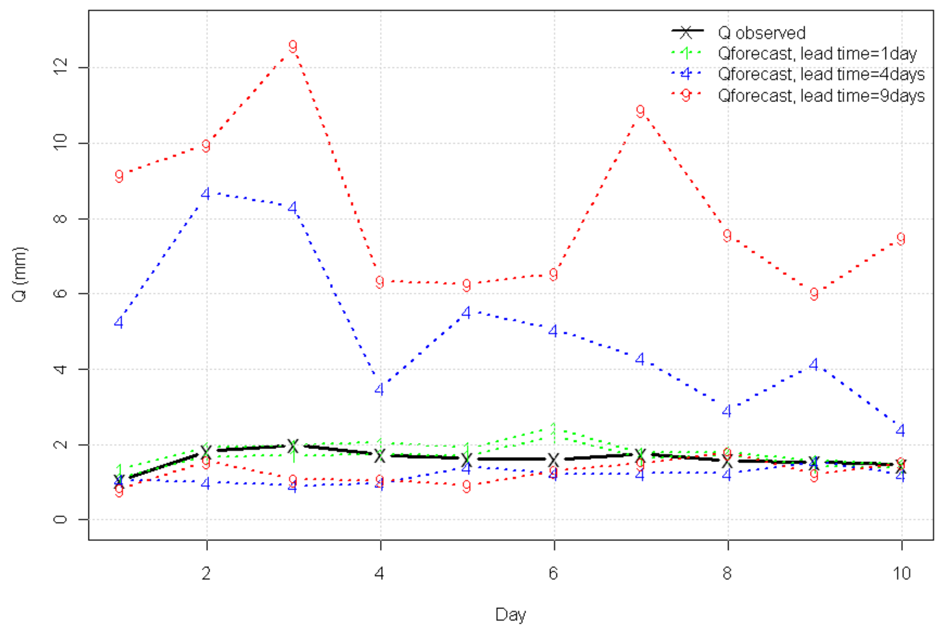


Figure 35: The upper and lower limits of forecast discharges changing with forecast precipitation at different lead times

In summary, taking into account the uncertainty of forecast precipitation together with initial condition makes the forecasts more reliable. The difference in impacts between the two tests decreases with lead times because of the larger uncertainty of forecast precipitation and the decrease in impact of initial condition at higher lead times. The trend of these uncertainties with threshold discharges is not obvious.

4.2.2 Uncertainty of model parameter

a. Uncertainty of parameterization and forecast precipitation

In Section 4 – Individual impact of uncertainty sources, when quantifying the uncertainty of parameterization only, the control member of ensemble forecast precipitation was used. In this section, in order to quantify the uncertainty of forecast precipitation and parameterization together, ensemble forecast of precipitation from PEARP and ECMWF is used. Applying the behavioural likelihood as defined in Section 3.2.4 (Nash values larger than 60%); it results in 633 behavioural parameter sets for catchment Allier and 97 for Ardèche.

With these numbers, it is only possible, due to computational time and memory constrains, to run with 11 forecast precipitation members from PEARP (633*11 and 97*11 simulations). However, with 51 members from ECMWF (633*51 and 97*51 simulations), it would be difficult, especially when these two uncertainties are propagated together with initial condition uncertainty (633*51*10 and 97*51*10 simulations). For that reason, the number of behavioural parameter sets need to be reduced.

One possible choice would be taking a higher threshold of likelihood value (for example NASH over 70% instead of 60%) to limit the number of simulations. However, look at Figure 36, the numbers of behavioural runs with NASH values over 70% is very small for both catchments: 40 for Allier and 18 for Ardèche. On the other hand, taking a different threshold of likelihood value would create differences with the previous results, and it would not be fair to compare all the tests. In the literature, reducing the behavioural parameter sets to limit the number of simulations has been done, for example in Larsbo and Jarvis (2005), 235 parameter sets were randomly selected from 30000 simulations when the number of behavioural simulations was over 235. In this research, therefore, for catchment Allier, 100 parameter sets are taken randomly from the distribution of 633 parameter sets' performance in Figure 36 (left). Since the number of behavioural simulations for catchment Ardèche is not too large (97 parameter sets), it is kept the same.

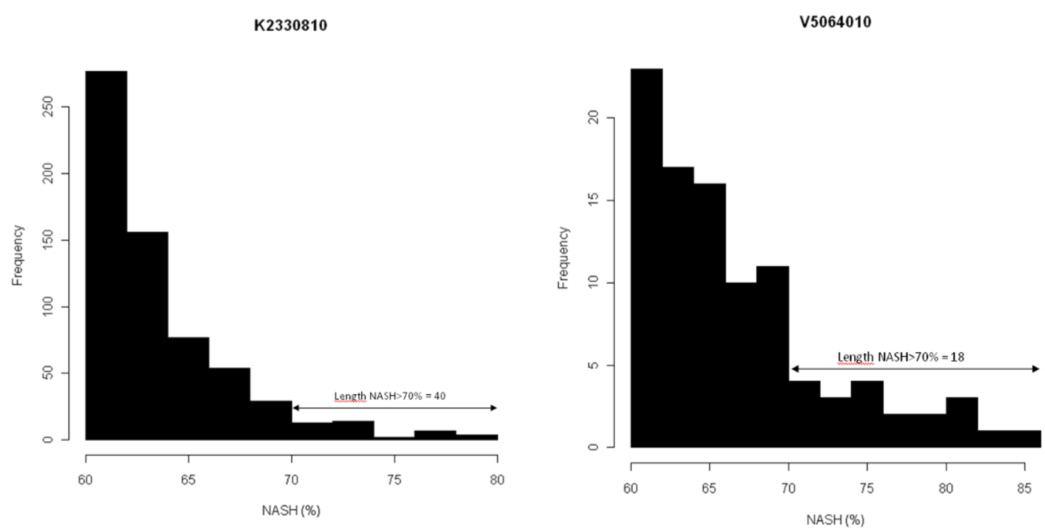
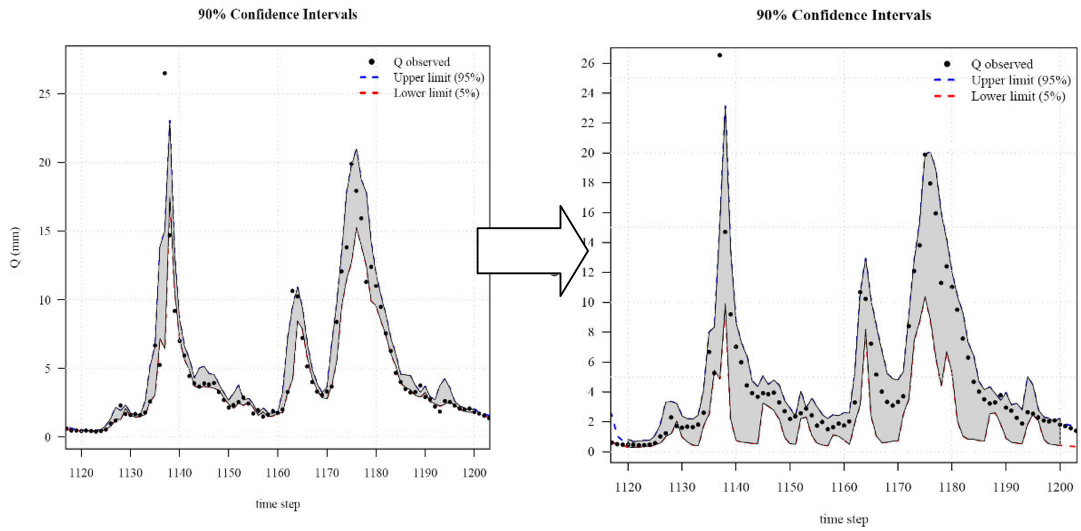


