Evaluation of discharge extremes in the Meuse river and her tributaries

4.3.2 Hourly

Similarly, the results of the model performance at the hourly timestep for the different stations are shown in Figure 4-23 to Figure 4-36. The main observations per station are summarised in Table 4-6. Overall, the observations made for the daily model in Table 4-4, also apply for the hourly model, except for the Meuse at Goncourt, where we see differences in model performance between both model timesteps. This could be related to the different input data (genRe for hourly versus E-OBS for daily) and the limited availability of open meteorological data in the French part of the Meuse basin.

Table 4-6 Summary of hourly model performance based on a visual inspection of the signature plots

Station	Catchment area [km^2]	Low flows	Average flows (mean monthly and cumulative flows)	High flows
Geul at Meerssen	338	good	overestimation of winter flows and cumulative flows	good
Meuse at Goncourt	364	good	underestimation of cumulative flows and winter flows	underestimation of the highest flow
Vesdre at Chaudfontaine	683	underestimation of low flows	good	underestimation of high flows
Ourthe at Tabreux	1607	good	good	good
Sambre at Salzinne	2842	good	good (slight underestimation of winter flows)	good
Meuse at Chooz	10120	good	model slightly underestimates winter flows and cumulative flows	good
Meuse at St Pieter	21233	good	good	good

The scores of several performance indicators (KGE of the daily flows, NSE of daily flows, NSE of log of the daily flows, NSE of the plotting position of the NM7Q and NSE of the plotting position of the max annual flow) are shown in Table 4-5.

Table 4-7 Summary of hourly model performance based on the scores of the performance indicators

	NSE	KGE	NSElog	NM7Q ([3])	MAXQ ([3])
Geul at Meerssen	0.12	0.5	0.17	-0.66	0.67
Meuse at Goncourt	0.68	0.62	0.86	0.8	0.13
Vesdre at Chaudfontaine	0.73	0.75	0.77	0.15	0.16
Ourthe at Tabreux	0.84	0.9	0.88	0.9	0.43
Sambre at Salzinnes	0.65	0.77	0.33	0.81	0.66
Meuse at Chooz	0.84	0.8	0.86	0.74	0.69
Meuse at St Pieter	0.83	0.85	0.82	0.47	0.54

[3] NM7Q refers to the NSE of the plotting position of the NM7Q and MAXQ refers to the NSE of the plotting position of the max annual flow.

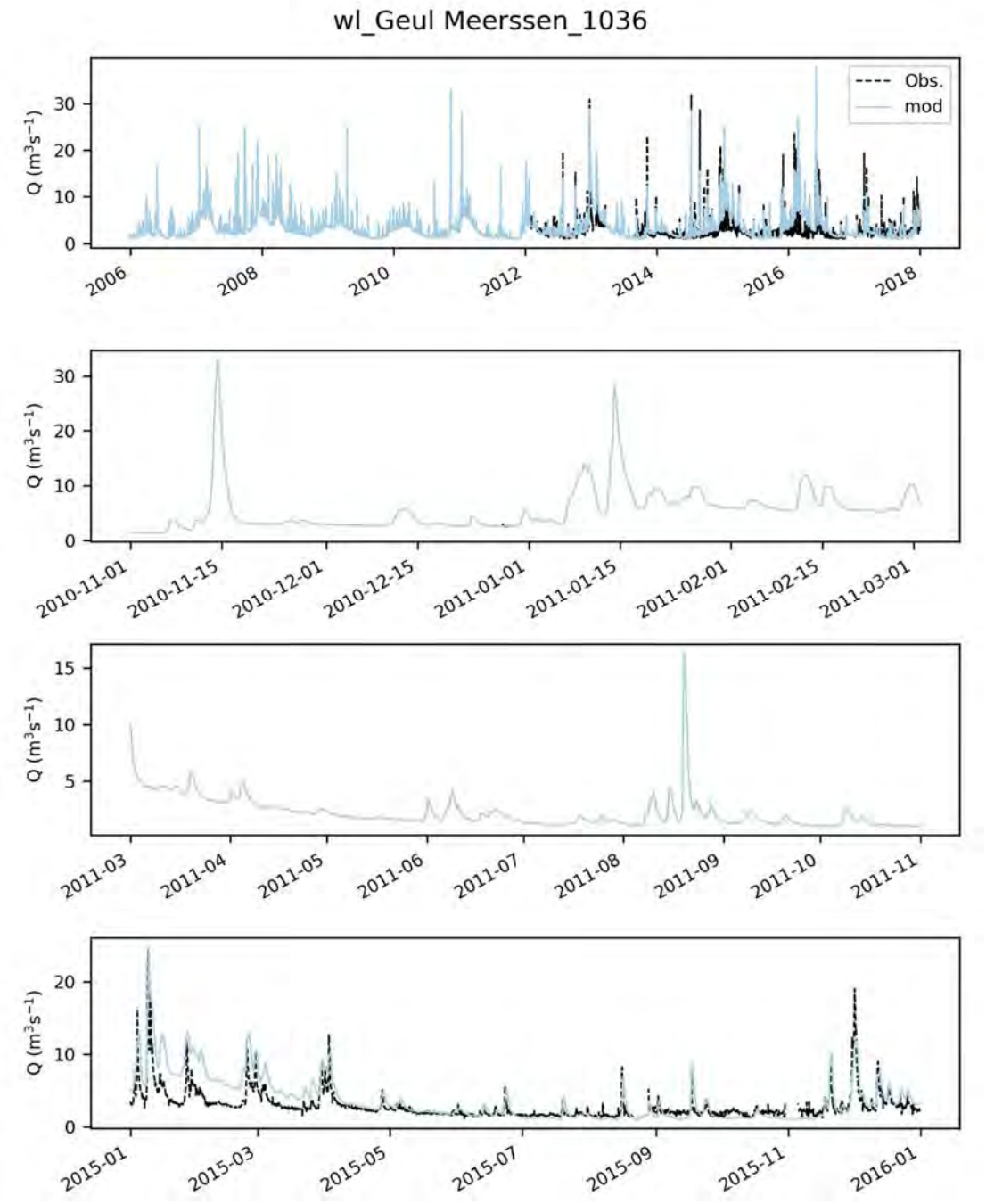


Figure 4-23 Modelled (blue) and observed (black) hydrographs for the Geul at Meerssen.

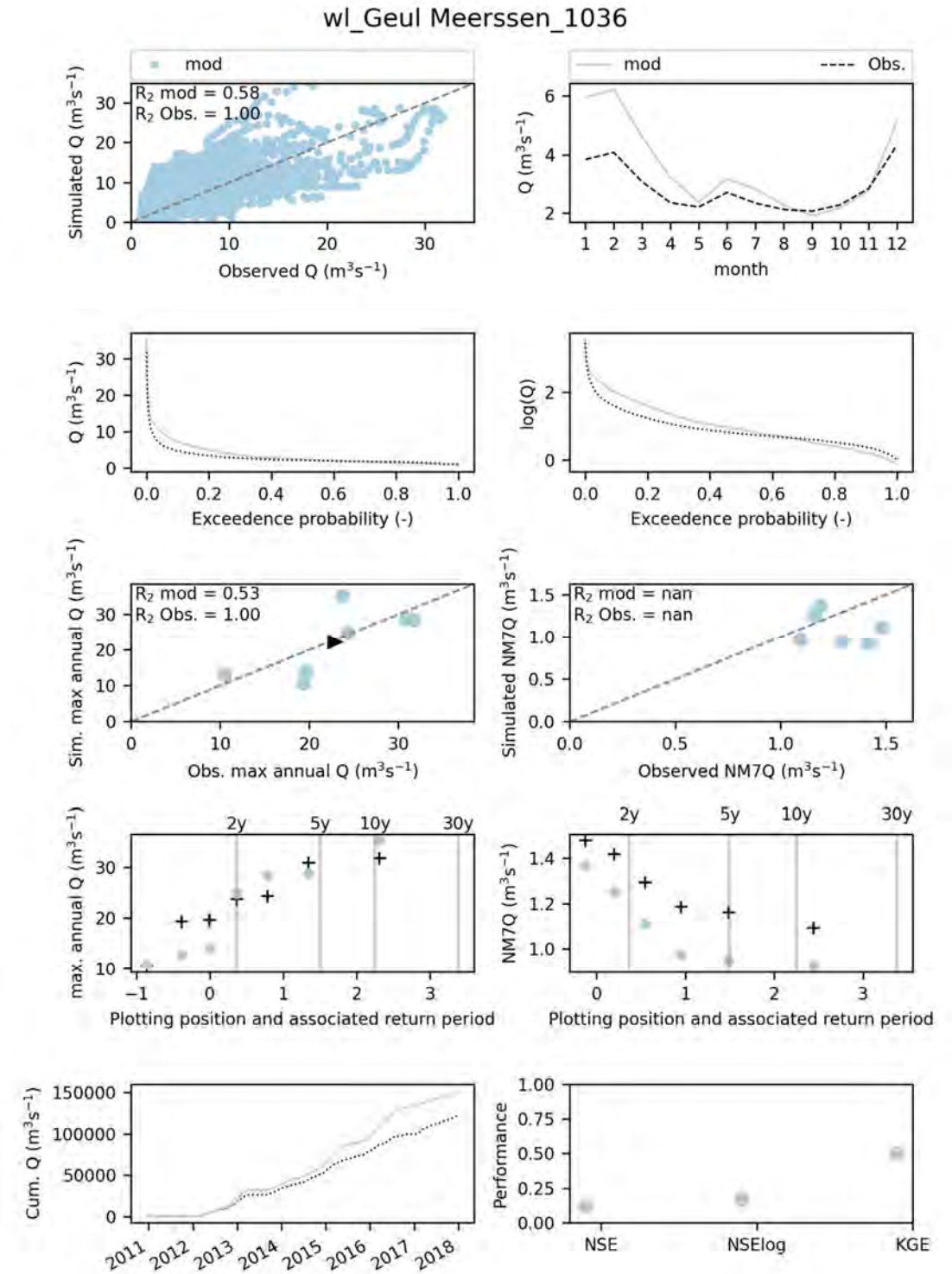


Figure 4-24 Modelled (blue) and observed (black) signatures for the Geul at Meerssen.

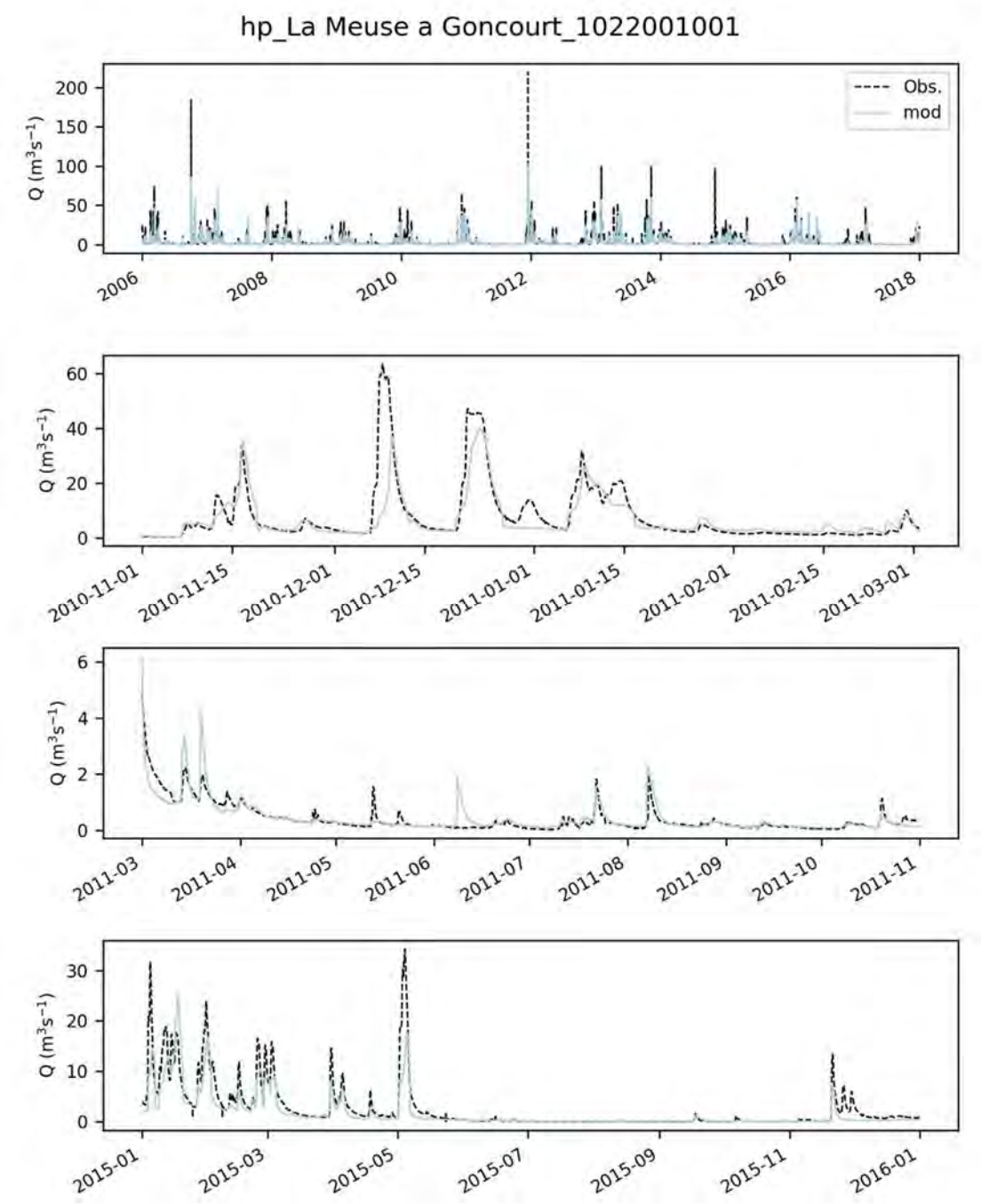


Figure 4-25 Modelled (blue) and observed (black) hydrographs for the Meuse at Goncourt.

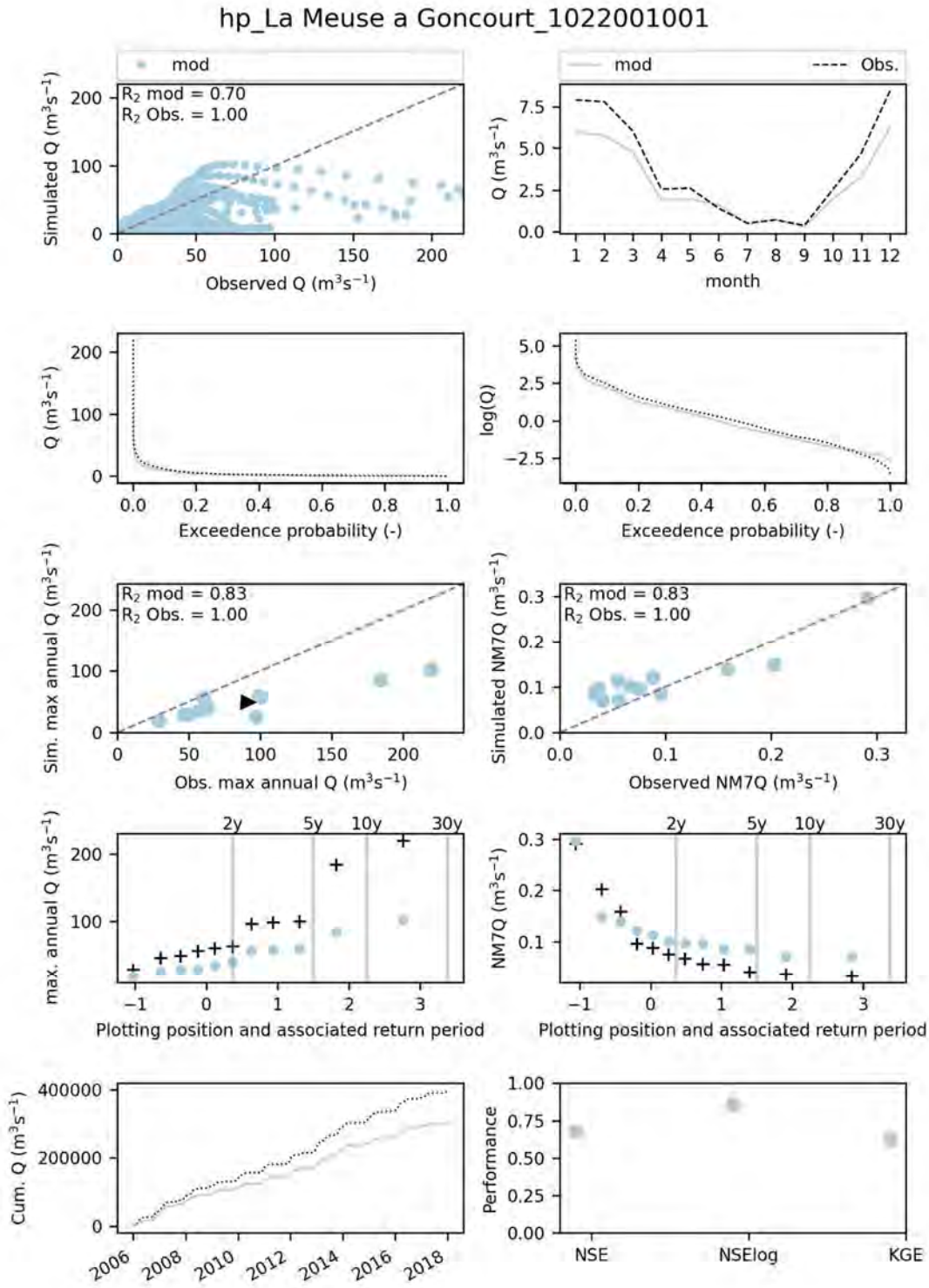


Figure 4-26 Modelled (blue) and observed (black) signatures for the Meuse at Goncourt.

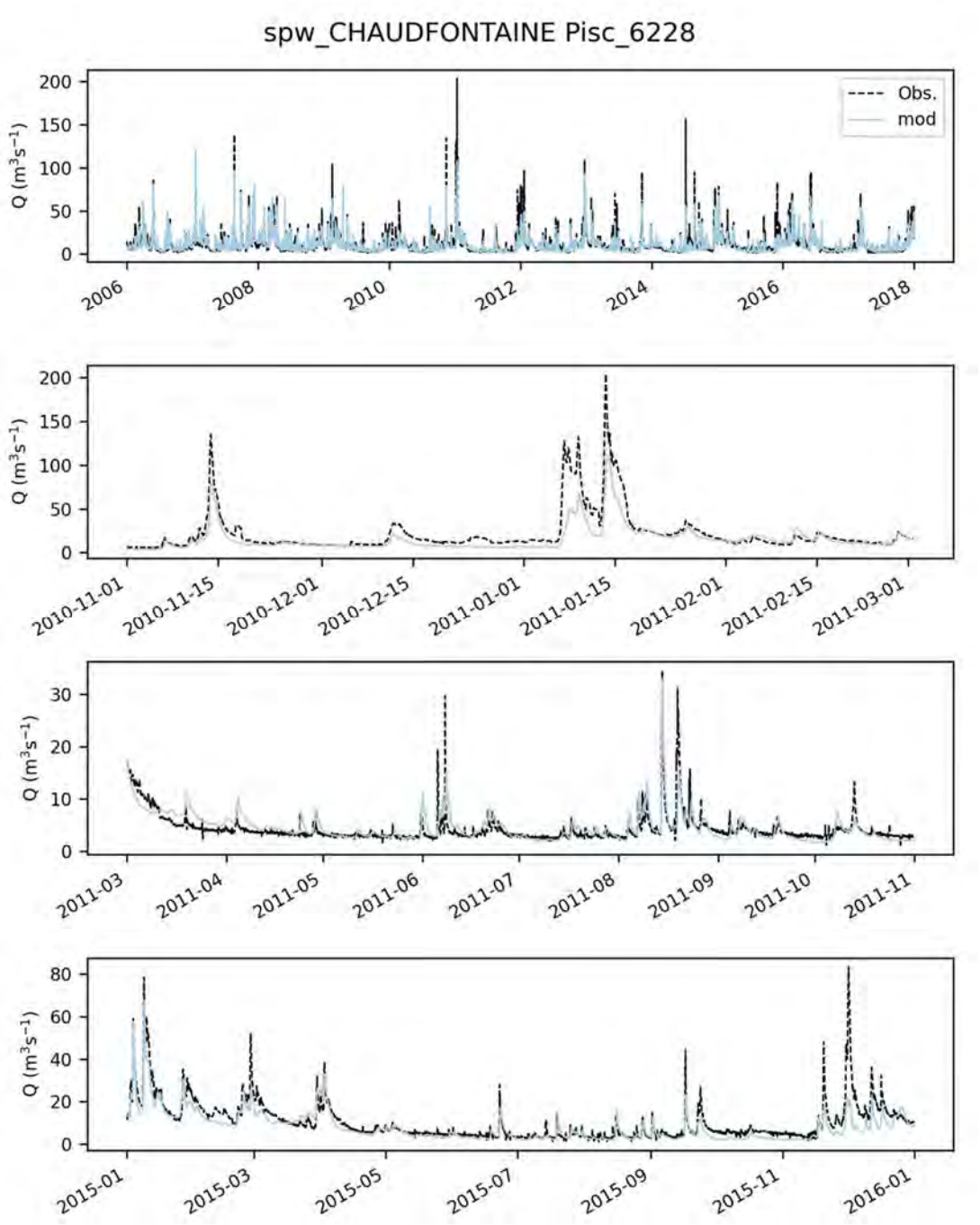


Figure 4-27 Modelled (blue) and observed (black) hydrographs for the Vesdre at Chaudfontaine.

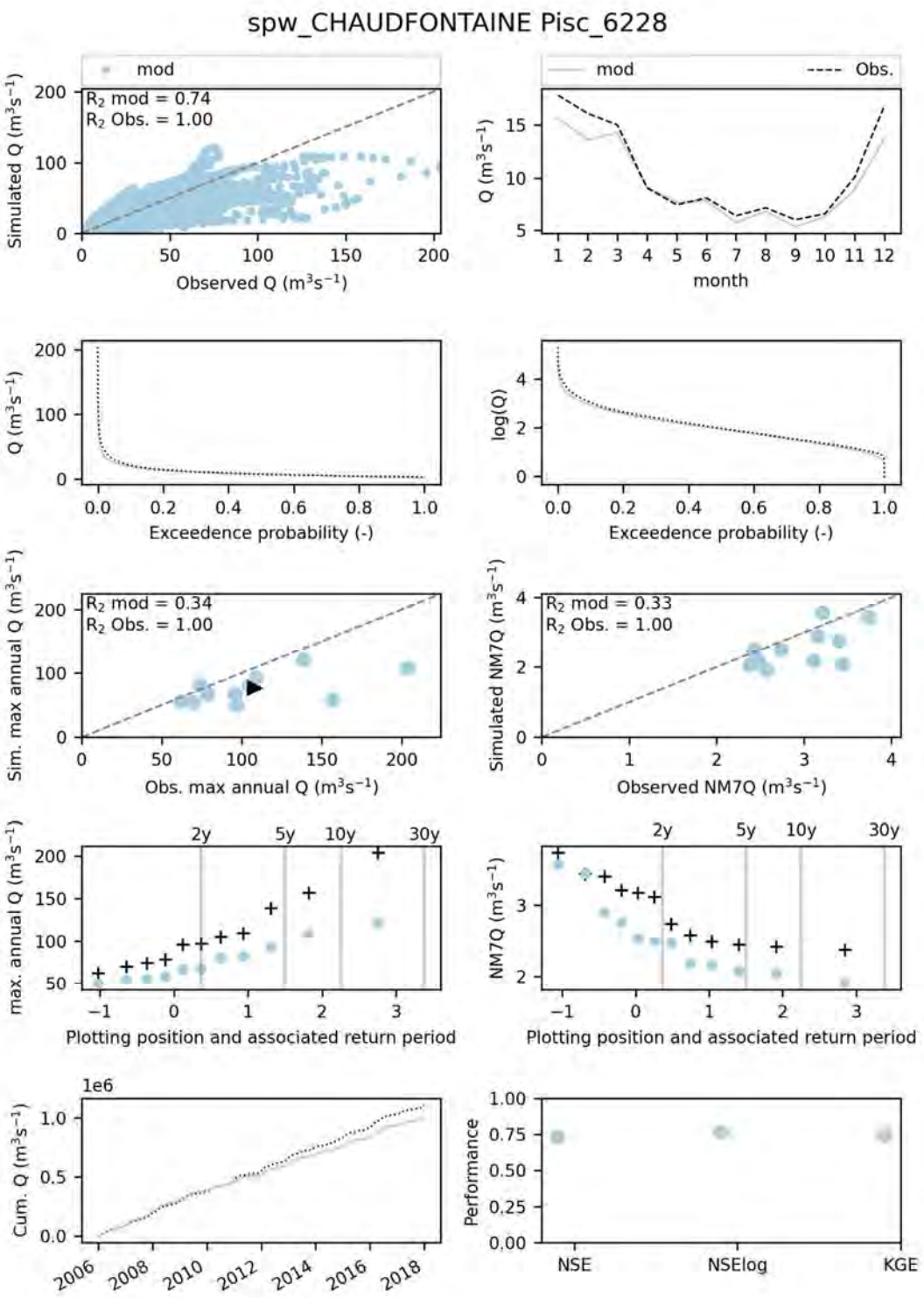


Figure 4-28 Modelled (blue) and observed (black) signatures for the Vesdre at Chaudfontaine.

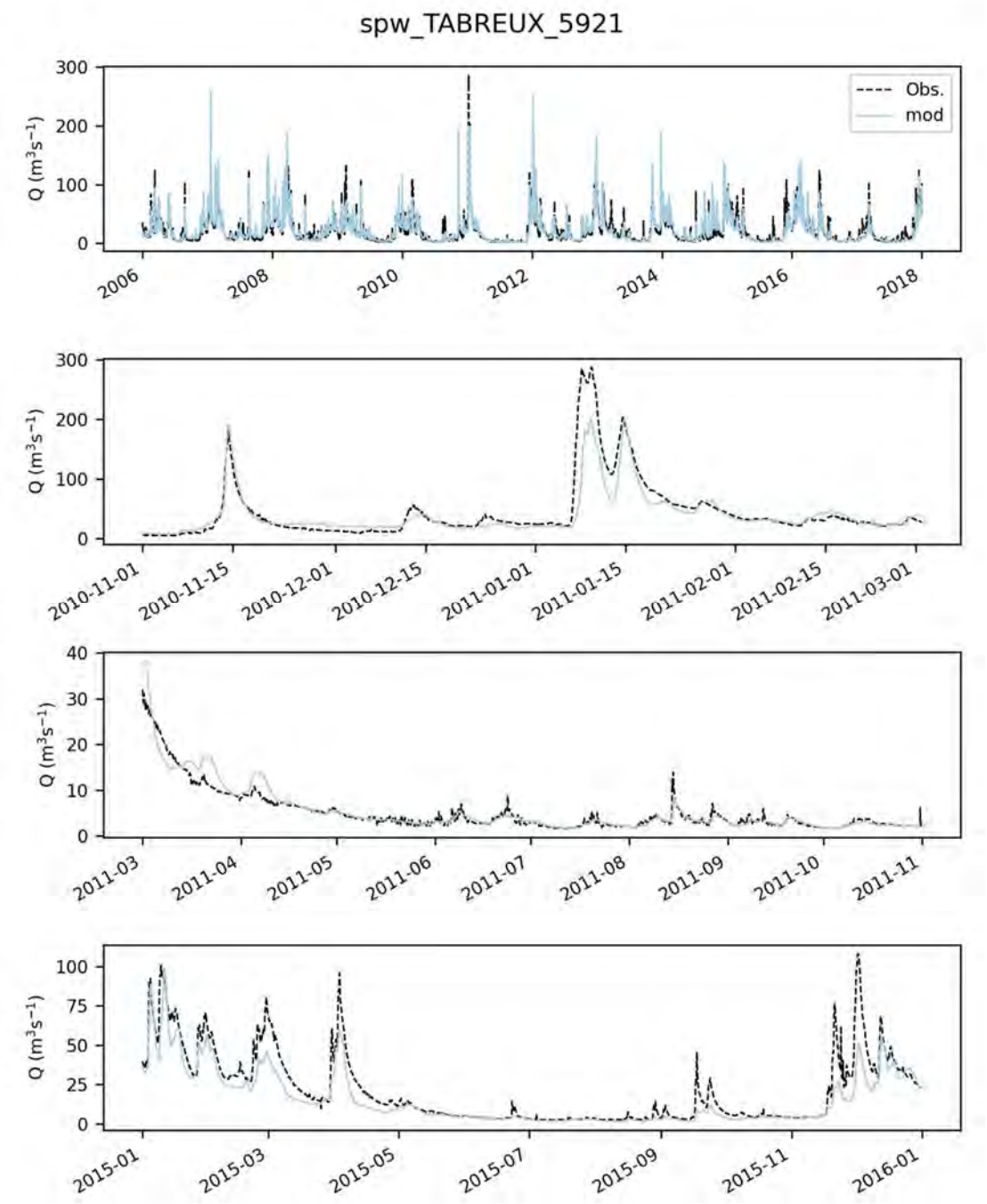


Figure 4-29 Modelled (blue) and observed (black) hydrographs for the Ourthe at Tabreux.

