Toward Coherent Accounting of Uncertainty in Hydrometeorological Modeling

Thèse

Antoine Thiboult

Doctorat en Génie des Eaux
Philosophiæ doctor (Ph.D.)

Québec, Canada

Résumé

La considération adéquate des différentes sources d'incertitude est un aspect crucial de la prévision hydrométéorologique. La prévision d'ensemble, en fournissant des informations sur la probabilité d'occurrence des sorties du modèle, représente une alternative séduisante à la prévision déterministe traditionnelle. De plus, elle permet de faire face aux différentes sources d'incertitude qui se trouvent le long de la chaîne de modélisation hydrométéorologique en générant des ensembles là où ces incertitudes se situent.

Le principal objectif de cette thèse est d'identifier un système qui soit capable d'appréhender les trois sources principales d'incertitude que sont la structure du modèle hydrologique, ses conditions initiales et le forçage météorologique, dans le but de fournir une prévision qui soit à la fois précise, fiable et économiquement attractive. L'accent est mis sur la cohérence avec laquelle les différentes incertitudes doivent être quantifiées et réduites. Notamment, celles-ci doivent être considérées explicitement avec une approche cohésive qui fasse en sorte que chacune d'entre elles soit traitée adéquatement, intégralement et sans redondance dans l'action des divers outils qui composent le système.

Afin de répondre à cette attente, plusieurs sous-objectifs sont définis. Le premier se penche sur l'approche multimodèle pour évaluer ses bénéfices dans un contexte opérationnel. Dans un second temps, dans le but d'identifier une implémentation optimale du filtre d'ensemble de Kalman, différents aspects du filtre qui conditionnent ses performances sont étudiés en détail. L'étape suivante rassemble les connaissances acquises lors des deux premiers objectifs en réunissant leurs atouts et en y incluant la prévision météorologique d'ensemble pour construire un système qui puisse fournir des prévisions à la fois précises et fiables. Il est attendu que ce système soit en mesure de prendre en compte les différentes sources d'incertitude de façon cohérente tout en fournissant un cadre de travail pour étudier la contribution des différents outils hydrométéorologiques et leurs interactions. Enfin, le dernier volet porte sur l'identification des relations entre les différents systèmes de prévisions précédemment créés, leur valeur économique et leur qualité de la prévision.

La combinaison du filtre d'ensemble de Kalman, de l'approche multimodèle et de la prévision météorologique d'ensemble se révèle être plus performante qu'aucun des outils utilisés séparément, à la fois en précision et fiabilité et ceci en raison d'une meilleure prise en compte de l'incertitude que permet leur action complémentaire. L'ensemble multimodèle, composé par 20 modèles hydrologiques sélectionnés pour leurs différences structurelles, est capable de minimiser l'incertitude liée à la structure et à la conceptualisation, grâce au rôle spécifique que jouent les modèles au sein de l'ensemble. Cette approche, même si utilisée seule, peut conduire à des résultats supérieurs à ceux d'un modèle semi-distribué utilisé de façon opérationnelle. L'identification de la configuration optimale du filtre d'ensemble de Kalman afin de réduire l'incertitude sur les conditions initiales est complexe, notamment en raison de l'identification parfois contre-intuitive des hyper-paramètres et des variables d'état qui doivent être mises à jour, mais également des performances qui varient grandement en fonction du modèle hydrologique. Cependant, le filtre reste un outil de première importance car il participe efficacement à la réduction de l'incertitude sur les conditions initiales et contribue de façon importante à la dispersion de l'ensemble prévisionnel. Il doit être malgré tout assisté par l'approche multimodèle et la prévision météorologique d'ensemble pour pouvoir maintenir une dispersion adéquate pour des horizons dépassant le court terme. Il est également démontré que les systèmes qui sont plus précis et plus fiables fournissent en général une meilleure valeur économique, même si cette relation n'est pas définie précisément.

Les différentes incertitudes inhérentes à la prévision hydrométéorologique ne sont pas totalement éliminées, mais en les traitant avec des outils spécifiques et adaptés, il est possible de fournir une prévision d'ensemble qui soit à la fois précise, fiable et économiquement attractive.

Abstract

A proper consideration of the different sources of uncertainty is a key point in hydrometeorological forecasting. Ensembles are an attractive alternative to traditional deterministic forecasts that provide information about the likelihood of the outcomes. Moreover, ensembles can be generated wherever a source of uncertainty takes place in the hydrometeorological modeling chain.

The global objective of this thesis is to identify a system that is able to decipher the three main sources of uncertainty in modeling, i.e. the model structure, the hydrological model initial conditions and the meteorological forcing uncertainty, to provide accurate, reliable, and valuable forecast. The different uncertainties should be quantified and reduced in a coherent way, that is to say that they should be addressed explicitly with a cohesive approach that ensures to handle them adequately without redundancy in the action of the different tools that compose the system.

This motivated several sub-objectives, the first one of which focusing on the multimodel approach to identify its benefits in an operational framework. Secondly, the implementation and the features of the Ensemble Kalman Filter (EnKF) are put under scrutiny to identify an optimal implementation. The next step reunites the knowledge of the two first goals by merging their strengths and by adding the meteorological ensembles to build a framework that issues accurate and reliable forecasts. This system is expected to decipher the main sources of uncertainty in a coherent way and provides a framework to study the contribution of the different tools and their interactions. Finally, the focus is set on the forecast economic value and provides an attempt to relate the different systems that have been built to economic value and forecast quality.

It is found that the combination of the EnKF, the multimodel, and ensemble forcing, allows to issue forecasts that are accurate and nearly reliable. The combination of the three tools outperforms any other used separately and the uncertainties that were considered are deciphered thanks to their complementary actions. The 20 dissimilar models that compose the multimodel ensemble are able to minimize the uncertainty related to the model structure,

thanks to the particular role they play in the ensemble. Such approach has the capacity to outperform more a complex semi-distributed model used operationally. To deal optimally with the initial condition uncertainty, the EnKF implementation may be complex to reach because of the unintuitive specification of hyper-parameters and the selection of the state variable to update, and its varying compatibility with hydrological model. Nonetheless, the filter is a powerful tool to reduce initial condition uncertainty and contributes largely to the predictive ensemble spread. However, it needs to be supported by a multimodel approach and ensemble meteorological forcing to maintain adequate ensemble dispersion for longer lead times. Finally, it is shown that systems that exhibit better accuracy and reliability have generally higher economic value, even if this relation is loosely defined.

The different uncertainties inherent to the forecasting process may not be eliminated, nonetheless by explicitly accounting for them with dedicated and suitable tools, an accurate, reliable, and valuable predictive ensemble can be issued.

Contents

List of Tables

List of Figures

En essayant continuellement on finit par réussir. Donc: plus ça rate, plus on a de chance que ça marche.

Les Shadoks – Jacques Rouxel

Remerciements

Mes premiers remerciements vont à mon directeur de thèse, François Anctil, qui m'a donné ma chance et su me faire confiance dans une nouvelle discipline alors que, fraîchement débarqué de physique, je ne connaissais pas grand-chose à la modélisation hydrologique. Je lui suis très reconnaissant d'avoir mis en place une véritable collaboration, en m'invitant à me responsabiliser, mais en restant toujours disponible afin de m'aiguiller et démêler les situations les plus épineuses. Merci François pour la dimension profondément humaine que vous donnez à l'enseignement et à la recherche.

Je tiens à remercier tous ceux qui ont été directement impliqués dans mon cursus universitaire, notamment les membres du jury, Vincent Fortin chercheur à Environnement Canada, Maria-Helena Ramos chercheuse à l'Institut National de Recherche en Sciences et Technologies pour l'Environnement et l'Agriculture, ainsi que Daniel Nadeau enseignant chercheur à l'Université Laval, qui ont contribué à l'amélioration de cette thèse, mais aussi les professeurs de l'université qui ont assisté aux différents examens qui ponctuent la thèse, Peter Vanrolleghem, Paul Lessard, Amaury Tilmant. J'adresse ma gratitude à l'équipe du Centre Expertise Hydrique du Québec et notamment Richard Turcotte et Simon Lachance Cloutier pour m'avoir fourni une partie importante des données nécessaires à la réalisation de mon doctorat ainsi que pour leur assistance. Je tiens aussi à remercier Marie-Amélie Boucher, enseignante chercheuse à l'UQAC, pour sa collaboration.

Un grand merci à tous les membres de ma famille qui m'ont soutenu tout au long de mon cursus universitaire et durant cette thèse.

Ce fut un réel plaisir de faire partie d'une équipe aussi sympathique. Merci à Darwin et son geekisme contagieux, Annie-Claude, Pascale pour son coup de pouce lors de ses stages, Mabrouk, Slim, Benoît, Jérôme, Islem, Flora, Anne. Une petite mention spéciale pour Greg qui a pris le temps de partager ses codes et son expérience avec le débutant que j'étais.

Un énorme merci à toutes les personnes qui m'ont réservé un chaleureux accueil à mon arrivée au Québec et avec qui j'ai eu un énorme plaisir à refaire le monde autour d'un café, d'un wrap au poulet ou d'une bière. Je pense notamment à Thibaud et Esther, Fred, Renaud, Jean, Sovanna, Étienne, Cédric, Loriane, Damien, Pocquet, Ian, Gaëlle, Boris, Queralt... Lors d'une thèse, il y a (malheureusement pour moi, heureusement pour eux) des amis qui quittent l'université, mais d'autres arrivent et prennent une relève d'enfer. Merci à mes amis arrivés un peu plus tard dans le département et qui ont fait que chaque nouvelle année fut aussi délectable que la précédente: Youen, Thomas the dude, Philippe, Ana que riquiña et Juan, Christine, Cyril.

Je me dois aussi de mentionner tous ceux qui ont contribué à ces joyeuses années en dehors de l'université. Merci à mon padawan moucheur Ben, JP, Emmanuelle, Aline et tout le reste de la clique ainsi qu'à mes colocs qui ont joyeusement rythmé mon quotidien, David, princesse Lily, Chloé et Spew.

Un grand merci à mes chums de rivière, beaucoup trop nombreux pour tous les citer, de m'avoir aidé à garder un équilibre ludo-universitaire raisonnable et d'avoir grandement contribué à mon intégration au Québec. Merci en particulier à Maxime, Maxime et Kath, Kevin, Nicolas R., Guillaume et Nicolas L. qui doit sûrement être le seul non-hydrologue à lire mes articles.

Mes derniers remerciements, et non des moindres, vont à Stéphie, pour son soutien lors de toutes ces années, pour ses adorables colis de survie du thésard, et qui, je ne me l'explique toujours pas, a fait preuve d'une motivation hors du commun pour la relecture de cette thèse. Mein Schatz.

Les remerciements sont à l'image d'un doctorat, on voudrait tout inclure et ne rien oublier, mais ça ne fonctionne jamais comme ça. Toutes mes excuses et merci à toi qui mérite aussi de figurer sur cette page.

À tous un grand merci et bonne lecture.

Preface

This thesis is based on several publications from which the first author is also the author of this thesis. Chapters have been slightly modified to avoid redundancy, mainly in chapters introduction and presentation of material.

A version of Chapter 1 has been published (Assessment of a multimodel ensemble against an operational hydrological forecasting system, A. Thiboult, F. Anctil, Canadian Water Resources Journal / Revue canadienne des ressources hydriques, Vol. 40, Iss. 3, 2015). A version of Chapter 2 co-authored by François Anctil is currently under review in Journal of Hydrology. A version of Chapter 3 co-authored by François Anctil and Marie-Amélie Boucher is published in Hydrology and Earth Science System Discussion (Hydrol. Earth Syst. Sci. Discuss., 12, 7179-7223, 2015, www.hydrol-earth-syst-sci-discuss.net/12/7179/2015/, doi: 10.5194/hessd-12-7179-2015) and is under reviewing for publication in Hydrology and Earth Science System.

Table 1.2, Figures 1.2 and Figures in Annex B and C were reproduced courtesy of Gregory Seiller.

Introduction

Although the equations that describe atmospheric circulation have been mastered for a long time, weather forecasting remains complex. Indeed, many phenomena are difficult to grasp because of the extreme sensitivity to initial conditions illustrated by Lorenz and chaos theory (Lorenz, 1963), or because of the non-linearity of the physical processes (Lions et al., 1992). For these reasons, weather forecasts are still uncertain and will likely remain so despite major dedicated efforts aimed at reducing this uncertainty.

Hydrological forecasting relies heavily on precipitation and temperature forecasts. Given the uncertainty that remains in precipitation forecast, predicting river streamflows is a daunting task. Hydrological forecasting adjoins additional complexity by implicitly or explicitly taking into account every process that lies between precipitation and the catchment outlet. All these physical processes are associated with some uncertainty, stochastic or deterministic. These uncertainties represent a real challenge for operational hydrological forecasting and scientific knowledge. Nonetheless, if uncertainties are well understood and properly handled, informative and reliable forecasts can be issued.

This thesis aims to evaluate, improve, and develop the understanding of operational forecasting techniques through the use of meteorological ensemble prediction, hydrological multimodel approach, and streamflow data assimilation.

Hydrological modeling: General statements

Hydrological modeling is used in many fields to deal with numerous issues. The very definition of hydrology even varies depending on the viewpoint and interests. Penman (1961) defines it asking the following question: "What happens to the rain?". Under this simple formulation actually lays a lot of complexity.

Indeed, although the water cycle is relatively well understood as a whole, the description and quantification of the physical processes and their uncertainties at the watershed scale is a major challenge in modeling. Components of the water cycle such as snowmelt, runoff, flow routing, evaporation, and interception are all elements that convey uncertainty (Elkadi, 1989; Goodrich and Woolhiser, 1991). These different aspects of the watershed dynamic are closely related to each other and are in practice extremely difficult to assess separately.

Hydrological models allow to understand and predict the watershed response to precipitation and to provide answers to practical case study of land use change, irrigation management, preservation of natural habitat, flood forecasting and mitigation, dam management, and water quality (Singh and Woolhiser, 2002). They also represent a considerable source of information for the understanding the watershed physics (Kirchner, 2006), specifically to identify interactions between climate, snow cover, land use, etc. The models are now a key component of management, engineering, and scientific research toolboxes that contribute to the overall understanding of watershed systems.

Development of the first components of what will eventually become hydrological models is often attributed to Mulvany in the middle of the XIXth century with the conceptualization of the rational method, followed by Sherman in the '30s with the introduction of the unit hydrograph, and Horton with the formalization of the infiltration process. Later, Penman (1948) and Thornthwaite (1948) formulated equations to describe evapotranspiration equations, which are now frequently used in hydrological modeling.

In the '50s, the numbers of new theories sprang up and contributed to modern hydrology. With the development of computers in the '60s, it became possible to set models at the catchment scale by integrating different hydrological processes. The Stanford Watershed Model is often presented as the first model to describe all the main dynamics of a watershed. Other famous models including TOPMODEL (Beven et al., 1984), SWMM (Metcalf and Eddy, 1971) and SHE (Abbott et al., 1986), for example, continue to be developed and are still used today. Singh and Woolhiser (2002) and Goodrich and Woolhiser (1991) draw up a more exhaustive list of hydrological models that are considered to have contributed substantially to the development of watershed modeling.

Quality of hydrological data is a crucial aspect for modeling, in particular their consistency and accuracy should be considered prior to model forcing (Silberstein, 2006). This is especially true with the use of new remote acquisition techniques such as satellites (Gourley and Vieux, 2006; Lorenc, 2003) and radars (Georgakakos et al., 2004; Gourley and Vieux, 2006; Ranzi et al., 2009; Turcotte et al., 2001). These techniques may allow to get round the lack of field measurements to better identify spatial properties of the watershed and its initial state and to provide a better picture of precipitation spatial distribution.

Also, the choice of an appropriate model may be conditioned by the type of data available – physical and fully distributed models usually necessitate more data than conceptual lumped models. Required data usually include hydrometeorological data for the forcing (solid and liquid precipitations, temperature, radiation, humidity, vapor pressure, solar radiation, and wind speed), land use, geology and topography (soil type, porosity, soil moisture, etc.).

Uncertainty: A primer

The definition of uncertainty varies depending on the user and field, and forecast expectation. It is defined here as " any deviation from the unachievable ideal of completely deterministic knowledge of the relevant system" (Walker et al., 2003). Uncertainty has three dimensions: location, level, and nature.

The notion of location refers to the structure of a system where it is possible to situate the various sources of uncertainty. Five locations are identified:

- the context: identification of boundaries, i.e. completeness of the modeling representation
- model uncertainty: the structural model uncertainty that comes from the failure in our understanding of the system and the technical model uncertainty
- inputs: external forces that lead to changes in the system
- parameter uncertainty: calibrated time invariant model components that are implicitly associated with input, structure, and calibration
- output: also called prediction error, which is aggregation of the uncertainties of the modeling chain

The level of uncertainty ranges between fully deterministic understanding that it is not possible to reach and total ignorance. We can define an axis of increasing level of uncertainty as such: determinism, statistical uncertainty, uncertainty about the scenario, recognized ignorance, total ignorance (Walker et al., 2003; Pappenberger and Beven, 2006). Ideally, all decisions should take into account the possibility of unknown changes, not just the uncertainty around the phenomenon known (Walker et al., 2003). Hence, the goal here is to reduce adverse impacts from the unexpected rather than hoping to eliminate them.

Finally, the nature of the uncertainty indicates if the uncertainty is due to the imperfection of knowledge that can be reduced by carrying out more fundamental research (the epistemic uncertainty) or the variability inherent to natural chaotic systems (uncertainty of variation).

Ensembles

The vast majority of hydrological models are deterministic and the error laying in their output can be hard to estimate properly. An alternative to the deterministic approach is the creation of ensembles.

Flood forecasting systems rely on rainfall estimations that can be gathered for example from rain gauges, radars for now-casting, or Numerical Weather Prediction (NWP) systems that allow to extend hydrological forecast beyond the catchment concentration time, typically up to 10 or 15 days ahead.

The past decade has witnessed a growing use of Ensemble Prediction Systems (EPS) that are a collection of deterministic forecasts. These can be used, among other applications, in operational flow forecasting (e.g. Dietrich et al., 2009; Georgakakos et al., 2004; Hopson and Webster, 2010; Renner et al., 2009; Roulin, 2007), for climate change studies (e.g. Christensen and Lettenmaier, 2007; Murphy et al., 2004; Seiller et al., 2012), and soil occupation investigation (e.g. Breuer et al., 2009; Huisman et al., 2009; Viney et al., 2009).

As such, many meteorological agencies have adopted ensembles (MEPS) like the Meteorological Service of Canada, the European Center for Medium-Range Weather Forecasts, the Japan Meteorological Agency. This move toward probabilistic forecast has been followed by flood forecasting agencies like the European Flood Alert System, the Finnish Hydrological Service, Georgia-Tech/Bangladesh project, and more.

Meteorological ensembles

Weather forecasting is a complex task because of the difficulties inherent to numerical modeling of the atmospheric circulation patterns, atmosphere and ground coupling, horizontal and vertical resolution issues, and the great sensitivity to initial conditions. To characterize these uncertainties, ensemble forecasting provides a sample of likely outcomes, i.e. several predictions for a given place at a given time.

Ensembles are usually generated from a set of initial conditions identified from an analysis. Thus, a set of typically ten to fifty initial conditions is obtained and is used to initiate simulations. Alternative techniques to form ensembles include the poor man's ensemble (Jasper et al., 2002; Davolio et al., 2008) that explicitly accounts for meteorological model structural and data assimilation error, lagged ensembles (Dietrich et al., 2009), and the creation of super ensembles such as the THORPEX Interactive Grand Global Ensemble (TIGGE, Park et al., 2008; He et al., 2009) that combines several MEPSs to provide extensive estimates of meteorological uncertainty. These ensembles show increasing performance (e.g., Cloke and Pappenberger, 2009; Charron et al., 2010; Buizza et al., 2007), which suggests forthcoming progress in hydrological forecasting.

Studies show that the meteorological ensemble forecasts are often more accurate than deterministic ones (Bourke et al., 2004; Palmer et al., 2004; Doblas-Reyes et al., 2000; Kang and Yoo, 2006), more consistent on consecutive days (Buizza, 2008) in addition to provide information about the uncertainty. However, in view of the MEPS remaining imperfections, it is possible and often desirable to process them prior to using, by correcting their bias (Christensen and Lettenmaier, 2007; Goddard et al., 2001) or dispersion (Boucher et al., 2015). For instance, Fortin et al. (2006) shows that it is possible to improve MEPSs' performance by processing their output, resulting in an improvement in temperature and precipitation prediction.

To reconcile meteorological and hydrological model scales, meteorological forecast frequently needs to be downscaled in order to provide higher resolution information. Several scaling techniques exist (Maraun et al., 2010) but are subdivided into two main groups: statistical and dynamical downscaling. Statistical downscaling aims to estimate statistical relationships between observed variables at local and larger scale variables, using neural networks, analog methods, multifractal cascades (Deidda, 2000), weather type (Boe et al., 2006),regression techniques, and disaggregation (Segond et al., 2006). On the other hand, dynamic scaling carries out a transfer from GMC outputs to regional models. Note that MEPS are mostly Global Circulation Models (GMC) but there exists also some regional MEPSs such as COSMO-LEPS (Molteni et al., 2001).

Hydrological ensembles

Hydrological Ensemble Prediction Systems (HEPS) rely on the same concept as the MEPS by providing a set of members that aims at grasping the sources of uncertainty. Several techniques are commonly used to create HEPS:

- a unique hydrological model with MEPS
- several models with deterministic forcing
- variants of the same model used together
- probabilistic data assimilation
- a combination of the above methods

There are many recent publications that build HEPS forecasts from a single hydrological model (e.g. Dietrich et al., 2009; He et al., 2009; Hopson and Webster, 2010; Jaun and Ahrens, 2009; Li et al., 2009; Randrianasolo et al., 2010; Ranzi et al., 2009; Renner et al., 2009; Thielen et al., 2009; Thirel et al., 2010b,a). These authors agree that the HEPS have substantial advantages over deterministic forecast or statistical ensembles[1], including greater reliability and better uncertainty handling. The uncertainty deciphered by the MEPS can be cascaded through the hydrological model (Pappenberger et al., 2005) and can be used in an operational context. First results of the project Distributed Model Intercomparison Project show that the simple average of all forecasts is very often better than any of the models within the ensemble (Reed et al., 2004; Smith et al., 2004).

Ensembles are generally better to predict extreme events even if they remain difficult to detect several days in advance. They allow not only to identify the most likely outcome but also to evaluate the probability of occurrence of rare and extreme events (Cloke and Pappenberger, 2009). Notheless, simulated flood peaks are often shifted in time or their amplitude is misjudged (Regimbeau et al., 2007).

When several models with fundamentally different structures are pooled, the subsequent ensemble is referred to as a multimodel ensemble. Clemen (1989) lists multimodel applications in a wide variety of fields (management, economics, social sciences, etc.) and states that the gain is important and obtained at low cost. The principle of multimodel is based on the fact that it is now broadly recognized that there is no model that provides the best results in every situation and that the increase of performance for a certain model and watershed is often realized at the expense of the performance on other watersheds (Oudin et al., 2006; Clark et al., 2008b; Reed et al., 2004; Smith et al., 2004). Thompson (1977) attributes the superiority of the multiple model approach by "the incontrovertible fact that two or more inaccurate but independent predictions of the same future events may be combined in a very specific way to yield predictions that are, on the average, more accurate than either or any of them taken individually". Results reported by Velázquez et al. (2011) show that simple pooling of existing models generally gives similar or superior performance than the best model of the ensemble.

A smaller number of studies, adopting a more comprehensive approach, combine meteorological and hydrological ensembles (Block et al., 2009; Velázquez et al., 2010, 2011). These forecasting systems have the advantage to take into account both the uncertainty of weather forecasting and hydrological models.

This type of ensemble characterizes more comprehensively uncertainties but exhibit a drawback in their larger size. The large number of members can be an obstacle in practical use, but solutions like member selection that allow to reduce ensemble size without loss of predictive skill were proposed (Brochero et al., 2011b,a; Marshall et al., 2005).

[1] Ensembles that are initially deterministic and made probabilistic by taking into account the assumed error. Members are probabilistically distributed around a mean.

Distributed models and lumped models

The question of the adequate spatial scale to represent catchment properties is still a controversial topic. With increasing computational power, the use of distributed models becomes more common, often at the expense of lumped models.

A distributed model takes into account the spatial variability of the geophysical characteristics of the watershed. The digital elevation models (geometry of the basin, the hydrographic network, topographic gradient) and digital terrain model provide required data, usually on a grid, to the hydrological models. An advantage of distributed models is that it is theoretically possible to have access to any variable at a specific location within the watershed. However, such a system requires a larger amount of data that may be difficult to gather. A second obstacle to their use is that their structure usually relies on a large number of parameters that make them more complex to calibrate.

Physically distributed models are necessary for understanding the processes that govern the dynamics of a watershed (Kirchner, 2006) but rarely improve upon lumped models performance in forecasting (e.g. Bormann et al., 2009; Refsgaard and Knudsen, 1996).

The spatial resolution faces the problem of variability. Sources of variability can be stochastic, deterministic, or both at the same time depending on the scale, and it is not possible to describe accurately the watershed according to a single characteristic length (Singh and Woolhiser, 2002). This justifies the simplifications made by hydrological models (lumped and distributed) as long as the behavior of the watershed is preserved.

Kavvas (1999) shows that large scale processes that are formed by a set of smaller scale processes, can be described directly at the resolution of interest since processes with high frequency at small scale are eliminated by averaging effect. Lumped models are partly based on this idea. It is assumed that it is not necessary to know all the physical phenomena at every location of the watershed in order to understand its evolution and to predict its behavior through its global dynamic. Lumped models therefore present an integrated response at the watershed scale. They generally require fewer inputs and are consequently much easier to calibrate and implement. In return, the validation of lumped models is less "complete" as it is only possible to compare a single value against observations, typically the flow at the outlet.

There are physical equations that can describe more or less accurately every step of the rainfall-runoff process but they are nevertheless difficult to apply and notably to integrate on large scale. It is convenient to use so-called conceptual models that reduce complexity. They are no longer based on a bottom-up description, but on an overall description of the system. Hydrological processes that are located between rainfall and the outlet are no longer expressed explicitly with physical equations but through concepts. The structure of these models is simpler, they have fewer parameters, and are therefore easier to calibrate (Perrin et al., 2001).

A third type of model, the semi-distributed ones, achieves a trade-off between the two types mentioned above. Semi-distributed models simulate hydrological processes at a scale where catchment properties are deemed homogeneous – the relatively homogeneous hydrological unit. Smith et al. (2004), Breuer et al. (2009), Bormann et al. (2009), and Refsgaard and Knudsen (1996) show that the different types of models may reach similar performance if a preliminary calibration is performed.

Snow modeling

Snow accumulation and melting modeling requires a particular attention since the process contributes significantly to the spring freshet. The melting of the snowpack can cause extreme events and proper identification of snowpack properties is necessary to enable prevention of adverse events. Solid precipitation delays the reaction of the watershed over periods that can vary widely. Modeling the accumulation of snow and its melting requires a specific module, the snow accounting routine. Snow and hydrological models faces similar issues, namely the adequate level of complexity to accurately describe relevant processes, the definition of spatial and temporal resolution, the identification of parameters and their calibration. Concealing the importance of snowmelt in streamflow forecasting leads to poor performance (e.g. Nicolle et al., 2011; Valery et al., 2014b).

Klemeš (1988) identifies several difficulties inherent to snow modeling. First, observations that are critical for model calibration and validation are hardly obtainable and often scarce. Also, the accuracy of the measurements is often dubious because of the large range of inherent difficulties that are associated with this type of data. Lastly, the data representativeness is often questionable as, for example, wind and shade can affect snowpack thickness and maturation, respectively. Although remote measurement techniques can provide a more reliable estimate than simple spot measurements, these techniques are still perfectible and interpretation of results is still tricky (Turcotte et al., 2001).

The snow accounting routine must perform several operations at each time step:

- interpolate meteorological data available for snow cover

- separate solid and liquid precipitations
- calculate the melting rate at different points
- integrate snow melting at catchment scale to get the runoff that directly contributes to streamflow
- update the snow covered area and/or snowpack

Melting description can be done by either conceptual models with reservoirs or fully distributed physical models. Gurtz et al. (2003), Martin (2005), and DeBeer and Pomeroy (2009) advocate a thermodynamic approach based on mass conservation equations and energy balance. On the other hand, conceptual modules favor the degree day approach since air temperature and the main components of the energy balance are highly correlated (Ohmura, 2001; Eckert, 2002; Hock, 2003; Kienzle, 2008). In this case, the only variable of interest is the temperature that is supposed to describe the melting. Mellow (1999) concluded that one should not favor any particular type of model, but should adapt his approach based on available data and model use.

Separation of liquid and solid precipitation can be performed by setting a threshold temperature (typically set to 0º C but it can be adjusted for each watershed and season, Kienzle, 2008), or by setting a temperature range in which snow is represented as partially solid/liquid (Valery et al., 2014a,b; Nicolle et al., 2011).

An additional difficulty in snow melting estimation is its highly heterogeneous spatial distribution that is linked to dissimilar sun exposure and topographic influence. It is thus difficult to define melting rate that are characteristic of catchment scale. Ferguson (1999) diagnoses interpolation of input data as the first step to be considered for the study of snow. The decrease of temperature with increasing altitude, the variation of this gradient during the year, etc., are aspects that need to be considered. To describe this variability, conceptual snow accounting routine frequently discretize the catchment in several bands of equal altitude or rely on snow stock depletion curve.

Data assimilation

A brief overview

Although hydrological models gained in efficiency, our understanding of hydrology remains incomplete and model outputs remains consequently quasi-systematically different from observations. Data assimilation (DA) has the capacity to reduce these deviations from the ideal value. Liu and Gupta (2007) defines assimilation as the "procedures that aim to produce physically consistent representations or estimates of the dynamical behavior of a system by merging the information present in imperfect models and uncertain data in an optimal way to achieve uncertainty quantification and reduction". Practically, DA aims at improving simulation by including new observations in the modeling process as they become available. DA is opposed to the open loop scheme, where the model evolves only according to the input forcing and remains otherwise unperturbed.

Data assimilation is now broadly used in hydrometeorological sciences and begins to be used operationally since it proved to have the capability to improve simulation quality. DA has been applied in many fields and can be used to assimilate various sources of observations. These assimilated data can be obtained from direct field measurements, such as watershed discharge, thickness and state of the snowpack, or soil moisture (e.g., Seo et al., 2009; Clark et al., 2008b; Thirel et al., 2010a; DeChant and Moradkhani, 2011; Franz et al., 2014), but also from remote sensing of soil moisture estimates (Forman et al., 2012; Meier et al., 2011; Renzullo et al., 2014; Alvarez-Garreton et al., 2014), snow sensing (Kuchment et al., 2010), and radar forcing (Harader et al., 2012; Kim and Yoo, 2014).

DA typically acts on forecast initial conditions, namely the internal states of the hydrological models that contain, for instance, information about soil moisture, conceptual reservoir levels, or snow cover and snowpack state. When comparing observations and "model first guess", this difference is used to reinitialize the model to ensure to minimize the observed discrepancy. Therefore, DA improves model initial conditions with the possibility to enhance forecasting by providing better initial conditions. Also, DA differs from error "diagnostic only" tool as it does not only aim at quantifying uncertainty but also at reducing the differences between observation and simulation. In the case of probabilistic DA techniques, in addition to set the model back "on the right path", the assimilation can control the shape of the predictive function and usually contributes to provide more meaningful confidence intervals.

As uncertainties may be located at several levels in the forecasting system and may be of a different nature, DA techniques may take different forms. In particular, depending on the DA technique used, it can act at several places to correct model predictions (Refsgaard, 1997).

- input updating: typically concerns temperature and precipitation and is often iteratively performed by experimented modelers. It acts on state variables through manually perturbed forcing, or may be automatized through an optimization scheme.
- state variable updating: adjustment of state variable that can be carried out by either a substitution of the value if the observation corresponds directly to the nature of the state variable or by more sophisticated techniques if the model state is hidden.

- parameters updating: modify originally time invariant parameters (obtained through calibration). This technique is often considered controversial but can compensate calibration error and explicitly account for parameter uncertainty.
- output updating: deterministic or probabilistic correction of model outputs. It relies on the assumption that outputs are frequently correlated through time. A major advantage resides in its simplicity since it may be associated to a post-processing step.

Occasionally, DA may be performed simultaneously at several places, like dual state-parameter updating (Vrugt et al., 2005; Moradkhani et al., 2005; Nie et al., 2011).

In the past decade, state updating techniques became dominant thanks to their efficiency. A broad range of techniques have been created with many variations that claim to improve upon the initial technique they are derived from. However, they usually fall into one of the following category: variational assimilation, particle filtering, and Kalman filtering approaches.

Variational assimilation uses past observations included in a specified temporal window to update state variables. It relies on the definition of a cost function representing the aggregated model error that should be minimized (Seo et al., 2003; Lorenc, 2003; Seo et al., 2009; Lee et al., 2011; Abaza et al., 2014).

The particle filter estimates recursively model states by providing a complete probability distribution of state variables. It does not update states based on observations but rather performs a resampling of the sets of states, named particles, to create an accurate posterior distribution with according weights (DeChant and Moradkhani, 2012; Thirel et al., 2013; Moradkhani et al., 2006).

Kalman filters

The Kalman filter (KF), the extended Kalman filter (EKF), and the ensemble Kalman filter (EnKF) belong to sequential assimilation methods. Sequential data assimilation rests upon studies of the system states in order to find a set that statistically fits best the observations. These methods rely on the Markov property, i.e. that future states depend only upon the present states and the equations that describe the evolution of the process. They intend to update one or more of the states of hydrological models and may be applied to high dimensional systems. In their basic form, the state-space models are based on the assumption that all probability distributions involved are Gaussian.

In 1960, Kalman describes a recursive solution for linear filtering problems with discrete data, which was named later the Kalman Filter (KF). The field of potential applications is wide and KF efficiency has been demonstrated from satellite trajectory correction to economy purpose and communication.

The extended Kalman filter (EKF, Jazwinski, 1970) realizes an improvement over the KF in the relaxation of the linearity constraint. Indeed, hydrological processes and models are recognized to be broadly non linear and this filter has been logically applied to hydrological modeling (Katul et al., 1993; Walker and Houser, 2001) and assimilation of snow cover (Sun et al., 2004). However, EKF presents major difficulties in its implementation, mainly in the derivation of linear tangent operators and convergence (Evensen, 1992; Bouttier, 1994) and felt in disuse for hydrological applications.

The ensemble Kalman filter was introduced in 1994 (Evensen) and then corrected (Burgers et al., 1998) to offer a stochastic alternative to the EKF. Unlike KF and EKF, in the EnKF, errors are statistically represented by a cloud of points that is propagated by the evolution model without any required linearization. The Monte Carlo approach enables to identify the error covariances that are necessary to compute the updated states values. Once this step achieved, the analysis is similar to the traditional KF.

Like most DA technique, the EnKF requires some prior knowledge about the uncertainty contained in the modeling process. This is particularly apparent in the specification of the perturbations of the model inputs and observations even if little attention has been dedicated to this topic (Moradkhani et al., 2005). All states variables are generally updated for lumped or semi-distributed models, but few studies like Vrugt et al. (2006) argue that updating only 3 out of the 9 model states leads to an improvement of 50% with the SAC-SMA model and that the remaining error is mostly uncorrelated.

Objectives

Despite dedicated efforts, quantification and reduction of the uncertainty between meteorological forecasts and predicted discharges remains imperfect. Several aspects of the rainfall-runoff modeling are still considered as problematic and many authors stress out that more work is needed to address these shortcomings.

It is now widely recognized that quantification and reduction of uncertainty have critical importance. Nonetheless, there is still no existing framework for uncertainty handling that makes consensus, as existing ones remain often incomplete. Cloke and Pappenberger (2009) concluded that a key challenge will be to elaborate an optimal framework that will make use of formal and explicit statistical treatment. Bourdin et al. (2012) substantiated this statement by emphasizing that the existing frameworks are typically confined to decipher one or two sources of uncertainty and that progress will be realized by developing broader uncertainty accounting. Therefore, an optimal framework should take into account all the hydrometeorological uncertainties, regardless of their location.

Even though hydrological model performance increased, none of them was shown to be superior in all circumstances. Moreover, we cannot observe a consequent increase in individual model performance over the last decade indicating that hydrology may have hit a hurdle in the improvement of model structure (Todini, 2007; Singh and Woolhiser, 2002). Although the use of multiple models is usually accompanied by higher computational cost than single model simulation, it exhibits a higher flexibility and robustness that is so far unmatched by any deterministic model. Further investigation of the multimodel approach should be carried out as most studies focused on the assessment of the ensemble gain over the individual models composing it, or on land use change, and little work has been done in a context of operational forecasting. Also, the role that individual models composing the multimodel pool and capacity of the multimodel approach to decipher structural uncertainty deserves more attention.

Meteorological ensembles are now recognized for their ability to represent the meteorological uncertainty. Since many meteorological agencies now provide meteorological ensemble products, these data become more easily available. As a consequence, flood forecasting agencies and research are progressively moving from deterministic forcing to ensemble prediction systems. Another important aspect raised by Cloke and Pappenberger (2009) is that the majority of studies using MEPS focus only on the prediction of a particular event and more rarely on long time series. The value of hydrological ensembles was demonstrated for flood events but little work addresses both streamflow forecasting through the whole year (thus including the lowest flows) and on the prediction of larger events. Moreover, despite the generalization of the use of such ensembles, most HEPS are still only generated by MEPS forcing and the combination of ensemble forcing with other tools remains under exploited.

Liu and Gupta (2007) insists that an integrated framework for the consideration of uncertainty that facilitates the implementation of data assimilation in a more cohesive, coherent, and systematic way should be established. Among others, Vrugt and Robinson (2007) draws positive conclusions about the use of the Ensemble Kalman Filter in an operational context. However, it remains unclear what sources of uncertainty are explicitly accounted for with the EnKF as generated ensembles are not systematically reliable. According to Liu and Gupta (2007), there is a lack of experience, knowledge and guidelines in the implementation of data assimilation techniques. They add that assimilation used alone is not sufficient to decipher all sources of uncertainty. Thus, there is room for improvement in the understanding, quantification, and reduction of uncertainty to reach an optimal framework for hydrological data assimilation.

General thesis objective

This thesis aims to identify a suitable framework to explicitly decipher, characterize, and reduce the main sources of uncertainty of the hydrometeorological modeling chain, in a coherent way in order to issue accurate, reliable, and valuable forecast. To obtain a coherent description and reduction of uncertainty, the different sources of uncertainty need to be addressed individually, with specific tools in a transparent way, avoiding "black boxes". Therefore, the action of these tools should not overlap each other by deciphering twice the same uncertainty and possibly compensating for other unaddressed uncertainties. This relates to Kirchner's statements (2006): "Getting the right answers for the right reasons". To achieve the aim, four goals are defined.

Objective 1: Multimodel approach in an operational context

The first objective is to set up 20 hydrological models on catchments at hand and to force them with deterministic meteorological forecasts to assess the benefits related to the additional information provided by a hydrological multimodel approach. At this stage, data assimilation will be performed with a simple output updating technique. This allows to investigate the multimodel contribution for handling rainfall-runoff model structural uncertainty and to explore the role that ensemble members play. To go beyond traditional assessment against best member or ensemble mean, the multimodel ensemble is also compared to Hydrotel, a distributed hydrological model that is currently used operationally for watershed management and climate change studies.

This objective presents an operational interest since it is based on a substantial database that will broaden the knowledge about multimodel, simple data assimilation, and applications for regions where snow has a great influence on the hydrological regime. This objective is the basis of the project since the subsequent elements of a more comprehensive forecasting framework will be built upon directly on this base.

It is expected that ensemble constituted by the 20 hydrological models exhibits performance at least comparable to that of Hydrotel.

Objective 2: Implementation and investigation of the EnKF

The second objective aims at identifying EnKF parametrization in order to reduce uncertainty related to hydrological initial conditions. To do so, the EnKF will be implemented on the 20 hydrological models to perform model state updating based on streamflow assimilation.

There is no clear guideline concerning the EnKF optimal setting, notably the specification of model input and output perturbations that are needed to account for the various sources of uncertainty and the identification of the state variables that should be updated. Thus a comprehensive screening of required perturbations will be carried out for each model to ensure streamflow simulations accuracy and reliability. Also, a systematic testing of all possible state variable combinations will be tested to identify the ideal set of state variables that has to be included in the updating process. Additionally, a particular attention will be paid to the EnKF and hydrological model adequacy. To minimize additional interferences with meteorological forecast uncertainty, perfect forecast (i.e. observations) will be used to force the models.

This objective should contribute to a better understanding of the EnKF in a hydrological context and more specifically, to the identification of the amount of uncertainty it can decipher in simulation and forecasting modes.

Objective 3: Accounting for the three main sources of uncertainty

The third step of this thesis intends to create a framework that explicitly accounts for the three major sources of uncertainty in hydrometeorological modeling, i.e. meteorological forcing, model structure, and initial condition uncertainty.

The benefits that the EnKF, MEPS forcing, and multiple models bring to the forecasting system will be investigated by either combining or alternating these tools. This will lead to the creation of several dissimilar systems. It thus becomes possible to investigate precisely their role and contribution to the quality of the forecast prediction as a function of the forecasting horizon, but also to study their interactions. Particular attention will be paid to the forecast reliability as it reflects the capacity to properly take into account the different sources of uncertainty. This framework can be regarded as potentially operational since it uses MEPS.

Finally, a framework that offers fully probabilistic description of the main sources of uncertainty should be obtained and it is expected that streamflow prediction are both accurate and reliable. Also, the identified forecasting system should decipher uncertainties in a coherent manner, by specific action of the different tools to address their dedicated sources of uncertainty in a cohesive way, without overlapping in their actions and without compensation for uncertainty that may not be addressed. If these requirements are fulfilled, resorting to post processing techniques should not be necessary to achieve reliability.

Objective 4: Investigation of the economic value of dissimilar systems

The last objective aims to investigate the economic value of the different systems that were developed while fulfilling objective 3. Thus, it becomes possible to assess the economic value of systems that decipher a different amount of total uncertainty with the forecasting tools previously mentioned. The relation between economic value and forecast quality (accuracy and reliability) will also be investigated.

For this purpose, several early warning systems based on the previously developed forecasting systems will be assessed with the relative economical value that is a flexible theoretical framework that scales the forecast value between the no-warning and perfect forecast cases.

Practically, this objective should help to quantify the economic gain that can be obtained through increasing system complexity required to decipher the different sources uncertainty and to relate economical value and forecast quality.

Chapter 1

Assessment of a multimodel ensemble against an operational hydrological forecasting system[1]

1.1 Abstract

Ensemble forecasts present an alternative to traditional deterministic forecasts by providing information about the likelihood of various outcomes. An ensemble can be constructed wherever errors are likely to occur within a hydrometeorological forecasting chain. This study compares the hydrological performance of a multimodel ensemble against deterministic forecasts issued by an operational forecasting system, in terms of accuracy and reliability. This comparison is carried out on 38 catchments in the province of Québec for more than 2 years of 6-day-ahead forecasts. The multimodel ensemble is comprised of 20 lumped conceptual models pooled together, while the reference forecast originates from an operational semi-distributed model. The results show that probabilistic forecast outperforms its deterministic counterpart and the deterministic operational forecast system, thanks to the role that each member plays inside the multimodel ensemble. This analysis demonstrates that the multimodel ensemble is potentially an operational tool, even though the specific setup for this study still suffers from underdispersion and needs to take into account additional sources of uncertainty to reach an optimal framework.

[1]The content of this chapter has been published in Canadian Water Resources Journal / Revue canadienne des ressources hydriques, Vol. 40, Iss. 3, 2015. A. Thiboult, F. Anctil, Assessment of a multimodel ensemble against an operational hydrological forecasting system. Authorship: A. Thiboult and F. Anctil designed the experimental setup and performed the analysis. Coding and simulations were carried out by A. Thiboult. The text has been written by A. Thiboult and revised by F. Anctil.

1.2 Introduction

Streamflow forecasting is a cornerstone for water management, civil protection and reservoir operation (Krzysztofowicz, 1999; Block et al., 2009; Dietrich et al., 2009; Ramos et al., 2010). Despite efforts dedicated toward the development of efficient operational systems, attaining accuracy and reliability remains a daunting challenge as the sources of uncertainty in the hydrometeorological chain are many (Walker et al., 2003).

Uncertainty sources are spread from the initialization of meteorological prediction systems to decision-making tools. Among them, it is generally admitted that the main sources include meteorological forcing, hydrological initial conditions, time-invariant parameters and model structure (Ajami et al., 2007).

For the last few years, in seeking reliable simulations, the hydrometeorological community has been progressively shifting from deterministic simulations to probabilistic ones based on ensembles: a representative sample of the possible future outcomes. An ensemble allows framing uncertainty but, moreover, its median is frequently more accurate than any of the members (Velázquez, 2010). This stems mainly from the fact that inaccurate models presenting uncorrelated errors may be combined in a way that is on average better than the members taken individually (Thompson, 1977). These ensembles can be built wherever uncertainty exists along the hydrometeorological modeling chain.

Many meteorological agencies have now adopted ensemble forecasting (e.g. the Meteorological Service of Canada, the European Center for Medium-Range Weather Forecasts, the Japan Meteorological Agency and more). This venue directly benefits hydrological forecasting for issuing probabilistic hydrological predictions based on an ensemble weather forecast that explicitly takes into account the meteorological forcing uncertainty (Cloke and Pappenberger, 2009; Velázquez et al., 2009). Many studies claim that ensembles allow better decision making and outperform deterministic forecasts (e.g. Boucher et al., 2012; Abaza et al., 2014).

The uncertainty in hydrological initial conditions has been intensively examined, mainly via data assimilation to reinitialize a model on externally measured variables, creating a pertinent set of initial conditions for the next time step (Liu and Gupta, 2007). Data assimilation may be used to, among other things, update model states (Clark et al., 2008b; Seo et al., 2009), possibly along with parameters (Moradkhani et al., 2005), or other variables like snowpack or some hydraulic information (DeChant and Moradkhani, 2011).

A large part of the total uncertainty arises from the hydrological modelling elements of the chain, where model parameters, conceptualization and structure add up to form an aggregate of uncertainty difficult to decipher and predict (Walker et al., 2003).

Model parameter uncertainty has been extensively investigated. Beven and Binley (1992) pioneered providing multiple possible answers through the Generalized Likelihood Uncertainty Estimation (GLUE), which allows producing an ensemble of parameter sets equally likely (from a mathematical point of view) to describe the catchment behavior. Because hydrologists have yet to identify a perfect model structure, there is no reason that a particular set of parameters should represent the "truth," so GLUE allows the estimation of uncertainty related to parameters through the equifinality principle (Beven and Binley, 1992; Beven and Freer, 2001). Many other techniques have been proposed to estimate uncertainty related to parameters, focusing on calibration (Vrugt et al., 2003), temporal variability of parameters (Thiemann et al., 2001), spatial variability (Feyen et al., 2008), combining stochastic methods and expert knowledge (Dietrich et al., 2009) and varying objective functions (Yapo et al., 1998; Gupta et al., 1998), sometimes combined with assimilation (Vrugt et al., 2005; Liu and Gupta, 2007).

Gourley and Vieux (2006) carried out a study to identify the sources of uncertainty in a hydrometeorological chain. They argued that dealing with input and parameter uncertainties may not be sufficient for encompassing the streamflow forecast error and that using different conceptualizations would be a more appropriate strategy to overcome this issue. This would create a way out of the endless quest for the perfect model and allow harnessing the inaccuracies of existing models in an ensemble forecasting framework to issue better estimates of predictive error. Georgakakos et al. (2004) combined a set of 11 calibrated and uncalibrated models forced by radar observations to issue forecasts for six catchments, qualifying their set-up as a potential operational tool. Clark et al. (2008a) confirmed that uncertainty arises from the structure of the model itself by assessing 79 unique model structures built out from four pre-existing hydrological models. This framework led them to conclude that "it is unlikely that a single model structure provides the best streamflow simulation for multiple basins" (see also Ajami et al., 2006; Breuer et al., 2009). They emphasized that taking into account the structural uncertainty is as important as considering parameter uncertainty. It seems that the main limitation to a multimodel ensemble may be the lack of dissimilarity between structures, which leads to underdispersed forecasts. This statement is substantiated by Viney et al. (2009), who argued that "the best ensembles are not necessarily those containing the best individual models," but those bringing diversity.

Ajami et al. (2006) investigated several simple means for merging members, namely the simple model average, the multimodel superensemble and the weighted average, and concluded that despite their simplicity, the ensemble average generally performs better than any member taken individually. More sophisticated methods, like Bayesian model averaging, were explored by Raftery et al. (2005) and Duan et al. (2007) to create probabilistic forecasts. Ajami et al. (2007) exploited the integrated Bayesian uncertainty estimator scheme to combine input, parameter and structural errors, and confirmed that merging the outputs leads to a more reliable forecast since considering only parameter uncertainty resulted in a large underdispersion. Velázquez (2010) confirmed that retaining all the output members and offering a fully probabilistic forecast is preferable over an aggregate of outputs, preserving all the information available.