Figure 36: Histogram of NASH values for catchment Allier (left) and Ardèche (right) – performance of the model with 633 and 97 behavioural parameter sets respectively

These randomly selected parameter sets are used for forecasting with 51 members of the ECMWF ensemble precipitation prediction system. The numbers of behavioural parameter sets are kept the same (633 for catchment Allier) when forecasting with PEARP as this system consists of only 11 ensemble members of forecast precipitation.

Figure 37 shows the 90% confidence intervals for catchment Ardèche when uncertainty of parameterization is taken into consideration only or together with uncertainty of forecast precipitation. The results show that the confidence intervals become larger when adding the uncertainty of forecast precipitation together with that of parameterization.



With only the uncertainty of parameterization

With the uncertainty of parameterization and forecast precipitation

Figure 37: 90% forecast confidence intervals for catchment Ardèche: a large flood event from 21/10/2008 to 20/11/2008 (Lead time = 1 day)

Table 8: The number of observed discharge falling in 90% confidence intervals of forecast when adding uncertainty of parameterization and forecast precipitation

Lead time (days)	Number of runs (number of parameter sets x number of forecast precipitation)	% Q observed fall within 90% forecast confidence intervals	% Q observed exceeding Q90% fall within 90% forecast confidence intervals	Number of runs (number of parameter sets x number of forecast precipitation)	% Q observed fall within 90% forecast confidence intervals	% Q observed exceeding Q90% fall within 90% forecast confidence intervals
	Allier			**Ardèche**		
1	6963 (633x11)	82.0	100	1067 (97x11)	67.9	77.0
2	6963 (633x11)	60.0	82.0	1067 (97x11)	73.1	79.5
3	5100(100x51)	55.2	73.6	4947 (97*51)	77.2	75.5
4	5100(100x51)	55.2	77.8	4947 (97*51)	83.0	67.0
5	5100(100x51)	55.6	84.7	4947 (97*51)	84.0	62.3
6	5100(100x51)	55.3	83.3	4947 (97*51)	84.1	60.4
7	5100(100x51)	56.0	83.3	4947 (97*51)	84.0	57.5
8	5100(100x51)	54.1	88.7	4947 (97*51)	84.1	58.5
9	5100(100x51)	54.4	85.7	4947 (97*51)	84.6	57.5

The percentage of observed discharge falling in the 90% confidence intervals of forecast and the percentage of observed discharge exceeding Q90% falling in the intervals is shown in Table 8. It can be seen that the percentage is large at lead times of 1 and 2 days; but it is getting smaller at higher lead times for catchment Allier. This is probably due to the reduction of simulations that affects the outcome; when reducing the number of behavioural parameter sets from 633 to 100, the upper and lower limits of the forecast might be also affected. From this, it can be seen that the GLUE method is very much dependant on the parameter sets that are used.

b. Uncertainty of calibration period and forecast precipitation

The impact of calibration period uncertainty is quantified by taking different calibration periods within the period that calibration data is available. Due to the lack of data in one catchment, this is only done for two catchments Allier and Ardèche.

In overall, the resulted BS shows same trend for all study catchments when accounting for the uncertainty of calibration period, forecast precipitation and initial condition; the forecast quality is getting lower when lead time increases (BS increases, see Figure 38). On Figure 38, the BS and its components for catchment Allier when accounting for forecast precipitation uncertainty (Test 2), forecast precipitation and initial condition uncertainty (Test 7), forecast precipitation and calibration period uncertainty (Test 8) are shown at lead times of 1, 3, 5, 7 and 9 day(s). It can be seen that accounting for the uncertainty of forecast precipitation and calibration period results in the most reliable forecasts among those three tests. The significance of the difference among three tests becomes larger with lead times, showing that at higher lead times, accounting for calibration period uncertainty can largely improve the forecasts.

For catchment Ardèche, the impact of calibration period uncertainty on the forecast is not as large as for catchment Allier, implying the site independence of the impact of uncertainty.

