An intercomparison of flood forecasting models for the Meuse River basin

Louise Arnal

An intercomparison of flood forecasting models for the Meuse River basin

Louise Arnal

Supervisors: Albrecht Weerts (Deltares), Maria-Helena Ramos (IRSTEA) and Maarten J. Waterloo (Vrije Universiteit Amsterdam)

1^{st} of September 2014

Master Research Project in Hydrology, code AM1104, 27 ECTS, with Extension of Master Thesis in Hydrology, code 450151, 12 ECTS

'This February 1953 file photo shows an aerial view of a windmill pump elevated above the floodwaters in the coastal village of Oude Tonge in The Netherlands. It took the collapse of dikes, drowning deaths of more than 1,800 people, and evacuation of another 100,000 in 1953 for the Dutch to say "Never again!" They have since constructed the world's sturdiest battery of dikes, dams and barriers. No disaster on that scale has happened since.' (Photos, 2012)

Contents

Abstract

In a world where floods are an important hazard, the incentive for flood forecasting, promoted by the advances in earth observations' quality, is growing. In this context, while models are developed and used primarily on a national level, it has appeared essential to share data and models internationally. The present research inscribes itself into this scope. Seven years of hourly simulations and two weeks of hourly hindcasts with a lead time of five days, with and without an error correction procedure, were executed using the Delft-FEWS forecasting platform for four flood forecasting models and five sub-basins of the Meuse River basin. These models included two lumped conceptual models, the HBV-96 model, operationally used in the Netherlands and in Germany and the GR4H model, the hourly version of the GR4J model, used in France; and two fully-distributed physically based models, the wflow_hbv and wflow_sbm models developed at Deltares, the Netherlands. The quality of the simulations and of the hindcasts produced by each model for each of the five sub-basins was analysed and compared based on a combination of graphical techniques and performance metrics. This intercomparison framework brought to light several results. It has been shown that the Regnie interpolation method improved the HBV-96 model simulation performance for high flows compared to the Kriging interpolation method, highlighting the importance of data interpolation. Furthermore, the conceptual HBV-96 and GR4H models produced more reliable simulations and hindcasts than the physically-based wflow models. Finally, the use of high resolution input data did not show any clear improvement of the simulations and hindcasts quality. As part of a larger project, this research has set the starting point for further investigations regarding the disparities in the models results identified.

Chapter 1

Introduction

Flooding is an event characterised by the physical phenomenon of slow or fast submersion of a zone usually dry. It can have a partially artificial origin due to anthropogenic alterations of the environment (forest denudation, dams...) and/or be controlled by meteorological events such as precipitation, temperature variations (snow melt) (PAGASA, 2004).

While, based on data acquired from 1980 to 2008, floods kill on average over 6700 people, affect more than 96 million people and lead to damage costs of approximately 10 billion euros per year worldwide (UNISDR, 2013). It has become evident that flood modelling and flood forecasting systems are crucial for the improvement of safety and protection against this natural hazard (WMO, 2011a).

Over the years, forecasting models have been developed in Germany, The Netherlands and France. In Germany, the federal Institute of Hydrology (BfG), and in the Netherlands, the Dutch Water Management Centre (WMCN), use the semi-distributed HBV-96 model as their basis for flow forecasting. Simultaneously, in France, the national research institute in sciences and technologies of the environment (IRSTEA, formerly known as Cemagref) and the French flood forecasting centres (SCHAPI-SPC) use lumped rural engineering (GR) models for flow forecasting of the Meuse River basin. Moreover, with the prevailing advances in earth observations' quality (precipitation, temperature, evapotranspiration, soil moisture, etc) the incentive to enhance the use of distributed hydrological models has increased greatly. The recent development of two distributed hydrological models at Deltares (Dutch research institute), the wflow_hbv and the wflow_sbm models, supports this.

Although these modelling systems are becoming more efficient with time, they will never be perfect due to several factors, such as uncertainty in weather observations and predictions, models accuracy to represent real systems, forecasts lead time [1],

[1] Lead time is the time that separates the moment when the threat is recognised (e.g.: here flooding

etc (Verkade and Werner, 2011).

By comparing the performance of flood forecasting models, essential outcomes regarding the relation between flood modelling and flood forecasting limitations and the above mentioned factors that contribute to those limitations can be obtained; these outcomes being the key to an improvement of the flood forecasting models (WMO, 2011b).

Within the framework of a larger project, which leading drivers are to improve flood forecasting by adding more models to the Delft-FEWS forecasting platform and by sharing data (both historical and operational) within and across national boundaries, the specific objectives of this research were to compare several flood forecasting models (HBV-96, GR4H, wflow_hbv and wflow_sbm) in simulations and hindcasts, for several sub-basins of the Meuse River basin, in order to:

- disentangle the link between model performance and model structure, assessing the limits of model complexity, and
- explore the benefits of using high resolution forcing data

For this purpose, simulations and hindcasts were executed within Delft-FEWS and analysed for the four different models for five sub-basins of the Meuse River basin.

So far, these four forecasting models have only been compared on a local scale. This project is thus unique in the sense that it enters in the context of international-wide intercomparisons for flood forecasting models used in Germany, The Netherlands and France. Furthermore, using the Meuse River basin as the basis of the intercomparison is an appropriate choice due to its transnational boundaries.

Moreover, this research, carried out with Deltares and IRSTEA, originates and embeds itself in the scope of international projects on the Meuse such as the international 'Amice' project; symposiums such as the 'Mini-Symposium: Towards improved transnational hydrological modelling of the Meuse basin' (Archambeau, 2013); as well as international agreements such as the 'Accord international sur la Meuse' approved in Gent, Belgium in 2002 according to which the members (the governments of Germany. Belgium, France, the Luxembourg and The Netherlands) are guided, within their actions, by the principle of prevention (flood forecasting being a step towards prevention; de Wallonie, 2002).

In a first part, the research area, the models and the forecasting platform used for the purpose of this research are presented. In a second part, the methods implemented and their corresponding results are described. Subsequently, the results are discussed and the main conclusions of the research are stated. Finally, recommendations are given with the aim of proposing a clear continuation of this research.

---

event) and when the predicted event takes place (Verkade and Werner, 2011).

Chapter 2

Research area

2.1 Meuse River basin

Geographical location

The Meuse River is a 875 km long river, originating in Pouilly-en-Bassigny, North-East of France. In the Netherlands, it divides near Heusden into the Afgedamde Maas and the Bergse Maas. The Afgedamse Maas joins the Waal distributary branch of the river Rhine at Woudrichem, which merges with the Bergse Maas (then called Amer) downstream, to form the Hollandsch Diep (Figure 2.1). Its total basin area is 33,000 km^2, partially covering France, Luxembourg, Belgium, Germany and The Netherlands (Figure A.1, Appendix A).

Figure 2.1: Meuse River in the Netherlands, (WWasser, 2013)

Geology

Three major soil types predominate this river basin: *i)* sedimentary Mesozoic rocks found from the Meuse River source until its confluence with the Chiers River, *ii)* Paleozoic metamorphic rocks from the confluence with the Chiers River until the Belgian-Dutch frontier, and *iii)* Cenozoic unconsolidated sedimentary rocks from the Belgian-Dutch border until its outlet, with an increasing tidal deposits contribution towards the outlet of the Meuse River basin (Tu, 2006).

Land use

The land use of the Meuse River basin consists mainly of agricultural land (pastures and arable land), covering up to 60% of the total basin area, while forested and urban areas cover the rest of the basin (approximately 30% and 10% of the total

basin area respectively; De Wit et al., 2007).

Climate

A temperate maritime climate dominates the Meuse River basin, with cool summers and temperate winter periods. As a result, the precipitation is fairly regular throughout the year, but presents a large spatial variability due to the elevation distribution over the whole area.

The elevation ranges from less than 100 m downstream, where low precipitation values of 700 to 800 mm are found, to more than 500 m in the Ardennes, with high annual precipitation values of 1000 to 1300 mm (Tu, 2006).

Hydrology

The Meuse River is a rain-fed river, although snow is a component of its regime during the winter months. Its mean discharge rate is about 250 m^3/s, with high flows - that eventually lead to floods - in winter around 3000 m^3/s and low flows in summer of 10 m^3/s (Tu, 2006).

The Meuse discharge regime is characterised by temporal variations (as seen previously), as well as spatial variations. The latter are controlled by the landscape and man-made flow regulating structures.

The upstream section of the Meuse River, until Charleville-Mézières (Figure A.1, Appendix A) is marked by a hilly landscape with large floodplains. This landscape, together with the numerous weirs present in this area and the existence of a lateral canal alongside certain segments of the river, lead to substantial flood mitigation downstream.

The central section of the Meuse River, delimited by Charleville-Mézières and Liège, crosses the peculiar mountainous landscape of the Ardennes Massif (Figure A.1, Appendix A). Throughout this part, the Meuse River flows through narrow steep valleys and its tributaries show very high river gradients. These features give rise to high flood waves, partially attenuated by weirs.

Finally, the downstream section of the river presents very wide floodplains and numerous weirs, the Meuse gradually turning into a typical lowland river (Tu, 2006).

2.2 Sub-basins selection

For the purpose of this research project, a selection of five sub-basins of the Meuse was made. The purpose of this selection is to offer a sample of sub-basins presenting large contrasts in basin area, precipitation heights as well as location within the Meuse catchment (upstream or downstream), in order to examine the effect of model definition on simulations and forecasts quality. The following sub-basins were selected:

- The Meuse in Saint-Mihiel (2543 km^2)
- The Meuse in Stenay (3914 km^2)
- The Chiers in Carignan (1978 km^2)
- The Semoy in Haulmé (1338 km^2)
- The Ourthe in Tabreux (1597 km^2)

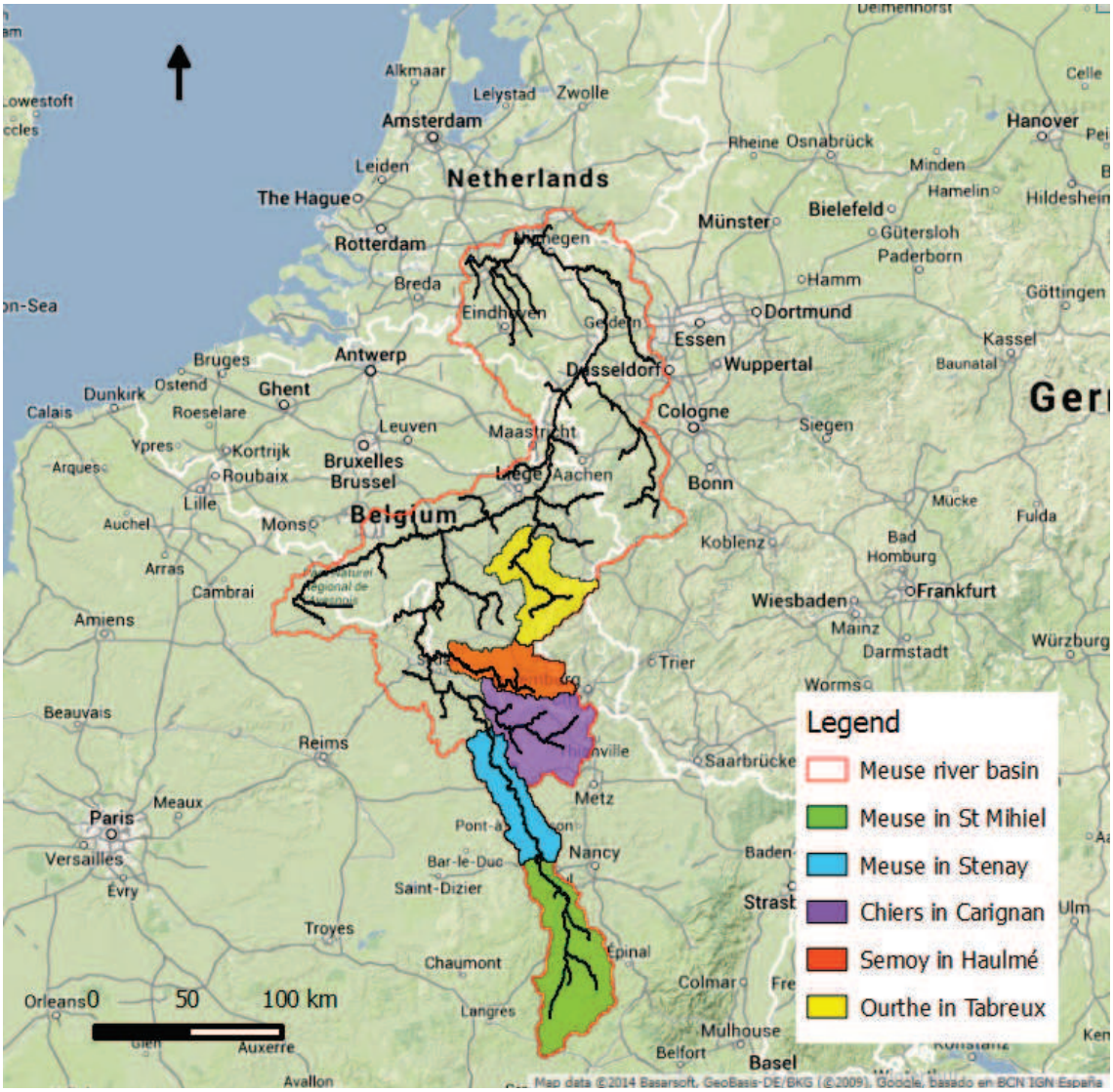


Figure 2.2: Selected Meuse sub-basins used for the models intercomparison

Chapter 3

The models

> *Models are 'a selected simplified version of a real system and phenomena that take place within it, which approximately simulates the excitation-response relationships that are of interest' (Bear and Cheng, 2010).*

The models considered for the purpose of this research all aim at simulating river discharge from input climatic conditions, their main differences lying in the models structures. In this section, a short overview of the latter is given.

3.1 HBV-96

The HBV-96 model is the operational model running within the Delft-FEWS platform at Deltares (see Chapter 4) (Berglov, 2009). HBV-96 is a conceptual semi-distributed hydrological model, originally created at the Swedish Meteorological and Hydrological Institute (SMHI) in 1996 from a re-evaluation of the lumped HBV hydrological model (Lindstrom et al., 1997; SMHI, 1999) (Figure B.1, Appendix B).

The model runs with hourly time steps over sub-basins (about 40 km^2) using the following input meteorological data: hourly precipitation and air temperature and potential evapotranspiration monthly mean averages. Computation of precipitation and air temperature is done through an optimal interpolation method, and potential evapotranspiration is corrected using altitude. Sub-basins are subdivided into zones defined by elevation and vegetation spatial distribution. Four different land use classes are used within the model: open areas, forests, lakes and glaciers (Berglov, 2009).

HBV-96 is structured in multiple routines: *i*) the precipitation routine, *ii*) the soil moisture routine and *iii*) the runoff response routine. The latter is composed of two

zones, an upper non-linear reservoir and a lower linear reservoir. The upper zone reservoir generates a quick runoff flux, while the lower reservoir produces as an output a baseflow component. The outputs of the reservoir are then summed prior to being subjected to the MAXBAS tranformation function (Figure B.2, Appendix B). This function is a triangular weighting function and is used as a routing function to compute simulated runoff. Between sub-basins, the routing is achieved by the use of modified Muskingum's equations and delay parameters (lag and damps) (Lindstrom et al., 1997; Berglov, 2009).

The HBV model has applications in flood modelling and operational flood forecasting in more than fifty countries all over the world. From multiple comparison framework, it has been established that the latter produced moderate to good flood simulations and forecasts. This quality, adjoined to its simplicity, conferred by its low number of parameters, has made the HBV-96 model adequately adjustable to an extensive number of cases, especially to large sub-basins such as the Rhine river basin (Berglov, 2009; Gorgen et al., 2010; De Wit, 2005; Drogue et al., 2010).

Many developments have been made for the operational use of the HBV-96 for flood forecasting within Delft-FEWS. These include the improvement of: the data interpolation method (Weerts et al., 2008), the evaporation calculations and the precipitation and temperature updating for forecasting (Berglov, 2009).

3.2 GR4H

The GR4H model is the hourly version of the GR4J daily model (Mathevet, 2005). It is a four-parameter empirical lumped rainfall-runoff model created at IRSTEA in the early 1980s. This model aims at reproducing a catchment's stream response to water inputs events by an ensemble of conceptual (non-physical) processes (Figure C.1, Appendix C; Perrin et al., 2003, 2007).

The GR4H model runs with hourly precipitation data (P_k) and average potential evapotranspiration data (E), lumped over the catchment area, as inputs and has four parameters (X_1: the production store maximum capacity (mm), X_2: the groundwater exchange coefficient (mm), X_3: the routing store maximum capacity (mm) and X_4: the time base of the unit hydrographs (hours)) optimized for each basin (Perrin et al., 2003, 2007).

Two conceptual reservoirs are present within this model, a production store and a routing store. The production store behaves conceptually similarly to an unsaturated reservoir, controlling the water infiltration into the ground. The routing store, as its name indicates, acts like a groundwater reservoir and participates to the water routing out of the system.

Water is routed linearly by two unit hydrographs (UH1 and UH2), which contribution is fixed. Indeed, 90% of the water is routed by UH1 and 10% of the water is routed by UH2. While conceptual, this routing method enables to simulate the time lag between a rainfall event and its corresponding peak discharge. This is done through the ordinates of the unit hydrographs, defined by S curves (SH1 and SH2), that distribute the water entering the unit hydrographs over time (Figure C.2, Appendix C) (Perrin et al., 2003, 2007). Nevertheless, these unit hydrographs can be related to physically based hydrological fluxes, being respectively baseflow and quick runoff.

For this project, the GR4H model does not present any snow routine as it was not yet ready on the hourly time step. This should however not be problematic, considering that the snowmelt water contribution to the total discharge does not exceed 15% for the sub-catchments selected. Nonetheless, simulations were analysed carefully for periods during which the contribution of snowmelt water was relatively high.

Numerous intercomparison projects including the GR4H model have lead to the conclusion that the latter performed satisfactorily on average and that an increase in the number of parameters did not show significant improvements of the streamflow simulations quality (Perrin et al., 2003; Mathevet, 2005; Gorgen et al., 2010; Lerat et al., 2006; Perrin, 2000). Nonetheless, several comparisons have also shown that more flexible models perform better than the GR4H model (Demirel et al., 2013; Esse, 2012; Esse et al., 2013; Pushpalatha et al., 2011).

The GR4H model is not used as thus at IRSTEA for operational forecasting purposes. Instead, the forecasting model used nowadays, called GRP, was born from several research activities which lead to the modification of the original GR4H structure, as well as the addition of several assimilation procedures to the latter (Tangara, 2005; Berthet, 2010).

3.3 OpenStreams models

OpenStreams is a project initiated by Deltares. Still in progress, this project's aim is to enhance hydrological models (or model components) re-use in different frameworks. The models or parts of models can be used independently through a command line interface or they can be associated through the use of a layer inherent to OpenStreams, which aim is to execute a standard interface that can process different programming languages (Schellekens, 2012).

Within OpenStreams there is wflow, a distributed hydrological modelling platform. The wflow models are written as a set of PCRaster python scripts imbricated into each other (Wesselung et al., 1996).

The maps, input data, parameters and models states needed for the computation of the models are organized within folders, they constitute the models 'database'.

The computation of the models is done over grid cells, using static PCRaster maps. For the purpose of this project, these maps have a 1 km by 1 km resolution. They are created using a digital elevation model (dem) and a land use map of the considered basin. These maps include: a catchment delineation map, dem maps, a map of the gauges, the original land use map (the land use classes are the following: infrastructures, arable land/pastures/permanent crops, forested areas, other nature, wetlands and water bodies), a local drainage direction (ldd) map, a soil map, a rivers map, a sub-catchment delineation map and maps of the spatial distribution of physical parameters used for the models computations.

Input data are hourly precipitation, potential evapotranspiration and temperature point data interpolated over the maps grid cells.

Parameters are distributed over the catchment in terms of the catchment's soil and vegetation classes.

Model states encompass all the information needed (reservoir contents, fluxes, etc) for starting the model from a selected moment in time. They are kept in the model's memory after each time step, for the following step.

The routing procedure within the wflow models consists of a kinematic wave function (Shultz et al., 2008), which is an approximation of Saint-Venant's equations. This function routes water over the ldd network and depends on several parameters, such as channel width and the manning roughness coefficient.

The wflow platform presently includes the wflow_hbv and the wflow_sbm models; a short description of these distributed hydrological models is given thereafter (Schellekens, 2013).

3.3.1 wflow_hbv

The wflow_hbv model is a fully distributed version of the HBV-96 model, where the original routing function (MAXBAS) was traded for a kinematic wave function (Schellekens, 2013). A complete diagram of the hydrological processes involved in the wflow_hbv model is presented in the Appendix D (Figure D.1), together with a description of the model's variables (Figure D.2, Appendix D). This diagram was produced based on the wflow_hbv python script.

This model, based on nineteen parameters, includes a snow routine.

For this project, the *Seepage*, the *ForecQ_qmec* and the *Inflow* fluxes were deactivated.

The interception values of the wflow_hbv model were involuntarily set to zero prior

to the model simulations and hindcasts. This was discovered at a late stage of the project, disabling any additional model runs. Therefore, the wflow_hbv model results were analysed but did not permit any reliable comparison with the other model results.

Due to their recentness, there is no published literature referring to the wflow_sbm and the wflow_hbv models performances, apart from the wflow documentation (Schellekens, 2013).

3.3.2 wflow_sbm

The wflow_sbm model originates from the distributed topog_sbm simple bucket model developed by Vertessy and Elsenbeer (Vertessy et al., 2000) in the sense that it follows the same soil modelling concept. The following description of the wflow_sbm model is based on the wflow documentation (Schellekens, 2013) and on the wflow_sbm python script from which a diagram of the processes inherent to the model was made (Figures D.3 and D.4, Appendix D).

The originality of the wflow_sbm model lies in its direct runoff component. Indeed, the precipitation water that does not infiltrate into the soil (*FreeWaterDepth*) runs off directly onto the soil's surface (similar to an overland flow) and contributes to the *Kinematic wave*.

The water that infiltrates reaches the soil water reservoir which consists of a bucket composed of an unsaturated store (*UStoreDepth*) and a saturated store (*FirstZoneDepth*), working intrinsically (Vertessy et al., 2000). Part of the water present in the saturated store seeps to the surface (*ExfiltWater* and *SubCellGWRunoff*) and is added to the *Kinematic wave*. The other part of the saturated store water content moves horizontally between cells (*FirstZoneFlux*) based on a D8 network [1], acting as a groundwater flow (Schellekens, 2013).

Part of the surface runoff water can reinfiltrate (*Reinfilt*) in the unsaturated store.

This model, dependent on twenty parameters, includes a snow routine.

For the purpose of this project, the *Inflow* and the *Reinfilt* fluxes were deactivated.

[1] D8 network principle: water flow from a cell to the surrounding cells is controlled by the slope between the two cells. The slope is determined from a ldd network, itself based on a digital elevation model DEM (Kiss, 2004).

Chapter 4

The forecasting platform

For the purpose of this study, the simulations and forecasts were conducted using the forecasting platform Delft-FEWS (Flood Early Warning System). It is used both operationally and in stand alone and has applications in about forty operational centers and eight different hydrological fields [1].

The main initial objective of its development was to offer a flexible platform where data and models could be collected for operational forecasting. Its originality relies essentially on its flexibility; being data-centric, this platform provides the possibility to store data on a database, while hydrological modelling is handled by external models linked to the platform.

Delft-FEWS, its current form introduced in 2002-2003, produces forecasts through a series of steps defined by XML files: data import and storage, data processing (accumulation, aggregation, merging of data, etc), model execution (through general adapters) and finally data export (Werner and Weerts, 2012; Deltares, 2010; Werner et al., 2013). All these steps can be processed as workflows using the platform's user interface (Figure E.1, Appendix E), which also allows for the selection and visualization of data as maps and graphics (Figure E.2 and Figure E.3, Appendix E).

This platform, due to its flexibility, offers an objective environment for the present models intercomparison project.

[1] Further information is provided on the following webpage: http://oss.deltares.nl/web/delft-fews/fews-maps.

Chapter 5

Methods

5.1 From GR4H to wflow_gr4

In order to be executable within Delft-FEWS, the GR4H model had to be incorporated in the wflow platform according to a pre-defined framework (Schellekens, 2013).

Thus, the original FORTRAN code was first reprogrammed in the PCRaster Python dynamic language. Subsequently, the wflow_gr4 model's database (static maps, initial states, parameters) was created, prior to the model's link to the platform.

Creation of the static maps

A set of static maps with a resolution 1 km by 1 km was first generated for each of the five sub-basins using as initial maps the dem and land use maps of the wflow_hbv and wflow_sbm models for the Meuse River basin.

Creation of the initial states

Initial states (the two unit hydrographs water content, the routing store and the production store water content) were then obtained for each sub-basin for the wflow_gr4 model by running simulations on the command line using one week of forcing precipitation and evaporation data. These states were used at the start of the simulation period as cold initial states by the wflow_gr4 model within Delft-FEWS.

Creation of the parameters - Calibration

A set of parameters (X1, X2, X3 and X4) for each of the five sub-basins was finally obtained through the wflow_gr4 model calibration, performed using a proprietary software developed at IRSTEA, called AirGR. The calibration was based on observed hourly discharge measurements using the Nash-Sutcliffe model efficiency

coefficient (NSE) as objective function (formula proposed in Nash and Sutcliffe (1970)).

The sources of the forcing data used for the calibration are shown in Table 5.1, presented underneath.

Table 5.1: Forcing data of the wflow_gr4 model calibration

Sub-basins	Precipitation	Evaporation	Discharge
The Meuse in Saint-Mihiel	Reanalysis data: Pluviometer 1 hr + 24 hrs + 24 radars		Hourly from the Banque Hydro[a]
The Meuse in Stenay			
The Semoy in Haulmé	SAFRAN daily + pluviometer hourly = SAFRAN hourly		Hourly from the Banque Hydro
The Chiers in Carignan			
The Ourthe in Tabreux	Hourly, retrieved from FEWS-Rivieren by Willem van Verseveld (Deltares)		

[a] http://hydro.eaufrance.fr/

The AirGR software follows a calibration procedure structured in two steps. First, a prefiltering step is performed based on a given set of parameters. Here, three parameter sets were provided. At the end of this step, the parameter set offering the best fit is chosen as a starting point for the second calibration procedure. This second step consists of a steepest descent local search. By changing parameters individually, this algorithm aims at producing the best fit for the preselected calibration coefficient. When the pace between two iterations becomes too small, the procedure stops, delivering the last set of parameters produced.

The operational hbv, the wflow_hbv and the wflow_sbm models were acquired already calibrated.