Deltares
Royal Netherlands Meteorological Institute Ministry of Infrastructure and the Environment
Rijkswaterstaat Ministry of Infrastructure and Water Management
Deltares

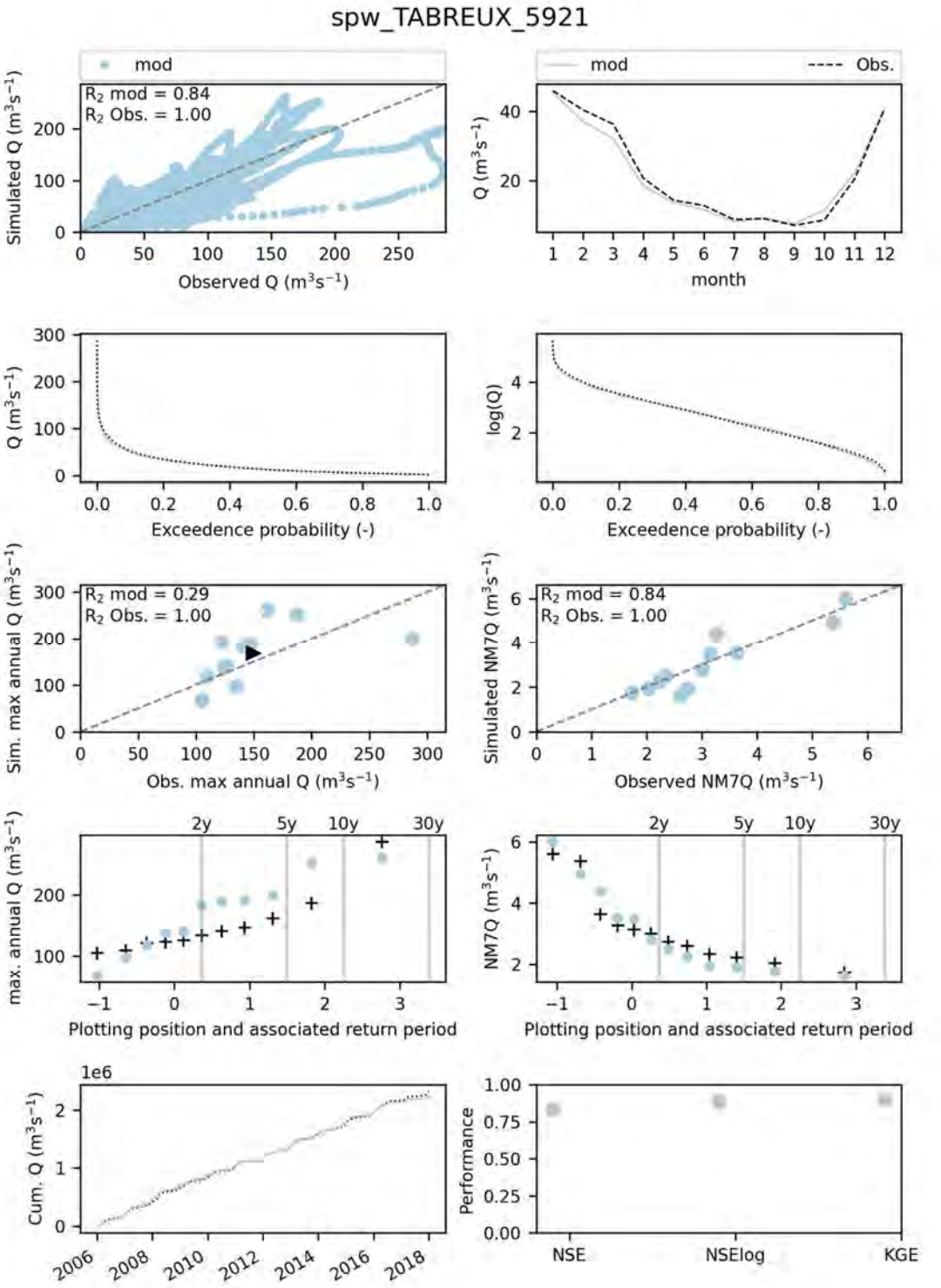


Figure 4-30 Modelled (blue) and observed (black) signatures for the Ourthe at Tabreux.

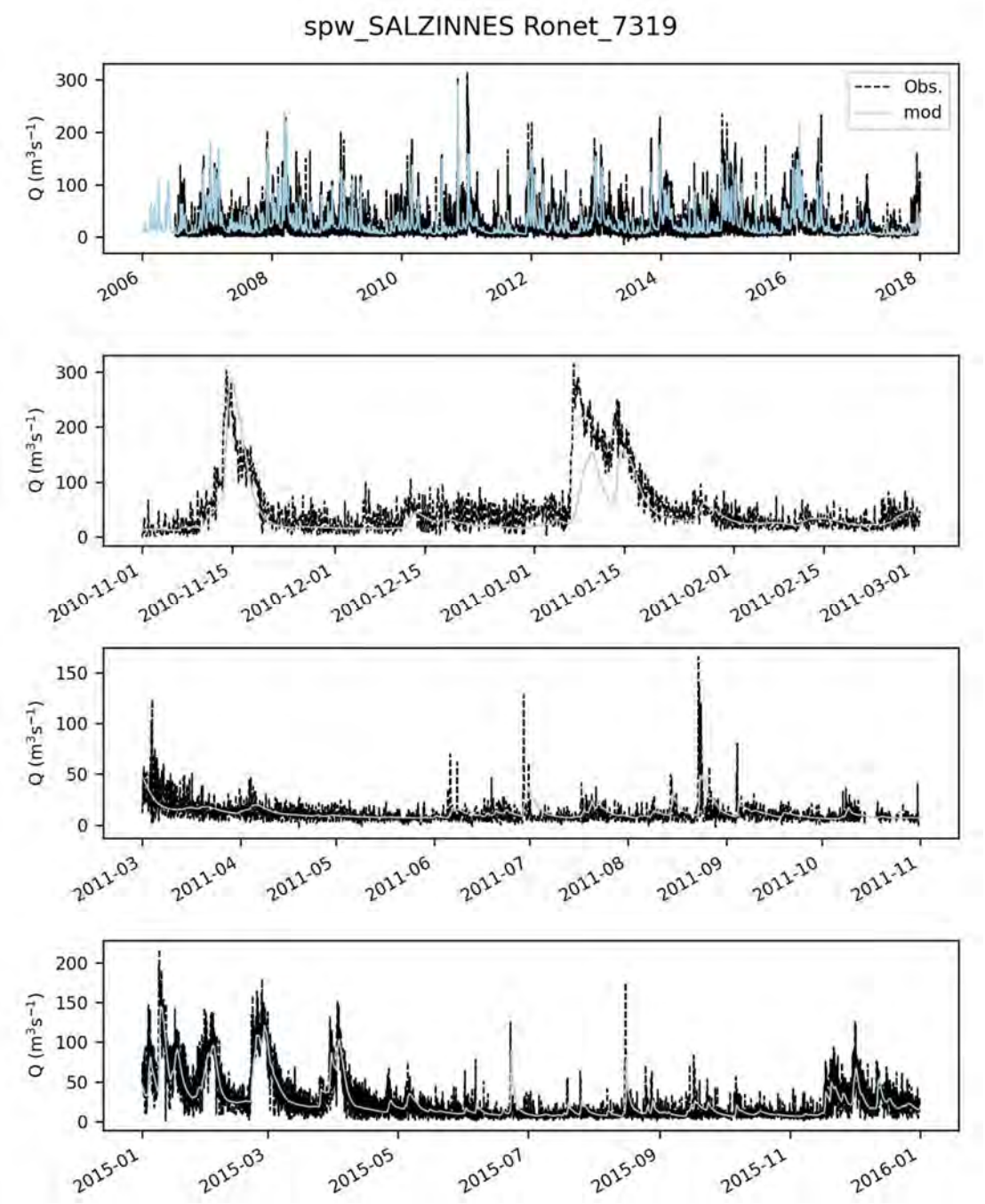


Figure 4-31 Modelled (blue) and observed (black) hydrographs for the Sambre at Salzinnes. The strong short-term variations in the observed series are caused by upstream weir operations.

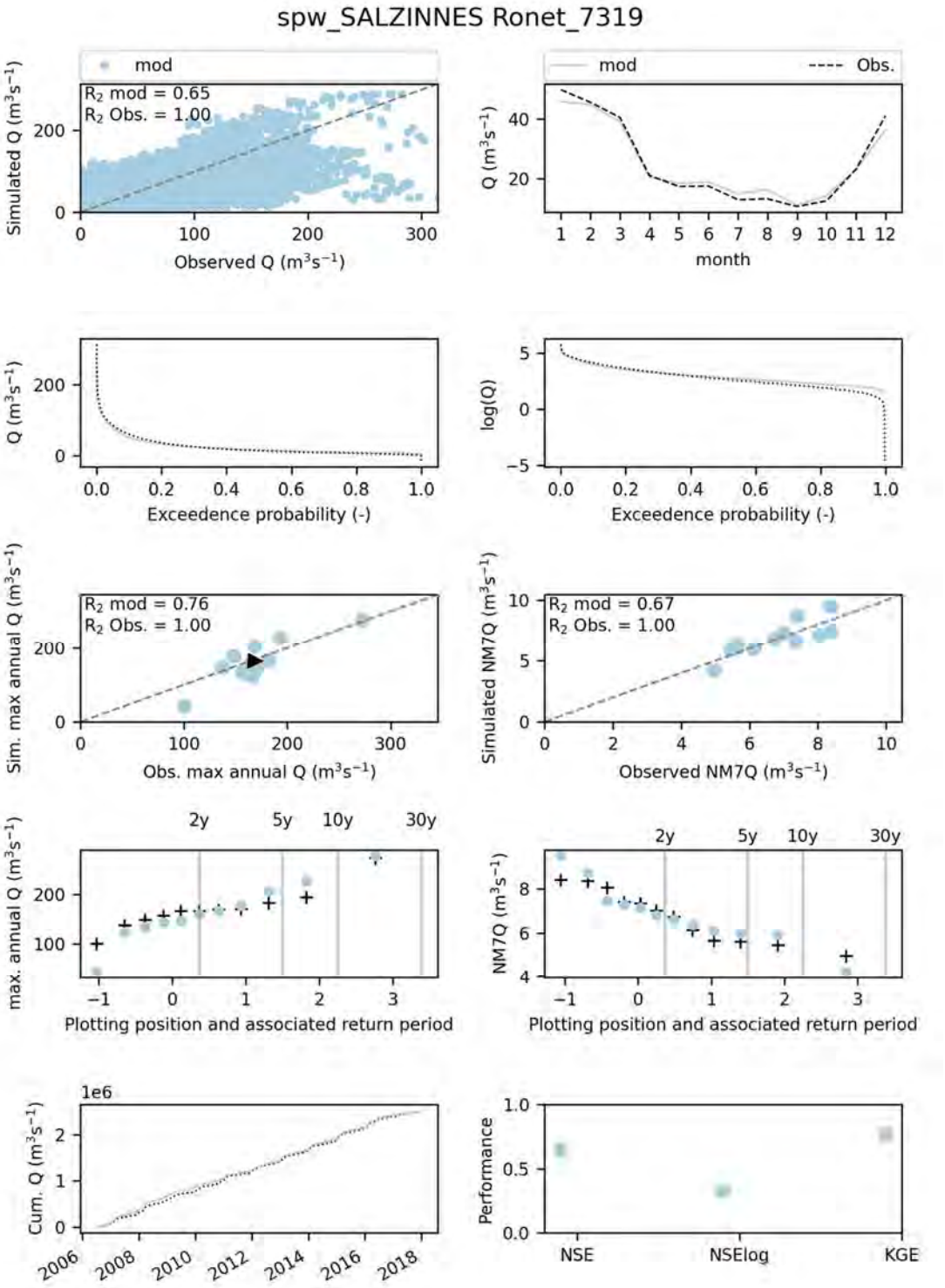


Figure 4-32 Modelled (blue) and observed (black) signatures for the Sambre at Salzinnes. For the Sambre at Salzinnes, we show the maximum annual flow using a rolling mean of 1 day to account for variations due to upstream weir operations.

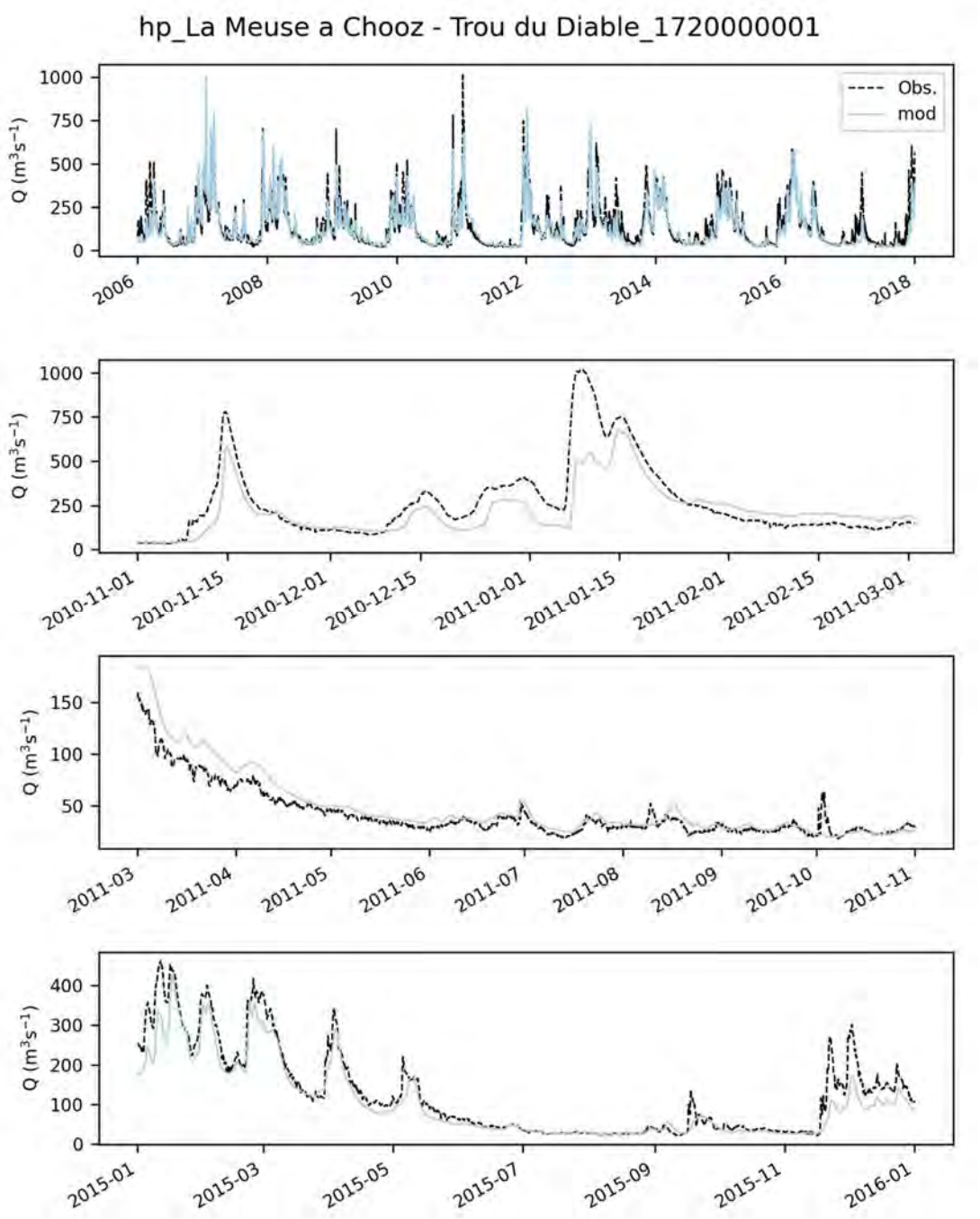


Figure 4-33 Modelled (blue) and observed (black) hydrographs for the Meuse at Chooz.

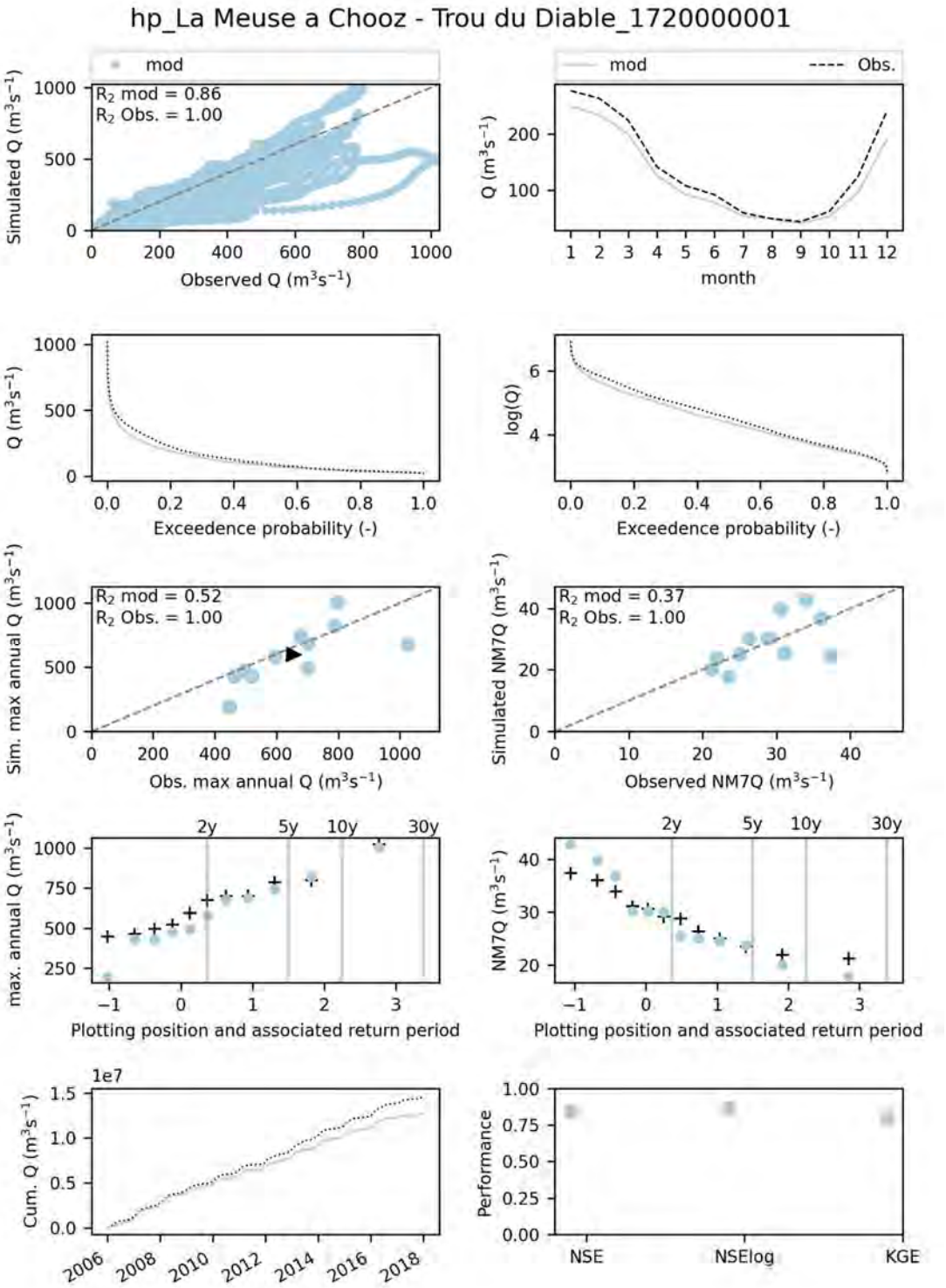


Figure 4-34 Modelled (blue) and observed (black) signatures for the Meuse at Chooz.

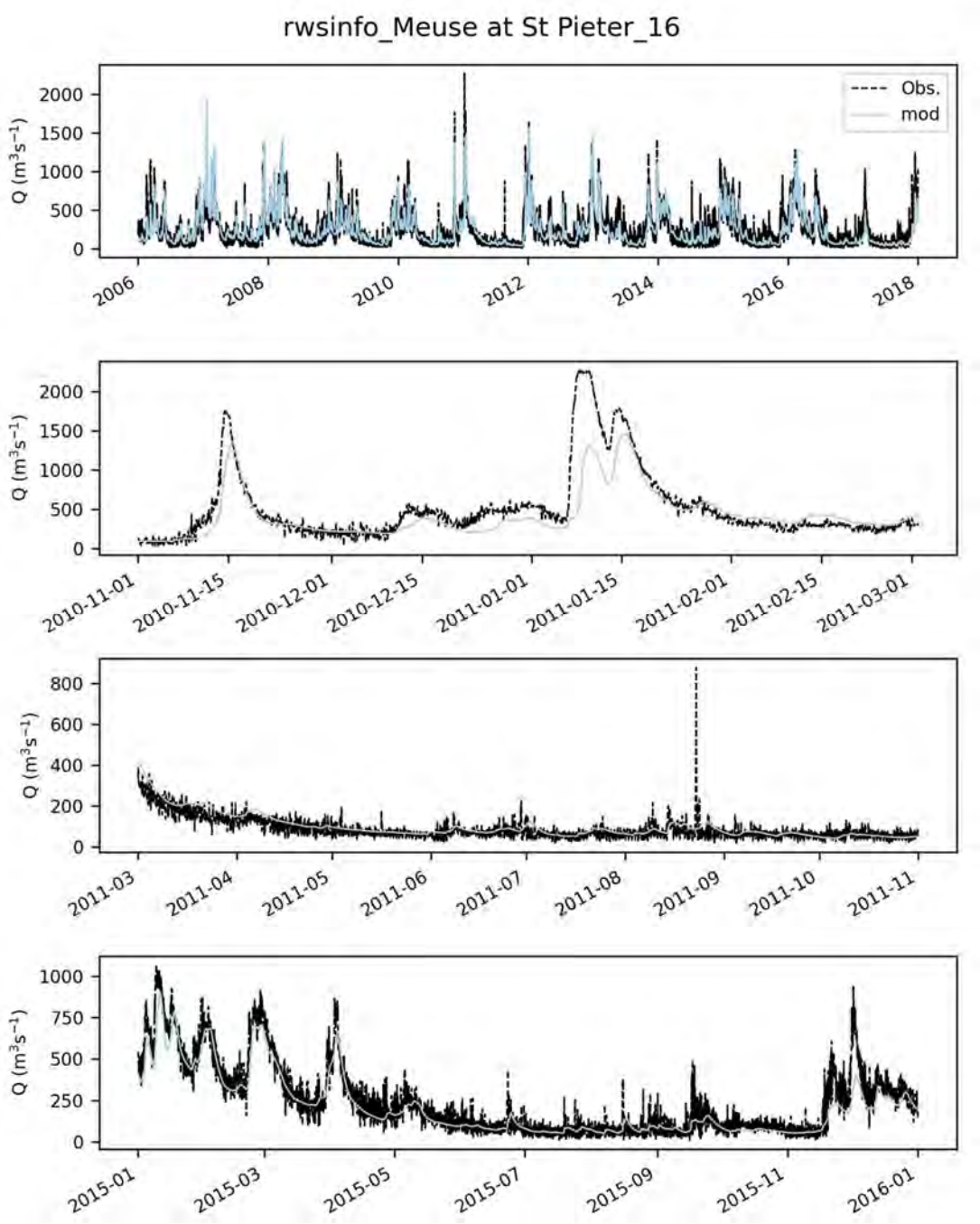


Figure 4-35 Modelled (blue) and observed (black) hydrographs for the Meuse at St Pieter. Also here, large short-duration fluctuations are caused by upstream weir operations.

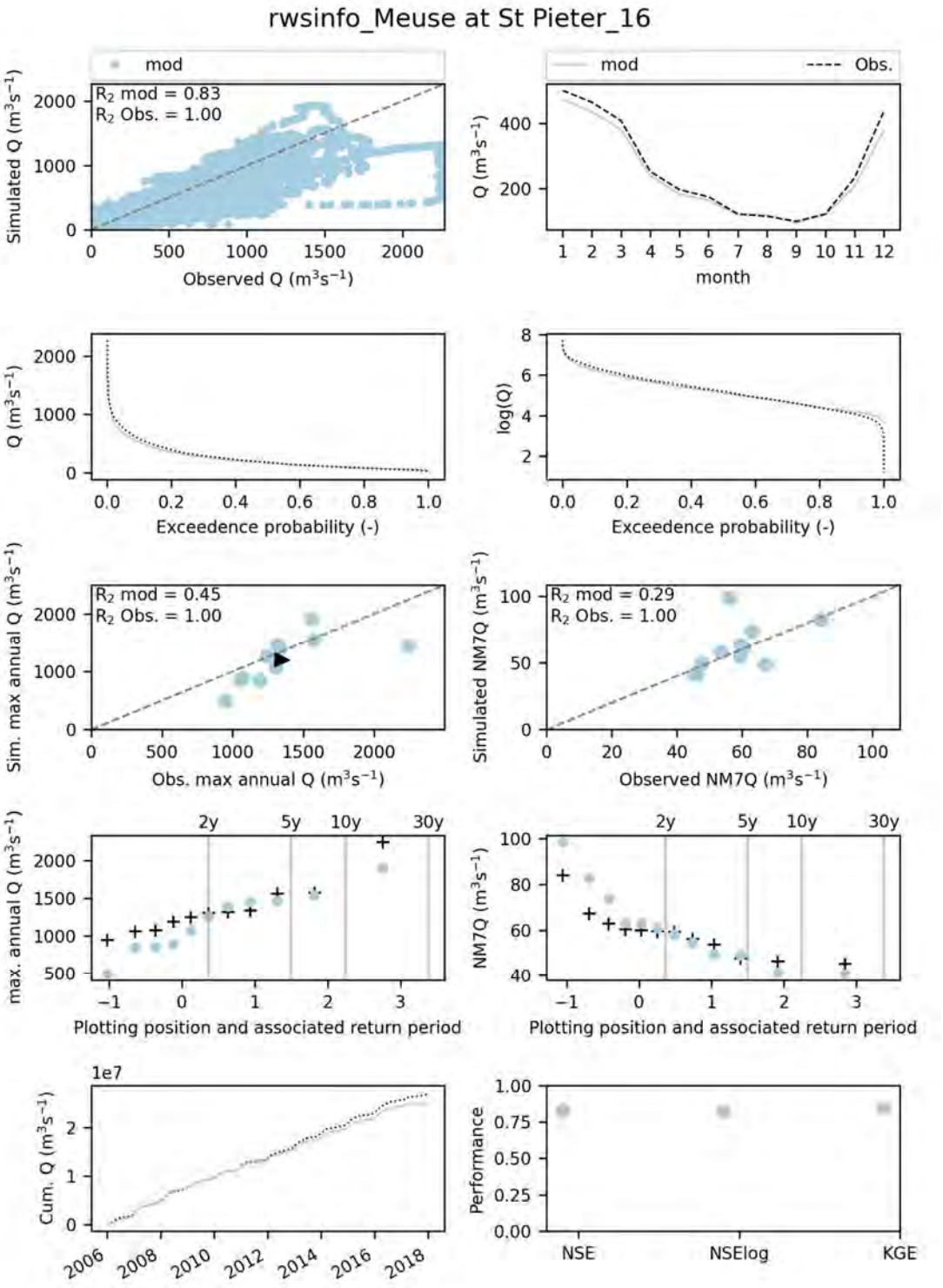


Figure 4-36 Modelled (blue) and observed (black) signatures for the Meuse at St Pieter. For the Meuse at St Pieter, we show the maximum annual flow using a rolling mean of 1 day to account for variations due to upstream weir operations.