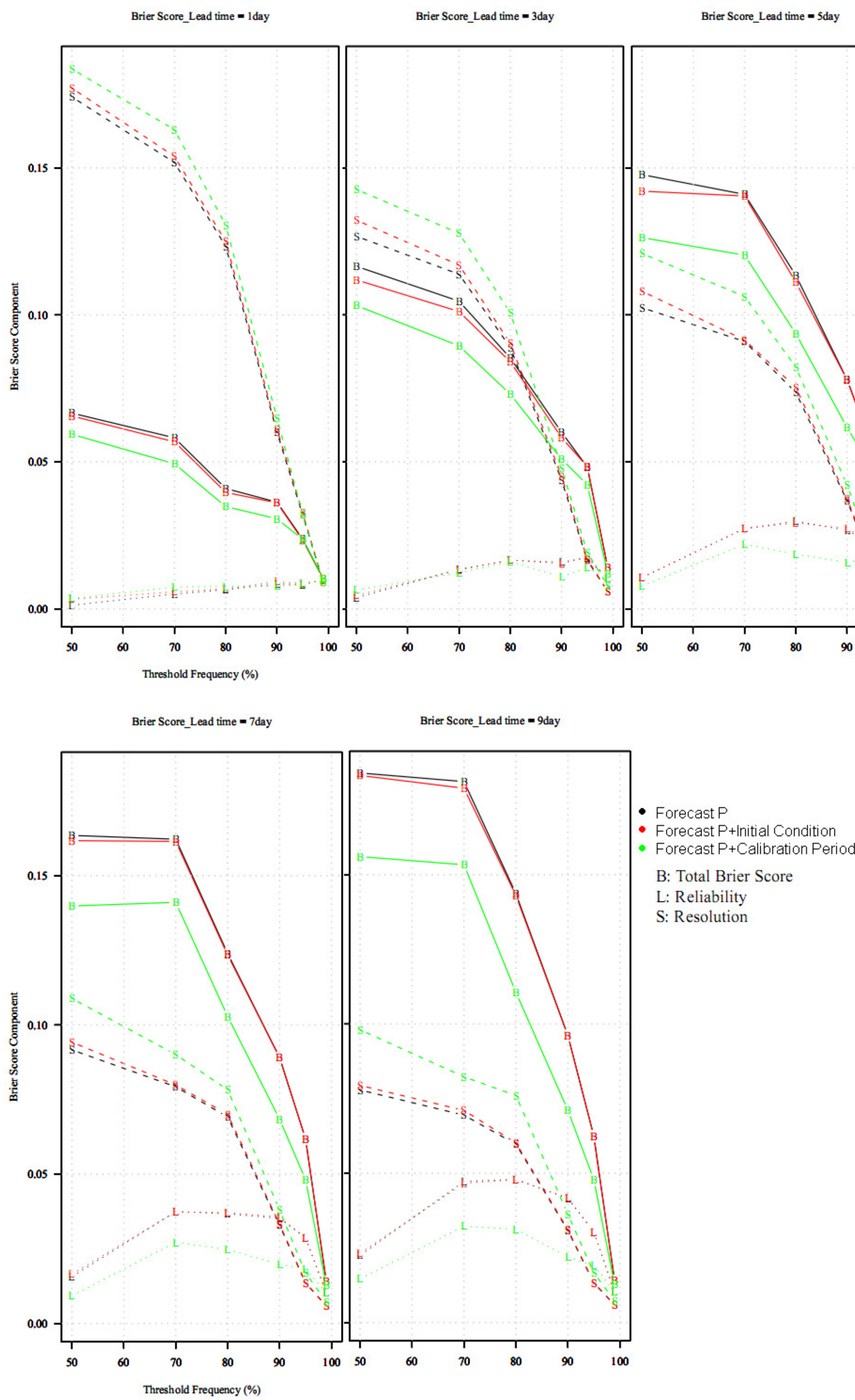


Figure 38: Brier scores at lead times 1 day, 3, 5, 7 and 9 days for catchment Allier

4.2.3 Accounting for all sources of uncertainty

After quantifying the impact of individual uncertainties on flood forecasting, the input precipitation is found to have almost no effect. The forecast precipitation, initial condition, parameterization and calibration period uncertainty shows significant impacts. Therefore, those uncertainties are propagated through the forecasting model for two catchments Allier and Ardèche.

Propagating the uncertainty of calibration period and parameterization together through the model might lead to an over estimation of uncertainty. Therefore, the uncertainty propagation is done with two different combinations of uncertainty sources: (1) forecast precipitation, initial condition, parameterization, and (2) forecast precipitation, initial condition, calibration period uncertainty. By doing this way, the impact of calibration period and parameterization uncertainty can be compared.

However, the result of parameterization uncertainty from GLUE method appears in the shape of confidence intervals which are drawn from a large number of ensemble discharge while the result from other uncertainties are considered with the reliability diagram and BS. It would be difficult to compare those uncertainties. On the other hand, the multiplication propagation would combine 51 members of ECMWF precipitation forecast, 10 members of initial conditions and 100 members of parameter sets, it would make 51000 simulations and would cost enormous amount of time to work on that. Therefore, for propagating the uncertainty of parameterization, 10 parameter sets are chosen randomly from the distribution of behavioural likelihood for both two catchments so that it is possible to compare to combinations of uncertainties with the reliability diagram and BS.

Figure 39 shows the BS values for both two catchments at lead time of 7 days. The uncertainty sources taken into account are forecast precipitation + initial condition, forecast precipitation + calibration period, forecast precipitation + initial condition + calibration period, forecast precipitation + initial condition + parameterization. The results show difference behaviour for two catchments. Combination 2 (forecast precipitation, initial condition, calibration period uncertainty) has larger impact on the forecast probability in catchment Allier, when propagating these uncertainties through the forecast, BS becomes the smallest. Combination 1 (forecast precipitation, initial condition, parameterization) does not show good impact for catchment Allier; taking into account these uncertainties makes the forecast quality becomes even worse than accounting for only 2 uncertainties (forecast precipitation, initial condition).

Meanwhile, for catchment Ardèche, a reversed trend is observed, propagating combination 1 through the forecast improves the forecast quality while combination 2 is not as good.