Link to the platform

Finally, the wflow_gr4 model was linked to the Delft-FEWS platform through a general adapter XML file, on the same basis as the other three models, already coupled to the latter prior to the beginning of this research project.

The wflow_gr4 model, although still being lumped, worked on a grid basis, using gridded data as input. The total runoff obtained for each grid cell was then summed in order to obtain the sub-basins discharge at the outlet.

5.2 Simulations

5.2.1 Data import and transformation

Prior to running the model simulations within Delft-FEWS, the data were imported into the platform before being processed. These two steps were indeed primordial for the data availability and usability by the models.

Precipitation (Figure F.1, Appendix F) and temperature data was imported to the platform from the following sources: DWD-Synop, DWD–TTRR, KNMI–Synop, LMW, MSW–Maas, MeteoFrance, REGNIE-HYRAS. The daily potential evaporation was obtained from monthly averages of the Meuse sub-basins.

Subsequently, a data preprocessing step was executed. A concise description of the transformation methods used for the operational hbv model and the wflow models respectively is presented thereafter.

The operational HBV-96 preprocessing scheme

For this model, two precipitation interpolation techniques were used.

- Precipitation
    1. Regnie

       The Regnie tranformation procedure is based on inverse distance interpolation methods. From monthly mean values of daily precipitation of a background grid, hourly areal mean precipitation values assigned to the HBV centroids were obtained[1] (Figure F.2, Appendix F) (Weerts et al., 2008).

    2. Kriging

       The Kriging preprocessing method relies on a linear Kriging interpolation method [2] (Figure F.3, Appendix F). From precipitation synoptic time series, scalar precipitation data assigned to the hbv centroids were computed (Figure F.3, Appendix F).

- Temperature

[1] The allocation of values to hbv centroids is done by the Thiessen polygon method which assigns the weights of each scatter point for each basin (Minnet, 2014)

[2] The Kriging interpolation function is a Gaussian process regression of data interpolation between available values at spatially distributed locations. Several functions exist within the Kriging interpolation method; such as the exponential, Gaussian, linear, spherical and power functions (Minnet, 2014).

The temperature preprocessing was done via a spherical Kriging interpolation method. Temperature time series were transformed in order to obtain hourly temperature data for the hbv centroids (Figure F.4, Appendix F).

- Evaporation

  The evaporation was used as it was imported, without any transformation, by this model.

The wflow preprocessing scheme

- Precipitation

  The output of the operational hbv precipitation transformation was used as a starting point for the wflow precipitation transformation method. The hourly precipitation centroid scalar values were interpolated to grids corresponding to the sub-basins using a linear Kriging interpolation method (Figure F.5, Appendix F).

- Temperature

  For the temperature transformation module, the output of the operational hbv temperature preprocessing was here used again as a starting point for the wflow temperature transformation module. The hourly temperature centroid scalar values were interpolated to grids corresponding to the sub-basins using a spherical Kriging interpolation method Figure F.6, Appendix F).

- Evaporation

  The imported centroid scalar daily evaporation was interpolated to hourly sub-basins grids values using an inverse distance transformation.

5.2.2 Simulations execution

The simulations were executed for all fours models (HBV-96 run twice, once with the Regnie precipitation transformation data and once with the Kriging precipitation transformation data, wflow_sbm, wflow_hbv and wflow_gr4) using the previously transformed input data. These simulations were performed on an hourly basis, with a one day interval, from the 1^{st} of January 2006 at 1 am until the 28^{th} of February 2013 at 11 pm. The year 2006 was chosen as a warm-up year. This daily interval was selected in order to save the models states once a day in the database for subsequently running the hindcasts.

5.2.3 Performance assessment

In order to assess the simulations performance an optimal combination of graphical techniques and performance metrics were computed on the following period: 1^{st} of January 2007 at 1 am until the 28^{th} of February 2013 at 11 pm.

Graphical techniques

A selection of six graphics was made, these were created for each model and each sub-basin and include a plot of:

- Simulated and observed discharge timeseries
- The cumulative discharge timeseries
- The discharge residual timeseries
  - The discharge residual was calculated by substracting the simulated discharge from the observed discharge. Thus, a positive number indicates that the simulated discharge is lower than the observed discharge and vice versa.
- The cumulative residual timeseries
- The simulated versus the observed discharge
- The RMSE, MSE ranking versus the MAE ranking
  - The models were classified in function of their MAE, RMSE and MSE scores (from 1 to 5, 1 representing the model with the best score, 5 the model with the worst score). A dot on the one to one line signifies that the model has the same ranking for those criteria. When a model plots under the one to one line, it suggests that the latter model makes errors of larger magnitude than the model plotting under the line.

These graphics were chosen in order to allow a complete investigation of models performances temporal dynamics to promote the identification of patterns and sources in error occurrence.

Performance metrics

A selection of five continuous performance metrics was made, focusing on different aspects of the simulations. The following criteria were computed for each model and each sub-basin:

- The Mean Absolute Error (MAE):

The MAE measures the average magnitude of the simulation errors with the following formula:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|S_i - O_i| \tag{5.1}$$

With,

N: the number of elements

S_i: the simulated discharge for element i

O_i: the observed discharge for element i

The MAE ranges from 0 to ∞, 0 being the perfect score, and has the same units as discharge.

Being a linear score, all the errors are weighted equally. However, this error metric does not indicate the direction of the deviation from the observed measurements (Koh and Ng, 2009; Sigbritt, 2001).

- The Root Mean Square Error (RMSE):

  The RMSE is also a measure of the average magnitude of the simulation errors, according to the following formula:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(S_i - O_i)^2} \tag{5.2}$$

  The RMSE ranges from 0 to ∞, 0 being the perfect score, and has the same units as discharge.

  Being a quadratic score, the RMSE gives a higher weight to larger errors. Like the MAE, the RMSE does not indicate the direction of the deviation (Koh and Ng, 2009; Sigbritt, 2001).

- The Mean Squared Error (MSE):

  The MSE measures the mean squared difference between the simulations and the observations according to the following formula:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(S_i - O_i)^2 \tag{5.3}$$

  It ranges from 0 to ∞, 0 being the perfect score, and its units are the square of the discharge.

As well as the RMSE, this metric is a quadratic score, giving higher weight to the larger errors. Like the MAE and the RMSE, the MSE does not indicate the direction of the deviation (Koh and Ng, 2009).

- The Mean Error (ME):

  The ME measures the average forecast error with the following formula:

$$ME = \frac{1}{N}\sum_{i=1}^{N}(S_i - O_i) \tag{5.4}$$

  It ranges from -∞ to ∞, 0 being the perfect score, and has the same units as discharge.

  The ME indicates the direction of the deviations. A positive score signifies that the total simulated discharge is overestimated; reciprocally, a negative value suggests that the total simulated discharge is underestimated. Nonetheless, one has to be careful with this metric, as a perfect score does not suggest a perfect simulation per se, errors might be compensating each other (Koh and Ng, 2009).

Intercomparison framework

Based on the above graphical techniques and performance metrics, an intercomparison framework was established in order to analyse specific characteristics of the models simulations and relate those to models structures (for the following models: HBV-96 using the Kriging interpolation method and the wflow models). This framework is structured in two main parts:

1. Overall statistical models performance assessment:
   (a) Error magnitude:
      i. Mean absolute error
      ii. Root mean square error
      iii. Mean square error
      iv. Graphic of RMSE, MSE ranking versus MAE ranking
   (b) Under or overestimation of total discharge:
      i. Graphic of cumulative discharge timeseries
      ii. Graphic of cumulative residual timeseries
      iii. Mean error

2. Model dynamics performance assessment: peakflows and low flows:
    i. Graphic of simulated and observed discharge timeseries
    ii. Graphic of discharge residual timeseries
    iii. Graphic of simulated versus the observed discharge

This same framework was used to assess the effect of data transformation on model simulations, by comparing the simulated discharge of the HBV-96 model based on the Kriging interpolation method (HBV Kriging) with the HBV-96 model based on the Regnie interpolation method (HBV Regnie).

As a result of the simulations performance assessment a higher simulated streamflow quality could be expect from the physically based distributed models (wflow_hbv and wflow_sbm) in contrast to the lumped models (HBV-96 and wflow_gr4).

5.3 Hindcasts

5.3.1 Data import and transformation

Prior to running the hindcasts, precipitation and temperature COSMO-LEPS data were imported into the Delft-FEWS platform and interpolated according to the Kriging interpolation method described previously (see Section 5.2.1).

Subsequently, the interpolated precipitation was perturbed in order to obtain fifteen different precipitation scenarios. These scenarios were obtained by adding and multiplying the original COSMO-LEPS interpolated precipitation data with a varying factor. These ensembles permitted the production of probabilistic hindcasts.

5.3.2 Hindcasts execution

The hindcasts were executed for the four models (HBV-96, wflow_sbm, wflow_hbv and wflow_gr4) on an hourly basis, from the 5^{th} of January 2011 at 1 am until the 20^{th} of January at 1 am. The hindcasts were produced with a lead time of five days.

5.3.3 Error correction

Furthemore, an error correction procedure was applied to the models hindcasts. This correction procedure, called autoregressive (AR) error correction method, is

an automatic algorithm that combines the simulated discharge, the observed discharge and the forecast discharge time series of a specific location in order to update the forecast discharge for each time step. By establishing the residual between the observed and the forecast discharge, the so called model error, this algorithm predicts the future error in order to correct for it (Minnet, 2014; Broersen and Weerts, 2005).

5.3.4 Performance assessment

For an optimal determination and comparison of the models performance in forcasting mode, a combination of two performance metrics, computed over the entire hindcast period using the Ensemble Verification Software (EVS), were used to analyse several quality attributes of the probabilistic forecasts(Brown et al., 2010).

- The Mean Continuous Ranked Probability Skill Score ($\overline{CRPSS}$):

  The $\overline{CRPSS}$ is a measure of the ensemble skill. It computes the average square error of a probability forecast for the entire time series according to the following formula:

  $$\overline{CRPSS} = \frac{\overline{CRPS}_{REF} - \overline{CRPS}_{MAIN}}{\overline{CRPS}_{REF}} \tag{5.5}$$

  Where REF refers to the reference forecast and MAIN refers to the forecasts to be evaluated. The EVS uses as default a sample climatology for the reference forecast. This sample climatology corresponds to the average of the forecasts to be evaluated.

  The $\overline{CRPS}$ is calculated as follows:

  $$\overline{CRPS} = \frac{1}{N}\sum_{i=1}^{N}\int_{x=-\infty}^{x=+\infty}(F_i^f(x) - F_i^o(x))^2 dx \tag{5.6}$$

  With,

  N: the number of elements

  $F_i^f(x)$: the forecast probability cumulative distribution function for element i

  $F_i^o(x)$: the observation cumulative distribution function for element i

  The $\overline{CRPSS}$ ranges from -∞ to 1, 1 being the best score, and does not have any units. If the $\overline{CRPSS}$ is negative, it signifies that the reference fore-

casting system gives a better forecast discharge than the forecasting model (Brown et al., 2010; Sigbritt, 2001).

The $\overline{CRPSS}$ gives indications regarding the limits of predictability of a forecasting system (when this one reaches a score of 0) as well as it is a measure of the reliability, of the resolution and of the uncertainty of the considered hindcasts ($\overline{CRPSS} = reliability + resolution - uncertainty$). Where **reliability** defines the degree of agreement between the frequency of occurrence of an event and the forecast probability; **resolution** refers to the ability of the forecast to resolve a set of observed events into subsets with different frequency distributions; and **uncertainty** is the distribution of the observed discharge (Sigbritt, 2001; Koh and Ng, 2009).

- The forecast errors by observed value:

  The forecast errors are a measure of the ensemble distribution. They are measured according to the following method.

  Firstly, for each observed value, the hindcasts ensembles are pooled in ten different bins, defined by deciles of forecast probability (from 0 to 1). Thereafter, for each bin, the forecast error is calculated by substracting the observed discharge from the ensemble member and plotted against the ascending observed discharge.

  This graphic displays the ensemble distribution of the errors between the hindcast discharges and the observed discharges, shown by the vertical spread of the forecast errors; as well as it gives an overview on the ensemble mean of the hindcasts (over or underestimation of the observed flow).

Intercomparison framework

An intercomparison framework was established on the previously described metrics. This was done in order to assess and compare the forecasting performance of the HBV, wflow_gr4, wflow_hbv and wflow_sbm models (with and without the error correction algorithm). This framework is here again structured in two parts:

1. Ensemble skill:
   i. Graphic of the $\overline{CRPSS}$
2. Ensemble distribution:
   i. Graphic of the forecast errors by observed value

Considering that the HBV-96 model is used operationally within the Delft-FEWS platform, with the previously described error correction procedure, for flood forecasting at Deltares a higher quality of hindcasts could be expected from the latter

model. The other models not being used at all operationally, or not as thus in the case of the wflow_gr4 model (see Section 3.2).

The conjunction of models performance assessment in simulation mode and in forecasting mode thus enabled the creation of a complete picture of the models' behaviour.

Chapter 6

Results

6.1 From GR4H to wflow_gr4

The implementation of the wflow_gr4 model within the wflow platform is presented on the following website: http://schj.home.xs4all.nl/html/.

Creation of the parameters - Calibration

A set of parameters was obtained for each Meuse subcatchment previously selected, according to the calibration procedure described in Section 5.1 (see Table 6.1).

As seen in the following Table, the calibration was performed on different periods of time for each subcatchment, depending on the forcing data availability. For the Ourthe in Tabreux, a calibration period overlapping the simulation period was selected due to the unavailability of hourly data for this subcatchment prior to 2002.

The calibration gave high Nash-Sutcliffe coefficients, the lowest result being of 0.87 for the Ourthe in Tabreux. The calibration results are illustrated with plots of simulated and observed discharge for the run calibration period, presented in the Appendix G (Figures G.1 to G.5).

Table 6.1: Calibration periods and parameters used in the wflow_gr4 model for the five sub-catchments

Sub-basin name	Calibration period		Calibration parameters				NSE
	Warm up	Run	X1 (mm)	X2 (mm)	X3 (mm)	X4 (hour)	
The Semoy in Haulmé	1995 - 1996	1996 - 2005	156.8	0.33	212.7	51.2	0.89
The Chiers in Carignan	1995 - 1996	1996 - 2005	357.8	-0.2	225.9	43.8	0.9
The Meuse in Saint-Mihiel	1987 - 1997	1997 - 2007	257.8	-0.53	171.8	93.5	0.89
The Ourthe in Tabreux	2002 - 2005	2005 - 2011	143.3	-0.77	270.4	23.2	0.87
The Meuse in Stenay	1987 - 1997	1997 - 2007	268.9	-0.66	307	152.6	0.92

6.2 Simulations

In this section, the models simulations obtained for each sub-basin with the four different models (HBV with the Kriging and the Regnie interpolation methods, wflow_gr4, wflow_hbv and wflow_sbm) will be analysed according to the inter-comparison framework described in Section 5.2.3.

6.2.1 Overall statistical model performance assessment

Error magnitude

Looking at Table 6.2, presenting the mean absolute errors based on the model simulations for the five sub-basins, the following observations can be made.

On average, the wflow_gr4 and the HBV Regnie models give the lowest mean absolute error, while the wflow_hbv model shows the highest mean absolute error.

Therefore, the wflow_gr4 and the HBV Regnie models produced overall a simulated discharge closer to the observed discharge than the other models and the wflow_hbv model gave a simulated discharge further off from the observed discharge than the other models.

Furthermore, we can conclude that the HBV-96 simulations with the Regnie interpolation method are in most cases closer to the observed discharge than the HBV Kriging simulations.

Table 6.2: Mean absolute error in m^3/s from Equation 5.1 for the different models

Sub-basins	HBV Kriging	HBV Regnie	wflow_gr4	wflow_hbv	wflow_sbm
The Meuse in Saint-Mihiel	12.5	10.6	11.8	19.0	13.1
The Meuse in Stenay	19.1	15.2	15.5	25.1	17.5
The Chiers in Carignan	7.0	7.4	9.9	17.9	12.2
The Semoy in Haulmé	11.7	10.9	8.5	9.9	15.2
The Ourthe in Tabreux	7.4	6.3	6.3	13.1	8.0

Subsequently, by comparing the root mean square errors (Table 6.3) and the mean square errors (Table 6.4) with the mean absolute errors (Table 6.2), further conclusions can be made regarding the error magnitude of the simulations for each model and each sub-basin (Figure 6.1).

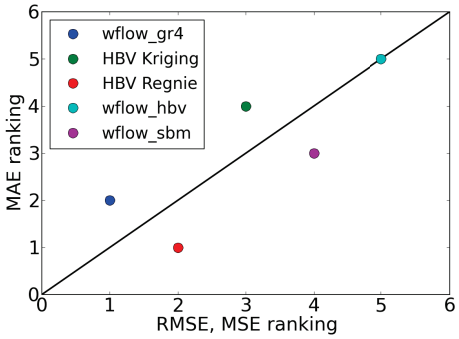
Figure 6.1 compares the ranking of each model for the RMSE, MSE and the MAE criteria. Overall, we can conclude that the wflow_sbm model made errors of large magnitude as it plots high on the RMSE, MSE ranking. On the contrary, the wflow_gr4 and the HBV Regnie models are on average relatively low on the RMSE, MSE ranking; suggesting that they made errors of smaller magnitude than the other models.

Table 6.3: Root mean square error in m^3/s from Equation 5.2 for the different models

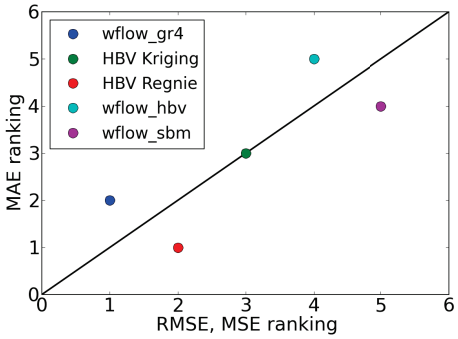
Sub-basins	HBV Kriging	HBV Regnie	wflow_gr4	wflow_hbv	wflow_sbm
The Meuse in Saint-Mihiel	24.3	20.9	19.4	24.6	26.1
The Meuse in Stenay	31.7	26.1	25.5	35.7	32.3
The Chiers in Carignan	12.5	12.4	14.0	21.4	18.8
The Semoy in Haulmé	21.6	19.7	18.8	18.4	29.2
The Ourthe in Tabreux	14.8	12.2	11.7	18.2	25.3

Table 6.4: Mean square error in m^6/s^2 from Equation 5.3 for the different models

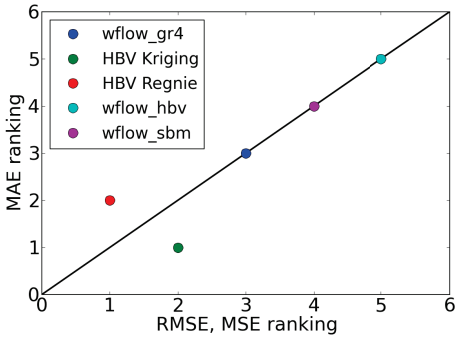
Sub-basins	HBV Kriging	HBV Regnie	wflow_gr4	wflow_hbv	wflow_sbm
The Meuse in Saint-Mihiel	590.5	436.6	374.5	603.3	681.4
The Meuse in Stenay	1004.3	680.8	651.2	1275.3	1041.8
The Chiers in Carignan	155.8	153.6	196.6	459.9	354.6
The Semoy in Haulmé	466.9	386.4	353.9	337.5	855.2
The Ourthe in Tabreux	220.1	149.9	137.4	329.7	642.3



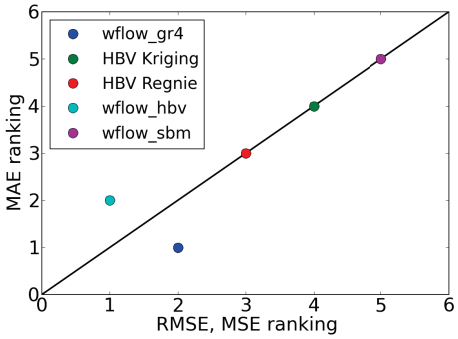
(a) The Meuse in Saint-Mihiel



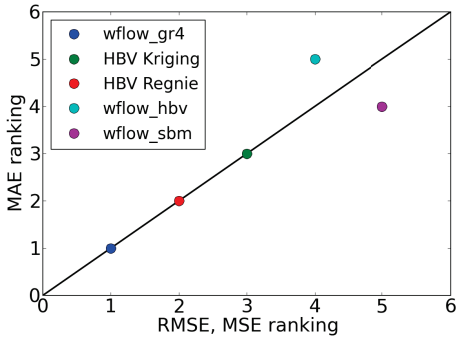
(b) The Meuse in Stenay



(c) The Chiers in Carignan



(d) The Semoy in Haulmé



(e) The Ourthe in Tabreux

Figure 6.1: Plots of the RMSE/MSE ranking versus the MAE ranking

Under- or overestimation of total discharge

From the figures of the cumulative discharge and the cumulative residual (Figures H.1 to H.5, Appendix H), as well as the mean error table (Table 6.5), the overall under- or overestimation of total discharge can be identified for each model and each sub-basin.

The HBV-96 model simulations, both with the Kriging and the Regnie transformation methods, show negative mean error values for all sub-basins except for the Chiers in Carignan, suggesting that they underestimated the total discharge for those four sub-catchments for which they present a negative mean error.

The wflow_gr4 model and the wflow_sbm model show similar results. The total simulated discharge was underestimated for the Meuse in Saint-Mihiel, the Meuse in Stenay and the Semoy in Haulmé, for which they both present negative mean errors; and overestimated for the Chiers in Carignan and the Ourthe in Tabreux, presenting positive mean error values.

Finally, the wflow_hbv model displays large positive mean error values for all sub-basins except for the Semoy in Haulmé. This model therefore overestimated largely the total discharge for all the sub-basins, except for the Semoy in Haulmé for which it underestimated it slightly. Overall, the wflow_hbv model shows the largest over- or underestimations of total discharge.

Given the positive mean error values for the Chiers in Carignan, surpisingly all the models show an overestimation of the total discharge for this sub-basin.

Table 6.5: Mean error in m^3/s from Equation 5.4 for the different models

Sub-basins	HBV Kriging	HBV Regnie	wflow_gr4	wflow_hbv	wflow_sbm
The Meuse in Saint-Mihiel	-6.8	-4.0	-0.4	12.3	-1.4
The Meuse in Stenay	-14.4	-10.2	-4.2	14.3	-6.4
The Chiers in Carignan	2.2	4.1	8.8	16.7	8.5
The Semoy in Haulmé	-9.8	-8.6	-1.9	-0.6	-4.4
The Ourthe in Tabreux	-1.0	-3.7	2.2	10.2	2.6

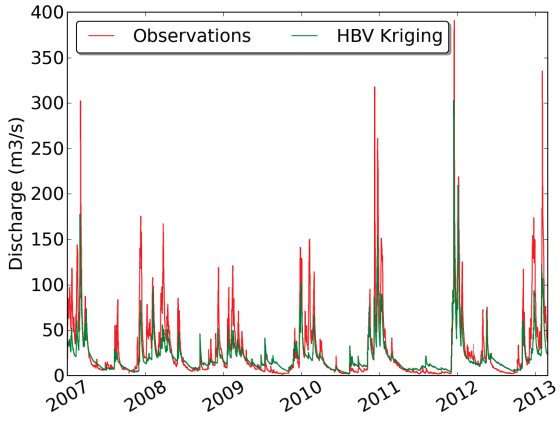
6.2.2 Model dynamics performance assessment: peakflows and low flows

A close look at model simulations will be carried out in order to identify the patterns in under or overestimation of total discharge for each model by characterising their behaviour during high and low flows respectively. For this purpose, the simulations for the Meuse in Saint-Mihiel were used as representative for the general model performances.

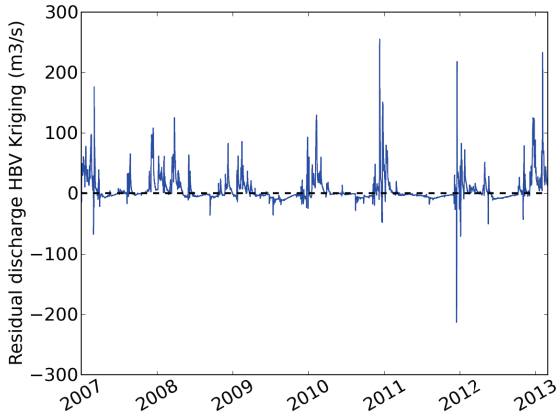
HBV Kriging model simulations results

From Figure 6.2a, presenting the observed and the simulated discharge timeseries of the HBV Kriging model for the Meuse in Saint-Mihiel, a general picture of this model behaviour during both high flows and low flows can be drawn. In a first place, it can be seen that the overall observed discharge dynamic was seemingly well captured by the HBV Kriging model. Nonetheless, several dissimilarities between the modelled and the observed discharge are distinguishable. For the last four low flow events shown in Figure 6.2a, from the year 2009 until the year 2012 included, the HBV Kriging model produced a slightly higher discharge than the observed discharge. This suggests that this model tends to overestimate low flows. Moreover, all the observed peakflows present in Figure 6.2a show a simulated discharge lower than the observed discharge. This is especially noticeable for the three high flow extreme events of the year 2011 until 2013 included. The HBV Kriging model therefore appears to be underestimating high flows.

Figure 6.2b depicts the discharge residual of the HBV Kriging model for the Meuse in Saint-Mihiel. When correlating the results given by the latter figure with the results previously described of Figure 6.2a, a complete image of the HBV Kriging model behaviour during high and low flows, as well as the magnitude of the under- and overestimations made by the model can be obtained. Several positive residual peaks, ranging from 100 m^3/s until 2010, to around 250 m^3/s for the last three years of simulations (2010 to 2013), can be appreciated in Figure 6.2b, corresponding to the high flow events of Figure 6.2a. This suggests that the HBV Kriging model underestimated the discharge during those peakflow events. Furthermore, a large negative residual peak reaching a value of approximately -200 m^3/s can be distinguished at the end of the year 2011, signifying that the model overestimated the observed discharge for this specific event. This might be due to the timing of the peak, indeed, the model simulated a peak either too early or too late compared to the observed peak. Regarding low flows, marginally negative residual values are observable in Figure 6.2b, corresponding to the low flow events of Figure 6.2a, suggesting a slight overestimation of the observed discharge by the HBV Kriging model during low flow conditions.