To the authors' knowledge, despite the attention that ensemble and, more specifically, multimodel ensemble modelling have aroused, comparisons have mostly been carried out between an ensemble and one or several members composing the ensemble. These studies have helped to recognize the benefit associated with the use of several members, but the comparisons are mostly not made in an independent way. There is a lack of comparisons of a multimodel ensemble against an independent model, i.e. a model which is not part of the multimodel ensemble. Comparing a multimodel ensemble to one of its members provides information only about the ensemble's relative performances, and one should ensure that the reference model (the model used for comparison) is proven in order to evaluate the gain of the system under scrutiny.

This study aims to assess gains related to the additional information provided by a hydrological multimodel ensemble over an operational deterministic model, and to investigate multimodel properties. The available data set is larger than that used in most previously mentioned papers, striving for more global and generalized information instead of focusing on a single event of interest. As Andreassian et al. (2009) recommended, tests are carried out on demanding conditions. Hydrological forecasts are issued for catchments that are notably affected by snow accumulation and spring freshet.

Hydrotel, a semi-distributed hydrological model operationally utilized for public dam management (Turcotte et al., 2004) by the Centre d'Expertise Hydrologique du Québec (CEHQ), is driven by deterministic weather forecasts from Environment Canada over a 2.5-year period to create the reference hydrological forecast. The multimodel ensemble is generated from 20 lumped hydrologic models selected for their diversity and forced by the same meteorological inputs as the operational system, to ensure a fair comparison.

The paper is organized as follows: a brief description of the catchments and computing tools is provided in the Methodology section, followed by a comparison between hydrologic forecasts in terms of performance and reliability. Further investigation into the multimodel ensemble is presented at the end of the Results section. Concluding statements are provided in the last section.

1.3 Methodology

The case study catchments and data are presented along with the description of scores selected to assess the performance and reliability of the forecasts, followed by a description of the output updating used to correct simulated streamflows.

1.3.1 Catchments and hydrometeorological data

The case study spans over 387,000 km^2 with coverage primarily of the province of Quebec, but also Ontario and the states of New York and Vermont (Figure 1.1). The area is situated between 43°15'N and 52°20'N latitudes and 68°85'W and 81°20'W longitudes, where the climate is classified as wet continental and the hydrologic regime is dominated by a spring freshet. The flows are structured into three major river systems: the Outaouais River, the Saguenay River and the lower part of the St. Lawrence River downstream of its confluence with the Outaouais River. More specifically, the database includes 38 catchments of various sizes and discharge levels (Table 1.1 and Table A.1). Snow data are not directly available as no continuous in situ measurements are carried out. To estimate the solid precipitation (Table 1), all available precipitation and temperature observations are used to force the snowmelt module. The annual mean is computed next from simulated snow accumulation time series.

Table 1.1: Mean annual characteristics of the catchments

	Catchment area (km^2)	Streamflow (m^3/s)	Solid precipitation (mm)	Total precipitation (mm)
Min	236	5	218	985
Max	15342	299	501	1544
Median	1275	27	349	1184

Climatological time series (of minimum and maximum daily temperatures and 24-hour total precipitation) were provided on a 1220-point grid (of 0.1° resolution) by the CEHQ. These data were created by applying kriging to observations made within the study area, and applying an elevation-based temperature correction of -0.005°C/m. The data set extends from 1969 to 2010, yet the 1990-2000 period was chosen for the calibration of the various hydrological models.

Hydrotel processes meteorological inputs. Thiessen polygons are used to define weights associated to each grid point situated inside each catchment. In order to drive all hydrological

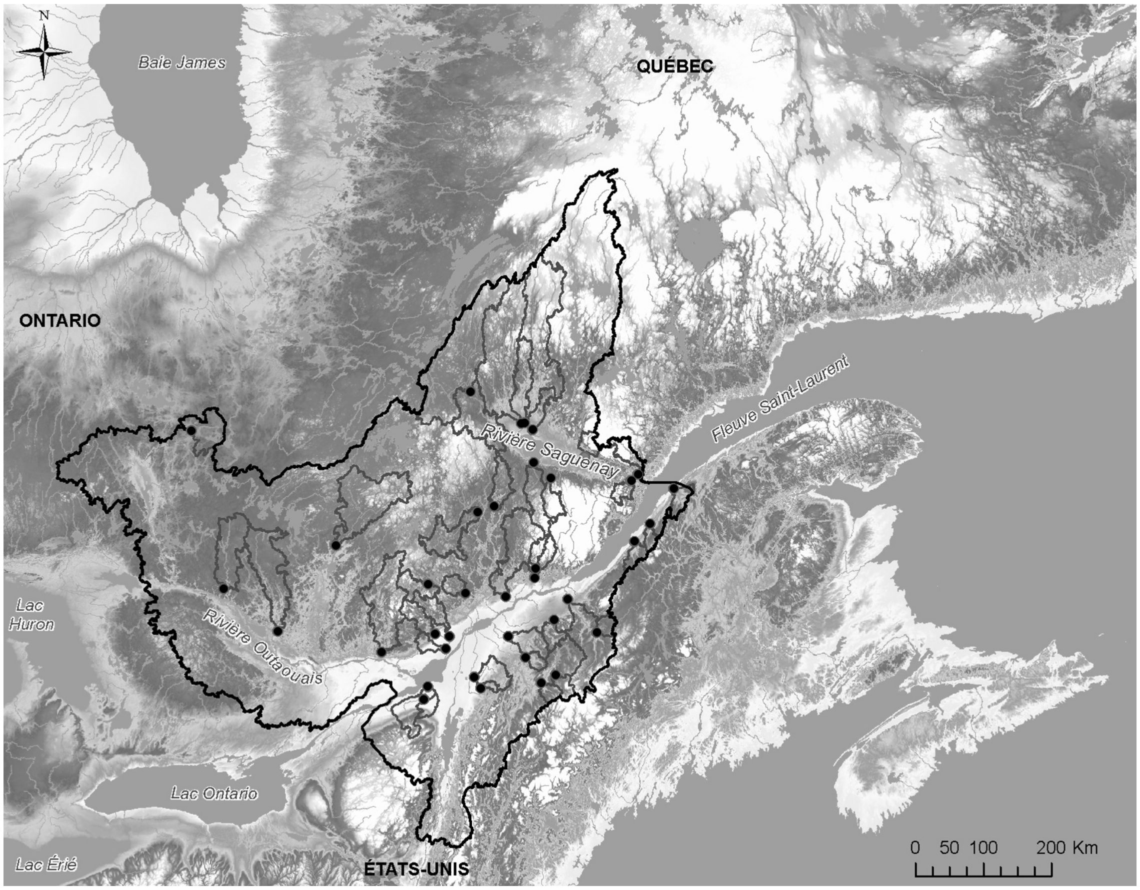

Figure 1.1: Catchments and hydrological stations.

models with the same meteorological information, the semi-distributed model is run first and the resulting evaluation of the climatology over the entire catchment is provided next to the lumped models.

1.3.2 Precipitation and temperature forecasts

The deterministic meteorological forecasts were issued by the Canadian Meteorological Center regional model over a 2-day horizon at spatial and temporal resolutions of 15 km and 3 hours, respectively. The forecast lead time is extended to 6 days by adjoining forecasts from the 35 km resolution global model. All forecasts from October 2008 to December 2010 were disaggregated by the CEHQ to the 0.1° grid of the climatological data by the nearest neighbor method (Gaborit et al., 2013). For the sake of this study, the meteorological data are aggregated to a daily time step.

The 3 years preceding the forecast period are used for model spin-up. Models are then forced with observations to bring their states to values that are representative of the catchment conditions at the beginning of the period of interest. In forecasting mode, hydrological forecasting and continuous simulation alternate. The forecast is issued by forcing the models with the meteorological forecast from $t = 0$ to $t = 6$. Models are then forced with observations from $t = 0$ to $t = 1$ to create the new initial condition set for the forecasts from $t = 1$ to $t = 7$. These steps are repeated over the entire forecasting period.

1.3.3 Hydrological models

Hydrotel is a semi-distributed model that simulates typical hydrological variables and processes such as soil water content, snow accumulation and melt, evapotranspiration, vertical water balance and surface runoff. It was conceptualized from a physical description of the catchment by Fortin et al. (1995). The model is mostly used by the CEHQ for the management of public dams in the Province of Québec (Turcotte et al., 2004). The present study relies on a calibration performed by the CEHQ using the Shufle Complex Evolution (SCE; Duan et al., 1992).

The multimodel ensemble pools 20 lumped models chosen for their structural diversity. The initial model selection was carried out by Perrin (2000) to investigate structures and parameter complexity performance following an extensive bibliographic research. This selection was revised by Seiller et al. (2012) for hydrological projections. Careful attention has been given to favor parsimony: model complexity should remain low except if more complexity provides a substantial gain in performance and robustness. This also ensures limiting issues related to parameter uncertainty and equifinality (Beven and Binley, 1992) during the calibration process, as the number of parameters is kept low.

The retained models rely on conceptual reservoirs to describe the principal processes of the hydrological cycle, and were developed in different contexts as some were initially intended for daily or monthly simulation or specifically for Nordic hydrology, for instance.

Some original models needed to be modified to match the ensemble frame. Potential evapotranspiration and snow accumulation and melting are computed externally. Additional model modifications have been performed upon model parameters, structure or space discretization. For some models, the number of calibrated parameters has been reduced by fixing parameters identified according to developers' comments, or after a sensitivity analysis. Simplifications in the structure may have also been carried out. The new structures were kept if they provided better results than original models. A time-delay function for routing is implemented for models that do not possess one in order to simulate catchments with a time of concentration greater than 1 day. Finally, a parametric logistic function is added to the models that required other catchment characteristics than hydrometeorological series.

Figure 1.2 summarizes the repartition of parameters and reservoirs of the 20 models. The final selection exhibits models with low to moderate complexity (four to 10 calibrated parameters and two to seven reservoirs). A more detailed description of model modifications can be found in Perrin (2000). Acronyms are used to emphasize that the models may substantially differ from their original version.

The lumped models are driven by precipitation and potential evapotranspiration. The latter is calculated from the empirical formulation proposed by Oudin et al. (2005), based on the daily mean air temperature and the calculated extraterrestrial radiation.

A single snow module (Cemaneige, Valery et al., 2014b) is used prior to running the lumped models. The two-parameter module accumulates solid precipitation and relies on a degree-day approach modulated by an energy balance index. It discretizes the catchment into five bands of equal elevation. The precipitation provided to the lumped models is thus the sum of the liquid precipitation and snowmelt water. The snowmelt module and individual models are calibrated together. The parameter sets retained for the snow module are influenced by the hydrologic model used and their parameterization. The hydrologic ensemble is therefore coupled with a snow accounting process that takes into account the snow module parameter uncertainty.

The lumped models are calibrated using the SCE method and the root mean squared error (RMSE) of square-rooted streamflow as the objective function over the same 10-year period (1990 to 2000) selected by the CEHQ for the calibration of the Hydrotel model.

1.3.4 Hydrologic forecast correction via streamflow assimilation

A simple assimilation technique is used to adjust hydrological forecasts. Output updating (Refsgaard, 1997) compares forecasted and observed streamflow on the date of forecast. Then the difference between observed and simulated streamflow at time t=0 is subtracted from each day of the 6-day forecast with a damping coefficient that depends on lead time, starting with 1 the first day and decreasing to 0 on day 6:

$$Q_{upd}(t=i) = (Q_{obs}(t=0) - Q_{for}(t=0))\sigma(i) + Q_{for}(t=1) \tag{1.1}$$

where $Q_{upd}(t = 1)$ is the updated forecast of the i^{th} lead day, $Q_{obs}(t = 0)$ the observed streamflow at the time of forecast emission, $\sigma(i)$ the empirically determined damping coefficient which decreases by 0.2 unit every day (i.e. σ = 1 at t = 1, σ = 0.8 at t = 2, etc.) and $Q_{for}(t = i)$ the streamflow forecast at the i^{th} day.

Table 1.2: Main characteristics of the 20 lumped models (from Seiller et al., 2012)

Model acronym	Number of parameters	Number of reservoirs	Derived from
M01	6	3	BUCKET (Thornthwaite and Mather, 1955)
M02	9	2	CEQUEAU (Girard et al., 1972)
M03	6	3	CREC (Cormary and Guilbot, 1973)
M04	6	3	GARDENIA (Thiery, 1982)
M05	4	2	GR4J (Perrin et al., 2003)
M06	9	3	HBV (Bergström and Forsman, 1973)
M07	6	5	HYMOD (Wagener et al., 2001)
M08	7	3	IHACRES (Jakeman et al., 1990)
M09	7	4	MARTINE (Mazenc et al., 1984)
M10	7	2	MOHYSE (Fortin and Turcotte, 2007)
M11	6	4	MORDOR (Garçon, 1999)
M12	10	7	NAM (Nielsen and Hansen, 1973)
M13	8	4	PDM (Moore and Clarke, 1981)
M14	9	5	SACRAMENTO (Burnash et al., 1973)
M15	8	3	SIMHYD (Chiew et al., 2002)
M16	8	3	SMAR (O'Connell et al., 1970)
M17	7	4	TANK (Sugawara, 1979)
M18	7	3	TOPMODEL (Beven et al., 1984)
M19	8	3	WAGENINGEN (Warmerdam et al., 1997)
M20	8	4	XINANJIANG (Zhao et al., 1980)

It is out of the scope of this study to explore statistical post-processing methods specific to hydrological ensemble forecasts for achieving a better representation of the uncertainty, such as regressions (e.g. Gneiting et al., 2005), kernel dressing (e.g. Roulston and Smith, 2003), Bayesian model averaging (BMA Raftery et al., 2005) and Bayesian processor of ensemble (BPE Krzysztofowicz and Maranzano, 2004). Exploring a raw multimodel ensemble allows us to investigate the role played by each hydrological model.

1.3.5 Scores

The hydrological forecasts are assessed in terms of accuracy, resolution and reliability. Reliability indicates the degree of statistical consistency between probability forecasts and the observed frequency of occurrence of a particular event. For instance, a reliable 80% interval should on average contain the observation eight times out of 10. Resolution evaluates the ability of discriminating between two events which are different. In the case of deterministic forecast, resolution is a measure of the distance between forecast and observation.

Traditional deterministic scores like mean absolute error (MAE) cannot be used when using probabilistic forecast, but distinctive scores need to be used in order to compare forecast

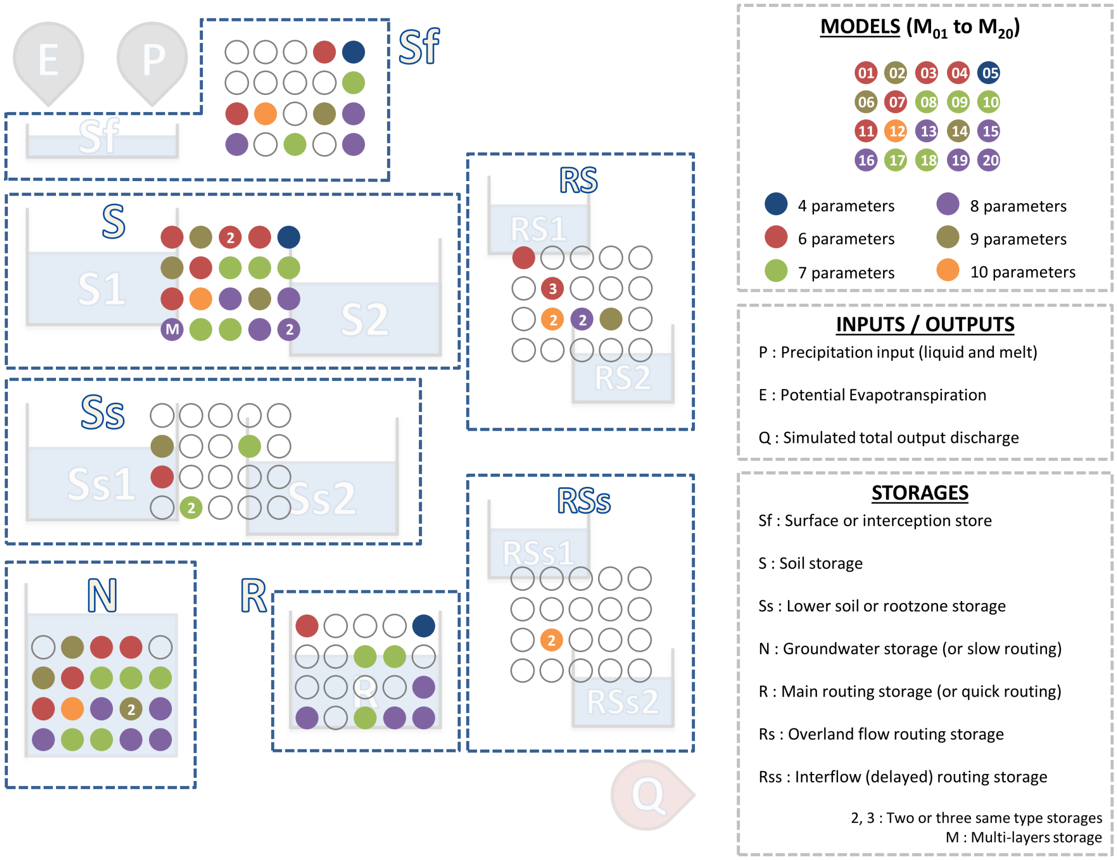

Figure 1.2: Illustration of model structural diversity (from Seiller et al., 2012)

probability and the observation frequency.

Reliability and resolution are both assessed by the continuous ranked probability score (CRPS Matheson and Winkler, 1976) (CRPS), which is the integral form of Brier's score:

$$BS = \frac{1}{N}\sum_{t=1}^{N}(p_t(x) - H_t)^2 \tag{1.2}$$

where $p_t(x)$ is the forecasted probability of occurrence of an event x and H_t is the heavy side function which is 1 when the event happens, 0 otherwise, and N the length of the time series.

The CRPS is "sensitive to the average ensemble spread and the frequency and magnitude of the outliers" (Hersbach, 2000).

$$CRPS(F_t, x_t) = \int_{-\inf}^{+\inf}(F_t(x) - H(x \geq x_t))^2 dx \tag{1.3}$$

where $F_t(x)$ is the predictive cumulative distribution function for day t, x is the predicted variable, x_t is the corresponding observed value and H is the heavy side function. CRPS evaluates the performance for one time step, so the MCRPS is defined as the average of CRPS over the entire period. An advantage of this score is that it can be compared to the MAE (Gneiting and Raftery, 2007), and consequently allows the comparison of deterministic and probabilistic forecasts.

The ranked histogram, also known as the Talagrand diagram, is used to evaluate reliability. It is an easy way to visualize how the ensemble and its members are located with respect to the observation. It assesses the reliability of the ensembles by subdividing them into boxes which are delimited by the members of the ensemble. The representation gives the frequency at which the observation falls into a specific bin. A flat histogram usually implies a reliable ensemble since the observation is equally likely to be situated in between any two members in average. More rarely, a rank histogram may appear uniform even in the presence of a conditional bias (Hamill, 2001).

The rank histogram is also used in this study in a non-traditional way to track each member among the ensemble. Instead of looking where the observation falls within the ensemble, attention is paid to where the individual members fall within the ensemble. This produces one rank histogram per member, hereafter called model rank histograms. Flatness is not necessary in that context because shape is not related to reliability anymore, but only to the place where the member under consideration is situated most frequently. This tool allows for an investigation of the role of each member of the ensemble (i.e. each different lumped hydrologic model).

1.4 Results

This section addresses three aspects. The first consists of a qualitative hydrograph comparison. The second concerns gains that a multimodel ensemble provides over the deterministic hydrological prediction system. The last one investigates the role and contribution of individual members in the multimodel ensemble.

1.4.1 Hydrograph Analysis

An example of streamflow prediction from 7 February to 7 June for the Dumoine River is displayed in Figure 1.3. The hydrographs represent the observed and simulated rate of discharge according to time. Unlike traditional forecast hydrographs where the origin of the hydrograph represents the first forecasting day and the x-axes represents the lead time, the hydrographs display the concatenation of values for the same lead times (1, 3 and 6 days ahead). Percentages on the figure and the grey shades associated with them denote the theoretical confidence interval of the multimodel ensemble. The theoretical confidence interval would be equal to the "true" confidence interval in the case where the multimodel ensemble is perfectly reliable.

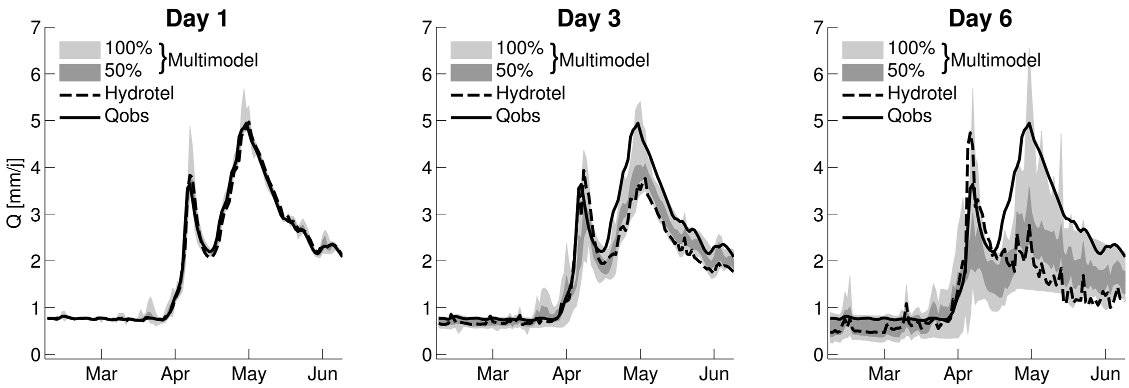

Figure 1.3: Multimodel ensemble and Hydrotel hydrographs for the Dumoine River. The grey shades depict the percentage of the theoretical confidence interval.

The principal advantage of the multimodel ensemble over the deterministic hydrological prediction system is that it is more often capable of predicting events which are harder to forecast (i.e. rare or extreme events or two consecutive streamflow peaks, as in Figure 1.3). Indeed, using many distinct models increases the possibility of one or few of them encompassing an event. For example, on 29 April for the 6th day in Figure 1.3, two models out of 20 surpass the observed streamflow, while Hydrotel strongly underestimates the observation. Multiple models also possess the ability to compensate for an erroneous prediction made by some of the ensemble members. For example, on the first week of April for the 6th day, one lumped model largely underestimates the streamflow, though the 50% ensemble theoretical confidence interval is barely affected.

Snowmelt simulations are indirectly influenced by the multimodel. Even if there is only a single snowmelt module, snowmelt uncertainty is partially taken into account thanks to the use of several parameter sets leading to differences in the snow accumulation and melt (Figure 1.4 – only the middle elevation band is represented, as the behavior of the different bands is very similar because of the weak catchment elevation variations). The dynamic of the simulated snow cover by the different snow module parameterizations does not differ substantially, since it still relies on a single structure (snow water equivalents for each snow module tend to be parallel) but provides an estimate of uncertainty about the snow cover depth.

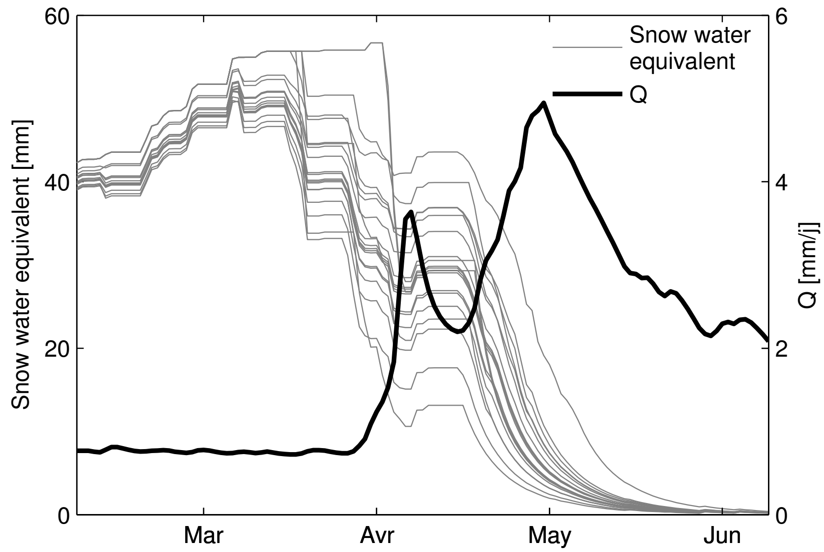

Figure 1.4: Simulated snow water equivalent stock for the different snowmelt module parameter sets and the middle elevation band for the Dumoine River.

1.4.2 Deterministic and probability forecast performance

Levels of performance of probabilistic and deterministic forecasts are reported in Figure 1.5, in which catchments are sorted by increasing Hydrotel MAE. The MAE allows a comparison of deterministic forecasts (the multimodel ensemble MAE is computed from the median value of the ensemble), while the MCRPS assesses multimodel ensemble forecast. As mentioned earlier, one may compare MAE and MCRPS.

Performance varies with catchments and horizons, but multimodel MAE is systematically better than Hydrotel MAE for all 38 catchments. This demonstrates that, even if the multimodel forecast is reduced to its median, the multimodel outperforms traditional deterministic forecast. This arises from the fact that models, if they are uncorrelated – or poorly correlated, tend to cancel out each other's errors. The gain lies in the combination of models, and not in the individual performances of the model used in the multimodel ensemble as shown in Figure 1.6. Individual models exhibit performance that are different but close to each other while the combination of them clearly stands above all 20 models. Also, Hydrotel is among the best models with constant performances for the validation period for all catchments, but is outperformed by the ensemble. These results are in agreement with those in Figure 1.3. The multimodel ensemble reduces the risk inherent in relying on a single (possibly misleading) model (Hagedorn et al., 2005).

There is also a net gain in retaining all multimodel outputs. When the multimodel is reduced to its median (multimodel MAE), performance decreases. This underlines the added value

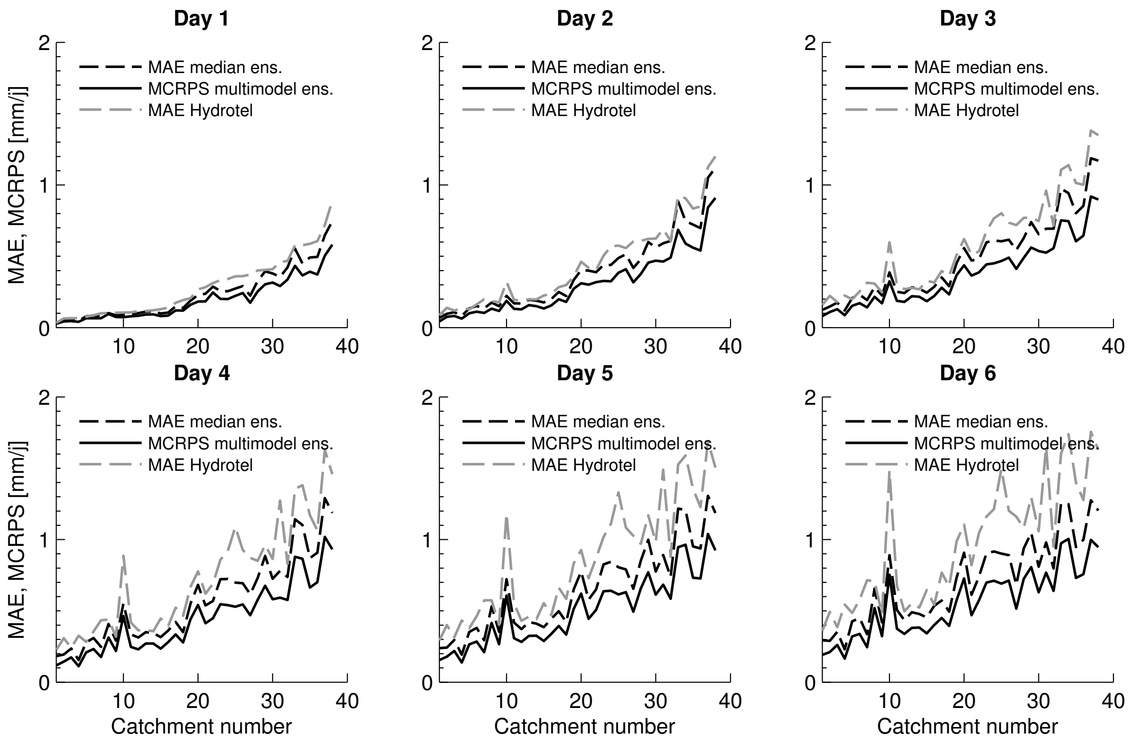

Figure 1.5: Mean Absolute Error (MAE) and Mean Continuous Ranked Probability score (MCRPS) of the deterministic and probabilistic forecasts for all 6-day horizons and catchments sorted by increasing Hydrotel MAE.

of the probabilistic forecast over the deterministic one, i.e. considering a probability density function rather than a single point forecast. Note that the difference between Hydrotel MAE and MCRPS scores grows as performance decreases.

The reliability of the multimodel is assessed. Figure 1.7 shows rank histograms pooled over all catchments. The multimodel ensemble is consistently underspread. The output updating preserves the rank histogram shape for the different lead times, but harms the spread (the members are corrected individually regardless of their initial position in the ensemble and they are squeezed around a certain value). The predictive uncertainty of the whole system is therefore lower than the multimodel uncertainty estimation without the output updating. The multimodel ensemble reliability may possibly be improved using an assimilation technique that better preserves the spread, and using ensemble meteorological forcing.

1.4.3 Ensemble member characteristics

Model rank histograms are grouped in Figure 1.8. They depict the role played by each lumped model (i.e. each member) in the multimodel ensemble, by illustrating where they tend to fall with respect to other ensemble members. A uniform histogram indicates that the

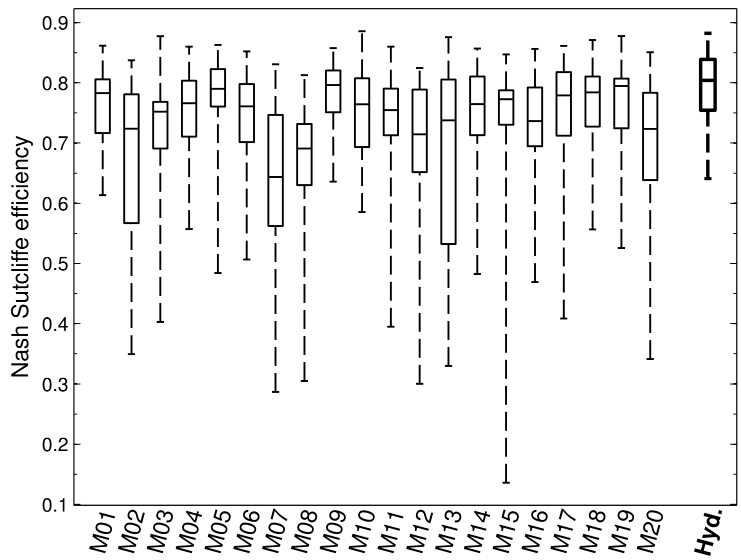

Figure 1.6: Individual models and Hydrotel Nash Sutcliffe efficiency for a 10-year validation period and the 38 catchments.

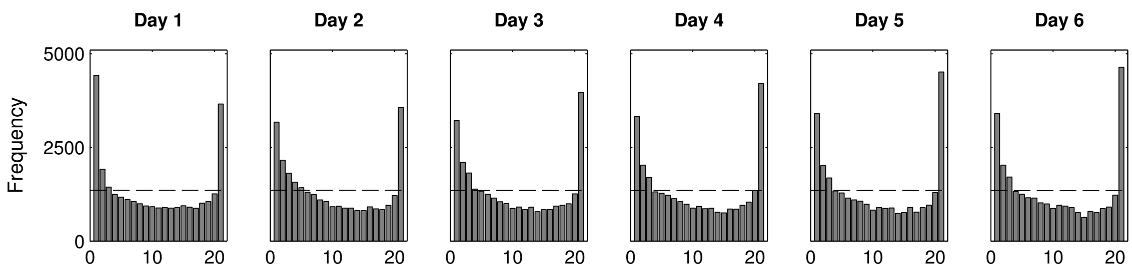

Figure 1.7: Multimodel ensemble rank histograms for daily discharges for each lead time, combining the time series from all catchments.

model occupies every rank equally, while a heterogeneous one identifies a model's preference to occupy different areas of the ensemble's spread. Contrary to the standard rank histograms illustrated in Figure 1.7, they do not assess reliability. Therefore, it is not intended to obtain a flat histogram, it is used as a tool to investigate the member position within the ensemble.

The rank histograms of each model in Figure 1.8 are sorted in ascending MAE values in order to relate the role of the model within the ensemble to its individual performance. The best model is situated in the top left corner, the worst on the bottom right corner. Only the 6^{th} lead time is shown, as it is representative of model characteristics at other lead times.

The different aspects of the rank histograms illustrate the way that the models complement one another. Some models seem more liable to capture outliers (higher or lower part of the ensemble) while others remain centered. The sorting of the models according to the level

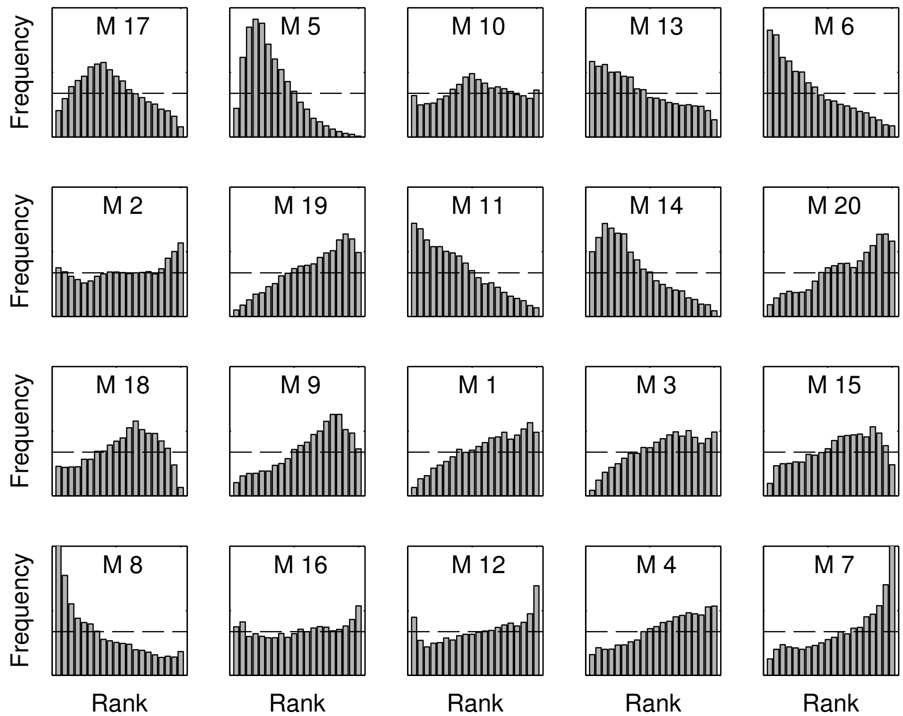

Figure 1.8: Model rank histograms sorted by increasing mean absolute error (MAE) values, from left to right and then top to bottom, for all catchments for the 6-day lead time. Each subplot represents the frequency of falling into a specific rank for the model under consideration (vertical axes) with its corresponding rank (horizontal axes).

of their individual performance does not reveal any particular behaviors. Interestingly, no model presents a pronounced U-shaped rank histogram, which means that no model has a structure/conceptualization that allows it to contribute to both ends of the predictive distribution.

1.5 Conclusion

Issuing accurate and reliable hydrological forecasts is still an outstanding challenge. Probabilistic forecasts grow in popularity in particular because they provide information about the uncertainty in the hydrometeorological modeling chain and the resulting forecasts. Within the many sources of uncertainty, hydrologic model structure and conceptualization are among the dominant ones. A multimodel ensemble that samples these sources of error has been proposed to tackle these issues simultaneously.

This study provided a comparison between a semi-distributed deterministic forecast system against one derived from an ensemble of 20 lumped conceptual models chosen for their dissimilar conceptualizations. The multimodel ensemble was also compared to its deterministic counterpart (i.e. ensemble median) to identify the benefits of using probabilistic forecasts. Finally, the characteristics of the multimodel ensemble itself were assessed, exploring the role that each member plays with respect to the rest of the ensemble.

Ensemble forecast was assessed in terms of resolution and reliability. It performs better than its deterministic counterpart in terms of resolution assessing MCRPS and MAE. Even if the multimodel ensemble only takes into account conceptualization and structural uncertainty (when driven by deterministic meteorological forecasts), it provides overall more accurate streamflow prediction than deterministic forecast systems do. Multimodel ensembles also present the advantage of being more likely to encompass events which are harder to predict. Moreover, the hydrological multimodel allows an indirect partial handling of snowmelt simulation error through the calibration process. Relying on several models increases the chance that at least one of them will be able to forecast any specific event (particularly rare extreme events). According to a rank histogram analysis, multimodel ensembles tend to be reliable, but they still occasionally lack spread. Attention has also been given to the role that the members of the multimodel ensemble play. No common pattern between model accuracy and rank was identified, suggesting that members play different roles inside the ensemble and that they contribute in diverse ways.

Despite its success with respect to an operational model, this multimodel setup (exploiting very simple model structures) may still be improved by considering other sources of uncertainty, such as meteorological uncertainty. Since snow accumulation and melt may have a great influence on the spring freshet simulation, the multimodel approach could be extended for those processes too. Generating larger ensembles by including different snow modules may serve to partially encompass this potentially large source of uncertainty. Finally, the streamflow assimilation technique could be improved by taking into account the ensemble's spread.

Chapter 2

Seeking optimal tuning of the Ensemble Kalman Filter[1]

2.1 Abstract

Forecast reliability and accuracy is a prerequisite for successful hydrological applications. This aim may be attained by using data assimilation techniques such as the popular Ensemble Kalman filter (EnKF). Despite its recognized capacity to enhance forecasting by creating a new set of initial conditions, implementation tests have been mostly carried out with a single model and few catchments leading to case specific conclusions. This paper performs an extensive testing to assess ensemble bias and reliability on 20 conceptual lumped models and 38 catchments in the Province of Québec with perfect meteorological forecast forcing. The study confirms that EnKF is a powerful tool for short range forecasting but also that it requires a more subtle setting than it is frequently recommended. The success of the updating procedure depends to a great extent on the specification of the hyper-parameters. In the implementation of the EnKF, the identification of the hyper-parameters is very unintuitive if the model error is not explicitly accounted for and best estimates of forcing and observation error lead to overconfident forecasts. It is shown that performance is also related to the choice of updated state variables and that all states variables should not systematically be updated. Additionally, the improvement over the open loop scheme depends on the watershed and hydrological model structure, as some models exhibit a poor compatibility with EnKF updating. Thus, it is not possible to conclude in detail on a single ideal manner to identify an optimal implementation; conclusions drawn from a unique event, catchment, or model are likely to be misleading since transferring hyper-parameters from a case to another may be hazardous. Finally, achieving reliability and bias jointly is a daunting challenge as the optimization of one score is done at the cost of the other.

[1] A modified version of this chapter has been published in Journal of Hydrology Vol 529, Iss 3, 2015. A. Thiboult, F. Anctil, On the difficulty to optimally implement the Ensemble Kalman filter: An experiment based on many hydrological models and catchments. Authorship: A. Thiboult and F. Anctil designed the experimental setup and performed the analysis. Coding and simulations were carried out by A. Thiboult. The text has been written by A. Thiboult and revised by F. Anctil.

2.2 Introduction

Despite the modelling advances in representing hydrological processes and providing more accurate streamflow forecasts, there is still a need for reducing and quantifying uncertainty. Most hydrological prediction systems are still deterministic and provide only the most likely outcome without addressing estimates of their uncertainty. The sources of uncertainty stem from multiple places in the hydrometeorological chain such as in inputs, initial conditions, parameter estimation, model structure, and outputs (e.g. Ajami et al., 2007; Salamon and Feyen, 2010; Liu and Gupta, 2007; Liu et al., 2012) and these uncertainties should be deciphered to enhance model predictive abilities and reliability for efficient decision making (Ramos et al., 2010).

A broad range of techniques has been developed to control uncertainty at different levels such as the Generalized Likelihood Uncertainty Estimation (GLUE), Shuffle Complex Evolution Metropolis algorithm (SCEM) for parameter uncertainty (Beven and Binley, 1992; Vrugt et al., 2003) and BMA combination technique for structural uncertainty (Jeremiah et al., 2011; Duan et al., 2007; Parrish et al., 2012; Ajami et al., 2007). Proper initial conditions are frequently identified as one of the main factors that contributes to an accurate forecast (DeChant and Moradkhani, 2011; Lee et al., 2011). Among others, data assimilation (DA) is commonly used in hydrometeorology to reduce initial condition uncertainty and proved to be a useful tool for modelling. DA incorporates observations into the numerical model to issue an analysis, which is an estimation of the best current state of the system. This has not only been largely applied to remote sensing for snow (Kuchment et al., 2010), soil moisture estimates (Forman et al., 2012; Meier et al., 2011; Renzullo et al., 2014; Alvarez-Garreton et al., 2014) or hydraulic information (Bailey and Bau, 2012), but also to update radar forcing (Harader et al., 2012; Kim and Yoo, 2014). Many applications also use in situ observations such as catchment discharge, snowpack measurements, or soil moisture to update models (e.g., Seo et al., 2009; Clark et al., 2008b; Thirel et al., 2010a; DeChant and Moradkhani, 2011; Franz et al., 2014). In addition, DA may be coupled with parameter optimization (Vrugt et al., 2005; Moradkhani et al., 2005; Nie et al., 2011).

Sequential DA techniques such as particle filter and the Kalman filter family are frequently used for recursive updating of the states of a system, each time an observation is made available. Among them, the ensemble Kalman filter (EnKF, Evensen, 1994) proved to be a powerful tool for hydrological forecasting (DeChant and Moradkhani, 2012; Rakovec et al., 2012; Vrugt and Robinson, 2007; Weerts and El Serafy, 2006; Abaza et al., 2014) that is effective and reliable enough for operational use (Andreadis and Lettenmaier, 2006). Several studies claim that they developed techniques that improved upon traditional EnKF (e.g., Clark et al., 2008b; Whitaker and Hamill, 2002) by focusing on the relaxation of constrains of traditional

EnKF implementation, or by explicitly including time lag between the soil moisture and the discharge in the updating process (Li et al., 2013, 2014; McMillan et al., 2013).

A key feature of EnKF is the proper specification of hyper-parameters (perturbations of inputs and outputs) and model states to be updated (Moradkhani et al., 2005). In most studies, EnKF implementation is based on an a priori selection of the hyper-parameters and updated states combination, which is then scarcely justified. Noteworthy exceptions are Moradkhani et al. (2005) and Chen et al. (2013), but these studies are very specific as they are performed on a single model and one or two catchments. Accurate perturbations representing error estimates are crucial since the EnKF updating scheme is based on the weighting of the model and observation relative error. However this specification is complex in practice as the different sources of uncertainty experience strong interactions (Moradkhani et al., 2006; Hong et al., 2006; Kuczera et al., 2006). Several attempts to account explicitly for structural error have been reported, for example by directly adding perturbations to the state variables (Reichle et al., 2002; Vrugt et al., 2006; Clark et al., 2008b), or by updating model parameters (Moradkhani et al., 2005; Vrugt et al., 2005; Naevdal et al., 2003).

Moreover, despite encouraging results, DeChant and Moradkhani (2012) points that little research has been done to examine the effectiveness and robustness of EnKF and that "studies need to provide a more rigorous testing of these techniques than has previously been presented". Another issue that needs consideration is that EnKF performance is mostly discussed as 'standalone', regardless of the influence of the coupling with the hydrological model. This is mainly due to the fact that EnKF is often tested on a single model. Thus, the question of adequacy between the DA technique and the model is rarely assessed.

The present study aims at identifying EnKF parametrization to reduce and quantify optimally the uncertainty related to initial conditions in a forecast mode. A second scope addresses the question of EnKF and hydrological model adequacy. In order to achieve this, the analysis is conducted on 20 structurally dissimilar lumped conceptual models, 38 catchments, 12 hyper-parameter sets, and all possible combinations of the state variables to strive for general results. Finally, the effectiveness of identifying the best EnKF parametrization without exploring all combinations is discussed.

Section 2.3 presents EnKF's basics, models, basins and scores. Section 2.4 presents the results of the DA techniques followed by a discussion and the conclusion statements are provided in section 2.5.

2.3 Material and methods

2.3.1 Hydrological models, snowmelt modules, and PET

The EnKF is tested individually on the 20 lumped conceptual models that were presented in Section 1.3.3. Because they are based on diverse hydrological concepts and present different degrees of complexity (4 to 10 calibrated parameters and 2 to 7 reservoirs to represent perceptual and conceptual hydrologic processes), they allow to test the EnKF in a comprehensive manner according to structure diversity (see Table 1.2).

The models exploit various conceptualizations and thus their parameters and state variables perform particular roles in simulating rainfall-runoff processes. Their reservoirs may describe systems ranging from precipitation interception to routing (or more conceptual functions). The role of state variables is not detailed in the article for concision purpose. For the same reason, the state variable values before and after the analysis step will not be discussed here but only the outputs of the models, i.e., simulated streamflow will be considered. For further details on state variable meaning, refer to Perrin (2000).

The potential evapotranspiration formulation and the snow accounting routine are the same as in Chapter 1.

2.3.2 State updating and EnKF implementation

EnKF addresses explicitly initial conditions uncertainty by creating an ensemble of possible model reinitializations by updating state variables according to a recursive Bayesian estimation scheme. It estimates the true probability density function of the model states conditioned by the observations.

The evolution of the model state variables vector x may be described through time with a non-linear forward operator M driven by the previous states, the deterministic forcing u that includes an error term ζ_t, and the (time-invariant) model parameters θ to which a model error η is added. The η error term does not include only state variable error but also implicitly other sources of error such as the structural and parameter error or forcing error.

$$x_t = M(x_{t-1}, u_{t-1}, \theta) + \eta_t \tag{2.1}$$

States and observations z are related through the following expression

$$z_t = H(x_t, \theta) + \epsilon_t \tag{2.2}$$

with H being the observation function and ϵ_t the observation error.

The EnKF relies on an approximation of Bayesian rule to identify the conditional density of the model states, $p(x_t|z_{1:t})$, given the previous time steps observations $z_{1:t}$, where x_t is the state vector that contains the model states. EnKF needs several realizations (N members) to derive the model error matrix. As the real true state is unknown, it is approximated by the ensemble mean:

$$\overline{x_t} = \frac{1}{N}\sum_{i=1}^{N} x_t^i \tag{2.3}$$

where i refers to the i$^{\text{th}}$ member. The model error matrix is thus defined as the difference between the true state and the single hydrological model realisations:

$$\mathbf{E}_t = (x_t^1 - \overline{x_t}, x_t^2 - \overline{x_t}, ..., x_t^N - \overline{x_t}) \tag{2.4}$$

Therefore, the model covariance matrix can be defined as:

$$\mathbf{P}_t = \frac{1}{N-1}\mathbf{E}_t\mathbf{E}_t^T \tag{2.5}$$

When an observation is available, model states are updated ($\mathbf{X}^+$) as a combination of the prior states $\mathbf{X}^-$ and the difference between the prior estimate $\mathbf{H}_t\mathbf{X}_t^-$ and observation.

$$\mathbf{X}_t^+ = \mathbf{X}_t^- + \mathbf{K}_t(z_t - \mathbf{H}_t\mathbf{X}_t^-) \tag{2.6}$$

The Kalman gain $\mathbf{K}_t$ represents the relative importance of the observation error with respect to the prior estimate (i.e. model simulation) and acts as a weighting coefficient. $\mathbf{R}_t$ denotes the covariance of the observational noise.

$$\mathbf{K}_t = \mathbf{P}_t\mathbf{H}_t^T(\mathbf{H}_t\mathbf{P}_t\mathbf{H}_t^T + \mathbf{R}_t)^{-1} \tag{2.7}$$

Since the identification of the Kalman gain equation members is complex, the term $\mathbf{P}_t\mathbf{H}_t^T$ is approximated by the forecasted covariance between the model states and the simulation estimates and $\mathbf{H}_t\mathbf{P}_t\mathbf{H}_t^T$ by the variance of the estimate.

$$\mathbf{P}_t\mathbf{H}_t^T = \frac{1}{N-1}\sum_{i=1}^{N}(x_t^i - \overline{x_t})(\mathbf{H}_t x_t^i - \overline{\mathbf{H}_t x_t})^T \tag{2.8}$$

$$\mathbf{H}_t\mathbf{P}_t\mathbf{H}_t^T = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{H}_t x_t^i - \overline{\mathbf{H}_t x_t})(\mathbf{H}_t x_t^i - \overline{\mathbf{H}_t x_t})^T \tag{2.9}$$

A more detailed description of EnKF equations and mathematical background can be found in Evensen (2003). In this study, the filter has been implemented following Mandel's (2006) computational recommendations.

A critical point in the EnKF implementation is a proper identification of the errors ϵ, ζ, and η because they will determine the observation predictive distribution. In a vast majority of cases, model error is not directly identifiable. Users need to estimate it by using stochastic perturbations through ϵ and ζ computation if no more direct estimation of model error η is made through perturbation of states or updating state variable and parameters conjointly. Note that adding perturbations to states and parameter updating are also subject to inaccuracies since they are "based on order-of-magnitude considerations, and may therefore be statistically unreliable" (Liu et al., 2012). In the present study, only ϵ and ζ are considered." Errors are assumed to be normally distributed with zero mean but their variances need to be put under scrutiny.