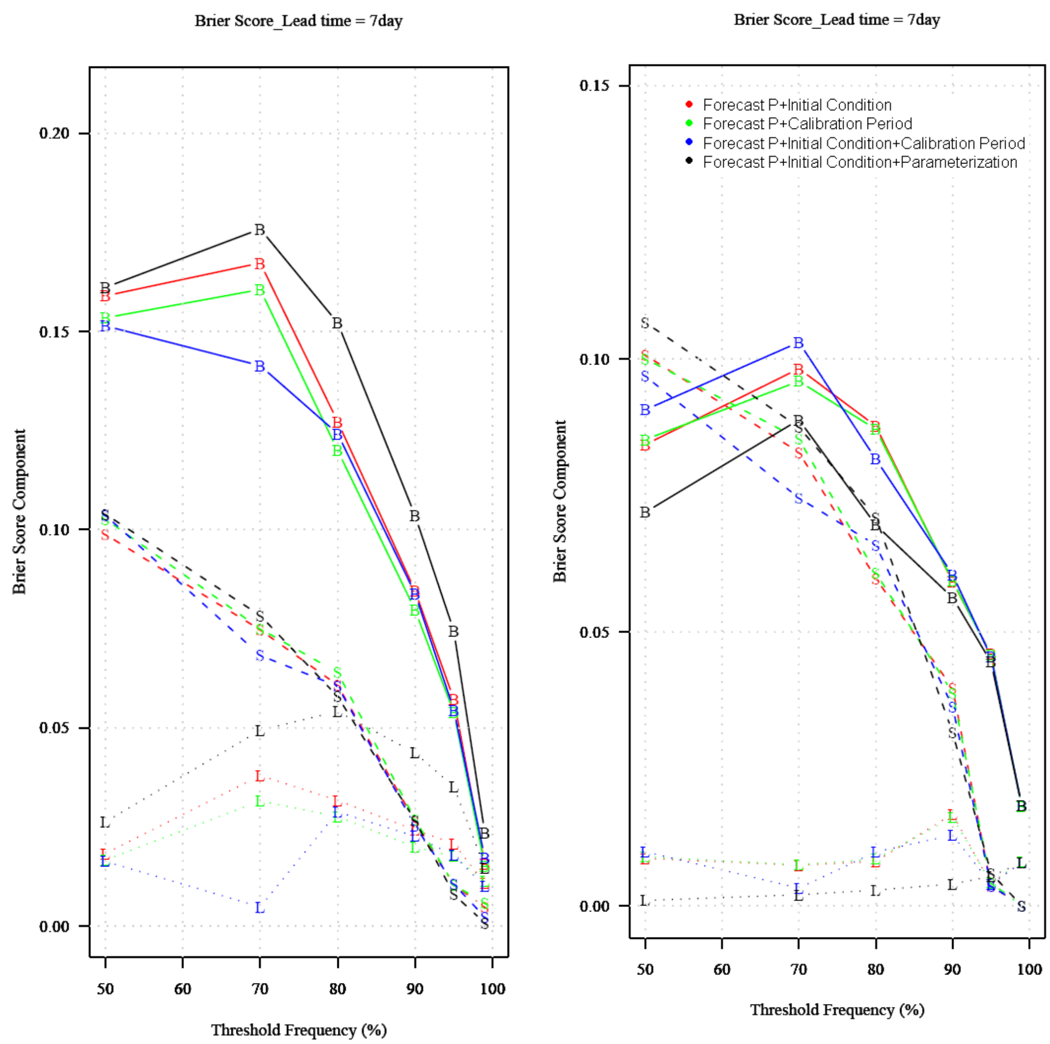


Figure 39: Brier score for catchment Allier (left) and Ardèche (right) at lead time of 7 days

The same trend can be seen on reliability diagrams. For catchment Allier (Figure 40), when accounting for uncertainties from forecast precipitation, initial condition and calibration period, the points come closest to the diagonal which means forecast probability comes closest to observed frequency. Most of the forecasts are over-estimated as the points are located under the diagonal to the side of the forecast probability.

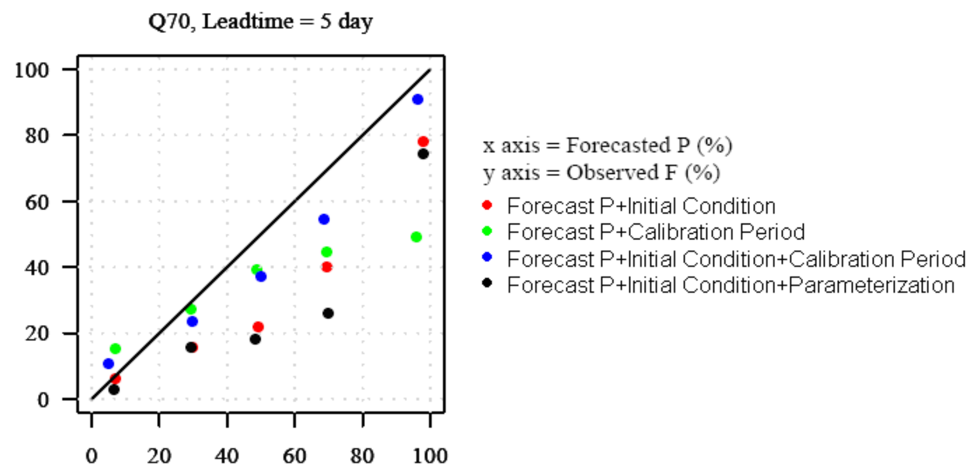


Figure 40: Reliability diagram for catchment Allier at lead time of 5 days exceeding Q70%

For catchment Ardèche, the impact of accounting for uncertainties from forecast precipitation, initial condition and parameterization is significantly the largest. As seen on Figure 41, when accounting for those uncertainties, the points (black) almost stay on the diagonal. The results are not the same for all lead times and all threshold discharge but same trend is always observed.

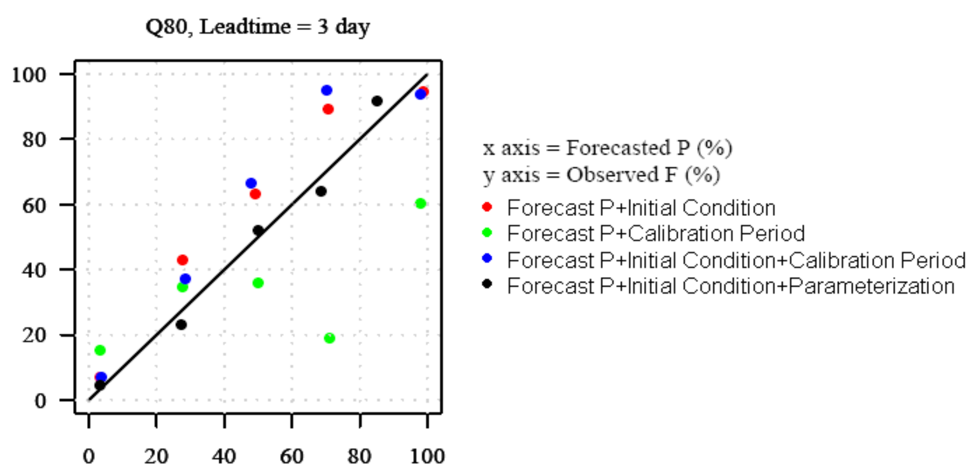


Figure 41: Reliability diagram for catchment Ardèche at lead time of 3 days exceeding Q80%

The bad performance of the forecast in catchment Allier when uncertainties from forecast precipitation, initial condition and parameterization are propagated into the forecast might come from the limitation of the parameter sets.