(a) Simulated and observed discharge timeseries



(b) Discharge residual timeseries, $Q_{obs} - Q_{sim}$

Figure 6.2: HBV Kriging model results for the Meuse in Saint-Mihiel. (a) Simulated and observed discharge timeseries. (b) Discharge residual timeseries

Figure 6.3 presents the simulated versus the observed discharge for the HBV Kriging model for the Meuse in Saint-Mihiel. The discharge corresponding to low flow events appear to be plotting imperceptibly above the one to one line on the latter figure. This indicates that the HBV Kriging model slightly overestimated the discharge during observed low flow conditions. The larger the discharge becomes, the broader the scatterplot appears on Figure 6.3, suggesting that the HBV Kriging model made larger errors for higher discharges. Nonetheless, this scatterplot is positioned principally underneath the one to one line, which shows a tendency of the model to underestimate high flows. Three striking dotted lines, appearing under the one to one line, correspond to high underestimations of peakflow events by the model. These might be correlated to the three last extreme peakflow events depicted in Figure 6.2a, which were underestimated by the HBV Kriging model. Additionally, a striking peak can be appreciated above the one to one line, reaching

a simulated discharge of about 320 m^3/s while the observed discharge was approximately 100 m^3/s. This event was thus overestimated by the model and can be linked to the highly negative residual peak observed in Figure 6.2b and associated to a timing error of the discharge peak.

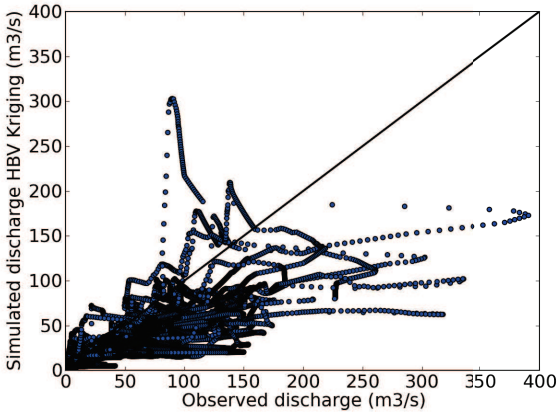


Figure 6.3: Simulated versus observed discharge of the HBV Kriging model for the Meuse in Saint-Mihiel

Looking at the discharge simulations for the event of January 2011 (Figure 6.4), an important event in the River Meuse flooding history, the previously described HBV Kriging simulation characteristics can be observed. Indeed, the dynamics of low flows was well captured by the HBV Kriging model. Nonetheless, the model appears to have overestimated the discharge slightly during low flow conditions. Furthermore, the peakflows were all underestimated by the model for this specific event, combined with a small underestimation of the peaks time lag.

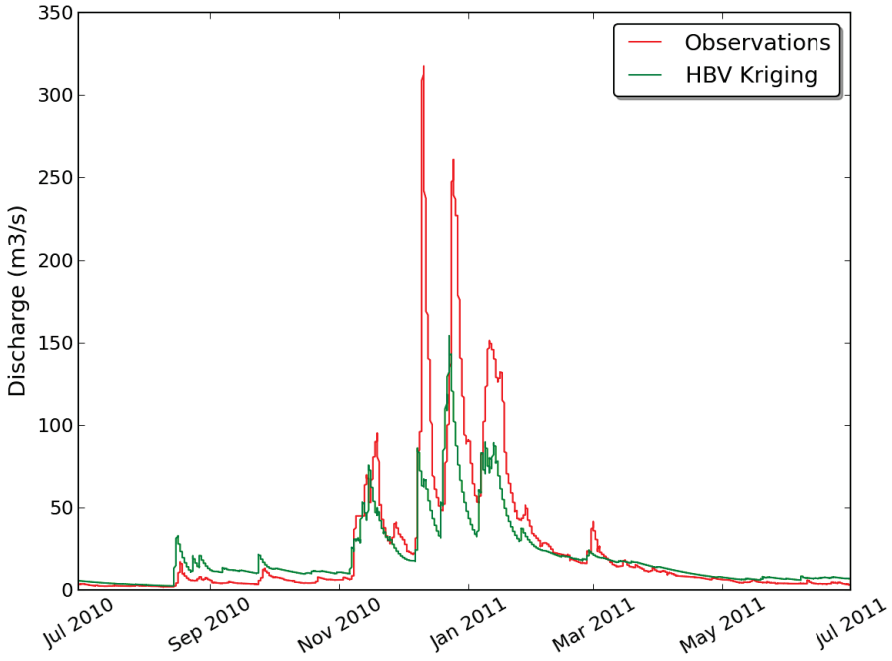


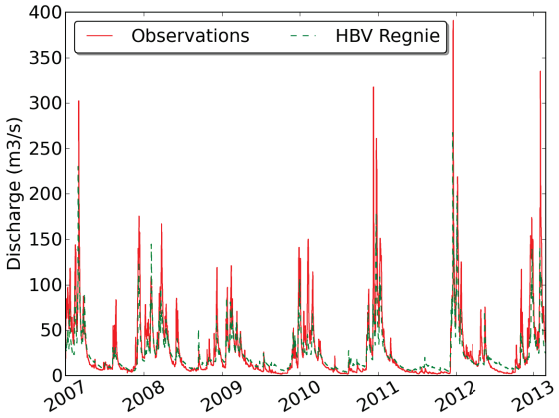
Figure 6.4: Simulated and observed discharge event, HBV Kriging model for the Meuse in Saint-Mihiel

HBV Regnie model simulations results

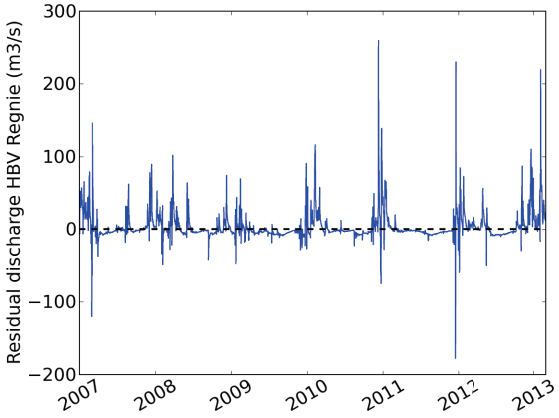
Figure 6.5a illustrates the simulated and the observed discharge timeseries of the HBV Regnie model for the Meuse in Saint-Mihiel. Based on this figure, the observed discharge dynamics appear to be well captured by the model, especially for the low flows, where only small discrepancies can be seen between the simulated and the observed discharge from 2009 until 2013. Contrastingly, almost all peakflows were underestimated by the HBV Regnie model, except for a singular peakflow event situated at the beginning of the year 2008. Similar discharge characteristics were highlighted by Figure 6.2a for the HBV Kriging model simulations.

By correlating Figure 6.5b, depicting the discharge residual timeseries of the HBV Regnie model for the Meuse in Saint-Mihiel, with Figure 6.5a, positive discharge residuals corresponding to observed peakflows on Figure 6.5a are appearent. These residuals range from approximately 100 m^3/s until 2010, to about 250 m^3/s for the last three years of simulations, 2010 to 2013, and correspond to an underestimation of the observed peakflows by the HBV Regnie model. These results are similar to the results for the HBV Kriging model shown in Figure 6.2b. Interestingly, a negative residual of about -200 m^3/s can also be observed at the end of the year 2011, during a high flow event, which was also observed for the HBV Kriging model

(Figure 6.2b) and can as thus be associated to an overestimation of the simulated discharge by the HBV Regnie model due to an error in the timing of the peak by the latter. Additionally, two similar negative residuals of lower magnitude than the previous one can be seen at the beginning of the year 2007 and at the end of 2010, reaching values of approximately -110 m^3/s and -80 m^3/s respectively. These peaks correspond to an overestimation of the observed discharge due to a timing error of the peakflow and were not present for the HBV Kriging model (Figure 6.2b). Therefore, the HBV Regnie model appears to have a higher tendency of time lag under- or overestimation than the HBV Kriging model does. Based on Figure 6.5b, slightly negative residual discharge values can be distinguished and attributed to the low flow periods in Figure 6.5a. The latter observation suggests, such as for the HBV Kriging model, a small overestimation of the observed low flows by the HBV Regnie model.



(a) Simulated and observed discharge timeseries



(b) Discharge residual timeseries, $Q_{obs} - Q_{sim}$

Figure 6.5: HBV Regnie model results for the Meuse in Saint-Mihiel. (a) Simulated and observed discharge timeseries. (b) Discharge residual timeseries

Figure 6.6 presents the simulated versus the observed discharge of the HBV Regnie model for the Meuse in Saint-Mihiel. In this figure, the low flows plot imperceptibly above the one to one line, supporting the previous observations about the slight overestimation of low discharges by the HBV Regnie model, similarly to the HBV Kriging model. Moreover, the width of the scatterplot appears to get broader towards higher observed flows, while it situates itself under the one to one line. This indicates a larger magntiude of the discrepancies between the simulated and the observed discharge for high flows as well as the tendency to underestimate the latter, as pointed out previously. Nevertheless, this scatterplot is situated closer to the one to one line than the scatterplot corresponding to the HBV Kriging model and illustrated in Figure 6.3. Thus, the HBV Kriging model underestimated high discharges more largely than the HBV Regnie model did. The same peaks are hereafter noticeable under the one to one line, for extremely high discharge events, as were observed for the HBV Kriging model (Figure 6.3). These correspond to the underestimated peakflows by the HBV Regnie model presented in Figure 6.5a. Furthemore, several peaks appear above the one to one line, the largest reaching a simulated discharge of about 280 m^3/s, while the observed discharge approximated 100 m^3/s. These peaks can be associated to the negative residuals illustrated in Figure 6.5b, where the HBV Regnie model overestimated the observed flow. This overestimation, also highlighted by Figure 6.3 for the HBV Kriging model, is nonetheless lower for the HBV Regnie model.

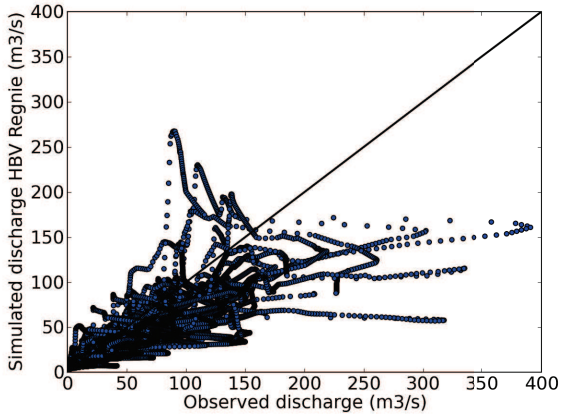


Figure 6.6: Simulated versus observed discharge of the HBV Regnie model for the Meuse in Saint-Mihiel

Figure 6.7 shows an observed and simulated discharge event by the HBV Kriging and Regnie models, for the Meuse in Saint-Mihiel. This figure illustrates all the previous comparative results of the HBV Kriging and Regnie simulations characteristics. Indeed, while the discharge dynamics was similarly well captured by both models and while both models equally marginally overestimated low flows; the HBV Kriging model appears to have underestimated peakflows more largely

than the HBV Regnie model did. This was found in Section 6.2.1, where the HBV Regnie model produced less errors and of lower magnitude on average than the errors produced by the HBV Kriging model, suggesting that the latter model made larger more frequent errors, corresponding to discrepancies between observed and simulated discharge during peakflows.

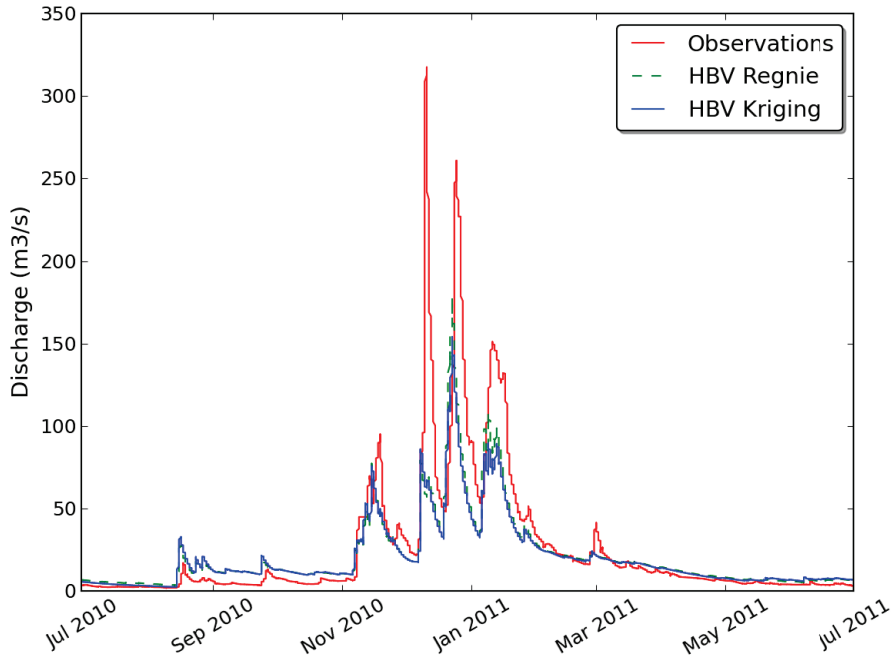


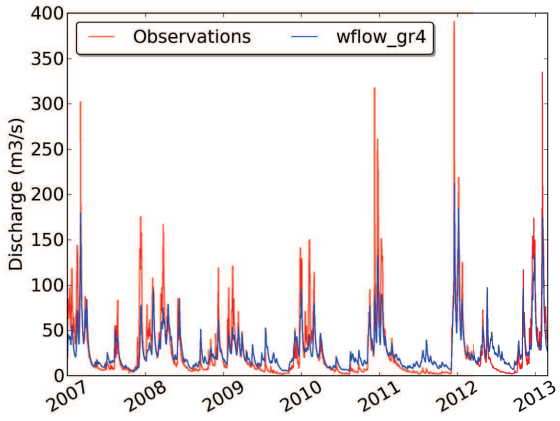
Figure 6.7: Simulated and observed discharge event, HBV Kriging and Regnie models for the Meuse in Saint-Mihiel

wflow_gr4 model simulations results

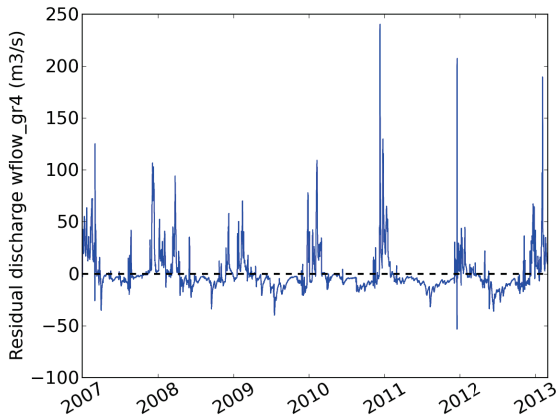
Figure 6.8a presents the wflow_gr4 model simulated discharge and the observed discharge timeseries for the Meuse in Saint-Mihiel. Based on this figure, the wflow_gr4 model simulations present a similar dynamic to the observed discharge. Nonetheless, this model overestimated all the low flow events, while all the peakflows were underestimated, simlarly to the HBV Kriging and Regnie models.

As shown by the discharge residual timeseries of the wflow_gr4 model for the Meuse in Saint-Mihiel (Figure 6.8b), positive residual values, ranging from approximately 100 m^3/s until 2010 to 250 m^3/s from 2011 to 2013, similarly to the HBV models simulations (Figures 6.2b and 6.5b), correspond to high observed discharges present in Figure 6.8a. This suggests that the wflow_gr4 model underestimated high flow events with the same approximate magnitude as the HBV models.

A large negative residual of approximately -50 m^3/s can as well be observed at the end of the year 2011 on Figure 6.8b, during a peakflow event presented in Figure 6.8a. The latter was also observed, with a nonetheless larger magnitude, for the HBV models simulations (Figures 6.2b and 6.5b). It could be due here again to an overestimation of the observed flow by the wflow_gr4 model due an over- or underestimation of the peak time lag. Moreover, Figure 6.8b displays negative residual values of around -10 m^3/s on average, that can be associated to the overestimation of low flows by the wflow_gr4 model. These negative values are larger than the values observed for the HBV models (Figures 6.2b 6.5b), indicating that the wflow_gr4 model made larger errors for low flows than the latter models.



(a) Simulated and observed discharge timeseries



(b) Discharge residual timeseries, $Q_{obs} - Q_{sim}$

Figure 6.8: wflow_gr4 model results for the Meuse in Saint-Mihiel. (a) Simulated and observed discharge timeseries. (b) Discharge residual timeseries

Figure 6.9 illustrates the simulated versus the observed discharge of the wflow_gr4 model for the Meuse in Saint-Mihiel. Low flows plot above the one to one line, more perceptibly than the HBV models simulations (Figures 6.3 and 6.6), sug-

gesting, as stated above, that the wflow_gr4 model overestimated low flows more largely than the HBV models did. Contrastingly, high observed flows are situated under the one to one line, illustrating the wflow_gr4 model tendency to underestimate high flows, as seen in Figures 6.8a and 6.8b. Singular peaks plotting under the one to one line, corresponding to high observed discharge values, correspond to the extremely high discharge events underestimated by the wflow_gr4 model. These characteristics were also noted for the HBV models (Figures 6.3 and 6.6). Furthermore, an isolated event can be seen above the one to one line, reaching a simulated discharge of about 180 m^3/s where the observed discharge reached a value of approximately 130 m^3/s. This event can be correlated to the negative residual pointed out in Figure 6.8b. The latter corresponds to the same observed event as the one overestimated by the HBV models and attributed to an under- or overestimation of the observed peak time lag. Nevertheless, it appears as an isolated event for the wflow_gr4 model, and of lower magnitude than for the HBV models simulations, where this appeared more frequently.

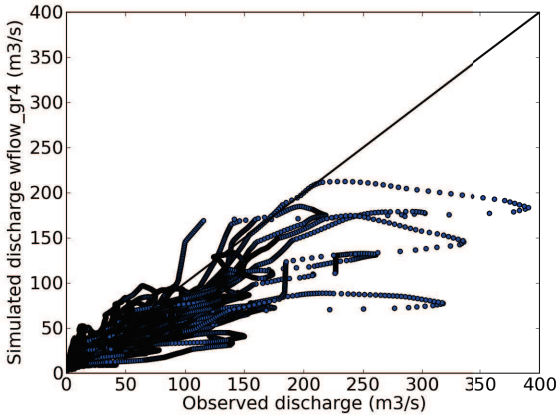


Figure 6.9: Simulated versus observed discharge of the wflow_gr4 model for the Meuse in Saint-Mihiel

Based on the following figure, Figure 6.10, presenting an observed peakflow event simulated by the wflow_gr4 model for the Meuse in Saint-Mihiel, the previously stated characteristics of the wflow_gr4 model simulations can be observed. Low flows were overestimated by the wflow_gr4 model, more significantly than by the HBV models (Figure 6.7) and peakflows were underestimated by the model, such as for the HBV models (Figure 6.7). Nevertheless, the discharge event situated in November 2010 was almost fully grasped by the wflow_gr4 model. Additionally, while the discharge dynamics appeared well captured by the wflow_gr4 model in Figure 6.8a, a contrasting result can be observed hereafter. Indeed, the simulated discharge appears smoothed, while during high flows it exhibits a characteristic bell-shape distribution.

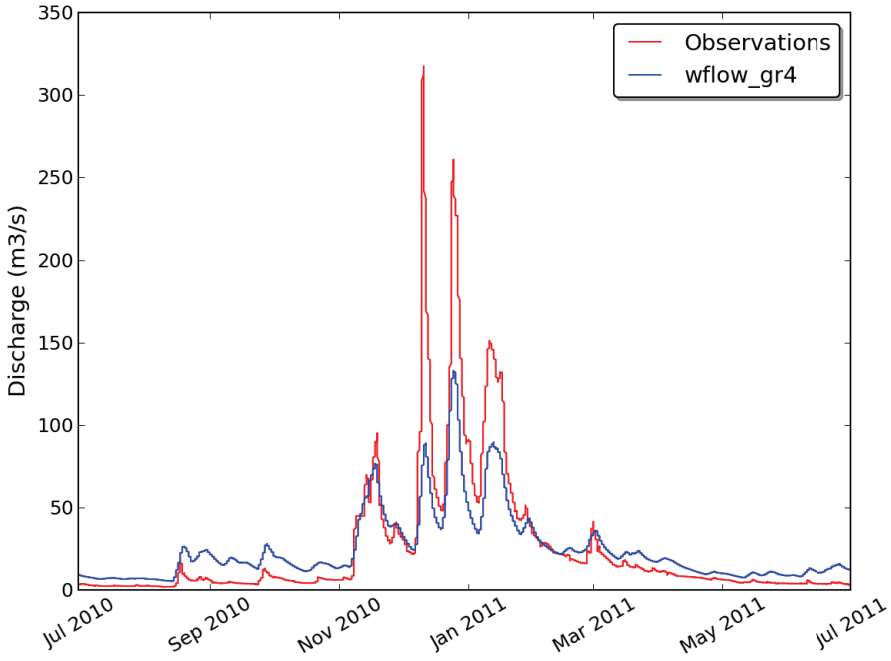


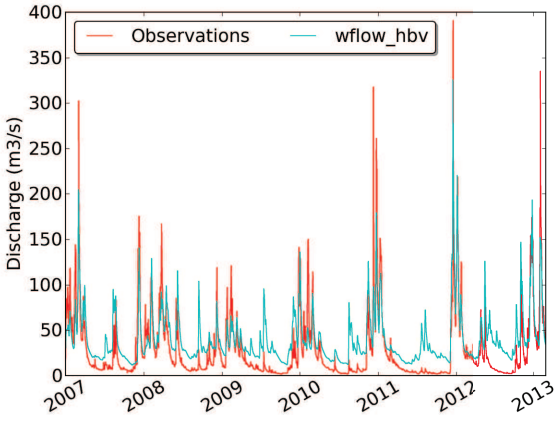
Figure 6.10: Simulated and observed discharge event, wflow_gr4 model for the Meuse in Saint-Mihiel

wflow_hbv model simulations results

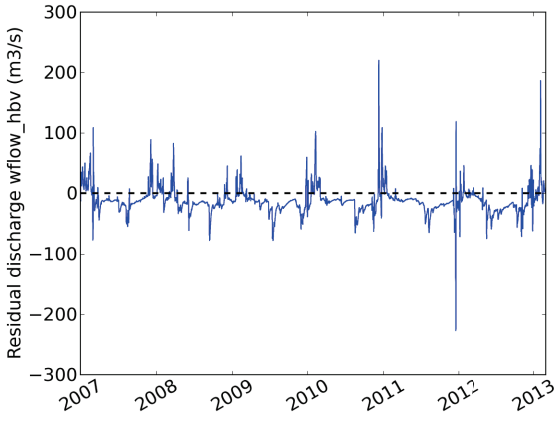
The prevailing information contained in Figure 6.11a depicts the tendency of the wflow_hbv model simulations during high and low flows for the Meuse in Saint-Mihiel. From the latter figure, a high overestimation of the low flows by the wflow_hbv model can be distinguished, coupled to an underestimation of high flows.

From Figure 6.11b, presenting the discharge residual timeseries of the wflow_hbv model for the Meuse in Saint-Mihiel, positive residual peaks of approximately 100 m^3/s until 2010 (included) and of about 220 m^3/s and 190 m^3/s at the end of the year 2010 and at the beginning of the year 2013 respectively, can be observed. These positive values correspond to peakflow events present in Figure 6.11a and therefore suggest that the wflow_hbv model underestimated those peakflows. Nevertheless, those positive residuals presented in Figure 6.11b, display, on average, a lower magnitude than for the previously described models (the HBV and the wflow_gr4 models). This indicates that the wflow_hbv model underestimated the discharge for peakflow events less than the latter models. Furthermore, a large negative residual peak can be observed at the end of the year 2011, reaching a value of approximately -230 m^3/s. The same peak was observed for the HBV models and the wflow_gr4 model, but appears of higher magnitude for the wflow_hbv model,

and could be attributed to an overestimation of the observed flow due to an over- or underestimation of the peak time lag by this model. Additionally, an average negative residual of -50 m^3/s can be remarked, corresponding to the observed low flows of Figure 6.11a. This implies a large overestimation of low flow discharges by the wflow_hbv model, greater than for the HBV and the wflow_gr4 models, as indicated by the negative residual average.



(a) Simulated and observed discharge timeseries



(b) Discharge residual timeseries, $Q_{obs} - Q_{sim}$

Figure 6.11: wflow_hbv model results for the Meuse in Saint-Mihiel. (a) Simulated and observed discharge timeseries. (b) Discharge residual timeseries

Figure 6.12 displays the simulated versus the observed discharge for the wflow_hbv model for the Meuse in Saint-Mihiel. Low flows appear significantly above the one to one line, conveying a characteristical large overestimation of low observed flows by the wflow_hbv model. For high observed flows, data tend to be positioned slightly below the one to one line, suggesting an underestimation of high discharges by the wflow_hbv model. Large peaks are appearant under the one to one line, for high observed discharges, and correspond to peakflow events underestimated by

the wflow_hbv model. Similar peaks were detected for the HBV and the wflow_gr4 models (Figures 6.3, 6.6 and 6.9). In addition, a peak can be observed above the one to one line, reaching a simulated discharge of approximately 340 m^3/s for an observed discharge of 100 m^3/s. This discharge overestimation by the wflow_hbv model could be associated to the large negative residual noticed in Figure 6.11b at the end of the year 2011.

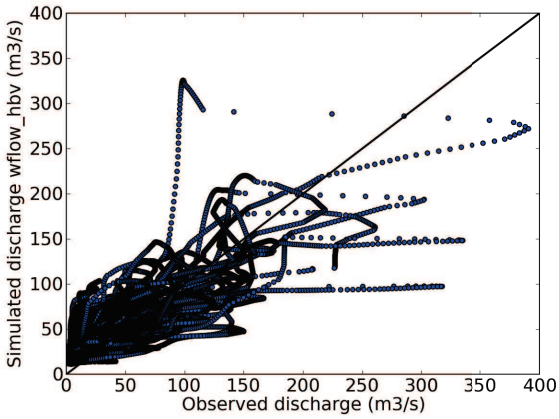


Figure 6.12: Simulated versus observed discharge of the wflow_hbv model for the Meuse in Saint-Mihiel

Figure 6.13 exposes the previously stated characteristics of the wflow_hbv model simulations for the Meuse in Saint-Mihiel. Indeed, while low flows were largely overestimated by the wflow_hbv model, high flows appear underestimated by the latter. Nevertheless, the observable peakflows of this specific event were captured slightly more completely than they were by the HBV and the wflow_gr4 models.