Hyper-parameters define the statistical properties of the forcing and observations ensembles. This study concentrates on the influence of the uncertainty in precipitation and temperature forecasts and in streamflow observation. Three precipitation perturbations (with a standard deviation corresponding to 25%, 50%, and 75% of the initial precipitation forecast magnitude), two streamflow perturbations (with a standard deviation corresponding to 10% and 25% of the observation), two temperature perturbations (standard deviation of 2ºC and 5ºC) are evaluated. These perturbations are centered around the perfect forecast or the observation. We thus obtain 12 sets of hyper-parameters. All errors are assumed to be uncorrelated. Note that the potential evapotranspiration is not directly perturbed but it is computed by the Oudin et al. formula (2005) forced with a temperature ensemble creating a subsequent set of perturbed PET values.

The present updating scheme relies on the Markov property that asserts that the future of the system is dictated only by the present state, not on the anterior sequence of observations. Model states are consequently updated according to the instantaneous covariance between states and the current streamflow observation while observations that preceded it are not incorporated. Li et al. (2013) affirms that this assumption may harm updating performance of models that incorporate unit-hydrograph routing but do not affect models that include dynamic routing stores, which is the case for 19 of the 20 models used in this study. Only model 5 (GR4J) is based on a unit-hydrograph approach.

Prior to the hyper-parameters evaluation, the number of members composing the EnKF ensemble is investigated. Four sizes (25, 50, 100 and 200 members) are tested on two sets of hyper-parameters. The experiment concerning the number of members was not conducted on all hyper-parameter sets to reduce computational cost.

The number and the combination of states to be updated are next put under scrutiny. Batch testing is used to investigate all states (reservoirs) combinations for each model regardless of their physical meaning. The number of possible combinations thus depends on the model at hand, varying from 3 for the 2-reservoir models up to 127 for the 7-reservoir model. As all combinations of state variables and hyper-parameters are tested, some cases turned out to be unrealistic and prone to make the EnKF unstable. This difficulty was overcome by setting back unrealistic states within their theoretical boundaries identified during calibration.

The EnKF is used to update daily model's states whenever streamflow observations are available. The model is then forced with the perfect meteorological forecast to issue a 10-day hydrological forecast. This lead time is sufficiently long to be able to see the effect of DA vanishing. A series of tests (not shown here) indicated that after 10 days, the influence of data assimilation is almost negligible for almost every model and catchments. Thus extending the forecast would bring no additional information.

Finally, every 20 models are tested on 38 catchments, 12 different hyper-parameter sets, and all possible reservoir combinations.

2.3.3 Scores

Probabilistic scores offer the possibility to evaluate more than individual member or ensemble mean and provide to the forecaster a better picture of the forecast probability distribution quality. Probabilistic forecasts should be assessed both in terms of accuracy and reliability to assess where the verification is situated among the ensemble, how the frequency of forecasted events corresponds to the frequency at which events are observed, and the gain of ensemble over deterministic forecast.

The Normalized Root-mean-square error Ratio (NRR) is used to quantify the spread of the ensemble with regard to its predictive skills (Murphy, 1988). A value of 1 indicates an appropriate spread, while greater and smaller values than 1 reflect too narrow and wide ensembles respectively. The NRR is function of the observation y_t, the ensemble forecast average, the number of members in the ensemble N, and time t.

$$NRR = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\left[\frac{1}{N}\sum_{n=1}^{N}\hat{y}_t^n\right] - y_t^n\right)^2}}{\frac{1}{N}\left\{\sum_{n=1}^{N}\sqrt{\frac{1}{T}\left[\sum_{t=1}^{T}(\hat{y}_t^n - y_t^n)^2\right]}\right\}\sqrt{\frac{N+1}{2N}}} \tag{2.10}$$

A complementary view of the NRR is the Spread Skill Plot (SSP) which is a graphical assessment that represents at the same time the bias of the ensemble, its spread and therefore its reliability. The SSP relies on the fact that the Root-Mean-Square Error (RMSE) should match the spread to achieve reliability (Fortin et al., 2014). In the case where the RMSE is greater than the spread, the ensemble is overconfident regarding to its predictive skills and vice versa.

The commonly used Nash Sutcliffe efficiency (Nash and Sutcliffe, 1970) is used to assess the bias of the median of the EnKF ensemble. A NSE value equals to 1 identifies a perfect prediction while a value below 0 indicates that the average observation is more skillful.

To evaluate the improvement or deterioration of the quality of the simulations, NSE and NRR gains are computed as:

$$G = \frac{S^{sim} - S^{ref}}{S^{opt} - S^{ref}} \tag{2.11}$$

with G the gain, S^{sim} the score after the state updating, S^{ref} the score without updating (the open loop), and S^{opt} the perfect score. In the case of the NRR, which is not a monotonic score (i.e. the minimization or maximization of the value does not systematically indicate an improvement or a decrease of the reliability), a substitution is performed to compute the gain. We consider that under and overdispersion should be penalized the same way which is reflected by the distance from the optimal score 1.

$$NRR^* = |NRR - 1| \tag{2.12}$$

NRR^* is then negatively oriented and bounded by 0 and can be used to compute the gain G.

2.3.4 Catchments and hydrometeorological data

The 38 watersheds used for this study are identical to the ones presented in Section 1.3.1.

2.3.5 Individual model performances

The study investigates EnKF implementation for many models i.e. the performance of the coupling of hydrological models and DA. It is not intended to compare models performances to each other. The 20 models are thus investigated separately and the comparison of models performance with or without EnKF updating is out of scope. However, one should recognize that models have different initial performance as shown on Figure 1.6. Their individual performance varies largely over the 38 catchments but no model consistently out performs or under performs the others in all situations. Best (or worst) results are frequently obtained by different models for different catchments. Therefore, performance after EnKF updating should be compared to each other only in terms of gain.

2.3.6 Meteorological forecast

The first scope of this paper is to reduce and quantify the uncertainty related to the watershed initial conditions for hydrological forecasting. Thus, to focus on that specific aspect, we do not use actual weather forecasts but meteorological observations to force the models. This ensures to minimize the error related to forcing as the remaining inaccuracies can be attributed to the representativeness of the measurements and the measurement errors rather than the many uncertainties related to weather forecasting. This forcing will be referred hereafter as 'perfect forecast'.

2.4 Results

2.4.1 An estimation of the required ensemble size

Ideally, one would propagate a large number of members to ensure to accurately sample the state variable probability density functions but this would increase drastically the computational cost. Thus, the first part of the study aims at identifying an approximation of the minimal number of member necessary to drive EnKF without performance loss. For this case, the hyper-parameters are fixed to $0.50P$ for precipitation, $0.1Q$ for streamflow, and 2º for temperature for graphical convenience, and different numbers of members are tested ($N = 25, 50, 100, 200$). The influence of the number of members has been carried out with other sets of hyper-parameters and led to the same conclusions.

Figure 2.1 shows simulations general behavior according to the number of members N. NSE and NRR are used to assess forecast accuracy and reliability, respectively. On the upper sub-plots are displayed 12768 points corresponding to all simulations per set of hyper-parameters, i.e. the results in simulation for every catchment, model and existing reservoir combination for a given model. As we want to quantify the effect of N on every simulation but also more precisely on best performing ones, the lower sub-plots display the best results by catchment and model. This implies to retain only the best state variable combination for updating. A single best simulation can be identified for a particular score for a catchment, but the simulation may be different whether it is assessed regarding its reliability or its bias. To overcome this selection issue, we retained the simulation offering the highest NSE among the three best NRR. This combined criterion ensures to keep the most reliable simulation in first place and then the lowest bias. Reliability is chosen as first order criterion to ensure to cover as well as possible the initial condition uncertainty without diminishing the EnKF spread.

Interest is set on forecasts with NSE and NRR close to 1, thus the vertical axis has been truncated for readability. Negative NSE are not shown even if they represent about 1.5% of the total number of simulations. Note that the NRR score is bounded for underdispersed distribution by $1/\sqrt{(N+1)/2N})$ and thus a NRR score below 0.8 or above 1.2 indicates doubtlessly a poorly reliable ensemble. For this purpose, an area is defined to delimit the ranges for deemed acceptable results ($0.8 < NRR < 1.2$ and $0.7 < NSE$) and is represented as a grey shade on Fig 2.1. The ratio ϕ of simulations having performance falling inside the aforementioned range to total number of simulations is displayed for every N. Additionally,

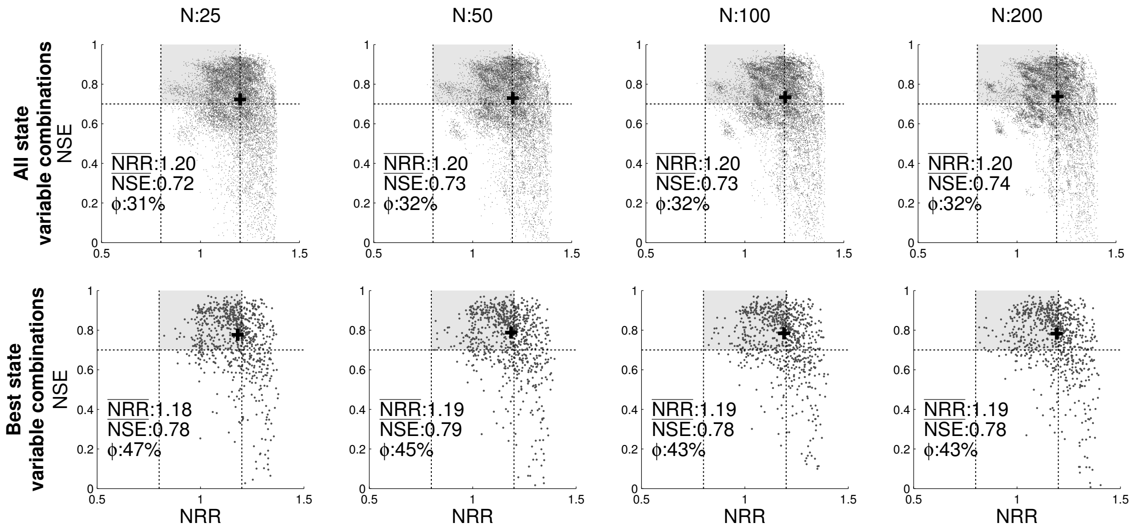

Figure 2.1: Influence of the number of members N on NSE and NRR in simulation

median NSE and median NRR of simulations are depicted as a cross on each plot. The range of acceptable results may seem permissive, but it has to be wide enough to encompass a reasonable number of models and watersheds. By defining a more demanding range, there is a risk that all the points inside it belong to a small number of model and catchment pairs and that the variations inside this range are only due to the state variable choice which therefore harms the representativeness of the ϕ ratio.

Results are very similar for different values of N, for all the simulations or only the best ones. The sampling of the states variable probability density function is more subject to stochastic errors when the number of member is low, but results are largely similar with various values of N. The weak difference between the number of members indicates that it is reasonable to keep only 25 or 50 members to limit sampling error. Further results will be presented for 50 members.

2.4.2 Influence of hyper-parameters

Figure 2.2 depicts the bias and reliability for the 12 hyper-parameter sets for day 1. Unlike the number of members, the additional error to forcing and observation is a driving parameter. The precipitation perturbations have the greatest influence on performance followed by temperature and streamflow. Note that the importance of the temperature perturbations has to be considered regarding the dominant role of the spring freshet for the studied watersheds.

For a given hyper-parameter set, the cloud of simulations is greatly dispersed on both reliability and bias axes. The diversity of performance is important and depends on model and

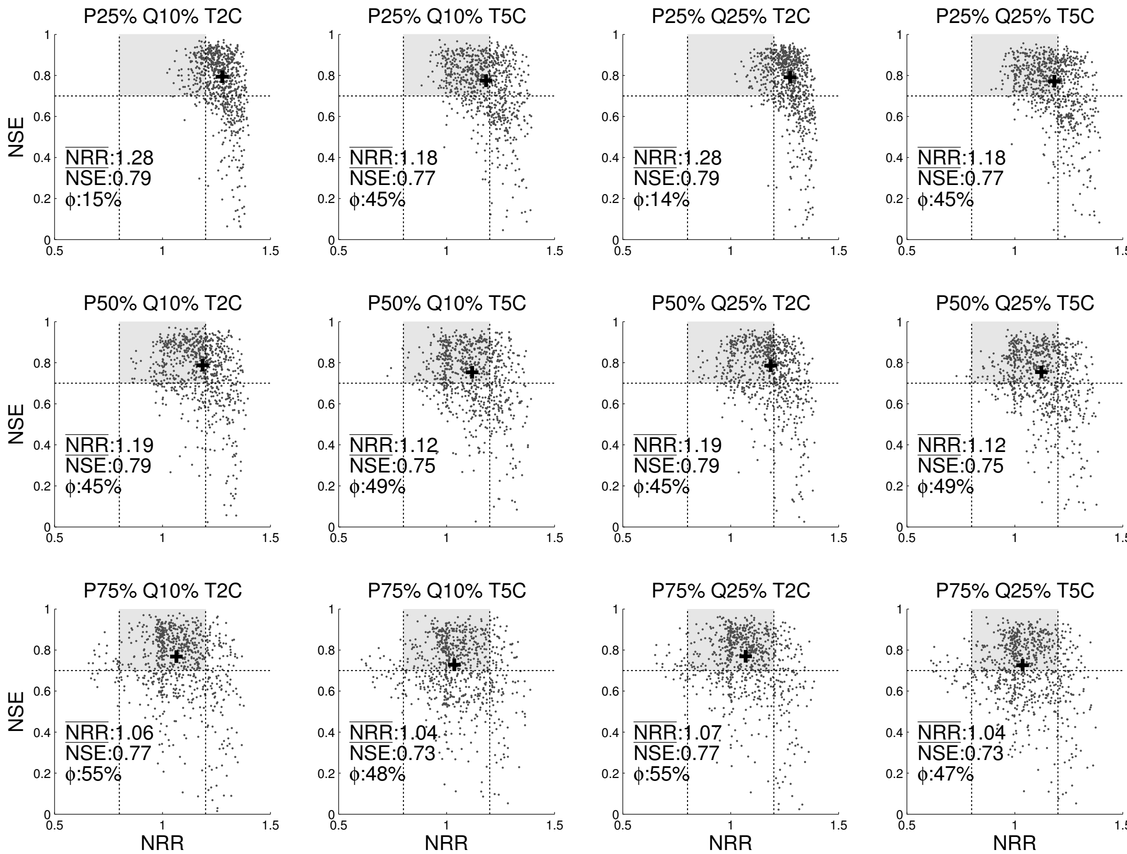

Figure 2.2: Influence of added perturbations to precipitations, temperatures and streamflows on NSE and NRR for day 1

catchment for a specific hyper-parameter set.

Reliability is more often achieved by using important perturbations. The lowest perturbation set (P 25%, T 2ºC, Q 10%) fails to encompass initial condition uncertainty as only 15% of simulations fall into the acceptable range of performance. Reliability best results are obtained with perturbations that are clearly unrealistically high to describe measurement uncertainty. However, few simulations using a hyper-parameter set that include low input perturbations, are close to perfect reliability. These simulations are very likely to become over-dispersive when the added noise magnitude increases, indicating that using too large perturbations also contributes to decrease reliability.

On the other hand, best NSE are found for lower perturbations. Though, the calibration of hyper-parameters is more sensitive to reliability as the drop in bias is less severe than the drop in reliability.

The simulations indicate that in a vast majority of cases, EnKF should not be used with best estimates of real forcing uncertainty as it will lead to under-dispersive ensemble if no additional technique is used to explicitly decipher other sources of uncertainty. To achieve optimal implementation, the noisy forcing has to take into account not only real forcing uncertainty but it needs to compensate for the model structural and parameter uncertainty. This contributes to drastically increase the difficulty of identifying the correct covariances.

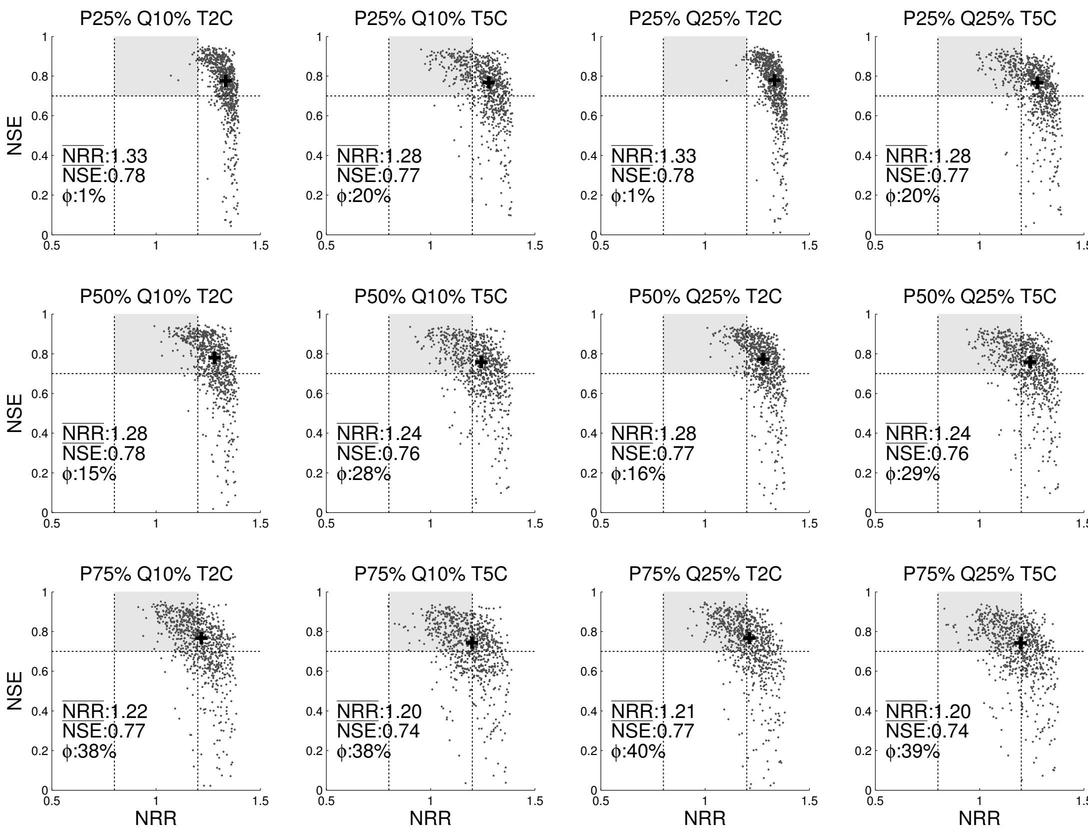

Figure 2.3: Influence of added perturbations to precipitations, temperatures and streamflows on NSE and NRR for day 3

Figure 2.3 displays the same bias-reliability representation in forecasting mode, but for the 3-day-ahead lead time where a global loss of reliability is observed. Ensembles become overconfident with increasing lead time but the bias remains approximatively the same. Only a 75% perturbation of the precipitation manages to provide more than 30% of acceptable results. For the 7$^{\text{th}}$ day (result not shown here), the percentage of acceptable results never exceeds 13%. This suggests that the uncertainty in initial conditions does not account for all sources of uncertainty, even if it encompasses more than real forcing and observation error. As this trend becomes more obvious with increasing lead time, it may also indicate that the information provided by the EnKF ensemble is not propagated during forecasting process.

Relative performance of hyper-parameters remains unchanged with lead time as the best performing hyper-parameter sets for reliability or bias are the same for day 1 or day 9. The model's performance in forecasting thus depends on the quality of the DA initialization.

Typical results are plotted on Figure 2.4 to illustrate the spread and the bias ($RMSE$) of the EnKF ensemble and the gain which is related to the difference between EnKF $RMSE$ and the open loop $RMSE$. The most common situations are cases a) and b) in which the ensemble is reliable or close to reliability for the first forecasting day as the spread matches or is close to match the EnKF $RMSE$. The spread diminishes quickly after the first forecasting day while the error of the ensemble increases. Although the ensemble spread should grow to match the increasing error, it collapses and the ensemble becomes overconfident. However, the error remains lower than the open loop forecasting confirming the gain provided by the EnKF.

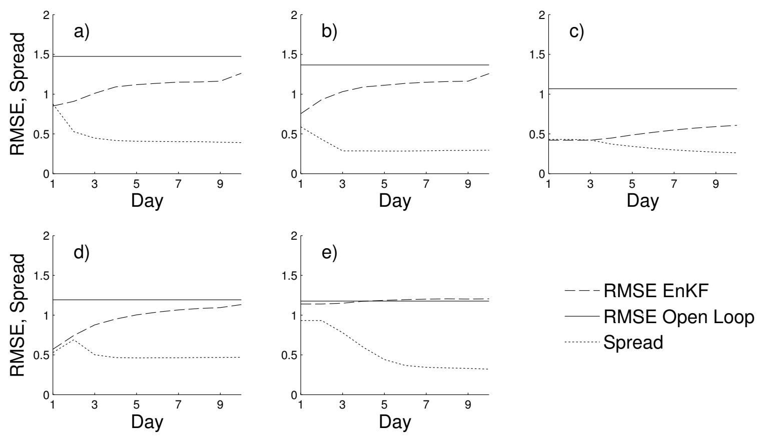

Figure 2.4: Typical spread-skill plots in forecasting. a) model 9 and watershed 28, b) model 5, watershed 28, c) model 1, watershed 36, d) model 5, watershed 26, e) model 7, watershed 13.

Less frequently, the spread may remain constant up to three days (case c) or for a very particular situation as in case d), which happens only for model number 5 (GR4J), the spread may increase for up to 2 days before eventually dropping. Finally, case e) reports a case where DA is unable to improve forecasting. In the latter case, EnKF simulation is quite similar to the open loop. This is explained by the state variable selection process where only the best state variable combination is kept. For this very particular model-watershed pair, DA works poorly and all state variable combinations deteriorate accuracy beyond the open loop performance. Consequently, the best simulation is achieved by the state variable combination influencing the least the streamflow and this combination is therefore selected. In an operational context and such case, the EnKF would not be used as it does not provide any gain and increases computational costs vainly.

In every cases, DA assimilation starts loosing its efficiency right after the spin up and the spread stops matching the EnKF $RMSE$ for longer lead times. This confirms that additional sources of uncertainty should be taken into account from the first forecasting day to achieve a reliable system and also implies that finding the best hyper-parameters only guarantees to find the optimal initialization without ensuring forecasting performances. Also, as the decrease in spread is always observed independently of model and catchment, it is possible to conclude that this behavior is specifically related to the EnKF.

To investigate the possible relation between models and hyper-parameters, Figure 2.5 shows the frequency at which a hyper-parameter set outperforms the other for a given model. Each of the 20 sub-plot corresponds to a model. The 12 hyper-parameters are referred to by numbers as following:

1: P25%Q10%T2ºC	2: P25%Q10%T5ºC	3: P25%Q25%T2ºC
4: P25%Q25%T5ºC	5: P50%Q10%T2ºC	6: P50%Q10%T5ºC
7: P50%Q25%T2ºC	8: P50%Q25%T5ºC	9: P75%Q10%T2ºC
10:P75%Q10%T5ºC	11:P75%Q25%T2ºC	12:P75%Q25%T5ºC

The bars represent the frequency at which a hyper-parameter set outperforms the other. The upper part and lower part of the figure refer to the bias and reliability respectively. For instance, the hyper-parameter set number 10 is the best one for approximatively 10% of the catchments for the bias and 35% for reliability for the model 1.

The repartition of the best performing hyper-parameters confirms that no hyper-parameter set performs better than others systematically for the NSE or the NRR and exceeds rarely 40% for any model. Thus, to ensure to get optimal updating performance, an optimal hyper-parameter should be chosen according to model and catchment.

An additional difficulty arises from the fact that bias and reliability are optimized by different hyper-parameters. Optimal NSE values are often obtained by low to moderate noise magnitude while the best NRR are obtained with higher perturbations. This highlights the challenge to optimize bias and reliability collectively during EnKF updating, leaving to the modeler the burden of prioritizing one over the other.

2.4.3 Influence of the choice of states variables

For the present section, results are shown for a particular hyper-parameters set P 50% Q 10% T 5% but similar conclusions could be drawn from the other tested sets.

This section addresses the question of identification of state variables that should be updated with the EnKF. For a N-state-variable-model, it exists $2^N - 1$ combinations and none is fa-

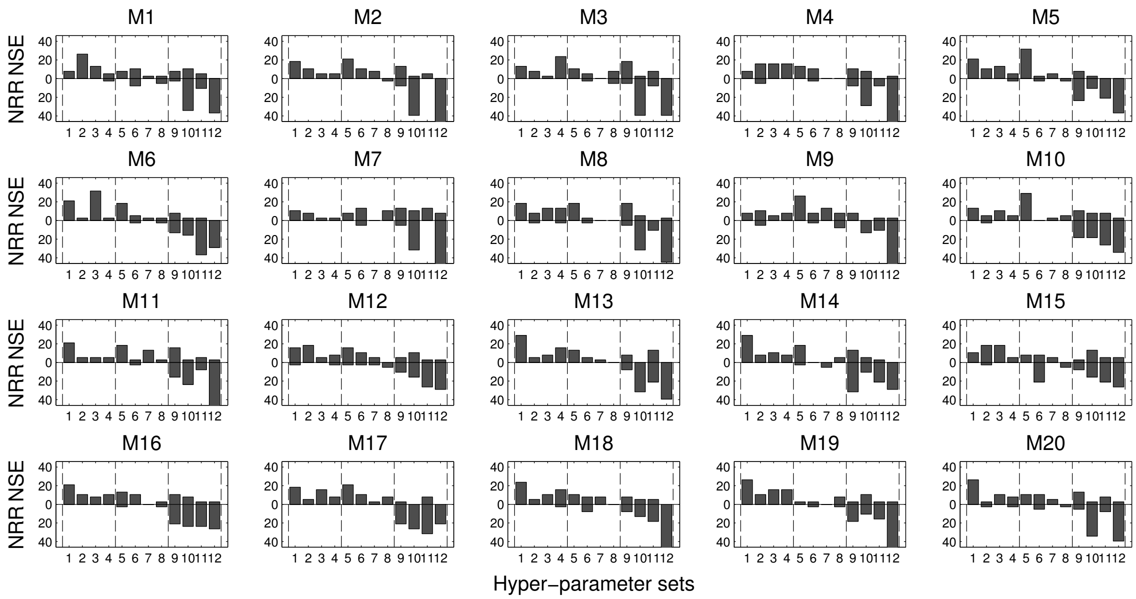

Figure 2.5: Frequency at which each of the 12 sets of hyper-parameters is better than others in term of accuracy (NSE) and reliability (NRR). Each boxplot corresponds to a model and results are displayed for day 3

vored during testing. Thus, the number of possibilities depends on the model; see Table 1.2 for the number of states.

As the reservoirs are situated at different levels in the models (from interception to routing), their individual updating is expected to affect differently model outputs; more precisely they affect the time-lag between state perturbation and the change in simulated streamflow. Seo et al. (2003) suggest to not perturb reservoirs concerning soil moisture as it is a long-term component that has an influence which lasts much longer than the longest operationally used lead time. On the contrary, Wöhling et al. (2006) encourage soil moisture updating as it will act on all lead times. Physically based models offer the possibility to deal with values that are theoretically measurable. The knowledge about these values allows to estimate critical values that are the most subject to uncertainty. Conceptual models states values do not refer directly to a measurable value and the identification of variable states for updating is thus complex. The amount of uncertainty related to these variables is hardly definable and there is no apparent clue to update a certain reservoir rather leaving others unperturbed.

Figure 2.6 presents the distribution of performance of every individual state combination per model, to illustrate the variability of NSE over the 38 catchments. Five models with different numbers of state variable are used to highlight the general 20 models behavior. Each box plot refers to a combination of updated state variable for a model. The box plot situated on the right of each sub-plot corresponds to the case where all model states are updated.

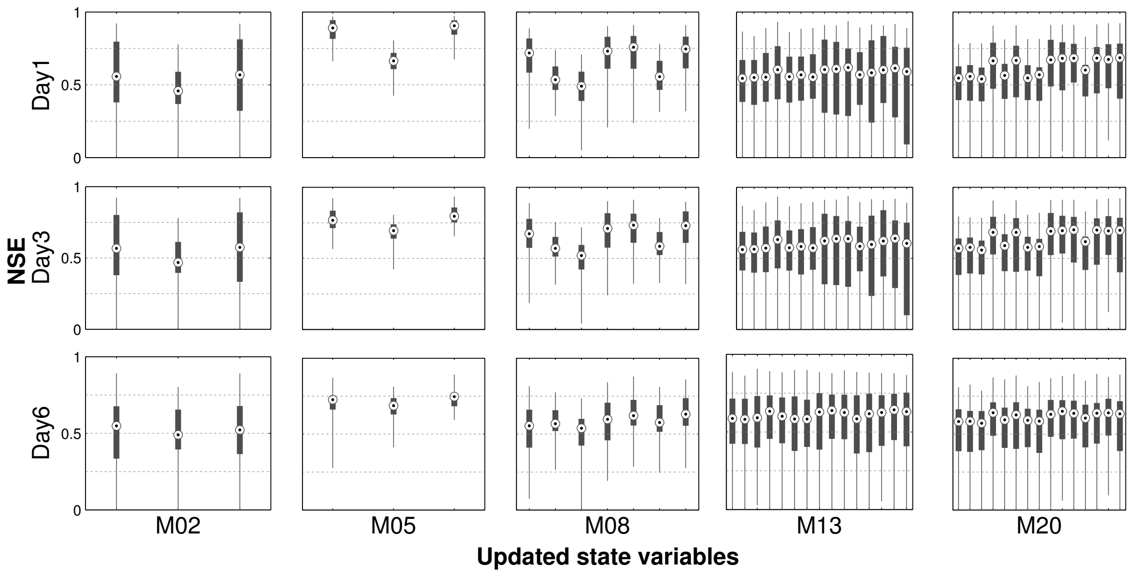

Figure 2.6: Distribution of NSE performance over the 38 catchments for model 2, 5, 8, 13, and 20 according to the updated states parameters for lead time 1, 3, and 6. The boxplot on the right of each sub-plot corresponds to the case where all states variable are updated. For details on the state variable combinations, see Table B.1

The main conclusion is twofold: the success of the updating procedure lies as much in the model as in the choice of the state variables. Most models best state combinations exhibit a median NSE higher than 0.75 for first lead time even if few models (model 2 in particular) seem to react poorly to the state updating. Best short-lead-time-model performance needs to be qualified as it frequently decreases with increasing lead time.

As median NSE values are frequently close to each other, it is possible to conclude that there is no obvious outperforming combination for any model – however, there are combinations that perform consistently poorly. Additional complexity in the choice of the best state combination to update arises from the performance variability over catchments for a specific updated set of states. As the median performance is close and the variability over catchments is important, it is very likely that one combination for a model on a catchment will be outperformed by another combination on another catchment.

As the state updating procedure is numerically implemented in the same way for all models, bad performance may be attributed to the suboptimal choice of updated model states or the potential inadequacy of the EnKF to a specific hydrologic model rather than the EnKF technique itself. On this subject, model number 2 open loop performances are often comparable with other models (see also Fig. 1.6) while its performance after updating are undoubtedly worse.

The question of best state set identification arises also as a function of the lead times (results not shown). In this study, we disregarded lead time specific states combinations since the use of different set of state combinations lead time dependent may improve performance for each lead time in average but it would imply to run in a parallel fashion several simulations for each lead time. An issue arising from such a technique would be the creation of discontinuities in the forecasting streamflows from one lead time to another.

Reservoirs which should be updated in priority are frequently –but not always– the closest reservoirs to the model outputs in the description of the rainfall-runoff process. The question of the number of reservoir to update is more complex as few global patterns emerge from the results. It is common practice to update all model states variable in lumped and semi-distributed models but this does not systematically lead to the best results. On Fig 2.6, model 13 illustrates this since the updating of some state variable sub-ensembles shows improved performance for first, second and third quartiles. Therefore, for some models, optimal updating may be obtained by leaving some stores unperturbed, for instance the routing store for model 13. Generally, the number of updated states remains rather low, never exceeding 4 even for high dimensional states models (model 7, 12 and 14). All states should be updated for model 5, 6, 10 and 19 but other models states should be partially updated. Models with a large number of state variables (high degree of freedom) are more prone to encounter equifinality issues as many outcomes frequently end up close for a specific conditions. This lead to an already known problem that requires the user to take an arbitrary decision or possibly to retain several combination with the associated computational cost increase. Also, likewise for traditional model parameter estimation, the identification of best set depend on the score used as objective function. Selecting states set based on a NSE criterion does not guarantee to maximize other accuracy scores, and even less to achieve highest possible reliability. Thus, different sets of updated states may capture more or less accurately specificities of the hydrograph.

2.4.4 Global and local updating schemes

Setting EnKF catchment specificities is possible and may be operationally conceivable and worth considering. This case is more computationally demanding as states identification needs to be carried out for all watersheds. Thus, the gain of such approach needs to be quantified to justify the increase in commitment. In the opposite case, the forecaster takes a risk relying on optimal updated states set identified from only one catchment if this set is transferred onto another catchment.

Figure 2.7 displays gains obtained by EnKF over open loop. Two EnKF updating schemes are compared:

- A global scheme: updating is carried out with a single set of state variables and hyper-parameter per model, identified as best according to the combined criterion in average over all catchments. The updated states and the hyper-parameters are the same regardless of catchment.

- A local scheme: updating is carried out with a different set of state variables and set of hyper-parameters for each catchment identified as the best set of state variable per catchment. The approach is thus catchment specific.

The gain between the two updating schemes is also examined.

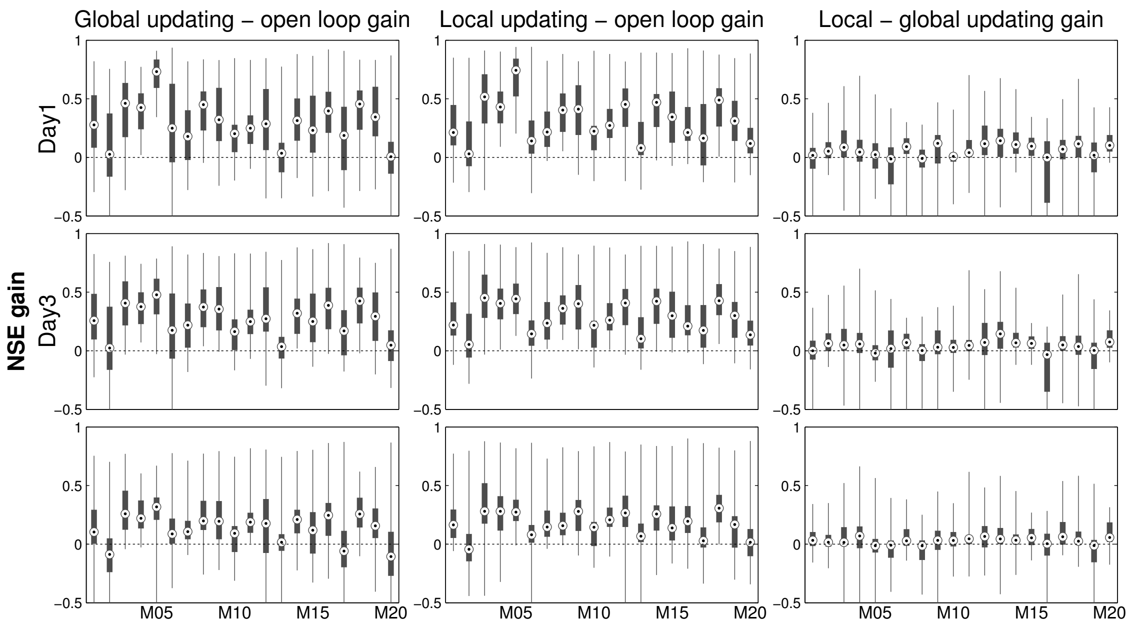

Figure 2.7: EnKF - open loop gains in NSE gains over the 38 catchments for global and locally defined hyper-parameters and state variable for day 1, 3 and 6.

Overall, both EnKF schemes enhance open loop forecasts in the vast majority of cases, from short to longer lead time. However, the gain in accuracy depends heavily on model and to a lesser extent on the global-local updating scheme. One can notice that models 2, 13 and 20 have a structure that react poorly to EnKF updating, especially for global states updating. The increase of computational resources may not be worth the potential gain in performance for the majority of catchments. Yet, these results are improved in the case where catchment specific state variable sets are used.

It is frequent –and normal– that the differences between the two updating schemes global/local for the same model are small. This is the case when a model has frequently the same best set of state variable over the 38 catchments which therefore turns to be the best in average over

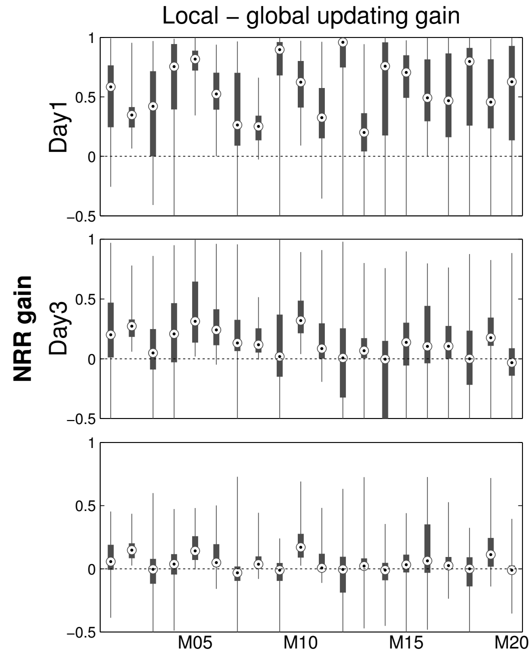

Figure 2.8: EnKF - open loop gains in NRR over the 38 catchments for global and locally defined hyper-parameters and state variable for day 1, 3 and 6.

catchments and explains the frequent small dispersion of the local/global updating gain. However extrema are high as they are obtained when the global updating fails badly on a catchment. The local scheme is logically better than global as it is designed to perform on all catchments but this does not ensure to be better than the global scheme in all cases. Indeed, even if the local updating is catchment specific, it is still averaged on lead times and thus the global state variable set may perform better for a particular lead time.

Figure 2.8 represents the gain between the local and the global updating procedure in reliability (no comparison is possible with the open loop as it is a deterministic forecast). The gain in reliability is consistently high for the first lead time as the second quartile is always positive and third quantile higher than 0.8 for most of the models. Some models, as the model 9, 12, 14 should be preferentially updated in a local way because their gain is substantial (third quartile is greater than 0.95). Interestingly, these models are among the most complex ones in the model pool and seem to require a more detailed setting to exploit optimally the EnKF for the first lead time. As with the NSE, the NRR gain decreases with lead time but stays mostly positive up to day 6 (see also Fig. 2.4). The gain may be negative for the reasons aforementioned with the NSE.

2.4.5 Influence of the catchments

To assess the importance of the catchments on forecasting performance, Figure 2.9 represents the models' NSE over the 38 watersheds. This complementary vision of Figure 2.7 reveals that catchments also have an influence on simulations that is as important as hyper-parameters, model structure, and the state variable selection.

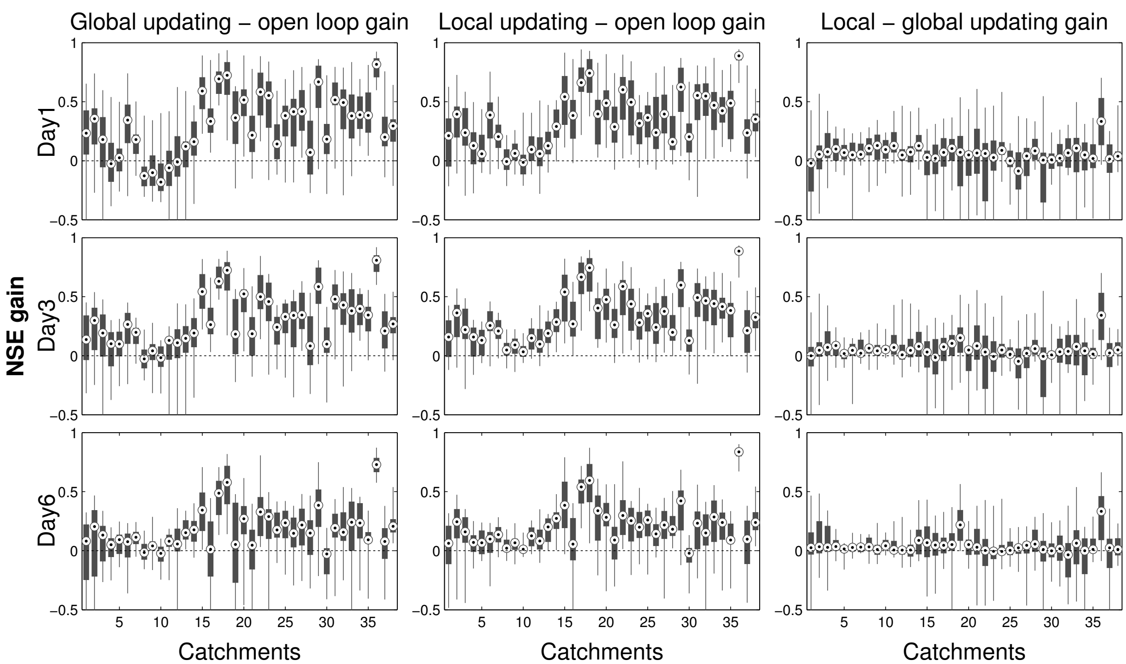

Figure 2.9: EnKF - open loop gains in NSE over the 20 models for global and locally defined hyper-parameters and state variable for day 1, 3 and 6 according to catchments

The majority of the catchment can benefit from EnKF updating, especially in the case where local updating is used. Yet, there is a disparity in the gain as few catchments display a negative median gain, namely catchments 4, 8, 9, 10, 11, and 12 for the first lead time and global updating and catchments 9 and 11 for local updating.

The gain diminishes with increasing lead time except for the catchments that exhibit a negative gain from day one. The underlying reason is that EnKF is not able to update correctly the state variables, attributing erroneous values to the state variable, combined with the fact that the updated state variables have a greater influence for short lead times.

EnKF performance and gain were compared to the available climatic data and catchments characteristics. Specifically, the average annual total and liquid precipitation, the area and the estimated concentration time were put under scrutiny. No clear correlation between these values and EnKF performance has been identified.

2.5 Conclusion and recommendations

This paper discusses the performance and implementation of EnKF in forecasting over a wide variety of catchments and rainfall-runoff lumped models. An extensive testing was carried out to assess EnKF state updating and how it relates to model, catchment, and lead times.

The results show that an optimal implementation of the EnKF is more complex than frequently suggested and that a detailed attention should be paid to the specification of hyper-parameters and updated state variable. While identification of the minimal number of members is relatively straightforward as a vast majority of models and catchment agree, there is no single and universal optimal EnKF implementation for any model. In practice, it is unlikely that the best state combination and hyper-parameter set in average are optimal for all watersheds. Unlike many case studies, it is not reasonable to recommend precise values, as the best EnKF settings are frequently case specific.

The hyper-parameters and more specifically, the perturbations of the inputs are frequently unintuitive to identify as there are often unrealistically high to implicitly account for other sources of uncertainty, especially parameter and structural uncertainty, and to eventually ensure model simulation reliability. An additional challenge arises from the difficulty to optimize reliability and ensemble median bias jointly as the improvement of one criterion is achieved at the expense of the other.

Models encounter important differences in their results and in the way they should be updated. Models with a high number of state variables (high degree of freedom) should receive an increased attention as they are more prone to encounter equifinality issues as many outcomes frequently end up close for a specific condition.

Regardless of the model, ensemble reliability decreases quickly with lead time as the ensemble spread drops from first days while the bias increases. This also underlines that taking into account explicitly initial condition uncertainty solely is not sufficient for medium range forecasting and that structural error and forcing error are dominant in modelling rainfall-runoff processes.

Despite these constrains, the gain that EnKF provides over open loop is substantial, especially if the optimization is carried out locally. The later implies a detailed testing of all combination to identify best performing EnKF implementation but is computationally more expensive. As the EnKF is not efficient with every model and catchment, we recommend to investigate data assimilation coupling with several models to go beyond EnKF - model structure compatibility issue.

Finally, we encourage EnKF users to perform a detailed analysis addressing the question of hyper-parameter and state variable selection of their system to ensure to make the most of EnKF. For further improvement, we also suggest to report explicitly the hyper-parameters and state variables they used to contribute to a better understanding of EnKF parametrization and to identify techniques that would allow to robustly identify the pertinent state variables that should be updated without the need to run all possible combinations.

Chapter 3

Accounting for three sources of uncertainty in ensemble hydrological forecasting[1]

3.1 Abstract

Seeking for more accuracy and reliability, the hydrometeorological community has developed several tools to decipher the different sources of uncertainty in relevant modeling processes. Among them, the Ensemble Kalman Filter, multimodel approaches and meteorological ensemble forecasting proved to have the capability to improve upon deterministic hydrological forecast. This study aims at untangling the sources of uncertainty by studying the combination of these tools and assessing their contribution to the overall forecast quality. Each of these components is able to capture a certain aspect of the total uncertainty and improve the forecast at different stage in the forecasting process by using different means. Their combination outperforms any of the tools used solely. The EnKF is shown to contribute largely to the ensemble accuracy and dispersion, indicating that the initial condition uncertainty is dominant. However, it fails to maintain the required dispersion throughout the entire forecast horizon and needs to be supported by a multimodel approach to take into account structural uncertainty. Moreover, the multimodel approach contributes to improve the general forecasting performance and prevents from falling into the model selection pitfall since models differ strongly in their ability. Finally, the use of probabilistic meteorological forcing was found to contribute mostly to long lead time reliability. Particular attention needs to be paid to the combination of the tools, especially in the Ensemble Kalman Filter tuning to avoid overlapping in error deciphering.

[1] A modified version of this chapter is published in Hydrology and Earth System Sciences Discussion, 12, 7179-7223, 2015, doi: 10.5194/hessd-12-7179-2015. A. Thiboult, F. Anctil, and M.-A. Boucher. Accounting for three sources of uncertainty in ensemble hydrological forecasting. Authorship: A. Thiboult and F. Anctil designed the experimental setup and performed the analysis. Coding and simulations were carried out by A. Thiboult. Meteorological data were provided by M.-A. Boucher. The text has been written by A. Thiboult and revised first by F. Anctil and then by M.-A. Boucher

3.2 Introduction

The complexity of hydrometeorological systems is such that it is not possible to perfectly represent their "true" descriptive physical processes, and even less to integrate them forward in time with mathematical models. These models are only an approximation of varying quality to represent and predict variables of interest, yet they proved to be skillful and useful for water resource management and hazard prevention (e.g. Bartholmes et al., 2009; Pagano et al., 2014; Demargne et al., 2014).

Inadequacies between simulation or predictions and observations can be largely attributed to the many sources of uncertainty that are located along the meteorological chain (e.g. Walker et al., 2003; Beven and Binley, 2014). Hence, it is admitted that improvement of the forecast ought to go through understanding and reducing the sources of uncertainty (e.g. Liu and Gupta, 2007). These sources have different nature that range from epistemic uncertainty due to the imperfection of our knowledge to variability uncertainty where the imperfections are due to the inherent system variability (e.g. Walker et al., 2003; Beven et al., 2008). They also differ in location, i.e. where they lay in the hydrometeorological modeling process: meteorological forcing, model parameter and structure, hydrological initial conditions, and, to a lesser extent, observations (Walker et al., 2003; Vrugt and Robinson, 2007; Ajami et al., 2007; Salamon and Feyen, 2010).

As all models are exposed to these sources of uncertainty, they necessarily lead to forecasts with imperfections. It is thus possible – and frequent – that several models can simulate the process of interest with the same accuracy. These simulations are equally likely in the mathematical sense; it is referred as the principle of equifinality (Beven and Binley, 1992).

Ensembles provide a probabilistic answer to the equifinality problem. They are a collection of deterministic predictions issued by different models to simulate the same event and attempt to produce a representative sample of the future. They can be built by a suitable method wherever a source of uncertainty needs to be put under scrutiny. Additionally, the ensemble means generally have better skills than deterministic systems and offer a better ability to forecast extreme events (e.g. Wetterhall et al., 2013).

As the sources of uncertainty differ in their location, nature and statistical properties, they need specific tools to be deciphered efficiently (Liu and Gupta, 2007). A wide range of methods have been developed in the past year to cater hydrological forecast needs.

At the beginning of the 90s, meteorologists pioneered the operational use of ensembles by constructing Meteorological Ensemble Prediction Systems (MEPS), mostly to take into account imperfect initial conditions that is a prime importance uncertainty source in view of the chaotic nature of the atmospheric physics. Several methods have been proposed to tackle this issue. For instance, to define the initial condition uncertainty, the European Center for Medium-Range Weather Forecasts (ECMWF) generates an ensemble by initiating their model with singular vectors (Molteni et al., 1996) to which a stochastic scheme is added to deal with the model physical parametrization uncertainty (Buizza et al., 1999).

The increasing accessibility of MEPS benefited to the hydrology community to issue probabilistic hydrological forecasts that take into account meteorological uncertainty forcing with Hydrological Ensemble Prediction Systems (HEPS, e.g. Cloke and Pappenberger, 2009; Brochero et al., 2011b; Boucher et al., 2012; Abaza et al., 2014).