4.3.3 Influence of timestep on model performance

This section shows the effect of the hourly timestep calibration on the daily timestep model performance for the catchment of the Ourthe at Tabreux. This was one of the main catchments we wanted to improve through the hourly timestep calibration as it was not performing well for both the highest and lowest flows. In addition, we show the results of the hourly and daily model performance before and after calibration for the Meuse at St Pieter as it is an important station for the Netherlands.

The model performance before and after calibration for the Ourthe at Tabreux for the hourly model is shown in Figure 4-37. Through model calibration, an improved performance for reproducing the minimum 7-days annual flows is achieved, while keeping the performance of the maximum annual flows relatively similar. High flows with return levels between approximately 2 and 5 years are still overestimated by the model.

When running the model calibrated at the hourly timestep with a daily timestep, we obtain the results shown in Figure 4-38 for the Ourthe at Tabreux. The improvements obtained after calibration in terms of low flow model performance also apply for the daily timestep model. The overestimation of the highest flows seen for the hourly timestep model are not clearly observed at the daily timestep. In contrast, the daily model seems to underestimate the streamflow with lowest (<2 years) and highest (>5 years) return levels and reproduces extremes with return levels between 2 and 5 years well. In the hourly model, the modelled plotting positions seem to be shifted upwards compared the daily model. Also, interesting to note is that the highest maximum annual flow is better simulated in the hourly model compared to the daily model. One of the reasons for this difference could be from the different datasets used to force the model at the daily and hourly timestep (E-OBS versus genRE).

It is interesting to note that although more computationally intensive, the calibration of the model at the hourly timestep results in improved or similar performance for both the hourly and daily timestep model. In the subsequent analyses, we therefore use the same parameter set for the model for both the hourly and daily analyses of extreme discharge frequencies but only vary the timestep setting from an hourly time step or a daily step.

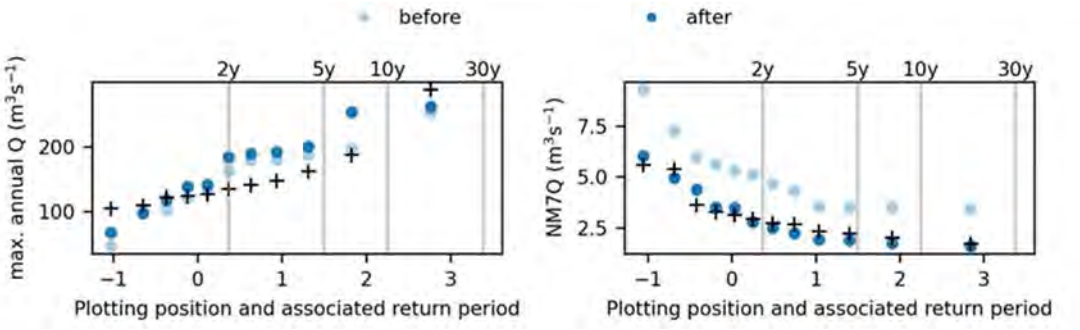


*Figure 4-37 Performance of the **hourly** model in reproducing annual maximum and minimum 7days flows (NM7Q) before and after the calibration (dark blue represents the model performance after calibration and light blue before calibration, while the black symbols show the observations) for the Ourthe at Tabreux for the period 2006-2017 (which corresponds to the genRE dataset available period).*

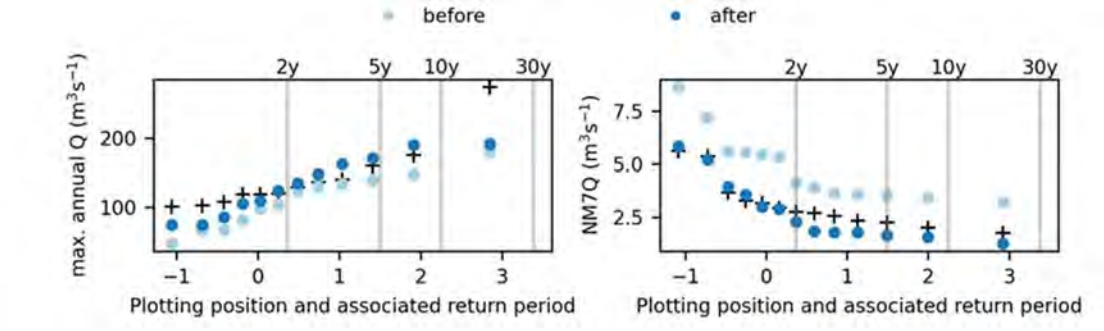


*Figure 4-38 Performance of the **daily** model in reproducing annual maximum and minimum 7days flows (NM7Q) before and after the calibration (dark blue represents the model performance after calibration and light blue before calibration, while the black symbols show the observations) for the Ourthe at Tabreux for the period 2006-2017 (which corresponds to the genRE dataset available period).*

In addition, we show these results also for the Meuse at St Pieter for the hourly time step model (Figure 4-39) and for the daily time step model (Figure 4-40) for the common period 2006-2017 (which corresponds to the genRE dataset period). For the Meuse at St Pieter, we see that the overall effect of calibrating all the internal nested catchments has resulted in a slight decrease of the maximum annual peak flows and this is seen for both the daily and the hourly model. Also, focusing on the highest maximum annual peak flow (left panels of Figure 4-39 and Figure 4-40), we see that the maximum modelled discharge at hourly timestep is closer to the observation compared to the daily timestep.

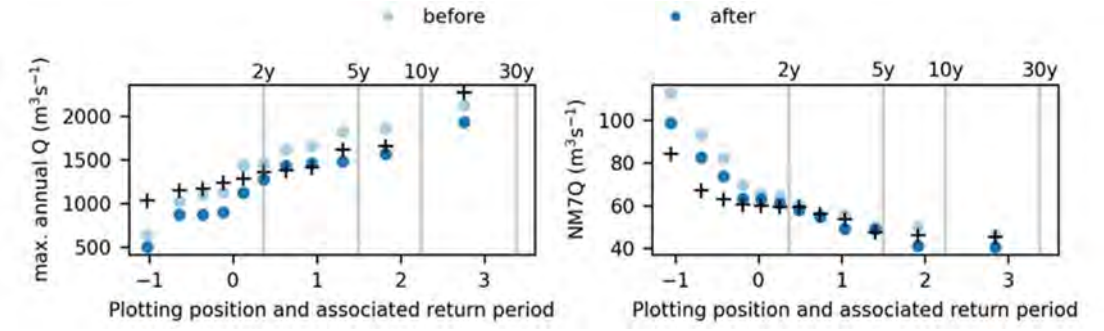


*Figure 4-39 Performance of the **hourly** model in reproducing annual maximum and minimum 7days flows (NM7Q) before and after the calibration (dark blue represents the model performance after calibration and light blue before calibration, while the black symbols show the observations) for the Meuse at St Pieter for the period 2006-2017 (which corresponds to the genRE dataset available period).*

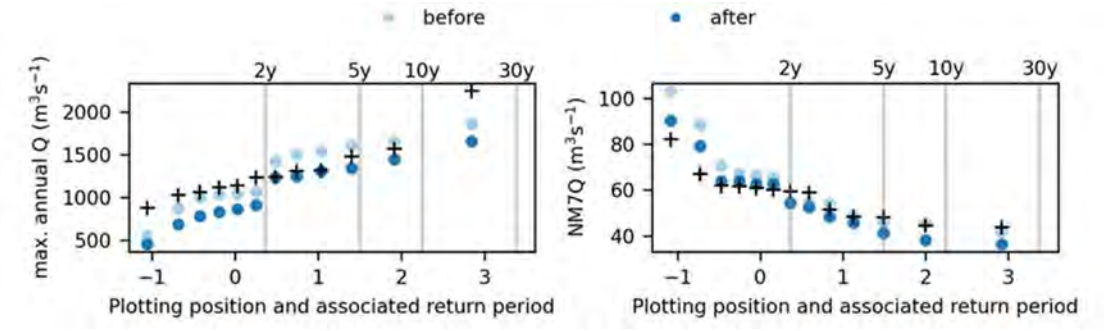


*Figure 4-40 Performance of the **daily** model in reproducing annual maximum and minimum 7days flows (NM7Q) before and after the calibration (dark blue represents the model performance after calibration and light blue before calibration, while the black symbols show the observations) for the Meuse at St Pieter for the period 2006-2017 (which corresponds to the genRE dataset available period).*

4.4 Implications and future work

The overall good performance of the model provides confidence to use the model for the next step of the project, which is the extreme discharge frequency analysis for the Meuse River and her tributaries. When analysing the results of the synthetic runs, it is helpful to consider how the model performs for the station of interest for the historic period to better understand the obtained results and place them into perspective.

Overall, the model has improved throughout the different steps: manual calibration, routing with the 1D floodplain schematization and automatic hourly calibration. However, the calibration may for specific stations result in improvements for specific aspects (e.g., mean monthly flow and cumulative flow) at the expense of other aspects of the hydrograph (e.g., maximum or minimum annual flows). Further improvements of model performance may, therefore, be obtained by further tuning parameters and/or improving process representation. One possibility is to further improve the reservoir schematization in the model using local data, a first analysis thereof is shown in Appendix 7C.

Concerning the automatic calibration of the hourly runs, due to the already high computational run times of the synthetic simulations using the RACMO dataset at the hourly time scale, we only selected one “most optimal” parameter set from the calibration analysis. Ideally, we would use an ensemble of parameter sets to better account for model parameter uncertainty, by selecting all parameter sets which perform above a defined threshold indicator. Alternatively, the GLUE analysis applied in Hegnauer et al. (2014) can be an interesting method to also consider parameter uncertainty. In addition, a different concept within the wflow framework could be used (such as the wflow_flextopo model which already exists for the Meuse basin) to also consider model structure uncertainty in the results of our extreme discharge frequency analysis. However, this would increase computational run times subsequently. In addition, it is relevant to also explicitly consider uncertainties in the observations, considering the type of measurement and the magnitude of the streamflow.

It is important to mention that the model was optimized for the basin area until St Pieter and for the Geul and the Rur. However, the tributaries between St Pieter and Mook (except for the Geul and the Rur) have not yet been optimized in wflow_sbm. This would be required for the model to be used for applications up to Mook.

Extreme discharges on the Meuse may be exacerbated through the coincidence of peaks from the tributaries. Understanding how the timing of the peaks from the tributaries during summer and winter conditions coincide with the streamflow on the Meuse requires further studying. This becomes increasingly important when assessing potential measures in the catchments that impact the hydrograph, such as increased storage capacity and wetland restoration.

Looking back at the recommendations made in Hegnauer et al. (2014), we have addressed two important aspects:

- Instead of using the lumped HBV model, we are now using a distributed and more physically based model. The distributed nature of the hydrological model allows us to extract spatial patterns in the extreme discharge frequency analysis, as shown in the next chapter.

Instead of using only the daily timestep, we demonstrate the use of the hourly timestep to better represent processes in the fast-responding tributaries of the Meuse basin.

5 Extreme discharge frequency analysis

5.1 Previous work

Previous extreme discharge return level estimates provided for St Pieter[4] in the Netherlands were based on a Gumbel distribution fit to 50,000 years of simulated discharge data from a hydrological model, driven by meteorological time series, produced by a stochastic weather generator, a methodology referred to as GRADE (Generator of Rainfall And Discharge Extremes, Hegnauer et al., 2014). The stochastic weather generator in turn uses historically observed time series. A complete description of the previous method and results is given in Hegnauer et al. (2014). Even though this approach allows to generate 50,000 years of climate and discharge time series, we identify the following limitations:

- Whereas the stochastic weather generator was a purely statistical model, in this study we use a physically-based climate model. The main advantage is that such an approach generates new plausible weather systems that were never observed in the past, which in turn can lead to new extreme discharge conditions. The stochastic weather data obtained from the regional climate model RACMO and referred to as the RACMO dataset is described earlier in section 3.2.3.
- Whereas the original GRADE approach relied on the conceptual semi-distributed hydrological model HBV (Lindström et al., 1997) at specific catchment outlets, discharge is now simulated with a physically-based spatially distributed model, i.e. the calibrated wflow_sbm model described in Chapter 4. This allows for extreme discharge estimates in the whole catchment. The framework of the wflow_sbm model also provides additional advantages compared to the HBV model concept. Spatial changes in soil moisture, total water storage, snow, evaporation, etc., can be explicitly included. Wflow_sbm model parameters are physically based and can be linked with our understanding of the climate and catchment characteristics, such as land use and climate (for e.g., the climate-derived rooting depth). This provides confidence when applying the model to other unforeseen conditions, as possible with climate change or synthetic simulations of the current climate. The simulation time, however, is longer than using the HBV model.
- The GRADE simulations were only performed at daily time scales because the weather generator was setup for a daily time step. However, a finer temporal resolution is particularly important in capturing fast-moving extreme discharge events, often characteristic of steep and small catchments. This is important, particularly for some of the Meuse tributaries, where the flood generating processes have a relatively short time scale.

Official extreme discharge return levels in the partner countries of the Meuse, such as in France, Belgium, and Germany, are based on different methods than GRADE, which is applied in the Netherlands, see also section 3.3.2. In general, observed discharge time series are used to extract discharge peaks used to either directly fit an extreme value distribution (EVD) or to statistically (i.e. not with a physically based model cascade) generate longer time series of discharge for the Netherlands to which either an EVD or empirical distribution is used to estimate given return periods.

[4] In Hegnauer et al. (2014), this location is referred to as Borgharen. See also section 3.3.1

5.2 Synthetic runs

An automated workflow is set-up using the Snakemake package (Mölder et al., 2021). The workflow is schematically shown in Figure 0-1 and further described here. First, time series of temperature, precipitation and potential evapotranspiration are read from the RACMO dataset, representing synthetic time series of the current climate, see also section 3.2.3. Second, the data is spatially resampled using a nearest neighbours' approach to downscale from a 12 km x 12 km rotated grid to the model resolution (approximately 1 km x 1 km). Unit conversion is also performed to the unit required for wflow. Third, this dataset is used as input for the hydrological model. The calibrated wflow_sbm model described in section 4.2 is run using this forcing dataset and simulated discharge is stored for the catchments of interest across the Meuse main river and its tributaries, as described in Chapter 2. For the simulated daily discharge, gridded outputs for the whole catchment are also stored for each time step as a NetCDF file. This was not feasible for the simulated hourly discharge due to the very large size of the file this would lead to. Hence, for the hourly time step, only time series at the selected catchments were stored. Runs were performed on a high performance computer using parallel computing with four threads to run the hydrological model. Running one ensemble member at the daily time step took approximately 8 hours and at the hourly time step approximately 20 hours. Finally, discharge peaks are extracted from the continuous time series to either fit an Extreme value distribution (EVD) or estimate return levels empirically.

5.3 Selection of extreme value models and theory

When a limited record length is available, extreme discharge return levels are calculated from an EVD fit to discharge peaks to statistically extrapolate return levels of discharges. The extreme value distribution models to apply are in turn dependent on the sampling method applied to extract peaks. Block maxima (BM) sampling extracts the highest value per constant block of time, for example, the highest value per year in the case of annual maxima (AM) or the highest value per month in the case of monthly maxima. Peaks over Threshold (PoT) sampling extracts the highest independent value above a certain threshold. Given an infinite size of samples, both methods should converge to similar distributions. Here, we focus on BM sampling, an easier method to analyse at large scale since selecting a proper threshold that leads to reliable distribution parameters across catchments of various sizes, is particularly computationally intensive.

When the sampled discharge peaks can be assumed to be independent and identically distributed (i.i.d.) and the size of the samples tends to infinity, the Fisher-Tippett Gnedenko theorem states that BM samples will converge towards the Generalized Extreme Value (GEV) distribution, described with the following cumulative distribution function $F(x)$ of extreme discharges, X:

$$F(x) = exp\left(-\left(1 - \gamma\left(\frac{x-\alpha}{\beta}\right)\right)^{1/\gamma}\right)$$

with $1 + {\gamma(\alpha - x)}/{\beta} \geq 0$ and $\gamma \neq 0$

where α, β and γ are the location, scale, and shape parameters, respectively. Here, α, γ can be any values but $\beta > 0$.

The location parameter α represents the center of the extreme value distribution. The scale parameter β corresponds to the spread of the extreme value distribution and the shape parameter γ the skewness of the distribution (or tail curvature). The shape parameter γ indicates the behaviour of the tail of the distribution and has a strong influence on the return levels of large return periods. This is illustrated in Figure 5-1. The value of the shape parameter γ determines the convergence to three possible sub-families of EVD.

If the shape parameter is positive, i.e. $\gamma > 0$, a type III (reversed Weibull) with a bounded tail applies; for a negative shape parameter, i.e. $\gamma < 0$, a type II (Fréchet) with a heavy tail applies and, if the shape parameter is zero, i.e. $\gamma = 0$, a type I (Gumbel) with a light tail applies. Note that there is no convention for the sign of the shape parameter. We follow the convention as applied in the statistical packages used and in accordance with the report from van Voorst and van den Brink (2023).

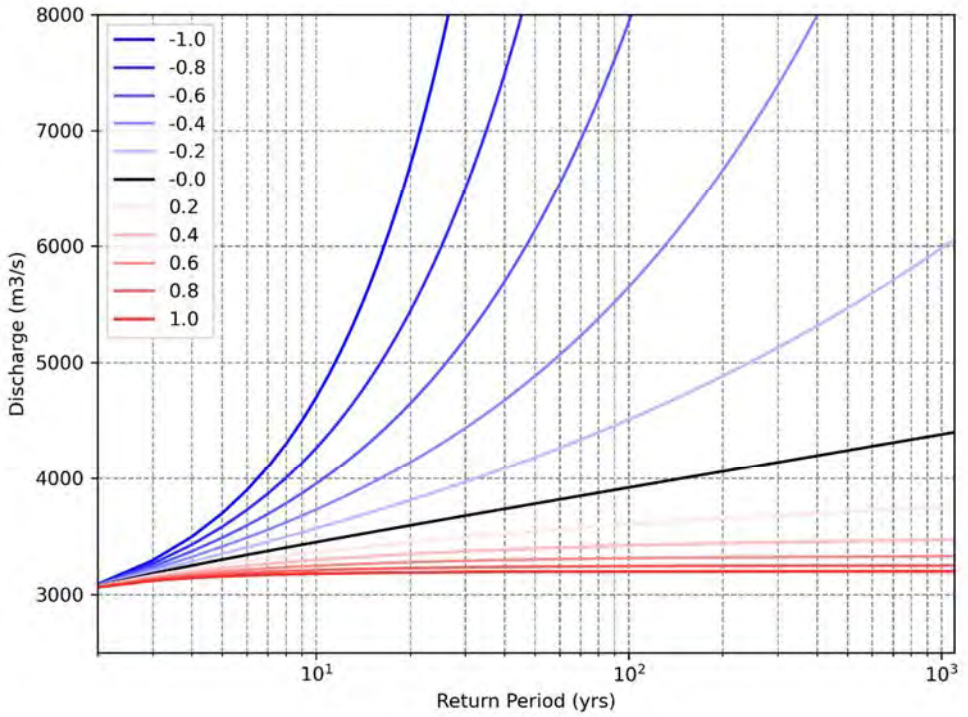


Figure 5-1: Exceedance probabilities obtained from fitting a Generalized Extreme Value (GEV) distribution with a location parameter of 3000, a scale parameter of 200, and varying shape parameters.

Estimation of the EVD parameters is done using the maximum likelihood method. Block maxima are extracted based on hydrological years, defined from October 1 to September 30. The summer season is defined as going from April until September included and the winter season from October to March included.

Record lengths used to fit the EVD have an impact of the uncertainties around large return period estimates. Frequency estimates are subject to noise due to natural variability. Hannart (2019) assesses that in order to estimate the 100-year return period with a 10% error requires 10,000 years of data (assuming the exact extreme value distribution is known). Bootstrapping methods are often applied to assess the confidence interval around return values from the impact of sampling uncertainty. A common rule of thumb is that reasonable results are obtained for return periods up to one half, one-third (Früh et al., 2010; Ludwig et al., 2023) or one-quarter (Pugh & Woodworth, 2014) of the length of the time series. Applying this rule here to the 1,040 year long dataset would mean that return levels are particularly uncertain for return periods higher than approximately 520 years, 340 years or 260 years, respectively.

Towards larger return periods, statistical uncertainty grows because the fit of the parameters relies on only a few extremes. The shape parameter, defining the tail of the distribution, is particularly sensitive to these extremes and found to converge slowly.

An advantage of the current approach is that with this record length, up to 1,024 hydrological years can be used for the extreme value analysis. This is a much longer series than typically offered by observed discharge records, which are often only a few decades or at most a century in length. Observed discharge data may also be impacted by trends in the catchment that in turn can directly impact parameter estimates from the EV fit. Decade long record lengths are usually too short to ensure convergence of the shape parameter, crucial to characterize the extreme value behaviour (Papalexiou & Koutsoyiannis, 2013).

When long time series are available, empirical return periods can be calculated directly from the analysis of the ranks of the peaks, removing the need to fit an extreme value distribution. Similar to fitting an EVD, return level estimates become uncertain around large return periods due to natural variability. The exceedance probability is estimated from the probability plotting position, as follow:

$$P = \frac{n + a}{\lambda\,(m + b)}$$

where m is the total number of peaks; n is the rank is ascending order of the peak from $n = 1$ for the largest peak to $n = m$ for the smallest; a and b are constants; λ is the expected number of peaks per year.

Among the many possibilities put forward in the literature for the values of a and b, we use here $a = -0.3$ and $b = 0.4$ (Bernard & Bos-Levenbach, 1955), to be in line with the GRADE method applied in the Netherlands (Hegnauer et al., 2014) and with the report from van Voorst and van den Brink (2023) analysing the synthetic extreme precipitation time series obtained from the RACMO model. For annual maxima, the expected number of peaks per year is by definition 1 (i.e., $\lambda = 1$).

Unless mentioned otherwise, for all the results presented in this section, a GEV extreme value distribution is applied to calculate the return periods of the synthetic extreme discharges modelled. The empirical return periods of the synthetic discharge annual maxima are shown with the plotting position mentioned above.

5.4 Extreme discharge return levels

Block maxima discharge peaks are extracted from both the daily and hourly runs from the 16 ensemble members forced with the bias corrected RACMO climate dataset, representing synthetic climate time series of 65 years of current climate characteristics from 1950 until 2014. Note that, because hydrological years are used instead of calendar years, this results in 64 valid years per members and thus a total of 64 * 16 = 1,024 years.

Detailed results are shown for the following catchments of interest, as identified during the workshop with the different stakeholders, see also section 2:

- Rur at Monschau (Germany)
- Geul at Meerssen (Netherlands)
- Meuse at Goncourt (France)
- Vesdre at Chaudfontaine (Belgium)
- Ourthe at Tabreux (Belgium)
- Sambre at Salzinne (Belgium)
- Meuse at Chooz (France)
- Meuse at St Pieter (Netherlands)

Other points of interest identified (the Geul at Hommerich, the Viroin at Treignes, the Ambleve at Martinrive, the Semois at Membre Pont, the Lesse at Gendron, the Rur at Stah, and the Meuse at St-Mihiel) are shown in Appendix D.

5.4.1 Influence of record length

Figure 5-2 compares the GEV fit applied to each member (64 hydrological years) versus when all ensembles are combined (1,024 hydrological years) for the model ran at the daily time step. Each black line corresponds to the GEV fit from one member whereas the red line corresponds to the fit to all ensembles combined. The black (red) shading around each curve represents the 95% confidence interval bounds for that specific fit and is obtained from 100 bootstrap samples. For each location, we observe a large spread in the tail behaviour, indicating that return level estimates from separate members can bring large differences, especially for large return periods (100-year return period and higher). This confirms the importance of long record lengths as stated above. Combining all ensemble members provides a coherent estimate, shown in red, within the spread of the 16 members but with a much smaller statistical uncertainty than when considering ensemble members separately, shown in grey. The daily simulated results are also in line with the observed discharge extremes at most locations. For locations with a short record length, such as the Rur at Monschau, the Meuse at Chooz or the Sambre at Salzinne, observed return periods deviate from estimates obtained when considering 1,024 hydrological years (shown with a red line). This is very likely due to the limited observation record length available at this location meaning that these empirical return periods are particularly uncertain. Interestingly, in almost all locations, we can find one fit from a member (black line) that follows the observed empirical return periods. This highlights that our results are in line with observations when considering natural variability and the short record length considered (65 years or less). The impact of the natural variability is already visible for return level estimates of the 10-year return period and strongly increases for larger return periods due to the impact of the different tail behaviour obtained from the fit of each member. The only exception is for Viroin at Treignes, shown in Figure D-1, when the model consistently underestimates discharge extremes. This is a result of the model performance at this location, which was found to underestimate high flows (see Table B-1). Generally, the simulated results are in line with return level statistics reported for the Meuse at Chooz, the Meuse at Goncourt, or the Meuse at St Pieter. At some locations, some of the largest discharges are observed in summer, for example at the Geul at Meerssen, the Rur at Monschau or the Vesdre at Chaudfontaine. This is further discussed in section 5.4.2.

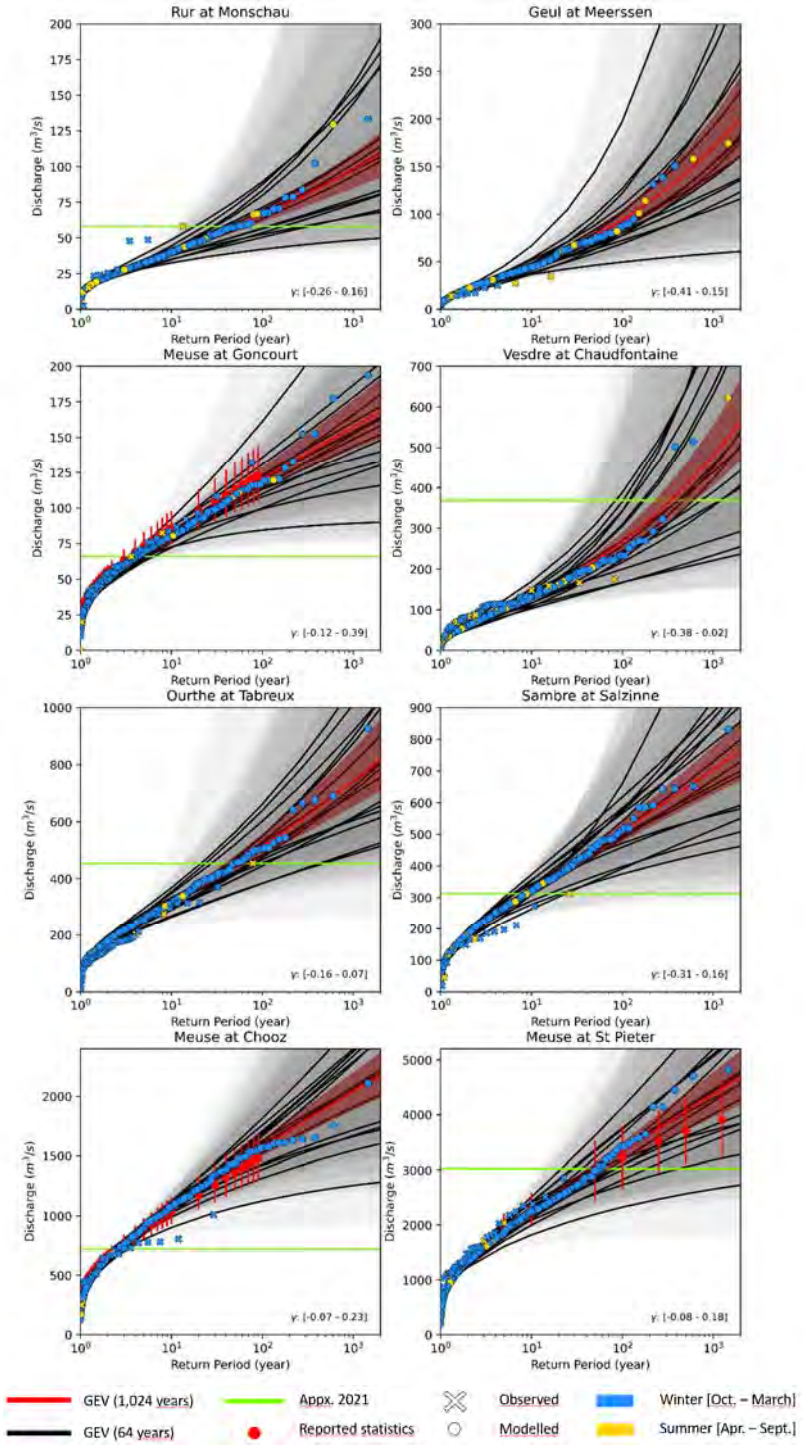


Figure 5-2: Extreme value distribution fit at selected locations across the Meuse catchment at a daily time step. Black lines represent the GEV fit obtained from one ensemble member (64 hydrological years) and the range of shape parameters is written at the bottom right of each figure. Red lines show the GEV fit obtained from combining all ensembles (1,024 annual maxima from hydrological year, shown as full dots). The shading represents the 95% confidence interval from the distribution parameters uncertainty obtained by bootstrapping from 100 bootstrap samples, shown here with some transparency level. Observed annual maxima of discharge peaks when present are shown with crosses. A blue (yellow) color indicates that the peak occurred in winter (summer). Recorded or estimated maximum discharge peak reached in July 2021 are shown in green. Note that these estimates may deviate from the real discharge peak observed during the event, see also section 3.3.1. Reported return levels (when present) with uncertainty are shown with red circles and vertical bars, respectively, see also section 3.3.2.