4.3 Discussions

The method used in this research, consisting of experimentally propagating different uncertainty sources through a flood forecasting system by multiplying the individual sources, shows some advantages comparatively to other methods reported in the literature (for example, the stochastic method proposed by Hostache et al. (2011) which quantifies the total predictive uncertainty without isolating individual sources of uncertainty, or the Bayesian approach proposed by Kavetski et al. (2006) which relies on the error models of the uncertainty source of interest, which is barely known). With the method used in this research, the individual uncertainty sources are well isolated and the user can choose which uncertainty sources to be propagated through the system. Additionally, if one method of uncertainty quantification fails to capture the uncertainties in the source been considered, the user can easily replace it by another method and test once more its propagation through the system. For example, the results of this research show that the input precipitation uncertainty does not have any impact on discharge forecast outcome. It might not be because of the small uncertainty of input precipitation but because of the method that is used to quantify this source of uncertainty cannot appropriately capture the uncertainties. With the flexible multiplying propagation method used here, one might change the method of input precipitation and the same process can be done.

The GLUE approach used to assess the uncertainty of parameterization has the advantage of being able to consider the interdependence among the model parameters. However, this method is time-consuming and requires high amounts of computational memory. The alternative solution of applying a limited number of simulations has been reported in the literature, but, in this research, it resulted in a decrease of the forecast performance when the number of runs was reduced. This may be the reason why, in one study catchment, the impact of GLUE parameterization was smaller than that of the parameter uncertainty evaluated with different calibration periods. However, it is just an assumption and further investigation needs to be done to confirm this finding.

The probabilistic evaluation tools used in this research, the Brier Score and the reliability diagram, showed high capability to assess the forecast outputs and proved to be useful tools to compare the different steps taken in uncertainty quantification and propagation. These tools are able to assess most of the desirable flow forecast properties (reliability and resolution). However, attention must be paid to the fact that the results are dependent on the number of forecasts exceeding each threshold considered in the computation of these statistical measures or falling in each probability bin considered in the evaluation of the reliability diagram. The scope of the data used should be large enough to have proper results. Moreover, the interpretation of the outcomes is sometimes not straight forward; for example, the results of Brier Score and reliability diagram show that the forecasts are more reliable with high flow which should not be the case and is the consequences of the small probability of exceeding the high flows of reference as well as forecast discharge.

Finally, the approach that is applied here for a specific forecasting system on some study areas but it can be also applied for other systems and areas. This approach is not limited in only the field of forecasting, it can be used also to quantify the uncertainty affecting other systems, such as hydrologic simulation. In that case, the components of the system are still the same except the absence of forecast precipitation.

5 Conclusions and recommendations

Uncertainties will always exist and are unavoidable in flood forecasting. Accounting for uncertainties can considerably improve the quality of the forecasts. Because the flood forecasting system is a complex system, many sources of uncertainty, which come from different components of the system, can propagate through the system and affect the quality of its forecasts. Quantifying all sources of uncertainty might be a fastidious task and unnecessary as some of the uncertainty sources may have almost negligible impacts on the forecast output. Furthermore, if uncertainties are not correctly quantified and propagated through the system, it might lead to an over estimation of the total forecast uncertainty. In this research, the main sources of uncertainty that may affect flow forecasts are studied.

The objectives of this research are to identify the sources of uncertainty which may play a significant role in flood forecasting; to quantify and propagate the main sources of uncertainty identified through a flow forecasting system; to evaluate, individually and together, the impact of uncertainty quantification on the forecast outcome.

Based on the results of uncertainty quantification and evaluation, this research aims at indicating the main sources uncertainty that should be propagated into flood forecasts to improve forecast quality in the study catchments.

The objectives are presented in terms of four research questions:

1. Which sources of uncertainty significantly affect flood forecasts?
2. How to quantify the important uncertainty sources that affect flood forecasts?
3. How to efficiently propagate those uncertainties through a forecasting model?
4. What is the impact of different sources of uncertainty on the quality of flood forecasts?

For flood forecasting, it is clear that the uncertainty needs to be quantified. By doing so, the impacts of different sources of uncertainty can be assessed. The uncertainty sources that have significant impacts on the forecasts need to be properly propagated into the forecast output. By propagating different sources of uncertainty into the forecasts, more information about the forecasts is added to the output, making it more reliable and, therefore, improving the quality of the forecasts. Distinguishing each source of uncertainty that stems from the individual components of a forecasting model is a challenge in flow forecasting uncertainty quantification. This research proposed an experimental approach to assess each individual source of uncertainty and combine them by their propagation through the forecasting system. Propagation is done with the multiplication of the individual uncertainty sources. In addition, probabilistic evaluation measures (Brier Scores and reliability diagrams) are investigated to assess the correctness of the total predictive uncertainty quantification.

5.1 Conclusions

The uncertainty coming from the spatially averaged precipitation which is used as input precipitation for the GRPE forecasting system has very small impact on the forecasts, compared with other sources of uncertainty such as initial condition uncertainty and parameter uncertainty. When taking this source of uncertainty into account, the change in the outcome is almost unnoticeable. Because of the insignificant impact of this source of uncertainty, at least when quantified with the method used in this research, it can be neglected when propagating the uncertainties through the forecasting chain.

The ensemble prediction systems of ECMWF and Météo-France, with 51 and 11 members of precipitation, respectively, were used to account for forecast precipitation uncertainty. Ten members of discharge ensembles, retrieved from the analysis of uncertainties from rating curves, were used to quantify the uncertainty of hydrologic initial conditions. They were also used as a reference to compare against the forecast outputs. The uncertainty of model parameters is considered in terms of uncertainty about the calibration period and uncertainty about the parameterization. On the one hand, the long period of available data was divided into 10 periods and the resulted parameter sets were used to quantify the uncertainty coming from the calibration period. On the other hand, the parameterization uncertainty was quantified using the GLUE approach, with a large number of simulations leading to the assessment of confidence intervals of the forecasts.