A lot of attention has been paid to the identification of hydrological model parameters and the non-uniqueness of the solutions. Among other technique, Vrugt et al. (2003) proposed the Shuffled Complex Evolution Metropolis Algorithm (SCEM-UA), a calibration technique that retains several sets of parameters instead of a single one for a more realistic assessment of parameter uncertainty. Beven and Binley (1992) suggested a more comprehensive approach for model acceptance or rejection with the Generalized Likelihood Uncertainty Estimation (GLUE) that allows to include different forms of competing models.

Gourley and Vieux (2006) assert that dealing only with input and parameter uncertainty is likely to issue unreliable forecast and that hydrological model structural uncertainty should be deciphered explicitly. This statement is substantiated by Clark et al. (2008b) who compares 79 unique model structures and concludes that a single structure is unlikely to perform better than the others in all situations. Poulin et al. (2011) adds that the structural uncertainty is larger than the parameter estimation uncertainty and provides more diverse outputs. Combining dissimilar hydrological model structures proved to possess a great potential (Breuer et al., 2009) even with simple combination patterns (Ajami et al., 2006; Velázquez et al., 2011; Seiller et al., 2012).

Initial condition uncertainty has also aroused scientific interest. Many studies using various data assimilation techniques to incorporate observations within the simulation processes demonstrated that the specification of catchment descriptive states is a crucial aspect of short and medium range forecasts (DeChant and Moradkhani, 2011; Lee et al., 2011). Among them, sequential data assimilation technique such as the Particle Filter (e.g. DeChant and Moradkhani, 2012; Thirel et al., 2013), the Ensemble Kalman Filter (e.g. Weerts and El Serafy, 2006; Rakovec et al., 2012) and variants (Noh et al., 2013; Chen et al., 2013; McMillan et al., 2013; Noh et al., 2014) substantially improve forecast over the open loop scheme, by reducing and characterizing the uncertainty in initial conditions.

Considerable efforts have been made in the development of these sophisticated techniques and this gave rise to many tools that have been individually tested useful. As Bourdin et al. (2012) points out, "To date, applications of ensemble methods in streamflow forecasting have typically focused on only one or two error sources [...] A challenge will be to develop ensemble streamflow forecasts that sample a wider range of predictive uncertainty". As underlined, the forecasting tools frequently tackle different sources of uncertainty and therefore do not exclude each other but can be seen as complementary, combining their assets to compose an overall better system.

The present study identifies three efficient tools, namely a hydrological multimodel approach, Ensemble Kalman Filter, and MEPS forcing that are used together to decipher the traditional hydrometeorological sources of uncertainty. The paper scope is to identify how they are complementary to each other, to assess their individual contribution to the hydrological forecast reliability and accuracy, and to eventually evaluate the possibility of achieving reliability without resorting to post-processing.

This is achieved by issuing a hindcast on 20 watersheds using the aforementioned techniques, either individually or combined, to investigate their specific role in the forecasting process. Each of them produces an ensemble that can be cascaded through the next ensemble technique in order to produce a larger ensemble that possesses a more comprehensive error handling. Finally, if all sources of error are accounted for, the ensemble should generate a forecast that is reliable (Bourdin et al., 2012).

This paper is organized as follow: section 3.3 presents the catchments, models, the Ensemble Kalman Filter basics and scores, section 3.4 sums up the systems specificities and their respective performances followed by a conclusion in section 3.5

3.3 Material and methodology

3.3.1 Catchments and hydrometeorological data

20 watersheds situated in the south of the Province of Québec have been selected for this study (Fig. 3.1). The catchments experience a mixed hydrological regime with a spring freshet resulting from the important winter snow cover and a lesser second peak in autumn.

The climatology of the catchments is varied, with a mean annual snow fall ranging from 2.9 meters to 4.5 meters and total precipitation fluctuates between 877 mm to 1236 mm. The size of the watersheds extends from 512 km^2 to 15342 km^2 and annual mean streamflow from 9

m^3/s to 302 m^3/s.

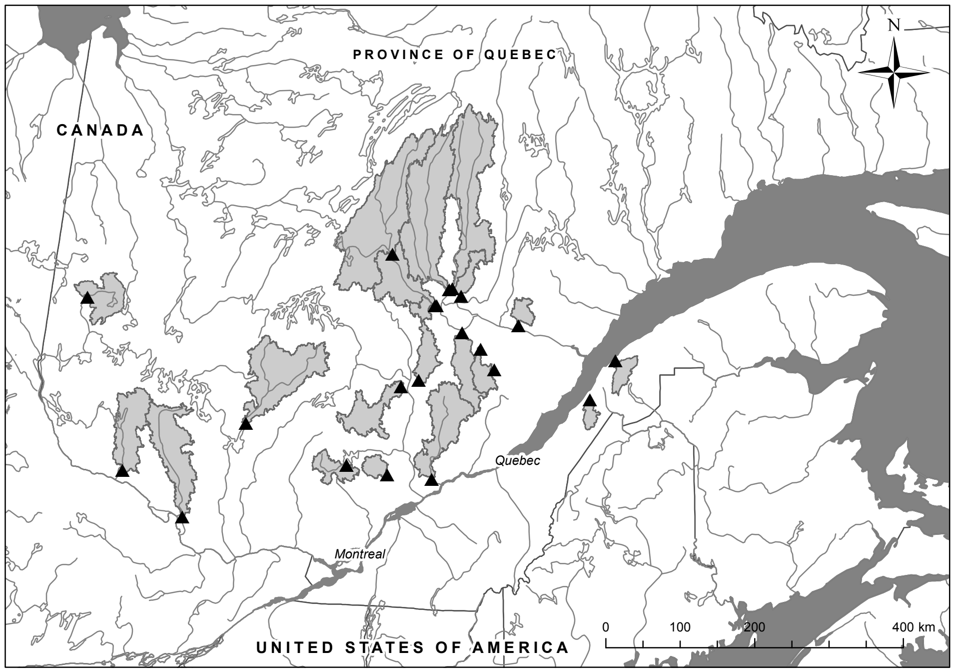

Figure 3.1: Spatial distribution of the watersheds

Daily total precipitation, maximum and minimum temperature and streamflows are provided by the Centre d'Expertise Hydrique du Québec. They performed kriging on the observations over a 0.1° resolution grid to which a temperature correction with an elevation gradient of -0.005°C/m is added. The data base is split into three periods: 1990-2000 for the calibration of the models, October 2005-October 2008 for the spin up, while November 2008-December 2010 is committed to the hydrological forecast assessment.

The MEPS used as inputs to the hydrological model were retrieved from the TIGGE database. The temperatures and precipitation forecasts from the European Center for Medium range Weather Forecasts (ECMWF) were chosen for this study. They are formed by 50 exchangeable members (Fraley et al., 2010) with a 6 hours time-step and a 10 day horizon. However, after conversion from Greenwich time to local Quebec time, the horizon reduces to 9 days. For the sake of the study and to match the common framework of the hydrological models, weather forecast is aggregated at a daily time step. The forecast is provided on a regular grid with a 0.5° resolution (N200 Gaussian grid) that is downscaled to a 0.1° resolution during data retrieval by using bilinear interpolation. As the rainfall-runoff models are lumped, a single representative point forecast is obtained for each MEPS member by averaging the grid points situated within the catchment boundaries. The weather forecast displays acceptable performance over the 20 selected catchments. In fact, in the initial group of 38 catchments, 18 displayed unsatisfactory performances so they were withdrawn from the experiment from the beginning, as pre-processing the meteorological inputs falls outside the scope of the project. When compared to the meteorological observations, rainfall and temperature $MCRPS$ over the 9 days (see sect. 3.3.4) remain below 3 mm and 3°C respectively for selected catchments.

An alternative to the ECMWF ensemble is used to simulate a deterministic meteorological forcing with equivalent theoretical skill. For this purpose, a single member is drawn randomly among the 50 exchangeable members.

3.3.2 Models, snow module and evapotranspiration

The multimodel ensemble, the snow accounting routine and the potential evapotranspiration formulation are similar to the ones presented in the previous Chapters.

3.3.3 Forecasting approaches

Two approaches are used and compared for forecasting, the open loop and the Ensemble Kalman Filter. Regardless of the method used, the meteorological observations over the three years preceding the forecast period are used for model spin up to bring models states to values that estimates the catchment conditions.

Open loop forecasting

When the open loop forecast is activated, the state variables are obtained in simulation mode and used as starting point to initiate the hydrological forecast. The simulation and forecast steps then alternate as follow: 1) the models are forced with observations up to the first day t of the forecast and 2) the models are next forced with the meteorological forecast to issue the hydrological prediction until $t + 9$. The procedure is repeated as the models are brought forward in time with the observations from t.

Ensemble Kalman Filter

The Ensemble Kalman Filter (EnKF) is a sequential data assimilation technique that uses a recursive Bayesian estimation scheme to provide an ensemble of possible model reinitializations. The model state variable vector $\boldsymbol{X}$ is updated according to its likelihood probability density function that is inferred by the observations z, $p(\boldsymbol{X}_t|z_t)$ with the indices t referring to the time.

When an observation becomes available, model states are updated ($\boldsymbol{X}^+$, the a posteriori estimation) as a combination of the predicted ($\boldsymbol{X}^-$, also called the a priori states) and the difference between the prior estimate of the variable of interest $\boldsymbol{H}\boldsymbol{X}^-$ and the corresponding observation z_t.

$$\boldsymbol{X}_t^+ = \boldsymbol{X}_t^- + \boldsymbol{K}_t(z_t - \boldsymbol{H}_t\boldsymbol{X}_t^-) \tag{3.1}$$

where $\boldsymbol{H}$ is the observation model that relates the state vectors and observations, and $\boldsymbol{K}$ is the Kalman gain matrix that defines the relative importance given to the output error respect to the prior state estimate.

The Kalman gain is defined with the model error covariance matrix $\boldsymbol{P}_t$ and the covariance of observation noise $\boldsymbol{R}_t$ as:

$$\boldsymbol{K}_t = \boldsymbol{P}_t\boldsymbol{H}_t^T(\boldsymbol{H}_t\boldsymbol{P}_t\boldsymbol{H}_t^T + \boldsymbol{R}_t)^{-1} \tag{3.2}$$

A detailed explanation of the EnKF mathematical background and concepts can be found in Evensen (2003). In this study, the filter has been implemented following Mandel (2006).

The EnKF is able to decipher catchment initial condition as it acts on variables after the spin up time, i.e. at the very start of the hydrological forecast. Thus, it is frequently presented as a tool that describes catchment descriptive states uncertainty such as soil moisture but it also implicitly takes into account model parameter and structural uncertainty as these are reflected in the model states and outputs errors. The forecast system comprises inaccuracies at several levels and consequently the error statistics that the EnKF uses to update state variables are not only intrinsic variability but also epistemic uncertainty that lay also in the value of the state variables.

The EnKF performance is highly influenced by its setting, in particular by the required noise specification of inputs and outputs (Noh et al., 2014) and also by the choice of the state variable vector (Li et al., 2011). This affects directly the spread of the ensemble and the corresponding uncertainty description. As the level of uncertainty varies from the model used and the simulated watershed, the optimal EnKF implementation also depends to a great extent on these aspects.

In this study, the EnKF is tuned to optimize reliability and accuracy per catchment and per model. The retained specification were identified after the extensive testing that has been carried out in Chapter 2. More precisely, two or three noise levels for each input and output were tested (a 25-50-75% standard deviation of the mean value with a gamma law for precipitation, 10-25-50% standard deviation of the mean value with the normal law for streamflow observations and 2-5° standard deviation with a normal law for the temperature). Additionally, as the choice of updated state variables is also a key component of the EnKF, all possible combinations of the state vector were tested with the 12 noise combinations described above. The retained EnKF setting were based on a two-step criterion; firstly the 3 settings that presented the best reliability are kept and then the one among them that led to the lowest bias. Therefore, the optimal setting may use unrealistically high perturbations that compensate partially for the structural error.

3.3.4 Scores

The continuous ranked probability score ($CRPS$, Matheson and Winkler, 1976) is a common verification tool for probabilistic forecasts that assesses accuracy and resolution. A cumulative distribution function is built based on the raw predictive ensemble, i.e. the collection of deterministic forecasts and then compared to the observation.

$$CRPS(F_t, x_{obs}) = \int_{-\infty}^{+\infty} (F_t(x) - H(x \geq x_{obs}))^2 \, dx \tag{3.3}$$

where $F_t(x)$ is the cumulative distribution function at time t, x the predicted variable, and x_{obs} is the corresponding observed value. The function H is the Heaviside function which equals O for predicted values smaller than the observed value, 1 otherwise. The $CRPS$ shares the same unit as the predicted variable x.

As the $CRPS$ assesses the forecast for a single time step, the $MCRPS$ is defined as the average $CRPS$ over the entire period. The $MCRPS$ can reduce to the Mean Absolute Error (MAE) if a single member is considered and thus it allows to compare deterministic and probabilistic forecasts (Hersbach, 2000; Gneiting and Raftery, 2007). Finally, a 0 value indicates a perfect forecast and there is no upper bound.

The reliability diagram (Stanski et al., 1989) is a graphical method to assess the reliability of a predictive ensemble by plotting forecasted against observed event frequencies. A perfectly reliable forecast is represented by a 45° line that indicates that forecasted and observed frequencies are equal. If the joint distribution curve differs from the perfect reliability lines, it indicates that the spread of the ensemble does not perfectly match its predictive skills. If the curve is situated above the perfect reliability line, this denotes an overdispersion of the ensemble, and an underdispersion in the opposite case.

The reliability is twofold. Since the reliability curve assesses the dispersion regarding the predictive skills of the ensemble, it is possible to have a perfectly reliable system with a low predictive capability in the case the dispersion is very high. For disambiguation, the ensemble spread is added to the plots.

Practically, one can define the deviation from perfect reliability by estimating a measure of distance between the forecast reliability curve and the perfect reliability line by computing the Mean Absolute Error (MAE) or Mean Square Error (MSE, Brochero et al., 2013). This dimensionless score allows to reduce the measure of reliability to a scalar. In the case where the MAE is used, it can be easily interpreted as the average distance between forecasted frequencies and the observed frequencies over all quantiles of interest. This verification score is henceforth referred as Mean absolute error of the Reliability Diagram, $MaeRD$.

Additional information about reliability can be obtained from the Spread Skill Plot (SSP, Fortin et al., 2014). It compares the Root Mean Square Error $RMSE$ and the square root of average ensemble variance that is a measure of the ensemble spread. The reliability is thus somehow decomposed into an accuracy error part and a spread component. Ideally, the spread should match the $RMSE$.

3.4 Results

Table 3.1 summarizes the specificities of the nine variants of the hydrometeorological forecast framework according to the three "forecasting tools": multimodel, EnKF, and ensemble meteorological forcing. Each of these switch may be activated or not and are marked as on/off in the table.

The multimodel switch dictates if the members issued by the 20 individual models are pooled together to create a single probabilistic forecast. In the case where the multimodel approach is not used, the models outputs are kept individually and 20 distinct ensembles – one per model – are considered.

Table 3.1: Description of the nine systems

Systems	A	B	C	D	E	F	G	H	H'
Multimodel	Off	Off	Off	Off	On	On	On	On	On
EnKF	Off	Off	On	On	Off	Off	On	On	On
Met. ensemble	Off	On	Off	On	Off	On	Off	On	On
Nb of members	(20x)1	(20x)50	(20x)50	(20x)2500	20	1000	1000	50000	50000

The EnKF switch indicates if sequential data assimilation or the open loop procedure is applied. When EnKF updating is used, an ensemble of 50 members is created from 50 likely initial condition sets identified by the filter. Otherwise, a single set of state variable values determined from the simulation is provided to the forecasting step. Note that the H and H′ system differ by the EnKF perturbations magnitude, where H uses perturbations that aim at optimizing the combined criterion while H′ uses lower perturbations that are deemed to be more realistic.

Lastly, the meteorological forcing employed during the forecast step can be either deterministic or probabilistic, using one randomly picked member or all 50 MEPS members.

These tools can be used alternatively or combined. For instance, if the EnKF and the meteorological ensemble forcing are used collectively, each of the 50 initial conditions sets will serve as starting point for each of the 50 meteorological forecast member creating a larger hydrometeorological ensemble that contains 2500 members.

We chose to disregard more complex or "hybrid" cases in this study, where for example, the final ensemble is composed with some models that benefit EnKF state updating while others are used in an open loop forecasting mode as these setups do not add additional information about the role of the tools, increase the degree of freedom for the system optimization and would shoot up computational costs.

The results for each of the nine systems applied to every catchment, lead time and possibly every model are not systematically detailed and compared to each other. The following graphs are deemed sufficient to interpret the role and benefits that the system components play on the forecast quality. Additional graphs representing the resolution and reliability of each system are provided in Annex E for readers who are interested in a specific set up.

To picture an overview of the results, Figure 3.2 represents the accuracy in terms of $MCRPS$ (or MAE for system A that is fully deterministic) and $MaeRD$. For graphical convenience, the full distribution of performance according to various factors is not displayed but only a single representative value. To reduce the whole of the results to a single scalar, the median performance has been considered. In the case where a multimodel approach is used, the median performance over the 20 catchments is displayed on the figure. Otherwise, when individual models are considered, firstly the median performing model is identified and then the median performance over the catchment is represented. This implies that the performance of individual models systems (A, B, C, and D) may refer to a different model for each lead time.

The four radar plots situated on the top of the figure illustrate the $MCRPS$ performance. As a reference, the center of the disk consist of the median $MCRPS$ value of the climatology over

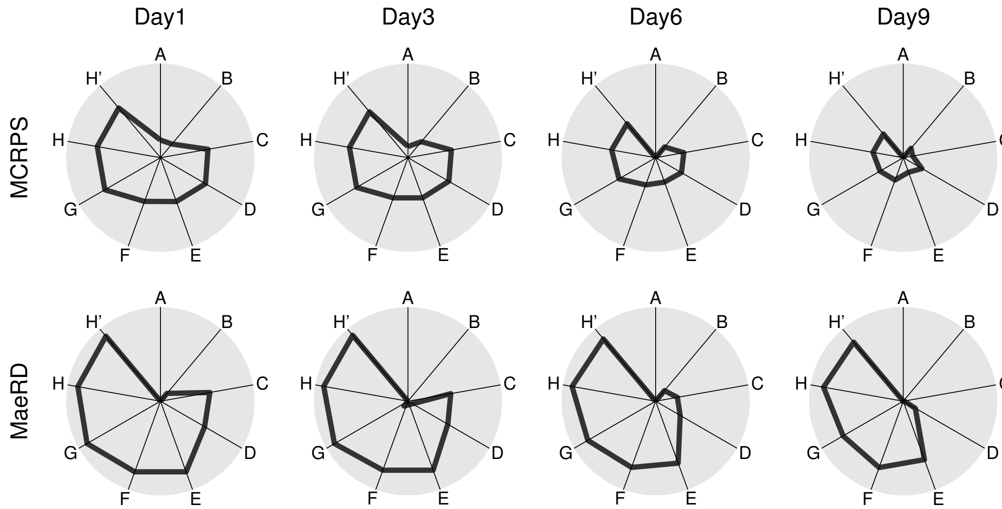

Figure 3.2: Synthetic results of the 9 systems that are referred by their code letter (see Table 3.1). The 4 top radar plots illustrate the *MCRPS* with the center indicating the climatology reference performance, and the perimeter representing a perfectly accurate simulation. The 4 bottom plots describe the measure of distance from perfect reliability, with the center indicating a *MaeRD*=0.5 while the perimeter corresponds to a perfect reliability. For detailed numerical values, see Table E.1 and Table E.2 in annex

the 20 catchments while the perimeter represent a perfect *MCRPS* equals to 0. The radius lines represent the nine systems described in Table 3.1 and are referred by their corresponding letter.

The nine systems present varying performance but all decrease logically with lead time. System A, which is deterministic, undoubtedly performs worse for every lead time. It is challenged from the 3rd day and is outperformed for medium range forecast by the hydrological climatology. System B presents a quite similar behavior to A but with a lower decrease of accuracy with lead time. System C may be considered as competitive for shorter lead times but looses quickly its edge. These preliminary results tend to indicate that simpler HEPS may not be appropriate to accurately forecast streamflows over a nine day horizon. However, all versions including the simpler version except system A are more informative than the climatology for all lead times. Systems G, H and H' stand out from the others for all lead times.

The second row in Figure 3.2 illustrates the reliability of each system. The center of the disk corresponds to a *MaeRD* equals to 0.5. System A is artificially placed at the center of the radar plot to denote that no reliability information is communicated since it is deterministic.

The reliability results shares similarities with the accuracy assessment. Simpler systems face difficulties to provide a reliable forecast. Despite the use of meteorological ensemble forcing, system B is far from providing the right dispersion. Systems C and D provide some information for short lead times but experience a substantial loss with increasing lead time. Once again, G, H and H' are performing best.

3.4.1 Multimodel approach and structural uncertainty

To assess the gain related to the multimodel approach, Figure 3.3 presents a comparison of the individual model *MAE* (A) and the *MCRPS* that pools all model output together (E). At this step, only the structural uncertainty is taken into account as the meteorological forcing is kept deterministic and no initial condition uncertainty estimation is provided for both cases. These systems are computationally cheap as they contain either 20x1 member or 20 members.

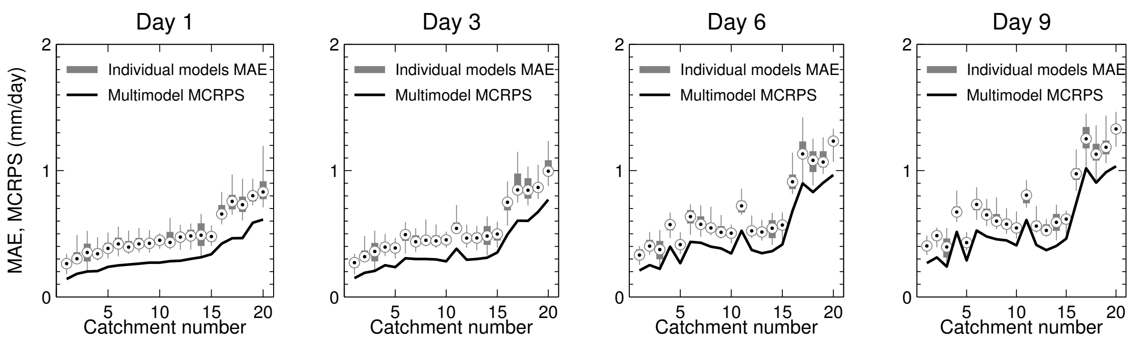

Figure 3.3: Comparison of individual models daily discharges *MAE* and multimodel *MCRPS* sorted by increasing multimodel *MCRPS* for the first day (version A vs E)

In Figure 3.3, each boxplot represents the distribution of performance (minimum, quantiles 0.25, 0.5, and 0.75, and maximum) of the 20 models while the curve details the multimodel accuracy. On the x axis, the 20 test catchments are sorted according to increasing multimodel *MCRPS* for the first lead time. This allows to notice that certain catchments exhibit a faster growing error.

The multimodel performs consistently better than the median performance of the model but also better than any model in the large majority of cases. Exceptions can be occasionally observed for catchment 3 and 17 where only one or two models outperform the ensemble. However, the best performing models differ from a catchment to another while the multimodel presents the advantage of being more robust than any of the models. This is explained by the varied individual model behaviors. Each model may grasp different specificities of the hydrograph by focusing more specifically on different (conceptual) hydrological processes. Consequently, the ensemble members – the models – have disparate errors. Whenever the mismatch between forecast members and observation is poorly correlated, their errors tend to cancel out each other.

Figure 3.4 presents the reliability of the system E. Each curve refers to one of the 20 catchments. As mentioned, the structural uncertainty of the hydrological models is solely explicitly taken into account by the combination of the models.

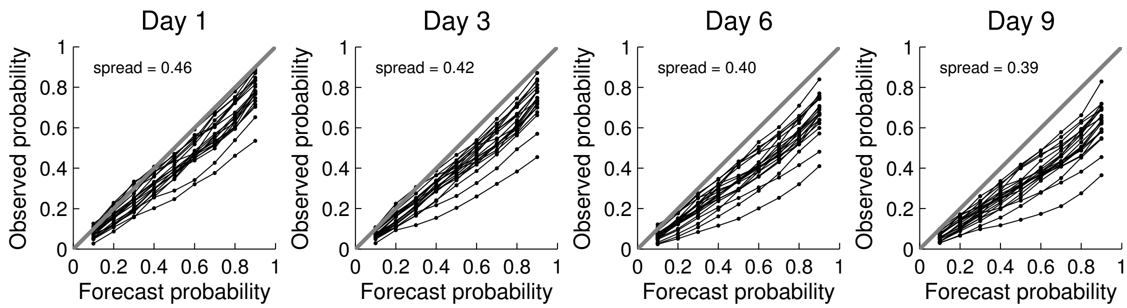

Figure 3.4: Reliability of the multimodel ensemble (system E) for all individual catchments. The spread represents the square root of mean ensemble variance averaged over all catchments.

System E is generally slightly over confident for all lead times and this trend becomes more apparent as the lead time increases. This is expected as the meteorological forcing uncertainty increases with time while the deterministic forcing does not support that aspect. One can notice that the reliability also depends on the catchments. For the first lead time, most of the catchments are close to reliability while there are two outliers for which accuracy skills do not match their corresponding spread. In fact, these catchments exhibits a constant hydrological wet bias – partially explained by a meteorological forecast wet bias that over-forecasts precipitations by almost 15% – that is not captured by any of the models even if the global tendency is respected. Consequently, the models errors are highly correlated and this prevents the members to form a performing ensemble. This bias indicates that the aggregation of the other sources of uncertainty drive the system toward an inaccurate state.

3.4.2 Data assimilation and initial condition uncertainty

Figure 3.5 illustrates the increase of performance related to the data assimilation by comparing systems E and G. System G improves upon E as it benefits from the EnKF data assimilation to handle the initial condition uncertainty. The models states are updated according to the last available observations and an ensemble is created for each model based on the probabilistic estimation of best initial conditions.

The EnKF provides considerable gain over open loop forecasts for all watersheds and reduces the number of lower performance watersheds. Data assimilation is in our case, par-

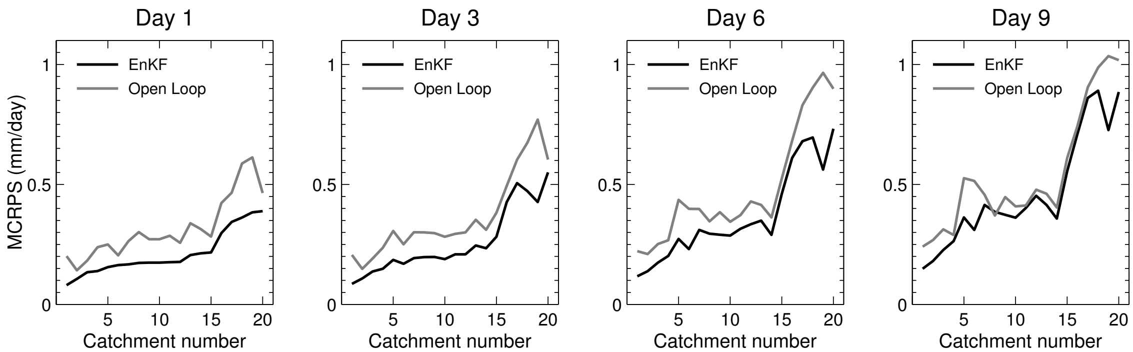

Figure 3.5: Comparison of open loop and EnKF multimodel *MCRPS* sorted by increasing EnKF *MCRPS* (system E vs G)

ticularly effective on catchments that present a systematic bias. This indicates that inaccuracies accumulated and stored during the spin up period in the state variable as the results of structural and forcing errors can be significantly reduced by providing adequate model reinitialization.

As the EnKF acts on model state variables right after the spin up period, it is not surprising to see its efficiency decreasing with lead time. This clarifies why the EnKF is beneficial for all lead times but that its skill decreases faster than the open loop scheme one. Moreover, the EnKF provides satisfactory initial condition distribution to minimize the error at the time the observation becomes available but does not sample the posterior states to be optimally integrated through time.

Figure 3.6 details the reliability of system G. There is a considerable increase of spread in comparison to system E for shorter lead time that goes beyond adequate dispersion and lead to a slightly overdispersed forecast for the first lead time. This was expected as the EnKF was initially implemented to maximize individual model reliability for system G (see Chapter 2). As the EnKF also takes into account the parameter and structural uncertainties and is combined with a multimodel approach, there may be a redundancy in the error deciphering. The structural error and the corresponding ensemble spread that it should describe may be somewhat accounted twice in that particular case. However, the overestimation of the ideal spread diminishes as the EnKF influence fades away quickly and the system goes back toward a better reliability for medium range forecast and underdispersion from days 4-5.

To explain the rapid decrease of reliability, Figure 3.7 displays the ensemble mean *RMSE* and the square root of average ensemble variance. This individual spread skill plot (one model and one catchment) is typical. The spread and the *RMSE* are close to a perfect match

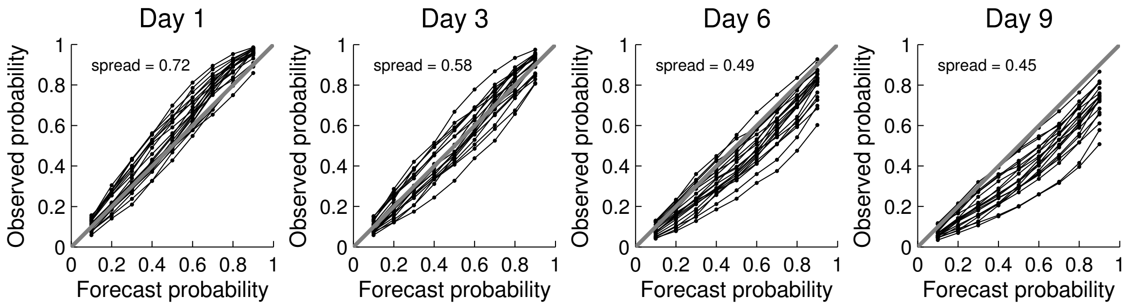

Figure 3.6: Reliability of the EnKF multimodel ensemble (system G) for all individual catchments. The spread represents the square root of mean ensemble variance averaged over all catchments.

for the first day indicating an appropriate dispersion, yet, they diverge rapidly. The reliability deterioration of the system is twofold: the increase of the ensemble mean bias and the decrease of the spread. The loss of hydrological predictive skill is coherent regarding that the meteorological accuracy diminishes with increasing lead time. Concerning the second point, in most cases, the ensemble of initial conditions that EnKF provides often differ little from each other – few percent – indicating that the posterior distribution of each parameter is rather narrow (DeChant and Moradkhani, 2012; Abaza et al., 2015). These dissimilarities are not large enough to provoke a divergence in the behavior of EnKF members during the forecasting step as the models are resilient. The different initial conditions thus tend to merge toward a certain value – often close the open loop one – which may not be accurate.

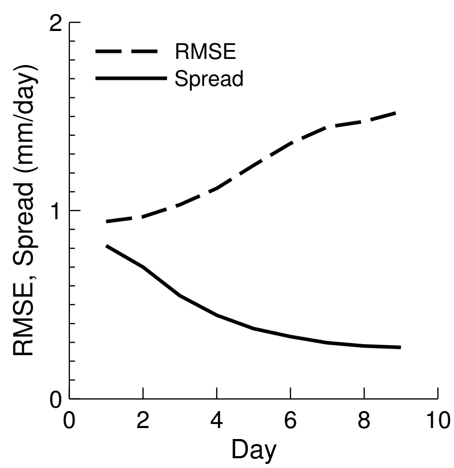

Figure 3.7: Typical Spread Skill plot of a single model EnKF ensemble

3.4.3 Contribution of the meteorological ensemble forcing

One step further in the system complexity is taken as the MEPS forcing is introduced. Figure 3.8 compares the *MCRPS* of systems G and H. They differ only in their meteorological forcing as the latter uses the 50 member probabilistic forecast. Difference between them is negligible until the 7th or 8th day where a slight improvement in performance can be noticed on most catchments. For these longer lead times, the probabilistic forcing is somewhat more efficient for the *MCRPS* but the main difference lies in the reliability (Figure 3.9). In fact, the reliability is substantially improved for the longest lead times when the meteorological uncertainty is provided to the system.

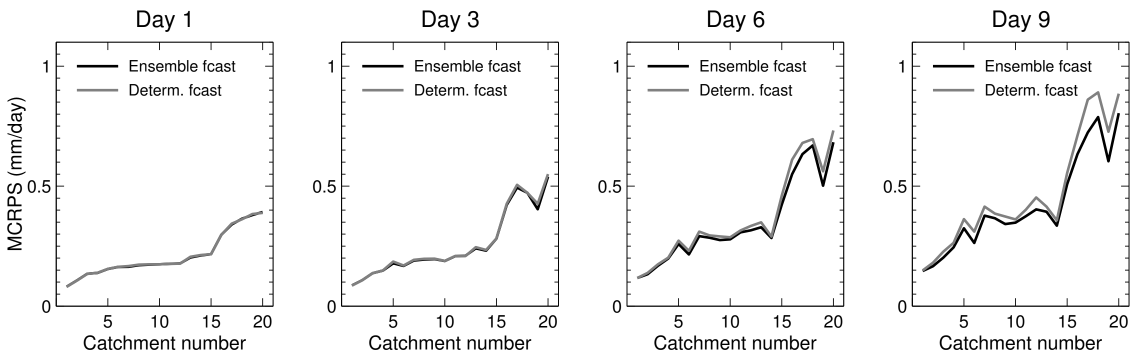

Figure 3.8: Comparison of EnKF multimodel *MCRPS* with deterministic and ensemble meteorological forcing (system G vs H)

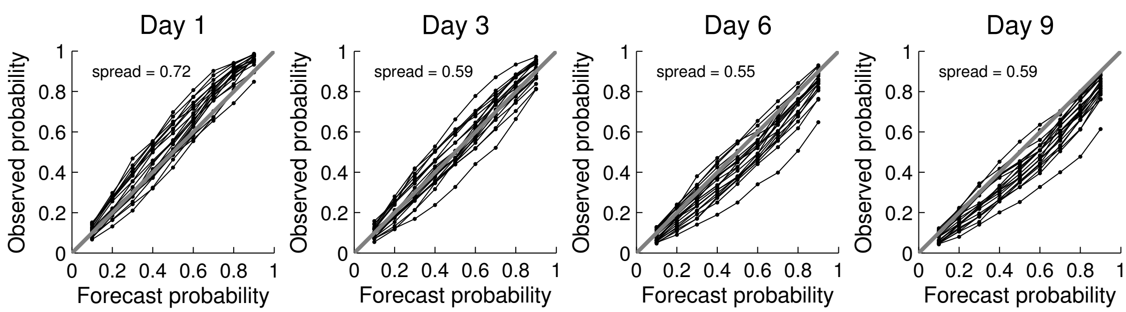

Figure 3.9: Reliability of the EnKF multimodel ensemble with MEPS forcing (system H)

The ECMWF MEPS dispersion grows with lead time and logically contributes to the HEPS spread accordingly. This is confirmed by comparing the spread of the G and H systems as they decrease at a different pace. While they are almost identical with a value of 0.58 mm/day and 0.59 mm/day respectively for the day 3, G spread drops to 0.45 mm/day for day 9 while the use of the MEPS maintains the spread to 0.59 mm/day. The evolution of the spread also indicates that the tool that contributes the most to the HEPS dispersion is the EnKF since the raw MEPS forcing is not able to balance the decrease of the spread induced by the EnKF.

The main sources of uncertainty – structure, initial conditions, and meteorological forcing – are cascaded through the different components of the forecasting system to provide better forecast than any of the systems previously described. Yet the system reliability is not perfect as the forecast for day 1 and day 9 are slightly overdispersive and underdispersive in addition to present sensitivity to the watersheds. To realistically represent the uncertainty of the system, the spread should grow with lead time as the future is more uncertain. This suggests that further improvement of this setup and particular application could be obtained with a more dispersed meteorological forcing.

3.4.4 Simplification of the framework

A potential drawback for operational use of such system is that it is computationally expensive as 50000 members are exploited to build it. The efficiency of a simpler system is assessed on Figure 3.10. Eight typical catchments are displayed in the sub plots to illustrate the conclusion. The box plots represent the *MCRPS* distribution of the 20 models results from system D that benefits EnKF state updating and MEPS forcing. Each of these models can be considered as a sub-ensemble of the large ensemble H driven by a single model instead of using a multimodel approach. This is a more consistent approach with the EnKF individual optimization that is carried out to aim for reliability for each model one at a time. The numbers at the top of the sub-plots refer to the model number that are better than the multimodel for each lead time.

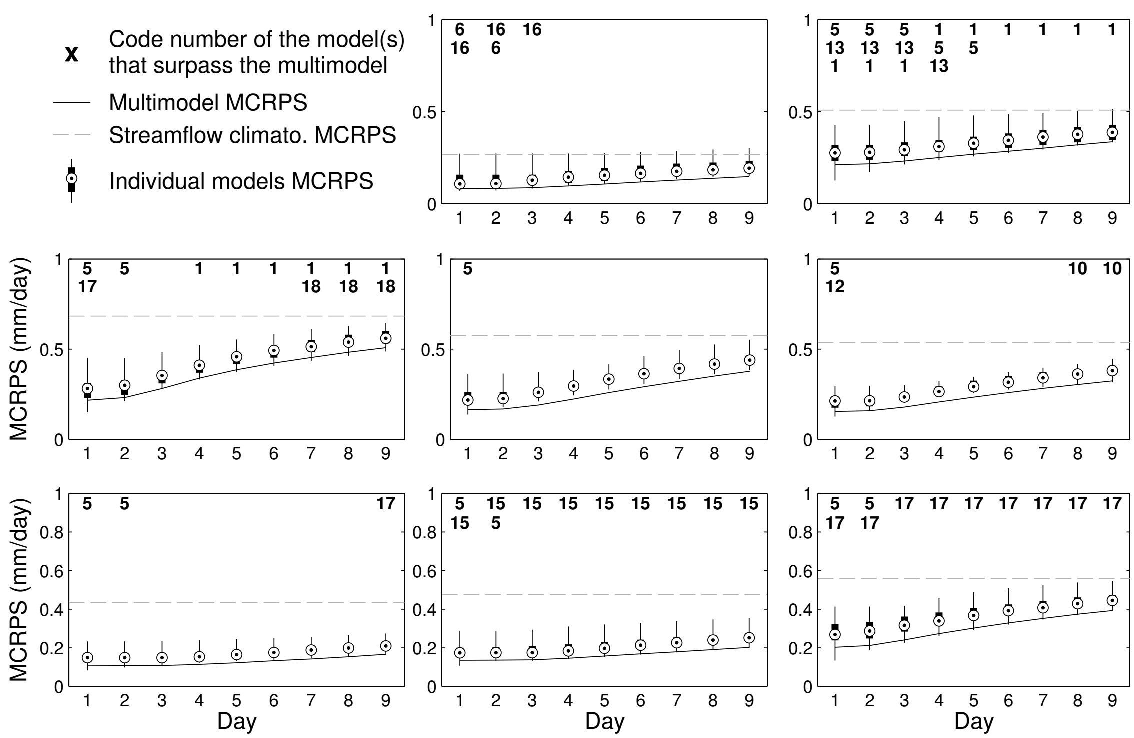

Figure 3.10: Comparative examples of the *MCRPS* on 8 watersheds of the EnKF individual models and the EnKF multimodel, both using MEPS forcing (system D vs H)

In Figure 3.10, sub-ensembles are more skillful than the hydrological climatology for all lead times but rarely outperform the multimodel forecast. More precisely, the median performing sub-ensemble is always poorer than the multimodel and only the best models among the 20 occasionally exhibit lower $MCRPS$. Individual models that outperform the multimodel frequently differ from a catchment to another and from a lead time to another. This emphasizes the difficulty to choose a priori a single model as half of the 20 models never behave better than the multimodel and only model 1, 5, and 17 perform better than the multimodel for several catchments. Choosing a sub-ensemble doubtlessly enhances the system computational requirements and eases operational implementation but relying on a single model may be misleading or, at least, minimize the expectation that one can have from the HEPS.

Figure 3.11 assesses the reliability of the same system with the $MaeRD$ score. Like for the previous plots, the box plots contain the 20 ensembles that correspond to the 20 models and are sorted by catchment with increasing multimodel $MaeRD$. Note that the $MaeRD$ does not provide precise information about dispersion but only about the distance from perfect reliability. Nevertheless, individual model ensemble may be either slightly over or underdispersive for the first lead time but are systematically underdispersive for longer lead times. On the other hand, system H can be either over or underdispersive depending on the watershed. Overdispersive forecasts, like for the catchment 20, can be recognized as they tend to become more reliable for longer lead time.

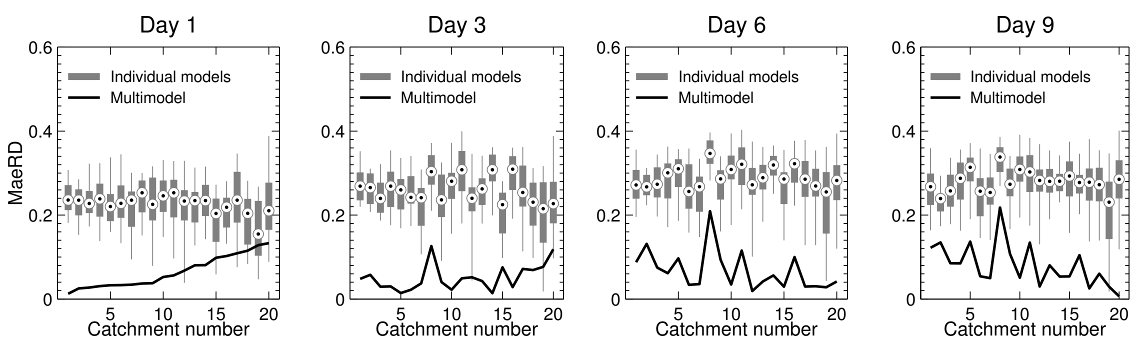

Figure 3.11: Comparison of the deviation from perfect reliability of EnKF individual models and the EnKF multimodel, both using MEPS forcing sorted by increasing EnKF multimodel *MaeRD* for the first day (system D vs H)

For the first lead time, the best individual model ensembles may be competitive with the multimodel but are already less efficient from day 3 and are drastically underdispersive for day 9. Even if the EnKF takes into account the structural uncertainty at $t = 0$, it loses its efficiency during the forecast. The information that the updated state sets contain about the structural uncertainty vanishes when the sets converge toward a common value. The multimodel approach, by its nature, allows to take over the role of the EnKF by dynamically preserving the required diversity.

3.4.5 Required EnKF perturbations

H' is identical to system H except that it relies on a different optimization of the EnKF. Instead of maximizing the combined criterion for individual models (see section 2), the EnKF noise specification is set lower to values that are more consistent with real uncertainties estimations of observed climatological and streamflow observations at catchment scale. Namely, precipitation is perturbed with a gamma law with a standard deviation of 25% of the mean value, temperatures with a normal law with a 2° standard deviation and streamflow observations with normal law with a 10% standard deviation.

This would correspond to a potential optimal EnKF implementation if the total uncertainty could be summarized to the input and output error and were perfectly identified, i.e. in a perfectly controlled environment with a negligible model structural error. Consequently, the structural error is theoretically only deciphered through the multimodel pooling. Yet this needs to be qualified as it is practically hard to untangle the source of uncertainty within the actual configuration of the EnKF but it reduces the risk that the tools effects overlap. By choosing these perturbations, the user also gets rid of a fastidious EnKF tuning by screening adequate perturbation (Chapter 2 and for e.g. Moradkhani et al., 2005) and hence simplifies the system implementation.

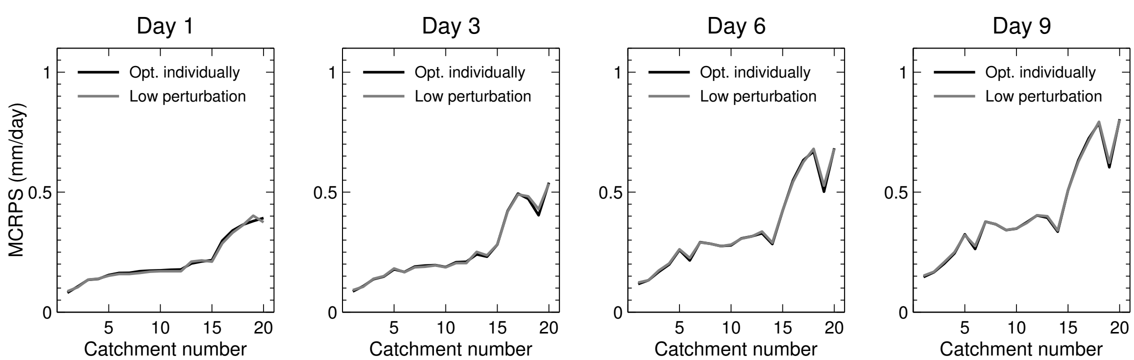

Figure 3.12: Comparison of EnKF multimodel MEPS systems using either individually optimized EnKF perturbations or lower input-output perturbations (system H vs H')

In Figure 3.13, system H' improves reliability for first lead times by reducing the overdispersion with a sensible decrease in the ensemble spread from 0.72 mm/day to 0.57 mm/day for day 1 without any degradation of the *MCRPS*. System H' maintains a more constant spread and reliability with increasing lead time as the main sources of uncertainty are more accurately deciphered specifically by their corresponding tool, leading to an overall better forecast.

Finally, there is still a difference in reliability between catchments indicating that it is unrea-

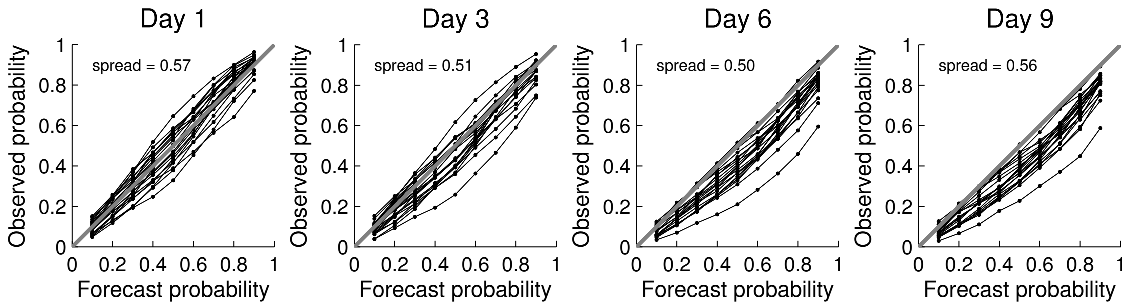

Figure 3.13: Reliability of the EnKF multimodel ensemble with MEPS forcing and lower input-output perturbations (system H')

sonable to assume that uncertainties are invariant from one catchment to another. Part of the misfit probably originates from the structures composing the multimodel ensemble that can be maladapted to simulate some catchments, from doubtful streamflow measurements, or from meteorological ensemble forcing that seems slightly underdispersive for this particular application.

3.5 Conclusion

This work investigates the contribution of three different probabilistic tools commonly used in hydrometeorological sciences. They are used conjointly and alternatively to identify their effect on the hydrological predictive ensemble and to untangle sources of uncertainty that are aggregated in the outputs.

Each of these tools is dedicated to capture a certain aspect of the total uncertainty. A multimodel approach is used to quantify and reduce explicitly the hydrological model error, the Ensemble Kalman Filter to decipher the uncertainty related to initial conditions and the meteorological ensemble to account for the forcing uncertainty.

The experiment shows that important gain may be achieved in terms of accuracy and reliability by adequately using these techniques. Their action differs substantially by their mean and range of action.

The EnKF provides accurate quantification of initial error but fails to maintain reliability as its effect fades out quickly after model spin up. The information about the structural uncertainty deciphered by the EnKF, which is contained in the state variable posterior distribution, is not propagated with time integration during the forecast step. However, the EnKF remains a key component of the system as it is the one that provides the most dispersion. This also indicates that the accumulation of past errors in the initial conditions is a dominant source of uncertainty.

The multimodel approach is able to partially compensate for the EnKF decreasing action by taking over the structural uncertainty. Moreover, the combination of independent models improves accuracy as their errors may cancel each other. Lastly, the use of ensemble meteorological forecast contributes to the reliability of medium range forecast by representing the meteorological forcing errors.

Their actions are complementary as they decipher different nature of uncertainty at different locations by acting at particular stages in the forecasting process. When combined, they need to be set according to the tools they are juxtaposed with to prevent overlapping actions. This is particularly the case for the EnKF that has important degree of freedom in its implementation. It can eventually be tuned with more realistic input perturbations by coupling with the multimodel ensemble and therefore, facilitate its implementation by relaxing the constraints of optimal perturbation screening.

Possible avenues for further improvements may be achieved through a multimodel state updating rather than individual model updating, i.e. by treating initial condition in a single step as a whole. Lastly, the meteorological forecast shown to be a little underdispersed for this application and could be possibly improved by applying suitable pre-processing techniques.