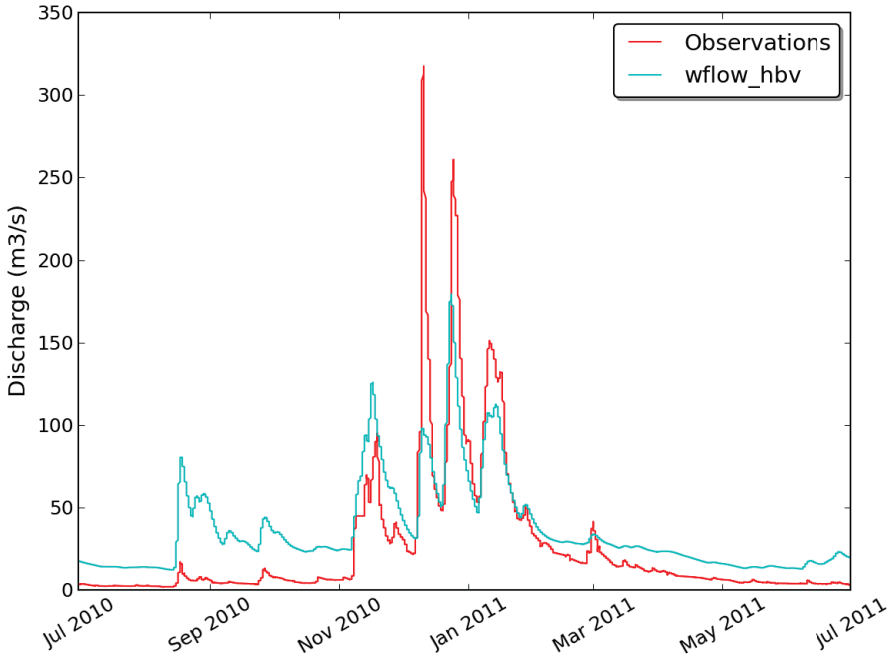


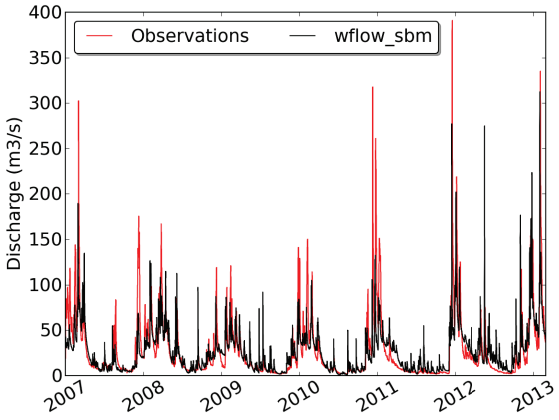
Figure 6.13: Simulated and observed discharge event, wflow_hbv model for the Meuse in Saint-Mihiel

wflow_sbm model simulation results

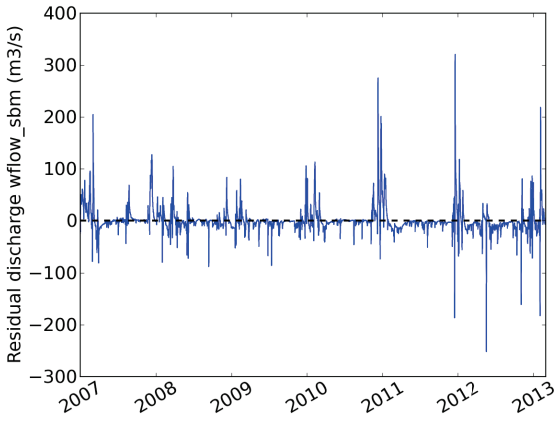
From Figure 6.14a, the simulated and the observed discharge timeseries for the Meuse in Saint-Mihiel, an outline of the wflow_sbm model behaviour during high and low flows can be drawn. Based on this figure, the wflow_sbm model did not appear to capture the observed discharge dynamics. The simulated low flows exhibit peaks where no peaks were observed, while peakflows were underestimated by the wflow_sbm model.

Figure 6.14b presents the discharge residual timeseries of the wflow_sbm model for the Meuse in Saint-Mihiel. From the latter figure, positive residual discharges ranging from 100 m^3/s to approximately 320 m^3/s are observable. These residuals can be correlated to the peakflows presented in Figure 6.14a and thus imply an underestimation of those peakflows by the wflow_sbm model. According to this postive residuals range, the wflow_sbm model displayed the highest underestimation of high flows compared to all the previously described models. Furthermore, multiple singular negative residual peaks, comprised between 0 and -100 m^3/s, on average, are existent. These negative peaks appear during low observed flows, suggesting that the wflow_sbm model overestimated low flows in a non systematic manner, as seen in Figure 6.14a. Larger negative residuals can be seen from the end of the year 2011 until the end of the simulations. These peaks range from

nearly -150 m^3/s to -250 m^3/s and can be associated to an overestimation of the observed discharge. The peak present at the end of the year 2011 was also observed for the previously described models (Figures 6.2b, 6.5b, 6.8b and 6.11b) and could be associated, similarly to the other models, to an under- or overestimation of the peak lag time. Moreover, the negative residual peak with a corresponding value of around -250 ^{3}/s, observed during observed low flows of the summer 2012, seems to be a simulation error.



(a) Simulated and observed discharge timeseries



(b) Discharge residual timeseries, $Q_{obs} - Q_{sim}$

Figure 6.14: wflow_sbm model results for the Meuse in Saint-Mihiel. (a) Simulated and observed discharge timeseries. (b) Discharge residual timeseries

From Figure 6.15, displaying the simulated versus the observed discharge for the wflow_sbm model for the Meuse in Saint-Mihiel, characteristical features of this model simulations can be observed. The figure shows a large spread of the discharge data, with low observed discharges plotting above the one to one line and high observed discharges plotting predominantly under the one to one line. This denotes a tendency of the wflow_sbm model to overestimate low flows and to un-

derestimate high flows in a non regular manner. Singular peaks can be appreciated above the one to one line and can be associated to the negative residual peaks in Figure 6.14b, corresponding to overestimations of the observed discharge. Moreover, peaks can as well be noticed under the one to one line for high observed discharges, similarly to the other models simulations results (Figures 6.3, 6.6, 6.9 and 6.12) and correspond to the underestimation of peakflows previously seen in Figure 6.14b, characterised by postive residuals.

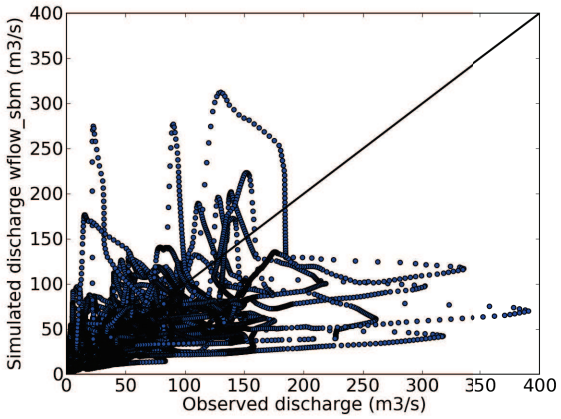


Figure 6.15: Simulated versus observed discharge of the wflow_sbm model for the Meuse in Saint-Mihiel

From Figure 6.16, a simulated and observed event for the wflow_sbm model for the Meuse in Saint-Mihiel, features of the wflow_sbm model simulations can be apprehended. Indeed, this model produced very dynamic simulations, particurly for low observed flows, during which singular peaks were simulated by the wflow_sbm model and led to an average overestimation of the low flows by this model. Furthermore, the peakflow events were largely underestimated by the model and a distinct underestimation of the time lag for peaks can be seen.

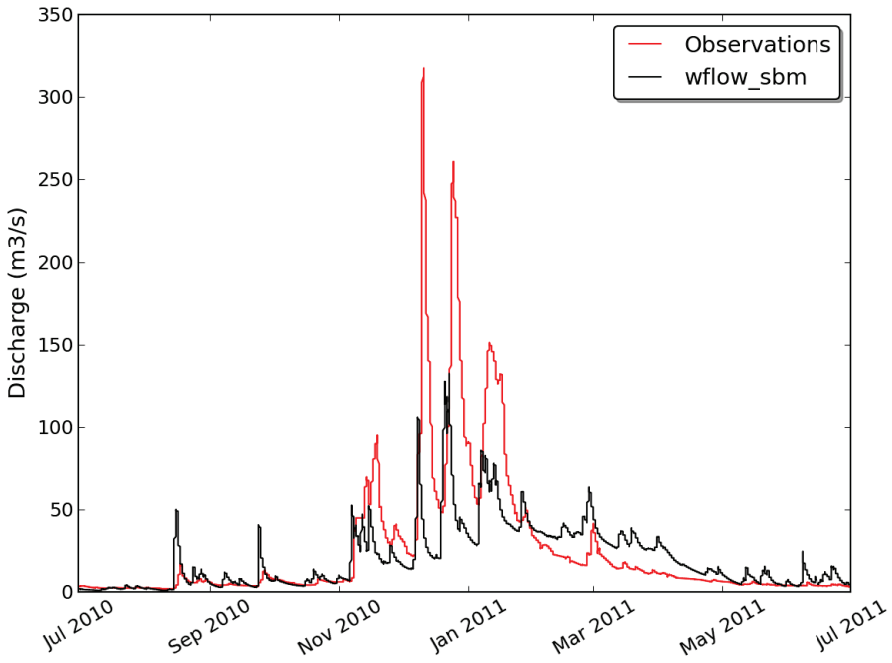


Figure 6.16: Simulated and observed discharge event, wflow_sbm model for the Meuse in Saint-Mihiel

For the HBV Kriging, HBV Regnie, wflow_gr4 and wflow_sbm models, the combination of the underestimation of high flows and the overestimation of low flows clarified the overall underestimation of total discharge seen in Section 6.2.1. Regarding the wflow_hbv model, the overall overestimation of total discharge (Section 6.2.1) can be explained by the slight underestimation of high flows and the large overestimation of low flows.

The Meuse in Saint-Mihiel was used as representative for the general models' dynamics in this Section. Nonetheless, contrasting results were obtained for the Chiers in Carignan, for which all the models overestimated the total discharge.

6.3 Hindcasts

In this section, hindcasts, produced for each sub-basin by the four different models, will be analysed according to the models ensemble skill and distribution.

Figure 6.17 shows the ensemble hindcasts obtained for the Chiers in Carignan with the HBV model without error correction, as an example of the hindcasts produced during this research. Based on this Figure, the HBV model hindcasts ensemble plot

majoritarily above the observed discharge, signifying that the HBV model overestimated the discharge for this specific event for the Chiers in Carignan.

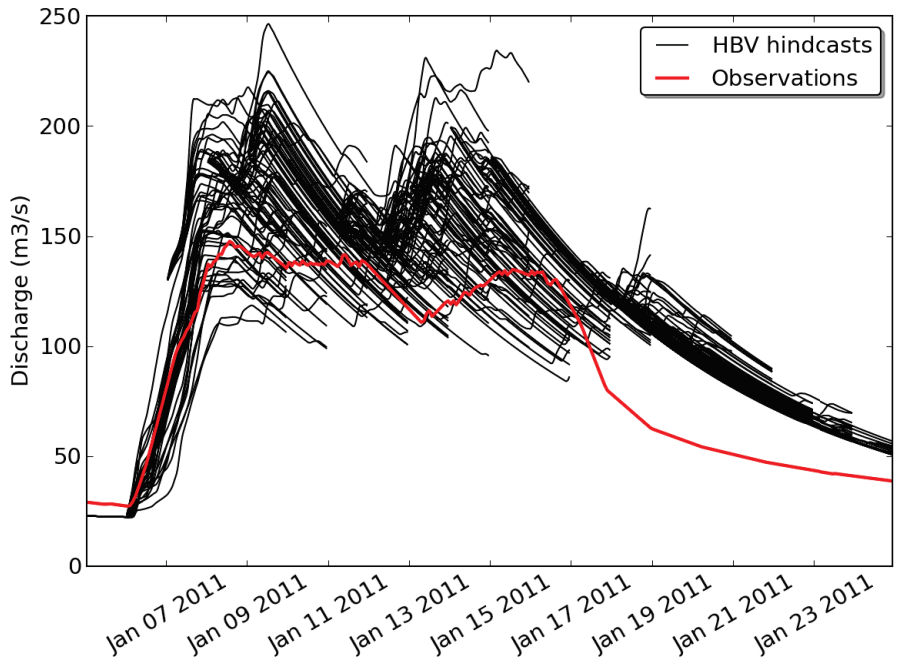


Figure 6.17: HBV ensemble hindcasts (without error correction) for the Chiers in Carignan

6.3.1 Ensemble skill analysis

In a first part, the models overall performance in forecasting mode will be examined. And the effect of the error correction on the limits of their ability to predict a given event will be displayed.

Hindcasts without error correction

Looking at the figures presenting the $\overline{CRPSS}$ for each model by forecast lead time (Figure 6.18), we can observe that all the models show on average a negative $\overline{CRPSS}$, especially for the Meuse in Stenay, suggesting that they present a poor resolution, as well as a poor reliability and a high uncertainty. Moreover, several patterns could be detected when comparing the models performances. Indeed, the wflow_hbv model displays the highest $\overline{CRPSS}$ for the Meuse in Saint-Mihiel (Figure 6.18a) and the Meuse in Stenay (Figure 6.18b). Whereas the HBV model presents the highest $\overline{CRPSS}$ for the Semoy in Haulmé (Figure 6.18d) as well as

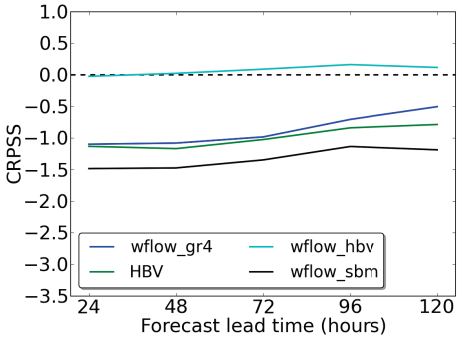
the Ourthe in Tabreux (Figure 6.18e). These models therefore produced the hindcasts with the highest ensemble skill for the respective sub-basins.

Contrastingly, the wflow_sbm model exhibits the lowest $\overline{CRPSS}$ for most of the sub-basins, excluding the Chiers in Carignan. Hence, this model generated the hindcasts with the lowest ensemble skill overall.

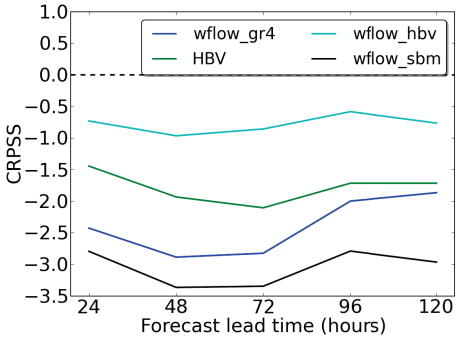
Additionally, the wflow_gr4 model demonstrates a medium $\overline{CRPSS}$ for all the sub-basins, compared to the other three models.

The Chiers in Carignan displayed divergent results from the other sub-basins in the sense that the wflow_gr4 model presents the highest $\overline{CRPSS}$ for this sub-basin. Furthermore, whereas the wflow_hbv and the HBV models display the highest $\overline{CRPSS}$ for the other sub-basins, they exhibit the lowest $\overline{CRPSS}$, on average, for the Chiers in Carignan.

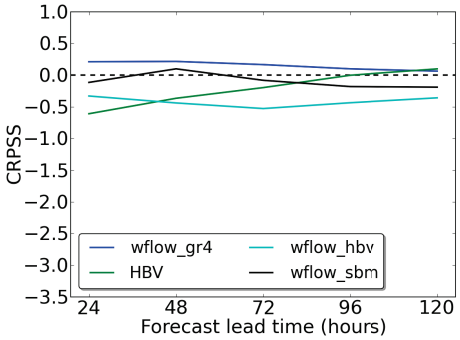
Investigating the $\overline{CRPSS}$ evolution with increasing lead time, no trend is observable. Indeed, a tendency of decreasing ensemble skill with lead time, shown by the decreasing $\overline{CRPSS}$, would have been expected. Nonetheless, this tendency can only be appreciated for two sub-basins, being the Semoy in Haulmé and the Ourthe in Tabreux.



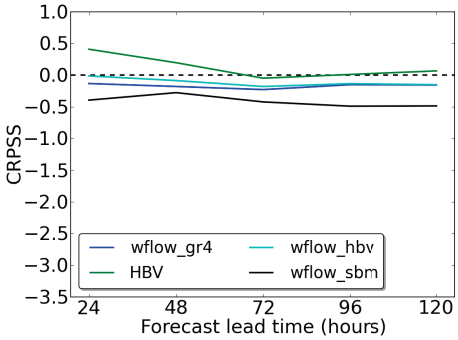
(a) The Meuse in Saint-Mihiel



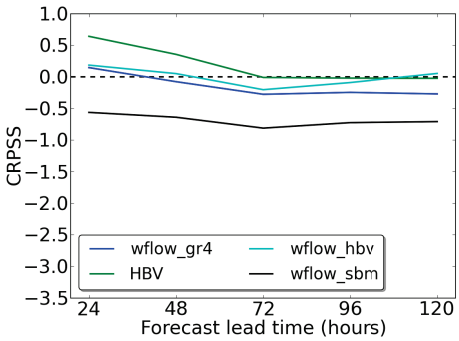
(b) The Meuse in Stenay



(c) The Chiers in Carignan



(d) The Semoy in Haulmé



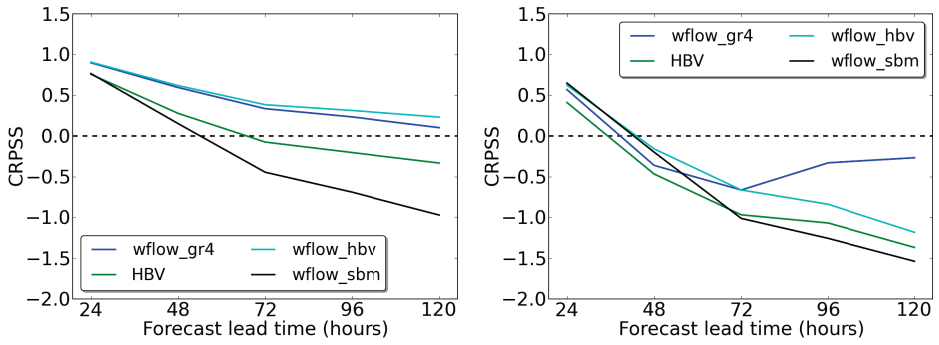
(e) The Ourthe in Tabreux

Figure 6.18: Plots of the $\overline{CRPSS}$ from Equation 5.5 with lead time

Hindcasts with error correction

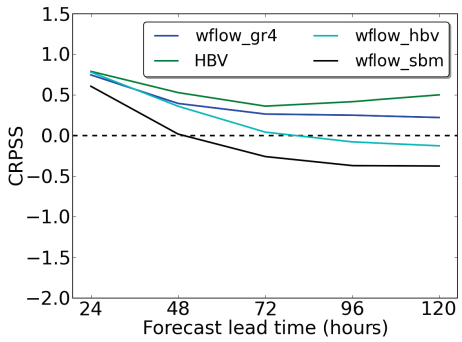
When comparing the hindcasts ensemble skills for all the models and all the sub-basins selected without and with error correction, a clear improvement can be distinguished for the hindcasts presenting an error correction procedure. Indeed, up to two days lead time, most models with an error correction presented a positive $\overline{CRPSS}$ (Figure 6.19). This suggests that all the models reliably predicted an observed discharge two days in advance. Additionally, a decrease of the $\overline{CRPSS}$ with increasing lead time, and thus of the ensemble skill, is appreciable on Figure 6.19 for the models with an error correction algorithm. This trend was not observable for the models without the error correction procedure.

From the time of predictability of each model (last lead time for which the model presents a positive $\overline{CRPSS}$), an average predictability limit could be computed for each model. On this basis, the HBV model showed the highest predictability limit, being able to forecast an event with approximately three and a half days lead time on average. The wflow_hbv and the wflow_gr4 models followed with predictability limits of on average slightly higher and slightly lower than three days lead times, respectively. As for the wflow_sbm model, it could predict events skillfully with a lead time of approximately two days on average, as indicated by Figure 6.19.

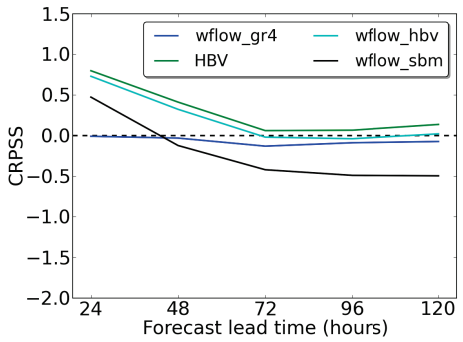


(a) The Meuse in Saint-Mihiel

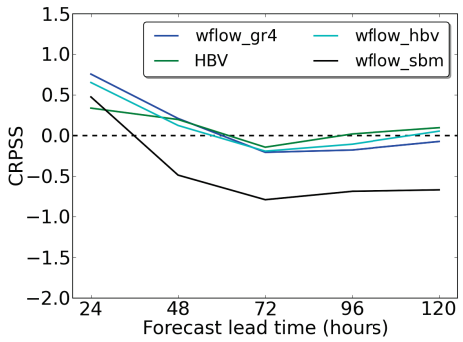
(b) The Meuse in Stenay



(c) The Chiers in Carignan



(d) The Semoy in Haulmé



(e) The Ourthe in Tabreux

Figure 6.19: Plots of the $\overline{CRPSS}$ from Equation 5.5 with lead time, models with error correction

6.3.2 Ensemble distribution analysis

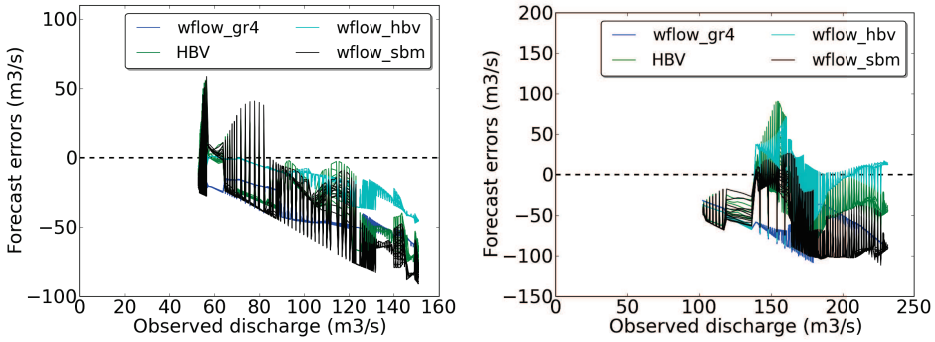
Hindcasts without error correction

Figures 6.20a to 6.21b display the forecast errors versus the observed discharge, for a twenty-four hours lead time, for four of the five sub-basins studied. The Chiers in Carignan exposed contrasting results from the other sub-catchments in Section 6.3.1, hence it was excluded from the ensemble distribution analysis. The following figures, providing informations on the hindcasts sharpness and errors, is an attempt towards the determination of the origin of the results presented in Section 6.3.1. From the latter, the models ensemble mean can be assessed. It can be seen that, overall, the models tend towards a negative forecast error, implying that they underestimated the observed discharge.

Figures 6.20a and 6.20b expose the forecast errors versus the observed discharge for a lead time of twenty-four hours for the Meuse in Saint-Mihiel and the Meuse in Stenay respectively. From these figures, it can be observed that the wflow_hbv model presents forecast errors overall closer to zero than the other models, especially appreciable with increasing observed discharges. This signifies that the wflow_hbv model produced hindcasts closer to the observed discharge for the Meuse in Saint-Mihiel and the Meuse in Stenay than the other models did, as was previously seen in Section 6.3.1 with the ensemble skill analysis. Furthermore, the spread in forecast errors for the wflow_hbv model, represented by the vertical lines in the following figures, appears to be relatively small, suggesting, to a certain extent, a high sharpness of its hindcast ensembles.

Considering the Meuse in Saint-Mihiel and the Meuse in Stenay (Figures 6.20a and 6.20b respectively), the wflow_sbm model produced the largest forecast errors for the latter, errors augmenting with increasing observed discharges. Likewise, this model displays the highest spread in forecast errors, indicating a very low sharpness of its hindcast ensembles. These observations are consistent with the low ensemble skill presented by the wflow_sbm model for the Meuse in Saint-Mihiel as well as for the Meuse in Stenay (Section 6.3.1).

The HBV and the wflow_gr4 models produced hindcasts with an average error range and ensemble distribution when comparing them with the hindcasts produced by the wflow_hbv and the wflow_sbm models for the Meuse in Saint-Mihiel and the Meuse in Stenay. Similar results were observed in Section 6.3.1, in the sense that the HBV and the wflow_gr4 models did not strike as being the best nor the worst models for those two sub-basins, regarding ensemble skill. Strikingly, the wflow_gr4 forecast errors spread appears to be surprisingly low for the Meuse in Saint-Mihiel and the Meuse in Stenay, suggesting a high sharpness of its hindcast ensembles.



(a) The Meuse in Saint-Mihiel

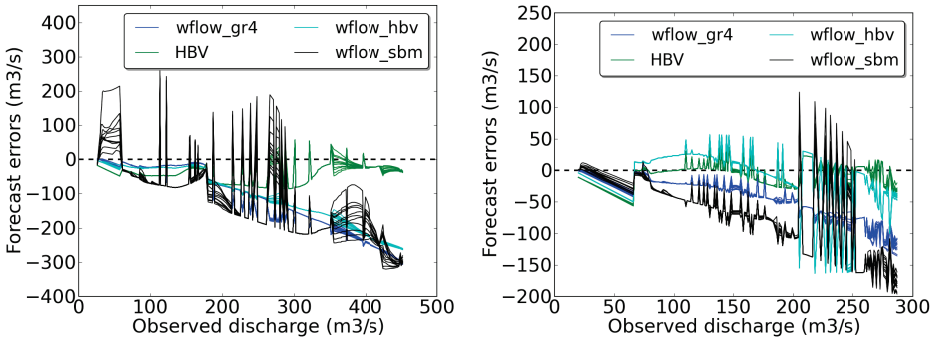
(b) The Meuse in Stenay

Figure 6.20: Forecast errors versus observed discharge, 24 hours lead time, for (a) the Meuse in Saint-Mihiel, (b) the Meuse in Stenay

Figures 6.21a and 6.21b expose the forecast errors versus the observed discharge for a lead time of twenty-four hours for the Semoy in Haulmé and the Ourthe in Tabreux respectively. For those sub-catchments, the HBV model presents forecast errors closer to zero than the other models, mostly appreciable for high observed discharges. This implies that the hindcasts produced by the HBV model for those sub-catchments were closer to the actual observed discharge than the hindcasts produced by the other models, as was observed in Section 6.3.1. Furthermore, the HBV model forecast errors present a relatively low spread, suggesting a high sharpness of this model's forecast ensembles.