A similar approach is applied to each grid cell from the Meuse catchment for the daily runs, where one GEV fit is done per ensemble member and another fit on the whole combined time series of synthetic discharge. Figure 5-3a shows the minimum shape parameter obtained across the 16 ensemble members for the daily runs (saved as gridded data) and Figure 5-3b the maximum shape parameter value obtained. Opposite signs of shape parameter are found across the members, confirming that 64 hydrological years is not sufficient to robustly characterise the tail of the extreme value distribution. It is expected that the extreme discharge behaviour should remain similar across the ensemble members. From Figure 5-1, it is shown that the sign of the shape parameter characterizes different extreme discharge tail behaviour. The variability of the sign of the shape parameter changes across the Meuse catchment, as shown in Figure 5-3c. In locations with a high variability, this means that return levels estimated for large return periods, such as the 100-year return period could vary a lot. Along the Meuse, the shape parameter is often positive (bounded tail) while across small tributaries, the shape parameter tends to often be negative (heavy tail) across the ensemble members.

Using all ensemble members, we obtain the shape parameters shown in Figure 5-3d. The Meuse river stands out, with a positive shape parameter along its whole course. Shape parameters close to 0 are found for the Sambre river, in both cases indicative of a light bounded tail behaviour. Using all ensemble members reduces the variability of the estimate of the shape parameters with most values being constrained between 0.3 and -0.4. Heavy tails, large negative values, can be found in small tributaries, highlighting the potential for very large discharge extremes locally. For example, river steepness can also significantly impact extremes. Figure 5-4 shows the obtained shape parameter alongside with the river slope and overall elevation from the hydrological model. This could, to some extent, explain why clear spatial patterns are observed in Figure 5-3d. In the Ardennes, where steep slopes are present, the shape parameter is negative, indicative of rapidly increasing discharge extremes for large return periods.

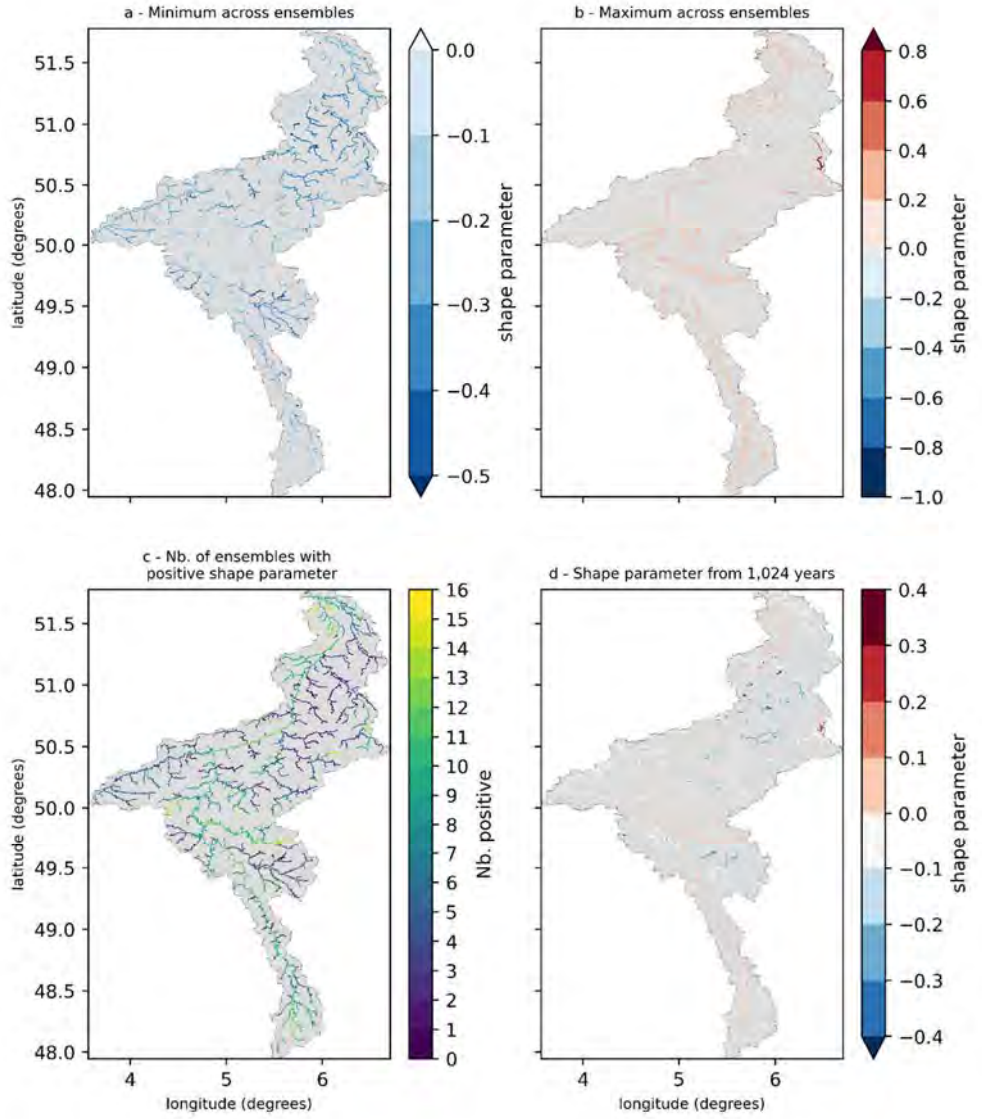


Figure 5-3: a) Minimum GEV shape parameter obtained from the 16 ensemble members for the daily runs. b) Maximum GEV shape parameter obtained from the 16 ensemble members for the daily runs. c) Number of ensembles with a positive GEV shape parameter. d) GEV shape parameter obtained by combining records from all ensembles, i.e. 1,024 hydrological years. A positive shape parameter indicates a bounded tail, a negative value a heavy tail and a value close to 0 a light tail, see also Figure 5-1.

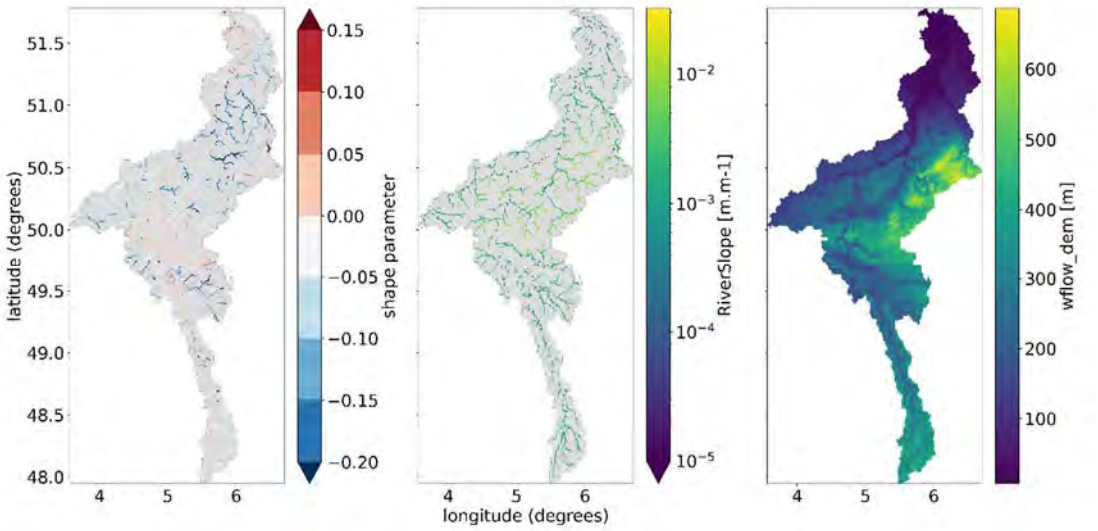


Figure 5-4: left panel: GEV shape parameter obtained by combining records from all ensembles, i.e. 1,024 hydrological years. A positive shape parameter indicates a bounded tail, a negative value a heavy tail and a value close to 0 a light tail (same as Figure 5-3d but with a different color scale). Middel panel: River slope from the hydrological model. Right panel: Digital elevation model (DEM) from the hydrological model

The link between the shape parameter and the catchment size and steepness is also clear from Figure 5-5 which compares a Gumbel fit (i.e. by definition with a shape parameter $\gamma = 0$) with a GEV fit (calculated shape parameter $\gamma < 0$ indicating upward tail curvature for small catchments) for the selected catchments. Small differences between return levels can be observed along large rivers (e.g., the Sambre at Salzinne, the Meuse at St Pieter and the Meuse at Chooz) indicating that the shape parameter is very close to 0. This is not the case for other smaller rivers where much heavier tails than Gumbel are observed such at the Vesdre at Chaudfontaine, the Geul at Meersen or the Rur at Stah. These results can be generally linked with subcatchment characteristics with the size of the catchment being one but not the sole characteristic influencing extreme discharges.

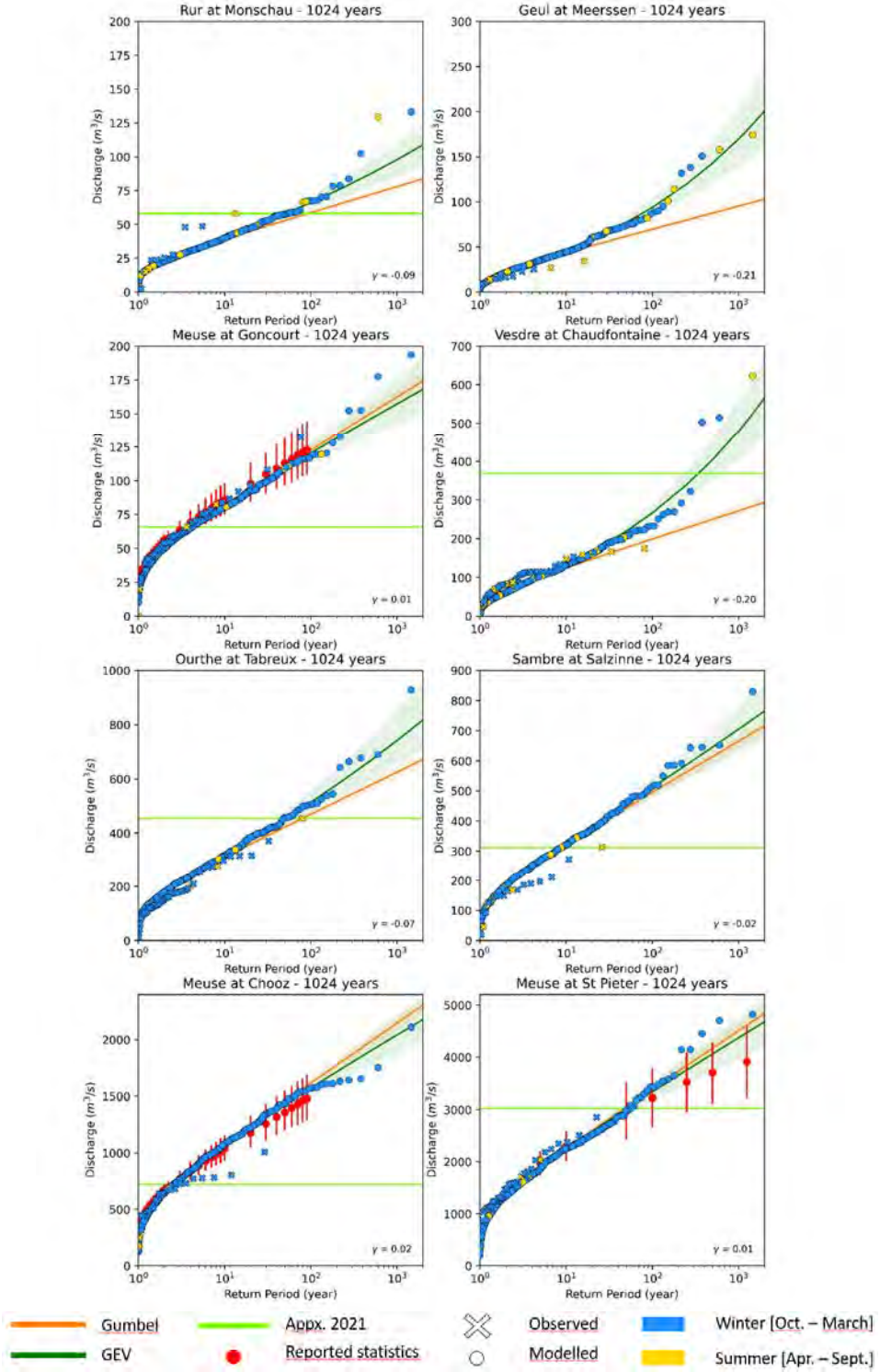


Figure 5-5: Extreme value distribution fit at selected locations across the Meuse catchment at a daily time step. Dark green lines represent the GEV fit obtained from combining all ensembles (1,024 hydrological years, shown as full dots) and blue lines the Gumbel fit. The shading represents the 95% confidence interval from the distribution parameters uncertainty obtained by bootstrapping from 100 bootstrap samples, shown here with some transparency level. Observed annual maxima of discharge peaks when present are shown with crosses. A blue (yellow) color indicates that the peak occurred in winter (summer). Recorded or estimated maximum discharge peak reached in July 2021 are shown in green. Note that these estimates may deviate from the real discharge peak observed during the event, see also section 3.3.1. Reported return levels (when present) with uncertainty are shown with red circles and vertical bars, respectively, see also section 3.3.2.

5.4.2 Influence of flood seasonality

Figure 5-2 and Figure 5-5 show that some extreme discharge events occur in summer for some locations, for example at the Geul at Meerssen, the Rur at Stah and at the Vesdre at Chaudfontaine. In these locations, these summer events represent some of the highest extreme daily discharge simulated. If summer events result from a different statistical population than winter events, this violates the assumption of identically distributed events made when fitting the extreme value distribution. In this subsection, we explore the potential importance of summer events and their impacts on estimated return levels.

Figure 5-6 shows the GEV shape parameter obtained when extracting discharge peaks irrespective of its occurrence during the hydrological year (Figure 5-6a) or considering only seasonal maxima (either summer or winter, Figure 5-6b and Figure 5-6c respectively). Large differences in shape parameters are observed when considering only summer maxima or winter maxima. Sampling summer maxima results in mainly negative shape parameters across the whole catchment, indicative of a heavy tail distribution. This behaviour is in line with the fact that some summer events can be particularly extreme and therefore deviate from typical high summer discharge events. Their impact on extreme discharge return levels is however dependent on the magnitude of those summer extremes compared to the winter extremes, as further detailed in this section. In general, considering only winter maxima results in a similar result as when considering the whole hydrological year. This indicates that extreme value distribution is dominated by winter maxima. However, in locations with the potential for large summer extremes, this results in a more negative shape parameter, indicating a heavier tail and a potential for extreme return levels when considering large return periods.

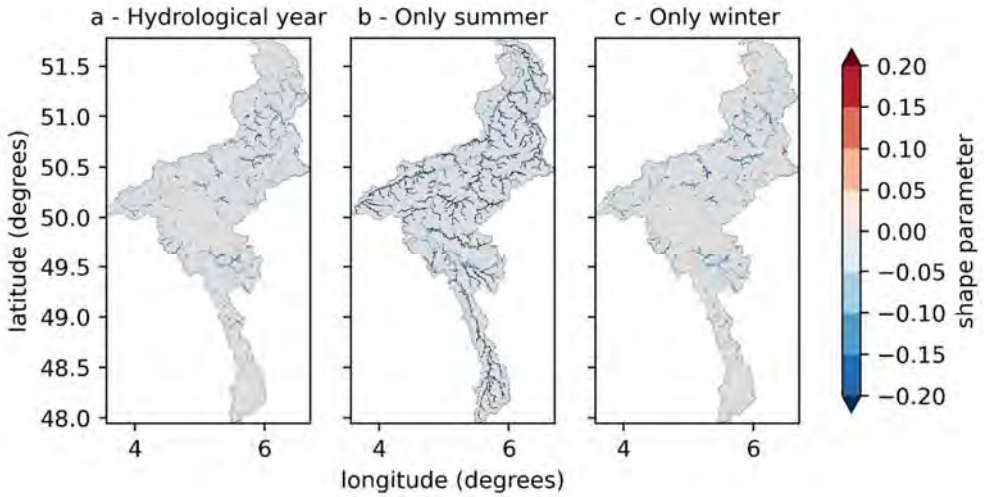
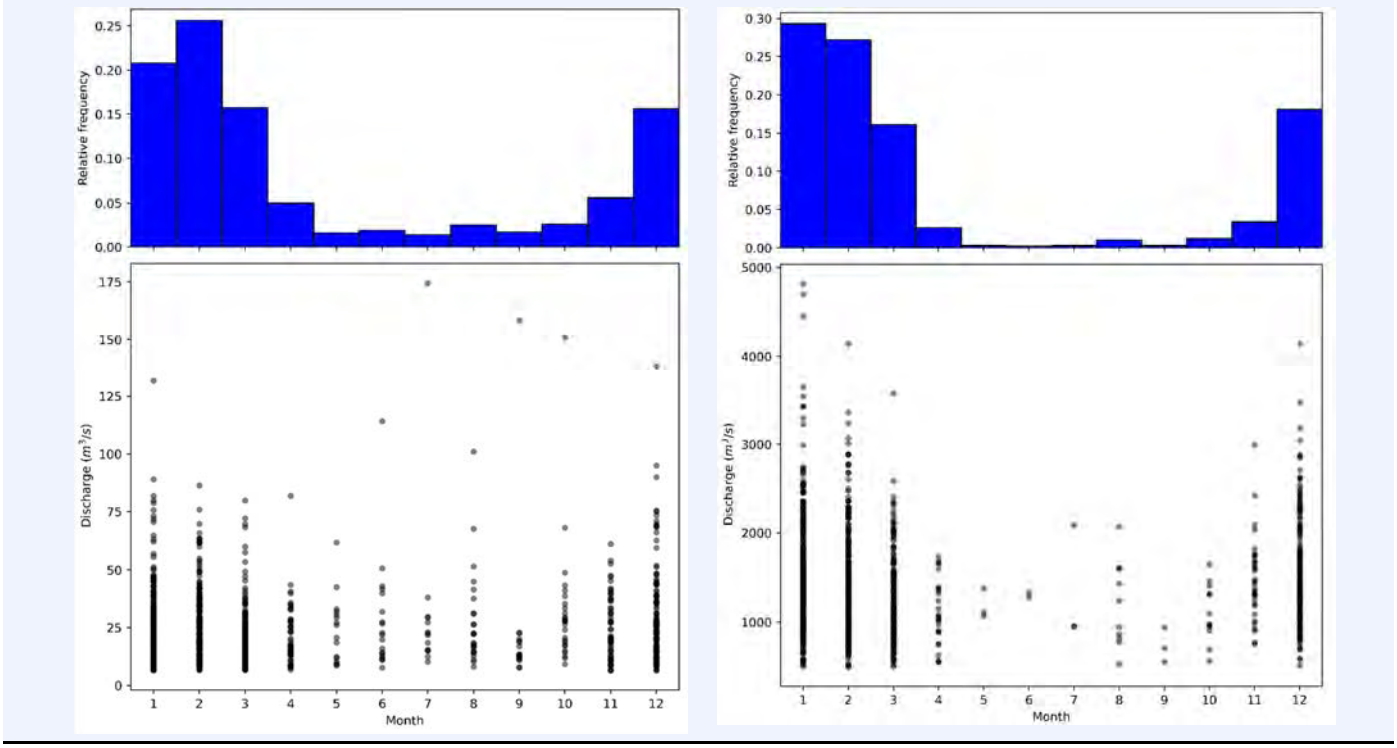


Figure 5-6: GEV shape parameter obtained from sampling discharge maxima from (a) 1,024 hydrological years, (b) summers only and (c) winters only. Summer is defined from April to September and winter from October to March.

The dominance of winter events in discharge maxima also appears clearly in the analysis of the month of occurrence of the hydrological annual maxima. Figure 5-7 shows a histogram of the relative frequency of the occurrence per month from the largest 1,000 annual maxima (i.e. almost all annual maxima) as well as the values of annual maxima for the Geul at Meerssen (left panel) and the Meuse at St Pieter (right panel). While the general shape of the distribution at the two locations is relatively similar when considering 1,000 highest maxima, the magnitude of the summer extremes when compared to the winter extremes can strongly differ. For the Geul at Meerssen, five out of the ten largest discharge events are in summer but this is not the case for the Meuse at St Pieter. This points to the fact that, for some locations such as the Geul at Meerssen, summer events can drive the tail of the extreme value distribution.



a) Geul at Meerssen

b) Meuse at St Pieter

Figure 5-7: Relative frequency of the month of occurrence and corresponding magnitude of the top 1,000 annual maxima of daily discharge for the Geul at Meerssen and the Meuse at St Pieter .

The summer, especially in the Ardennes, brings localized heavy rainfall and thunderstorms. Locally, this can lead to extremes, but the effect of a local extreme smooths out on the scale of the Meuse because other locations usually receive less rainfall. Winters however have predominantly stratiform rainfall, which results in rainfall being generally more evenly distributed over the Meuse catchment, although orographic enhancement still results in the Ardennes receiving more rainfall. Antecedent conditions such as soil moisture conditions will also impact discharges. Low evaporation in winter contributes to increased discharge magnitude. Moderate but frequent rainfall events can contribute to a wet period and therefore increase soil wetness.

Van Voorst and van den Brink (2023) analysed the seasonal patterns from the RACMO rainfall dataset, averaged for different catchment sizes from 680 km^2 to 21,300 km^2 for the Meuse at St Pieter. They found that, when considering an hourly accumulation period, summer rainfall extremes are always larger than winter rainfall extremes. When considering a daily accumulation period, summer rainfall extremes will often surpass winter rainfall extremes for large return periods (100-year return period and higher). When considering a 10-day accumulation period, this inflection point is moved to the 1000-year return period or higher. Such conclusions cannot be directly done for discharge extremes because catchment characteristics play an important role in generating discharge extremes and rainfall-runoff processes are not linear. Knowing the rainfall-runoff response time of the catchment provides a first indication of the rainfall events likely to cause extreme discharges. Very roughly, for catchment with a response time of one day, from the above, we can estimate that there are about 10 rainfall summer events of similar or stronger magnitude than winter events in the RACMO dataset. However, the spatial distribution of the rainfall over the catchment will influence the timing and the magnitude of the discharge event. For the same amount of rainfall and total duration of the event, a spatially uniform rainfall event over an elongated catchment will lead to a lower discharge extreme than when localized near the catchment outlet.

For example, the RACMO rainfall dataset contains 48-hour summer rainfall events more extreme than the July 2021 for the Meuse at St Pieter. However, they did not result in more extreme discharges than the July 2021 estimate because most rainfall was located more upstream in the catchment, and not in the fast responding Ardennes as observed in July 2021. Therefore, while the RACMO dataset can generate July 2021 event-like, the total synthetic record length of 1,040 years of the RACMO dataset is not long enough to assess its impact on summer extremes with confidence for large catchments such at the Meuse at St Pieter. To investigate this thoroughly, a much longer dataset should be generated. Alternatively, this could also be done by specifically focusing on generating plausible summer rainfall events.

5.4.2.1 Impact on extreme return levels

We highlight for one location of interest, the Rur (Roer) at Stah, how seasonality can influence return levels. Figure 5-8 shows return levels obtained from different calculation methods. In one case, shown in black, return levels are obtained from annual maxima of the hydrological years: in that case, the peak could have occurred in winter or in summer. In the other case, summer and winter maxima are extracted, and a GEV distribution is fit to each set of points, shown in red and blue respectively. The combined yearly exceedance frequency, referred to as combined GEV and shown as a purple dashed line in the figure, is then calculated as the sum of the respective exceedance frequencies (Dullaart et al., 2021; Palutikof et al., 1999) from the summer and the winter fit, as follows:

$$T(x) = \frac{1}{\frac{1}{T_s(x)} + \frac{1}{T_w(x)}}$$

where T is the return period in years of a given discharge level x. T_s and T_w refer to the respective return periods of the summer discharge extremes and winter discharge extremes.

Differences in return levels between the two calculation methods appear around the 100-year return period and increase as the return period increases. This is because around this threshold, the magnitude of the summer extreme is almost similar to the winter extreme magnitude. In some cases, as observed for the Geul at Meerssen, summer extreme magnitudes can even surpass winter extremes.

Note that this approach is not restricted to a GEV fit and other extreme value distributions could be used to combine summer and winter extremes.

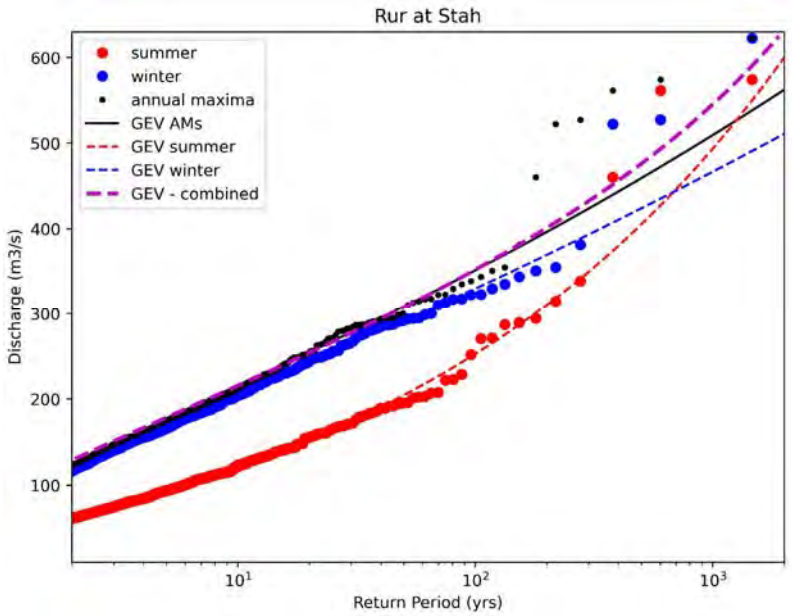


Figure 5-8: Combined GEV analysis for the Rur at Stah. Fitting a GEV model irrespective of summer and winter events leads to the black line (GEV AMs). Differentiating between winter and summer events as different populations leads to the blue and red line, respectively. These exceedance frequencies can then be recombined to lead to a new yearly frequency estimate of discharge extremes, shown in purple. Note that other extreme value distributions than the GEV distribution can be used.

5.4.3 Influence of timestep

So far, the results shown are extracted from the model runs performed at the daily resolution. For small catchments, it is hypothesized that subdaily time scales are of importance to better represent discharge extremes. Figure 5-9 and Figure 5-10 are similar to Figure 5-2 and Figure 5-5 but for the simulated discharge at an hourly timestep instead of a daily timestep. Similar findings as for the daily timestep can be made by visually comparing them. First, as for the daily timestep, combining ensemble members provide great value to reduce the statistical uncertainty of the GEV fit. Most simulated discharges are in line with the empirical frequency analysis derived from observations, especially considering the particularly short record length of observations. A noticeable difference is for the Meuse at Goncourt. While at the daily timestep, results were in line both with observations and officially reported statistics, this is not the case for the simulated hourly discharge. Hourly discharge return levels are consistently underestimated and not in line with the officially report statistics. This result is consistent with the reported model performance at this location for the hourly time step. Observed hourly extreme discharges and statistics are also lower for the Geul at Meerssen than the modelled extreme discharges. The best approach to correct for these differences would be to improve the model performance for these locations. However, if this is not possible a simpler bias correction could also be applied. A potential approach to deal with such bias could be to use the location and scale parameter from the observed time series but the shape parameters from the synthetic time series. In this way, bias of the model would be circumvented by local information but added value of using long synthetic time series would remain.