Chapter 4

Forecast quality and value in decision-making[1]

4.1 Abstract

In an operational context, efficient decision making is usually the ultimate objective of hydrometeorological forecasts. Because of the uncertainties that lay within the forecasting process, decisions are also subject to uncertainty. A comparison of five Early Warning Systems (EWS) based on different forecasting systems is performed to investigate how uncertainties affect the decision quality. These systems differ by the location of the sources of uncertainty and the total amount of uncertainty that they are expected to decipher. They are assessed with the Relative Economic Value, which is a flexible measure that assesses the economic benefits of the EWS. All systems provide a gain over the no-forecasting case for every horizon but most complex systems, which decipher more uncertainties, are found to reduce the most the expected damages. Systems with better accuracy and reliability are generally the most valuable even if this relation is loosely defined.

4.2 Introduction

Floods are recognized as one of the most devastating natural disasters. Related hazards and risks are considerable in numerous places and require adequate prevention measures. Governments and communities seek to reduce risk – the product of the hazard and its frequency of occurrence – uppermost in urban and industrial zones and to protect environmental and agriculture areas.

[1]This chapter has not been submitted to a journal yet. Authorship: A. Thiboult and F. Anctil designed the experimental setup and performed the analysis. Coding and simulations were carried out by A. Thiboult. The text has been written by A. Thiboult and revised by F. Anctil

Traditionally, risk prevention is preferred over relief for economical and human considerations (Rogers and Tsirkunov, 2010). Three types of answers for reducing flooding risk are advocated: the construction of dedicated infrastructures, the use of early warning systems (EWS) and the respect of environmental buffers. To date, the combination of these three approaches has shown to be the most efficient practical answer to minimize economic damages and loss of lives. In the case where all these answers are not available, an EWS is estimated to serve the best the neediest people (Rogers and Tsirkunov, 2010). Streamflow management is not limited to risk management since lower streamflows may also be of interest for agriculture, construction, energy, telecommunication, tourism, transport, logistics and water availability (Frei, 2010).

Merz et al. (2010) and Priest et al. (2011) noticed a progressive shift from structural measures (e.g. building dikes and retention basins) that aim at decreasing the occurrence and intensity of flood discharge toward non-structural measures that include a wide panel of responses such as the adoption of policies and laws, public awareness raising and education, for instance. This stresses the actual move from hazard prevention to risk prevention and follows the recommendation of Bavarian State Ministry for the Environment (2006) and Ministry Of Transport (2006) to enhance co-existence with rivers.

A non-structural measure, such as an evacuation alert, relies on the available mitigation time, i.e. the amount of time available between the beginning of the mitigation response and the moment the flooding occurs. A flood alert solely based on upstream observation will provide a mitigation time that is limited to the flood wave travel time, but an EWS based on meteorological and rainfall-runoff forecast, will have the capacity to greatly extend mitigation time. The longer the mitigation time is, the more the population has time to evacuate the flooded area or the authorities to take preventing actions. With increasing lead time, the amount of property and infrastructures that can be protected increases, but there is at the same time an increment in the number of costly false alarms (Rogers and Tsirkunov, 2010; Priest et al., 2011).

The maximal potential flood warning effectiveness provided by the forecast needs to be qualified, as many social considerations have to be accounted for (Molinari and Handmer, 2011). The steps that include the understanding of the alert, the action that may be taken and their effectiveness makes that there is a difference, possibly large, between maximal and effective damage reduction.

Nonetheless, numerous studies demonstrated that flood alerts are economically efficient (e.g. Priest et al., 2011; Molinari and Handmer, 2011; Verkade and Werner, 2011; Perrels et al., 2013). Frei (2010) estimated that benefits generated by weather services in Switzerland amount to some hundreds of millions of US$ per year. Similar results were obtained by Anaman and Lellyett (1996); Lazo and Chestnut (2002) and Leviäkangas et al. (2007) in other industrialized countries, i.e. ratios of invested and saved money oscillating between 1:4 and 1:6. Pappenberger et al. (2015) determined that the European Flood Awareness System (EFAS, Thielen et al., 2009; Bartholmes et al., 2009) which provides information to national authorities and to the Emergency Response Coordination Center of the European Commission up to 10 days ahead, reaps benefits as high as 400 Euros for every 1 Euro invested.

A complete warning system is typically composed of four facets: the risk knowledge, the monitoring-forecasting and warning, the dissemination and communication and the response capability (Rogers and Tsirkunov, 2010). The failure of one of these components will be cascaded through others and will consequently decrease or suppress the flood mitigation effectiveness. Among them, the forecasting step, always imperfect, is critical and complex. It is subject to many uncertainties because of the inaccuracies that lay in the mathematical representation of hydrometeorological systems, mainly in the system state and its dynamic behavior (e.g. Ajami et al., 2007; Salamon and Feyen, 2010; Liu and Gupta, 2007; Liu et al., 2012).

Ensemble prediction systems (EPS) provide a probabilistic answer that is able to incorporate different modeling sources of uncertainty. The approach is gaining in popularity and starts being used by operational agencies (see the review by Cloke and Pappenberger, 2009). In a decision making context, the value of an EPS proved to be efficient and is capable to improve upon traditional deterministic forecast (e.g. Richardson, 2000; Zhu et al., 2002; Verkade and Werner, 2011; Boucher et al., 2012; Stephens and Cloke, 2014), even if the communication of probabilistic forecasts remains a challenge (Ramos et al., 2010; Demeritt et al., 2013).

Despite the different investigations concerning EWS and decision making systems, no economic study provides a comprehensive comparison between different systems that goes beyond the probabilistic and deterministic forecasts confrontation (Verkade and Werner, 2011; Boucher et al., 2012). Moreover, the question of forecast economic value is often tackled alone and the relation between quality and value is rarely addressed. As Verkade and Werner (2011) point out, it is expected that the value will increase with increasing sharpness but effort should be dedicated to clarify what are the mandatory qualities that the forecast needs to improve its economic value.

This study adopts a simple framework to evaluate the economic gain that could be reached for the different forecasting systems presented in Chapter 3. These systems differ by the way they decipher the different sources of uncertainty and the amount of total uncertainty they handle. As a result, they vary in terms of performance and reliability. The framework allows to compare their economic value on the same database. The second aim of the study is to attempt to relate the hydrometeorological forecasting system quality and economic value.

Section 4.3 presents the methodology, including the hydrometeorological data, the description of the relative economic value framework, the decision rules and the scores. Results are presented Section 4.4 where the relative economic value and the relation between forecast quality and value is exposed. Finally, concluding statements are provided Section 4.5.

4.3 Methodology

4.3.1 Description of hydrometeorological data

This chapter uses the same hydrometeorological data than Chapter 3.

4.3.2 Hydrological model calibration

The models were individually calibrated on a 10-year period and the $RMSE$ on square-rooted streamflows as objective function. The calibration is consequently not designed to capture specifically higher discharges, which makes the present system assessment demanding as it focuses on the highest discharge percentiles.

4.3.3 Hydrological forecast economic value

Justification of the use of the relative economic value

Attempts to realistically define the damage associated to a particular river stage and avoided loss are subject to many approximations and errors and limited by the spatial and temporal boundaries (Merz et al., 2010). The intangible costs resulting from deaths or traumas, for instance, are hardly economically quantifiable by their nature. When damages are tangible, the cost evaluation is more straightforward but also subject to approximations as the flood indirect consequences may be difficult to identify. To properly quantify damages, the approach has to be sufficiently holistic to encompass all effective consequences that can be social, political and environmental (Merz et al., 2010).

To assess the economic gain related to protected values, Parker et al. (2007) uses an estimation of the proportion of moveable inventory within a property. The main limitation of this approach is the fact that there are plenty of other measures, potentially more efficient that can be taken to prevent damage loss. Moreover, the flood alert is not systematically followed by people and efficient preventing measures are not always taken. More generally, decisions are made under constrains and can be encumbered by cues that are fallible, ambiguous and altered by judgement (Choo, 2009).

This gave rise to the concept of maximum potential reduction of flood damage that relates the actual flood damage avoided to other factors that stands in the way of optimal mitigation (Parker, 1991). The relation is defined as the product of the maximal potential reduction for a perfect system, the probability that the forecast is issued sufficiently ahead to react, the fraction of concerned people that will respond to the warning and the fraction of people who will take effective measures. This product is estimated to 0.5 in UK by the Department for Environment, Food and Rural Affairs (Verkade and Werner, 2011).

To free ourselves from these constrains and approximations, we propose to use the Relative Economic Value (REV) and the cost-loss ratio to compare the different forecasting systems at hand. The REV is a more theoretical assessment of the value of a forecast as it is not based on real damage statistics but can be easily transferable to more practical cases providing an extensive knowledge of the area, residents, economical activities and goods at risk.

Cost-loss ratio

The cost-loss ratio (CLR) represents the ratio of the costs of mitigation and the avoidable losses due to adverse event and is defined as:

$$r = \frac{C}{L_a} \tag{4.1}$$

where r is the CLR, C is the cost of the mitigation warning response, and L_a the avoidable damage. In the following study, results are presents for values of r such as $0 \leq r \leq 1$, as it does not economically make sense to take preventive measures that are more expensive than avoidable damages. On the other hand, the CLR cannot equal 0 as operating an EWS already implies some cost and such hypothetical case would benefit from continuous warning as they are free. In practice, the cost-loss values are situated in a narrower range but the definition of realistic CLR is out of scope. This definition is convenient as it can theoretically encapsulate different costs (costs to set/initialize the EWS, costs of operation, and costs associated with the event mitigation), and all sources of avoidable loss. Therefore, a wide range of potential cases can be built upon this synthetic value.

Relative economic value

The Relative Economic Value (REV) is a dimensionless factor that scales between the case where no warning is issued and the perfect warning case (Zhu et al., 2002). A REV equals to 1 denotes the best possible forecast decision system while a REV equals to 0 indicates that the system does not provide gain over the no warning case. The REV is negative whenever the sum of costs of issued warnings is greater than the sum of avoided loss.

The estimation of the costs of an Early Warning System (EWS), a no-warning system and perfect forecast cases can thus be estimated from a contingency table. Table 4.1 describes the

Table 4.1: Contingency table with costs associated with each type of event.

		Forecast/action: Yes	Forecast/action: No
Observation	Yes	Hit (h) Mitigated Loss ($C + L_u$)	Miss (m) Loss ($L_a + L_u$)
	No	False Alarm (f) Cost (C)	Correct false (c) No Cost (–)

4 possible cases and their frequency of occurrence, hit (h), false alarm (f), missed events (m) and correct false (c), and the different types of expenses, i.e. avoidable loss (L_a), unavoidable loss (L_u), and operating cost (C).

If no warning system is available, the expected damages over the assessment period are given by:

$$E_{nowarn} = (h+m)(L_a + L_u) \tag{4.2}$$

The no-warning case can also be seen as a system that has the same skill as the climatology as the frequency of river stages exceeding the critical threshold is statistically low, i.e. sufficiently low that the user will not issue warning if no forecast information is available (Verkade and Werner, 2011).

On the other hand, the damages for a perfect forecast is the product of the number of events that lead to damages and the sum of the warning and response costs and the unavoidable losses.

$$E_{perfect} = (h+m)(C + L_u) \tag{4.3}$$

Finally, the expected damages for an early warning system includes the costs associated with correct alarms, missed and therefore unprotected adverse events, and the costs of false alarms.

$$E_{EWS} = h(C + L_u) + m(L_a + L_u) + fC \tag{4.4}$$

The expected damages with EWS, the no-warning, and the perfect warning system can be therefore compared with the REV. It is the equivalent of a skill score that scale between the optimal value $E_{perfect}$ and the reference value E_{nowarn}.

$$\begin{aligned} REV &= \frac{E_{nowarn} - E_{EWS}}{E_{nowarn} - E_{perfect}} \\ &= \frac{(h+m)L_a - (h+f)C - m(L_a)}{(h+m)L_a - (h+m)C} \\ &= \frac{(h+m) - (h+f)r - m}{(h+m)(1-r)} \end{aligned} \tag{4.5}$$

Optimal decision rule

The above decision-making framework accepts deterministic or probabilistic hydrological forecasts. If the decision to issue a warning is solely based on mathematical evidence, the forecaster will issue a warning if the expected damages with mitigation action are lower than the expected value without mitigation. The optimal decision rule is therefore

$$\begin{aligned} C + PL_u &< P(L_a + L_u) \\ \frac{C}{L_a} &< P \\ r &< P \end{aligned} \tag{4.6}$$

In the case of the deterministic forecast, the probability of exceeding the threshold is either 0 or 1 depending if the forecast is under or above the defined threshold.

4.3.4 Flood threshold

In practical cases, flooding thresholds are expressed in terms of river stages. However, as we do not possess rating curves nor critical river stages, we define a streamflow threshold and make the assumption that at each stage corresponds a unique streamflow.

A threshold is thus set for each catchment and defined as the quantile 0.9 of the observed streamflows on the assessment period. Unlike a streamflow value determined from a return period for instance, it allows to have the same number of adverse events on all catchments. Moreover, with a 2-year return period streamflow, almost 75% of the studied catchment would not experience any event needing mitigation measures during the testing period. Therefore, this streamflow threshold may not refer to flooding hazards but can be related to the needs for smooth functioning of many aspects in economy, administration and society (Frei, 2010) or dam management (Boucher et al., 2012).

4.3.5 System selection

For concision purpose and graphical convenience, only 5 out of the 9 systems presented in the previous chapter are retained. The economic value of systems A, B, C, E and H' will be presented. The choice is based on the differences in devising these systems.

- A: Most simple and fully deterministic case
- B: Ensemble weather forecast
- C: Probabilistic streamflow assimilation with EnKF
- E: Hydrological multimodel
- H': Most complex case. Combines multimodel, ensemble forcing, and data assimilation.

Table 4.2: Description of the five selected hydrological ensemble prediction systems

Systems	A	B	C	E	H'
Multimodel	Off	Off	Off	On	On
EnKF	Off	Off	On	Off	On
Met. ensemble	Off	On	Off	Off	On
Nb of members	1	50	50	20	50000

H' has been retained over the system H since it proved to be the most consistent with the description of uncertainty and offers overall better predictions. Other cases have not been kept because they are combinations of the selected systems. For systems A, B and C that do not include a multimodel approach, a model selection is necessary. For this purpose, the median performing model over catchment and lead times is kept. The model for the systems A and B is M09 and M06 for system C.

4.4 Results

4.4.1 System relative economic value

Figure 4.1 illustrates the REV for the 20 catchments, the 5 selected systems, and day 1. Each of the color line corresponds to a system and is function of the cost-loss ratio. The first observation is that the REV varies according to catchment which was expected as the systems simulate catchment behavior with various quality levels. Nonetheless, all systems remain economically efficient for most CLR values.

The second finding from Figure 4.1 is that the value of the forecast depends on the system. Indeed, system H' proves to be the most valuable since it provides the highest economic value for most catchments and CLR, and always (except for catchment 022301) exhibits positive REV even in the most demanding situations, i.e. for cost-loss ratios close to 1. It may be occasionally challenged by system C on a few catchments and specific CLR values. By contrast, systems A and B frequently the systems that offer the lesser improvement over the no-warning expected damage. These two systems only differ by their meteorological forcing, (i.e. probabilistic or deterministic) and exhibit identical REV for day 1. This indicates that picking randomly one member does not contribute to any difference in economic value for the first lead time as the difference between meteorological members is too little.

Occasionally, REV remains constant regardless of the CLR. This particular situation arises when no false alarm is issued. Also, an increase in REV can be sometimes noticed with increasing CLR. This is explained by a diminishing number of false alarms, which is related to the optimal decision rule. With increasing CLR, the degree of certainty of exceeding the

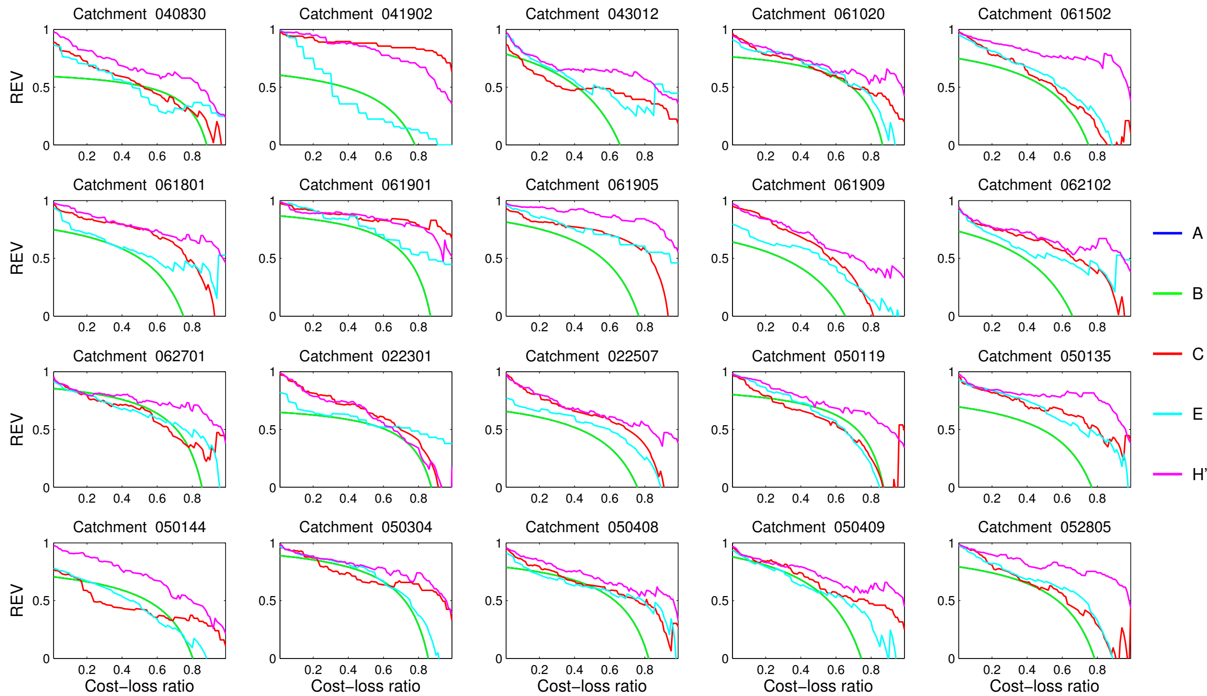

Figure 4.1: Relative Economic Value (REV) according to the cost-loss ratio for the 20 catchments, the 5 early warning systems and day 1. The differences between systems A and B results are not distinguishable.

streamflow threshold needs to be higher before issuing a warning and may therefore contribute to the rejection of a false alarm.

Figure 4.2 exhibits similar results than Figure 4.1 but for day 5. The value of forecast logically decreases as the number of false alarms and missed event increases. An increasing number of false alarm may be identified by a rapid decrease in REV for higher *CLR*. However, in an operational context, the deterioration of the forecast value needs to be qualified as the cost-lost ratio describing the situation is susceptible to be lower with increasing mitigation time. This refers to the notion of trade-off between the forecast value and lead time, a topic that is not addressed in this study. In fact, lead times used in this study are considerably longer than most of the study previously mentioned. For instance, the maximum mitigation time is 6 hours in Verkade and Werner (2011).

The superiority of system H' is confirmed in Figure 4.2. By contrast, system C, which was often the second best system in day 1, experiences a substantial decrease and henceforth is now challenged by system E. It is that the influence of the EnKF vanishes and may in some cases become worse than the open loop. System B that was indistinguishable from system A in day 1 gains in relative efficiency as it outperforms system A by taking into account the meteorological forcing uncertainty.

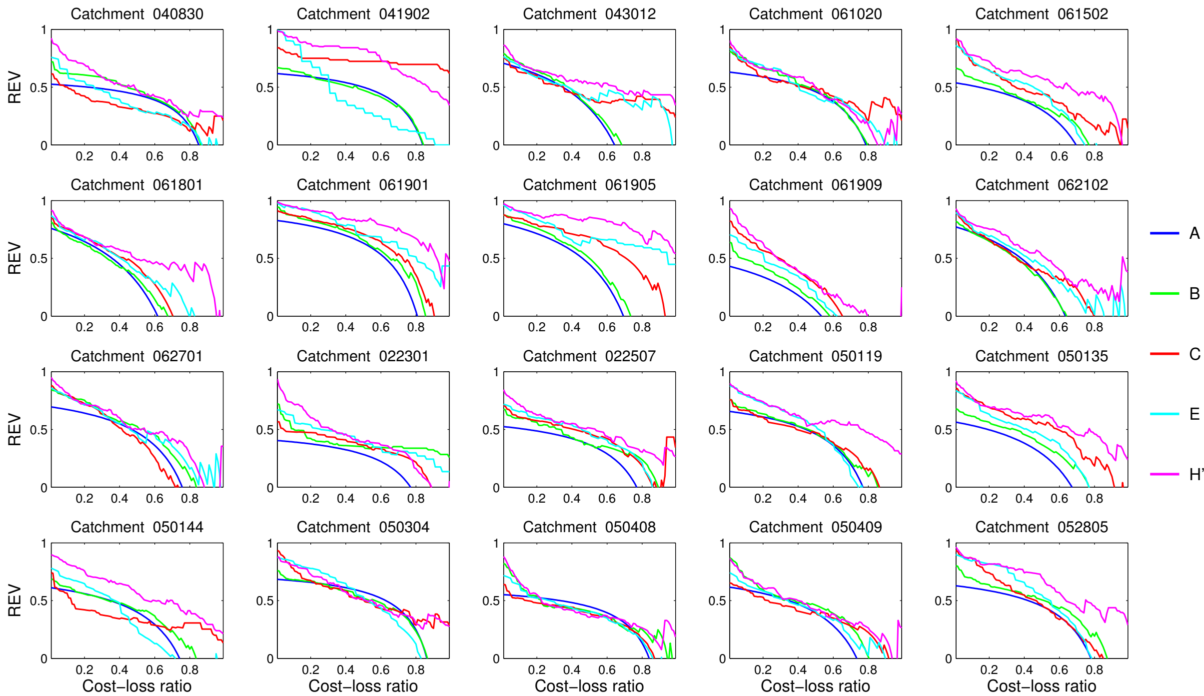

Figure 4.2: Relative Economic Value (REV) according to the cost-loss ratio for the 20 catchments, the 5 early warning systems and day 5.

General conclusion from this analysis indicates that, following the example of forecast quality (Chapter 3), the economic value of a system can be improved by accounting explicitly for the principal hydrometeorological sources of uncertainty. Therefore, more complex systems that decipher more sources of uncertainty, but that are eventually also more demanding in development, tuning and computational requirement, are overall more economically attractive. Moreover, the difference in cost among the five systems is possibly small compared to the costs engaged in mitigation actions, avoidable damages, or damage flood relief.

4.4.2 Relation between value and quality

This section aims relating the forecast economic value to accuracy and reliability. Figure 4.3 is based on the definition of a REV threshold set to 0.5 and its corresponding CLR (referred as CLR threshold). More precisely, the CLR threshold is equal to the CLR value when the REV curve falls lower than 0.5 for the first time, or in other terms, this measure indicates from which CLR a system has a relative economic value that becomes lower than 0.5. If the REV is always lower than 0.5, then the CLR threshold is 0. This allows to condense the information and to merge lead times. To be able to identify the first lead time, the corresponding marker is a circle while others are denoted by asterisks. The decrease of the performance in accuracy and reliability is monotonic, thus, other lead times are easily identified. Note that the X-axis has been truncated and therefore, not all simulation are systematically represented. Missing

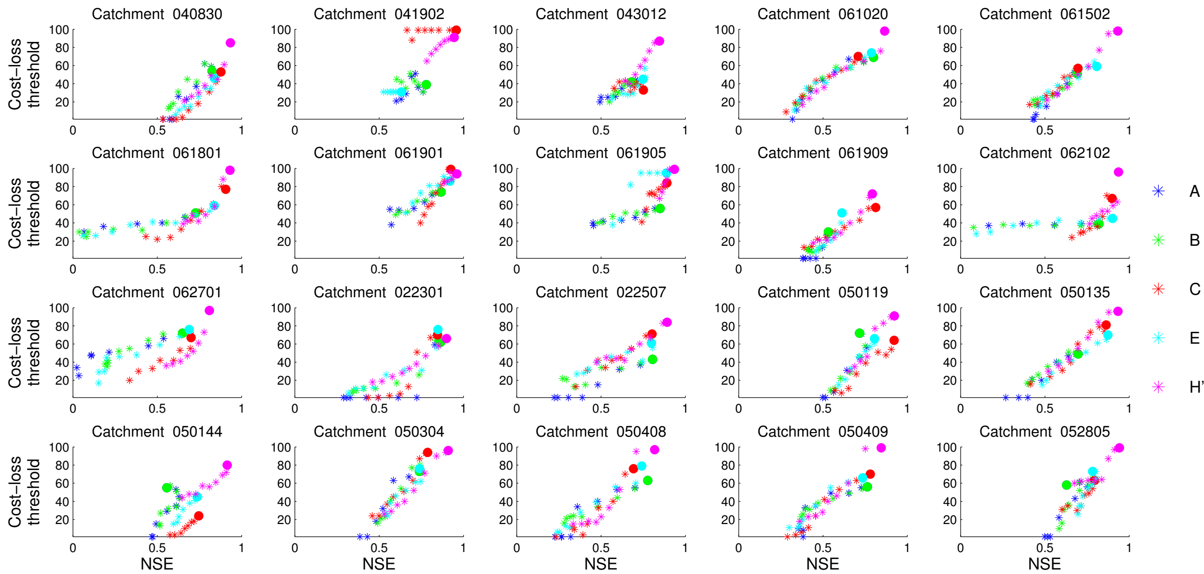

Figure 4.3: Relation between Nash Sutcliffe Efficiency (NSE) and the cost-loss threshold. The circles indicate day 1 results.

systems thus exhibit poor performance according to the NSE criterion.

Most catchments exhibit a correlation between NSE and CLR threshold, but these relations vary. For catchments 061020, 061502, and 050135, the relation between economic value and accuracy can be described relatively precisely by a simple polynomial relation, while relations for catchment 041902 or 062701 are less obvious. Highest NSE systems usually implies highest economic value but this may occasionally not be true where it can be over performed by few other simulations, as for catchments 041902, for instance. However, for a given CLR threshold it is frequent to observe different values of NSE. Thus, the relation between NSE and CLR threshold is not uniquely defined. This is particularly apparent for catchment 061801 where CLR threshold around 0.45 are obtained with NSE ranging from approximatively 0.8 to 0.2.

Once more, system H' clearly stands above the other systems both in terms of value and quality. Often, the CLR threshold of the H' for the second or third day (occasionally up to five days like for catchments 043012 and 050144) lies in the same area as other systems first lead time. This indicates that the equal-quality decision can be taken a few days earlier with most performing systems, while the most simple ones (A and B) are rarely competitive.

Figure 4.4 illustrates the distribution of CLR thresholds as a function of the NRR. System A is not represented as it is deterministic, so no associated reliability exists. A decrease in reliability is often accompanied by a decrease in expected economic value but the relation is

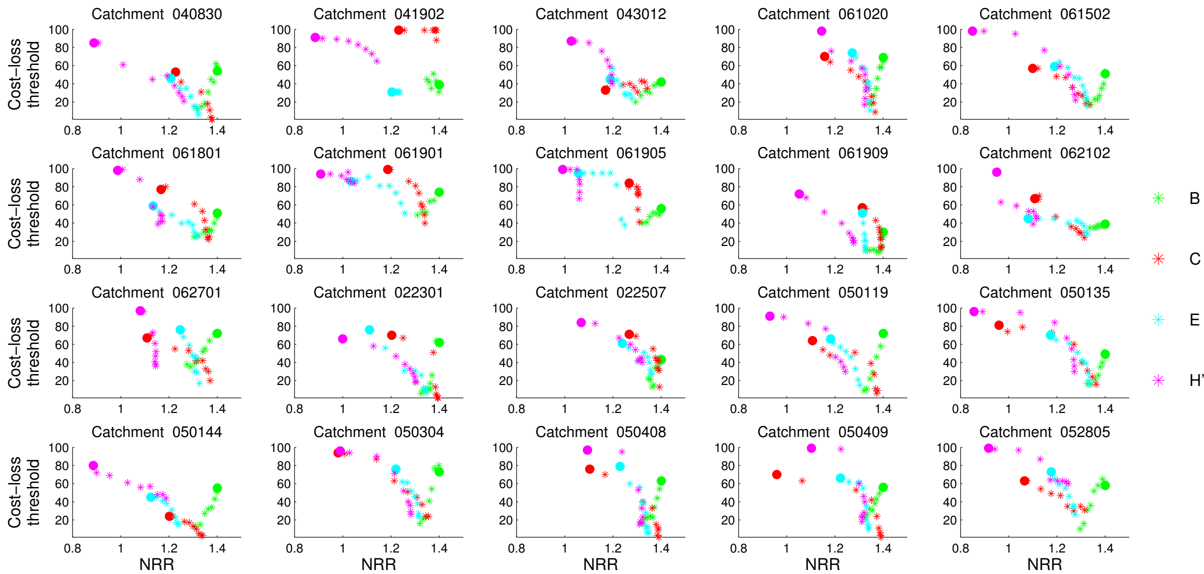

Figure 4.4: Relation between Normalized Root-mean-square-error Ratio (NRR) and cost-loss threshold. The circles indicate day 1 results.

not as evident as for the *NSE*. Also, system B does not follow the same logic as the others, since its spread is only generated by the meteorological forcing and logically increases with lead time. The increasing spread (and thus reliability in that case) combined with a drop in accuracy explains the reason why system B relation is oriented in the other direction. This also indicates that relying only on a reliability measure and to tune a system based solely on reliability maximization may be deceptive.

In the case we omit system B, the relation between reliability and economic value becomes clearer for most catchments. From there, it is worth noticing that the relations between economic value and *NRR* are explicit for the same catchments as the relation between economic value and *NSE*. Therefore, Figure 4.5 is used to verify if reliability and accuracy are correlated to each other, and to economic value together.

Figure 4.5 plots *NRR* against *NSE* for system B, C, E, and H' for all lead times. It reveals that *NSE* and *NRR* are correlated by systems. Again, the relation between reliability and accuracy of system B differs from the other ones because of the use of meteorological forcing alone. The relations actually differ according to catchments and systems, but generally a decrease in reliability is accompanied by a decrease in accuracy. This emphasizes that both aspects of forecast quality cannot be fully separated with the tested systems. The manner the forecasting systems were developed here makes that an improvement in the description of uncertainty with the use of the different hydrometeorological tools, improves reliability and accuracy together (see also Chapter 3).

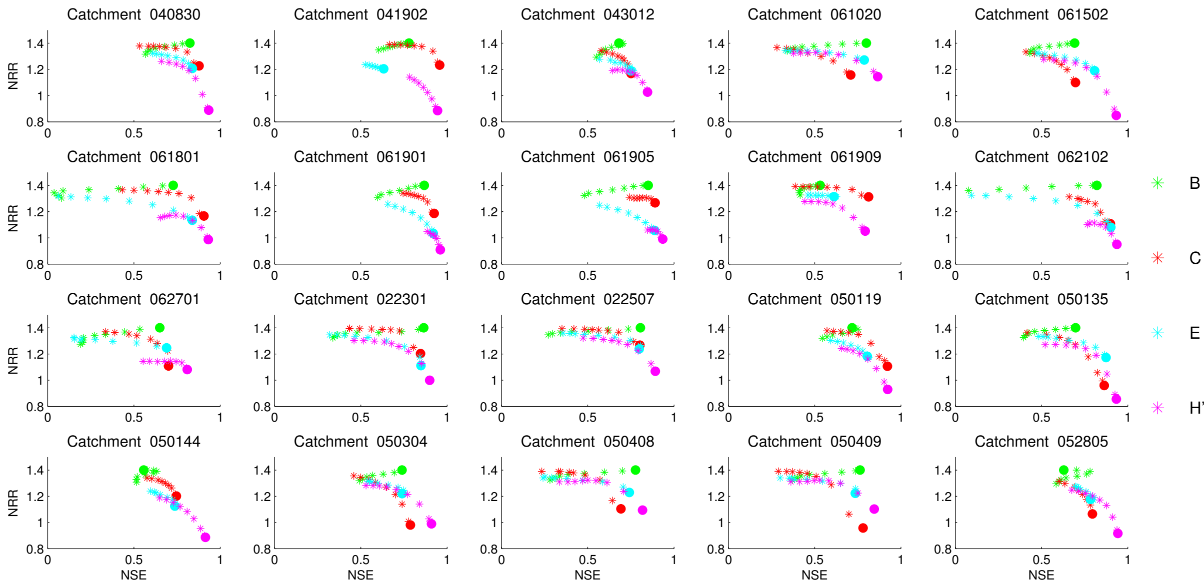

Figure 4.5: Relation between Normalized Root-mean-square-error Ratio (NRR) and Nash Sutcliffe Efficiency (NSE). The circles indicate day 1 results.

However, attention should be paid uppermost on accuracy, as even if reliability is shown to generally contribute to the REV for ensemble prediction, it is not a strict prerequisite since the deterministic system – although it provides the worse results among the tested EWS – still provides improvement over the no-warning case.

Finally, it becomes obvious that it not possible to precisely establish a relation between probabilistic forecast quality and value based on a conjoint measure of accuracy and reliability. Other scores have been tested (*MaeRD*, *RMSE*, *CRPS*) without any change in the conclusions. This explains that no clear threshold on accuracy and reliability from which the forecast is not valuable could be found. Murphy and Ehrendorfer (1987) proves with a more theoretical framework and the Brier score that the description of the value of a deterministic forecast cannot be fully described by a one-dimensional measure of accuracy, also referred as the accuracy/value envelope. This is verified in the present study as to a single measure of accuracy and reliability may correspond several economic values. Considering a second dimension in the forecast assessment through reliability measurements do not add more information as the two scores turn out to be not independent. Thus the confrontation of accuracy-reliability and value still lead to a multivalued function.

4.5 Conclusion

This paper presents (i) a comparison of five Early Warning Systems (EWS) in terms of economic value and (ii) an attempt to relate accuracy and reliability to economic value. Each of the EWS includes an imperfect forecasting component that differs from the others by the way it aims to decipher one or several sources of uncertainty and by their amount of total hydrometeorological uncertainty. Therefore, the systems vary in complexity, ranging from a single hydrological model to a system that includes a multimodel approach combined with ensemble Kalman filtering and meteorological ensemble forcing.

The assessment of the forecast economic value relies on a flexible theoretical framework where the relative economic value is used to scale the forecast value between the no-warning and perfect forecast cases. Warnings are issued and subsequent costly mitigation actions are carried out when the forecasted streamflow exceed a predefined threshold.

All systems are found to provide gain up to 9 days ahead for most catchments and cost-loss ratios. However, more complex systems provide higher economic value. By addressing specifically and adequately the three major sources of uncertainty in hydrometeorological modeling (i.e. initial condition, meteorological forcing and structural uncertainty), the expected damages are reduced the most efficiently for all lead times.

The study search for a relation between quality and value reveals that in most cases, better accuracy and reliability provide higher economic values. However, the link is loosely defined since for a given economic value, several NSE and NRR values exist. Therefore, using forecast quality only provides a rough estimate of the potential forecast value. This is also partly due to the fact that NSE and NRR are poor estimators of the economic value as, in practice, they are linked to one another, at least for the systems devised in this study.

Therefore the question of the identification of a relation between quality and value is still pending and additional work need to be dedicated to this topic as end-users are frequently interested in economic considerations. Further work is also required for more concrete applications, especially by a proper consideration of suitable mitigation measures, identification of actual flood damage avoided, and the possibility of improved decision rules.

General conclusion

Despite the improvement in the description of hydrometeorological processes that have been achieved during the past decades, comprehensive handling of uncertainty in modeling remains a daunting challenge. Sources of uncertainty with different characteristics persist at several locations in the hydrometeorological modeling chain and are aggregated in the forecasting predictive distribution, and therefore stand in the way of optimal use of forecasting and decision-making systems. Efforts have been dedicated to reduce uncertainty by developing a wide range of tools that capture a fraction of the total uncertainty, notably by resorting to ensemble techniques, i.e. a collection of deterministic forecasts. This shift from deterministic to probabilistic approach improved uncertainty deciphering but shortcomings need to be addressed as these techniques typically focus on a single source. To date, no framework to handle major sources of uncertainty have been identified.

This thesis aims to broaden the knowledge on the way to explicitly account for the three major sources of hydrometeorological forecasting uncertainty, namely the meteorological forcing, the hydrological model initial conditions, and the structural/conceptualization model. Particular attention should be paid to the coherence of the system to ensure that each sources of uncertainty is properly tackled by the corresponding and suitable component of the forecasting system, without overlapping neither in their action or compensation for unaddressed uncertainties. Such system is expected to provide, accurate, reliable, and economically valuable forecast.

To achieve this goal, a framework that assesses and reduces specifically each source of uncertainty that have different nature and different location is built from three dedicated tools. The Ensemble Kalman Filter (EnKF), the multimodel approach, and the meteorological ensemble forcing are all probabilistic tools that were selected to respectively decipher initial condition, structural, and forcing uncertainties. To identify their contribution to the reduction of the predictive uncertainty, they have been alternatively tested alone and combined, which allow to have a more proficient knowledge of their combination, a point that was rarely assessed.

Therefore, several systems were constructed by including different tool combinations leading to systems that theoretically decipher different amount of the total uncertainty. When combining all tools adequately, the forecasting system shown to be more accurate, more reliable with higher economic value than any other tool sub-combination even if the link between economic value and forecast quality is not straightforward. This also indicates that in the identified best system, modeling uncertainties are best estimated and reduced by the tools that fulfil their objectives.

A multimodel ensemble, composed by 20 dissimilar models, shown to contribute for explicit accounting of structural and conceptualization uncertainty. A comparison between the multimodel ensemble and a physically based semi-distributed model was carried out in an operational context with deterministic meteorological forcing and simple output updating. It revealed that the multimodel approach is more prone to lead to accurate forecast and correct representation of complex events, and can be therefore act as an operational solution. The probabilistic approach should be preferred over its deterministic counterpart not only because retaining every ensemble members leads to better performance than the ensemble median or any of the model taken individually in most cases, but also because it provides more information such as an estimation of the uncertainty through reliability measurements and prevents from the model selection pitfall. The multimodel approach is superior thanks to the different locations every model occupies in the ensemble, indicating that they perform roles that are different and may contribute to the ensemble quality and diversity. Moreover, in the case where the snow accounting routine is calibrated together with each individual hydrological model, its parameter values will be different for each model. Therefore, this multi-parametrization of the snow module will implicitly account for part of the snow accumulation and melt uncertainty through parameter uncertainty.

The superiority of meteorological ensemble prediction system for hydrological purpose is confirmed in the thesis. Unlike hydrological model pooling, the use of meteorological ensembles did not bring clear improvement over deterministic forcing in terms of accuracy, but remains a key element as it contributes to system reliability from medium range horizons and on.

Lastly, by efficiently reinitializing model states based on streamflow observations, the EnKF, if properly set, is able to provide suitable initial conditions that greatly improve predictions for shorter lead times. The EnKF deciphers initial condition uncertainty and contributes largely to the predictive ensemble accuracy and required dispersion. However the positive effects of the filter quickly fade out after spin up. The information that is contained in the different sets of state variables is poorly propagated through time because of hydrological model resiliency. As a result, the ensemble reliability decreases if the EnKF is used alone indicating that deciphering the accumulation of past error is a priority but is not sufficient for medium range forecasting. In fact, the EnKF is a powerful tool for short range forecasting but needs to be supported by a dynamical technique to account for other sources of uncertainty, such as multimodel and ensemble forcing.

In order to achieve optimal implementation, the EnKF setting is more subtle than frequently suggested. Indeed, the specification of the hyper-parameters is not straightforward, as they have to be set according to tools that are combined with the EnKF. In the case where the EnKF is used alone, input forcing perturbations will have to be screened in detail and will be set to unrealistically high values to additionally account for structural and parameter errors. In such situation, optimizing reliability and accuracy together is complex since the maximization of one aspect of forecast quality is achieved at the expense of the other. Also, the choice of the state variables to update may not be intuitive for the tested models. The optimal selection is often a sub-ensemble of model state variables and no global pattern emerges to ensure a flawless choice, especially for models with higher degree of freedom that are often more complex to tune finely. Another difficulty arises in the fact that both hyper-parameters values and the choice of state variables to update depend on models and catchments. Thus a transfer of an optimal EnKF tuning from a case to another may be hazardous. Finally, there are important differences in the expected gain according to models since none equally benefits from states updating.

The tools that were used for building the forecasting system are complementary because they allow to capture and decipher different uncertainties at different locations. Their combination is found to outperform any of them separately and to provide accurate, reliable, and valuable forecasts. However, particular attention in the manner to combine them should be paid to prevent unaddressed uncertainty or overlapping effect, especially in the EnKF, which presents flexibility in its implementation.

Prospects

In the framework that was suggested, models were reinitialized individually through the EnKF. As a consequence, error covariances and conditional density of the model states were estimated one model at a time. Therefore, the EnKF has updated the model states regardless of the ensemble quality and consistency, only with the goal to optimize the initial conditions of the model it is dealing with. Intuitively, this is not a major drawback since the ensemble quality depends on its members. By improving each member accuracy, the ensemble is likely to become more accurate. Hence, updating simultaneously the state variables of all models would provide a more coherent handling of the uncertainty and a better control over the ensemble spread. Possible EnKF finer tuning could be achieved through input and output perturbations. These perturbation would account for input and output only, without the need to hypothesizing that the error structure were compensated by state variables or parameters correction. Thus it would not be necessary to overestimate perturbations of the inputs to subsequently influence states variables like it is the case in Chapter 2, nor to call upon direct perturbation of state variables, nor to update state variables. Consequently, the conjoint model reinitialization would theoretically improve the initial condition of the system itself. The ensemble would not be therefore composed by a collection of models but by a multistructural system in its own. The suggestion of this novel system raises several questions.

To continue in the multistructural perspective, for better coherence, models should no longer be calibrated individually but collectively. This leads to a problem of identification of a suitable way to calibrate probabilistic forecast. A multicriterion that includes one or several measurements of accuracy and reliability could be considered. However, by calibrating simultaneously every model, the number of parameters increases dramatically leading to poorly constrained optimization, which is very likely to provoke equifinality issues.

Also, in this thesis, it is shown that the system that performs best includes 50.000 members. It is clear to us that such amount of values is not needed to sample the predictive density function, even if its shape is complex. Therefore, some member selection should be carried out to reduce ensemble size. This selection can be done at several levels. It appears that selecting meteorological members in a mathematically consistent way is complex as ensemble members are, in theory, exchangeable. A random sampling of forcing members could be intended, provided that there is no loss in the system predictive skills. An easier way to cut down in the ensemble size would be to perform a selection in the model pool. Despite 20 dissimilar model structures may be considered mandatory to provide a comprehensive description of the structural uncertainty, Chapter 2 shown that the compatibility between EnKF and model structure is not systematic. While the 20 models presented performance in open loop validation that are of the same order of magnitude, once models are updated, some exhibit poorer performance in comparison to the others. The model selection could be based on a backward greedy selection method or by a less systematic approach. We suggest for instance, instead on focusing directly on model performance or dissimilarities in the structure, to adopt a "negative approach", i.e. to focus on individual model misrepresentation of observations (individual model predictive error). This is also a key component of the multimodel approach as it relies on the possible compensation for model errors. Therefore, the emphasis should be paid on selection of models that present errors that are different. This probably requires a synthetic experimental framework to get rid of other sources of uncertainty such as forcing and observation.

The actual time step of the system (daily) may be considered too coarse for certain operational applications. Finer temporal resolution may be desirable, especially for small size or mountainous catchments that are likely to be subjects to flash floods and for hydraulic model coupling. As MEPS are already able to provide such temporal resolution, attention should be focused on the structure of the hydrological models. Models such as GR4J were modified in an hourly version by some simple modifications of model constants (GR4H) or by adding a reservoir for routing (GR5H). Such an approach could be extended to the 19 other models to convert them to a more suitable temporal resolution.

Decreasing the size of the ensemble is a critical aspect of the multistructural system updating since all states of all members should be stored simultaneously to derive the system error covariance matrix. As demonstrated in Chapter 2, the number of members is not a driving parameter for the model performances as long as it remains reasonably high. Further testing with fewer EnKF members is required. Also, even if meteorological studies indicated that EnKF is able to cope with high dimensional systems, this should be verified with the suggested framework.

Further improvement to the spring freshet forecasting could be obtained by dedicating more attention to snow processes. To do so, three main avenues are identified. In the presented framework, the uncertainty in the description of the snow processes was only deciphered through a multiparametrization of the snow module (Chapter 1) and state updating was carried out only on the hydrological model states leaving the snow stock unperturbed. Updating the Cemaneige snow reservoir could prevent situations where the freshet driven by snow melt is still going on, while Cemaneige reservoir ran dry, which is likely to happen since quantifying accurately the snowpack throughout the whole winter is subject to errors. The second possibility to improve upon the present snow routing accounting would be to account for the snow module structural and calibration uncertainty by updating the module calibrated parameters such as the melting factor. Such updating is expected to improve simulations where melting occurs on the catchment while the snow module does not reflect it and therefore allow, for instance, to provide more water to the hydrological model when needed. Finally snow accounting simulation could be enhanced by using additional observations than streamflows to assimilate. This requires more data, such as snow remote sensing or direct field observations but would allow to address the question of the snow independently of the rest of the hydrological processes. Therefore, the uncertainty could be reduced prior to cascading it to the hydrological models.

Moreover, by adding additional sources of observations, a step further could be done in untangling the modeling processes and their sources of uncertainties. In the thesis, it is not possible to state doubtlessly that each tool decipher only its source of uncertainty as they remain partly tangled. To overcome this issue, each process should be assessed independently implying more measurements at every step of the forecasting process.

Meteorological forcing is the element of the system that received the less attention. The assessment of the ECMWF forecast that was not shown in the thesis, revealed that the HEPS suffers both bias and frequent underdispersion. This was also observed by Verkade et al. (2013) and the author was able to achieve moderate gain with simple pre-processing techniques in precipitation forecasting but better results were obtained with temperature, which is a key input in snow modeling. The author adds that the gain in hydrological forecasting was negligible and this was probably partly related to errors in the initial condition that acts like a buffer. However, this may not be the case with the suggested system as the error in the initial condition should be smaller.