As for the Meuse in Saint-Mihiel and the Meuse in Stenay (Figures 6.20a and 6.20b), the wflow_sbm model shows the highest forecast errors as well as the largest spread of the latter for the Semoy in Haulmé and the Ourthe in Tabreux. Based on its ensemble distribution anaysis, combined with its ensemble skill analysis (Section 6.3.1), the wflow_sbm model produced the least reliable ensemble forecasts for those four sub-catchments studied.

For the sub-basins depicted in Figures 6.21a and 6.21b, the wflow_hbv and the wflow_gr4 models generated hindcasts with an average error range and ensemble distribution, when compared to the HBV and the wflow_sbm models hindcasts. The same information was conveyed by Figures 6.18d and 6.18e in Section 6.3.1, where both models exhibited an average ensemble skill for the Semoy in Haulmé and the Ourthe in Tabreux.



(a) The Semoy in Haulmé

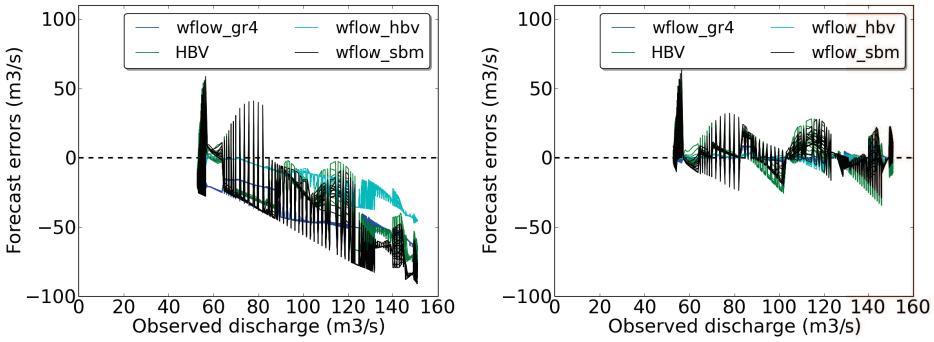
(b) The Ourthe in Tabreux

Figure 6.21: Forecast errors versus observed discharge, 24 hours lead time, for (a) the Semoy in Haulmé, (b) the Ourthe in Tabreux

Hindcasts with error correction

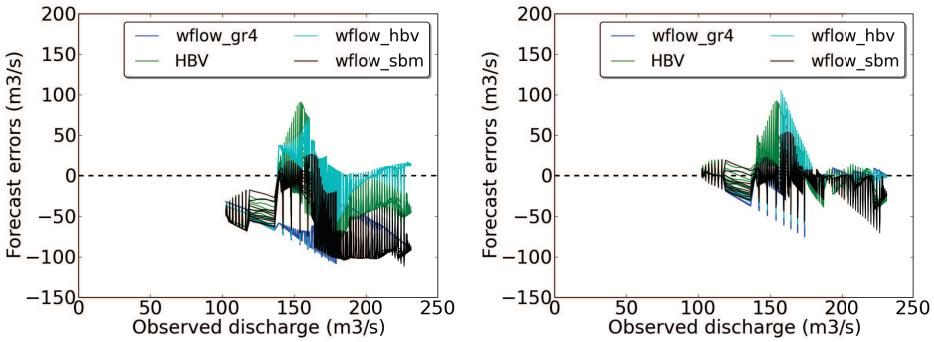
Figures 6.22 to 6.26 present the forecast errors by observed discharge, for a twenty-four hours lead time, for all the models and all the sub-basins without and with the error correction algorithm. These figures promote the investigation of the effect of the error correction on the hindcasts ensemble skill (seen in Section 6.3.1) in terms of their sharpness and errors.

Looking at Figures 6.22, 6.23 and 6.24, corresponding to the Meuse in Saint-Mihiel, the Meuse in Stenay and the Chiers in Carignan, respectively, the forecast ensemble errors appear to be closer to the zero line for the models with error correction than for the models without error correction. This is distinguishable mainly for high observed discharges. Furthermore, the spread of the ensemble seems smaller for the models with error correction than for the models without the error correction algorithm. These observations suggest that, for these three sub-basins, the error correction procedure induced a reduction of the forecast errors as well as an increase of the ensembles sharpness for all the models.



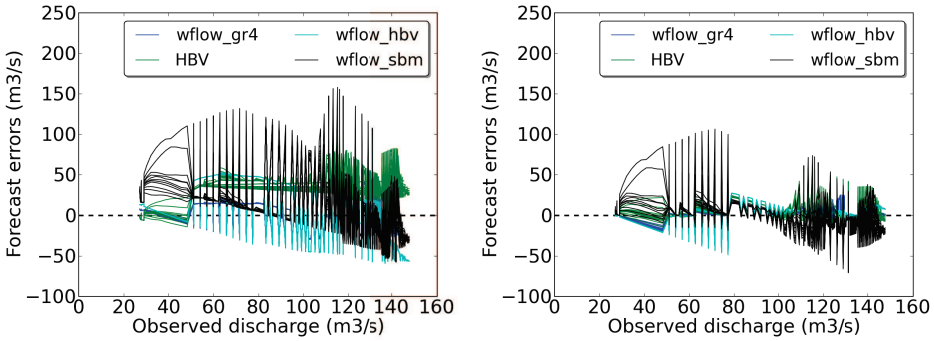
(a) Without error correction (b) With error correction

Figure 6.22: Forecast errors versus observed discharge, 24 hours lead time, for the Meuse in Saint-Mihiel (a) without error correction and (b) with error correction



(a) Without error correction (b) With error correction

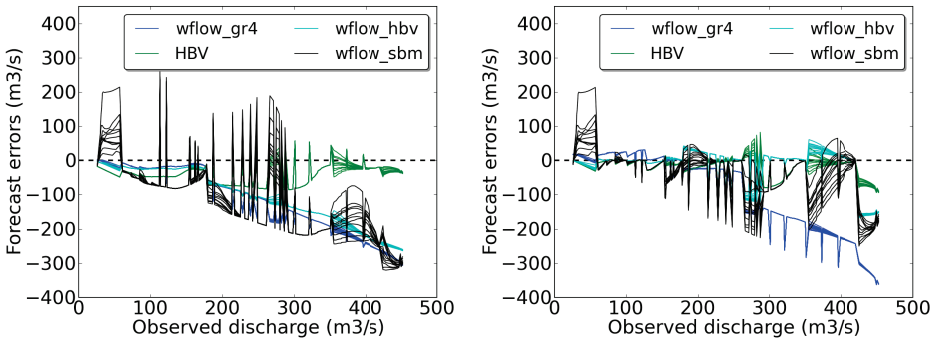
Figure 6.23: Forecast errors versus observed discharge, 24 hours lead time, for the Meuse in Stenay (a) without error correction and (b) with error correction



(a) Without error correction (b) With error correction

Figure 6.24: Forecast errors versus observed discharge, 24 hours lead time, for the Chiers in Carignan (a) without error correction and (b) with error correction

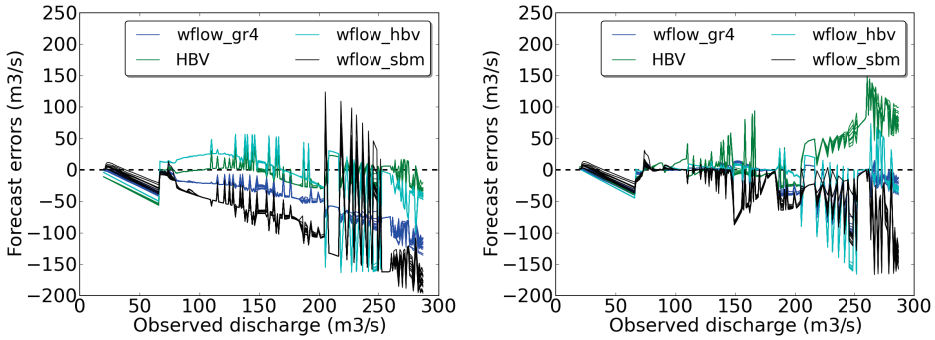
Based on Figure 6.25, for the Semoy in Haulmé, the error correction procedure seems to have given rise to dissimilar results for the different models. Indeed, the forecast errors for the wflow_hbv and the wflow_sbm models appear closer to zero with the error correction procedure than without the latter. Moreover, the wflow_sbm ensemble spread is clearly smaller with the error correction procedure than without this procedure. Contrastingly, the forecast errors for the HBV and the wflow_gr4 models appear greater with the error correction procedure and the spread of the wflow_gr4 ensemble forecasts is apparently larger with the error correction than without the latter. Therefore, whereas the error correction procedure induced positive results on the wflow_sbm and the wflow_hbv model hindcasts performances in terms of their ensemble forecasts errors and sharpness, opposite results seem to have occured for the HBV and the wflow_gr4 models for the Semoy in Haulmé.



(a) Without error correction (b) With error correction

Figure 6.25: Forecast errors versus observed discharge, 24 hours lead time, for the Semoy in Haulmé (a) without error correction and (b) with error correction

In the case of the Ourthe in Tabreux (Figure 6.26), dissimilarities in the models results with the error correction algorithm are also existent. Overall, the wflow_hbv, the wflow_gr4 and the wflow_sbm models forecasts errors are situated closer to the zero line with error correction than without the latter. The HBV model forecast errors, being closer to zero without the error correction algorithm, are highly positive for high observed discharges with the error correction algorithm. This implies an overestimation of high observed discharges by the HBV model, with the error correction procedure, for the Ourthe in Tabreux. Additionally, the spread in the ensemble forecasts for the HBV, the wflow_hbv and the wflow_sbm models seems to be larger for high observed discharges with the error correction procedure. These observations suggest that the wflow_gr4 model is the only model that benefited positively from the error correction procedure for the Ourthe in Tabreux in terms of its ensemble forecast errors and sharpness.



(a) Without error correction

(b) With error correction

Figure 6.26: Forecast errors versus observed discharge, 24 hours lead time, for the Ourthe in Tabreux (a) without error correction and (b) with error correction

Chapter 7

Discussion

Based on the simulations and hindcasts results obtained for the four models for the five sub-catchments selected, a general comparative overview of the models performances can be given and is presented in Table 7.1.

7.1 HBV-96 model performance

In simulation, the HBV model presented the best performance for high flows as well as for low flows, when compared to the other models simulations produced during this research. Similar results were found in the literature. Indeed, as stated in Section 3.1, the HBV model has shown a good to moderate performance for flood simulations and forecasts in several studies (Table 7.1; Berglov, 2009; Gorgen et al., 2010; De Wit, 2005; Drogue et al., 2010). For low flow conditions, the HBV model with the Kriging interpolation procedure as well as the Regnie interpolation procedure displayed a good performance. As regards to high flow conditions, both interpolation procedures produced simulations of a lower quality than during low flows, moderate for the Kriging interpolation and moderate to good for the Regnie interpolation. This is due to the slight underestimation of the discharge during high flow events by the HBV model, which was equally observed by Weerts et al. (2008); De Wit (2005) for the Moselle and the Rhine River basins. In addition, as observed in Section 6.2, the HBV model using the Regnie interpolation procedure showed better simulation performances, which was also suggested in Weerts et al. (2008).

The HBV-96 model displayed the least divergence with the observed discharge in hindcasts, similarly to simulations. It indeed showed a moderate efficiency for hindcasts without error correction algorithm and a good efficiency with the error correction procedure (Table 7.1). This model, used operationally at Deltares for forecasting purposes, has undergone several changes with the aim to increase

its efficiency for flood forecasting. These changes, referred to in Section 3.1, have shown positive results on the HBV model forecasts regarding their quality (Berglov, 2009; Weerts et al., 2008).

7.2 wflow_gr4 model performance

Wflow_gr4 is the model presenting the simplest and the least flexible structure of all the models used for this research. This simplicity and lack of flexibility, due to the small number of parameters and reservoirs, yielded an average simulation performance - of moderate effiency in high and low flows - when compared to the other models studied (Table 7.1). Nonetheless, this lack of flexibility was also the origin of recurrent discrepancies between the observed and the simulated discharge, as observed in Section 6.2.

Indeed, a striking feature of the wflow_gr4 simulations was the underestimation of peakflows and the overestimation of low flows. The origin of this moderately poor performance springs in the wflow_gr4 model's structure.

A first plausible origin of the wflow_gr4 model's moderate simulation performance could be the fixed unit hydrographs contributions, controlling the amount of baseflow and quick runoff (UH1 and UH2 respectively, Section 3.2). Similar conclusions were drawn by Esse (2012) and Pushpalatha et al. (2011), who suggested the creation of a second parallel routing store to offer more resilience in the GR4H model's structure.

Another source could be the conceptual water exchange function (Section 3.2), as this latter depends on parameter X_2, which is either positive or negative. A fifth parameter was introduced to the original GR4J model by Le Moine (2008), which gave rise to the GR5J model, in order to obtain a more realistic groundwater exchange. This new parameter (X_5), by allowing the groundwater exchange to fluctuate between negative and positive values, led to improvements in the low flow simulations.

A supplementary origin of discrepancy between the observed and the simulated discharge during peakflows could be the absence of a snow module for the wflow_gr4 model. However, the snowmelt water contribution to the total discharge does not exceed 15% for the sub-catchments selected, and should therefore not be greatly influential on the simulated discharge.

An additional characteristic of the wflow_gr4 model simulations, mentioned in Section 6.2.2, was the peakflows bell-shape. This attribute of the wflow_gr4 simulations is due to the linear routing by the two unit hydrographs (Section 3.2).

With respect to hindcasts, the wflow_gr4 model presented a moderate to low effiency without any error correction procedure, rising to a good effiency with the incorpo-

ration of an error correction algorithm (Table 7.1). These results can be related to studies carried at IRSTEA, with as a final purpose the development of a reliable flood forecasting model. Indeed, as presented in Section 3.2, numerous researches initiated from the GR4H model, which proved to be efficient in simulation but not as efficient in forecasting, leading to the creation of the GRP model. This was achieved through the modification of the GR4H model's original structure as well as the inclusion of a data assimilation scheme to the latter (Tangara, 2005; Berthet, 2010). The GRP model is nowadays used operationally at IRSTEA for flood forecasting purposes.

7.3 wflow_hbv model performance

The wflow_hbv model, being a fully distributed version of the HBV model, should have given similar or better results than the latter. Nonetheless, our results did not confirm this. Contrarily, as observed in Section 6.2, the wflow_hbv model produced simulations of low effiency for low flow conditions, due to a considerable overestimation of the observed low flows (Table 7.1). It nonetheless produced more reliable simulations for high flow conditions than the HBV model did, presenting overall a good performance for the latter (Table 7.1). This is due to the fact that the interception was inadvertently set to zero prior to the simulations runs. It was discovered at the end of the project, when time restrictions prevented running additional simulations.

Nevertheless, the wflow_hbv model showed a relatively good efficiency in forecasting, although and probably due to the fact that the interception was set to zero. With a moderate performance without any error correction algorithm, increasing to a good performance with the incorporation of this algorithm to the model (Table 7.1).

7.4 wflow_sbm model performance

A striking outcome of this project was the performance of the fully distributed wflow_sbm model, both in simulation and in hindcasting, compared to the performance of the lumped models (HBV and wflow_gr4). Indeed, overall, the wflow_sbm model showed a lower quality of discharge simulations and hindcasts. Regarding simulations, this model produced very large errors, leading to a low effiency during high flows and a moderate effiency for low flows (Section 6.2; Table 7.1). As for hindcasts, the combination of a poor skill with a poor sharpness of the ensemble forecasts produced made their quality low to moderate, without and with the error correction procedure respectively (Section 6.3; Table 7.1).

Obligeois et al. (2013); Smith et al. (2004); Reed et al. (2004) showed that lumped models could outperform distributed models, depending on the precipitation spatial variability and the catchment size. According to these authors, distributed models showed less discrepancy between the observed and the simulated discharge for large catchments and rainfall events of large spatial variability.

A key feature of the wflow_sbm model simulations was the very dynamic behaviour of the latter, producing peaks during the observed low flows; as well as the underestimation of high flows and the over- or underestimation of the peaks time lag (Section 6.2). Several hypothesis can be emitted in an attempt to clarify the origin of these phenomenons as no literature review was found on the subject.

A first reason of this dynamic behaviour could be the presence of the *FreeWaterDepth* storage, which, acting as overland flow, contributes directly to the *Kinematic wave*. This flux is present solely in the wflow_sbm model and could therefore be the cause of the dynamic simulations observed for the latter.

A second reason could be the presence of the *ExfiltWater* flux (see Figure D.3), over-contributing to the *Kinematic wave* and thus the discharge at the outlet of the sub-basin as a result of the low storage capacity of the model.

The consequence of both reasons would therefore be the presence of singular peaks during observed low flows, and the underestimation of high flows.

A solution to these discrepancies could therefore be the increment of the storage capacity of this model's reservoirs by supplementary calibration procedures.

Additionally, from Vertessy et al. (2000)'s work on the topog_sbm model, from which the wflow_sbm originates, similar findings to the previously stated features were made. Indeed, the topog_sbm model similarly underestimated high flows and their time lag.

Table 7.1: Performance assessment summary of the four models based on Chapter 6

Name of models		HBV	wflow_gr4	wflow_hbv	wflow_sbm
Original developer		SMHI	IRSTEA	Deltares	
Routing method		MAXBAS	Two unit hydrogaphs	Kinematic wave function	
Model type		Conceptual, semi-distributed	Conceptual, lumped	Physical, distributed	
Efficiency in simulations	High flows	Moderate with Kriging, good to moderate with Regnie	Moderate	Good	Low
	Low flows	Good with Regnie and Kriging	Moderate	Low	Moderate
Efficiency in hindcasts	Without error correction	Moderate	Moderate to low	Moderate	Low
	With error correction	Good	Good	Good	Moderate

7.5 Simulated peaks time lag error

For the simulations, all models produced at least one highly negative residual at the end of the year 2011 (Section 6.2.2). This residual was attributed to an overestimation of the observed discharge by the models, due to an over- or underestimation of a peak time lag. The recurrence of this residual suggests that the error might be situated in the quality of the input data used by all the models for running the simulations.

7.6 The case of the Chiers in Carignan

Both in simulation and in forecast, the Chiers in Carignan displayed distinctively different results than the other sub-basins (Sections 6.2 and 6.3). This could either suggest that the trend observed for the other four sub-basins is not applicable to the entire Meuse River basin, due to heterogeneity in geology, land use, etc, or that there was an error in the input data and/or the interpolation of the latter.

Chapter 8

Conclusions and recommendations

8.1 Conclusions

This research was conducted at Deltares and IRTSEA, with as main purpose to bridge the gap between models performances in simulations and hindcasts and models structures. To this effect, five sub-basins of the Meuse River basin were pre-selected in order to offer a spatial basis for the intercomparison framework. These sub-basins included the Meuse in Saint-Mihiel, the Meuse in Stenay, the Semoy in Haulmé, the Ourthe in Tabreux and the Chiers in Carignan. Additionally, four models were chosen, these were the semi-distributed conceptual HBV-96 model, the GR4H lumped and conceptual model and the fully-distributed physically-based wflow_hbv and wflow_sbm models.

The GR4H model from IRSTEA was first transcribed as a PCRaster Python model, from its original FORTRAN code, and calibrated based on hourly observed discharge measurements using the Nash-Sutcliffe model efficiency criterion (NSE) for the five pre-selected sub-catchments. The calibration produced an average NSE of 0.9 for all the sub-basins. This transcription, leading to the wflow_gr4 model, enabled its use within the intercomparative framework built during this project.

The performance of the four models was assessed and compared through the execution of seven years of hourly streamflow simulations and a two weeks event of hourly hindcasts (with and without error correction; with a lead time of up to five days), for the five sub-basins, using the Delft-FEWS forecasting platform. The data-centric forecasting platform provided an objective environment for this comparison, acting as an empty shell where forcing data was collected and interpolated prior to the models execution.

The wflow_hbv model interception values were unentendedly all set to zero prior to

the simulations and hindcasts executions. This was discovered at a late stage of the project, disabling any further model runs. Therefore, unfortunately, no comparison was legitimately possible between the wflow_hbv results and the other models' results. The wflow_hbv model results were nevertheless described.

The models performances in simulations were analysed and compared first quantitatively, based on their overall statistical performances (with the MAE, the RMSE and the MSE performance metrics) and on the under- or overestimation of the total discharge by the models. The models performances were then evaluated qualitatively, looking at the models' dynamism during peakflows and low flow conditions. The HBV model simulations were executed with two different interpolation methods, the Regnie and the Kriging interpolation techniques. This intercomparison framework was based on the Meuse in Saint-Mihiel simulation results, used as representative for all the sub-basins. Nonetheless, the Chiers in Carignan displayed contrasting results from the other sub-basins, which were attributed to either an error in the input data or the interpolation of the latter for this sub-basin, or to heterogeneity in the Meuse River basin geology, land use, etc (Section 7).

On average, the semi-distributed HBV model (with both interpolation techniques) and the lumped wflow_gr4 model presented the best efficiency in simulations. These models globally displayed the lowest error magnitudes. Both models presented a tendency to underestimate the total discharge, due to the overestimation of the low flows coupled to the underestimation of the high observed flows by the latter. Nonetheless, the HBV model showed a higher performance than the wflow_gr4 model, especially for low flow conditions. Furthermore, the observed discharge dynamics were well captured by the HBV model while they were not by the wflow_gr4 model, presenting a smoothed simulated discharge due to its routing method.

Additionally, it was observed that the Regnie interpolation method conferred the HBV simulations a higher performance than the Kriging interpolation technique for high flows. This suggests that both the quality of the forcing data and the interpolation technique used are primordial to the model performance.

The fully-distributed models, the wflow_hbv and the wflow_sbm models, presented on average the lowest efficiency in simulations. The wflow_hbv model overestimated largely the low observed flows due to the null interception values, leading to an overall discharge overestimation. Regarding the wflow_sbm model, it presented on average the highest discrepancies with the observed discharge due to a large underestimation of the peakflows, combined an overestimation of the low flows, attributed to an over-dynamicity of the model.

The models performances in hindcasts (with and without error correction) were analysed with a coupled investigation of their ensemble skill (based on the $\overline{CRPSS}$) and of their ensemble distribution. On average, all the models showed an underestimation of the total discharge as well as a low ensemble skill. Globally, the

HBV-96 and the wflow_hbv models presented the highest performance in hindcasts both with and without the error correction algorithm, showing the smallest forecasts errors and the largest sharpness of their forecasts. The wflow_hbv high performance in hindcasts was nevertheless linked to the null interception values. The wflow_sbm model displayed the lowest performance, both without and with the error correction algorithm, due to the large forecast errors produced by the later, coupled to a small sharpness of its forecasts. The wflow_gr4 model, being moderately performant in hindcasts, showed a very high sharpness of its ensembles.

The error correction algorithm, by increasing the hindcasts sharpness and decreasing the models errors (especially for high observed flows and more noticeably for one day lead time, showing a decreasing performance with increasing lead time trend), led to an increase of the ensemble skill as well as an increase of the ensemble distribution of the hindcasts produced by all the models.

A major result of this research is the performance of the fully-distributed physically-based wflow_sbm model compared to the performance of the semi-distributed conceptual HBV model annd of the lumped conceptual wflow_gr4 model, the wflow_hbv not being legitimately comparable. Indeed, the wflow_sbm model showed no clear improvements on the simulations and hindcasts efficiencies for the sub-basins selected, compared to the HBV and the wflow_gr4 models. On the contrary, the wflow_sbm model displayed the lowest performance, in simulations and in hindcasts. Section 7 constituted an attempt to clarify this unexpected result, which nonetheless deserves further research.

8.2 Recommendations

This intercomparison project, although relatively complete, could benefit greatly from further investigations in order to confirm the trends observed throughout this research.

As a first step, more sub-basins of the Meuse River basin could be added to this project. Indeed, as observed in Sections 6.2 and 6.3, the Chiers in Carignan displayed contrasting results to the other four sub-basins. Thus, this discrepancy merits additionnal studies.

Conjointly, the values of the wflow_hbv model interception should be fixed prior to running simulations and hindcasts in order to investigate the benefits of using distributed models alternately to lumped or semi-distributed models for flood modelling and forecasting.

Additionally, a longer period of hindcasts could be executed for all four models in order to validate the findings made during this project with solely two weeks of hindcasts.

Surprisingly low performances were obtained for the wflow_sbm and the wflow_hbv models, in simulations and in hindcasts, which were not expected. Therefore, an additional calibration of the kinematic wave function could be performed for both of these models, prior to running simulations and hindcasts for the latter.

And finally, the models fluxes quantities and reservoirs contents could be analysed in order to explore the conjectures made in Chapter 7 regarding the models structures influences on the simulated discharge.

Chapter 9

Acknowledgments

First of all, I would like to thank my supervisors, Albrecht Weerts, Maria-Helena Ramos and Maarten J. Waterloo, who guided me throughout this research by providing me with the tools and knowledge, and who reinforced my interest for this field.

Furthermore, I would like to express my gratitude to Laurène Bouaziz and Jan Verkade, who shared their office with me during those five months of internship at Deltares and who gave me great advices.

Without the help of Jaap Schellekens, I would not have grasped the complexity of the wflow models.

Thanks to Patricia Lopez for introducing me to EVS.

I would moreover like to acknowledge the unique atmosphere found at IRSTEA and the warm welcome of its Hydrology team.

A special thanks to Florent Lobligeois and David Brochard, who shared their office with me during this month spent at IRTSEA; to Laurent Coron, for his valuable help with AirGR; and to the ultimate frisbee lunch lectures (and their members), which were a great source of inspiration.

The occupants of the hydroroom (Asta Kunkel, Katerina Lampropoulou and Laurens Kaland) made the writing of my thesis more entertaining.

I would also like to state that without the support of my friends and family, this experience would not have been the same.

Last but not least, thanks to the kerkzaal occupiers, who put all their effort into fighting for the preservation of the Hydrology Master at the VU.

Bibliography

P. Archambeau. Hydrological modelling of the meuse basin, 2013. URL `http://hydrology.argenco.ulg.ac.be/home`.

J. Bear and A. H-D. Cheng. *Modeling groundwater flow and contaminant transport*. Springer Dordrecht, Heidelberg, London, New York, 2010.

G. Berglov. *Improvement HBV model Rhine in FEWS: final report*. SMHI, 2009.

L. Berthet. *Prevision des crues au pas de temps horaire: pour une meilleure assimilation de l'information de debit dans un modele hydrologique*. PhD thesis, AgroParisTech, 2010.

P. M. T. Broersen and A. H. Weerts. Automatic error correction of rainfall-runoff models in flood forecasting systems. In *Instrumentation and Measurement Technology Conference, 2005. IMTC 2005. Proceedings of the IEEE*, volume 2, pages 963–968. IEEE, 2005.