With the above described methods applied to data from three study catchments in France, it was found that the uncertainty of forecast precipitation, initial condition (discharge data) and parameters have a significant impact on the quality of the flood forecasts. These sources of uncertainty need to be considered when issuing a forecast in order to improve the quality of the predictions.

In general, the forecast precipitation uncertainty shows the largest impact on the flood forecasts, especially for high lead times; the initial condition uncertainty has strong effect at small lead times up to 2 days.

The relative significance of the parameter uncertainty evaluated using different calibration periods or the GLUE parameterization varies from catchment to catchment: the results show that, in one catchment, parameter uncertainty based on different calibration periods shows more significant influence on the quality of flood forecasts, while, in another catchment, it is the parameterization uncertainty that proved to have a larger impact. This shows that there may be a site dependence of the impacts of different uncertainty sources on flood forecasting.

5.2 Recommendations

This research recommends taking into account the uncertainty of forecast precipitation, initial condition and model parameter in flood forecasting to improve the reliability of the forecasts. The uncertainty of model parameters can be considered through different angles and by different approaches. In this research, the quantification of parameter uncertainty using different lengths of the calibration period was considered and proved to be a simple method that results in a large impact on the forecast outputs. Therefore, it is recommended to use such an approach for this source of uncertainty, especially if the computational capability available does not allow applying more sophisticated methods like the GLUE method for parameterization uncertainty.

The conclusions above are drawn based on the data used in this research, the study catchments selected, as well as on the forecasting system and the uncertainty quantification methods that were chosen. When quantifying the total predictive uncertainty in flood forecasting for other cases, the results might be different. However, since the results of this research are consistent with what has been reported in the literature, it is quite probably that, for other cases, similar results might be achieved. The more or less significant impact of uncertainty propagation in flood forecasting showed to be site dependent: different responses of flow forecasts were observed among the study catchments. More case studies would be necessary for a larger assessment of the impacts and to investigate if there is a general spatial trend.

In the scope of this research, uncertainty from model structure was not considered due to the time constrain. However, this is recognized as one of the main sources of uncertainty that may impact flow forecasts. This source of uncertainty should be considered in future research. This could be done, for instance, by applying the methods that are used here to different hydrologic forecasting model structures to quantify the impact of model structure uncertainty on the quality of the flood forecasts.

Concerning the evaluation measures used for assessing the uncertainty quantification, there is no single perfect evaluation tool for a multi dimensioned domain like ensemble forecast. It is necessary to use different tools to have a more comprehensive look into the uncertainties affecting flood forecasts and to better understand their impact on the quality of the forecast outcome.

6 References

American Meteorological Society Glossary. (2011). http://amsglossary.allenpress.com/glossary/browse?s=f&p=18. Retrieved date: April 12th, 2011.

BBC. (2002). "What's behind the weather?". http://news.bbc.co.uk/2/hi/science/nature/2190585.stm. Retrieved date: June 29th, 2011.

Berthet, L., Andréassian, V., Perrin, C., and Javelle, P. (2009). "How crucial is it to account for the antecedent moisture conditions in flood forecasting? Comparison of event-based and continuous approaches on 178 catchments". Hydrology and Earth System Sciences 13: 819–831.

Beven, K. and Binley, A. (1992). "The future of distributed models: Mode calibration and uncertainty prediction". Hydrological Processes 6: 278-298.

Brier, W. Z. (1950). "Verification of forecasts expressed in terms of probability". Monthly weather review 78(1): 1-3.

Brocker, J., and Smith, L. A. (2007). "Increasing the reliability of reliability diagrams". American Meteorological Society 22: 651-661. doi: 10.1175/WAF993.1.

Butt, M., Paynea, J.T, Kristensen, M., and Madsena, H. (2009). "An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation". Journal of Hydrology 298: 242–266.

Carpenter, T. M., and Georgakakos, K. P. (2004). "Impacts of parametric and radar rainfall uncertainty on the ensemble streamflow simulations of a distributed hydrologic model." Journal of Hydrology 298: 202–221.

Cloke, H. L., and Pappenberger, F. (2009). "Ensemble flood forecasting: A review". Journal of Hydrology 375(3-4): 613-626.

European Centre for Medium-Range Weather Forecasts. (2011). "The ECMWF Ensemble Prediction System". http://www.ecmwf.int/about/corporate_brochure/leaflets/EPS_2011.pdf. Retrieved date: April 27, 2011.

European Commission. (2011). "Flood and their impacts". http://ec.europa.eu/environment/water/flood_risk/impacts.htm. Retrieved date: April 12th, 2011.

Franz, K. J., and Houge, T. S. (2011). "Evaluating uncertainty estimates in hydrological models: borrowing the forecast verification community". Hydrology and Earth System Sciences Discussion 8: 3085-3131. doi: 10.5194/hessd-8-3085-2011.

Jin, X., Xu, C. Y., Zhang, Q., and Singh, V.P. (2010). "Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model". Journal of Hydrology 383: 147–155. doi: 10.1016/j.jhydrol.2009.12.028.

He, Y., Wetterhall, F., Cloke, H.L, Pappenberger, F., Wilson, M., Freer, J., and McGregor, G. (2009). "Tracking the uncertainty in flood alerts driven by grand ensemble weather predictions". Meteorological Applications 16: 91-101.

Hostache, R., Matgen, P., Montanari, A., Montanari, M., Hoffmann, L., and Pfister, L. (2011). "Propagation of uncertainties in coupled hydro-meteorological forecasting systems: A stochastic approach for the assessment of the total predictive uncertainty". Atmospheric Research 100: 263–274. doi: 10.1016/j.atmosres.2010.09.014.

Hughes, D.A., Kapangaziwiri, E., and Sawunyama, T. (2010). “Hydrological model uncertainty assessment in southern Africa”. Journal of Hydrology 387: 221-232.

Kavetski, D., Kuczera, D., and Franks, S.W. (2006). "Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory". Water Resources Research 42 (W03407). doi: 10.1029/2005WR004368.

Krzysztofowicz, R. (2001). “The case for probabilistic forecasting in hydrology”. Journal of Hydrology 249 (1-4): 2-9.