Appendix A

Watershed characteristics

Table A.1: Main characteristics of the catchments

Station number	River name	Region	Chap 1,2	Chap 3,4	Area (km2)	Mean ann. Q (m3/s)	Coeff. of variation	Mean ann. P (mm)	Mean ann. snow (cm)
022301	Trois Pistoles	Bas Saint Laurent	✓	✓	923	18	1.81	1109	382
022507	Du Loup	Bas Saint Laurent	✓	✓	512	10	1.47	1050	378
022704	Ouelle	Bas Saint Laurent	✓		784	16	1.90	1122	381
023402	Chaudière	Chaudière Appalaches	✓		5822	115	1.54	1099	306
023422	Famine	Chaudière Appalaches	✓		696	15	1.59	1142	328
024003	Bécancour	Centre du Québec	✓		917	22	1.40	1254	349
030101	Nicolet Sud-Ouest	Estrie	✓		550	12	1.56	1199	298
030103	Nicolet	Estrie	✓		1550	33	1.60	1186	291
030234	Eaton	Estrie	✓		647	13	1.67	1256	326
030282	Au Saumon	Estrie	✓		736	18	1.49	1262	352
030304	Noire	Montérégie	✓		1508	25	1.72	1143	256
030340	Yamaska Nord	Montérégie	✓		236	5	1.22	1244	277
030905	Chateaugay	Montérégie	✓		2503	36	1.64	1018	218
030907	Des Anglais	Montérégie	✓		638	8	2.12	970	207
040204	Rouge	Laurentides	✓		5460	104	0.98	1054	308
040830	Gatineau	Laurentides	✓	✓	6796	127	1.08	1023	332
041301	Coulonge	Outaouais	✓		5158	72	1.09	973	289
041902	Dumoine	Abibiti	✓	✓	3743	50	0.81	968	297
043012	Kinojévis	Abitibi	✓	✓	2572	39	1.12	921	324
050119	Matawin	Lanaudière	✓	✓	1383	24	1.11	1025	328
050135	Croche	Mauricie	✓	✓	1551	29	1.24	996	360
050144	Vermillon	Mauricie	✓	✓	2650	39	1.10	957	312
050304	Batiscan	Mauricie	✓	✓	4483	96	1.03	1162	381
050408	Saint-Anne	Capitale Nationale	✓	✓	1539	51	1.20	1412	502
050409	Bras du Nord	Capitale Nationale	✓	✓	643	19	1.21	1385	499
052212	Ouareau	Lanaudière	✓		1262	26	1.25	1133	376
052219	L'assomption	Lanaudière	✓		1287	24	1.29	1072	350
052233	De l'Achigan	Lanaudière	✓		642	11	1.55	1091	299
052805	Du loup	Mauricie	✓	✓	767	12	1.27	1020	332
060101	Petit Saguenay	Saguenay			695	15	1.51	986	398
061020	Aux Ecorces	Saguenay	✓	✓	1107	28	1.09	1236	450
061502	Métabetchouane	Saguenay	✓	✓	2202	48	1.19	1168	420
061801	Péribonka	Saguenay	✓	✓	1010	19	1.16	1000	376
061901	Ashuapmushuan	Saguenay	✓	✓	15342	300	0.92	984	379
061905	Ashuapmushuan	Saguenay	✓	✓	11200	227	0.88	1001	394
061909	Au Saumon	Saguenay	✓	✓	586	8	1.36	877	334
062101	Mistassibi	Saguenay	✓		8676	204	1.01	1012	422
062102	Mistassini	Saguenay	✓	✓	9534	200	1.08	1004	409
062701	Valin	Saguenay	✓	✓	761	24	1.13	1123	453
062802	Sainte Marguerite	Cote Nord			1093	31	1.33	1120	475

Appendix B

Model structures

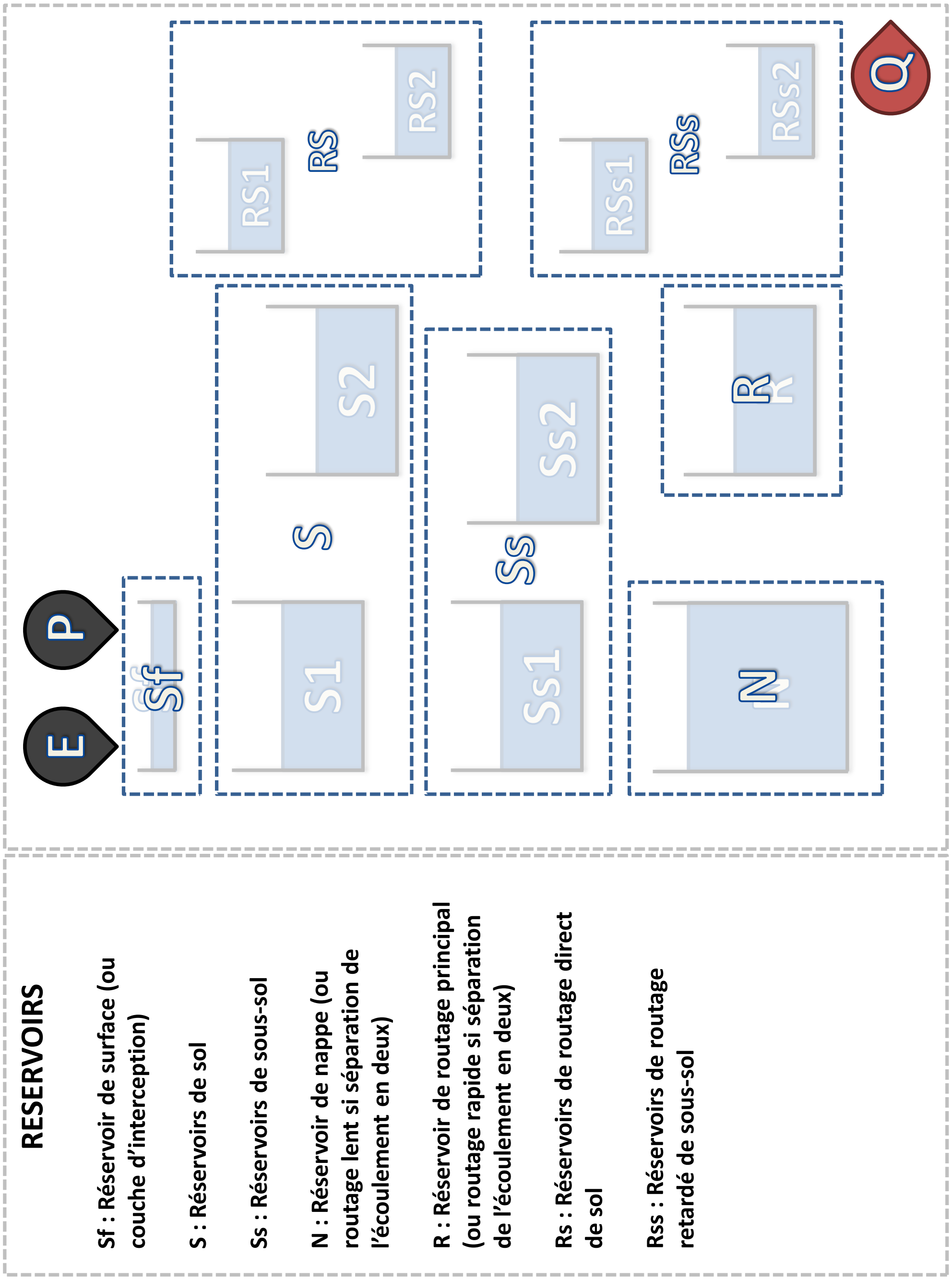

Figure B.1: Description of the models' structures (from Seiller, 2013)

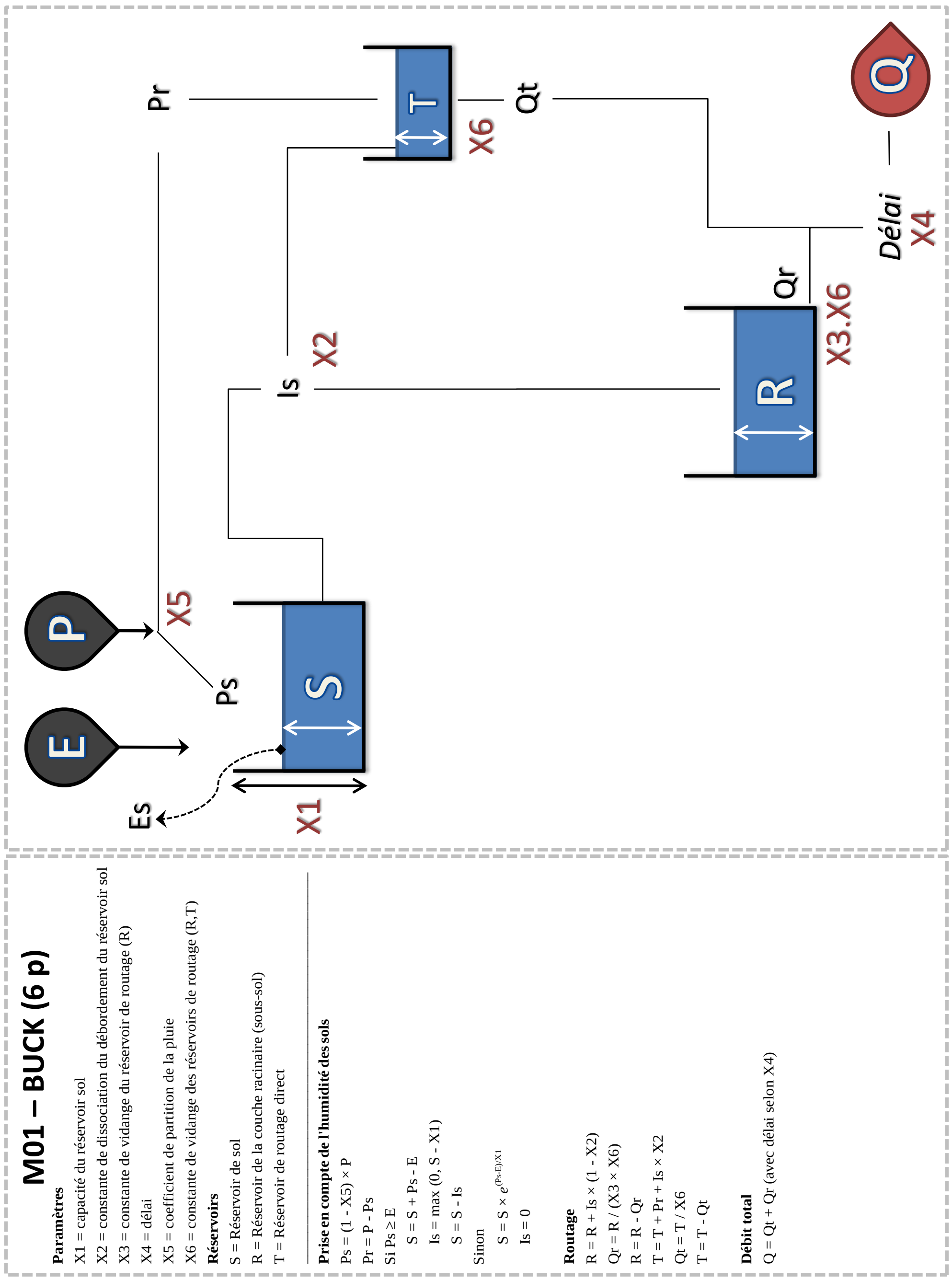

Figure B.2: Structure of model M01 (from Seiller, 2013)

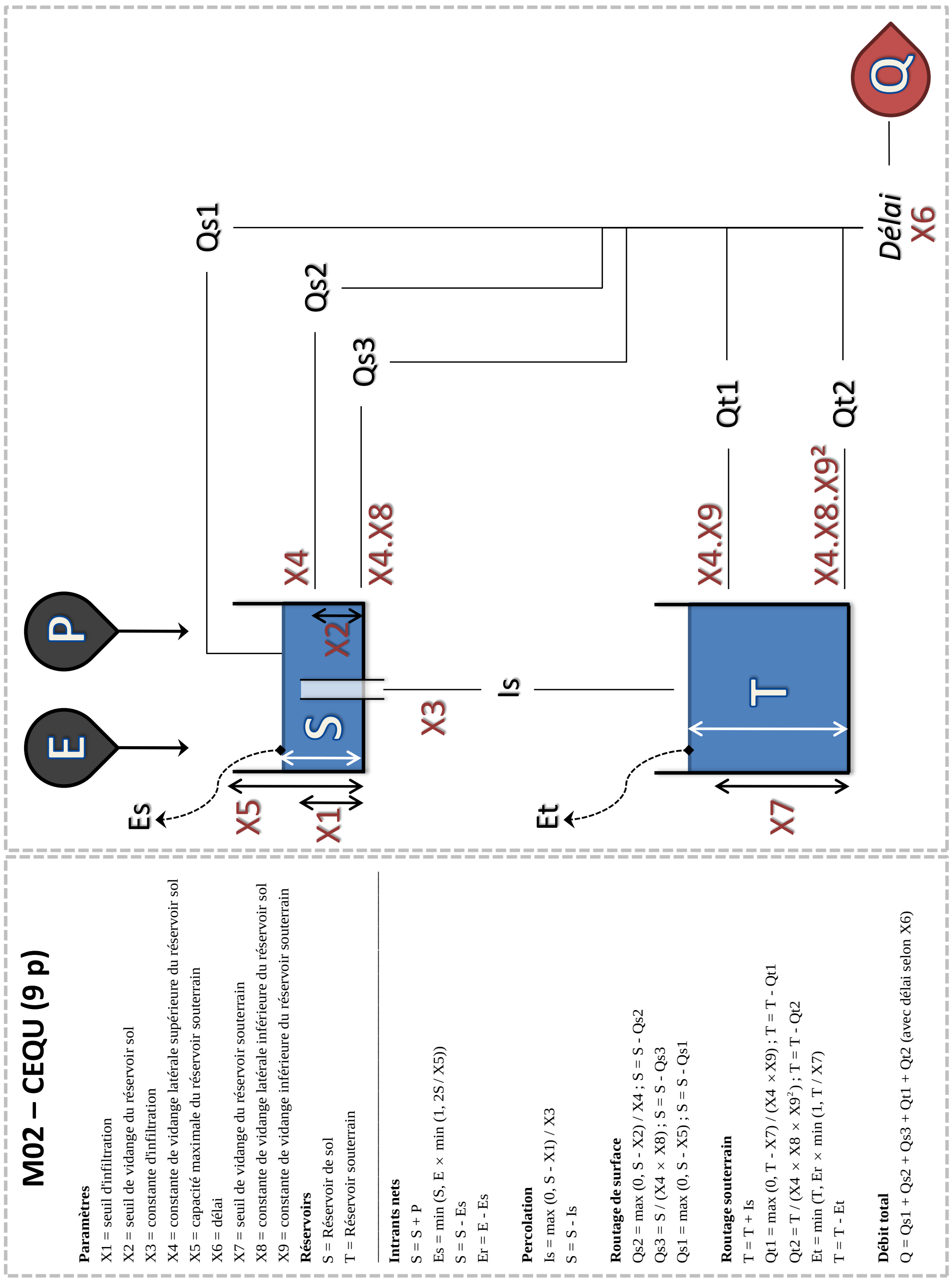

Figure B.3: Structure of model M02 (from Seiller, 2013)

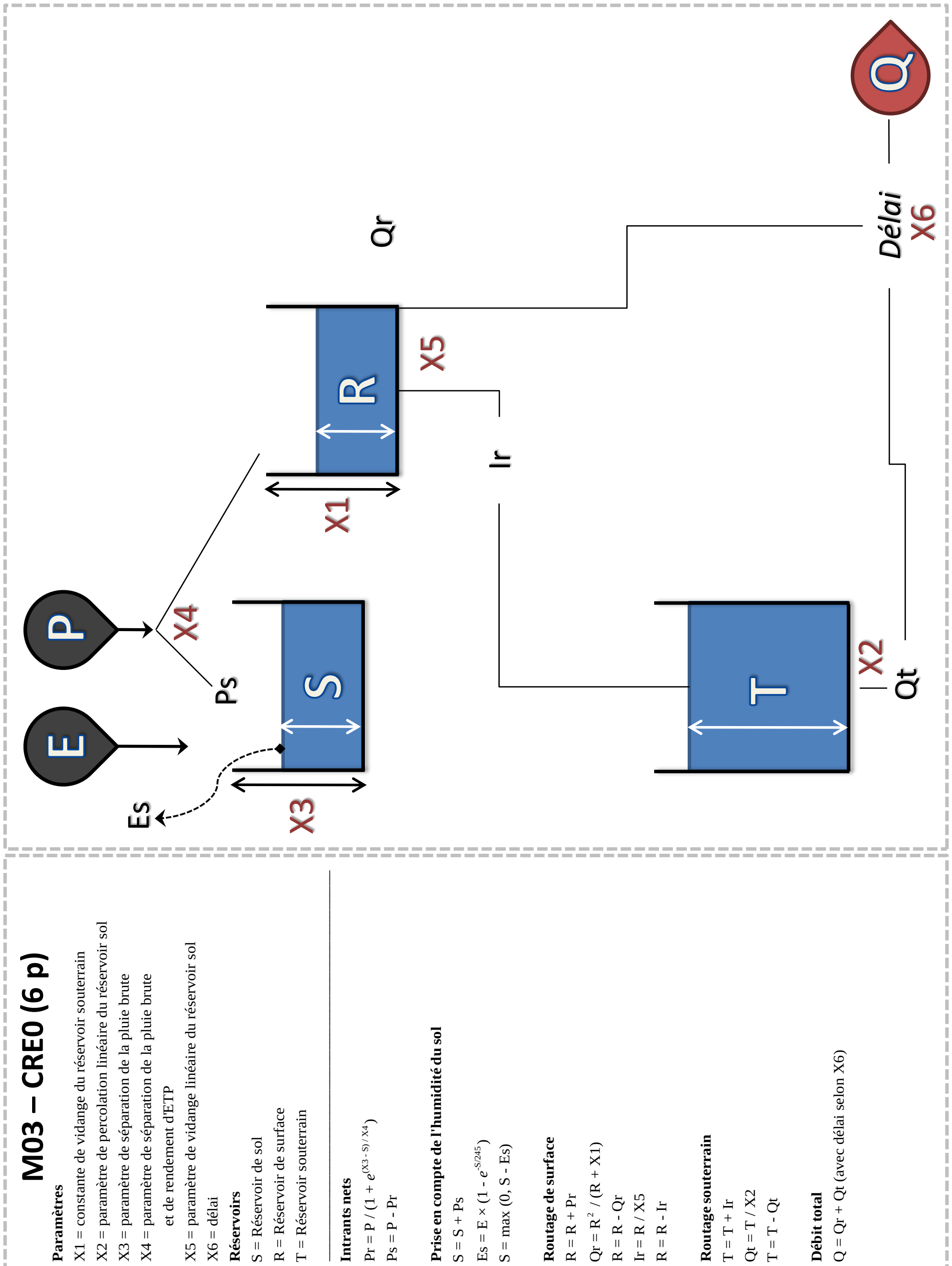

Figure B.4: Structure of model M03 (from Seiller, 2013)

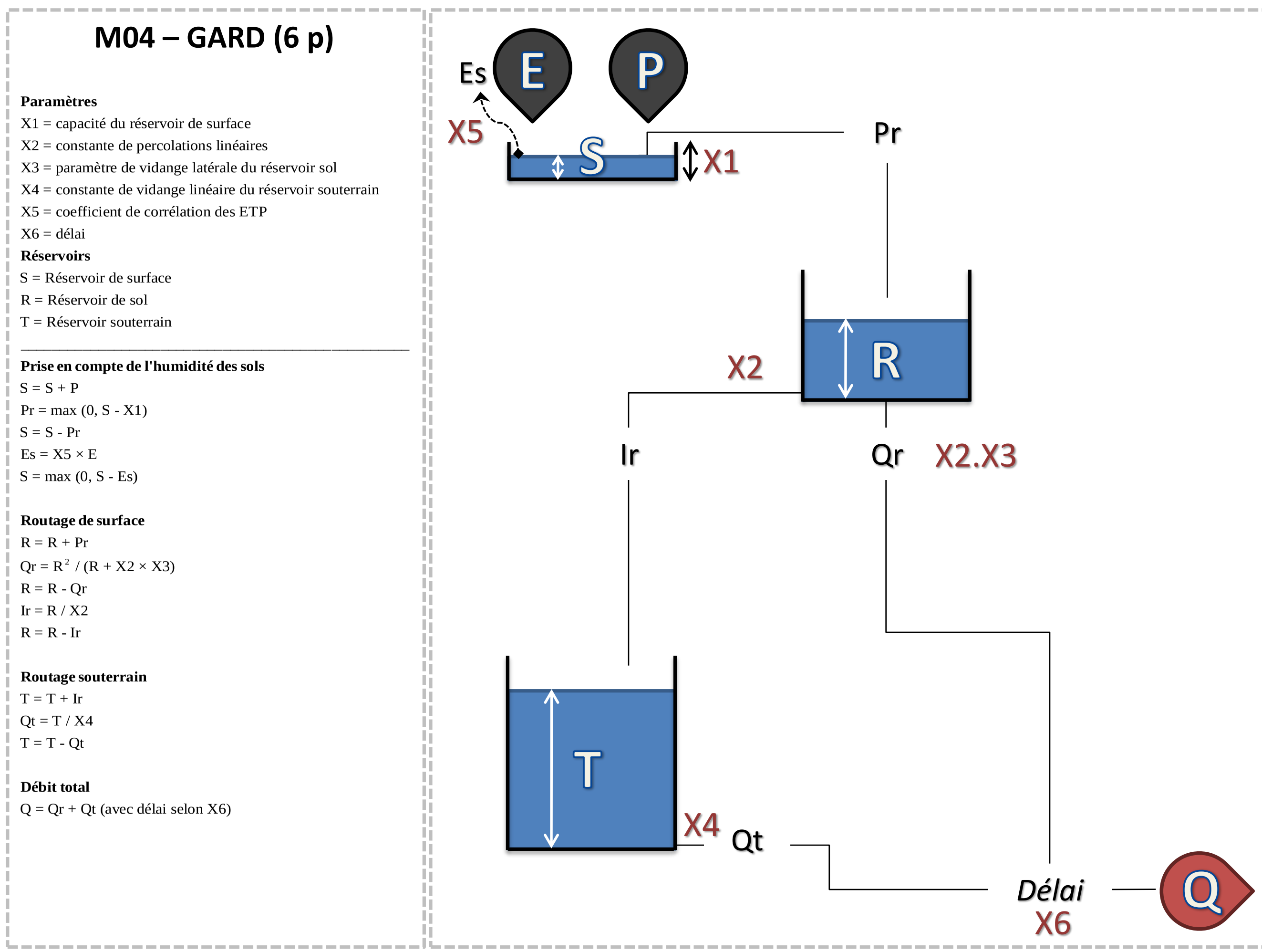

Figure B.5: Structure of model M04 (from Seiller, 2013)

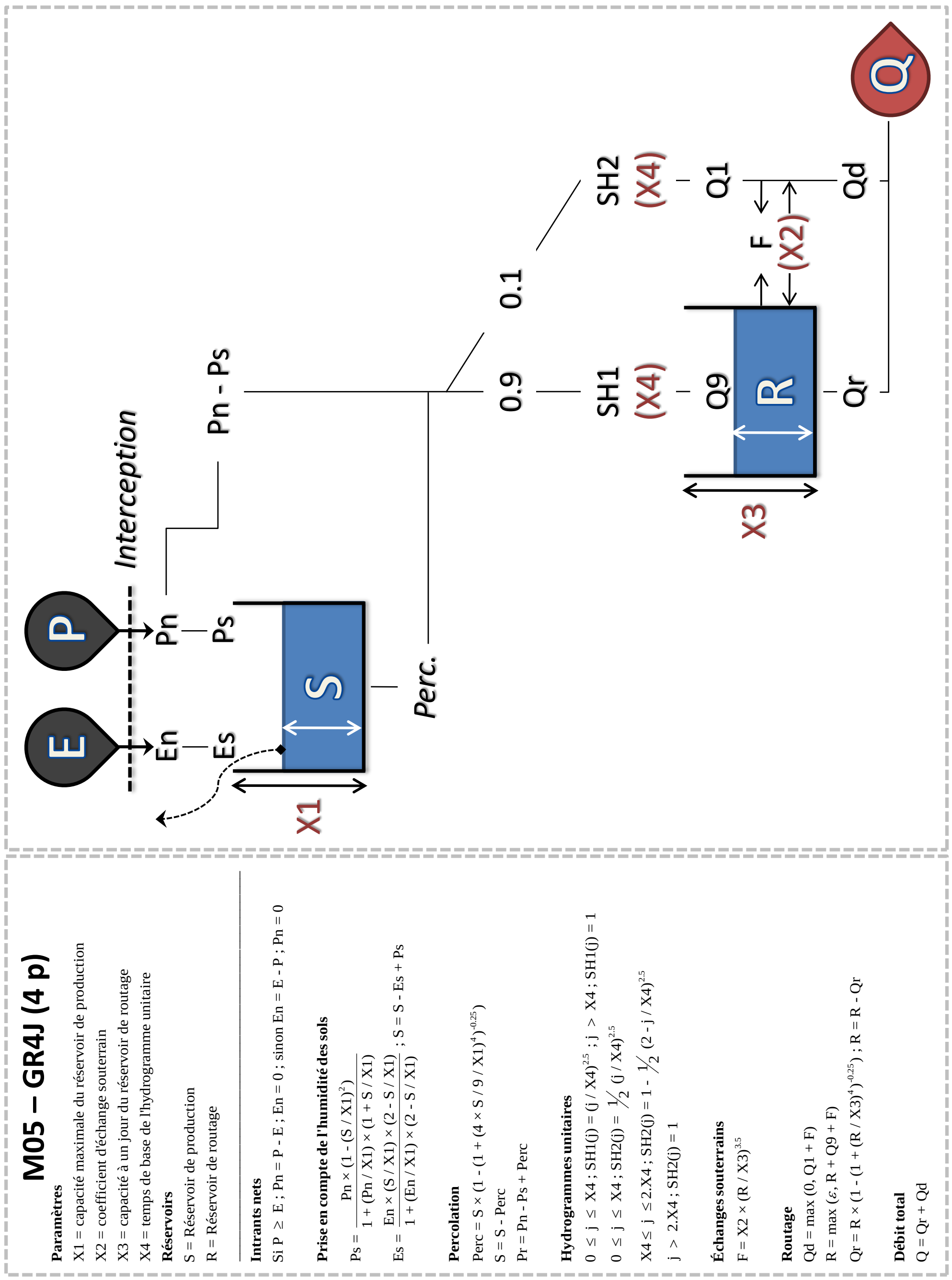

Figure B.6: Structure of model M05 (from Seiller, 2013)

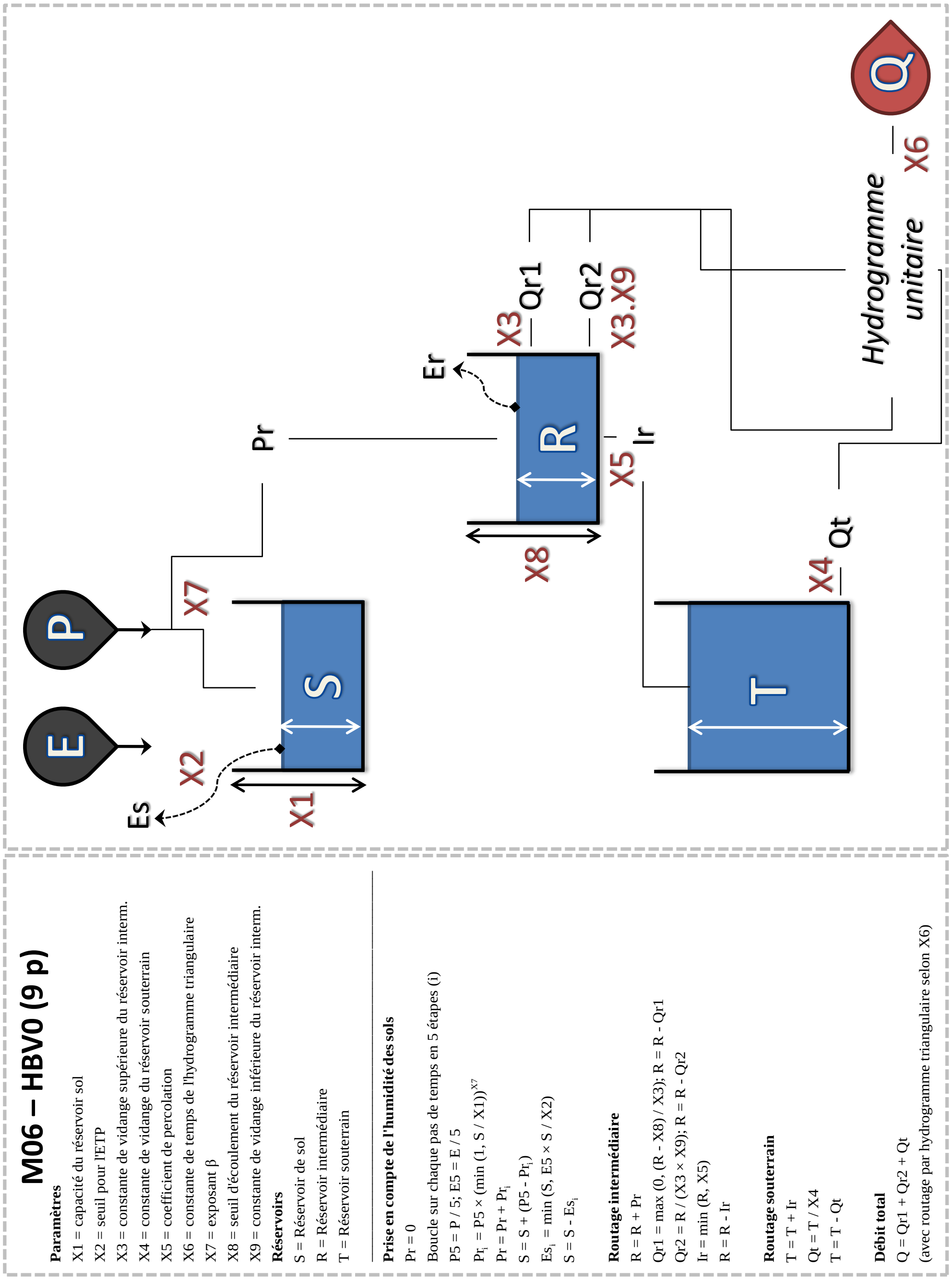

Figure B.7: Structure of model M06 (from Seiller, 2013)

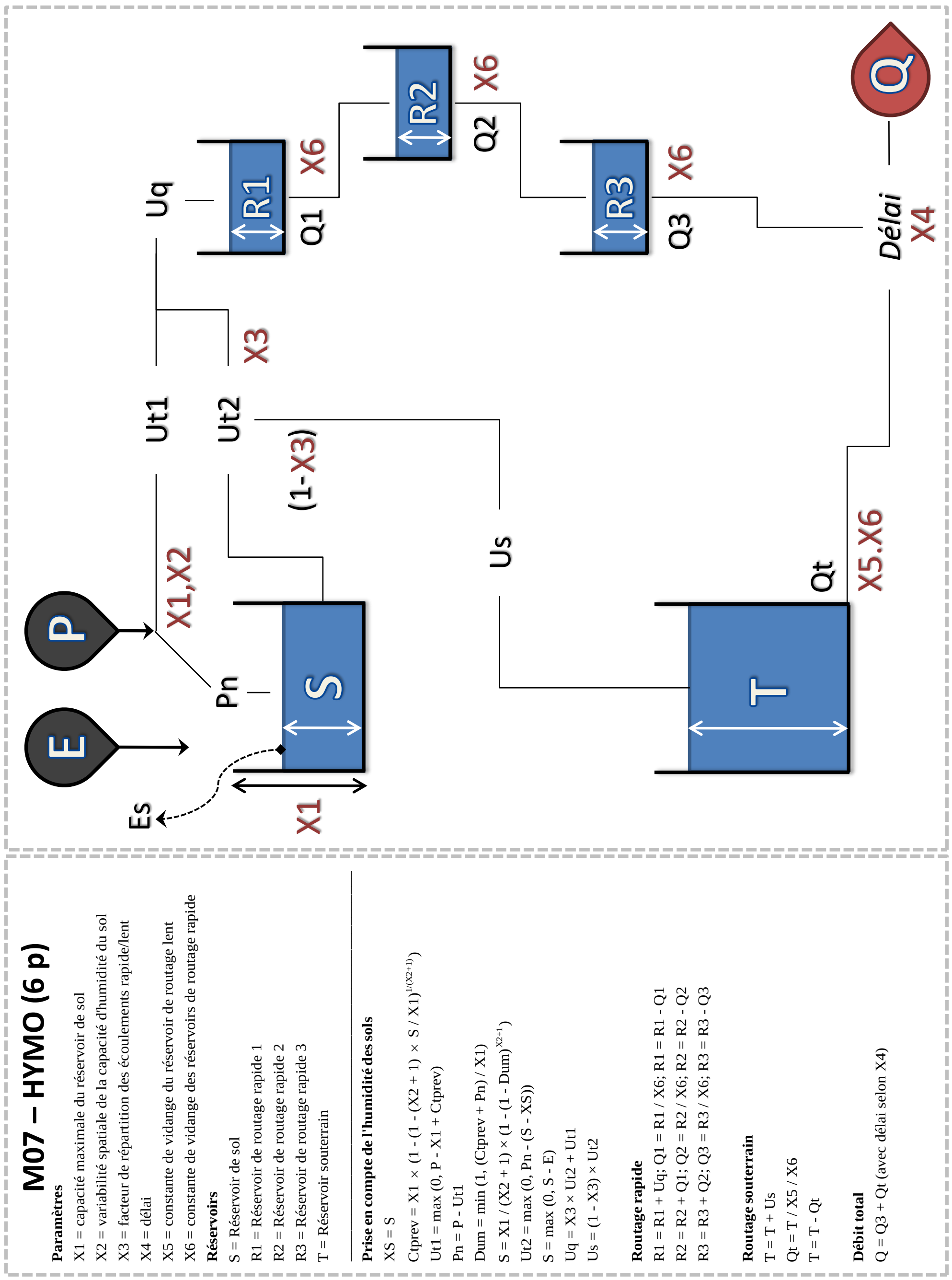

Figure B.8: Structure of model M07 (from Seiller, 2013)

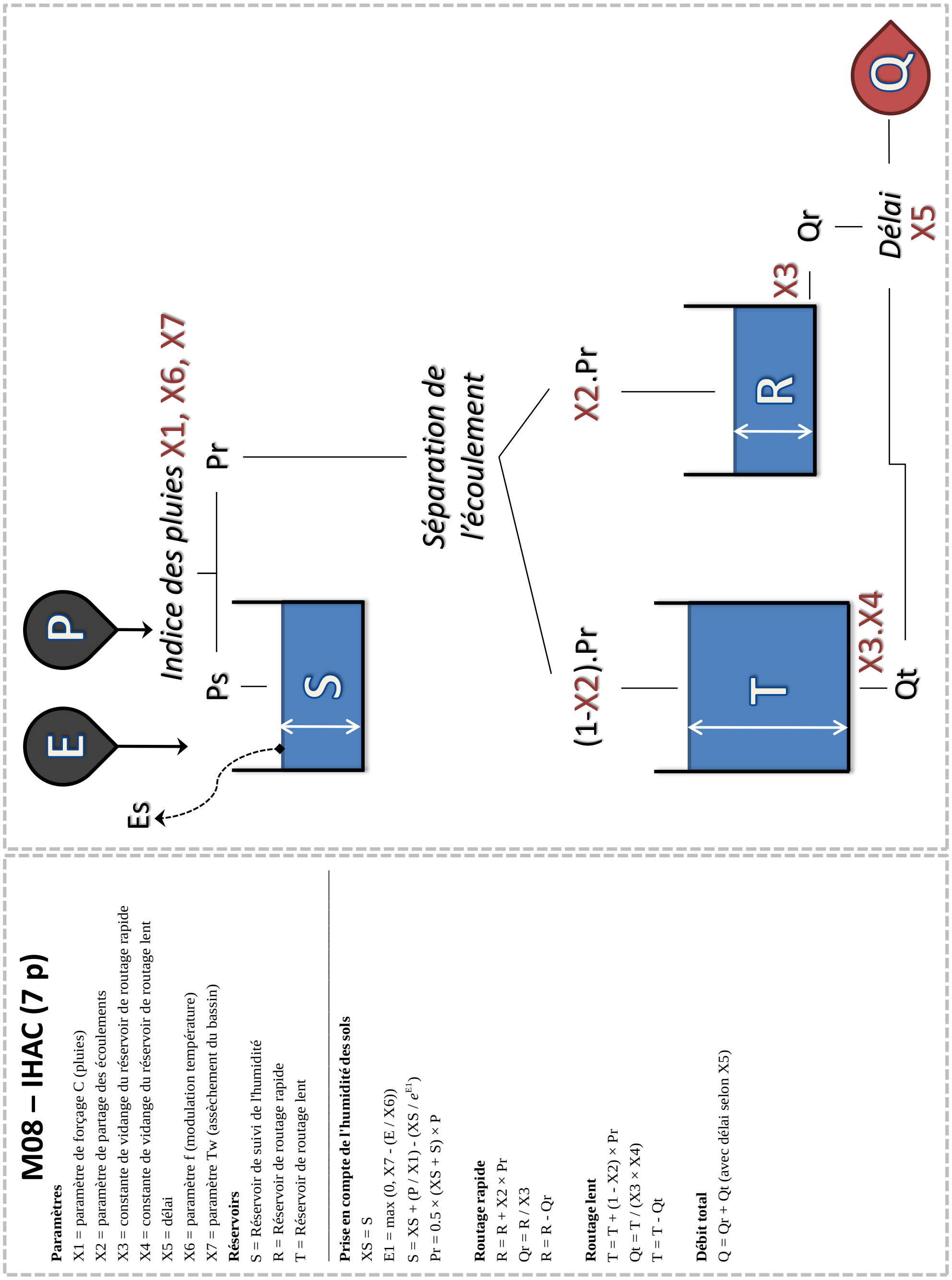

Figure B.9: Structure of model M08 (from Seiller, 2013)

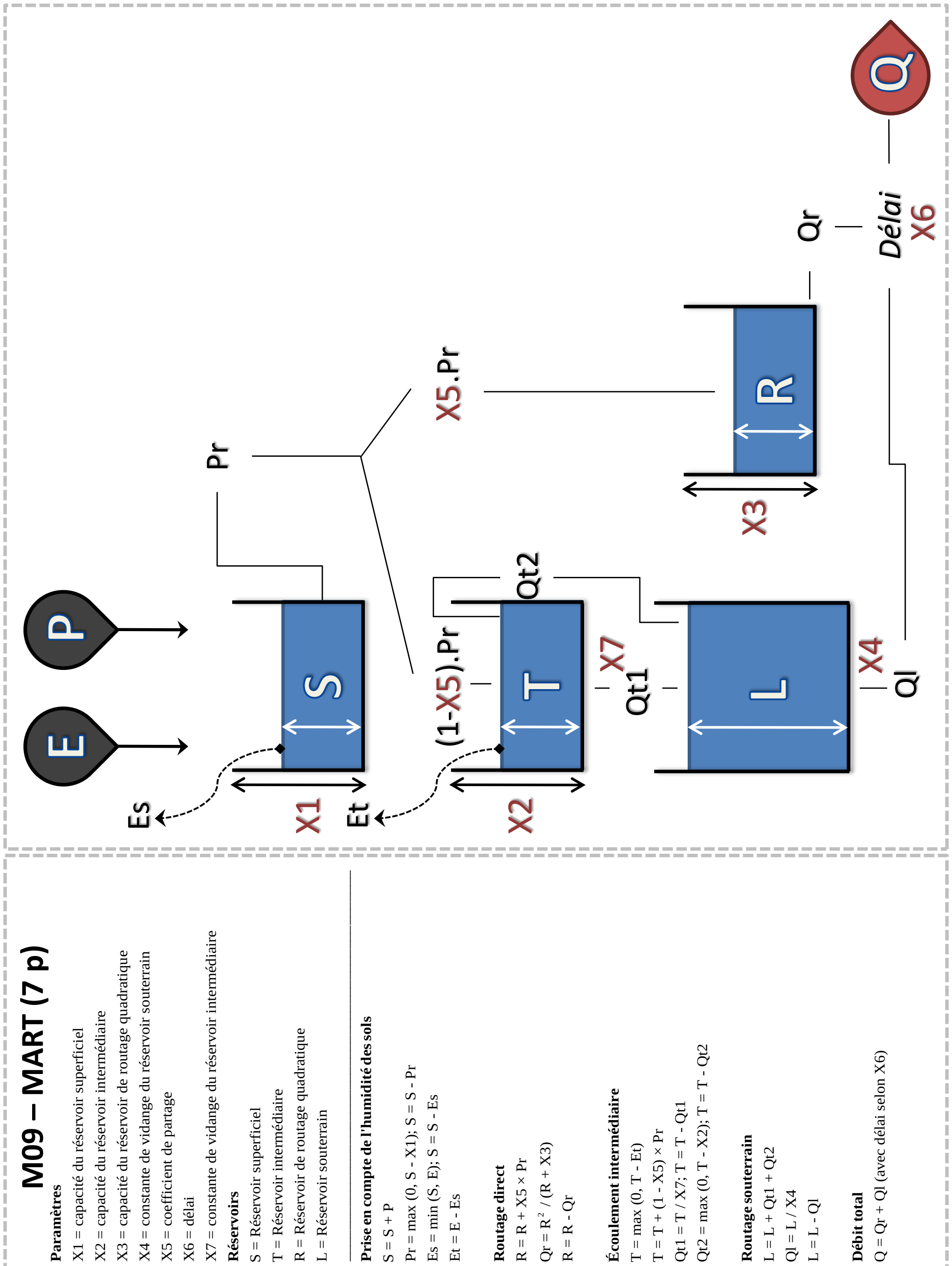

Figure B.10: Structure of model M09 (from Seiller, 2013)

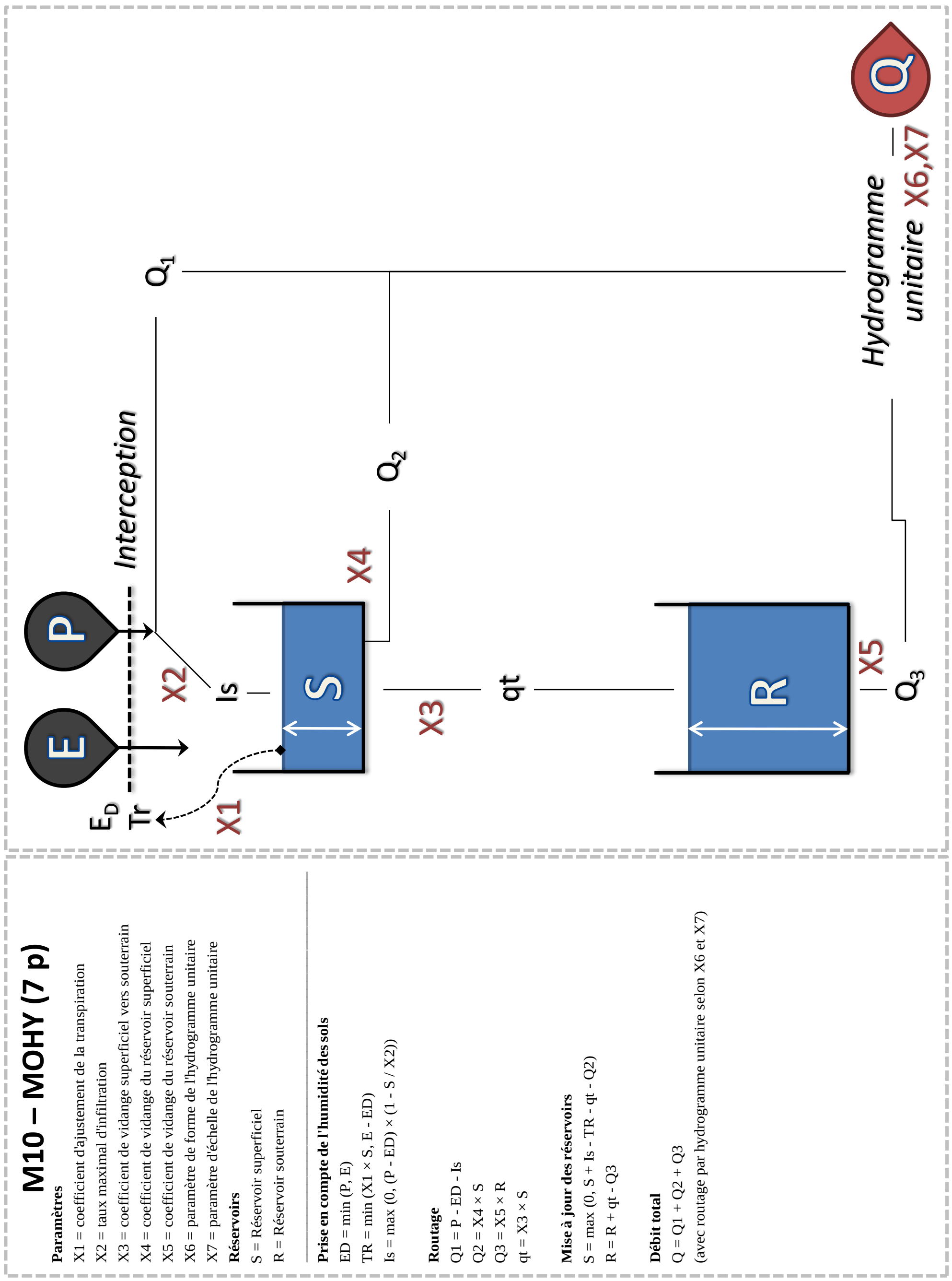

Figure B.11: Structure of model M10 (from Seiller, 2013)

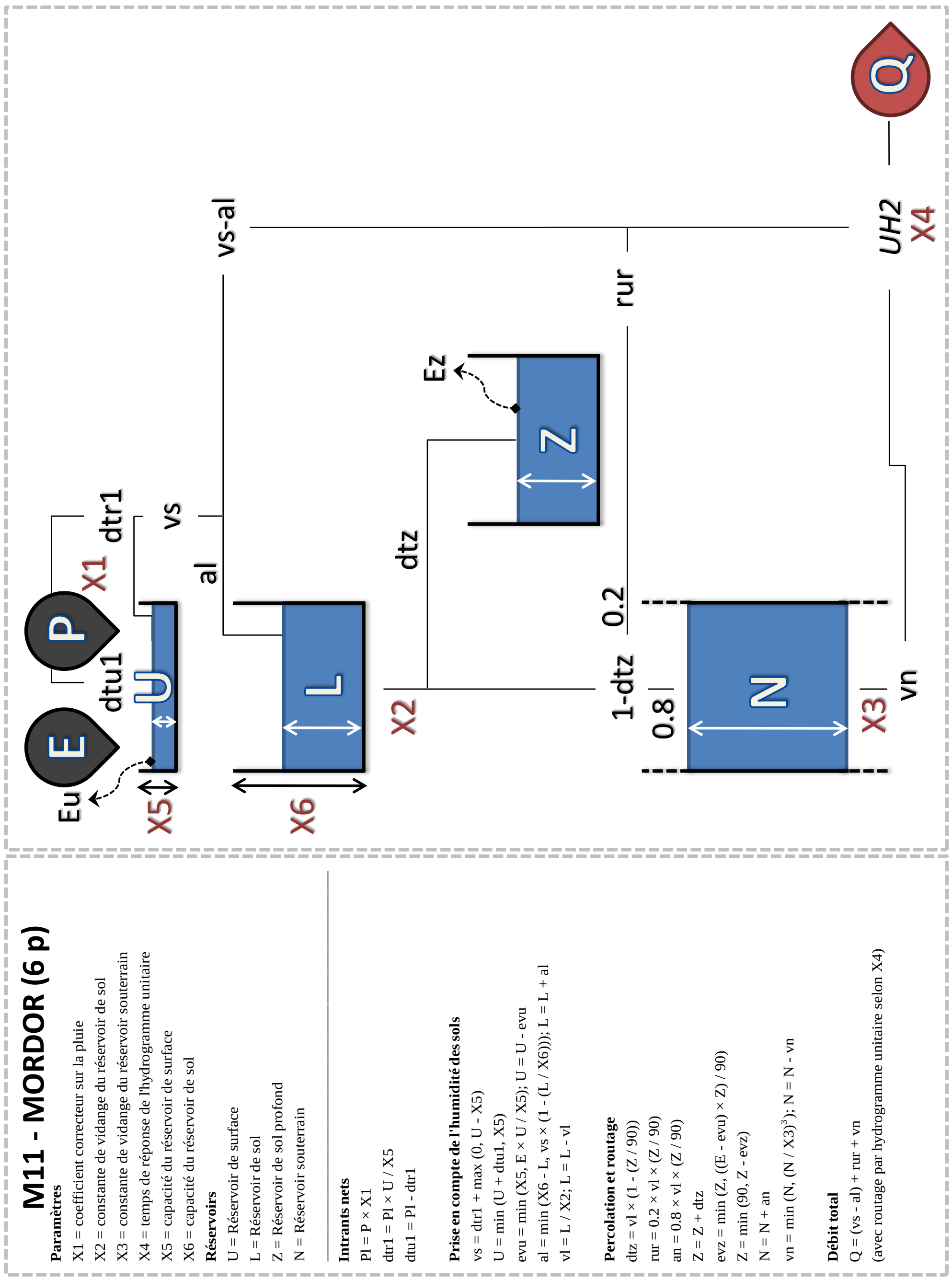

Figure B.12: Structure of model M11 (from Seiller, 2013)

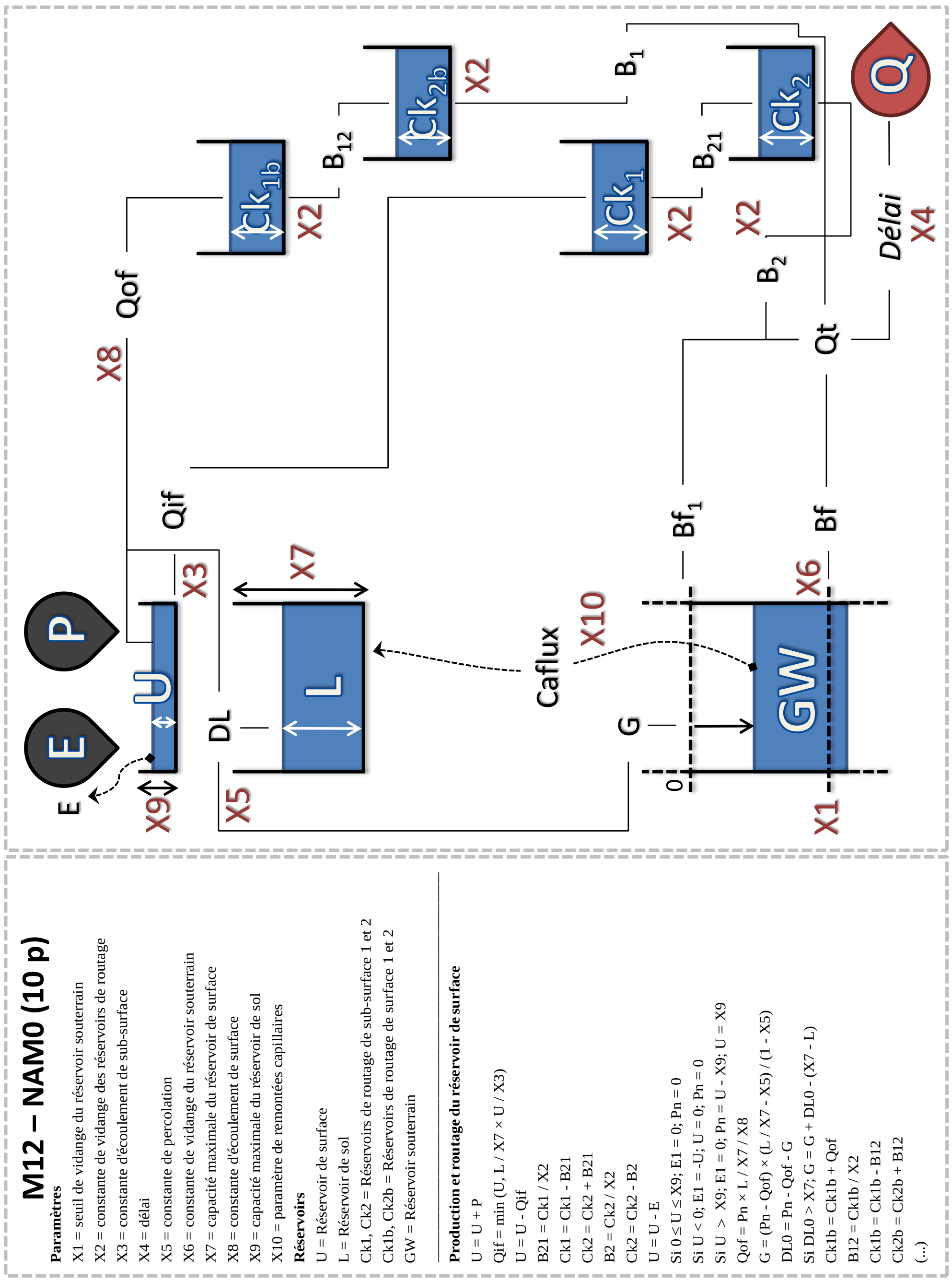

Figure B.13: Structure of model M12 (from Seiller, 2013)

M12 – NAM0 (10 p)

(...)