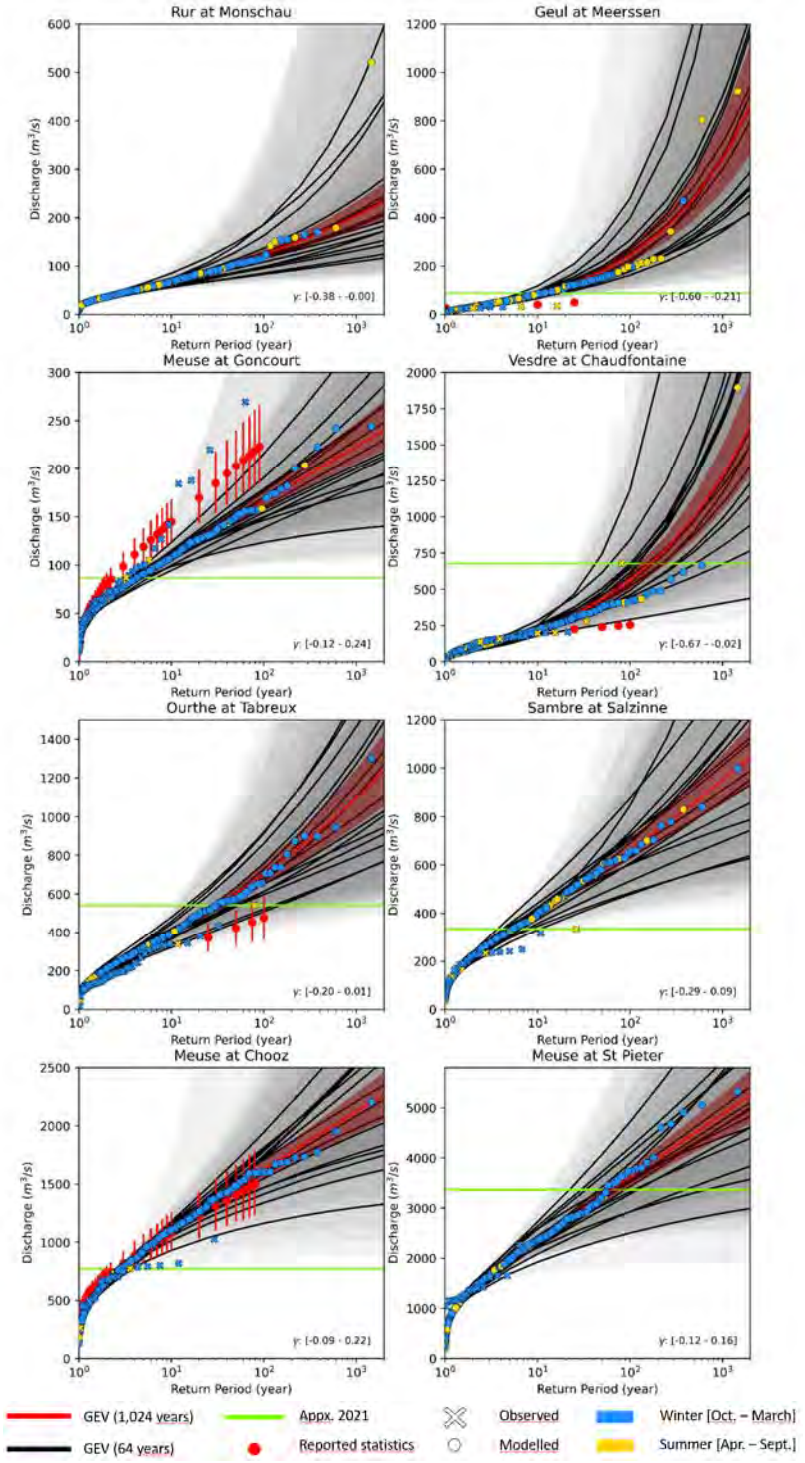


Figure 5-9:Extreme value distribution fit at selected locations across the Meuse catchment at an hourly time step. Black lines represent the GEV fit obtained from one ensemble member (64 hydrological years) and the range of shape parameters is written at the bottom right of each figure. Red lines show the GEV fit obtained from combining all ensembles (1,024 annual maxima from hydrological year, shown as full dots). The shading represents the 95% confidence interval from the distribution parameters uncertainty obtained by bootstrapping from 100 bootstrap samples, shown here with some transparency level. Observed annual maxima of discharge peaks when present are shown with crosses. A blue (yellow) color indicates that the peak occurred in winter (summer). Recorded or estimated maximum discharge peak reached in July 2021 are shown in green. Note that these estimates may deviate from the real discharge peak observed during the event, see also section 3.3.1. Reported return levels (when present) with uncertainty are shown with red circles and vertical bars, respectively, see also section 3.3.2.

Deltares
Royal Netherlands Meteorological Institute Ministry of Infrastructure and the Environment
Rijkswaterstaat Ministry of Infrastructure and Water Management
Deltares

Similar to the findings from daily timestep shown in Figure 5-5, Figure 5-10 indicates that the tail behaviour at the hourly timestep can deviate from the light tail (Gumbel distribution). However, this deviation is (as expected) more pronounced than for the daily timestep because the small-scale extreme event result in an even more extreme maximum at smaller time scales. The shape parameter in this case is heavily influenced by the most extreme events simulated. The highest simulated event for the Vesdre at Chaudfontaine, the Geul at Meersen and the Rur at Stah are extremely high, much higher proportionally than for the daily timestep. This results in a much steeper curvature of the GEV tail (or in other words a more negative shape parameter than for the daily timestep).

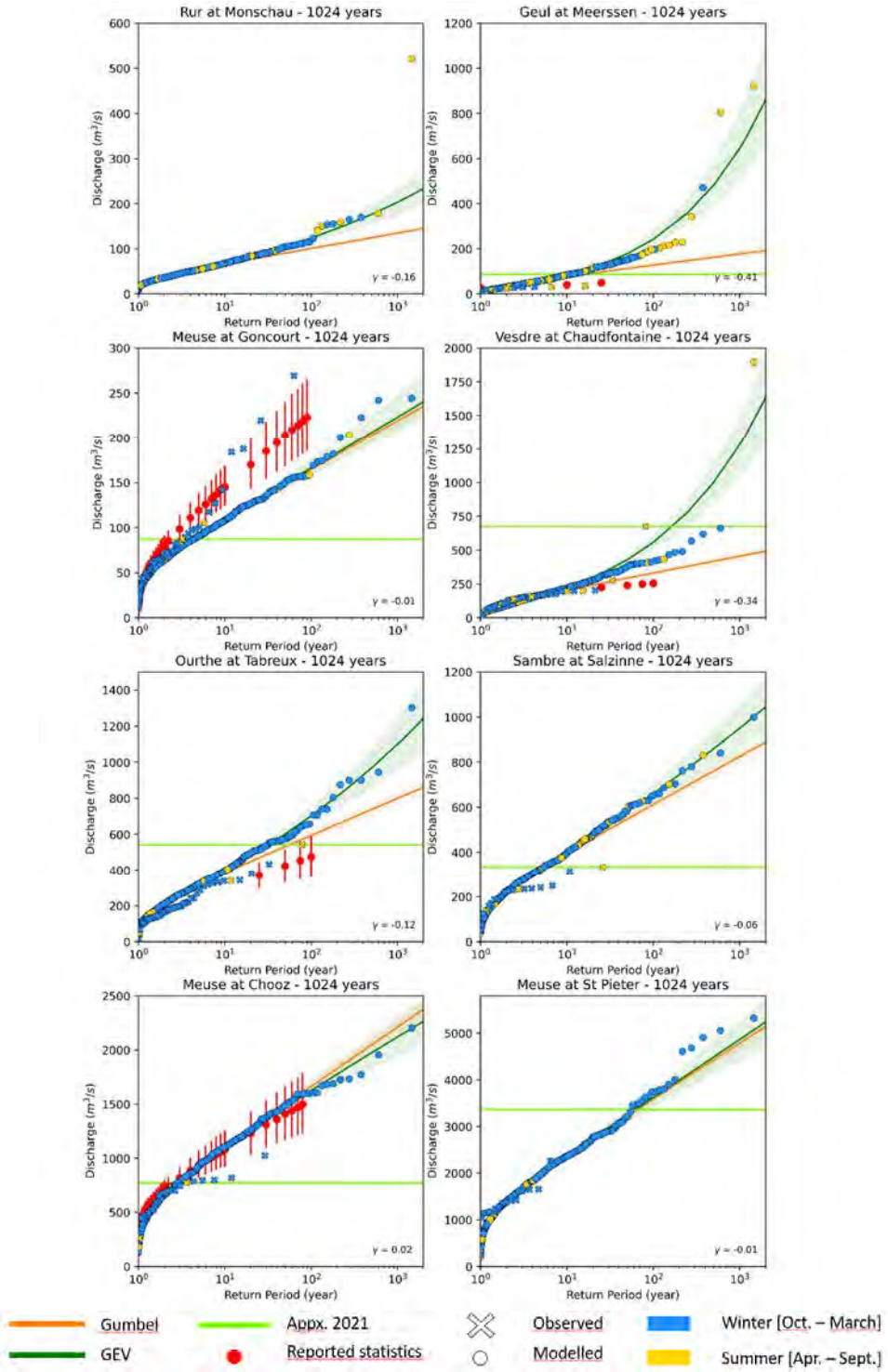


Figure 5-10: Extreme value distribution fit at selected locations across the Meuse catchment, based on hourly model results. Dark green lines represent the GEV fit obtained from combining all ensembles (1,024 hydrological years, shown as full dots) and blue lines the Gumbel fit. The shading represents the statistical uncertainty from the distribution parameters uncertainty. Observed discharge peaks when present are shown with crosses. Blue (yellow) colors indicates whether the peak occurred in winter (summer). Recorded July 2021 level when present are shown in green. Official statistics when present and reported uncertainty are shown with red circles and vertical bars, respectively.

These differences between simulated discharge extremes at the daily and hourly timestep are explicitly calculated for the empirical return periods, using the plotting position defined in section 5.3. The results are shown in Figure 5-11. Up to the 100-year return period approximately, the difference in hourly return levels is more or less linearly increasing. For larger return periods than the 100-year return period, large jumps can appear. This can result in hourly estimates being at least double the amount obtained from the simulated daily discharge extremes. Percentage differences with respect to daily return levels can be higher than 400% (Geul at Meerssen). For return periods lower than 100 years, percentage differences are either relatively constant (Sambre at Salzinne, Meuse at Goncourt, Meuse at Chooz, Meuse at St Pieter) or constantly increasing (Ourthe at Tabreux, Vesdre at Chaudfontaine, Geul at Meerssen). The lowest differences are observed for Meuse at Chooz and the Meuse at St Pieter, the two largest catchments considered.

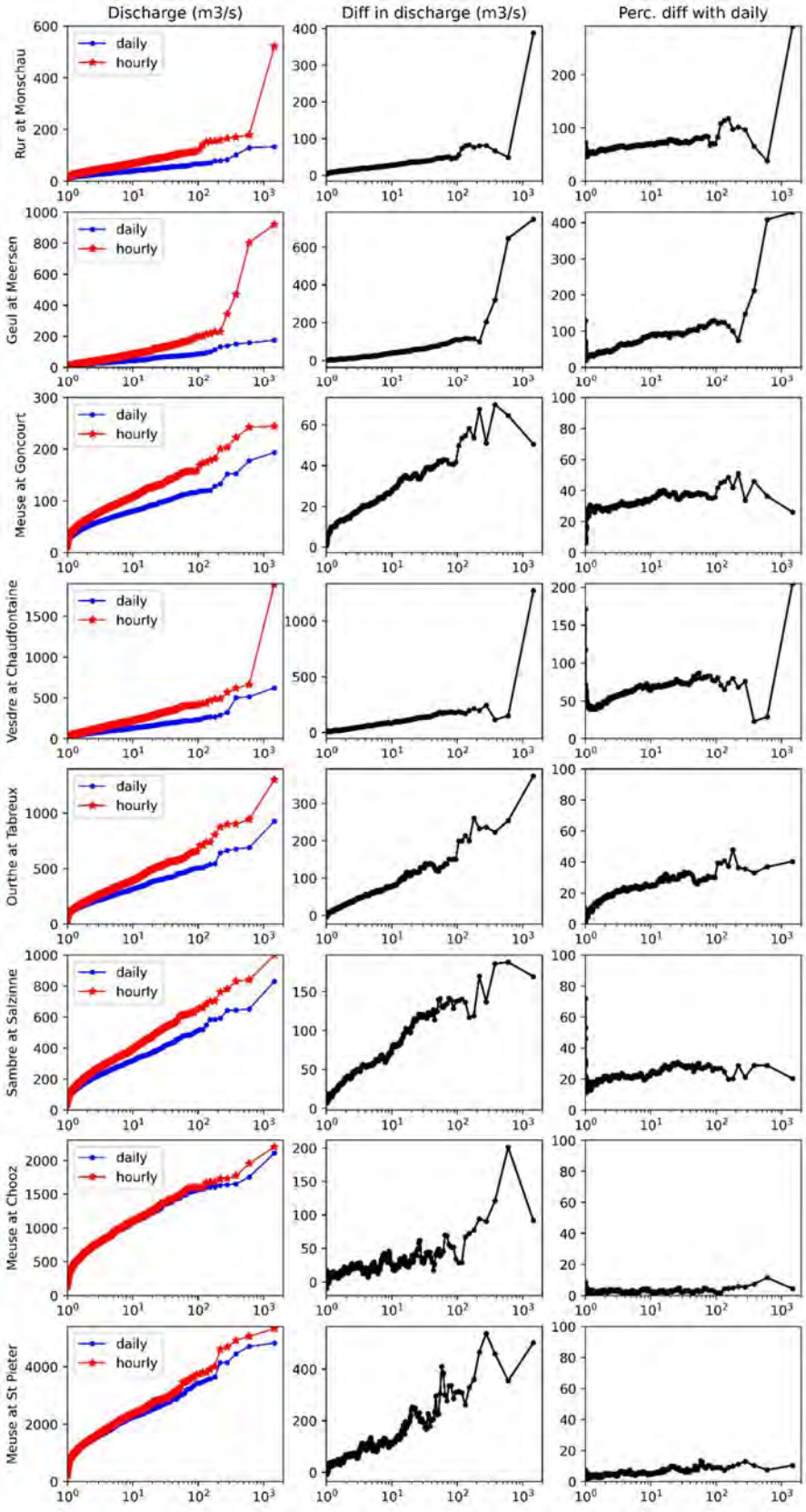


Figure 5-11. Left column - Empirical return levels obtained from 1,024 hydrological years at selected locations from the daily and hourly runs. Middle column: Difference in discharge between the two temporal resolutions. Right column: Relative percentage difference with respect to the empirical daily return levels. Catchments are ordered from the smallest catchment size (top row) to the largest catchment size (bottom row), see also Table 2-1.

Deltares
Royal Netherlands Meteorological Institute Ministry of Infrastructure and the Environment
Rijkswaterstaat Ministry of Infrastructure and Water Management
Deltares

There are various reasons for these differences between simulated hourly and daily discharge extremes. For extreme discharge events that last a few hours (less than a day), one can expect a large difference between hourly and daily time scales. In this case, the rising and falling limb of discharge hydrographs are not properly captured by running the model at a daily time scale. If the response time of a catchment is lower than a day, an extreme precipitation event will be attenuated when modelled at the daily time step and consequently the daily discharge peak as well. On the opposite, one can expect little difference between the two time scales at locations where discharge events last multiple days. In this case, the variation around the hourly discharge peak shows little variations with the day of maximum discharge due to the slowly varying hydrographs. This phenomenon is most likely observed for the Meuse at St Pieter. Discharge hydrographs duration is intrinsically linked with the response time of the catchment considered.

The difference in model performance between the hourly and daily time resolution is not necessarily consistent, as shown by comparing Table 4-4 and Table 4-6. Extreme discharges can be well captured for a time scale but over or underestimated for another time scale, with no constant bias between the two time scales. For example, difference in performance can affect summer or winter discharge extremes differently as some processes in the catchment may be better represented for one season than for the other. If this bias is not constant, the ranking of the discharge extremes will differ. This is shown for the Geul at Meerssen and the Sambre at Salzinnes in Figure 5-12 below. Focusing on the largest 10 annual maxima discharge (i.e. the highest 10 values in the scatter plot), there is a difference between the daily and hourly time scale. For the Sambre at Salzinnes, two out of the top ten annual maxima happen in summer at an hourly time scale but not at all the daily time scale. For the Geul at Meerssen, the proportion of the highest peaks happening in summer is not the same for the hourly and the daily time scale (i.e. four out of ten largest annual maxima happen in summer at the daily resolution versus eight out of the ten largest peaks for the hourly resolution). Again, this points to the fact that small catchments may produce a very high flood peak with a short thunderstorm in summer that cannot be properly captured at the daily average time scale.

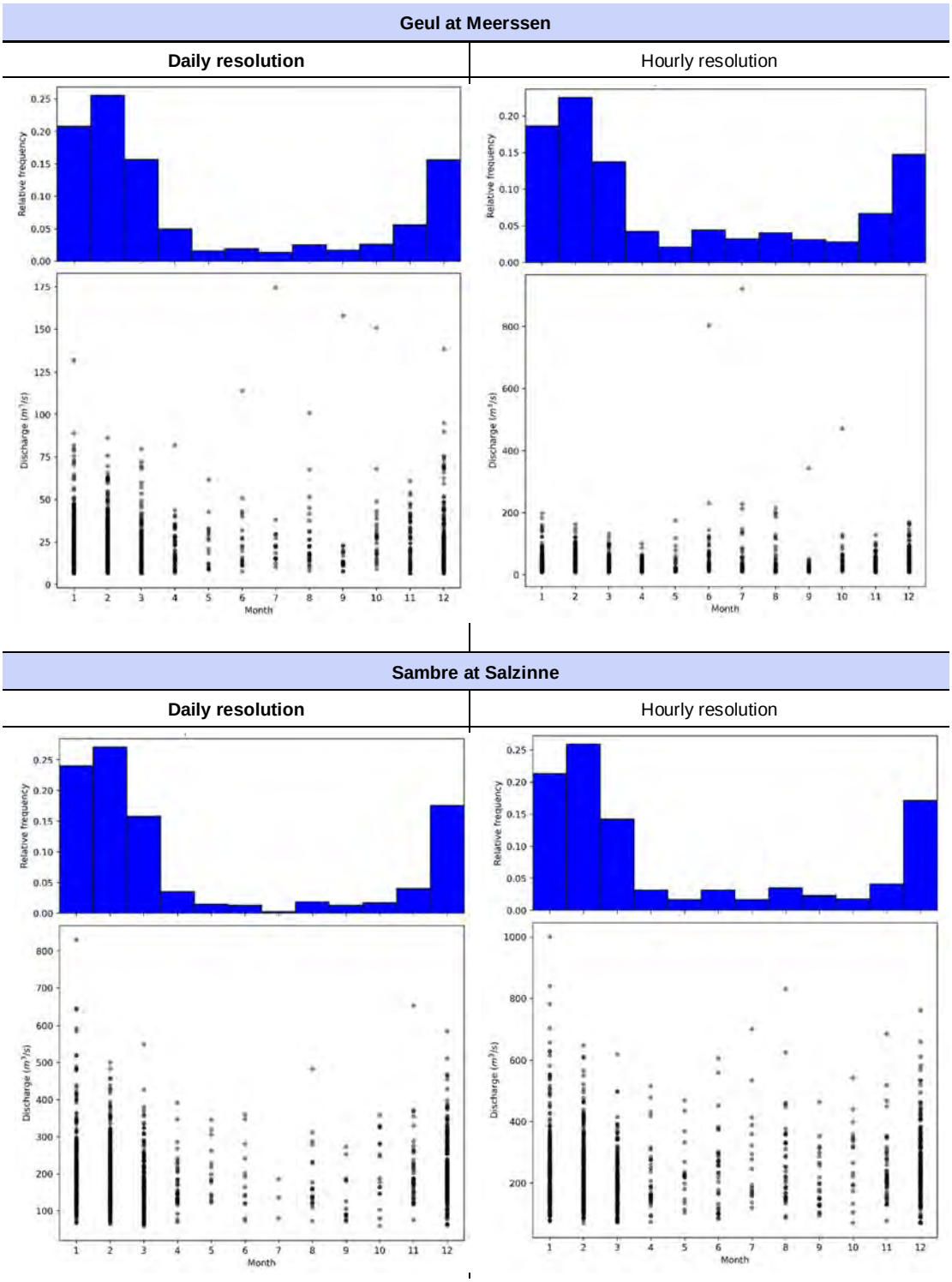


Figure 5-12: Relative frequency of the month of occurrence of the discharge extremes for the Geul at Meerssen (top) and the Sambre at Salzinne (bottom) for a daily time step (left column) and hourly time step (right column).

5.4.4 Return discharge levels for given return periods of selected tributaries

Table 5-1 and 5-2 list the obtained return levels from fitting the GEV extreme value distribution to the annual maxima from the hydrological years for the daily and hourly discharge peaks, respectively. Note that based on the previous findings of this section, these values should be interpreted with care for the 1000-year return period since large uncertainties exist from the limited sample size and due to the possible presence of summer extremes driving the behaviour of the tail of the distribution. The values for the other catchment of interests are given in Table D-2 and Table D-3 in Appendix 7D.4.

Table 5-1: Obtained return levels from fitting GEV and 95% confidence interval from the simulated daily discharges

Location of interest	5 year return level (m^3/s)		10 year return level (m^3/s)		100 year return level (m^3/s)		1000 year return level (m^3/s)	
	daily	95% CI	daily	95% CI	daily	95% CI	daily	95% CI
Rur at Monschau (Germany)	33	[32, 33]	40	[39, 41]	66	[61, 71]	98	[85, 112]
Geul at Meerssen (Netherlands)	35	[34, 36]	46	[45, 48]	94	[86, 105]	170	[146, 205]
Meuse at Goncourt (France)	69	[68,70]	82	[80,83]	120	[114, 127]	157	[143,173]
Vesdre at Chaudfontaine (Belgium)	100	[98,103]	131	[127,137]	266	[241, 304]	479	[401,603]
Ourthe at Tabreux (Belgium)	260	[257,263]	316	[309,323]	513	[479, 547]	741	[658, 829],
Sambre at Salzinne (Belgium)	268	[265, 271]	325	[319,333]	511	[485, 547]	704	[644, 791]
Meuse at Chooz (France)	899	[890,909]	1066	[1045,1087]	1571	[1489, 1658]	2042	[1868, 2236]
Meuse at St Pieter (Netherlands)	1881	[1852,1900]	2237	[2172,2280]	3329	[3073, 3507]	4368	[3825, 4772]

Table 5-2: Obtained return levels from fitting GEV and 95% confidence interval from the simulated hourly discharges

	5 year return level (m^3/s)		10 year return level (m^3/s)		100 year return level (m^3/s)		1000 year return level (m^3/s)	
Location of interest	**hourly**	**95% CI**	**hourly**	**95% CI**	**hourly**	**95% CI**	**hourly**	**95% CI**
Rur at Monschau (Germany)	54	[54, 55]	68	[66, 70]	124	[112, 136]	203	[170, 240]
Geul at Meerssen (Netherlands)	57	[56, 58]	83	[79, 87]	242	[210, 270]	644	[503, 782]
Meuse at Goncourt (France)	91	[90, 92]	109	[107, 112]	166	[156, 177]	223	[201, 248]
Vesdre at Chaudfontaine (Belgium)	159	[156, 162]	221	[214, 229]	557	[504, 618]	1279	[1074, 1538]
Ourthe at Tabreux (Belgium)	318	[314, 323]	398	[388, 411]	701	[653, 770]	1099	[969, 1297]
Sambre at Salzinne (Belgium)	327	[323, 330]	402	[393, 411]	660	[621, 703]	950	[855, 1059]
Meuse at Chooz (France)	924	[916, 933]	1098	[1079, 1117]	1625	[1552, 1705]	2120	[1965, 2300]
Meuse at St Pieter (Netherlands)	1984	[1957, 2007]	2377	[2318, 2431]	3621	[3379,3855]	4862	[4328, 5413]

5.5 Implications and future work

The use of the synthetic climate dataset to generate long time series of discharges with the improved hydrological model of the Meuse provides the following advantages. First, estimates of discharge return levels anywhere in the catchment are now possible. Second, the use of the 16 independent climate ensemble members resulted in 1,040 years of simulated discharges, or 1,024 hydrological years. We find that such extended record length is crucial to reduce the statistical uncertainty from discharge return levels. Third, at the daily time step where a complete spatial analysis was possible, we observe a clear spatial pattern of the shape parameter across the Meuse catchment. The shape parameter describes the tail of the extreme value distribution and is an indicator of extreme discharge behaviour. Therefore, the spatial pattern of the shape parameter is linked to the different extreme hydrological signatures present in the catchment.

While robust estimates from the location and scale parameter can be made based on record lengths of a few decades long, this is not the case for the shape parameter, a key parameter that describes the tail of the GEV distribution and can greatly impact return level estimates from large return periods (higher than 100-year return period). We find that using the ensemble members greatly narrows down this uncertainty and leads to converged shape parameters.

This could be very valuable in helping refine extreme value fit based for observed time series (of limited record length). We have shown that this is of most importance in smaller basins, where high return periods extremes are much higher than lower return periods (heavy tailed GEV), and to an even greater extent in cases where summer extremes are an important consideration in determining extreme return levels. In locations where systematic biases are observed between observed and simulated discharges, an error correction model could be applied. For example, the location and scale parameter from the observed discharge times series could be used to correct the centre and spread of the extreme value distribution while the shape parameter, representing the skewness of the distribution (or tail curvature) could be extracted from the simulated discharge since it has less statistical uncertainty.

Differences are found between discharge return levels at the daily and hourly time steps. This emphasizes the importance of performing extreme value analysis at relevant flood time scales locally. For small tributaries and large return periods, this difference can be significant, more than 350% of the daily return level. In locations with flat and elongated flood waves lasting multiple days, this difference is less important. Understanding extreme discharge hydrograph characteristics and how they travel through the catchment will be a key future aspect to explore in understanding peak discharges through the Meuse catchment.

The analysis of the 1,024 hydrological years reveals that the vast majority of discharge extremes happen in winter and as a result dominate the fit of the extreme value distribution. This is also corroborated by analysing the shape parameters obtained from summer maxima and winter maxima. Interestingly, summer extremes result in distribution with heavier tails than winter extremes, everywhere in the Meuse catchment. At some locations, the highest discharge extremes are clustered in summer. If that is the case, the traditional approach of using annual maxima irrespective of the season is ill-posed and not applicable anymore. However, the number of intense summer events remain limited even from the 16 ensemble members. While the RACMO dataset contain rainfall events similar or worse than the July 2021 event, the total synthetic record length of 1,040 years of the RACMO dataset is not long enough to assess its impact on summer extremes with confidence for large catchments (approximately larger than 2,500 km^2) such at the Meuse at St Pieter. Only a few summer rainfall events of similar or worse magnitude are present and this is not enough to assess the impact of the spatial variability of rainfall events on extreme discharges for a catchment such as the Meuse at St Pieter.

Separating the synthetic generation of summer and winter extremes may be an interesting avenue to compensate for this bias and more explicitly explore the role of summer events for very large return periods (higher than 100 year) where large uncertainties remain. Current frequency analysis approaches based only on winter extremes but where summer extremes dominate the most extreme return level most likely under- or overestimate the highest return periods. Since some flood mitigation decision and design are directly based on estimates of high return periods (100-year or higher), this can have importance consequence for decision makers. For example, the Netherlands mainly focusses on the winter period for the assessment of flood defences in the BOI program. The July 2021 event has shown that understanding the summer discharges is also essential for flood risk. This uncertainty should be better reflected in current estimates and, if possible, be removed by more in-depth analysis in future work.

6 Conclusions and future work

This study provides extreme discharge return level estimates based on a physically-based modelling chain using long synthetic meteorological time series of the current climate and a spatially-distributed hydrological model. In brief, the modelling chain is successfully applied for the Meuse basin with more in-depth analysis at selected catchments of interests. This report focused on the improvement of the performance of the distributed hydrological model for the Meuse and on the obtained estimates of extreme discharge return levels. Discharge values for low to medium return periods (around the 10- to 30-year return period, depending on the length of the observations) are in line with current estimates from tributaries of the Meuse based on observations. For the Meuse at St Pieter, where a statistical weather generator was applied, the 100-year return level at the daily time step obtained is 3,329 m^3/s, which is in line with previous estimates of 3,220 m^3/s (Hegnauer et al., 2014). Most importantly, the extreme discharges for high return periods (up to return periods of 300 years considering the total record length of the synthetic discharge extremes) are more robust with this model cascade, and the behaviour of the tail of the distribution much better understood and put into perspective of the season. We showed that, with the long synthetic discharge time series obtained, distinct summer extremes and winter extremes can be found and used to identify mixed distributions in small and medium catchments. This in turn revealed that events such as the July 2021 event are found back in our new distribution functions and must be accounted for in design. In large catchments, approximately higher than 2,500 km^2, the synthetic rainfall time series are not long enough to fully capture the impact of summer rainfall extreme events on summer discharge extremes. While synthetic rainfall events of similar or even worse magnitude than the July 2021 event are present, they are not enough events to cover the spatial heterogeneity possible in these large catchments and therefore their impact on the frequency of summer discharge extremes remain uncertain. Finally, the results show some bias in some areas where further optimization of the hydrological model is still needed.