Kuczera, G., Renard, B., Thyer, M., and Kavetski, D. (2010). “There are no hydrological monsters, just models and observations with large uncertainties!”. Hydrological Sciences Journal, 55(6): 980-991. doi: 10.1080/02626667.2010.504677.

Larsbo, M., and Jarvis, N. (2005). "Simulating Solute Transport in a Structured Field Soil: Uncertainty in Parameter Identification and Predictions". Journal of Environmental Quality 34: 621-634.

Lepioufle, J. M. (2009). “Modélisation spatio-temporelle d'un champ de pluie. Application aux pluies journalières du bassin versant de la Loire”. PhD Thesis, INPG Grenoble, Cemagref, 168 p.

Magnusson, S. E. (1997). “Uncertainty analysis: Identification, Quantification and Propagation”. Report, Lund Institute of Technology.

Mascaro, G., Vivoni, E. R., and Deidda, R. (2010). “Implications of Ensemble Quantitative Precipitation Forecast Errors on Distributed Streamflow Forecasting”. Journal of Hydrometeorology 11: 69-86. doi: 10.1175/2009JHM1144.1.

Maskey, S., Guinot, V., and Price, R. K. (2004). "Treatment of precipitation uncertainty in rainfall-runoff modelling: a fuzzy set approach". Advances in Water Resources 27: 889-898. doi: 10.1016/j.advwatres.2004.07.001.

McMillan, H., Freer, J., Pappenberger, H., Krueger, T., and Clark, M. (2010). “Impacts of uncertain river fl M data on rainfall-runoff model calibration and discharge predictions”. Hydrological Processes 24: 1270–1284. doi: 10.1002/hyp.7587.

Murphy, A. H. (1973). "A new vector partition of the probability score". Journal of Applied Meteorology 12: 595-600.

Nicolau J. (2002). “Short-range ensemble forecasting”. In Proceedings WMO/CBS Technical Conferences On Data Processing and Forecasting Systems, Cairns, Australia, 2-3 Dec, 4p.

Olsson, J., and Lindstrom, G. (2008). "Evaluation and calibration of operational hydrological ensemble forecasts in Sweden". Journal of Hydrology 350: 14-24. doi: 10.1016/j.jhydrol.2007.11.010.

Oudin, L., Michel, C., and Anctil, F. (2005). "Which potential evapo-transpiration input for a lumped rainfall-runoff model?: Part 1—Can rainfall-runoff models effectively handle detailed potential evapo-transpiration inputs?". Journal of Hydrology 303: 275-289. doi: 10.1016/j.jhydrol.2004.08.026.

Pappenberger, F., Ghelli, A., and Buizza, R. (2009). "The Skill of Probabilistic Precipitation Forecasts under Observational Uncertainties within the Generalized Likelihood Uncertainty Estimation Framework for Hydrological Applications". Journal of Hydrometeorology 10: 807-819.

Pappenberger, F., Beven, K. J., Hunter, N.M., Bates, P.D., Gouweleeuw, B.T, Thielen, J., and de Roo, A.P.J. (2005). "Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the European Flood Forecasting System (EFFS)". Hydrology and Earth System Sciences 9(4): 381-393.

Pappenberger, F., Beven, K. J., de Roo, A.P.J., Thielen, J., and Gouweleeuw, B.T. (2004). “Uncertainty analysis of the rainfall runoff model LisFlood within the Generalized Likelihood Uncertainty Estimation (GLUE)”. International Journal of River Basin Management 2(2): 123–133.

Ramos, M. H., Mathevet, T., Thielen, J., and Pappenberger, F. (2010). "Communicating uncertainty in hydro-meteorological forecasts: mission impossible?". Meteorological Applications. doi: 10.1002/met.202.

Ramos, M.H., Thielen, J., and Pappenberger, F. (2008). "Utilisation de la prévision météorologique d'ensemble pour la prévision hydrologique opérationnelle et l'alerte aux crues". Proc. Colloque SHF Prévisions hydrométéorologiques, Lyon, 18-19/11/2008, 191-199.

Ramos, M.H., Leblois, E., and Creutin, J. D. (2006). “From point to areal rainfall: linking the different approaches for the frequency characterization of rainfalls in urban areas”. Water Science and Technology 54 (6-7): 33-40.

Randrianasolo, A., Ramos, M-H, Thirel, G., Andréassian V., and Martin, E. (2010). “Comparing the scores of hydrological ensemble forecasts issued by two different hydrological models”. Atmospheric Science Letters 11.

Renard, B., Ramos, M. H., Pham, T. V., Thyer, M., Le Coz, J., Branger, F., Kavetski, D., and Kuczera. G. (2011). "Tracking the role of streamflow uncertainty in hydrological ensemble predictions". European Geosciences Union General Assembly. (Poster)

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W. (2010). "Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors". Water Resources Research 46, W05521. doi: 10.1029/2009WR008328.

Renner, M., Werner, M.G.F., Rademacher, S., and Sprokkereef, E. (2009). "Verification of ensemble flow forecasts for the River Rhine". Journal of Hydrology 376: 463–475. doi: 10.1016/j.jhydrol.2009.07.059.

Rossa, A., Liechti, K., Zappa, M., Bruen, M., Germann, U., Haase, G., Keil, C., and Krahe,P. (2010). “The COST 731 Action: A Review on Uncertainty Propagation in Advanced Hydro-meteorological Forecast Systems”. Atmospheric Research. doi: 10.1016/j.atmosres.2010.11.016.

Schaake, J., Franz, K., Bradley, A., and Buizza, R. (2006). "The Hydrologic Ensemble Prediction Experiment (HEPEX)". Hydrology and Earth System Sciences Discussions 3: 3321–3332.

Shaw, E. (1994). "Hydrology in practice". Taylor & Francis. ISBN 0 7487 4448 7.

Smith, K. and Ward, R. (1998). "Floods: physical processes and human impacts." John Wiley & Sons, Ltd. 24(13).

Tangara, M. (2005). "Nouvelle méthode de prévision de crue utilisant un modèle pluie-débit global". Thèse de Doctorat, Cemagref Antony, EPHE, Paris, 374 pp.