$B1 = Ck2b / X2$
$Ck2b = Ck2b - B1$

Production et routage du réservoir de sol

$L = L + DL0$
$L = \max(0, L - E1 \times L / X7)$
Si $L > X7$; $L = X7$

Production et routage du réservoir souterrain

$GW = GW - G$
$Bf = \max(0, (X1 - GW) / X6))$
$GW = GW + Bf$
$Bf1 = \max(0, -GW + 0.1)$

Remontées capillaires

$Caflux = (1 - L / X7)^{0.5} \times (X10 / GW)^{2}$
$L = L + Caflux$; $GW = GW + Caflux$

Débit total

$Q = Bf + Bf1 + B1 + B2$ (avec délai selon X4)

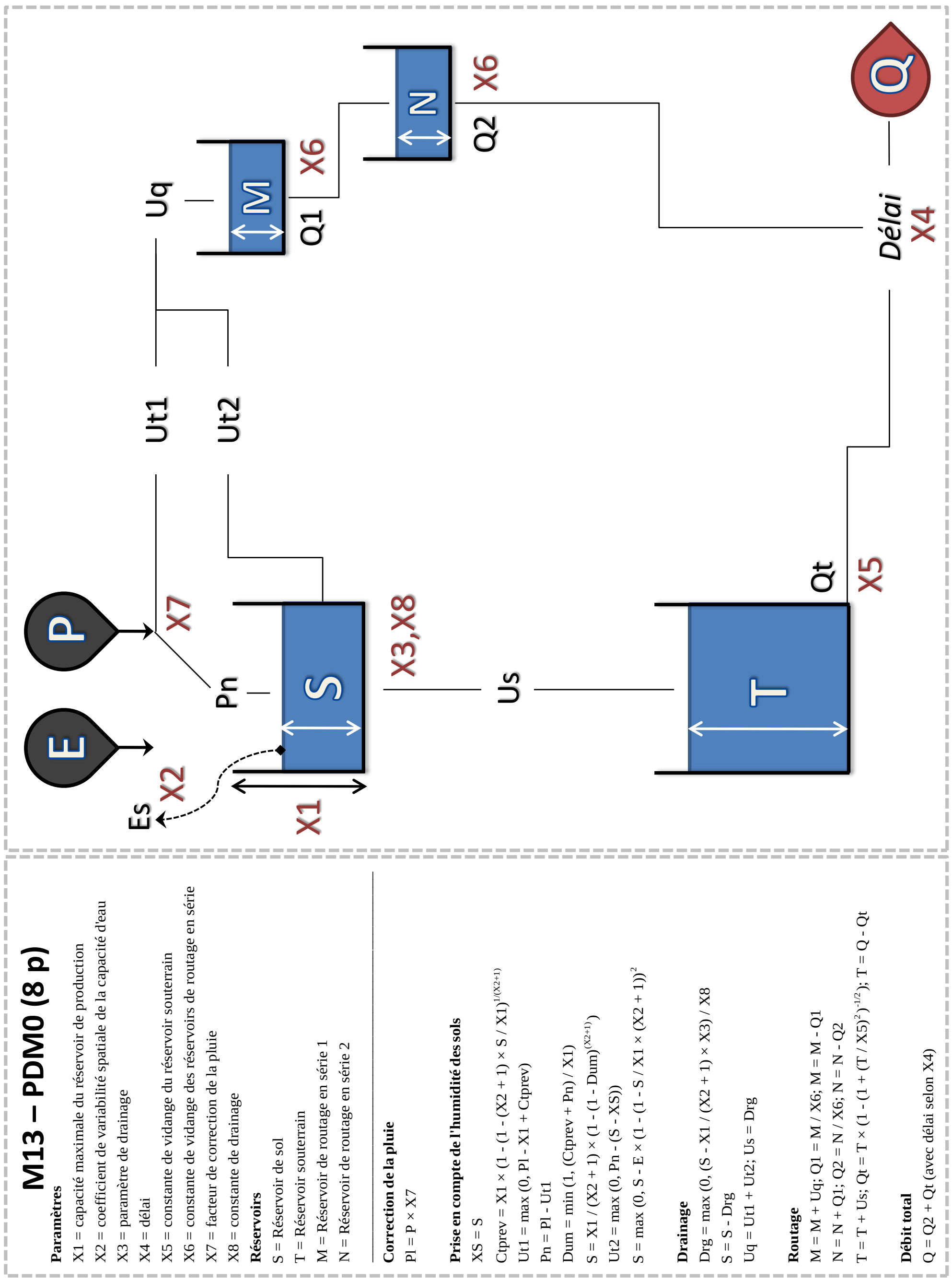

Figure B.14: Structure of model M13 (from Seiller, 2013)

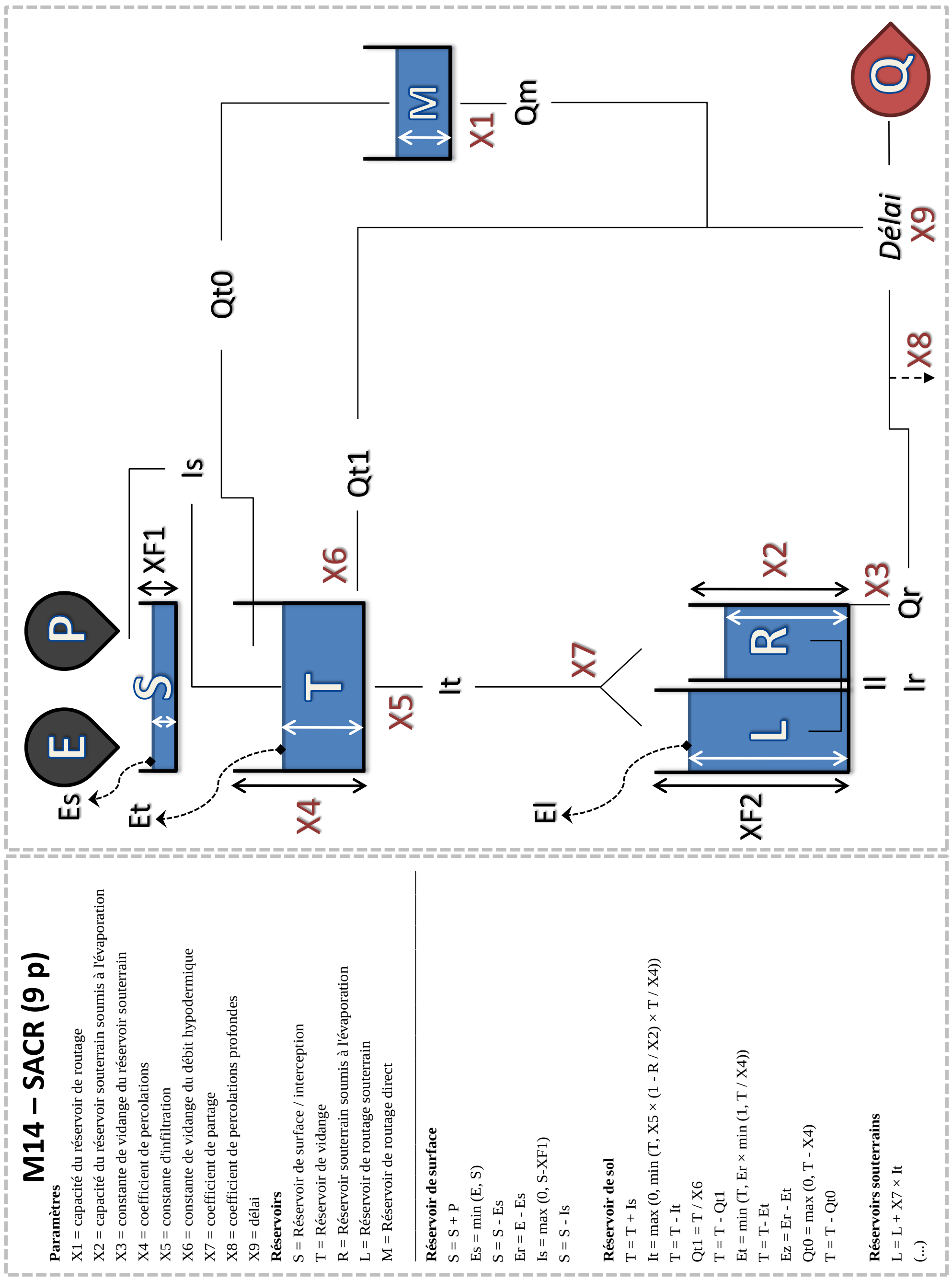

Figure B.15: Structure of model M14 (from Seiller, 2013)

M14 – SACR (9 p)

(...)

```
Il = max (0, L - XF2)
L = L - Il
R = R + (1 - X7) × It + Il
El = (Ez × L) / (XF1 + XF2)
L = L - El
Si L < 0
    Ir = min (-L, max (0, R - (X2 - XF2)))
    L = max (0, L + Ir)
    R = R - Ir
Qr = R / X3
R = R - Qr
Qr = Qr / X8
```

Routage direct

```
M = M + Qt0
Qm = M / X1
M = M - Qm
```

Débit total

Q = Qr + Qm + Qt1 (avec délai selon X9)

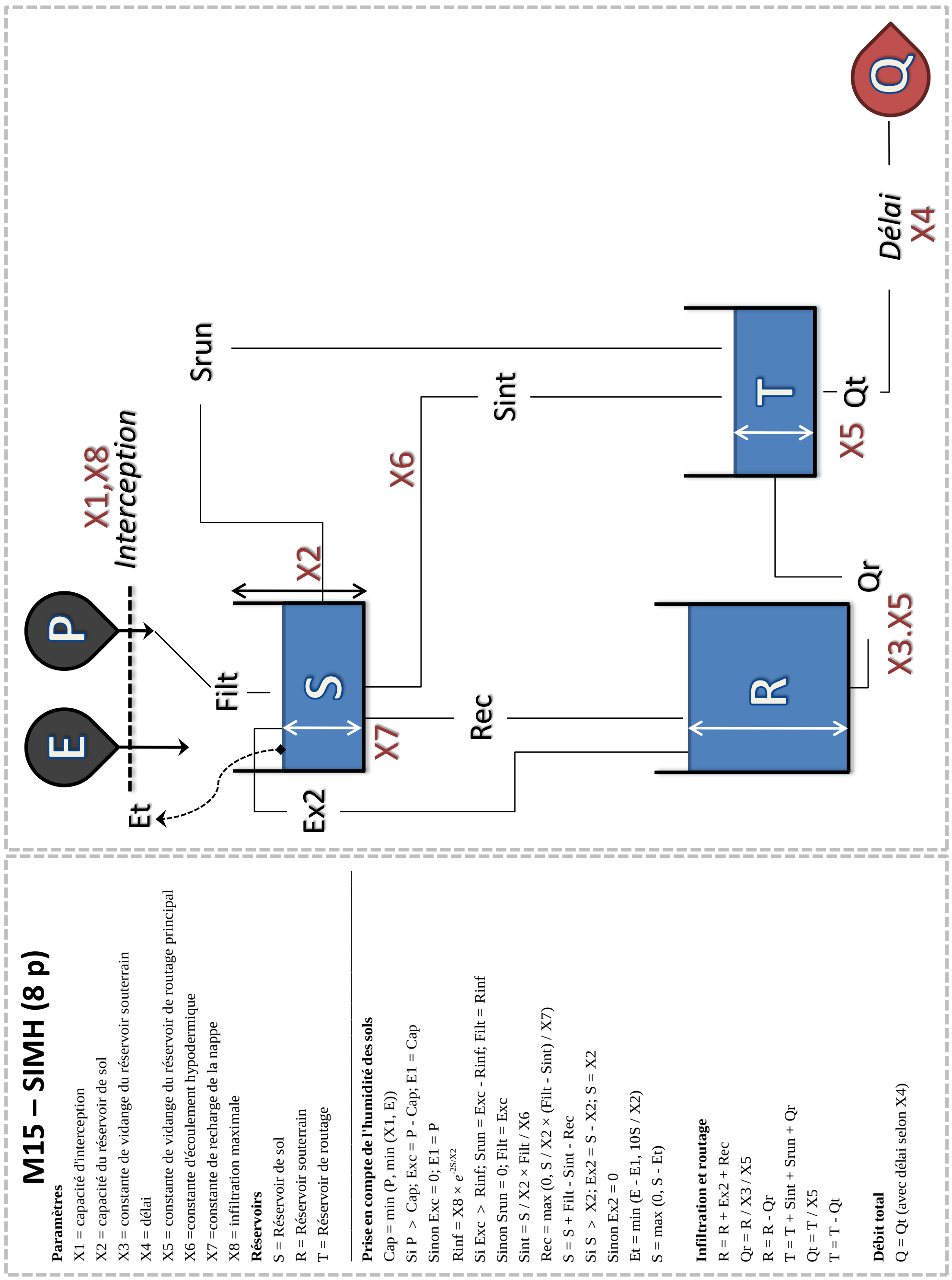

Figure B.16: Structure of model M15 (from Seiller, 2013)

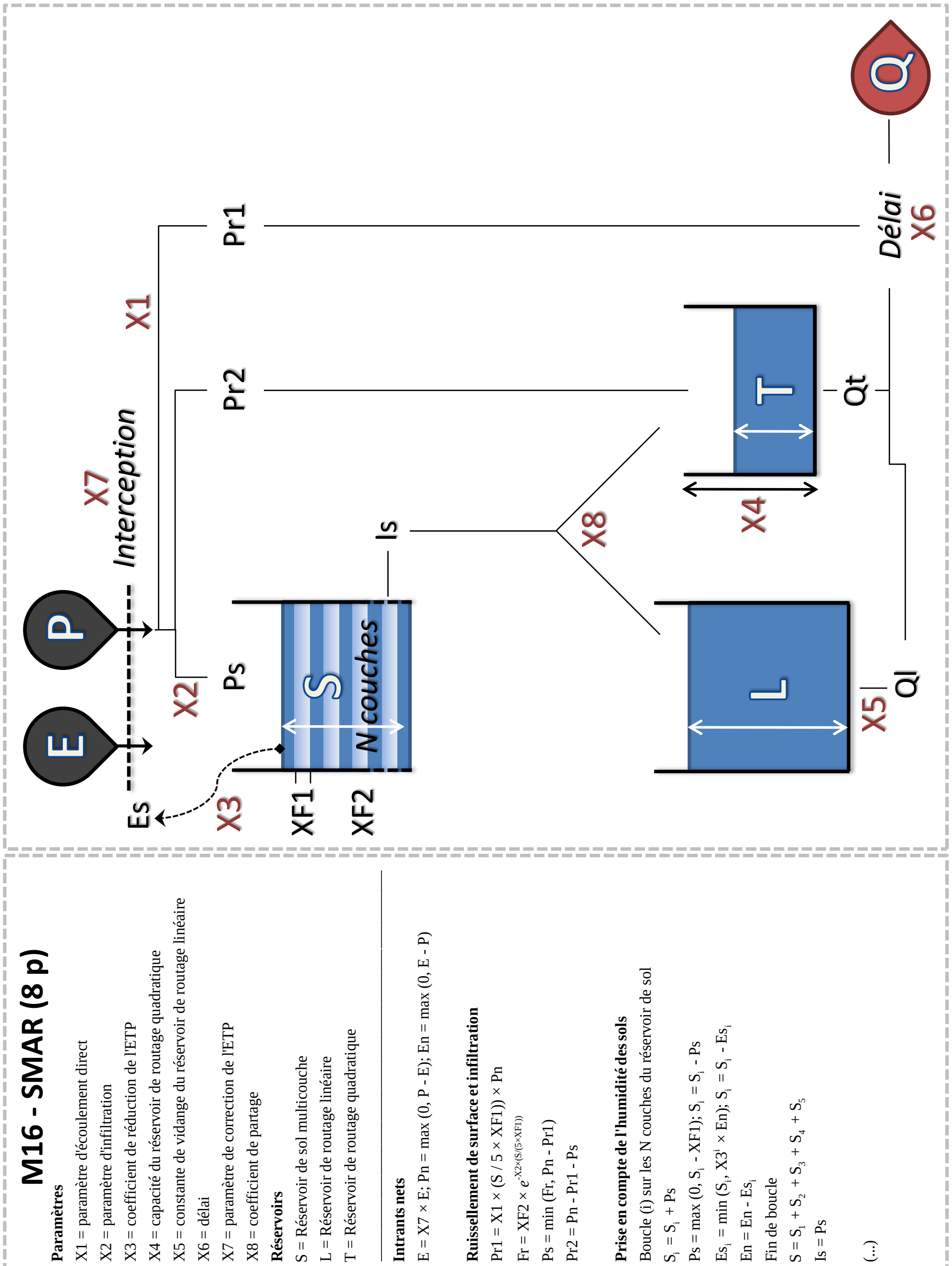

Figure B.17: Structure of model M16 (from Seiller, 2013)

M16 - SMAR (8 p)

(...)

Routage linéaire

$L = L + (1 - X8) \times Is$

$Ql = L / X5$; $L = L - Ql$

Routage quadratique

$T = T + X8 \times Is + Pr2$

$Qt = T^2 / (T + X4)$; $T = T - Qt$

Débit total

$Q = Pr1 + Ql + Qt$ (avec délai selon X6)

(avec N = 16 ; XF1 = 25 mm ; XF2 = 200 mm)

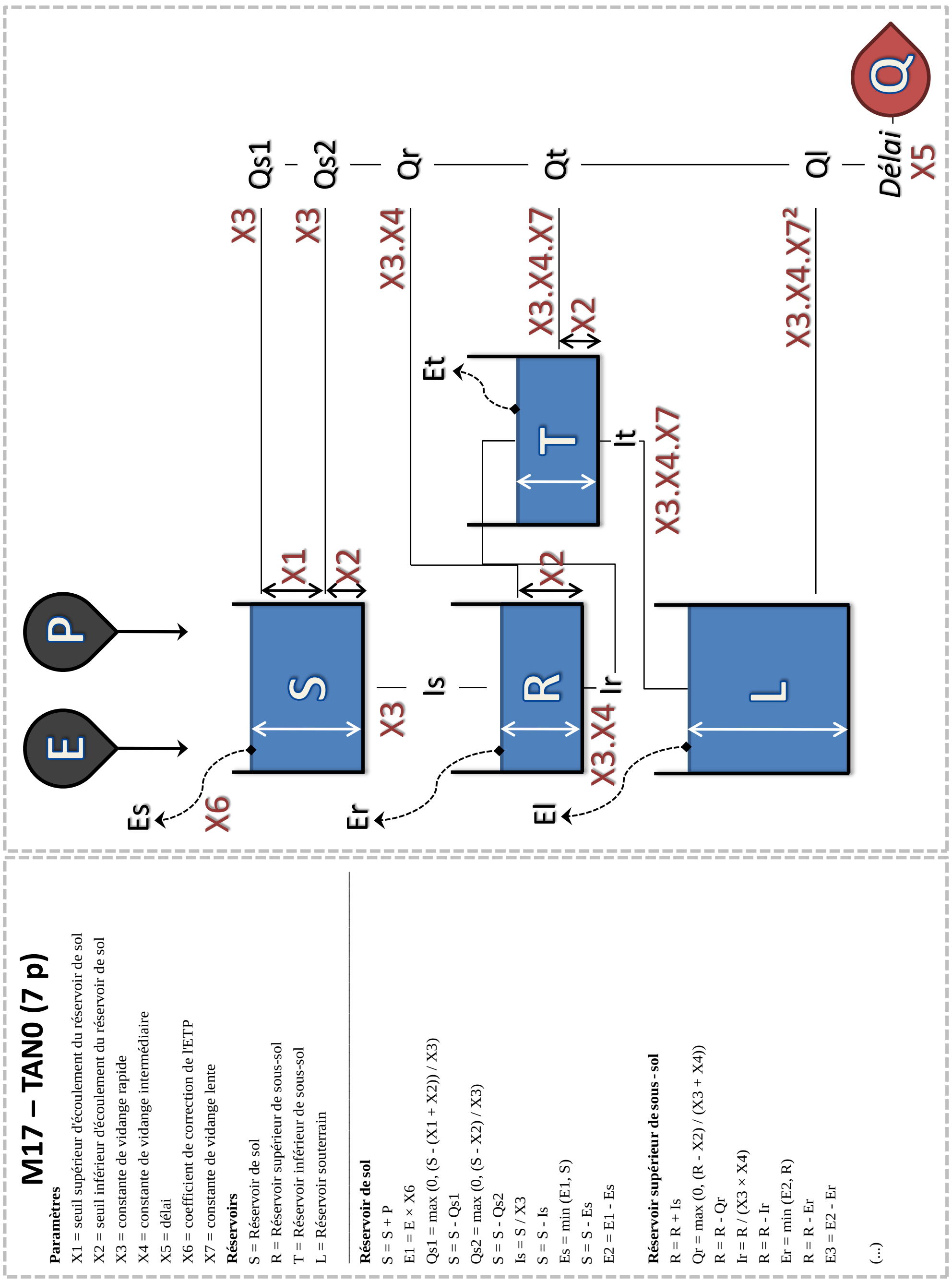

Figure B.18: Structure of model M17 (from Seiller, 2013)

M17 – TAN0 (7 p)

(...)

Réservoir inférieur de sous - sol

T = T + Ir

Qt = max(0, (T - X2) / (X3 × X4 × X7))

T = T - Qt

It = T / (X3 × X4 × X7)

T = T - It

Et = min (E3, T)

T = T - Et

E4 = E3 - Et

Réservoir souterrain

L = L + It

Ql = L / (X3 × X4 × $X7^2$)

L = L - Ql

El = min (E4, L)

L = L - El

Débit total

Q = Qs1 + Qs2 + Qr + Qt + Ql (avec délai selon X5)

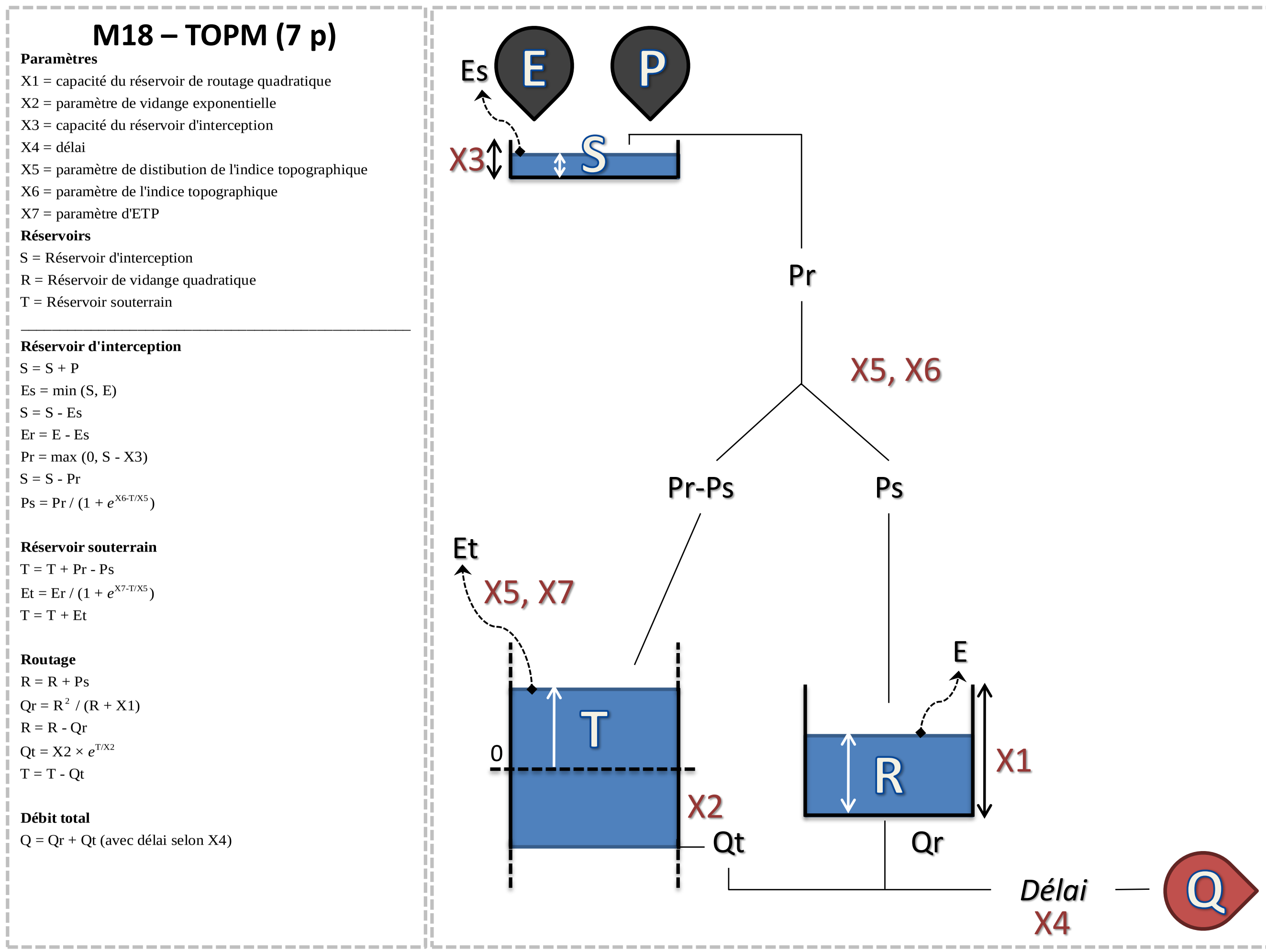

Figure B.19: Structure of model M18 (from Seiller, 2013)

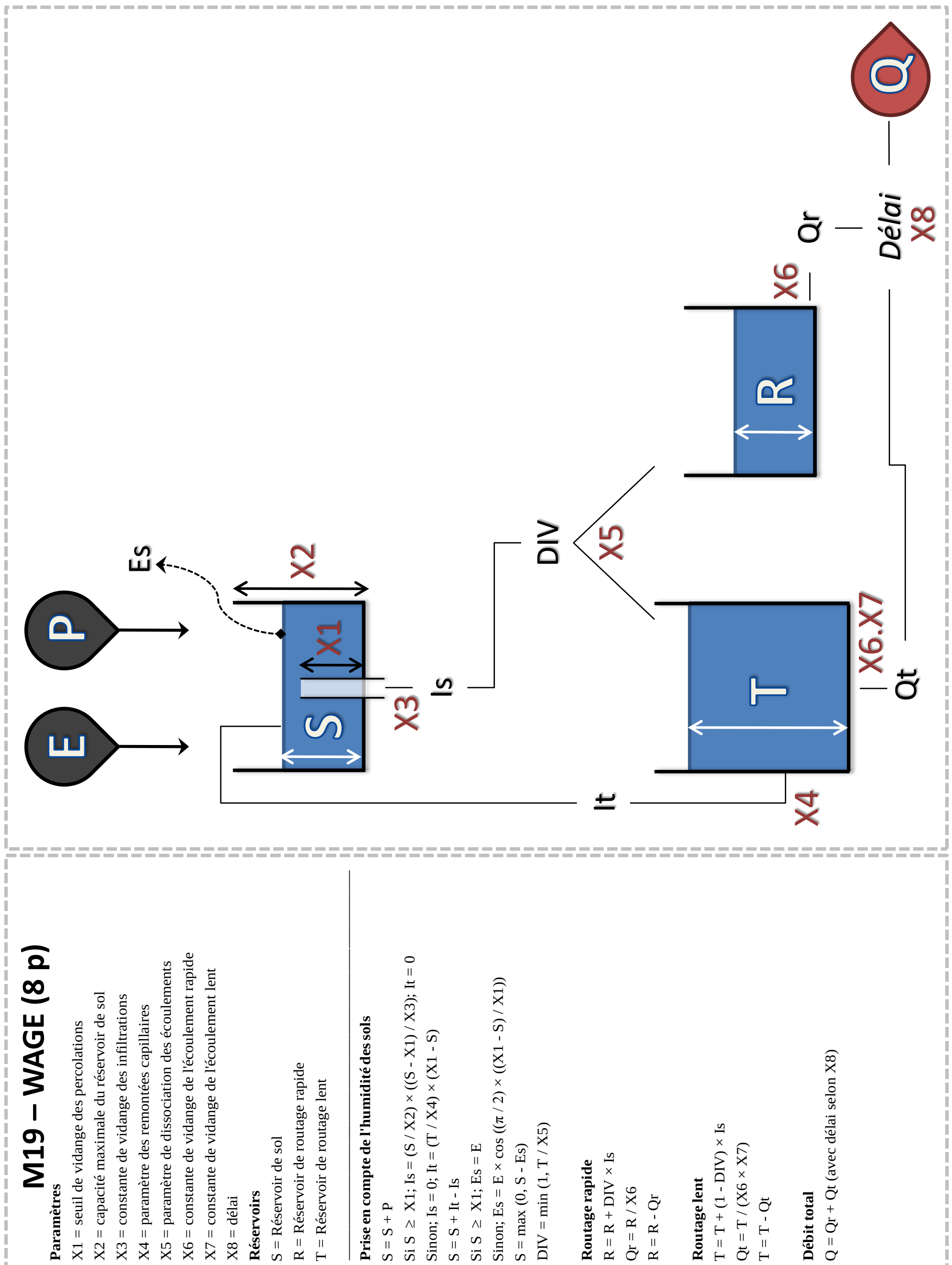

Figure B.20: Structure of model M19 (from Seiller, 2013)

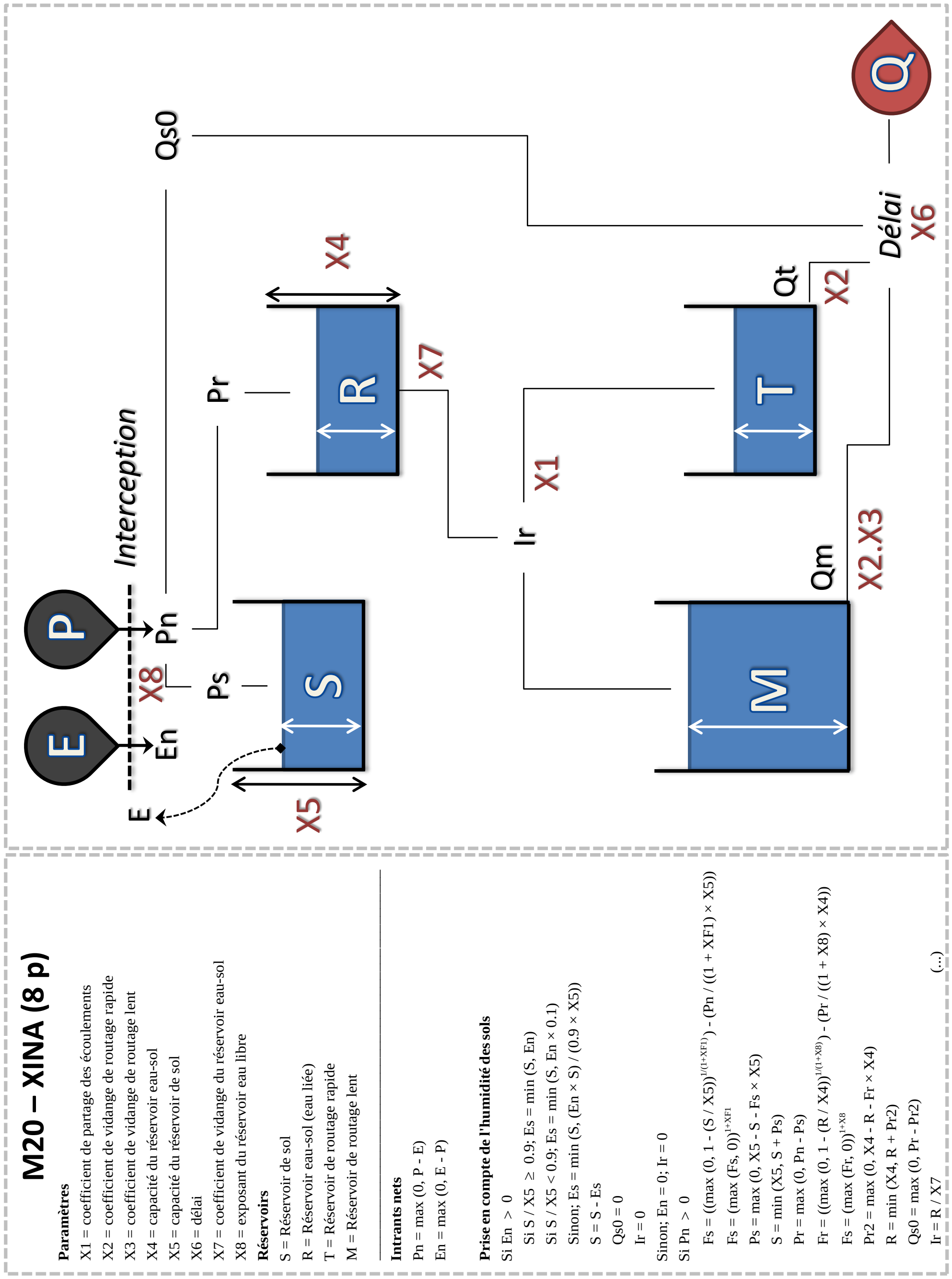

Figure B.21: Structure of model M20 (from Seiller, 2013)

M20 – XINA (8 p)

(...)

Sinon; $P_n = 0$; $I_r = 0$; $P_r = 0$; $P_s = 0$; $Q_{s0} = 0$

Mise à jour et routage

$R = R - I_r$
$T = T + I_r \times X1$
$Q_t = T / X2$
$T = T - Q_t$
$M = M + I_r \times (1 - X1)$
$Q_m = M / (X2 \times X3)$
$M = M - Q_m$

Débit total

$Q = Q_{s0} + Q_t + Q_m$ (avec délai selon X6)
(avec $XF1 = 0.25$)

Table B.1: State variable combinations that are updated for the 5 models presented in Figure 2.6

Possible state variable combinations	Models: M02	M05	M08	M13	M20
1	S	R	R	M	M
2	T	S	S	N	R
3	S,T	S,R	T	S	S
4			R,T	T	T
5			S,R	M,N	R,M
6			S,T	S,M	R,T
7			S,R,T	S,N	S,M
8				S,T	S,R
9				T,M	S,T
10				T,N	T,M
11				S,M,N	R,T,M
12				S,T,M	S,R,M
13				S,T,N	S,R,T
14				T,M,N	S,T,M
15				S,T,M,N	S,R,T,M

Appendix C

Snow module structure

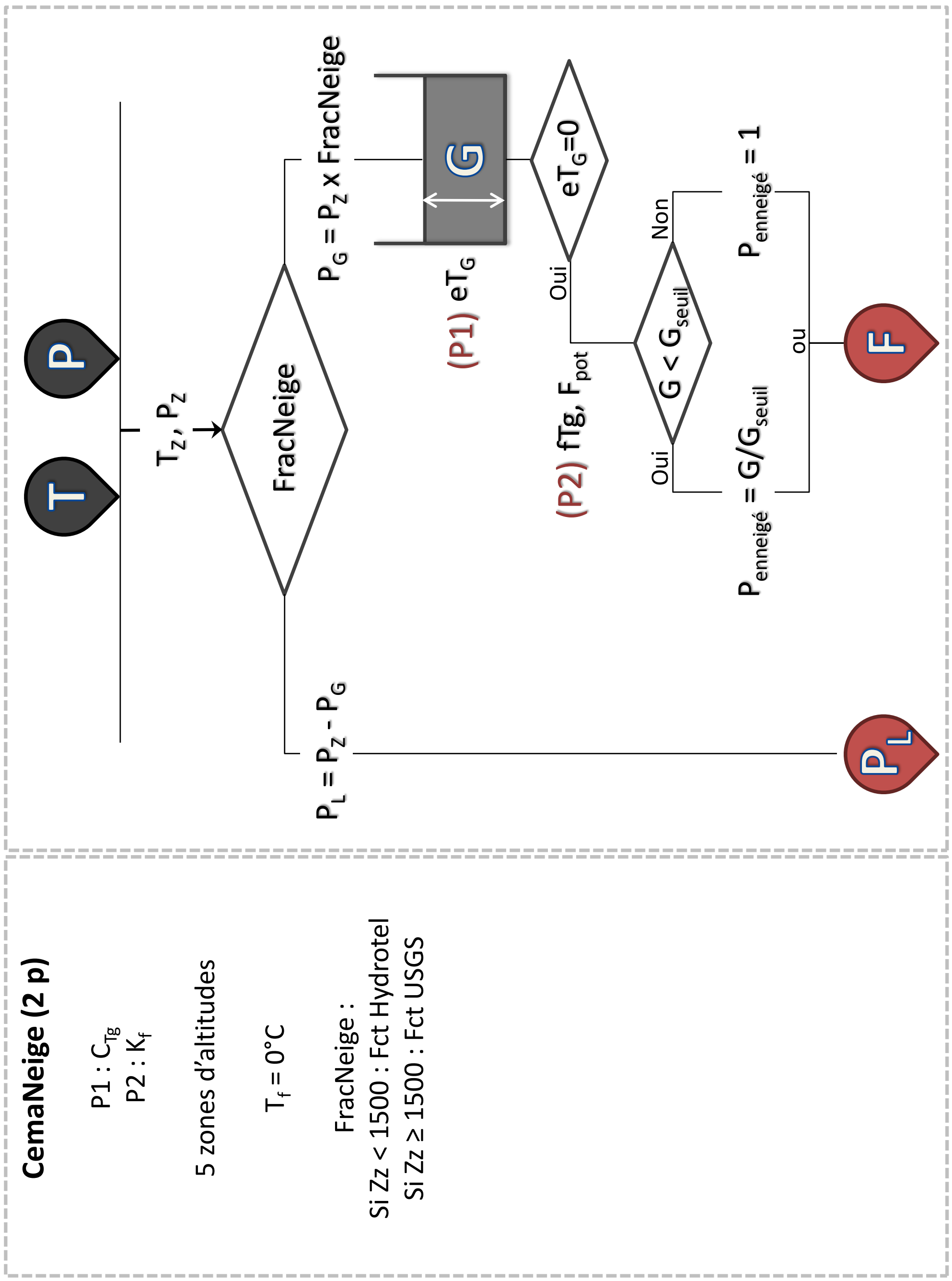

Figure C.1: Structure of Cemaneige (from Seiller, 2013)

Appendix D

Ensemble Kalman Filter updating scheme

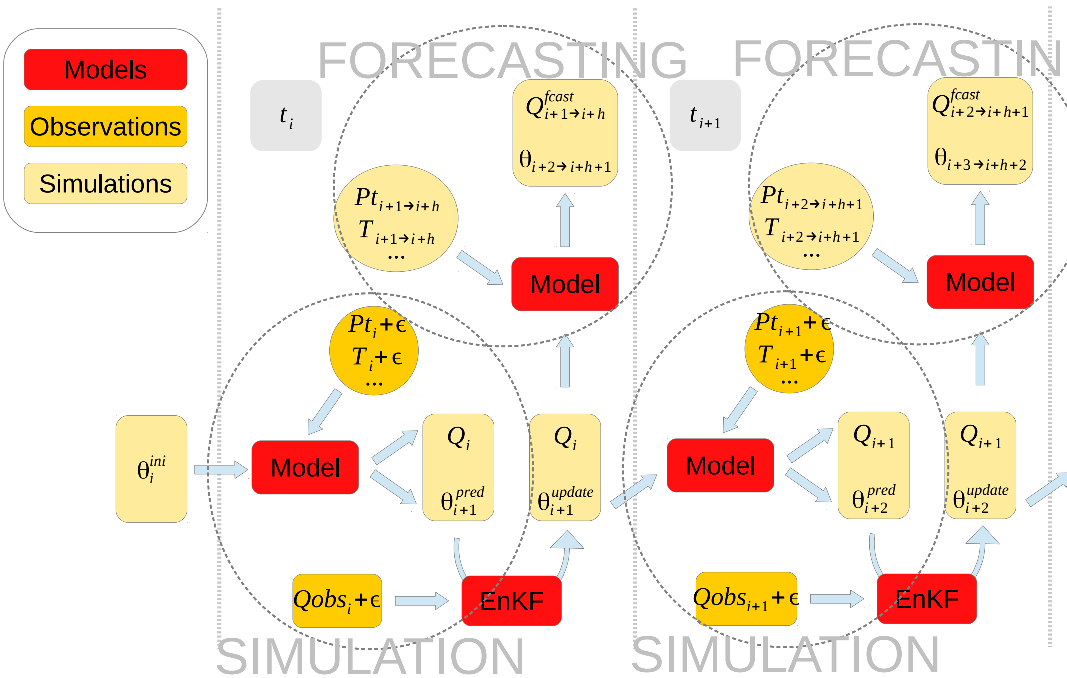

Figure D.1: Schematic representation of state updating with the Ensemble Kalman Filter

Appendix E

Accuracy and reliability of the nine systems

Table E.1: Mean continuous ranked probability score (*MCRPS*) values of the nine systems in average over the 20 catchments

	Day 1	Day 3	Day 6	Day 9
A	0.45	0.48	0.59	0.65
B	0.44	0.43	0.47	0.48
C	0.27	0.29	0.38	0.49
D	0.25	0.27	0.38	0.42
E	0.28	0.30	0.40	0.46
F	0.28	0.30	0.38	0.41
G	0.18	0.20	0.30	0.39
H	0.18	0.20	0.29	0.37
H′	0.17	0.20	0.29	0.37

Table E.2: Deviation from perfect reliability with the mean absolute error of the reliability diagram (*MaeRD*) of the eight systems in average over the 20 catchments

	Day 1	Day 3	Day 6	Day 9
B	0.45	0.53	0.43	0.50
C	0.23	0.27	0.38	0.49
D	0.23	0.26	0.35	0.43
E	0.10	0.11	0.15	0.17
F	0.10	0.11	0.13	0.13
G	0.05	0.05	0.08	0.13
H	0.05	0.05	0.05	0.07
H′	0.05	0.04	0.07	0.09

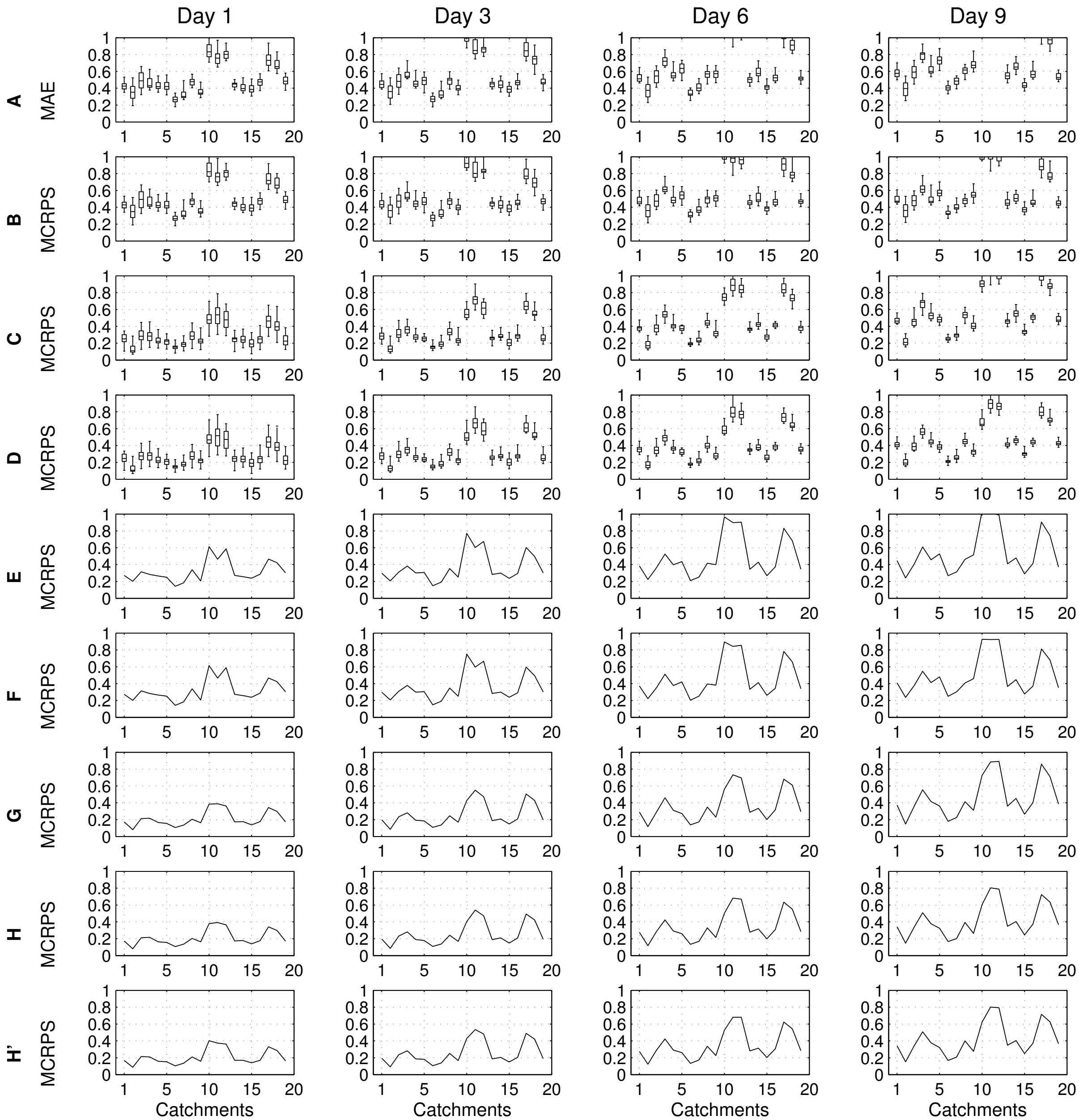

Figure E.1: Continuous Ranked Probability Score (CRPS) and Mean Absolute Error (MAE) of the 9 systems according to the 20 catchments

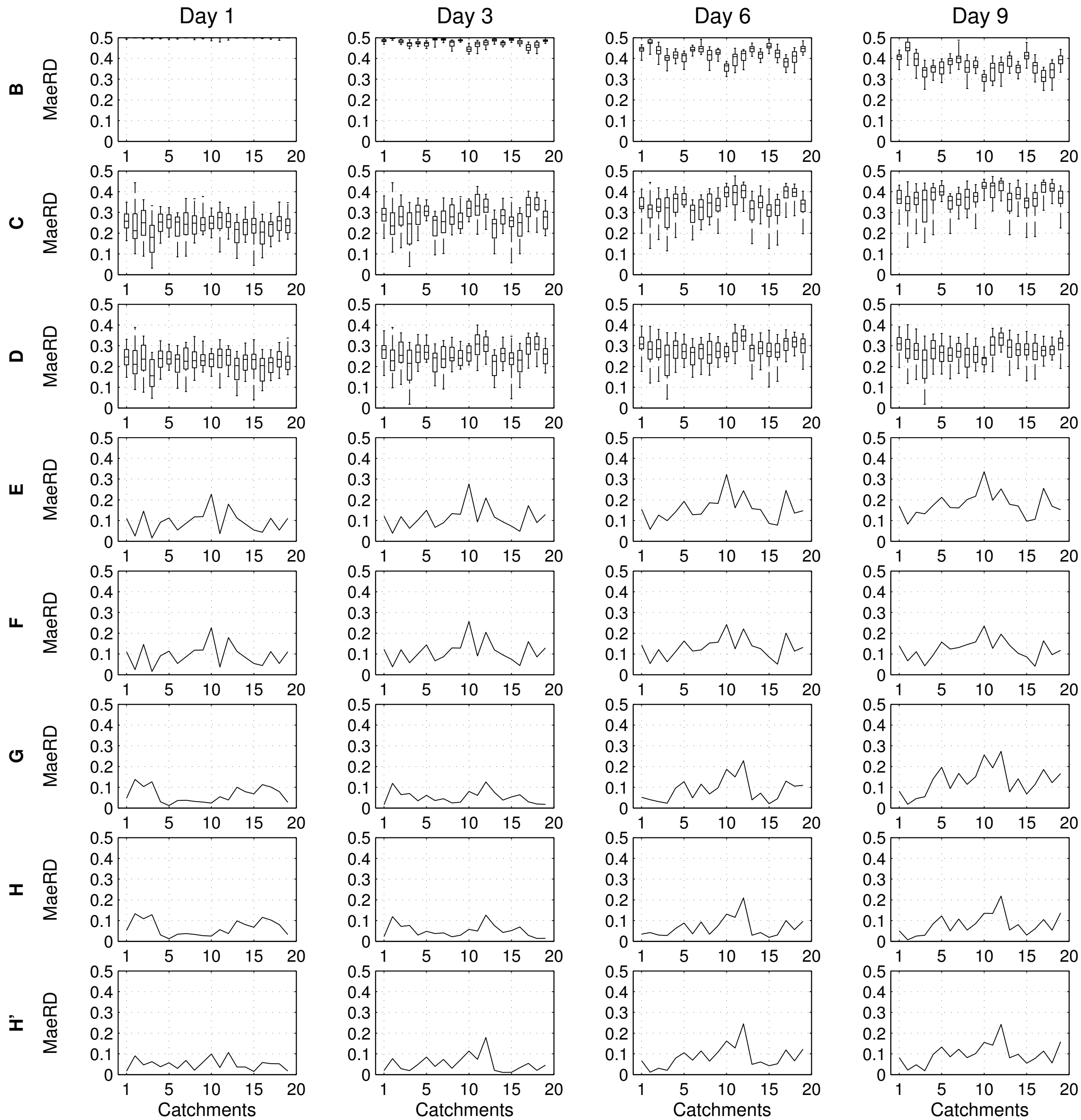

Figure E.2: Mean Absolute Error of the Reliability Diagram (MaeRD) of systems B, C, D, E, F, G, H, and H' according to the 20 catchments

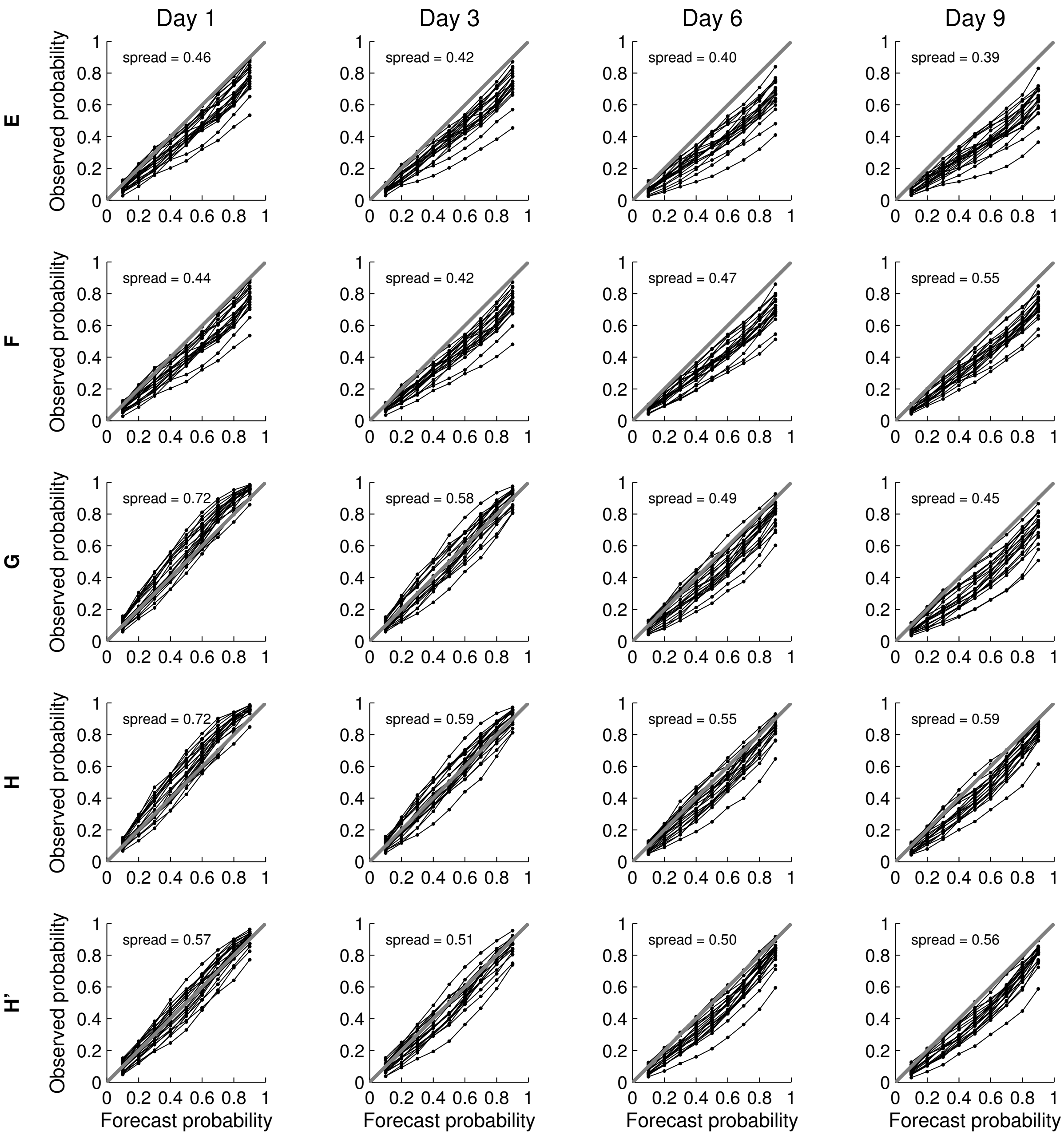

Figure E.3: Reliability Diagram of the systems E, F, G, H, and H′ according to the 20 catchments. The spread represents the square root of mean ensemble variance averaged over all catchments.