The conclusions and recommendations for future work are worked out in more detail in the following sections. Section 6.1 focuses on the technical findings, limitations, and suggestions for future work where section 6.2 discusses the implications of our findings for decision makers.

6.1 For technical purposes

6.1.1 Main findings

Based on the results from this study, we can conclude the main following technical points:

- The shape parameter, describing the tail of the extreme value distribution is an indicator of extreme discharge behaviour. We show clear patterns of the shape parameters across the Meuse (Figure 5-3d). Smaller and steeper tributaries (e.g. the Vesdre at Chaudfontaine, the Geul at Meerssen) exhibit heavy tails (negative shape parameter of -0.20/-0.34 and -0.21/-0.41 for the daily/hourly time step, meaning a steep upward curving of the tail) while larger river basins exhibit light tails (shape parameter close to zero as in the Gumbel distribution).
- Using longer synthetic time series reduces the statistical uncertainty around extreme value distribution parameters. For 1,040 years of data, estimates up until approximately 300 year return can be robustly estimated. This is particularly beneficial for the estimate of the shape parameter, which requires long time series in order to reach convergence.

Where model bias is found in the model results, the shape parameter derived from our analysis could be used by stakeholders in combination with the location and scale parameters derived from (much shorter) observation times series. In detail, the location and scale parameter from the observed discharge times series could be used to correct the center and spread of the extreme value distribution while the shape parameter, representing the skewness of the distribution (or tail curvature) could be extracted from the simulated discharge since it has less statistical uncertainty. In the longer term, further improvements of the hydrological model in areas of poor performance are favoured and recommended to align both model results and observed results without the need to apply such method to bias correct results in a postprocessing step.

- The considered time step of the hydrological model influences extreme return levels substantially. The extent of this behaviour is dependent on catchment characteristics. On average, we find a difference of 20% higher magnitude for the hourly time step compared to daily for the catchment of interests with a catchment size lower than ~2,500 m^2 but this can vary up to 400% for the Geul at Meerssen for the 100-year return period. In general, smaller tributaries are more sensitive to the time step of the model. This is expected as these catchments have faster response times, typically lower than a day.
- At certain locations, mainly in smaller and steep catchments, summer events drive the most extreme discharge events, i.e. the tail of the extreme value distribution. Having long synthetic discharge time series allows us to show the separate contribution of summer and winter discharge extremes at smaller catchments. The station Rur at Stah is a good example of such behaviour (Figure 5-8). At this location, discharge magnitudes of low return periods are lower in summer than in winter (i.e. this statistically translates to a lower summer location parameter than winter location parameter). However, above a certain return period, the empirical distribution of the summer extremes may become higher than winter extremes. This is most likely indicative of specific weather and flood mechanisms for these summer events. This can have strong implications for return level estimates as the considered population of discharge extremes is not identically distributed anymore. This calls for tailored extreme value analysis where seasonality is explicitly considered. If that is the case, the traditional approach of using annual maxima irrespective of the season is ill-posed and not applicable anymore. For larger catchments, such as the Meuse at St Pieter, even though the synthetic rainfall time series contain events of similar or worse magnitude than the July 2021 event, they are not enough events to capture their contribution to summer extreme discharge events. Further work is needed to quantify the role of summer extreme rainfall events in extreme summer discharge.

6.1.2 Limitations and future work

Based on the results from this report, we also identified the current limitations and suggestions for future work:

- Focusing more specifically on extreme summer events remains difficult given the return periods at which they occur. In that sense, estimating the influence of summer events on the highest return periods (larger than 300-year return period) is not enough from 1,040 years of simulations. The analysis could be repeated with longer meteorological datasets, such as the SEAS5 dataset, the RACMO dataset which was used for the climate change impact assessment for the Netherlands done in 2014. This would not only help capture more summer events but also assess the robustness of the spatial pattern of the shape parameter derived in this study. An alternative approach could be to use the RACMO dataset in a statistical weather generator, to generate new synthetic rainfall events, other spatial resampling techniques or deep learning methods. These approaches however, should ensure that the generated rainfall events are physically plausible.

- Even though extreme summer events may be particularly important and stem from other physical mechanisms, these events are underrepresented or absent from the observed time series used to calibrate the hydrological model. Therefore, future work should focus on tailoring calibration towards those events.
- We recommend further optimization of the hydrological model at the hourly time steps since it resulted in improvements for both the hourly and daily time scales. Furthermore, an hourly time scale is needed to properly represent discharge extremes in smaller and steeper catchment.
- In general, uncertainties across the whole modelling chain could be quantified by increasing the number of climate models and hydrological models or parameter realization.
- In future work, the discharge dataset could also be used in other relevant aspects for flood response or drought analysis. Having continuous discharge time series allow to explore the spatial dependence between flood extremes in the Meuse. This can be particularly relevant to understand the interactions between the flood peaks happening in the main river with respect to the tributaries. It also allows to better document the dynamics and natural variability of low flows.

6.2 For decision making

Looking back at the estimation of the return period of the 2021 event, these results provide important insights for decision making:

- Under a traditional approach, i.e. using annual maxima and a Gumbel fit, the return period of this event would likely result in an overestimation for smaller tributaries. In these smaller catchments, the summer events drive the most extreme discharges events, i.e. the tail of the extreme value distribution.
- We find other summer events of the same or worse magnitude than what observed in July 2021 in many small to medium catchments. Natural variability is, therefore, important and not sufficiently present in current approaches for extreme discharge frequency analysis as they rely on short observations time series or lack physical processes when time series are lengthened with a statistical weather generator. These extreme events should be considered in flood response and for design purposes as they are physically plausible.
- The physically-based distributed modelling chain is essential to derive spatial patterns of extreme discharge behaviour and to increase our understanding of the complex interactions between topography, catchment size, meteorological event types, etc.
- Extreme discharge behaviour may be driven by multiple processes. For example, in some locations, extremes of similar magnitude can happen both in summer and winter. However, the underlying processes and meteorological conditions for each season are potentially very different. In such case, these two populations should be addressed separately for return level estimation, which has strong implications for design. This was shown for a small catchment. For larger catchments, longer synthetic rainfall time series are needed to also account for the spatial complexity introduced from the catchment size and the catchment response time. Scaling up this approach should be further investigated.
- In locations where very extreme return periods are of interests, for example the 10,000 year return period or higher in the Netherlands, the 1,040 years of synthetic discharge extremes is not enough to estimate these very high return periods with certainty. For this, much longer extreme discharge time series are needed. Different options are mentioned here, such as repeating this analysis with longer meteorological time series, using a statistical weather generator to extend the time series, or other spatial resampling techniques or deep learning to generate new extreme weather events to be modelled in a hydrological model.

Each alternative has specific challenges in their implementation and special care should be taken to ensure that the generated weather events are physically plausible.

- The results show a high level of confidence for the climate and hydrological model. Although some locations can benefit from further improvements, the insights derived above apply at every location and are crucial for decision making in the Meuse. Flood mitigation strategies and design are based on estimates of high return levels, for example the 100-year return period or higher. The July 2021 event has shown that understanding extreme summer discharges and their influence on flood return levels is essential. In locations where extreme summer discharges can be of similar magnitude as winter extremes, neglecting or grouping them will result in an over- or underestimation of return levels.
- Other external drivers, of importance for return level estimates, such as climate and land cover changes, should be further considered. Although outside of the scope of this current report, the current methodology is suitable for such analysis and could be done in future work.
- The continuous discharge time series across the whole Meuse catchment modelled for this report can be used to investigate other topics of importance for decision makers. For example, they can help quantify the importance of the spatial dependence between the main river and her tributaries. More generally, the RACMO model provides valuable joint time series of different climate drivers that could also be used to analyse the role of climate variability for both the Meuse catchment, as done here, but also at the coast. Joint time series of wind and rainfall from RACMO could be used to investigate the role of the natural variability in the dependence between storm surge and river discharge at the coast.

7 References

Adams, J. M., Gasparini, N. M., Hobley, D. E. J., Tucker, G. E., Hutton, E. W. H., Nudurupati, S. S., & Istanbulluoglu, E. (2017). The Landlab v1. 0 OverlandFlow component: a Python tool for computing shallow-water flow across watersheds. *Geoscientific Model Development*, *10*(4), 1645–1663.

Andreadis, K. M., Schumann, G. J. P., & Pavelsky, T. (2013). A simple global river bankfull width and depth database. *Water Resources Research*, *49*(10), 7164–7168. https://doi.org/10.1002/wrcr.20440

Bocharov, G. (2022). *pyextremes v2.3.1. Available at https://georgebv.github.io/pyextremes/*.

Bouaziz, L. (2020). *Evaluation of hydrological models of the Meuse river basin*.

Bouaziz, L. (2021). *Internal processes in hydrological models, a glance at the Meuse basin from space*.

Bouaziz, L., & Buitink, J. (2022). *Developments of the wflow Meuse and Rhine hydrological models in 2022*.

Bouaziz, L. J. E., Aalbers, E. E., Weerts, A. H., Hegnauer, M., Buiteveld, H., Lammersen, R., Stam, J., Sprokkereef, E., Savenije, H. H. G., & Hrachowitz, M. (2022a). Ecosystem adaptation to climate change: The sensitivity of hydrological predictions to time-dynamic model parameters. *Hydrology and Earth System Sciences*, *26*(5), 1295–1318. https://doi.org/10.5194/hess-26-1295-2022

Bouaziz, L. J. E., Aalbers, E. E., Weerts, A. H., Hegnauer, M., Buiteveld, H., Lammersen, R., Stam, J., Sprokkereef, E., Savenije, H. H. G., & Hrachowitz, M. (2022b). Ecosystem adaptation to climate change: The sensitivity of hydrological predictions to time-dynamic model parameters. *Hydrology and Earth System Sciences*, *26*(5), 1295–1318. https://doi.org/10.5194/hess-26-1295-2022

Bouaziz, L. J. E., Steele-Dunne, S. C., Schellekens, J., Weerts, A. H., Stam, J., Sprokkereef, E., Winsemius, H. H. C., Savenije, H. H. G., & Hrachowitz, M. (2020). Improved understanding of the link between catchment-scale vegetation accessible storage and satellite-derived Soil Water Index. *Water Resources Research*. https://doi.org/10.1029/2019WR026365

Bouaziz, L., Weerts, A., Schellekens, J., Sprokkereef, E., Stam, J., Savenije, H., & Hrachowitz, M. (2018). Redressing the balance: Quantifying net intercatchment groundwater flows. *Hydrology and Earth System Sciences*, *22*(12), 6415–6434. https://doi.org/10.5194/hess-22-6415-2018

Bourgault, P., Huard, D., Smith, T. J., Logan, T., Aoun, A., Lavoie, J., Dupuis, É., Rondeau-Genesse, G., Alegre, R., Barnes, C., Laperrière, A. B., Biner, S., Caron, D., Ehbrecht, C., Fyke, J., Keel, T., Labonté, M.-P., Lierhammer, L., Low, J.-F., … Whelan, C. (2023). xclim: xarray-based climate data analytics. *Journal of Open Source Software*, *8*(85), 5415. https://doi.org/10.21105/joss.05415

Bruwier, M., Erpicum, S., Pirotton, M., Archambeau, P., & Dewals, B. J. (2015). Assessing the operation rules of a reservoir system based on a detailed modelling chain. *Natural Hazards and Earth System Sciences*, *15*(3), 365–379. https://doi.org/10.5194/nhess-15-365-2015

Chbab, H. (2017). *Basisstochasten WTI-2017 : statistiek en statistische onzekerheid*.

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., & Jones, P. D. (2018). An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets. *Journal of Geophysical Research: Atmospheres*, *123*(17), 9391–9409. https://doi.org/10.1029/2017JD028200

Couasnon, A., Scussolini, P., Tran, T. V. T., Eilander, D., Muis, S., Wang, H., Keesom, J., Dullaart, J., Xuan, Y., Nguyen, H. Q., Winsemius, H. C., & Ward, P. J. (2022). A Flood Risk Framework Capturing the Seasonality of and Dependence Between Rainfall and Sea Levels—An Application to Ho Chi Minh City, Vietnam. *Water Resources Research*, *58*(2). https://doi.org/10.1029/2021WR030002

De Almeida, G. A. M., & Bates, P. (2013). Applicability of the local inertial approximation of the shallow water equations to flood modeling. *Water Resources Research*, *49*(8), 4833–4844. https://doi.org/10.1002/wrcr.20366

de Boer-Euser, T., McMillan, H. K., Hrachowitz, M., Winsemius, H. C., & Savenije, H. H. G. (2016). Influence of soil and climate on root zone storage capacity. *Water Resources Research*. https://doi.org/10.1002/2015WR018115

de Wit, M. (2008). *Van regen tot {M}aas* (M. Joenje, Ed.; p. 54). Uitgeverij Veen Magazines.

Dewals, B., Erpicum, S., Pirotton, M., & Archambeau, P. (2021, April). The July 2021 extreme floods in the Belgian part of the Meuse basin. . *Hydrolink Magazine*.

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arsouze, T., Bergman, T., Bernardello, R., Boussetta, S., Caron, L.-P., Carver, G., Castrillo, M., Catalano, F., Cvijanovic, I., Davini, P., Dekker, E., Doblas-Reyes, F. J., Docquier, D., Echevarria, P., Fladrich, U., … Zhang, Q. (2022). The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6. *Geoscientific Model Development*, *15*(7), 2973–3020. https://doi.org/10.5194/gmd-15-2973-2022

Dullaart, J. C. M., Muis, S., Bloemendaal, N., Chertova, M. V., Couasnon, A., & Aerts, J. C. J. H. (2021). Accounting for tropical cyclones more than doubles the global population exposed to low-probability coastal flooding. *Communications Earth & Environment*, *2*(1), 1–11. https://doi.org/10.1038/s43247-021-00204-9

Eilander, D., Boisgontier, H., Bouaziz, L. J. E., Buitink, J., Couasnon, A., Dalmijn, B., Hegnauer, M., de Jong, T., Loos, S., Marth, I., & van Verseveld, W. (2023). HydroMT: Automated and reproducible model building and analysis. *Journal of Open Source Software*, *8*(83), 4897. https://doi.org/10.21105/joss.04897

European Environment Agency. (2018). *Corine Land Cover (CLC) 2018, Version 2020_20u1*. https://land.copernicus.eu/pan-european/corine-land-cover/

Früh, B., Feldmann, H., Panitz, H.-J., Schädler, G., Jacob, D., Lorenz, P., & Keuler, K. (2010). Determination of Precipitation Return Values in Complex Terrain and Their Evaluation. *Journal of Climate*, *23*(9), 2257–2274. https://doi.org/10.1175/2009JCLI2685.1

Gao, H., Fenicia, F., & Savenije, H. H. G. (2023). HESS Opinions: Are soils overrated in hydrology? *Hydrology and Earth System Sciences*, *27*(14), 2607–2620. https://doi.org/10.5194/hess-27-2607-2023

Günther, A. (2023). *International hydrogeological map of Europe*. BGR Hannover (Federal Institute for Geosciences and Natural Resources).

Hannart, A. (2019). *Introduction to Extreme Value Analysis*. https://www.wcrp-climate.org/images/summer_school/Agenda_and_Program_Downloadable_Lectures/Summerschool_Nanjing_lecture_Extreme_value_theory_Hannart.pdf

Hartgring, S. (2023). *On forecasting the Rur river*.

Hegnauer, M., Beersma, J. J., van den Boogaard, H. F. P., Buishand, T. A., & Passchier, R. H. (2014). *Generator of Rainfall and Discharge Extremes (GRADE) for the Rhine and Meuse basins - Final report of GRADE 2.0*.

Hengl, T., De Jesus, J. M., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. In *PLoS ONE* (Vol. 12, Issue 2). https://doi.org/10.1371/journal.pone.0169748

Hooghart, J. C., & Lablans, W. N. (1988). Van Penman naar Makkink: een nieuwe berekeningswijze voor de klimatologische verdampingsgetallen. *De Bilt, Royal Netherlands Meteorological Institute (KNMI), De Bilt, the Netherlands*.

Horn, G., & Hurkmans, R. (2022). *Analyse & Maatregelen wateroverlast Roer*.

Hoshino, T., Shimizu, K., Hegnauer, M., & Yamada, T. J. (2022). *Evaluating the Impact of Climate Change on the Return Period of Flood Peak Discharge over The Tokachi River Basin, Northern Japan by Using a Massive Ensemble Climate Dataset*.

Imhoff, R. O., van Verseveld, W. J., van Osnabrugge, B., & Weerts, A. H. (2020). Scaling Point-Scale (Pedo)transfer Functions to Seamless Large-Domain Parameter Estimates for High-Resolution Distributed Hydrologic Modeling: An Example for the Rhine River. *Water Resources Research*, *56*(4), 1–28. https://doi.org/10.1029/2019WR026807

Klein, A. C. (2022). *Hydrological Response of the Geul Catchment to the Rainfall in July 2021 (MSc thesis)*. http://repository.tudelft.nl/.

Lehmkuhl, F., Schüttrumpf, H., Schwarzbauer, J., Brüll, C., Dietze, M., Letmathe, P., Völker, C., & Hollert, H. (2022). Assessment of the 2021 summer flood in Central Europe. *Environmental Sciences Europe*, *34*(1), 107. https://doi.org/10.1186/s12302-022-00685-1

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., & others. (2011). High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment*, *9*(9), 494–502.

Lin, P., Pan, M., Allen, G. H., de Frasson, R. P., Zeng, Z., Yamazaki, D., & Wood, E. F. (2020). Global Estimates of Reach-Level Bankfull River Width Leveraging Big Data Geospatial Analysis. *Geophysical Research Letters*, *47*(7). https://doi.org/10.1029/2019GL086405

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

Deltares
Royal Netherlands Meteorological Institute Ministry of Infrastructure and the Environment
Deltares
Rijkswaterstaat Ministry of Infrastructure and Water Management

Ludwig, P., Ehmele, F., Franca, M. J., Mohr, S., Caldas-Alvarez, A., Daniell, J. E., Ehret, U., Feldmann, H., Hundhausen, M., Knippertz, P., Küpfer, K., Kunz, M., Mühr, B., Pinto, J. G., Quinting, J., Schäfer, A. M., Seidel, F., & Wisotzky, C. (2023). A multi-disciplinary analysis of the exceptional flood event of July 2021 in central Europe - Part 2: Historical context and relation to climate change. *Natural Hazards and Earth System Sciences*, *23*(4), 1287–1311. https://doi.org/10.5194/nhess-23-1287-2023

Messager, M. L., Lehner, B., Grill, G., Nedeva, I., & Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications*, *7*(1), 13603.

Mohr, S., Ehret, U., Kunz, M., Ludwig, P., Caldas-Alvarez, A., Daniell, J. E., Ehmele, F., Feldmann, H., Franca, M. J., Gattke, C., Hundhausen, M., Knippertz, P., Küpfer, K., Mühr, B., Pinto, J. G., Quinting, J., Schäfer, A. M., Scheibel, M., Seidel, F., & Wisotzky, C. (2023). A multi-disciplinary analysis of the exceptional flood event of July 2021 in central Europe - Part 1: Event description and analysis. *Natural Hazards and Earth System Sciences*, *23*(2), 525–551. https://doi.org/10.5194/nhess-23-525-2023

Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Lee, S., Twardziok, S. O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., & Köster, J. (2021). Sustainable data analysis with Snakemake. *F1000Research*, *10*, 33. https://doi.org/10.12688/f1000research.29032.2

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Neal, J., Schumann, G., & Bates, P. (2012). A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas. *Water Resources Research*, *48*(11). https://doi.org/10.1029/2012WR012514

Palutikof, J. P., Brabson, B. B., Lister, D. H., & Adcock, S. T. (1999). A review of methods to calculate extreme wind speeds. *Meteorological Applications*, *6*(2), 119–132. https://doi.org/10.1017/S1350482799001103

Papalexiou, S. M., & Koutsoyiannis, D. (2013). Battle of extreme value distributions : A global survey on extreme daily rainfall. In *Water Resources Research* (Vol. 49, Issue 1, pp. 187–201). https://doi.org/10.1029/2012WR012557

Pugh, D., & Woodworth, P. (2014). *Sea-level science: understanding tides, surges, tsunamis and mean sea-level changes.* . Cambridge University Press.

Schulzweida, U. (2022). *CDO User Guide (2.1.0)*. https://doi.org/10.5281/zenodo.7112925

Schweckendiek, T., & Slomp, R. (2018). Risk-based safety standards and safety assessment tools in the Netherlands. *Wasserbauwerke Im Bestand-Sanierung, Umbau, Ersatzneubau: Dresdner Wasserbaukolloquium*, 33–44.

Slomp, R., Knoeff, H., Bizzarri, A., Bottema, M., & de Vries, W. (2016). Probabilistic Flood Defence Assessment Tools. *E3S Web of Conferences*, *7*, 03015. https://doi.org/10.1051/e3sconf/20160703015

Stucky. (2021). *Analyse indépendante sur la gestion des voies hydrauliques lors des intempéries de la semaine du 12 juillet 2021 : Lot 1 - factualisation. Technical Report, Document nr 5875/4001a.*

van der Veen, R. (2021). *Topafvoeren hoogwater Maas juli 2021.*

van Dorland, R., Beersma, J., Bessembinder, J., Bloemendaal, N., van den Brink, H., Brotons Blanes, M., Drijfhout, S., Haarsma, R., Krikken, F., Le Bars, D., Lenderink, G., van Meijgaard, E., Reerink, T., Selten, F., Severijns, C., Siegmund, P., Sterl, A., Overbeek, B., de Vries, H., … van der Wiel, K. (2023). *KNMI National Climate Scenarios 2023 for the Netherlands*.

van Meijgaard, E., Van Ulft, L. H., Van de Berg, W. J., Bosveld, F. C., Van den Hurk, B. J. J. M., Lenderink, G., & Siebesma, A. P. (2008). *The KNMI regional atmospheric climate model RACMO, version 2.1* .

van Osnabrugge, B., Weerts, A. H., & Uijlenhoet, R. (2017). genRE: A Method to Extend Gridded Precipitation Climatology Data Sets in Near Real-Time for Hydrological Forecasting Purposes. *Water Resources Research*, *53*(11), 9284–9303.

Van Verseveld, W. J., Weerts, A. H., Visser, M., Buitink, J., Imhoff, R. O., Boisgontier, H., Bouaziz, L., Eilander, D., Hegnauer, M., Ten Velden, C., & Russell, B. (2022). Wflow_sbm v0.6.1, a spatially distributed hydrologic model: from global data to local applications. *Geoscientific Model Development (Discussions)*, *182*. https://doi.org/10.5194/gmd-2022-182

van Voorst, L., & van den Brink, H. (2023). *An evaluation of the use of regional climate model data applied to extreme precipitation in the Meuse basin*.

Vertessy, R. A., & Elsenbeer, H. (1999). Distributed modeling of storm flow generation in an Amazonian rain forest catchment: Effects of model parameterization. *Water Resources Research*, *35*(7), 2173–2187. https://doi.org/10.1029/1999WR900051

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Vorogushyn, S., Apel, H., Kemter, M., & Thieken, A. (2022). Analyse der Hochwassergefährdung im Ahrtal unter Berücksichtigung historischer Hochwasser. *Hydrol. Wasserbewirts.*, *66*, 244–254.

Wang-Erlandsson, L., Bastiaanssen, W., Gao, H., Jägermeyr, J., Senay, G., van Dijk, A., Guerschman, J., Keys, P., Gordon, L., & Savenije, H. (2016). Global root zone storage capacity from satellite-based evaporation data. *EGU General Assembly Conference Abstracts*, *18*, 14504.

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., & Pavelsky, T. M. (2019). MERIT Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography Dataset. *Water Resources Research*, *55*(6), 5053–5073. https://doi.org/10.1029/2019WR024873

A Catchment of interests and datasets

Figure A-1 shows the selected catchments, identified as points of interests by the partners and stakeholders, for which the results are shown in Appendix.

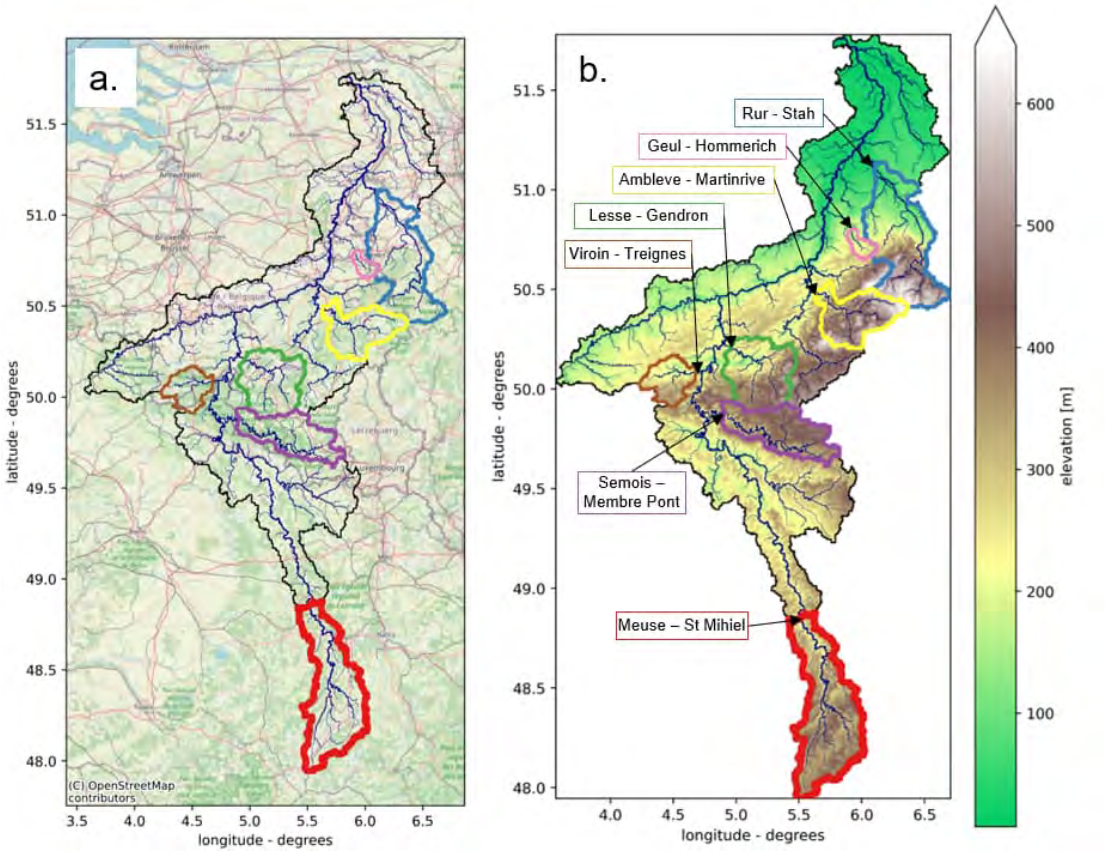


Figure A-1: Selected catchments for the Meuse river and her tributaries, with background map (a) and topography (b) presented in the appendix.

A.1 Recorded and estimated discharge for July 2021

Figure A-2 and Figure A-3 show the daily and hourly discharge time series, when available, for the selected catchments of interest. At the Rur at Monschau, the Rur at Stah and the Geul at Hommerich, the gauge did not record any peak discharges around the time of the event. For the Geul at Meerssen, an hourly peak discharge estimate of 88 m^3/s for the event is used based on the study from van der Veen (2021). For the Rur at Stah, a hydraulic simulation estimated the instantaneous discharge peak to be around 354 m^3/s while a manual measurement during the flood event led to an estimate of 267 m^3/s but it is unsure whether this was the peak (Horn & Hurkmans, 2022). Hartgring (2023) found a maximum hourly peak discharge of 292 m^3/s based on an hourly simulation using the wflow model for this area and rainfall data from the KMI (RADFLOOD21 dataset). We use the largest value reported of 354 m^3/s as a rough estimate of the hourly discharge peak. Note that we treat the daily estimate as unreliable since we do not have the hourly peak hydrograph for that location. Finally for the Meuse at St Pieter, we summed up the estimates from Meuse at St Pieter Noord and the estimates at the Albert canal (see also section 3.3.1).