J. D. Brown, J. Demargne, D-J. Seo, and Y. Liu. The ensemble verification system (EVS): a software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations. *Environmental Modelling Software*, 25(7), 2010.

Service Public de Wallonie. Accord international sur la meuse, 2002. URL `http://environnement.wallonie.be/legis/international/accord004.htm`.

M. De Wit. Estimation of extreme floods of the river meuse using a stochastic weather generator and a rainfall runoff model. *Hydrological Sciences Journal des Sciences Hydrologiques*, 2005.

M. J. M. De Wit, H. A. Peeters, P. H. Gastaud, P. Dewil, K. Maeghe, and J. Baumgart. Floods in the meuse basin: event descriptions and an international view on ongoing measures. *International Journal of River Basin Management*, 5(4): 279–292, 2007.

Deltares. Software Delft-FEWS, 2010.

M. C. Demirel, M. J. Booij, and A. Y. Hoekstra. Effect of different uncertainty

sources on the skill of 10 day ensemble low flow forecasts for two hydrological models. *Water Resources Research*, 2013.

G. Drogue, M. Fournier, A. Bauwens, F. Commeaux, O. De Keizer, D. Francois, E. Guilmin, A. Degre, S. Detrembleur, B. Dewals, M. Pirotton, D. Pontegnie, C. Sohier, and W. Vaneuville. Analysis of climate change, high-flows and lowflows scenarios on the meuse basin WP1 report action 3. Technical report, EPAMA, 2010.

W. R. Esse, C. Perrin, M. J. Booij, D. C. M. Augustijn, F. Fenicia, D. Kavetski, and F. Lobligeois. The influence of conceptual model structure on model performance: a comparative study for 237 french catchments. *Hydrology and earth system sciences*, 17, 2013.

W. R. van Esse. Demystifying hydrological monsters: can flexibility in model structure help explain monster catchments?, October 2012.

K. Gorgen, J. Beersma, H. Buiteveld, G. Brahmer, M. Carambia, O. de Keizer, P. Krahe, E. Nilson, R. Lammersen, and C. Perrin. *Assessment of climate change impacts on discharge in the Rhine River Basin: results of the RheinBlick2050 Project*. CHR, 2010.

R. Kiss. Determination of drainage network in digital elevation models, utilities and limitations. *Journal of Hungarian geomathematics*, 2, 2004.

T-Y. Koh and J. S. Ng. Improved diagnostics for NWP verification in the tropics. *Journal of Geophysical Research*, 114(D12), June 2009. ISSN 0148-0227. doi: 10.1029/2008JD011179.

N. Le Moine. *Le bassin versant de surface vu par le souterrain : une voie d'amelioration des performances et du realisme des modeles pluie debit?* Paris 6, January 2008.

J. Lerat, B. Abadie, V. Andreassian, and C. Perrin. Catchment domain model, part a. focus on the rhine basin. what to expect from simple approaches in hydrological modelling ? Technical report, CEMAGREF, May 2006.

G. Lindstrom, B. Johansson, M. Persson, M. Gardelin, and S. Bergstrom. Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology*, 201(1-4):272–288, December 1997. ISSN 0022-1694. doi: 10.1016/S0022-1694(97)00041-3.

F. Lobligeois, V. Andreassian, C. Perrin, P. Tabary, and C. Loumagne. When does higher spatial resolution rainfall information improve streamflow simulation? an evaluation on 3620 flood events. *Hydrology and Earth System Sciences Discussions*, 10(10):12485–12536, October 2013. ISSN 1812-2116. doi: 10.5194/hessd-10-12485-2013.

T. Mathevet. *Quels modeles pluie-debit globaux au pas de temps horaire? Developpements empiriques et intercomparaison de modeles sur un large echantillon de bassins versants*. PhD thesis, Ph. D. thesis, ENGREF, 463 pp, 2005.

A. Minnet. Delft-FEWS documentation, April 2014. URL `https://publicwiki.deltares.nl/display/FEWSDOC/Home`.

J. E. Nash and J. V. Sutcliffe. River flow forecasting through conceptual models part i a discussion of principles. *Journal of Hydrology*, 10(3):282–290, April 1970. ISSN 0022-1694. doi: 10.1016/0022-1694(70)90255-6.

PAGASA. Causes and types of floods, 2004. URL `http://pagasa.dost.gov.ph/genmet/floods/cause\_types.html`.

C. Perrin. *Vers une amelioration d'un modele global pluie debit*. PhD thesis, Institut National Polytechnique de Grenoble-INPG, 2000.

C. Perrin, C. Michel, and V. Andreassian. Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279(1–4):275–289, August 2003. ISSN 0022-1694. doi: 10.1016/S0022-1694(03)00225-7.

C. Perrin, M. Claude, and A. Vazken. Modeles hydrologiques du genie rural (GR), 2007.

Anderson Photos. Anderson photos, 2012. URL `http://www.independentmail.com/photos/2012/nov/25/150357/`.

R. Pushpalatha, C. Perrin, N. Le Moine, T. Mathevet, and V. Andreassian. A downward structural sensitivity analysis of hydrological models to improve low flow simulation. *Journal of Hydrology*, 411:66–76, December 2011. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2011.09.034.

S. Reed, V. Koren, M. Smith, Z. Zhang, F. Moreda, D-J. Seo, and DMIP Participants. Overall distributed model intercomparison project results. *Journal of Hydrology*, 298:27–60, October 2004. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2004.03.031.

J. Schellekens. Home OpenStreams deltares public wiki, 2012. URL `http://publicwiki.deltares.nl/display/OpenS/Home`.

J. Schellekens. OpenStreams: wflow documentation (release 0.91), 2013.

J. M. Shultz, C. E. Crosby, and A. J. McEnery. Kinematic wave technique applied to hydrologic distributed modeling using stationary storm events: an application to synthetic rectangular basins and an actual watershed. *Hydrology Days*, pages 116–126, 2008.

N. (Finnish Meteorological Institute) Sigbritt. Eumetcal, 2001. URL `http://www.eumetcal.org/resources/ukmeteocal/verification/www/english/courses/msgcrs/index.htm`.

SMHI. *IHMS, Integrated Hydrological Modelling System, Manual, Version 4.5*. Swedish Meteorological and Hydrological Institute, S-601 76 Norrkoping, Sweden, 1999.

M. B. Smith, V. I. Koren, Z. Zhang, S. M. Reed, J-J. Pan, and F. Moreda. Runoff response to spatial variability in precipitation: an analysis of observed data. *Journal of Hydrology*, 298:267–286, October 2004. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2004.03.039.

M. Tangara. *Nouvelle methode de prevision de crue utilisant un modele pluie debit global*. PhD thesis, EPHE, 2005.

M. Tu. Assessment of the effects of climate variability and land use change on the hydrology of the meuse river basin, 2006.

UNISDR. Disaster statistics flood hazards PreventionWeb.net, 2013. URL `http://www.preventionweb.net/english/hazards/statistics/?hid=62`.

J. S. Verkade and M. G. F. Werner. Estimating the benefits of single value and probability forecasting for flood warning. *Hydrology and Earth System Sciences*, 15(12):3751–3765, December 2011. ISSN 1607-7938. doi: 10.5194/hess-15-3751-2011.

R. Vertessy, H. Elsenbeer, Y. Bessard, and A. Lack. Storm runoff generation at la cuenca. *Spatial patterns in catchment hydrology: observations and modelling*, 271, 2000.

A.H. Weerts, D. Meissner, and S. Rademacher. Input data rainfall runoff model operational systems FEWSNL FEWSDE. Technical Report Q 4234, Dltares, Delft, December 2008.

M. Werner and A. Weerts. Running reforecasts (hindcasts) for verification with FEWS, 2012.

M. Werner, J. Schellekens, P. Gijsbers, M. van Dijk, O. van den Akker, and K. Heynert. The delftFEWS flow forecasting system. *Environmental Modelling & Software*, 40:65–77, February 2013. ISSN 1364-8152. doi: 10.1016/j.envsoft.2012.07.010.

C. G. Wesselung, D-J. KARSSENBERG, P. A. Burrough, and W. DEURSEN. Integrating dynamic environmental models in GIS: the development of a dynamic modelling language. *Transactions in GIS*, 1(1):40–48, 1996.

WMO. *Manual on flood forecasting and warning*. World Meteorological Organization, Geneva, 2011a. ISBN 9789263110725 9263110727.

WMO. WMO flood forecasting initiative: Development of a decision-aiding tool for the selection of flood forecasting models (draft), 2011b.

WWasser. Rhine course and river system, place names in english, 2013. URL http://en.wikipedia.org/wiki/File:Flusssystemkarte_Rhein_04.jpg.

Appendix A

Rhine river system

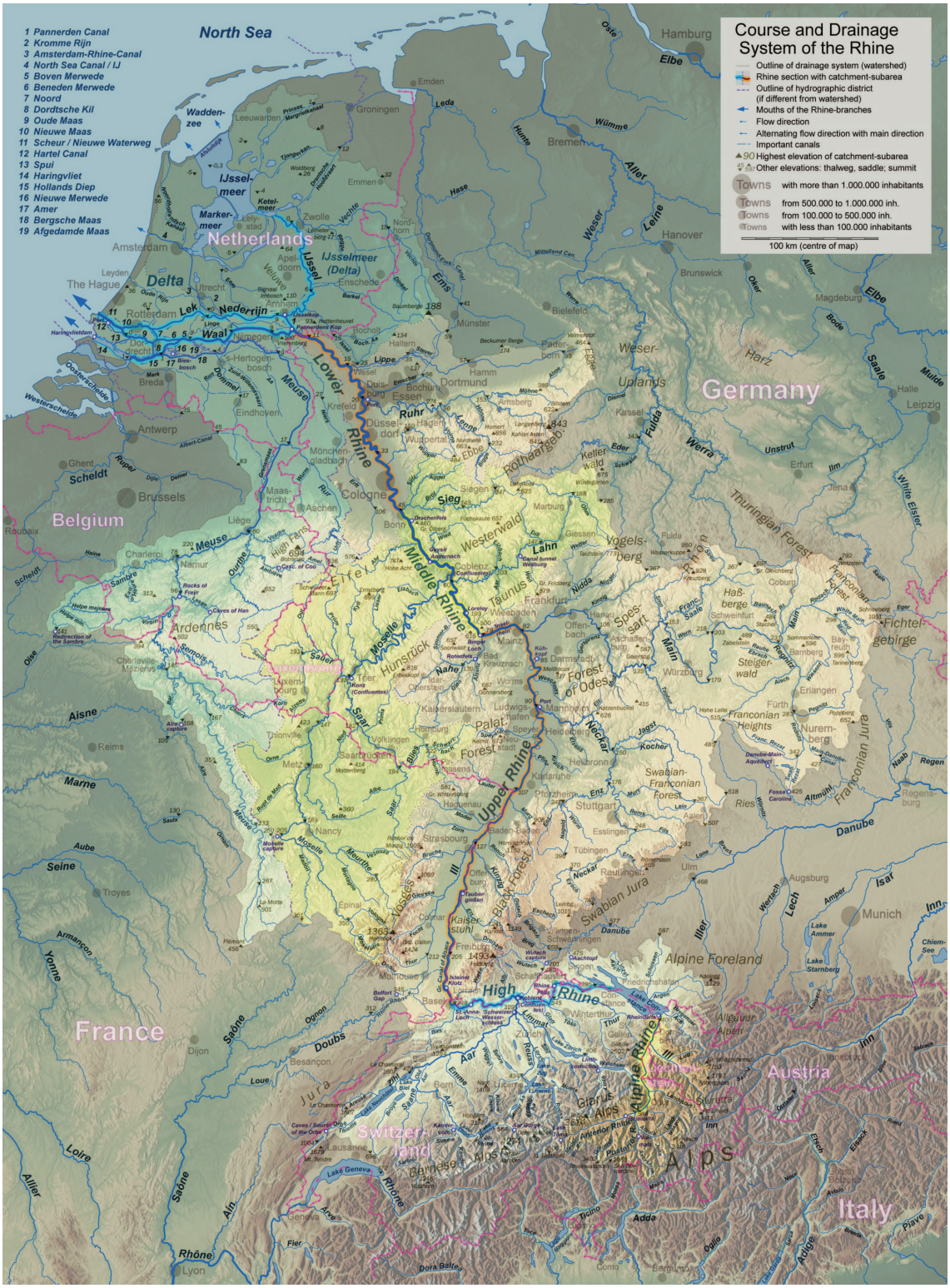


Figure A.1: Rhine and Meuse courses and river systems (WWasser, 2013)

Appendix B

HBV-96 model

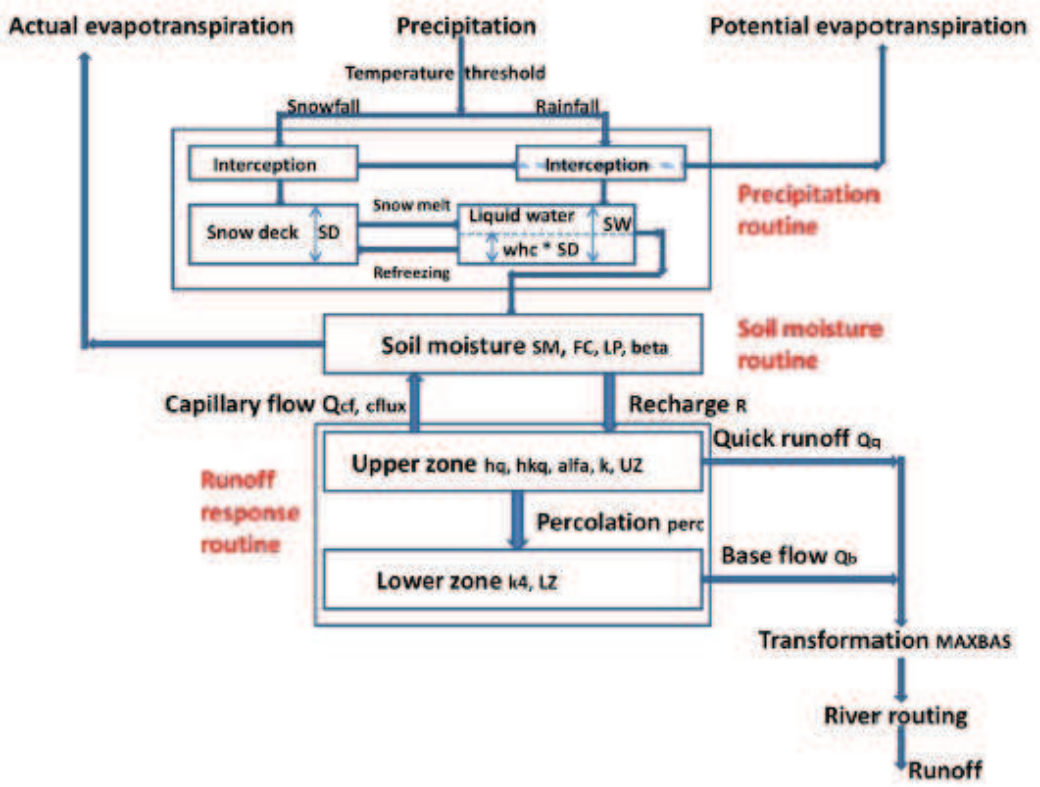


Figure B.1: Diagram of the HBV-96 model (Schellekens, 2013)

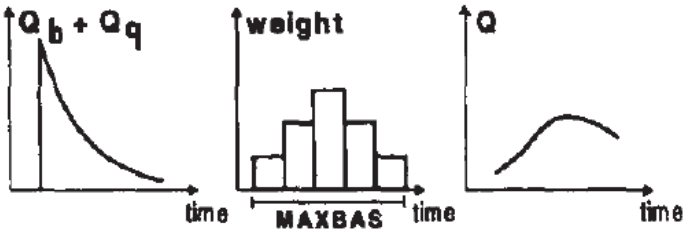


Figure B.2: MAXBAS transformation function, modified from (Lindstrom et al., 1997)

Appendix C

GR4J model

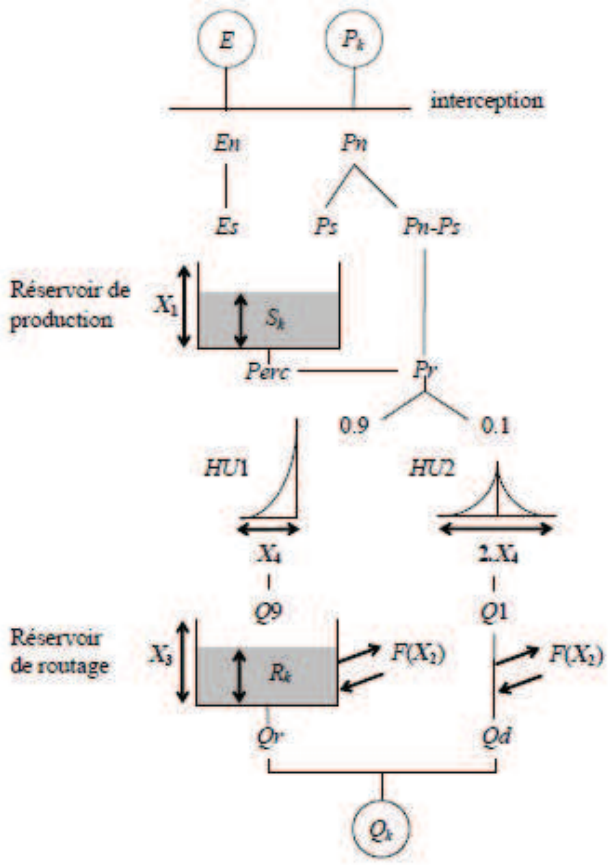


Figure C.1: Diagram of the GR4J model (Perrin et al., 2003)

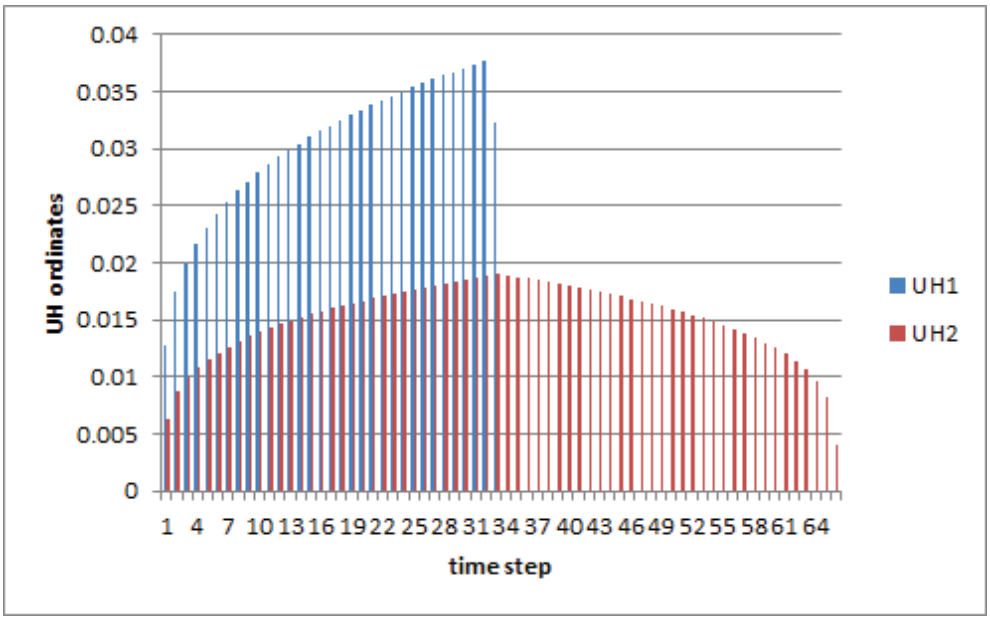


Figure C.2: Example of the GR4J model routing method by the unit hydrographs, with $X_4 = 32.85$ days (Perrin et al., 2003)

Appendix D

OpenStreams models

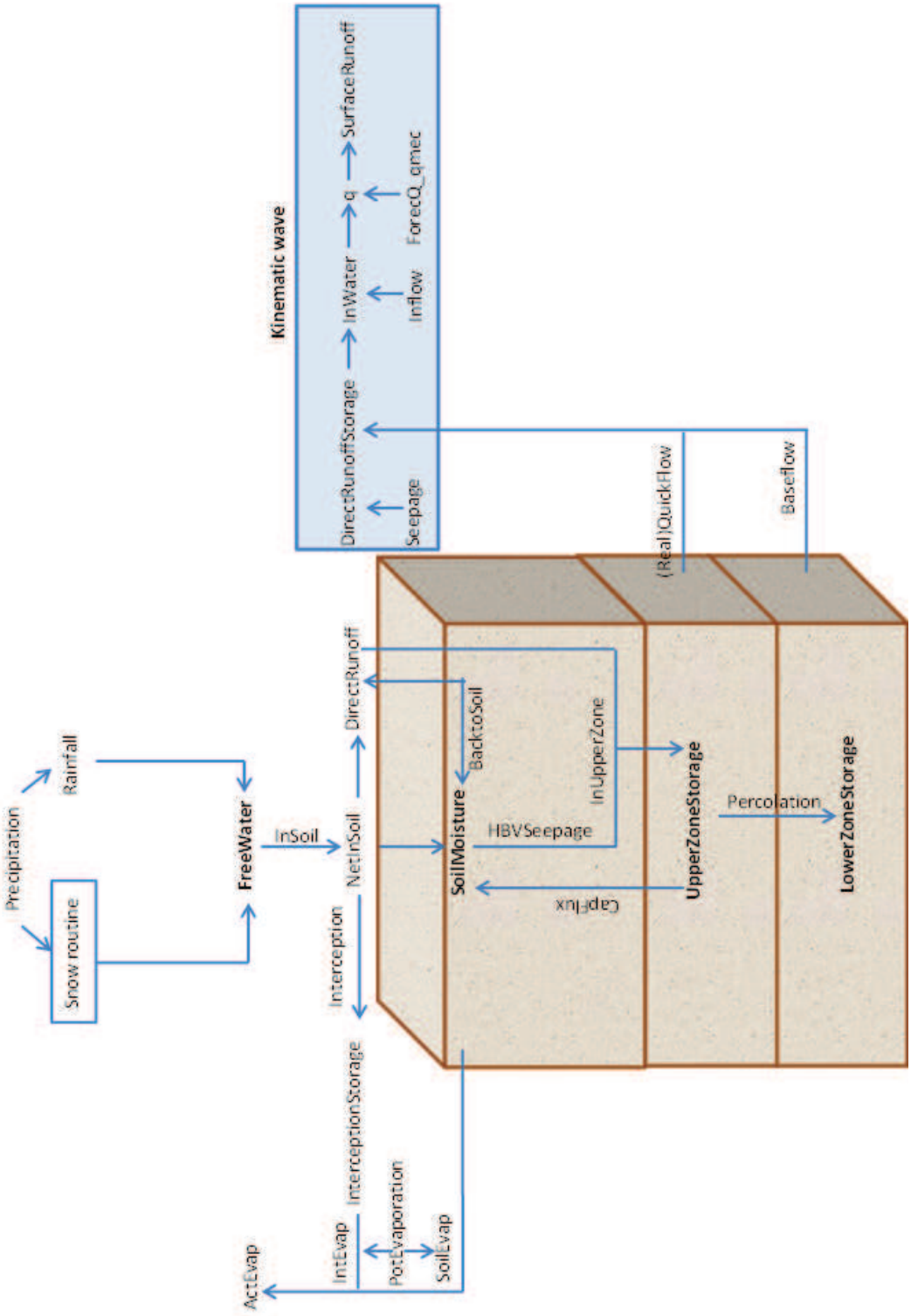


Figure D.1: Diagram of the wflow_hbv model per cell

RainFall: rainfall depth (mm)

FreeWater: free water content in snow (mm)

InSoil: abundant water in snow pack which goes into the soil (mm)

NetInSoil: net water which infiltrates into the soil (mm)

Interception: interception (mm)

InterceptionStorage: actual interception storage (mm)

IntEvap: actual evaporation from interception storage (mm)

PotEvaporation: potential evaporation (mm)

SoilMoisture: actual soil moisture (mm)

DirectRunoff: direct runoff (mm)

SoilEvap: actual soil evaporation (mm)

ActEvap: sum of evaporation components (mm)

Backtosoil: correction for extremely wet periods (mm)

HBVSeepage: runoff water from soil to the Upper zone (mm)

InUpperZone: total runoff water (mm)

CapFlux: capillary rise (mm)

UpperZoneStorage: water content of the Upper zone (mm)

Percolation: actual percolation to the lower zone (mm)

LowerZoneStorage: water content of the Lower zone (mm)

Baseflow: specific runoff (baseflow part) per cell (mm)

QuickFlow: specific runoff (quickflow part) (mm)

RealQuickFlow: quickflow if $K_{upper\,zone}$ is precalculated (mm)

DirectRunoffStorage: direction runoff (mm)

Seepage: seepage to the direct runoff storage (mm)

InWater: surface runoff (m^3/s)

Inflow: abstractions (m^3/s)

q: surface runoff per distance along the stream (m^3/s)

ForecQ_qmec: extra inflow to kinematic wave reservoir (m^3/s)

SurfaceRunoff: surface runoff in the kinematic wave (m^3/s)