Bibliography

Abaza, M., Anctil, F., Fortin, V., Turcotte, R., 2015. Exploration of sequential streamflow assimilation in snow dominated watersheds. Advances in Water Resources 80, 79–89.

Abaza, M., Garneau, C., Anctil, F., 2014. Comparison of sequential and variational streamflow assimilation techniques for short-term hydrological forecasting. Journal of Hydrologic Engineering 20, 1–11.

Abbott, M.B., Bathurst, J.C., Cunge, J.A., Oconnell, P.E., Rasmussen, J., 1986. An introduction to the European Hydrological System. Système Hydrologique Européen, "SHE". History and philosophy of a physically-based distributed modeling system. Journal of Hydrology 87, 45–59.

Ajami, N.K., Duan, Q., Gao, X., Sorooshian, S., 2006. Multimodel combination techniques for analysis of hydrological simulations: Application to Distributed Model Intercomparison Project results. Journal of Hydrometeorology 7, 755–768.

Ajami, N.K., Duan, Q.Y., Sorooshian, S., 2007. An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. Water Resources Research 43, 1–19.

Alvarez-Garreton, C., Ryu, D., Western, A.W., Crow, W.T., Robertson, D.E., 2014. The impacts of assimilating satellite soil moisture into a rainfall-runoff model in a semi-arid catchment. Journal of Hydrology 519, 2763–2774.

Anaman, K.A., Lellyett, S.C., 1996. Assessment of the benefits of an enhanced weather information service for the cotton industry in Australia. Meteorological Applications 3, 127–135.

Andreadis, K.M., Lettenmaier, D.P., 2006. Assimilating remotely sensed snow observations into a macroscale hydrology model. Advances in Water Resources 29, 872–886.

Andreassian, V., Perrin, C., Berthet, L., Le Moine, N., Lerat, J., Loumagne, C., Oudin, L., Mathevet, T., Ramos, M.H., Valery, A., 2009. HESS Opinions 'Crash tests for a standardized evaluation of hydrological models'. Hydrology and Earth System Sciences 13, 1757–1764.

Bailey, R.T., Bau, D., 2012. Estimating geostatistical parameters and spatially-variable hydraulic conductivity within a catchment system using an ensemble smoother. Hydrology and Earth System Sciences 16, 287–304.

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

Bavarian State Ministry for the Environment, 2006. River basin agenda for alpine space. Report. Bavarian State Ministry for the Environment.

Bergström, S., Forsman, A., 1973. Development of a conceptual deterministic rainfall-runoff model. Nordic Hydrology 4, 147–170.

Beven, K., Binley, A., 1992. The future of distributed models - Model calibration and uncertainty prediction. Hydrological Processes 6, 279–298.

Beven, K., Binley, A., 2014. GLUE: 20 years on. Hydrological Processes 28, 5897–5918.

Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology 249, 11–29.

Beven, K.J., Kirkby, M.J., Schofield, N., Tagg, A.F., 1984. Testing a physically-based flood forecasting model (TOPMODEL) for 3 UK catchments. Journal of Hydrology 69, 119–143.

Beven, K.J., Smith, P.J., Freer, J.E., 2008. So just why would a modeler choose to be incoherent? Journal of Hydrology 354, 15–32.

Block, P.J., Souza, F.A., Sun, L.Q., Kwon, H.H., 2009. A streamflow forecasting framework using multiple climate and hydrological models. Journal of the American Water Resources Association 45, 828–843.

Boe, J., Terray, L., Habets, F., Martin, E., 2006. A simple statistical-dynamical downscaling scheme based on weather types and conditional resampling. Journal of Geophysical Research 111, 1–20.

Bormann, H., Breuer, L., Giertz, S., Huisman, J.A., Viney, N.R., 2009. Spatially explicit versus lumped models in catchment hydrology - experiences from two case studies. pp. 3–26.

Boucher, M.A., Perreault, L., Anctil, F., Favre, A.C., 2015. Exploratory analysis of statistical post-processing methods for hydrological ensemble forecasts. Hydrological Processes 29, 1141–1155.

Boucher, M.A., Tremblay, D., Delorme, L., Perreault, L., Anctil, F., 2012. Hydro-economic assessment of hydrological forecasting systems. Journal of Hydrology 416, 133–144.

Bourdin, D.R., Fleming, S.W., Stull, R.B., 2012. Streamflow modelling: A primer on applications, approaches and challenges. Atmosphere-Ocean 50, 507–536.

Bourke, W., Buizza, R., Naughton, M., 2004. Performance of the ECMWF and the BoM ensemble prediction systems in the Southern Hemisphere. Monthly Weather Review 132, 2338–2357.

Bouttier, F., 1994. A dynamical estimation of forecast error covariances in an assimilation system. Monthly Weather Review 122, 2376–2390.

Breuer, L., Huisman, J.A., Willems, P., Bormann, H., Bronstert, A., Croke, B.F.W., Frede, H.G., Graff, T., Hubrechts, L., Jakeman, A.J., Kite, G., Lanini, J., Leavesley, G., Lettenmaier, D.P., Lindstrom, G., Seibert, J., Sivapalan, M., Viney, N.R., 2009. Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM). I: Model intercomparison with current land use. Advances in Water Resources 32, 129–146.

Brochero, D., Anctil, F., Gagne, C., 2011a. Simplifying a hydrological ensemble prediction system with a backward greedy selection of members - Part 1: Optimization criteria. Hydrology and Earth System Sciences 15, 3307–3325.

Brochero, D., Anctil, F., Gagne, C., 2011b. Simplifying a hydrological ensemble prediction system with a backward greedy selection of members - Part 2: Generalization in time and space. Hydrology and Earth System Sciences 15, 3327–3341.

Brochero, D., Gagne, C., Anctil, F., 2013. Evolutionary multiobjective optimization for selecting members of an ensemble streamflow forecasting model, in: Gecco'13: 15th annual conference on Genetic and Evolutionary Computation, pp. 1221–1228.

Buizza, R., 2008. The value of probabilistic prediction. Atmospheric Science Letters 9, 39–42.

Buizza, R., Bidlot, J.R., Wedi, N., Fuentes, M., Hamrud, M., Holt, G., Vitart, F., 2007. The new ECMWF VAREPS (Variable Resolution Ensemble Prediction System). Quarterly Journal of the Royal Meteorological Society 133, 681–695.

Buizza, R., Miller, M., Palmer, N., 1999. Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. Quarterly Journal of the Royal Meteorological Society 125, 2887–2908.

Burgers, G., van Leeuwen, P.J., Evensen, G., 1998. Analysis scheme in the ensemble Kalman filter. Monthly Weather Review 126, 1719–1724.

Burnash, R.J.C., Ferral, R.L., McGuire, R.A., 1973. A generalized streamflow simulation system - Conceptual modelling for digital computers.

Charron, M., Pellerin, G., Spacek, L., Houtekamer, P.L., Gagnon, N., Mitchell, H.L., Michelin, L., 2010. Toward random sampling of model error in the Canadian ensemble prediction system. Monthly Weather Review 138, 1877–1901.

Chen, H., Yang, D.W., Hong, Y., Gourley, J.J., Zhang, Y., 2013. Hydrological data assimilation with the Ensemble Square-Root-Filter: Use of streamflow observations to update model states for real-time flash flood forecasting. Advances in Water Resources 59, 209–220.

Chiew, F.H.S., Peel, M.C., Western, A.W., 2002. Application and testing of the simple rainfall-runoff model SIMHYD. pp. 335–367.

Choo, C.W., 2009. Information use and early warning effectiveness: Perspectives and prospects. Journal of the American Society for Information Science and Technology 60, 1071–1082.

Christensen, N.S., Lettenmaier, D.P., 2007. A multimodel ensemble approach to assessment of climate change impacts on the hydrology and water resources of the Colorado River Basin. Hydrology and Earth System Sciences 11, 1417–1434.

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

Clark, M.P., Slater, A.G., Rupp, D.E., Woods, R.A., Vrugt, J.A., Gupta, H.V., Wagener, T., Hay, L.E., 2008b. Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. Water Resources Research 44, 1–14.

Clemen, R.T., 1989. Combining forecasts - A review and annotated bibliography. International Journal of Forecasting 5, 559–583.

Cloke, H.L., Pappenberger, F., 2009. Ensemble flood forecasting: A review. Journal of Hydrology 375, 613–626.

Cormary, Y., Guilbot, A., 1973. Étude des relations pluie-débit sur trois bassins versants d'investigation, in: publications, I. (Ed.), Design of water resources projects with inadequate data: IAHS Madrid Symposium, pp. 265–279.

Davolio, S., Miglietta, M.M., Diomede, T., Marsigli, C., Morgillo, A., Moscatello, A., 2008. A meteo-hydrological prediction system based on a multi-model approach for precipitation forecasting. Natural Hazards and Earth System Sciences 8, 143–159.

DeBeer, C.M., Pomeroy, J.W., 2009. Modelling snow melt and snowcover depletion in a small alpine cirque, Canadian Rocky Mountains. Hydrological Processes 23, 2584–2599.

DeChant, C.M., Moradkhani, H., 2011. Improving the characterization of initial condition for ensemble streamflow prediction using data assimilation. Hydrology and Earth System Sciences 15, 3399–3410.

DeChant, C.M., Moradkhani, H., 2012. Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. Water Resources Research 48, 1–15.

Deidda, R., 2000. Rainfall downscaling in a space-time multifractal framework. Water Resources Research 36, 1779–1794.

Demargne, J., Wu, L.M., Regonda, S.K., Brown, J.D., Lee, H., He, M.X., Seo, D.J., Hartman, R., Herr, H.D., Fresch, M., Schaake, J., Zhu, Y.J., 2014. The science of NOAA's operational hydrologic ensemble forecast service. Bulletin of the American Meteorological Society 95, 79–98.

Demeritt, D., Nobert, S., Cloke, H.L., Pappenberger, F., 2013. The European Flood Alert System and the communication, perception, and use of ensemble predictions for operational flood risk management. Hydrological Processes 27, 147–157.

Dietrich, J., Schumann, A.H., Redetzky, M., Walther, J., Denhard, M., Wang, Y., Pfuetzner, B., Buettner, U., 2009. Assessing uncertainties in flood forecasts for decision making: prototype of an operational flood management system integrating ensemble predictions. Natural Hazards and Earth System Sciences 9, 1529–1540.

Doblas-Reyes, F.J., Deque, M., Piedelievre, J.P., 2000. Multi-model spread and probabilistic seasonal forecasts in PROVOST. Quarterly Journal of the Royal Meteorological Society 126, 2069–2087.

Duan, Q.Y., Ajami, N.K., Gao, X.G., Sorooshian, S., 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging. Advances in Water Resources 30, 1371–1386.

Duan, Q.Y., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resources Research 28, 1015–1031.

Eckert, N., 2002. Prise en compte des couverts neigeux temporaires au sein d'un modèle pluie débit. Thesis. Université Pierre et Marie Curie, Université Paris-Sud, École nationale des Mines de Paris et École Nationale du Génie Rural des Eaux et des Forêts.

Elkadi, A.I., 1989. Watershed models and their applicability to conjunctive use management. Water Resources Bulletin 25, 125–137.

Evensen, G., 1992. Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. Journal of Geophysical Research-Oceans 97, 17905–17924.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. Journal of Geophysical Research-Oceans 99, 10143–10162.

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

Ferguson, R.I., 1999. Snowmelt runoff models. Progress in Physical Geography 23, 205–227.

Feyen, L., Kalas, M., Vrugt, J.A., 2008. Semi-distributed parameter optimization and uncertainty assessment for large-scale streamflow simulation using global optimization. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques 53, 293–308.

Forman, B.A., Reichle, R.H., Rodell, M., 2012. Assimilation of terrestrial water storage from GRACE in a snow-dominated basin. Water Resources Research 48, 1–14.

Fortin, J., Moussa, R., Bocquillon, C., Villeneuve, J., 1995. Hydrotel, un modèle hydrologique distribué pouvant bénéficier des données fournies par la télédétection et des systèmes d'information géographique. Journal of Water Sciences / Revue des sciences de l'eau 8, 97–124.

Fortin, V., Abaza, M., Anctil, F., Turcotte, R., 2014. Why should ensemble spread match the RMSE of the ensemble mean? Journal of Hydrometeorology 15, 1708 – 1713.

Fortin, V., Favre, A.C., Said, M., 2006. Probabilistic forecasting from ensemble prediction systems: Improving upon the best-member method by using a different weight and dressing kernel for each member. Quarterly Journal of the Royal Meteorological Society 132, 1349–1369.

Fortin, V., Turcotte, R., 2007. Le modèle hydrologique MOHYSE, Note de cours pour SCA7420. Report. Département des sciences de la terre et de l'atmosphère, Université du Québec à Montreal.

Fraley, C., Raftery, A.E., Gneiting, T., 2010. Calibrating multimodel forecast ensembles with exchangeable and missing members using Bayesian Model Averaging. Monthly Weather Review 138, 190–202.

Franz, K.J., Hogue, T.S., Bank, M., He, M.X., 2014. Assessment of SWE data assimilation for ensemble streamflow predictions. Journal of Hydrology 519, 2737–2746.

Frei, T., 2010. Economic and social benefits of meteorology and climatology in Switzerland. Meteorological Applications 17, 39–44.

Gaborit, E., Anctil, F., Fortin, V., Pelletier, G., 2013. On the reliability of spatially disaggregated global ensemble rainfall forecasts. Hydrological Processes 27, 45–56.

Garçon, R., 1999. Modèle global Pluie-Débit pour la prévision et la prédétermination des crues. La Houille Blanche 7/8, 88–95.

Georgakakos, K.P., Seo, D.J., Gupta, H., Schaake, J., Butts, M.B., 2004. Towards the characterization of streamflow simulation uncertainty through multimodel ensembles. Journal of Hydrology 298, 222–241.

Girard, G., Morin, G., Charbonneau, R., 1972. Modèle précipitations-débits à discrétisation spatiale. Cahiers ORSTOM, Série Hydrologie 9, 35–52.

Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102, 359–378.

Gneiting, T., Raftery, A.E., Westveld, A.H., Goldman, T., 2005. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Monthly Weather Review 133, 1098–1118.

Goddard, L., Mason, S.J., Zebiak, S.E., Ropelewski, C.F., Basher, R., Cane, M.A., 2001. Current approaches to seasonal-to-interannual climate predictions. International Journal of Climatology 21, 1111–1152.

Goodrich, D.C., Woolhiser, D.A., 1991. Catchment hydrology. Report. International Union of Geodesy and Geophysics.

Gourley, J.J., Vieux, B.E., 2006. A method for identifying sources of model uncertainty in rainfall-runoff simulations. Journal of Hydrology 327, 68–80.

Gupta, H.V., Sorooshian, S., Yapo, P.O., 1998. Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. Water Resources Research 34, 751–763.

Gurtz, J., Zappa, M., Jasper, K., Lang, H., Verbunt, M., Badoux, A., Vitvar, T., 2003. A comparative study in modelling runoff and its components in two mountainous catchments. Hydrological Processes 17, 297–311.

Hagedorn, R., Doblas-Reyes, F.J., Palmer, T.N., 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting - I. Basic concept. Tellus Series A - Dynamic Meteorology and Oceanography 57, 219–233.

Hamill, T.M., 2001. Interpretation of rank histograms for verifying ensemble forecasts. Monthly Weather Review 129, 550–560.

Harader, E., Borrell-Estupina, V., Ricci, S., Coustau, M., Thual, O., Piacentini, A., Bouvier, C., 2012. Correcting the radar rainfall forcing of a hydrological model with data assimilation: application to flood forecasting in the Lez catchment in Southern France. Hydrology and Earth System Sciences 16, 4247–4264.

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

Hersbach, H., 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather and Forecasting 15, 559–570.

Hock, R., 2003. Temperature index melt modelling in mountain areas. Journal of Hydrology 282, 104–115.

Hong, Y., Hsu, K.L., Moradkhani, H., Sorooshian, S., 2006. Uncertainty quantification of satellite precipitation estimation and Monte Carlo assessment of the error propagation into hydrologic response. Water Resources Research 42, 1–15.

Hopson, T.M., Webster, P.J., 2010. A 1-10-Day Ensemble forecasting scheme for the major river basins of Bangladesh: Forecasting severe floods of 2003-07. Journal of Hydrometeorology 11, 618–641.

Huisman, J.A., Breuer, L., Bormann, H., Bronstert, A., Croke, B.F.W., Frede, H.G., Graeff, T., Hubrechts, L., Jakeman, A.J., Kite, G., Lanini, J., Leavesley, G., Lettenmaier, D.P., Lindstroem, G., Seibert, J., Sivapalan, M., Viney, N.R., Willems, P., 2009. Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM) III: Scenario analysis. Advances in Water Resources 32, 159–170.

Jakeman, A.J., Littlewood, I.G., Whitehead, P.G., 1990. Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. Journal of Hydrology 117, 275–300.

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

Jaun, S., Ahrens, B., 2009. Evaluation of a probabilistic hydrometeorological forecast system. Hydrology and Earth System Sciences 13, 1031–1043.

Jazwinski, A.H., 1970. Stochastic Processes and Filtering Theory.

Jeremiah, E., Sisson, S., Marshall, L., Mehrotra, R., Sharma, A., 2011. Bayesian calibration and uncertainty analysis of hydrological models: A comparison of adaptive Metropolis and sequential Monte Carlo samplers. Water Resources Research 47, 1–13.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. Transactions of the ASME – Journal of Basic Engineering 82, 35–45.

Kang, I.S., Yoo, J.H., 2006. Examination of multi-model ensemble seasonal prediction methods using a simple climate system. Climate Dynamics 26, 285–294.

Katul, G.G., Wendroth, O., Parlange, M.B., Puente, C.E., Folegatti, M.V., Nielsen, D.R., 1993. Estimation of in situ hydraulic conductivity function from nonlinear filtering theory. Water Resources Research 29, 1063–1070.

Kavvas, M.L., 1999. On the coarse-graining of hydrologic processes with increasing scales. Journal of Hydrology 217, 191–202.

Kienzle, S.W., 2008. A new temperature based method to separate rain and snow. Hydrological Processes 22, 5067–5085.

Kim, J., Yoo, C., 2014. Use of a dual Kalman filter for real-time correction of mean field bias of radar rain rate. Journal of Hydrology 519, 2785–2796.

Kirchner, J.W., 2006. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research 42, 1–5.

Klemeš, V., 1988. The modelling of mountain hydrology: the ultimate challenge, in: IAHS (Ed.), Strbské Pleso Workshop, p. 16p.

Krzysztofowicz, R., 1999. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. Water Resources Research 35, 2739–2750.

Krzysztofowicz, R., Maranzano, C.J., 2004. Hydrologic uncertainty processor for probabilistic stage transition forecasting. Journal of Hydrology 293, 57–73.

Kuchment, L.S., Romanov, P., Gelfan, A.N., Demidov, V.N., 2010. Use of satellite-derived data for characterization of snow cover and simulation of snowmelt runoff through a distributed physically based model of runoff generation. Hydrology and Earth System Sciences 14, 339–350.

Kuczera, G., Kavetski, D., Franks, S., Thyer, M., 2006. Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters. Journal of Hydrology 331, 161–177.

Lazo, J.K., Chestnut, L.G., 2002. Economic value of current and improved weather forecast in the U.S. household sector. Report. National Oceanic and Atmospheric Administration. Stratus Consulting.

Lee, H., Seo, D.J., Koren, V., 2011. Assimilation of streamflow and in situ soil moisture data into operational distributed hydrologic models: Effects of uncertainties in the data and initial model soil moisture states. Advances in Water Resources 34, 1597–1615.

Leviäkangas, P., Hautala, R., Räsänen, J., Öörni, R., Sonninen, S., Hekkanen, M., Ohlström, M., Venäläinen, A., Saku, S., 2007. Benefits of meteorological services in Croatia. Report. Finnish Meterological Institute, Helsinki.

Li, H.B., Luo, L.F., Wood, E.F., Schaake, J., 2009. The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting. Journal of Geophysical Research-Atmospheres 114, 1–10.

Li, Y., Ryu, D., Wang, Q., Pagano, T., Western, A., Hapuarachchi, P., Toscas, P., 2011. Assimilation of streamflow discharge into a continuous flood forecasting model, in: Bloschl, G., Takeuchi, K., Jain, S., Farnleitner, A., Schumann, A. (Eds.), The International Union of Geodesy and Geophysics XXV General Assembly, pp. 107–113.

Li, Y., Ryu, D., Western, A.W., Wang, Q.J., 2013. Assimilation of stream discharge for flood forecasting: The benefits of accounting for routing time lags. Water Resources Research 49, 1887–1900.

Li, Y., Ryu, D., Western, A.W., Wang, Q.J., Robertson, D.E., Crow, W.T., 2014. An integrated error parameter estimation and lag-aware data assimilation scheme for real-time flood forecasting. Journal of Hydrology 519, 2722–2736.

Lions, J.L., Temam, R., Wang, S.H., 1992. New formulations of the primitive equations of atmosphere and applications. Nonlinearity 5, 237–288.

Liu, Y., Weerts, A.H., Clark, M., Franssen, H.J.H., Kumar, S., Moradkhani, H., Seo, D.J., Schwanenberg, D., Smith, P., van Dijk, A., van Velzen, N., He, M., Lee, H., Noh, S.J., Rakovec, O., Restrepo, P., 2012. Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. Hydrology and Earth System Sciences 16, 3863–3887.

Liu, Y.Q., Gupta, H.V., 2007. Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. Water Resources Research 43, 1–18.

Lorenc, A.C., 2003. The potential of the ensemble Kalman filter for NWP - a comparison with 4D-Var. Quarterly Journal of the Royal Meteorological Society 129, 3183–3203.

Lorenz, E.N., 1963. Deterministic nonperiodic flow. Journal of the Atmospheric Sciences 20, 130–141.

Mandel, J., 2006. Efficient Implementation of the Ensemble Kalman Filter. Report. UCDHSC.

Maraun, D., Wetterhall, F., Ireson, A.M., Chandler, R.E., Kendon, E.J., Widmann, M., Brienen, S., Rust, H.W., Sauter, T., Themessl, M., Venema, V.K.C., Chun, K.P., Goodess, C.M., Jones, R.G., Onof, C., Vrac, M., Thiele-Eich, I., 2010. Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. Reviews of Geophysics 48, 1–34.

Marshall, L., Nott, D., Sharma, A., 2005. Hydrological model selection: a Bayesian alternative. Water Resources Research 41, 1–11.

Martin, E., 2005. Modélisation du manteau neigeux et applications dans les domaines du changement climatique et de l'hydrologie. Habilitation a diriger des recherches. Centre National de Recherche Météorologique, Grenoble.

Matheson, J.E., Winkler, R.L., 1976. Scoring rules for continuous probability distributions. Management Science 22, 1087–1096.

Mazenc, B., Sanchez, M., Thiery, D., 1984. Analyse de l'influence de la physiographie d'un bassin versant sur les paramètres d'un modèle hydrologique global et sur les débits caractéristiques à l'exutoire. Journal of Hydrology 69, 97–188.

McMillan, H.K., Hreinsson, E.O., Clark, M.P., Singh, S.K., Zammit, C., Uddstrom, M.J., 2013. Operational hydrological data assimilation with the recursive ensemble Kalman filter. Hydrology and Earth System Sciences 17, 21–38.

Meier, P., Froemelt, A., Kinzelbach, W., 2011. Hydrological real-time modelling in the Zambezi river basin using satellite-based soil moisture and rainfall data. Hydrology and Earth System Sciences 15, 999–1008.

Mellow, R., 1999. A synopsis and comparison of selected snowmelt algorithms. Report. US Army Corps of Engineers.

Merz, B., Kreibich, H., Schwarze, R., Thieken, A., 2010. Review article 'Assessment of economic flood damage'. Natural Hazards and Earth System Sciences 10, 1697–1724.

Metcalf, Eddy, 1971. Storm water management model. Report. Univ. Of Floridan and Water Resources Engeineers Inc.

Ministry Of Transport, 2006. Spatial planning key decision "Room for the river". Report. Ministry of Transport, Public Works and Water Management.

Molinari, D., Handmer, J., 2011. A behavioural model for quantifying flood warning effectiveness. Journal of Flood Risk Management 4, 23–32.

Molteni, F., Buizza, R., Marsigli, C., Montani, A., Nerozzi, F., Paccagnella, T., 2001. A strategy for high-resolution ensemble prediction. I: Definition of representative members and global-model experiments. Quarterly Journal of the Royal Meteorological Society 127, 2069–2094.

Molteni, F., Buizza, R., Palmer, T.N., Petroliagis, T., 1996. The ECMWF ensemble prediction system: Methodology and validation. Quarterly Journal of the Royal Meteorological Society 122, 73–119.

Moore, R.J., Clarke, R.T., 1981. A distribution function approach to rainfall runoff modeling. Water Resources Research 17, 1367–1382.

Moradkhani, H., Hsu, K., Hong, Y., Sorooshian, S., 2006. Investigating the impact of remotely sensed precipitation and hydrologic model uncertainties on the ensemble streamflow forecasting. Geophysical Research Letters 33, 1–5.

Moradkhani, H., Sorooshian, S., Gupta, H.V., Houser, P.R., 2005. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. Advances in Water Resources 28, 135–147.

Murphy, A.H., Ehrendorfer, M., 1987. On the relationship between the accuracy and value of forecasts in the cost–loss ratio situation. Weather Forecasting 2, 243–251.

Murphy, J.M., 1988. The impact of ensemble forecasts on predictability. Quarterly Journal of the Royal Meteorological Society 114, 463–493.

Murphy, J.M., Sexton, D.M.H., Barnett, D.N., Jones, G.S., Webb, M.J., Stainforth, D.A., 2004. Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature 430, 768–772.

Naevdal, G., Johnsen, L.M., Aanonsen, S.I., Vefring, E.H., 2003. Reservoir monitoring and continuous model updating using ensemble Kalman filter. Spe Journal 10, 66–74.

Nash, J.E., Sutcliffe, I., 1970. River flow forecasting through conceptual models. Part 1 - A discussion of principles. Journal of Hydrology 10, 282–290.

Nicolle, P., Ramos, M., Andreassian, V., Valery, A., 2011. Better predict snowmelt floods: evaluation of streamflow probabilistic forecasts over mountainous catchments in France, in: France, S.H.d. (Ed.), Colloque SHF : "L'eau en montagne, mieux observer pour mieux prévoir", pp. 163–170.

Nie, S., Zhu, J., Luo, Y., 2011. Simultaneous estimation of land surface scheme states and parameters using the ensemble Kalman filter: identical twin experiments. Hydrology and Earth System Sciences 15, 2437–2457.

Nielsen, S.A., Hansen, E., 1973. Numerical simulation of the rainfall-runoff process on a daily basis. Nordic Hydrology 4, 171–190.

Noh, S.J., Rakovec, O., Weerts, A.H., Tachikawa, Y., 2014. On noise specification in data assimilation schemes for improved flood forecasting using distributed hydrological models. Journal of Hydrology 519, 2707–2721.

Noh, S.J., Tachikawa, Y., Shiiba, M., Kim, S., 2013. Sequential data assimilation for streamflow forecasting using a distributed hydrologic model: particle filtering and ensemble Kalman filtering, in: Floods: from Risk to Opportunity, pp. 341–349.

O'Connell, P.E., Nash, J.E., Farrell, J.P., 1970. River flow forecasting through conceptual models, Part II - The Brosna catchment at Ferbane. Journal of Hydrology 10, 317–329.

Ohmura, A., 2001. Physical basis for the temperature-based melt-index method. Journal of Applied Meteorology 40, 753–761.

Oudin, L., Andreassian, V., Mathevet, T., Perrin, C., Michel, C., 2006. Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations. Water Resources Research 42, 1–10.

Oudin, L., Michel, C., Anctil, F., 2005. Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 1 - Can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs? Journal of Hydrology 303, 275–289.

Pagano, T.C., Wood, A.W., Ramos, M.H., Cloke, H.L., Pappenberger, F., Clark, M.P., Cranston, M., Kavetski, D., Mathevet, T., Sorooshian, S., Verkade, J.S., 2014. Challenges of Operational River Forecasting. Journal of Hydrometeorology 15, 1692–1707.

Palmer, T.N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M., Delecluse, P., Deque, M., Diez, E., Doblas-Reyes, F.J., Feddersen, H., Graham, R., Gualdi, S., Gueremy, J.F., Hagedorn, R., Hoshen, M., Keenlyside, N., Latif, M., Lazar, A., Maisonnave, E., Marletto, V., Morse, A.P., Orfila, B., Rogel, P., Terres, J.M., Thomson, M.C., 2004. Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). Bulletin of the American Meteorological Society 85, 853–872.

Pappenberger, F., Beven, K.J., 2006. Ignorance is bliss: Or seven reasons not to use uncertainty analysis. Water Resources Research 42, 1–8.

Pappenberger, F., Beven, K.J., Hunter, N.M., Bates, P.D., Gouweleeuw, B.T., Thielen, J., 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, 381–393.

Pappenberger, F., Cloke, H.L., Parker, D.J., Wetterhall, F., Richardson, D.S., Thielen, J., 2015. The monetary benefit of early flood warnings in Europe. Environmental science and policy 51, 278–291.

Park, Y.Y., Buizza, R., Leutbecher, M., 2008. TIGGE: Preliminary results on comparing and combining ensembles. Quarterly Journal of the Royal Meteorological Society 134, 2029–2050.

Parker, D., 1991. The damage-reducing effects of flood warnings. Report. Flood Hazard Research Centre, Middlesex University.

Parker, D.J., Tunstall, S.M., McCarthy, S., 2007. New insights into the benefits of flood warnings: Results from a household survey in England and Wales. Environmental Hazards 7, 193–210.

Parrish, M.A., Moradkhani, H., DeChant, C.M., 2012. Toward reduction of model uncertainty: Integration of Bayesian model averaging and data assimilation. Water Resources Research 48, 1–18.

Penman, H.L., 1948. Natural evaporation from open water, bare soil and grass, in: Royal Society of London Series A - Mathematical and Physical Sciences, pp. 120–145.

Penman, H.L., 1961. Weather, plant and soil factors in hydrology. Weather 16, 207–219.

Perrels, A., Frei, T., Espejo, F., Jamin, L., Thomalla, A., 2013. Socio-economic benefits of weather and climate services in Europe. Advances in Science and Research 10, 65–70.

Perrin, C., 2000. Vers une amélioration d'un modèle global pluie-débit. Thesis. Institut National Polytechnique de Grenoble - INPG.

Perrin, C., Michel, C., Andreassian, V., 2001. Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. Journal of Hydrology 242, 275–301.

Perrin, C., Michel, C., Andreassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. Journal of Hydrology 279, 275–289.

Poulin, A., Brissette, F., Leconte, R., Arsenault, R., Malo, J.S., 2011. Uncertainty of hydrological modelling in climate change impact studies in a Canadian, snow-dominated river basin. Journal of Hydrology 409, 626–636.

Priest, S.J., Parker, D.J., Tapsell, S.M., 2011. Modelling the potential damage-reducing benefits of flood warnings using European cases. Environmental Hazards-Human and Policy Dimensions 10, 101–120.

Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian model averaging to calibrate forecast ensembles. Monthly Weather Review 133, 1155–1174.

Rakovec, O., Weerts, A.H., Hazenberg, P., Torfs, P., Uijlenhoet, R., 2012. State updating of a distributed hydrological model with Ensemble Kalman Filtering: effects of updating frequency and observation network density on forecast accuracy. Hydrology and Earth System Sciences 16, 3435–3449.

Ramos, M.H., Mathevet, T., Thielen, J., Pappenberger, F., 2010. Communicating uncertainty in hydro-meteorological forecasts: mission impossible? Meteorological Applications 17, 223–235.

Randrianasolo, A., Ramos, M.H., Thirel, G., Andreassian, V., Martin, E., 2010. Comparing the scores of hydrological ensemble forecasts issued by two different hydrological models. Atmospheric Science Letters 11, 100–107.

Ranzi, R., Bacchi, B., Ceppi, A., Cislaghi, M., Ehret, U., Jaun, S., Marx, A., Hegg, C., Zappa, M., 2009. Real-time demonstration of hydrological ensemble forecasts in map d-phase. Houille Blanche-Revue Internationale De L Eau 5, 95–104.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.J., Participants, D., 2004. Overall distributed model intercomparison project results. Journal of Hydrology 298, 27–60.

Refsgaard, J.C., 1997. Validation and intercomparison of different updating procedures for real-time forecasting. Nordic Hydrology 28, 65–84.

Refsgaard, J.C., Knudsen, J., 1996. Operational validation and intercomparison of different types of hydrological models. Water Resources Research 32, 2189–2202.

Regimbeau, F., Habets, F., Martin, E., Noilhan, J., 2007. Ensemble streamflow forecasts over France. Report. ECMWF Newsletter.

Reichle, R.H., Walker, J.P., Koster, R.D., Houser, P.R., 2002. Extended versus ensemble Kalman filtering for land data assimilation. Journal of Hydrometeorology 3, 728–740.

Renner, M., Werner, M.G.F., Rademacher, S., Sprokkereef, E., 2009. Verification of ensemble flow forecasts for the River Rhine. Journal of Hydrology 376, 463–475.

Renzullo, L.J., van Dijk, A., Perraud, J.M., Collins, D., Henderson, B., Jin, H., Smith, A.B., McJannet, D.L., 2014. Continental satellite soil moisture data assimilation improves root-zone moisture analysis for water resources assessment. Journal of Hydrology 519, 2747–2762.

Richardson, D.S., 2000. Skill and relative economic value of the ECMWF ensemble prediction system. Quarterly Journal of the Royal Meteorological Society 126, 649–667.

Rogers, D., Tsirkunov, V., 2010. Costs and benefits of early warning systems in Global Assesment Report on Disaster Risk Reduction. Report. United Nations International Strategy for Disaster Reduction (UNISDR) and World Bank.

Roulin, E., 2007. Skill and relative economic value of medium-range hydrological ensemble predictions. Hydrology and Earth System Sciences 11, 725–737.

Roulston, M.S., Smith, L.A., 2003. Combining dynamical and statistical ensembles. Tellus Series A - Dynamic Meteorology and Oceanography 55, 16–30.

Salamon, P., Feyen, L., 2010. Disentangling uncertainties in distributed hydrological modeling using multiplicative error models and sequential data assimilation. Water Resources Research 46, 1–20.

Segond, M.L., Onof, C., Wheater, H.S., 2006. Spatiat-temporal disaggregation of daily rainfall from a generalized linear model. Journal of Hydrology 331, 674– 689.

Seiller, G., 2013. Évaluation de la sensibilité des projections hydrologiques au choix des outils hydrométéorologiques globaux conceptuels. Thesis. University Laval, Canada.

Seiller, G., Anctil, F., Perrin, C., 2012. Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions. Hydrology and Earth System Sciences 16, 1171–1189.

Seo, D.J., Cajina, L., Corby, R., Howieson, T., 2009. Automatic state updating for operational streamflow forecasting via variational data assimilation. Journal of Hydrology 367, 255–275.

Seo, D.J., Koren, V., Cajina, N., 2003. Real-time variational assimilation of hydrologic and hydrometeorological data into operational hydrologic forecasting. Journal of Hydrometeorology 4, 627–641.

Silberstein, R.P., 2006. Hydrological models are so good, do we still need data? Environmental Modelling & Software 21, 1340–1352.

Singh, V.P., Woolhiser, D.A., 2002. Mathematical modeling of watershed hydrology. Journal of Hydrologic Engineering 7, 270–292.

Smith, M.B., Seo, D.J., Koren, V.I., Reed, S.M., Zhang, Z., Duan, Q., Moreda, F., Cong, S., 2004. The distributed model intercomparison project (DMIP): motivation and experiment design. Journal of Hydrology 298, 4–26.

Stanski, H.R., Wilson, L.J., Burrows, W.R., 1989. Survey of common verification methods in meteorology. Report. Atmospheric Environment Service - Forecast Research Division.

Stephens, E., Cloke, H., 2014. Improving flood forecasts for better flood preparedness in the UK (and beyond). Geographical Journal 180, 310–316.

Sugawara, M., 1979. Automatic calibration of the tank model. Hydrological Sciences 24, 375–388.

Sun, C.J., Walker, J.P., Houser, P.R., 2004. A methodology for snow data assimilation in a land surface model. Journal of Geophysical Research-Atmospheres 109, 1–12.

Thielen, J., Bartholmes, J., Ramos, M.H., de Roo, A., 2009. The European Flood Alert System - Part 1: Concept and development. Hydrology and Earth System Sciences 13, 125–140.

Thiemann, M., Trosset, M., Gupta, H., Sorooshian, S., 2001. Bayesian recursive parameter estimation for hydrologic models. Water Resources Research 37, 2521–2535.

Thiery, D., 1982. Utilisation d'un modèle global pour identifier sur un niveau piézométrique des influences multiples dues à diverses activités humaines, in: Improvement of methods of long term prediction of variations in groundwater resources and regimes due to human activity, pp. 71–77.

Thirel, G., Martin, E., Mahfouf, J.F., Massart, S., Ricci, S., Habets, F., 2010a. A past discharges assimilation system for ensemble streamflow forecasts over France - Part 1: Description and validation of the assimilation system. Hydrology and Earth System Sciences 14, 1623–1637.

Thirel, G., Martin, E., Mahfouf, J.F., Massart, S., Ricci, S., Regimbeau, F., Habets, F., 2010b. A past discharge assimilation system for ensemble streamflow forecasts over France - Part 2: Impact on the ensemble streamflow forecasts. Hydrology and Earth System Sciences 14, 1639–1653.

Thirel, G., Salamon, P., Burek, P., Kalas, M., 2013. Assimilation of MODIS snow cover area data in a distributed hydrological model Using the particle filter. Remote Sensing 5, 5825–5850.

Thompson, P.D., 1977. How to improve accuracy by combining independent forecasts. Monthly Weather Review 105, 228–229.

Thornthwaite, C.W., 1948. An approach toward a rational classification of climate. Geographical Review 38, 55–94.

Thornthwaite, C.W., Mather, J.R., 1955. The water balance. Report. Drexel Institute of Climatology.

Todini, E., 2007. Hydrological catchment modelling: past, present and future. Hydrology and Earth System Sciences 11, 468–482.

Turcotte, R., Fortin, J.P., Bernier, M., Gauthier, Y., 2001. Developments for snowpack water equivalent monitoring using Radarsat data as input to the Hydrotel hydrological model, in: Owe, M.B.K.R.J.R.A. (Ed.), Remote Sensing and Hydrology 2000, pp. 374–376.

Turcotte, R., Lacombe, P., Dimnik, C., Villeneuve, J.P., 2004. Prévision hydrologique distribuée pour la gestion des barrages publics du Québec. Canadian Journal of Civil Engineering 31, 308–320.

Valery, A., Andreassian, V., Perrin, C., 2014a. 'As simple as possible but not simpler': What is useful in a temperature-based snow-accounting routine? Part 1 - Comparison of six snow accounting routines on 380 catchments. Journal of Hydrology 517, 1166–1175.

Valery, A., Andreassian, V., Perrin, C., 2014b. 'As simple as possible but not simpler': What is useful in a temperature-based snow-accounting routine? Part 2 - Sensitivity analysis of the Cemaneige snow accounting routine on 380 catchments. Journal of Hydrology 517, 1176–1187.

Velázquez, J., Anctil, F., Perrin, C., 2010. Performance and reliability of multimodel hydrological ensemble simulations based on seventeen lumped models and a thousand catchments. Hydrology and Earth System Sciences 14, 2303–2317.

Velázquez, J.A., 2010. Evaluation of hydrological ensemble prediction systems for operational forecasting. Thesis. University Laval, Canada.

Velázquez, J.A., Anctil, F., Ramos, M.H., Perrin, C., 2011. Can a multi-model approach improve hydrological ensemble forecasting? A study on 29 French catchments using 16 hydrological model structures? Advances in Geosciences 29, 33–42.

Velázquez, J.A., Petit, T., Lavoie, A., Boucher, M.A., Turcotte, R., Fortin, V., Anctil, F., 2009. An evaluation of the Canadian global meteorological ensemble prediction system for short-term hydrological forecasting. Hydrology and Earth System Sciences 13, 2221–2231.

Verkade, J.S., Brown, J.D., Reggiani, P., Weerts, A.H., 2013. Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales. Journal of Hydrology 501, 73–91.

Verkade, J.S., Werner, M.G.F., 2011. Estimating the benefits of single value and probability forecasting for flood warning. Hydrology and Earth System Sciences 15, 3751–3765.

Viney, N.R., Bormann, H., Breuer, L., Bronstert, A., Croke, B.F.W., Frede, H., Graeff, T., Hubrechts, L., Huisman, J.A., Jakeman, A.J., Kite, G.W., Lanini, J., Leavesley, G., Lettenmaier, D.P., Lindstroem, G., Seibert, J., Sivapalan, M., Willems, P., 2009. Assessing the impact of land use change on hydrology by ensemble modelling (LUCHEM) II: Ensemble combinations and predictions. Advances in Water Resources 32, 147–158.

Vrugt, J.A., Clark, M.P., Diks, C.G.H., Duan, Q., Robinson, B.A., 2006. Multi-objective calibration of forecast ensembles using Bayesian model averaging. Geophysical Research Letters 33, 1–6.

Vrugt, J.A., Diks, C.G.H., Gupta, H.V., Bouten, W., Verstraten, J.M., 2005. Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. Water Resources Research 41, 1–17.

Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, S., 2003. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resources Research 39, 1–18.

Vrugt, J.A., Robinson, B.A., 2007. Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. Water Resources Research 43, 1–15.

Wagener, T., Boyle, D.P., Lees, M.J., Wheater, H.S., Gupta, H.V., Sorooshian, S., 2001. A framework for development and application of hydrological models. Hydrology and Earth System Sciences 5, 13–26.

Walker, J.P., Houser, P.R., 2001. A methodology for initializing soil moisture in a global climate model: Assimilation of near-surface soil moisture observations. Journal of Geophysical Research-Atmospheres 106, 11761–11774.

Walker, W.E., Harremoës, P., Rotmans, J., van der Sluijs, J.P., van Asselt, M.B.A., Janssen, P., Krayer von Krauss, M.P., 2003. Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. Integrated Assessment 4, 5–17.

Warmerdam, P.M., Kole, J., Chormanski, J., 1997. Modelling rainfall-runoff processes in the Hupselse Beek research basin, in: IHP-V, Technical Documents in Hydrology, pp. 155–160.

Weerts, A.H., El Serafy, G.Y.H., 2006. Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. Water Resources Research 42, 1–17.

Wetterhall, F., Pappenberger, F., Alfieri, L., Cloke, H.L., Thielen-del Pozo, J., Balabanova, S., Danhelka, J., Vogelbacher, A., Salamon, P., Carrasco, I., Cabrera-Tordera, A.J., Corzo-Toscano, M., Garcia-Padilla, M., Garcia-Sanchez, R.J., Ardilouze, C., Jurela, S., Terek, B., Csik, A., Casey, J., Stankunavicius, G., Ceres, V., Sprokkereef, E., Stam, J., Anghel, E., Vladikovic, D., Eklund, C.A., Hjerdt, N., Djerv, H., Holmberg, F., Nilsson, J., Nystrom, K., Susnik, M., Hazlinger, M., Holubecka, M., 2013. HESS Opinions "Forecaster priorities for improving probabilistic flood forecasts". Hydrology and Earth System Sciences 17, 4389–4399.

Whitaker, J.S., Hamill, T.M., 2002. Ensemble data assimilation without perturbed observations. Monthly Weather Review 130, 1913–1924.

Wöhling, T., Lennartz, F., Zappa, M., 2006. Technical Note: Updating procedure for flood forecasting with conceptual HBV-type models. Hydrology and Earth System Sciences 10, 783–788.

Yapo, P.O., Gupta, H.V., Sorooshian, S., 1998. Multi-objective global optimization for hydrologic models. Journal of Hydrology 204, 83–97.

Zhao, R.J., Zuang, Y.L., Fang, L.R., Liu, X.R., Zhang, Q., 1980. The Xinanjiang model. IAHS Publications 129, 351–356.

Zhu, Y.J., Toth, Z., Wobus, R., Richardson, D., Mylne, K., 2002. The economic value of ensemble-based weather forecasts. Bulletin of the American Meteorological Society 83, 73–83.