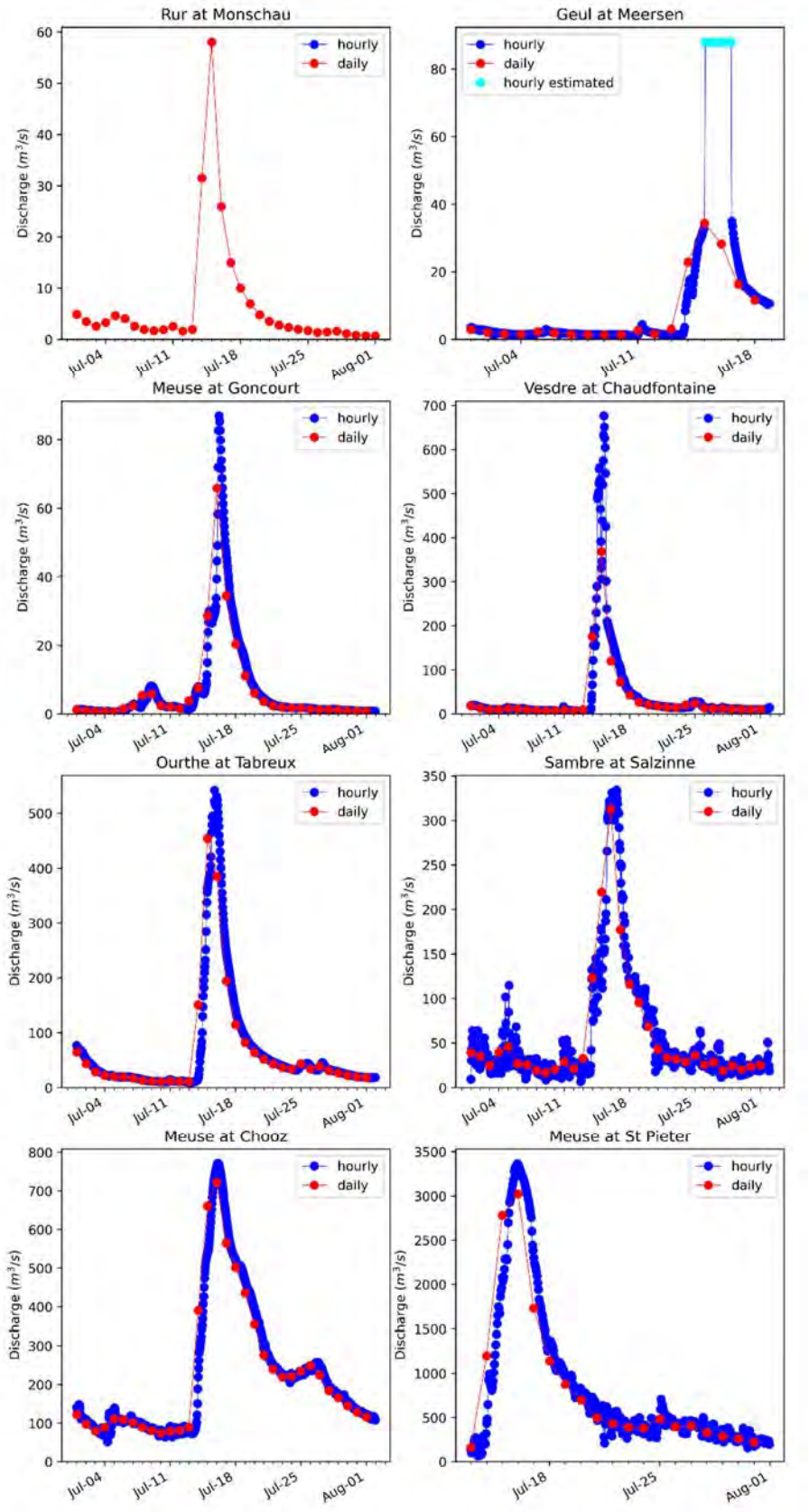


Figure A-2: Recorded or estimated discharge for the selected catchment of interests shown in the main report between July 1st and August 1st 2021.

Deltares
Royal Netherlands Meteorological Institute Ministry of Infrastructure and the Environment
Rijkswaterstaat Ministry of Infrastructure and Water Management
Deltares

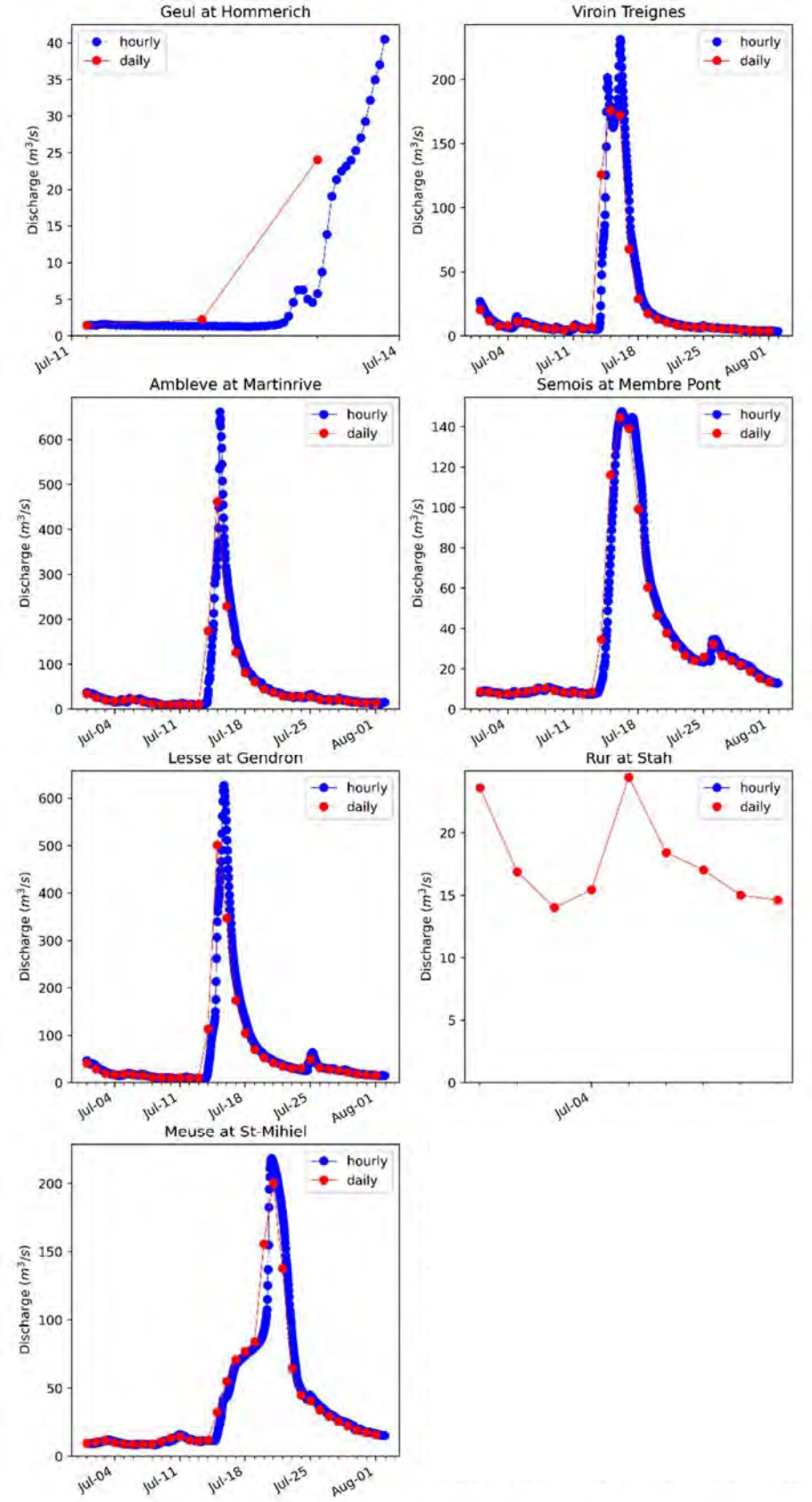


Figure A-3: Recorded or estimated discharge for the selected catchment of interests shown in the appendix, section D, between July 1st and August 1st 2021.

Deltares
Royal Netherlands Meteorological Institute Ministry of Infrastructure and the Environment
Deltares
Rijkswaterstaat Ministry of Infrastructure and Water Management

B Model performance – additional stations

B.1 Daily

In a similar way as in the main report, the model performance is shown for several additional stations in the appendix. For the daily model timestep, these stations are shown in Table B-1 and include the following:

- Geul at Hommerich
- Viroin Treignes
- Ambleve at Martinrive
- Semois at Membre Pont
- Lesse at Gendron
- Rur at Stah
- Meuse at St-Mihiel

The main findings in terms of visual evaluation of model performance and performance indicators are summarised in the tables below.

Table B-1 Summary of daily model performance based on a visual inspection of the signature plots

Station	Catchment area [km2]	Low flows	Average flows (mean monthly and cumulative flows)	High flows
Geul at Hommerich	151	Slight overestimation by the model	Good (slight overestimation by the model of winter monthly flows and cumulative flows)	Good (slight underestimation by the model)
Viroin at Treignes	548	Good (slight underestimatio n by the model)	Underestimation of winter monthly flows and cumulative flows	Underestimation of maximum annual flows
Ambleve at Martinrive	1068	Good	Good	Good but slight underestimation of the max. annual flows with highest return periods
Semois at Membre-Pont	1226	Good	Underestimation of winter mean monthly flows and cumulative flows	Underestimation, especially for higher return periods
Lesse at Gendron	1286	Good	Good	Good (except for an underestimation of the max. annual flows with highest return periods)
Rur at Stah	2152	Overestimation of lowest flows	Overestimation of mean monthly flows and of cumulative flows	Good
Meuse at St-Mihiel	2551	Good	Good, slight underestimation of winter monthly flows and cumulative flows	Good

Table B-2 Summary of daily model performance based on the scores of the performance indicators

	NSE	KGE	NSElog	NM7Q	MAXQ
Geul at Hommerich	0.7	0.81	0.67	0.14	0.8
Viroin at Treignes	0.77	0.67	0.84	0.8	0.37
Ambleve at Martinrive	0.79	0.85	0.85	0.96	0.81
Semois at Membre-Pont	0.8	0.63	0.88	0.84	0.48
Lesse at Gendron	0.83	0.82	0.86	0.9	0.78
Stah	0.61	0.73	0.53	-15.9	0.87
Meuse at Saint-Mihiel	0.81	0.82	0.85	0.89	0.9

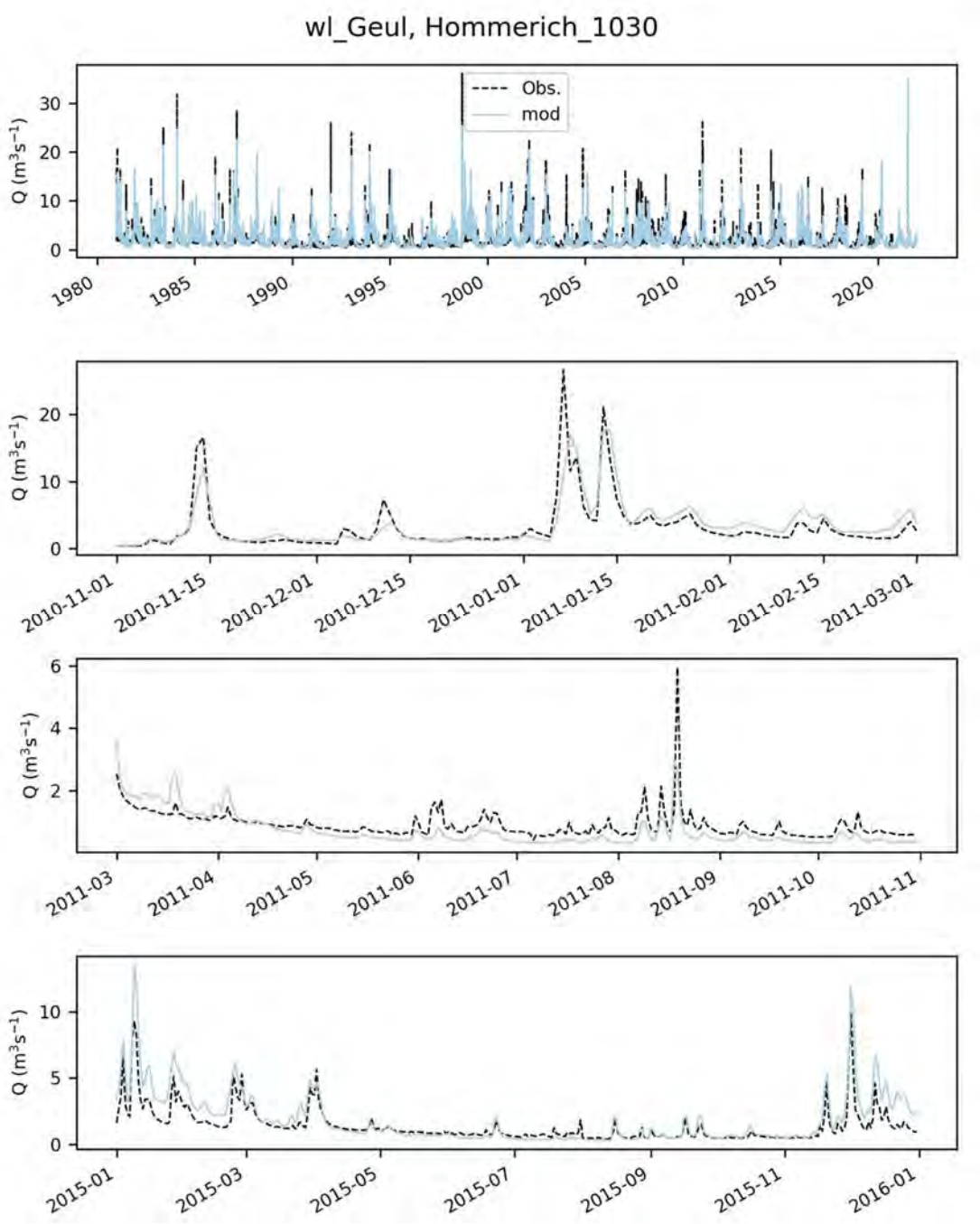


Figure B-1 Modelled (blue) and observed (black) hydrographs for the Geul at Hommerich

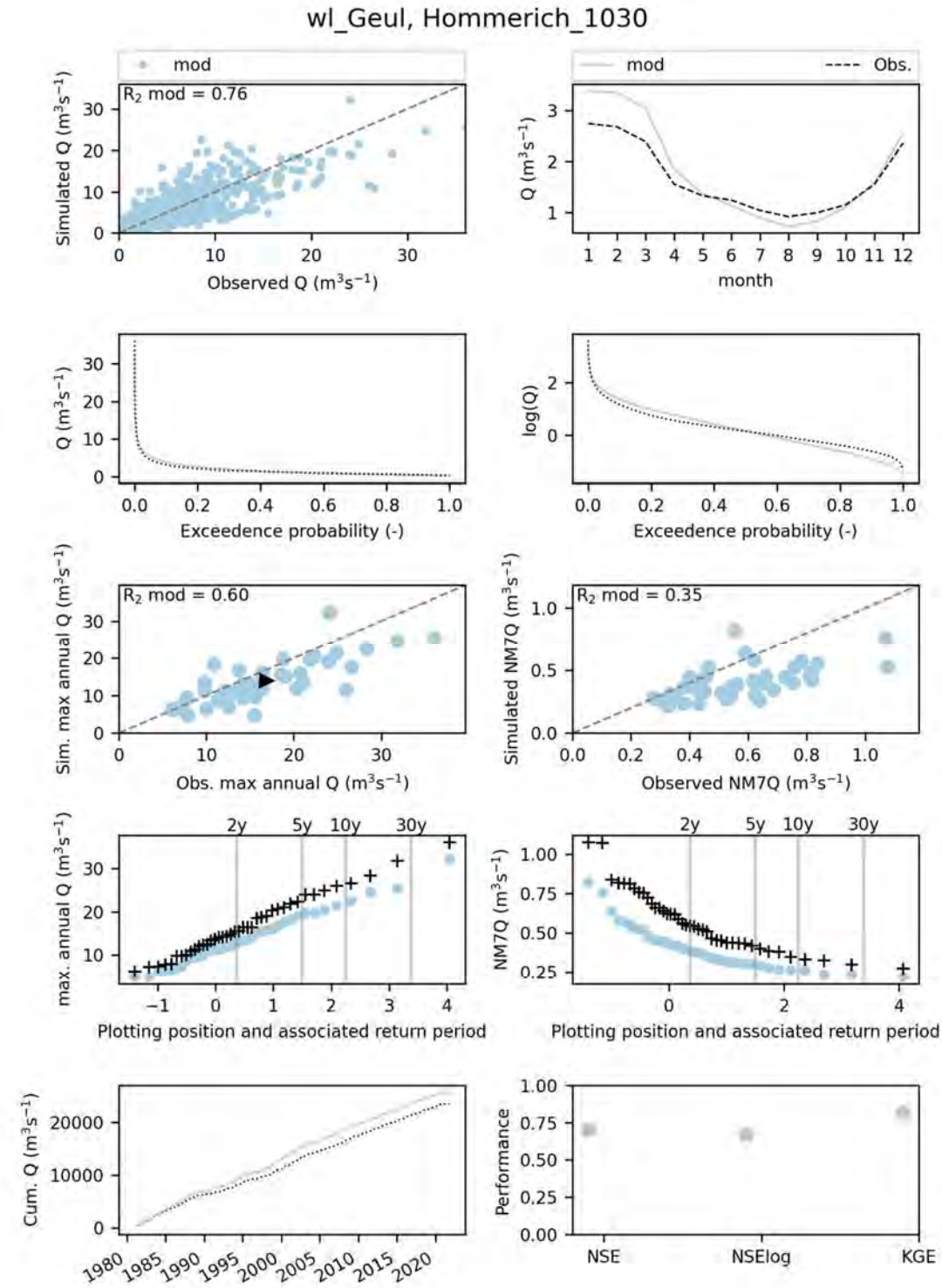


Figure B-2 Modelled (blue) and observed (black) signatures for the Geul at Hommerich

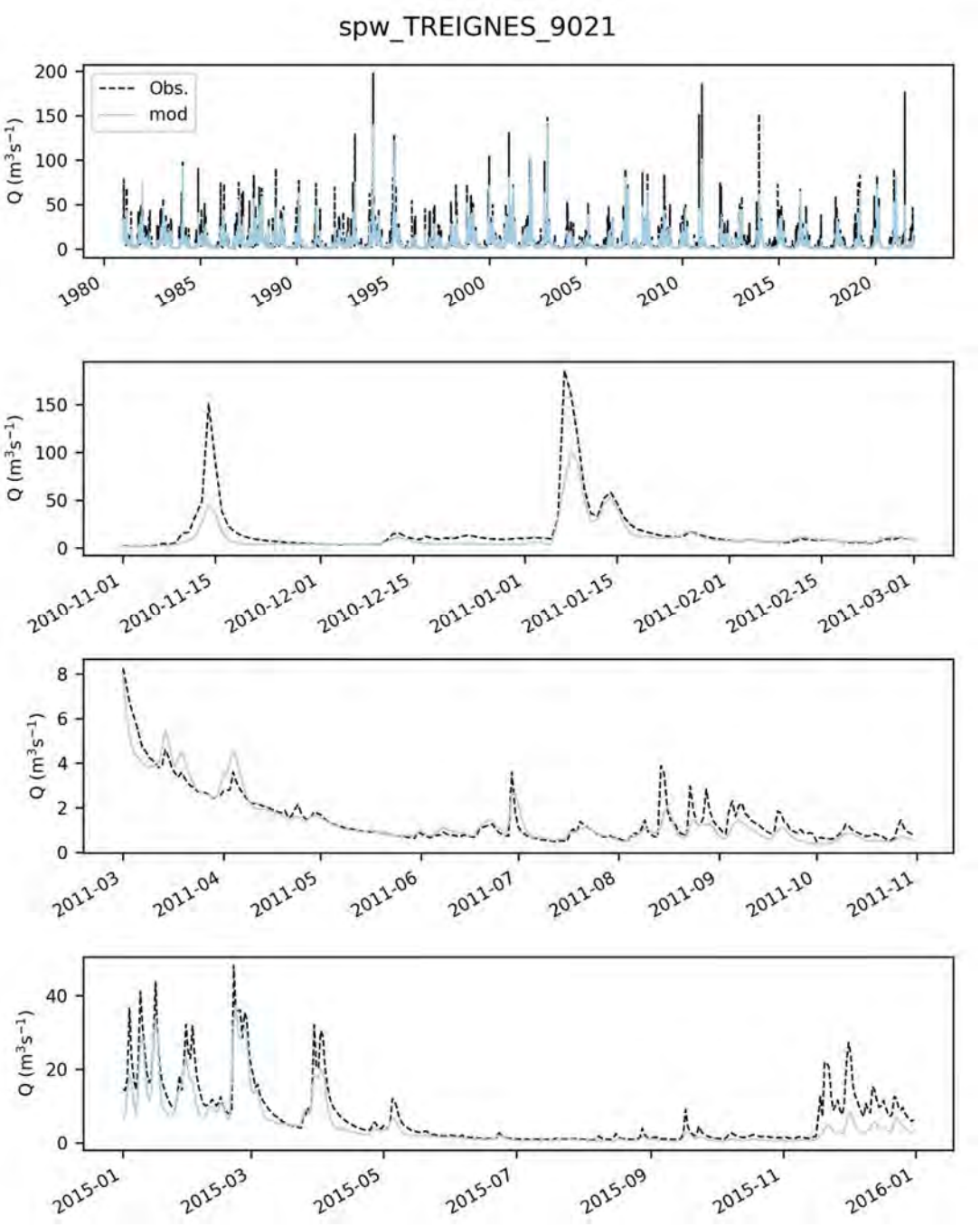


Figure B-3 Modelled (blue) and observed (black) hydrographs for the Viroin at Treignes

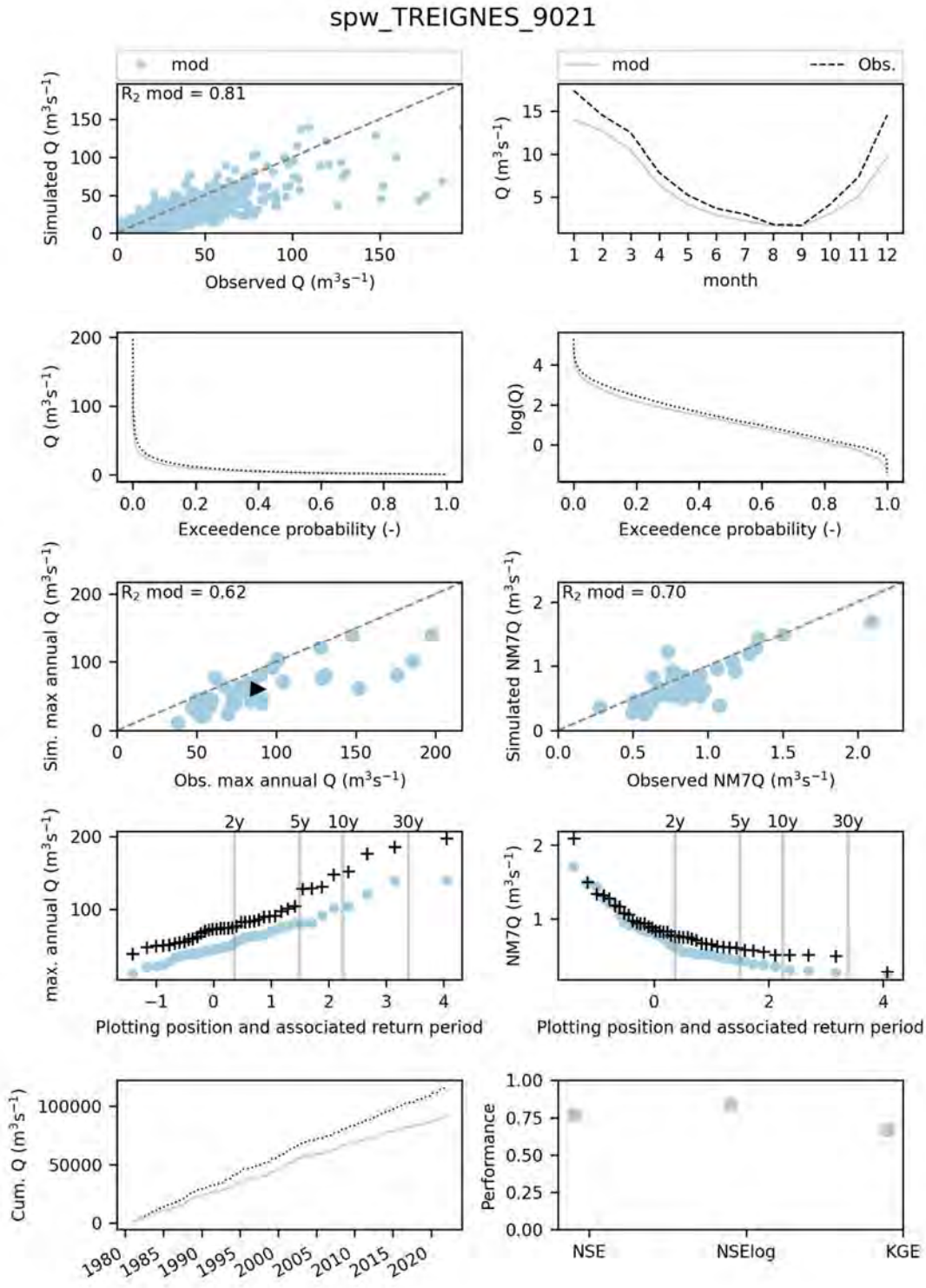


Figure B-4 Modelled (blue) and observed (black) signatures for the Viroin at Treignes

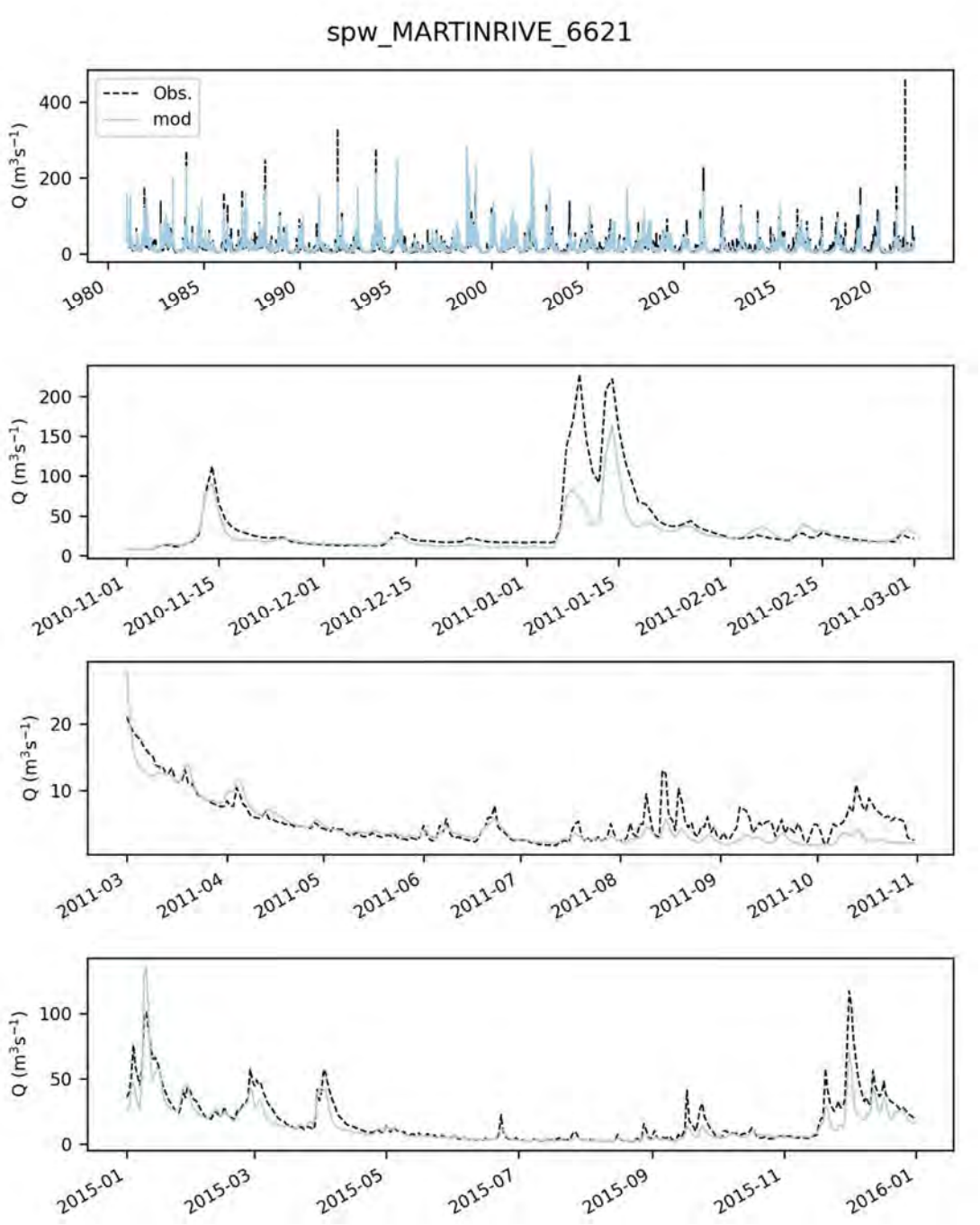


Figure B-5 Modelled (blue) and observed (black) hydrographs for the Ambleve at Martinrive