Figure D.2: Description of the wflow_hbv variables

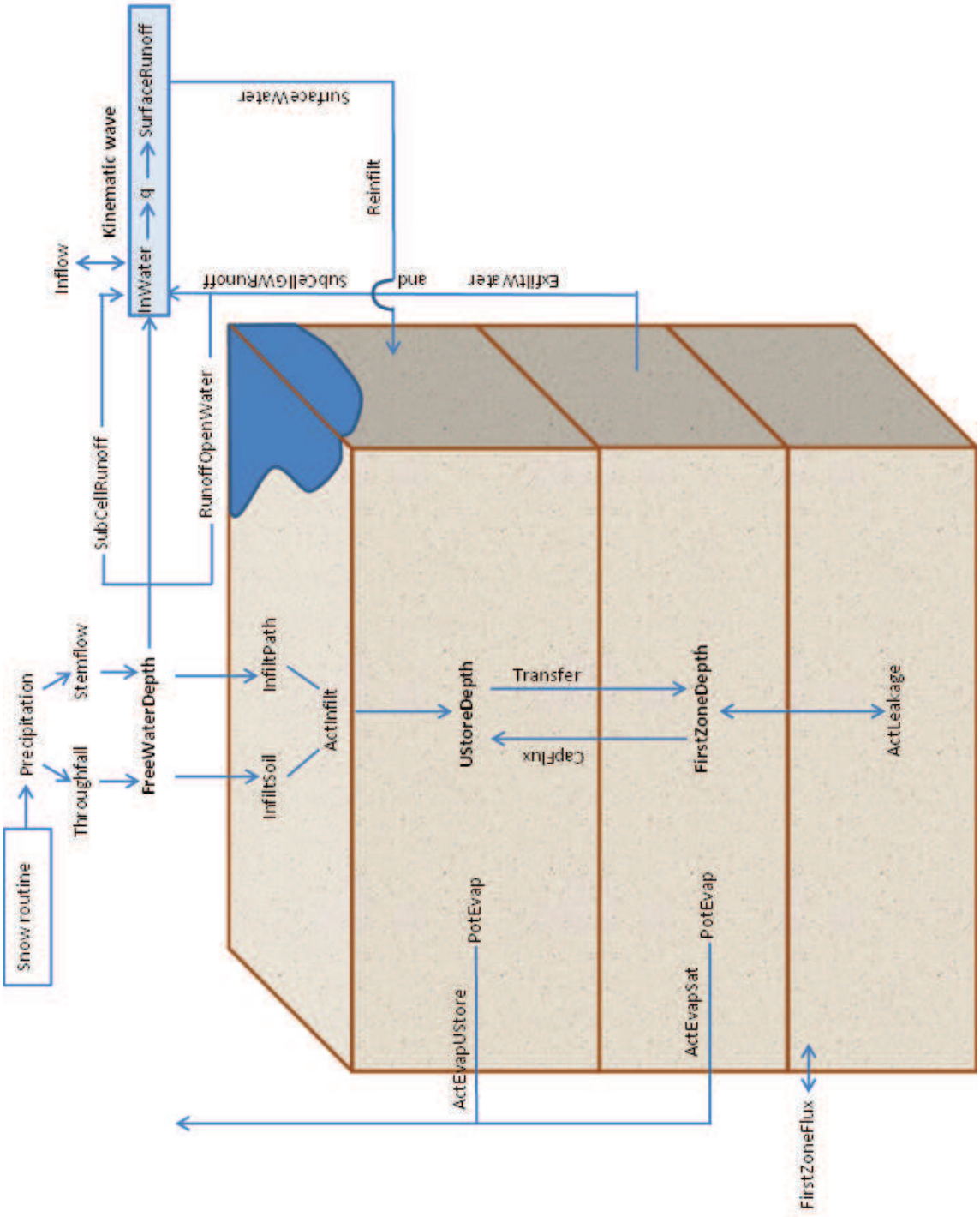


Figure D.3: Diagram of the wflow_sbm model per cell

FreeWaterDepth: freewater storage (mm)

SubcellRunoff: runoff water from the freewater storage per subcell (mm)

RunoffOpenWater: runoff onto water boddies and river network (mm)

InfiltSoil: infiltration in undisturbed soil (mm)

InfiltPath: infiltration in compacted areas (paths) (mm)

ActInfilt: total infiltration (mm)

UStoreDepth: amount of available water in the unsaturated zone (mm)

Reinfilt: water that re-infiltrates to the unsaturated zone (mm)

SurfaceWater: surface water from surface runoff (mm)

CapFlux: capillary flux from the saturated to the unsaturated zone (mm)

Transfer: downward flux from the unsaturated to the saturated zone (mm)

Potvap: potential evapotranspiration (mm)

ActEvapUStore: actual evaporation from the unsaturated zone (mm)

FirsZoneDepth: amount of water available in the saturated zone (mm)

SubCellGWRunoff: runoff water from the saturated zone per subcell (mm)

ActLeakage: actual leakage or seepage from/to the saturated zone (mm)

ActEvapSat: actual evaporation from the saturated zone (mm)

FirstZoneFlux: horizontal (downstream) transport of water from the saturated zone (mm)

ExfiltWater: exfiltration from the saturated store (mm)

InWater: surface runoff (m^3/s)

q: surface runoff per distance along stream (m^3/s)

SurfaceRunoffDyn: surface runoff in the dynamic reservoir (m^3/s)

SurfaceRunoff: surface runoff in the kinematic wave (m^3/s)