Thirel, G., Martin, E., Mahfouf, J. F., Massart, S., Ricci, S., Regimbeau, F., and Habets, F. (2010). “A past discharge assimilation system for ensemble streamflow forecasts over France – Part 2: Impact on the ensemble streamflow forecasts”. Hydrology and Earth System Science 14 : 1639–1653. doi: 10.5194/hess-14-1639-2010.

Thirel, G., Rousset-Regimbeau, F., Martin, E., and Habets, F. (2008). "On the Impact of Short-Range Meteorological Forecasts for Ensemble Streamflow Predictions". Journal of Hydrometeorology. doi: 10.1175/2008JHM959.1.

Uhlenbrook, S., Seibert, J., Leibundgut, C., and Rodhe, A. (1999). "Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structure". Hydrological Sciences 44(5): 779-797.

Verbunt, M., Walser, A., Gurtz, J., Montani, A., and Schar, C. (2007). “Probabilistic Flood Forecasting with a Limited-Area Ensemble Prediction System: Selected Case Studies”. Journal of Hydrometeorology 8: 897-909. doi: 10.1175/JHM594.1.

Walker, W. E., Harremoes, P., Rotman, J., van der Sluijs, J.P., van Assel, M.B.A., Janssen, P. and Krayer Von Krau, M.P. (2003). “Defining Uncertainty A Conceptual Basis for Uncertainty Management in Model-Based Decision Support”. Integrated Assessment 4 (1): 5-17.

Warmink, J. J., Van der Klis, H., Booij, M. J., and Hulscher, S. J. M. H. (2010). “Identification and Quantification of Uncertainties in a Hydrodynamic River Model Using Expert Opinions”. Water Resource Manage. doi: 10.1007/s11269-010-9716-7.

Weeink, W. H. A. (2010). "Threshold for flood forecasting and warning". University of Twente, the Netherlands. Master thesis.

Weijs, S. V., Schoups, G., and van de Giesen, N. (2010). "Why hydrological predictions should be evaluated using information theory". Hydrology and Earth System Sciences: 14, 2545–2558. doi: 10.5194/hess-14-2545-2010.

WMO. (2011). Forecast verification – issues, methods and faq. WWRP/WGNE Joint Working Group on Verification. http://www.cawcr.gov.au/projects/verification/. Retrieved date: August 2nd, 2011.

Xiong, L., and O’Connor, K. M. (2008). "An empirical method to improve the prediction limits of the GLUE methodology in rainfall–runoff modelling". Journal of Hydrology 349: 115-124. doi: 10.1016/j.jhydrol.2007.10.029.

Appendix

Detail description of the uncertainty quantification and propagation tests

Test No.	Uncertainty accounted in flood forecast	Catchment applied	Test code	No. of members of Q forecast	Lead time (days)	Data used: Precipitation: Forecast	Data used: Precipitation: Simulation	Data used: Initial condition	Data used: Model parameters
1	Control forecast (CF) (No uncertainty is taken into account)	Allier Ardèche Arc	**CT**	1	1 to 2	CF of PEARP	POBS	QOBS	Calibrated and validated in 1958-2005
					3 to 9	CF of ECMWF			
2	Forecast precipitation	Allier Ardèche Arc	**AR**	11	1 to 2	PEARP	POBS	QOBS	Calibrated and validated in 1958-2005
			EU	51	3 to 9	ECMWF			
3	Initial condition	Allier Ardèche Arc	**QI**	10	1 to 2	CF of PEARP	POBS	QI	Calibrated and validated in 1958-2005
					3 to 9	CF of ECMWF			
4	Calibration period	Allier Ardèche	**PR**	10	1 to 2	CF of PEARP	POBS	QOBS	Calibrated parameters for different periods from 1958 to 2005 (10 periods in total)
					3 to 9	CF of ECMWF			
5	Parameterization	Allier	**GL**	633	1 to 2	CF of PEARP	POBS	QOBS	Behavioural parameters from 125000 runs
		Ardèche		97	3 to 9	CF of ECMWF			
6	Input precipitation	Ardèche	**PI**	10	1 to 2	CF of PEARP	PI	QOBS	Calibrated and validated in 1958-2005
7	Forecast precipitation + Initial condition	Allier Ardèche Arc	**AR_QI**	11*10	1 to 2	PEARP	POBS	QI	Calibrated and validated in 1958-2005
			EU_QI	51*10	3 to 9	ECMWF			
8	Forecast precipitation + Calibration period	Allier Ardèche	**AR_PR**	11*10	1 to 2	PEARP	POBS	QOBS	Calibrated parameters for different periods from 1958 to 2005 (10 periods in total)
			EU_PR	51*10	3 to 9	ECMWF			
9	Forecast precipitation + Parameterization	Allier Ardèche	**AR_GL**	11*633	1 to 2	PEARP	POBS	QOBS	Behavioural parameters from 125000 runs
				11*97					
			EU_GL	51*100	3 to 9	ECMWF	POBS	QOBS	100 parameter sets selected randomly from 633 behavioural sets
				51*97					Behavioural parameters from 125,000 runs
10	Forecast precipitation + Initial condition + Calibration period	Allier Ardèche	**AR_QI_PR**	11*10*10	1 to 2	PEARP	POBS	QI	Calibrated parameters for different periods from 1958 to 2005 (10 periods in total)
			EU_QI_PR	51*10*10	3 to 9	ECMWF			
11	Forecast precipitation +Initial condition + Parameterization	Allier Ardèche	**AR_QI_GL**	11*10*10	1 to 2	PEARP	POBS	QI	10 sets of parameters chosen randomly from the distribution of behavioural runs
			EU_QI_GL	51*10*10	3 to 9	ECMWF			

*** Explanation of the notation:**

No.	Code	Type of data	Catchment	Period of data	Number of members
1	POBS	Observed precipitation	Allier Ardèche Arc	11/03/2005 to 31/07/2009	1
2	PEARP	Ensemble forecast precipitation from Meteo France		11/03/2005 to 31/07/2009	11
3	ECMWF	Ensemble forecast precipitation from European Centre for Medium-Range Weather Forecasts (ECMWF)		11/03/2005 to 31/07/2009	51
4	QI	Ensemble discharges estimated from rating curve		11/03/2005 to 31/07/2009	10
5	QOBS	Observed discharge		depending on catchments	1
6	PI	Spatially averaged precipitation	Ardèche	2000 to 2008	10