Figure D.4: Description of the wflow_sbm variables

Appendix E

The forecasting platform

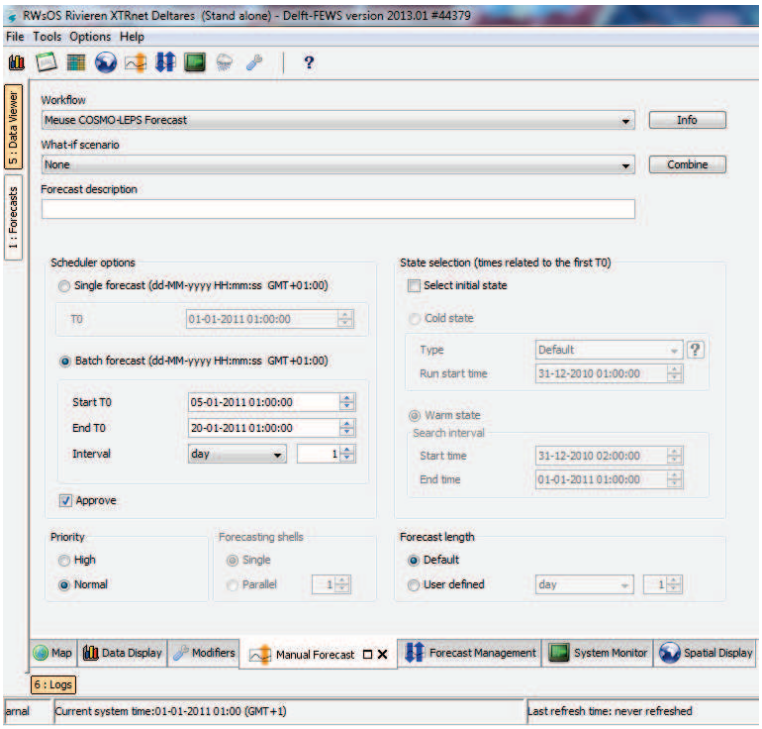


Figure E.1: Delft-FEWS manual forecast

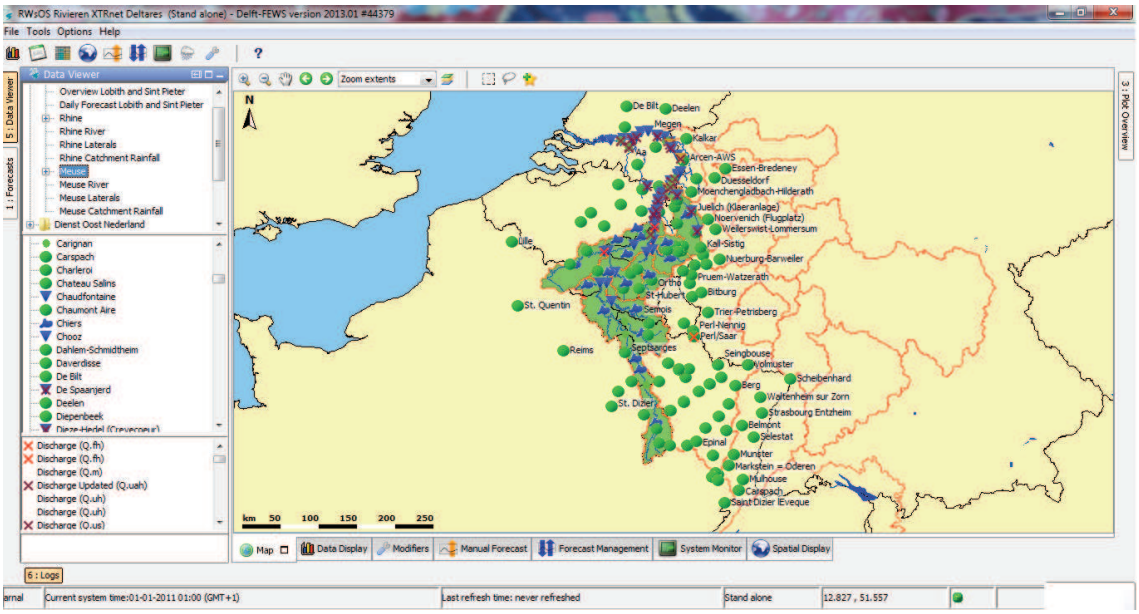


Figure E.2: Delft-FEWS map, the Meuse River basin

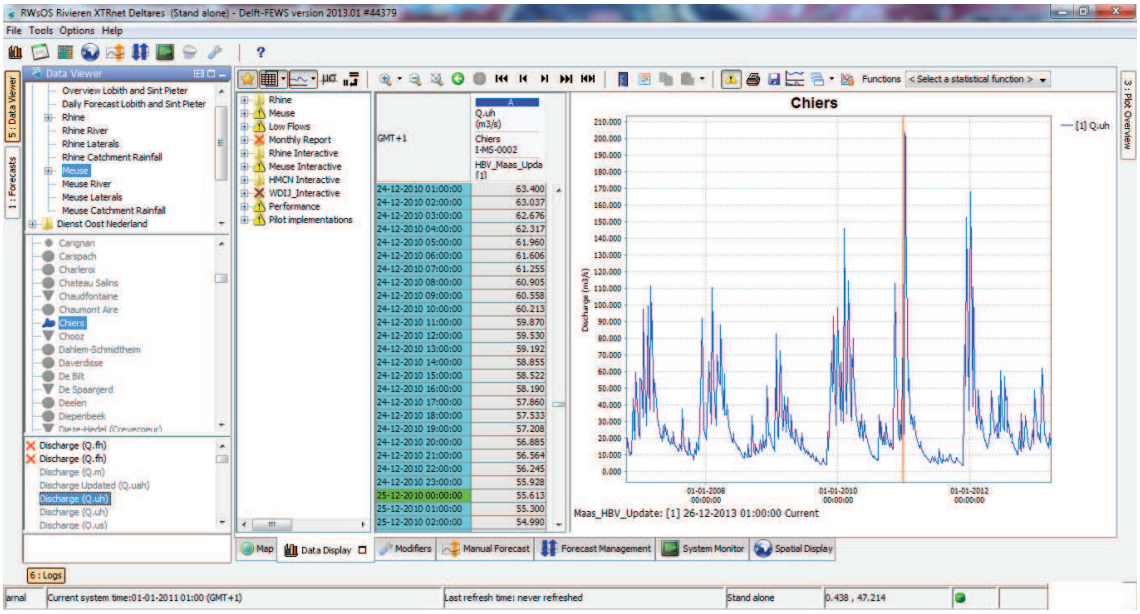


Figure E.3: Delft-FEWS data display

Appendix F

Data import and transformation

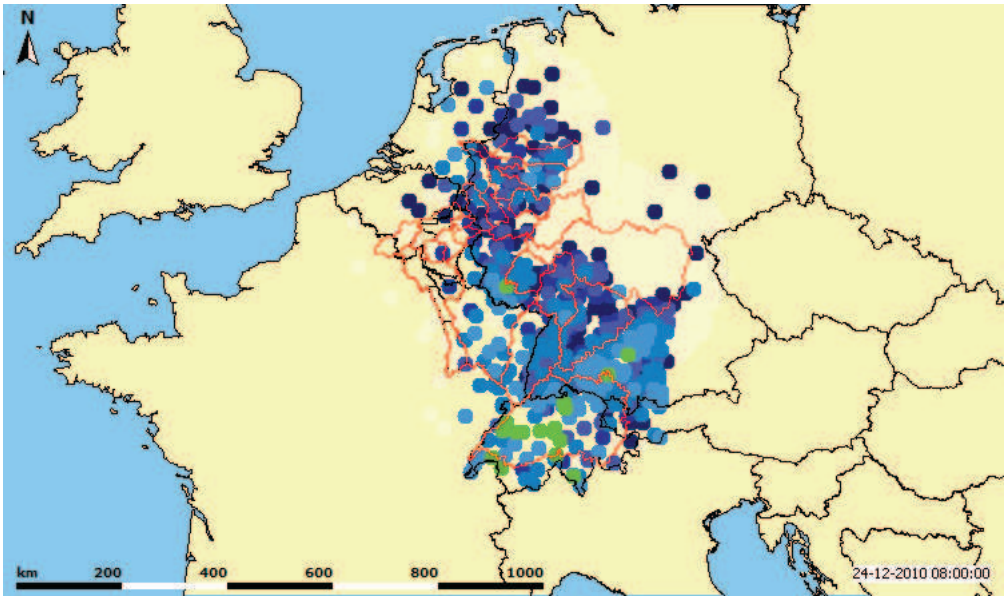


Figure F.1: Observed precipitation at the stations locations

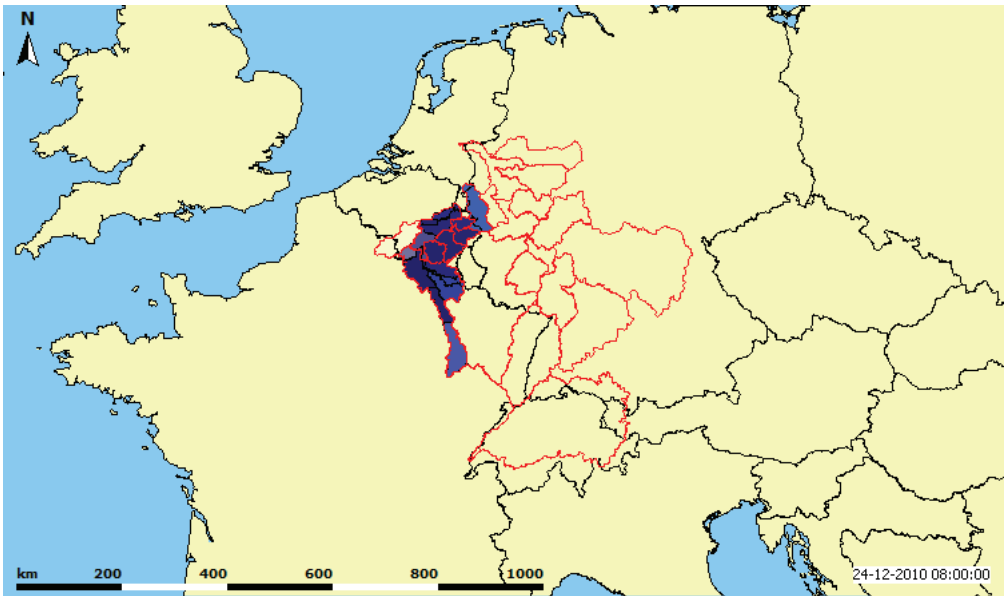


Figure F.2: Regnie precipitation ouput for the operational hbv model

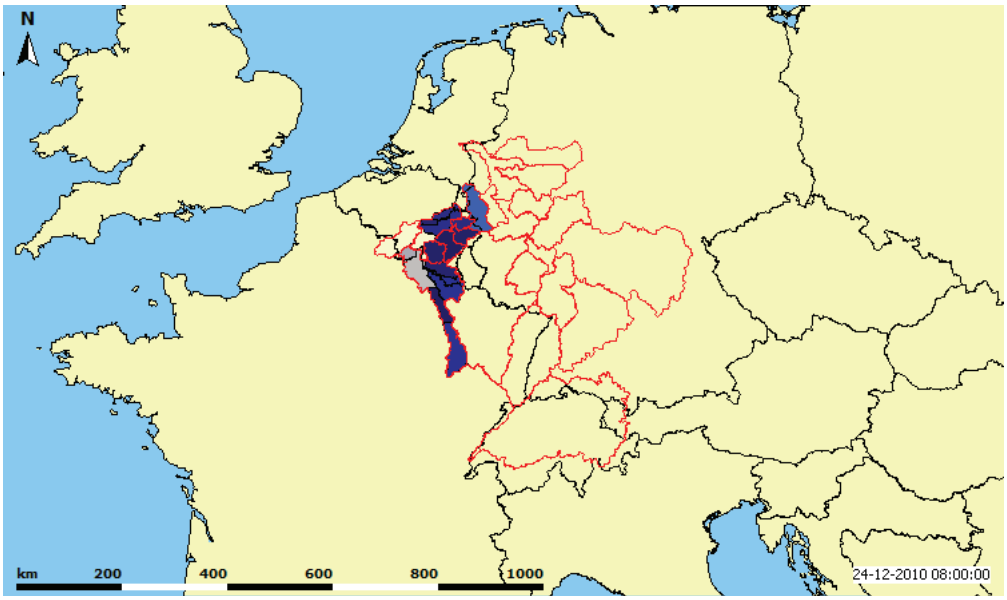


Figure F.3: Kriging precipitation ouput for the operational hbv model

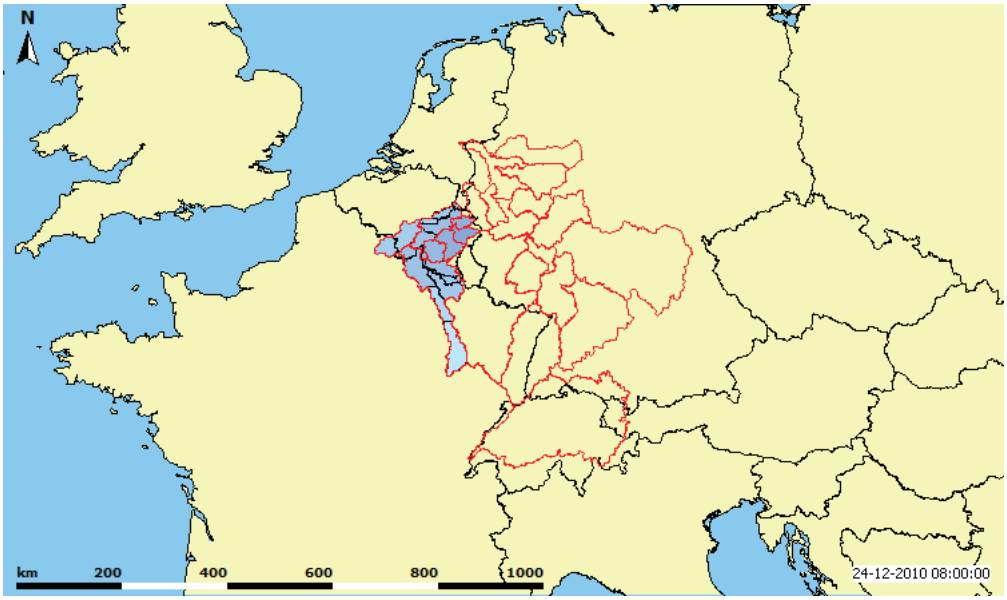


Figure F.4: Kriging temperature ouput for the operational hbv model

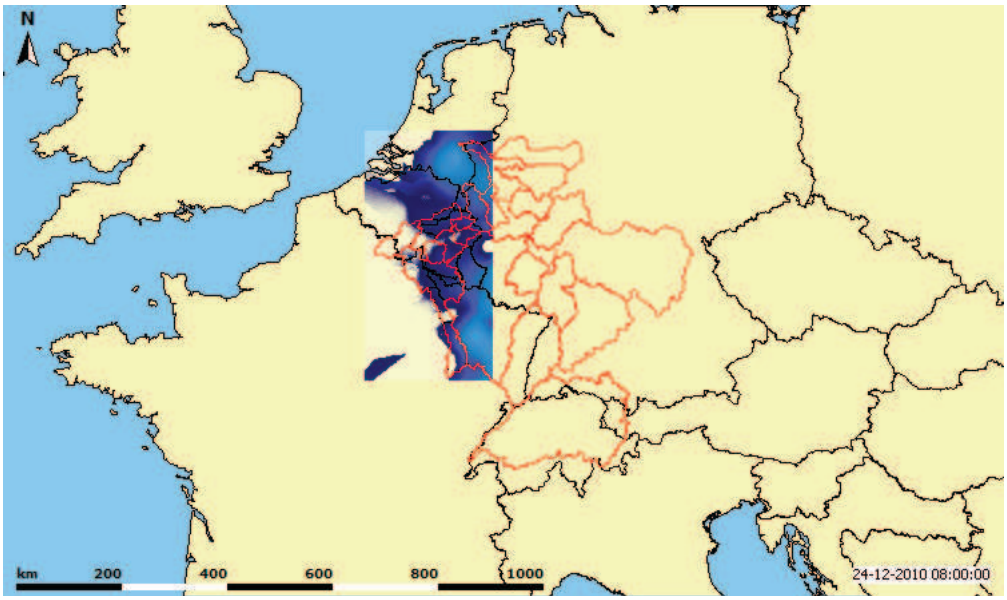


Figure F.5: Kriging precipitation ouput for the wflow models

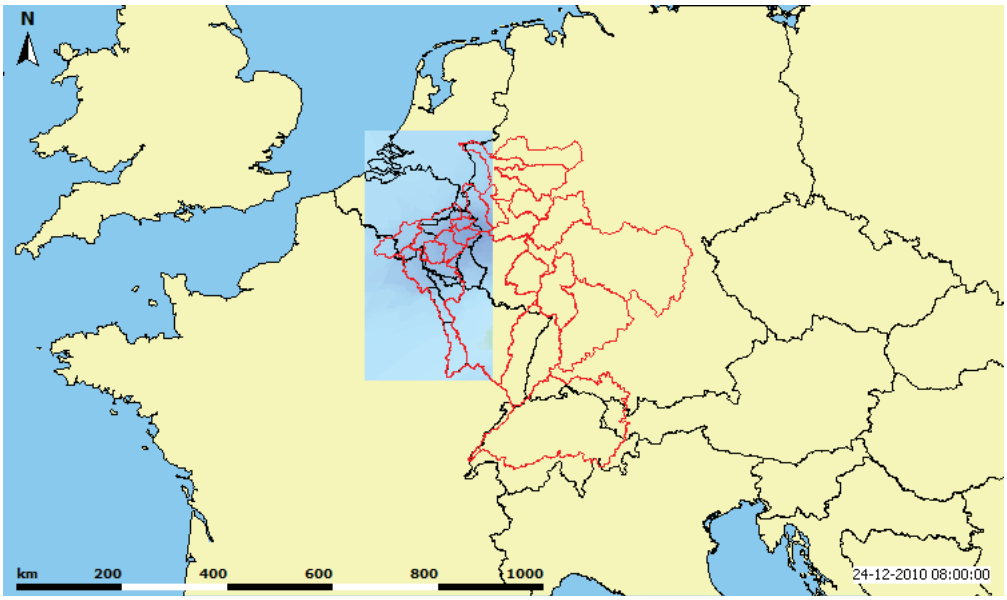


Figure F.6: Kriging temperature ouput for the wflow models

Appendix G

Calibration of the wflow_gr4 model

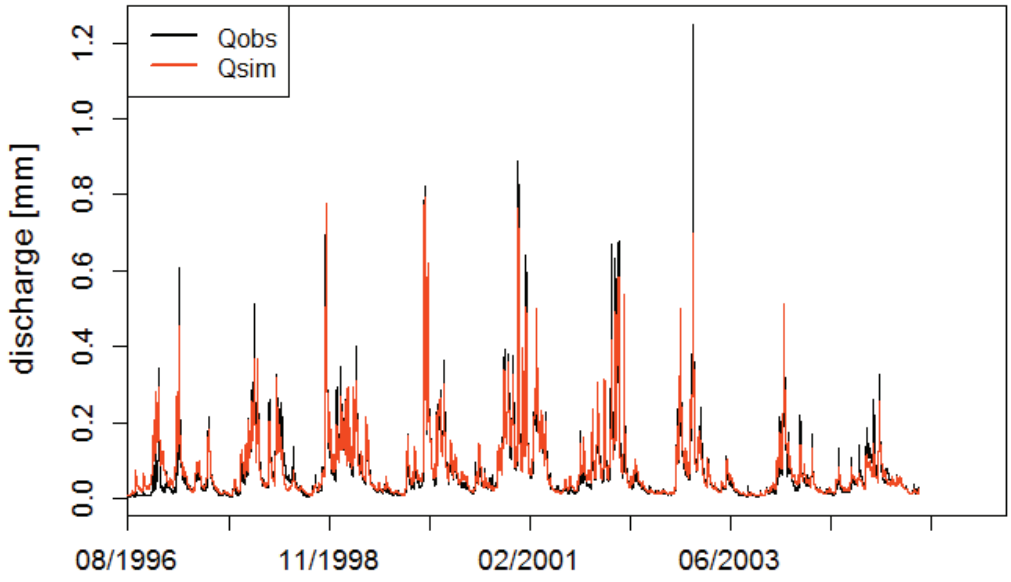


Figure G.1: Calibration run period for the Semois in Haulmé

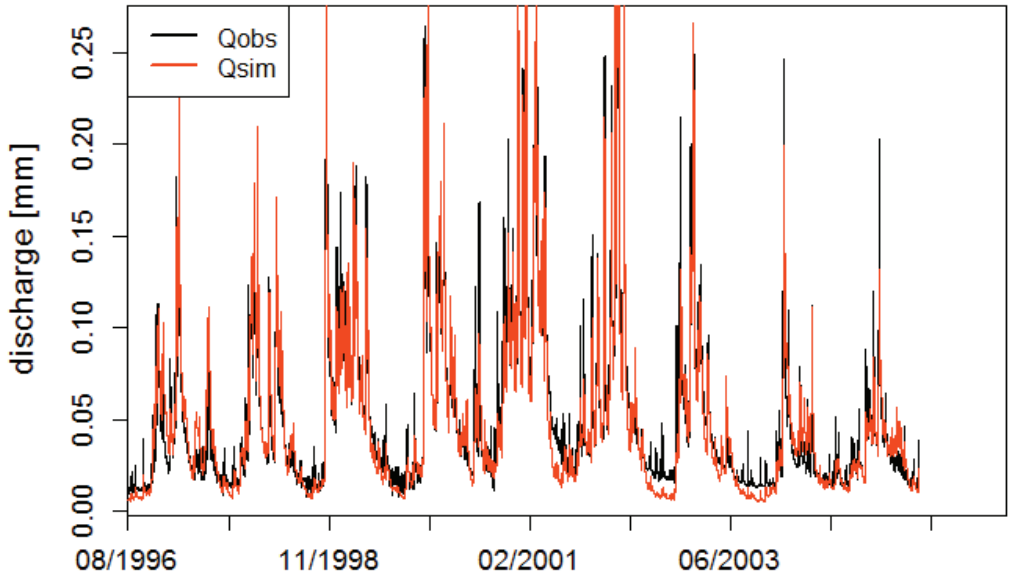


Figure G.2: Calibration run period for the Chiers in Carignan

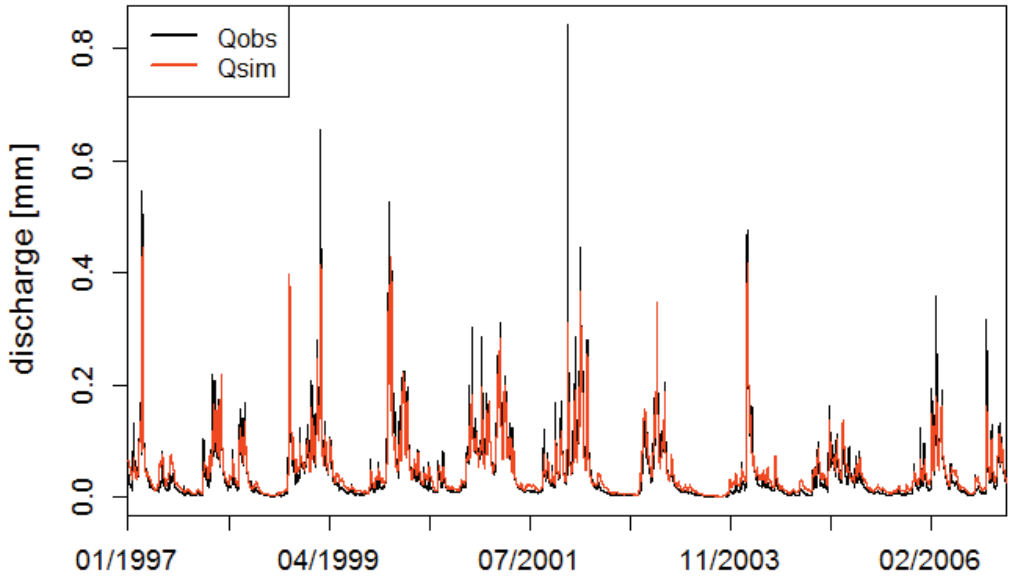


Figure G.3: Calibration run period for the Meuse in Saint-Mihiel

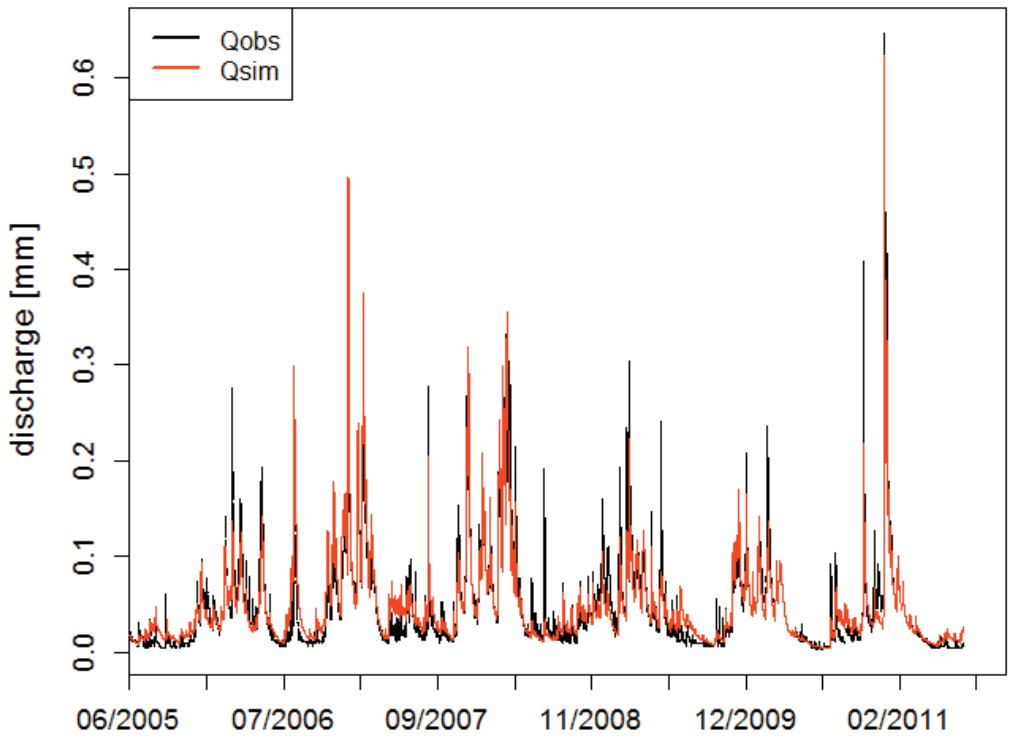


Figure G.4: Calibration run period for the Ourthe in Tabreux

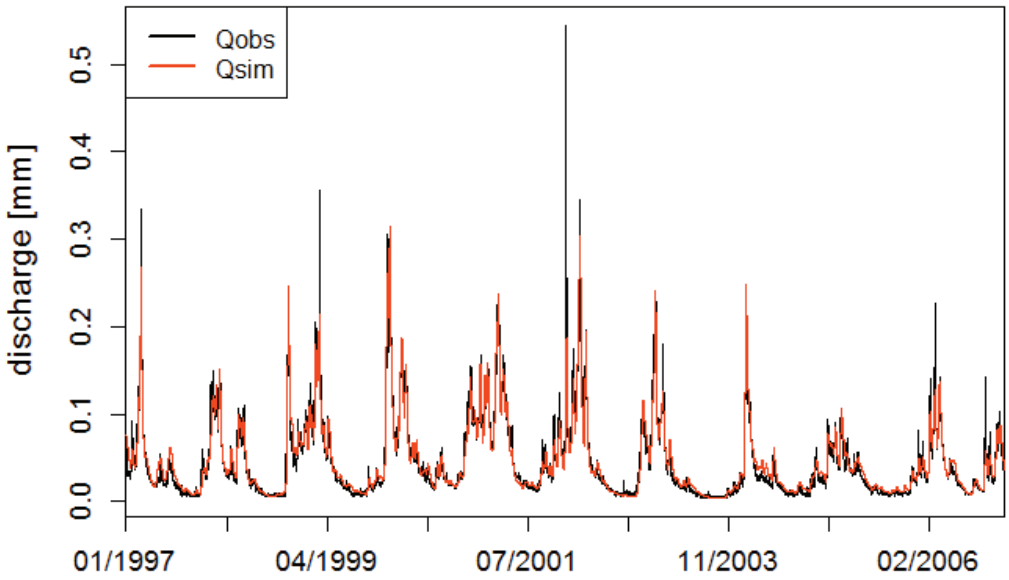
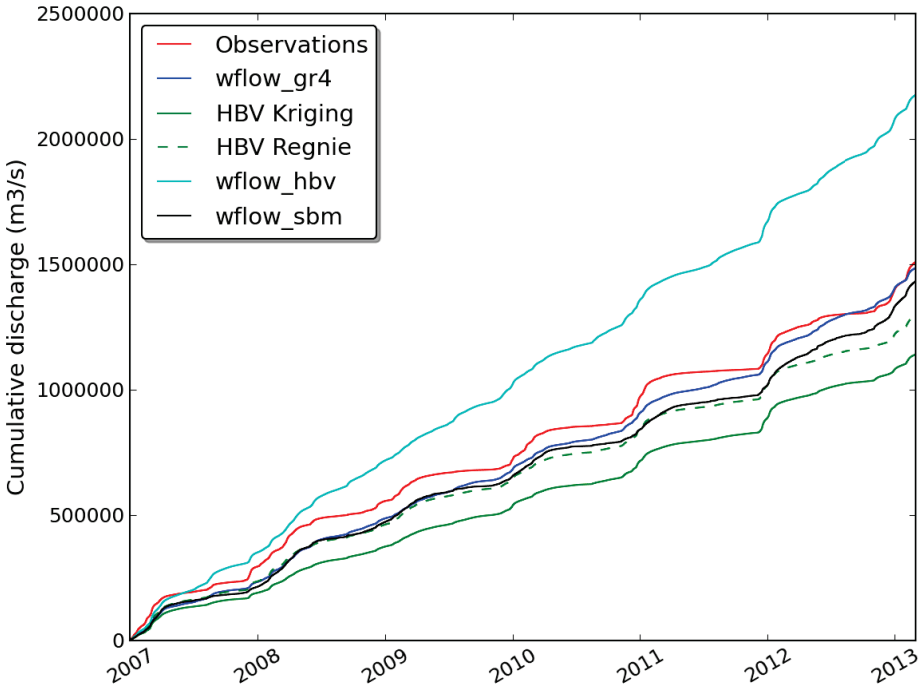


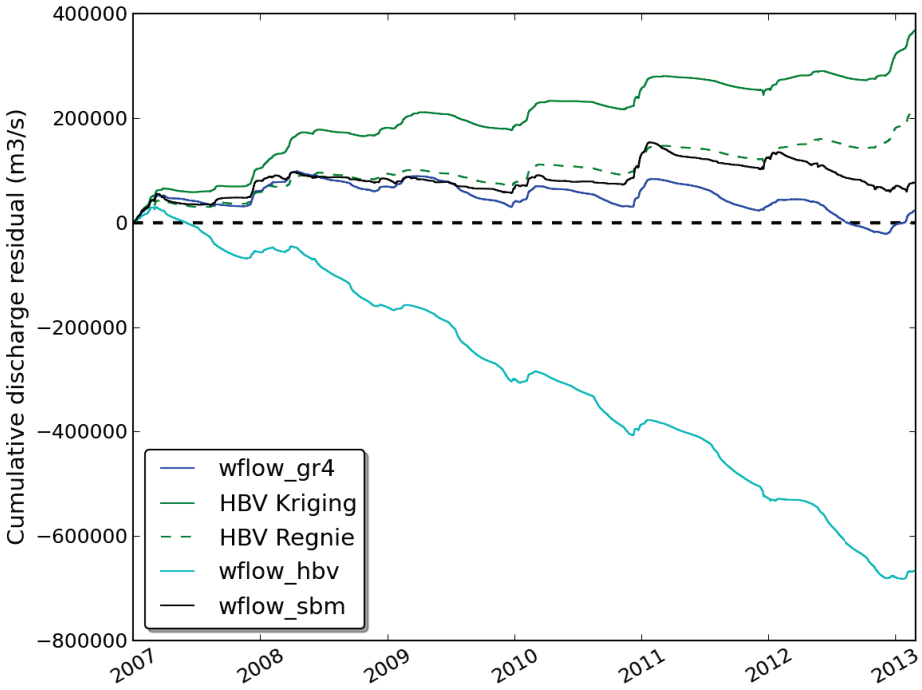
Figure G.5: Calibration run period for the Meuse in Stenay

Appendix H

Simulations results

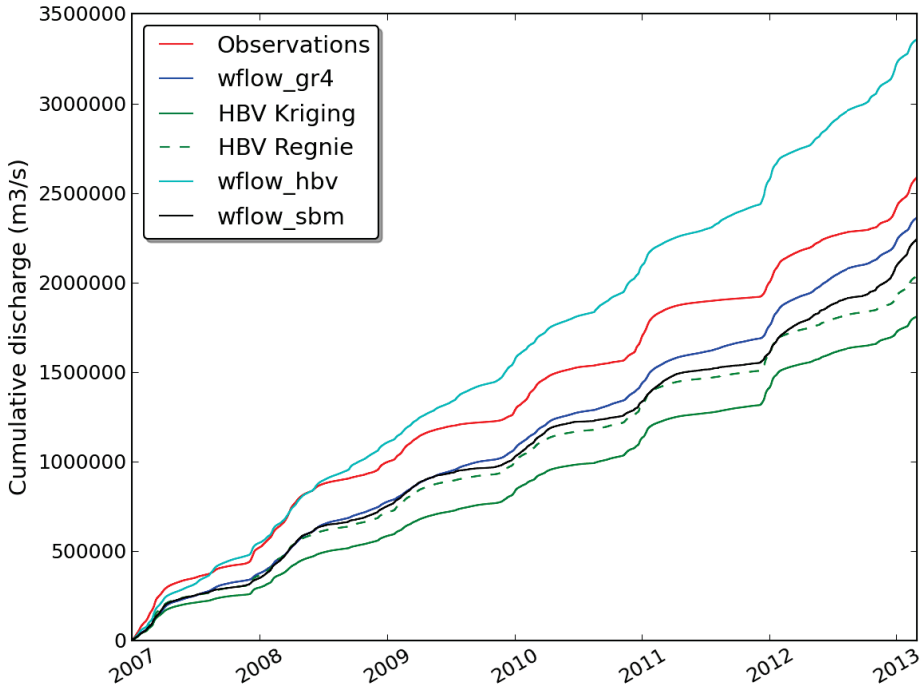


(a) Cumulative discharge

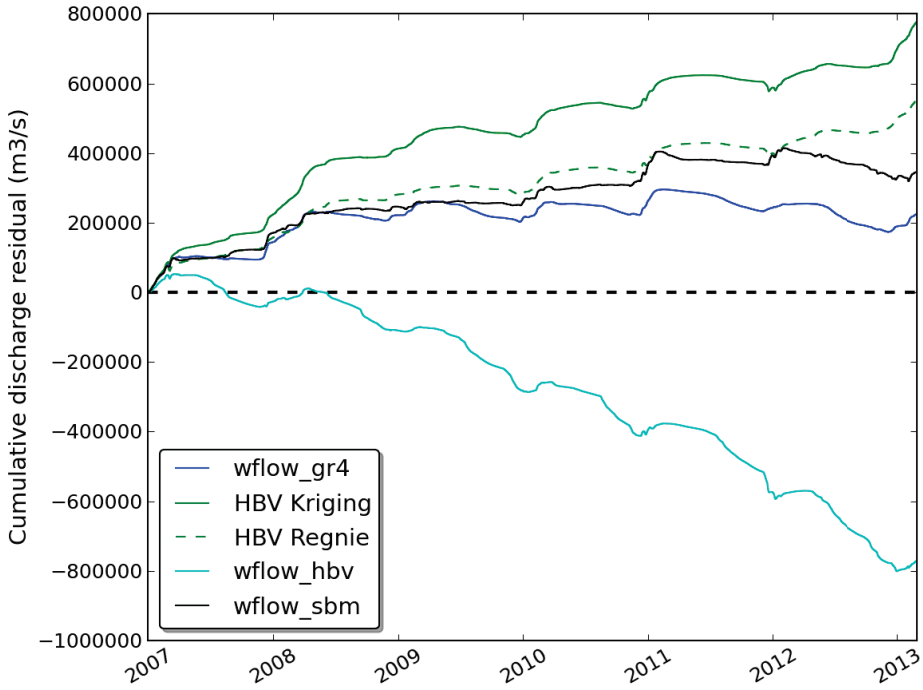


(b) Cumulative residual

Figure H.1: The (a) cumulative discharge and the (b) cumulative residual for the Meuse in Saint-Mihiel

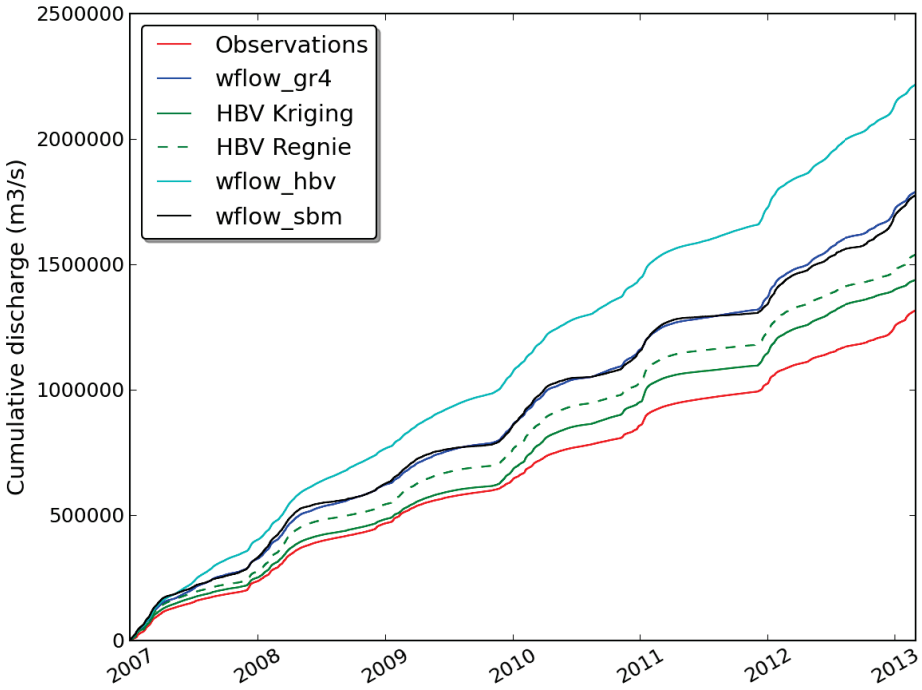


(a) Cumulative discharge

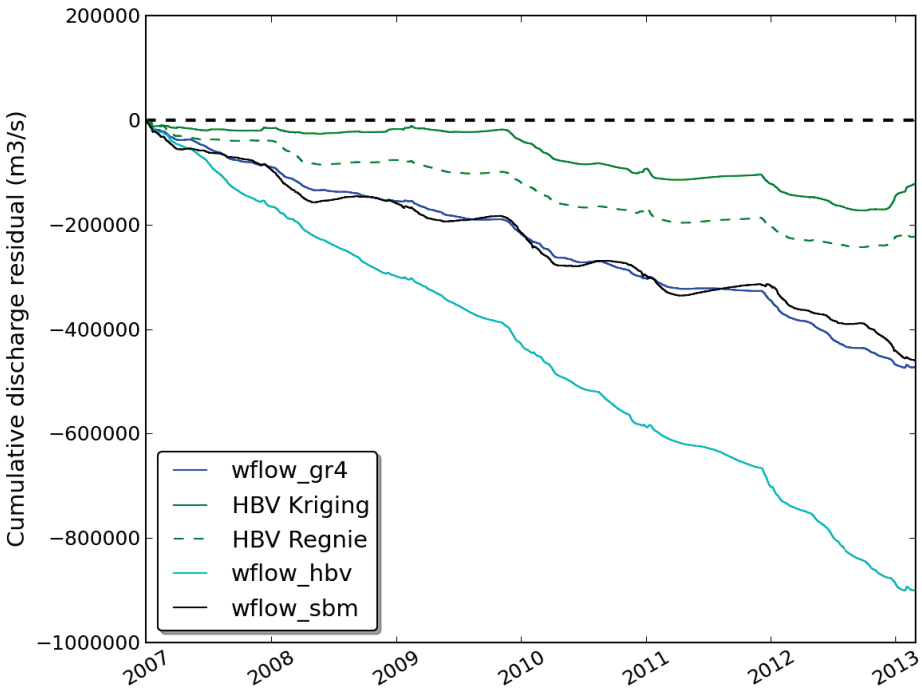


(b) Cumulative residual

Figure H.2: The (a) cumulative discharge and the (b) cumulative residual for the Meuse in Stenay

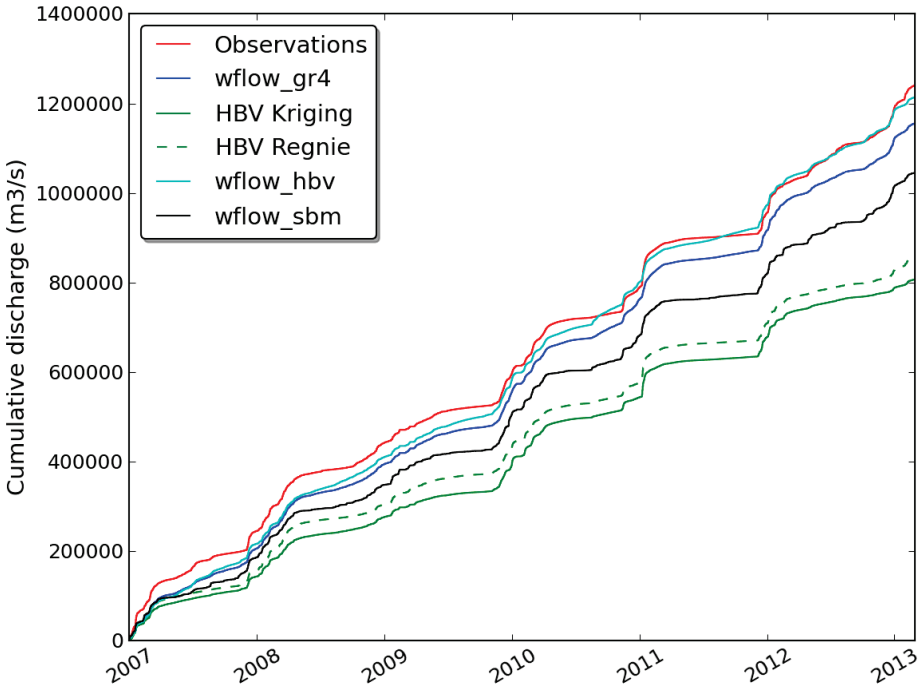


(a) Cumulative discharge

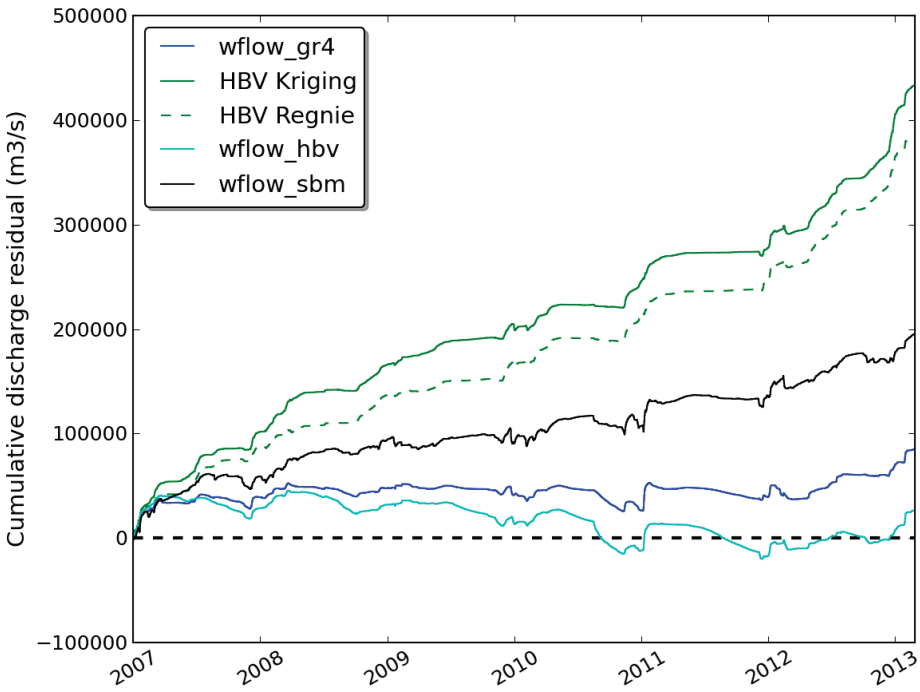


(b) Cumulative residual

Figure H.3: The (a) cumulative discharge and the (b) cumulative residual for the Chiers in Carignan

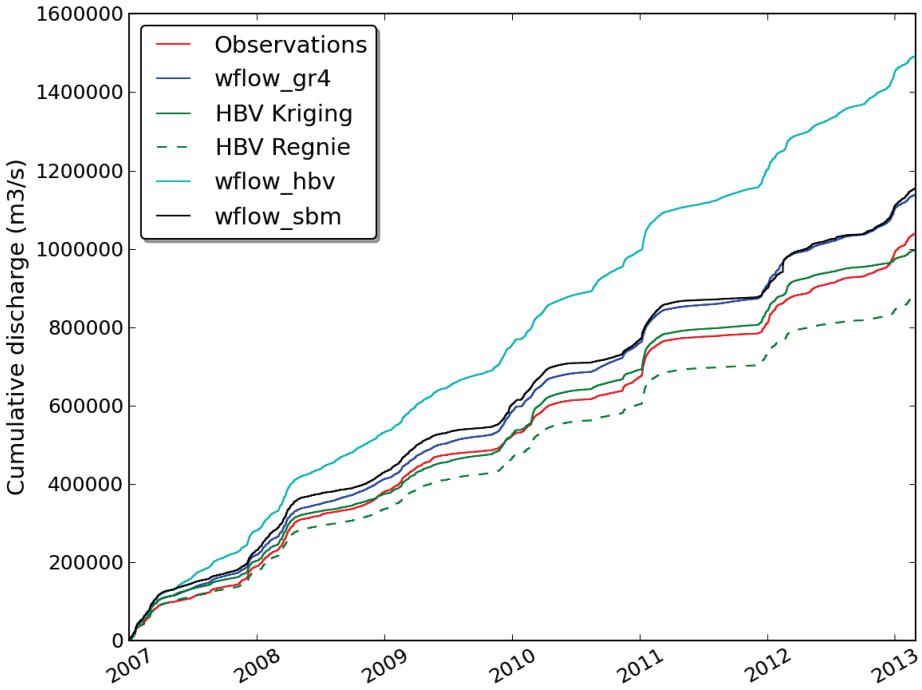


(a) Cumulative discharge

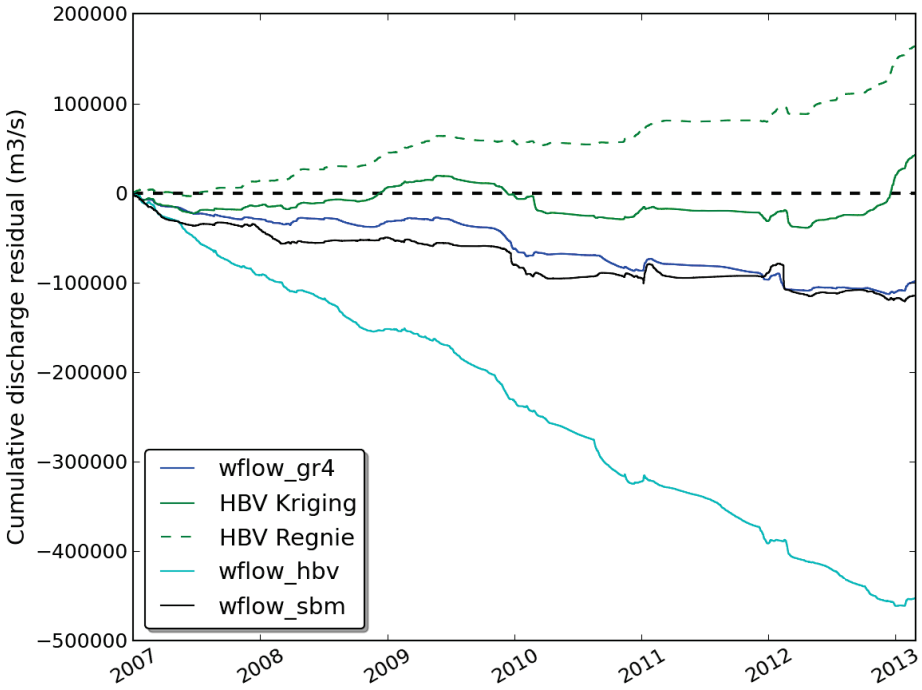


(b) Cumulative residual

Figure H.4: The (a) cumulative discharge and the (b) cumulative residual for the Semoy in Haulmé



(a) Cumulative discharge



(b) Cumulative residual

Figure H.5: The (a) cumulative discharge and the (b) cumulative residual for the Ourthe in Tabreux