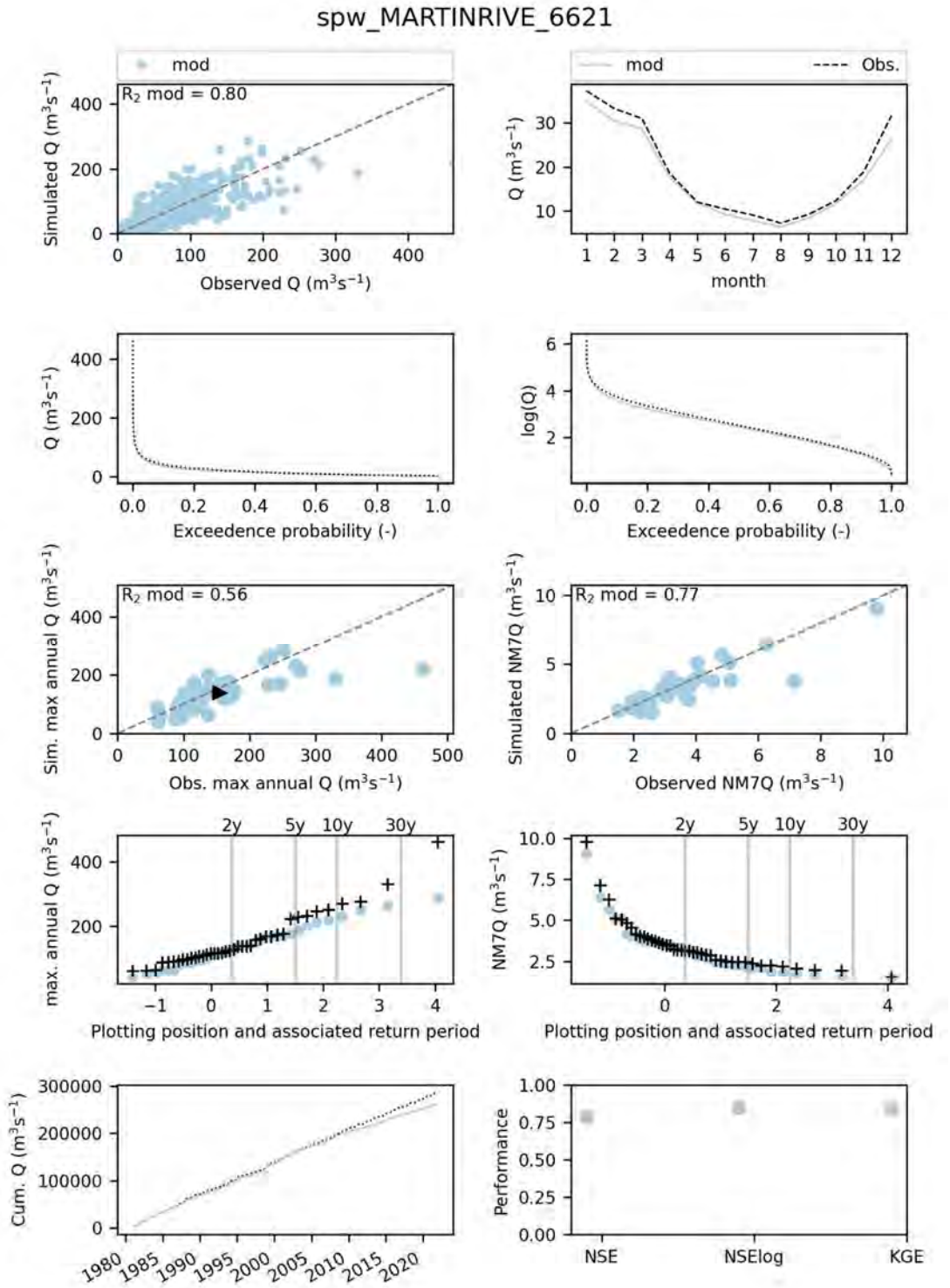


Figure B-6 Modelled (blue) and observed (black) signatures for the Ambleve at Martinrive

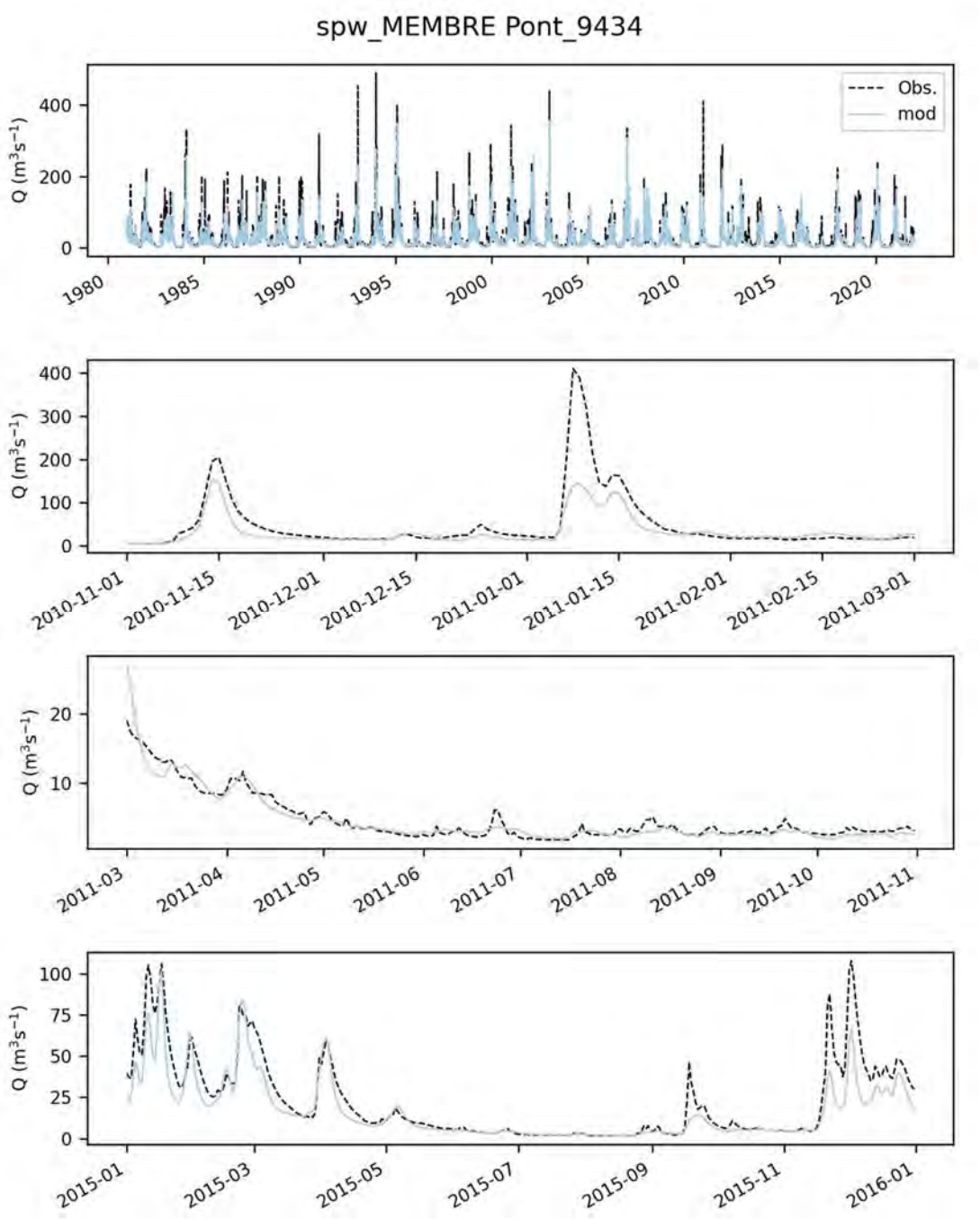


Figure B-7 Modelled (blue) and observed (black) hydrographs for the Semois at Membre-Pont

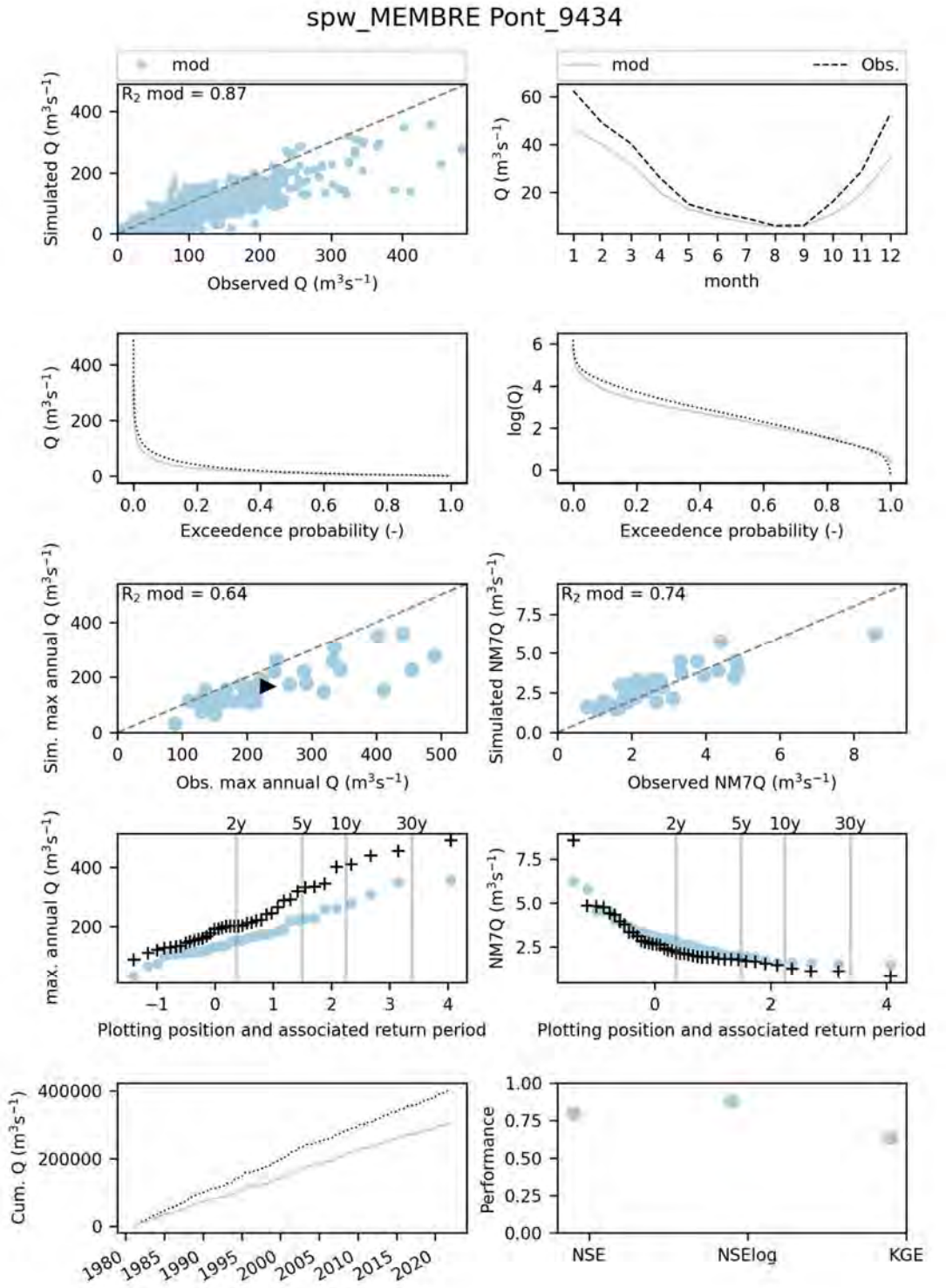


Figure B-8 Modelled (blue) and observed (black) signatures for the Semois at Membre-Pont

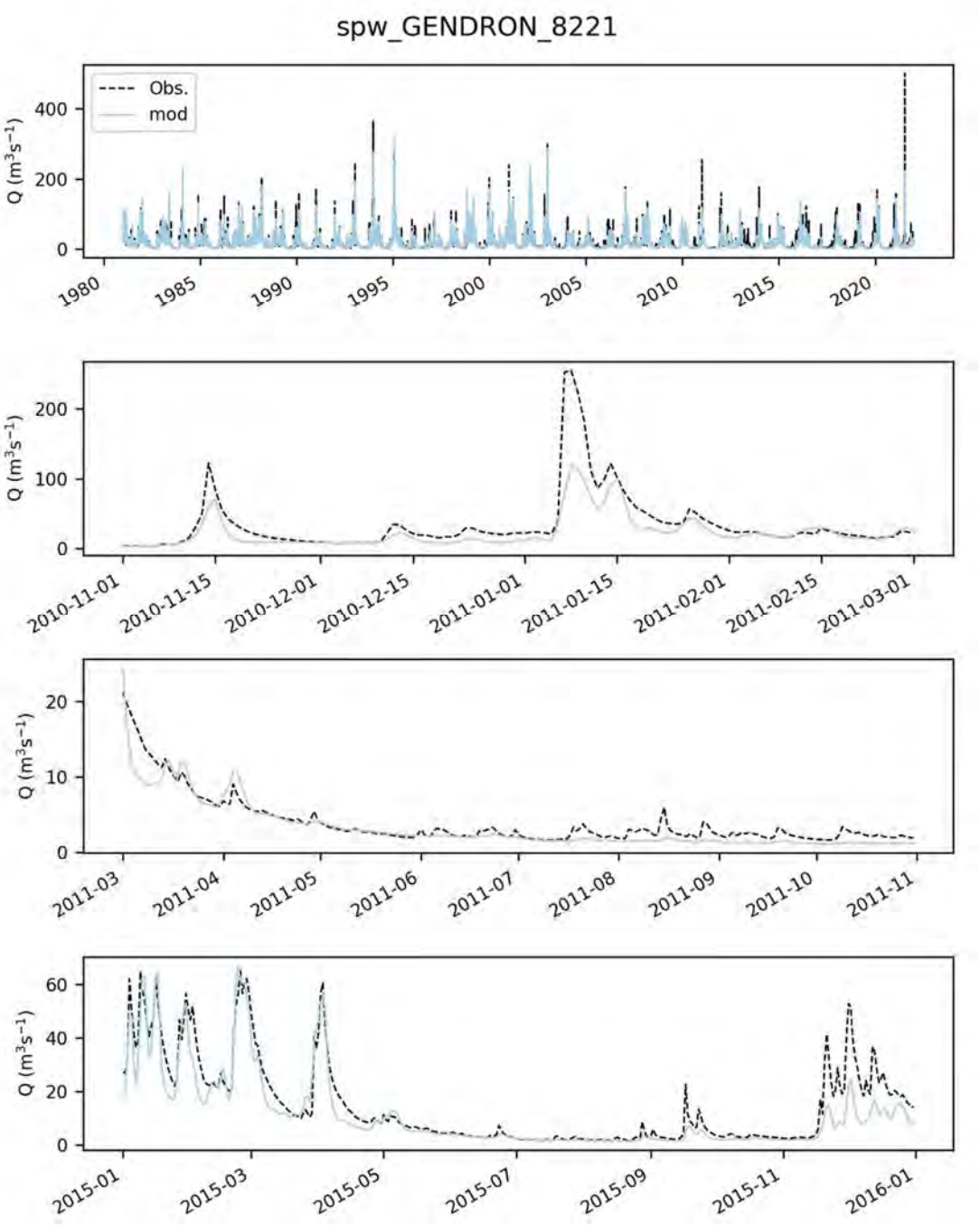


Figure B-9 Modelled (blue) and observed (black) hydrographs for the Lesse at Gendron

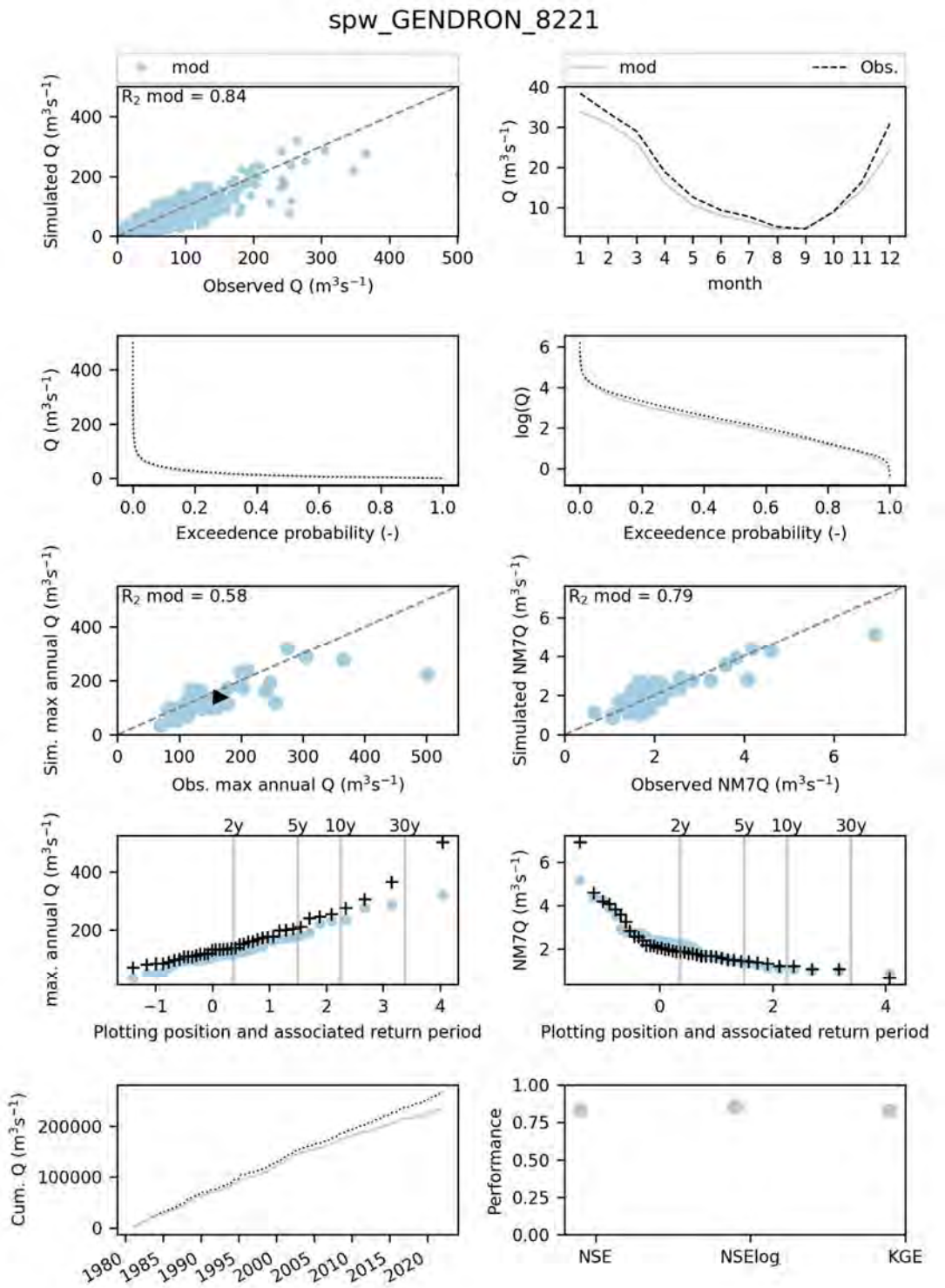


Figure B-10 Modelled (blue) and observed (black) signatures for the Lesse at Gendron

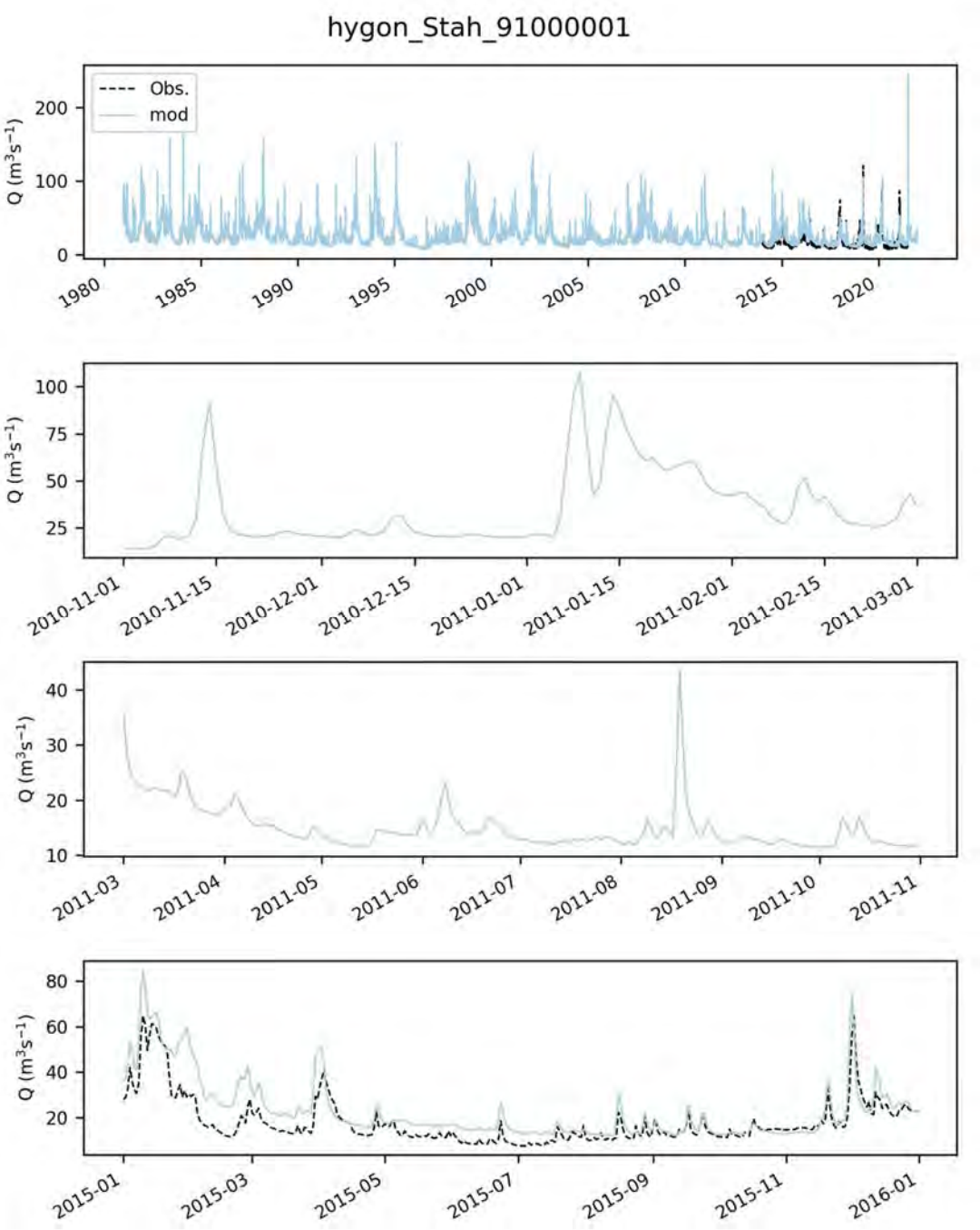


Figure B-11 Modelled (blue) and observed (black) hydrographs for the Rur at Stah

Deltares
Royal Netherlands Meteorological Institute Ministry of Infrastructure and the Environment
Rijkswaterstaat Ministry of Infrastructure and Water Management
Deltares

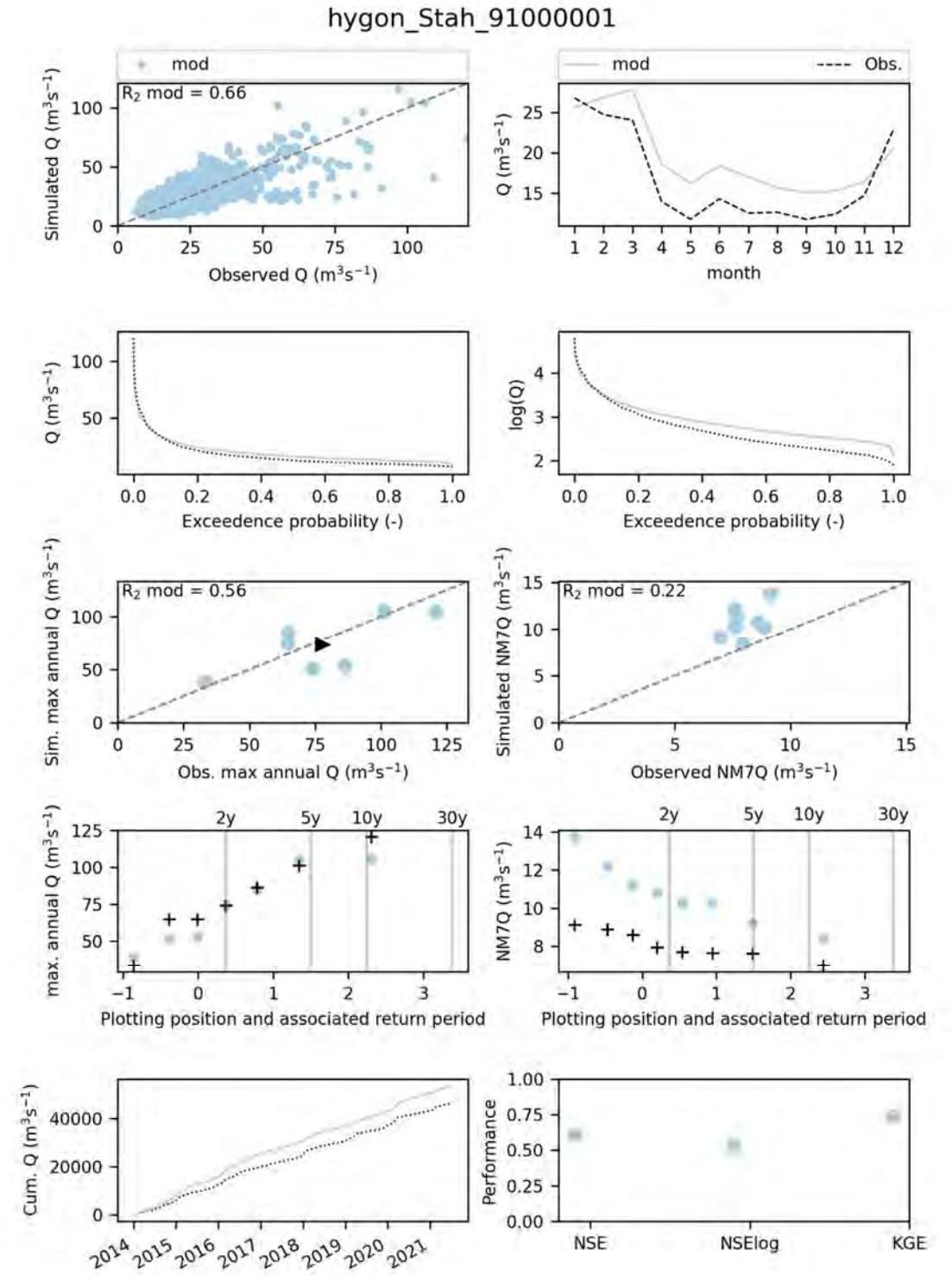


Figure B-12 Modelled (blue) and observed (black) signatures for the Rur at Stah

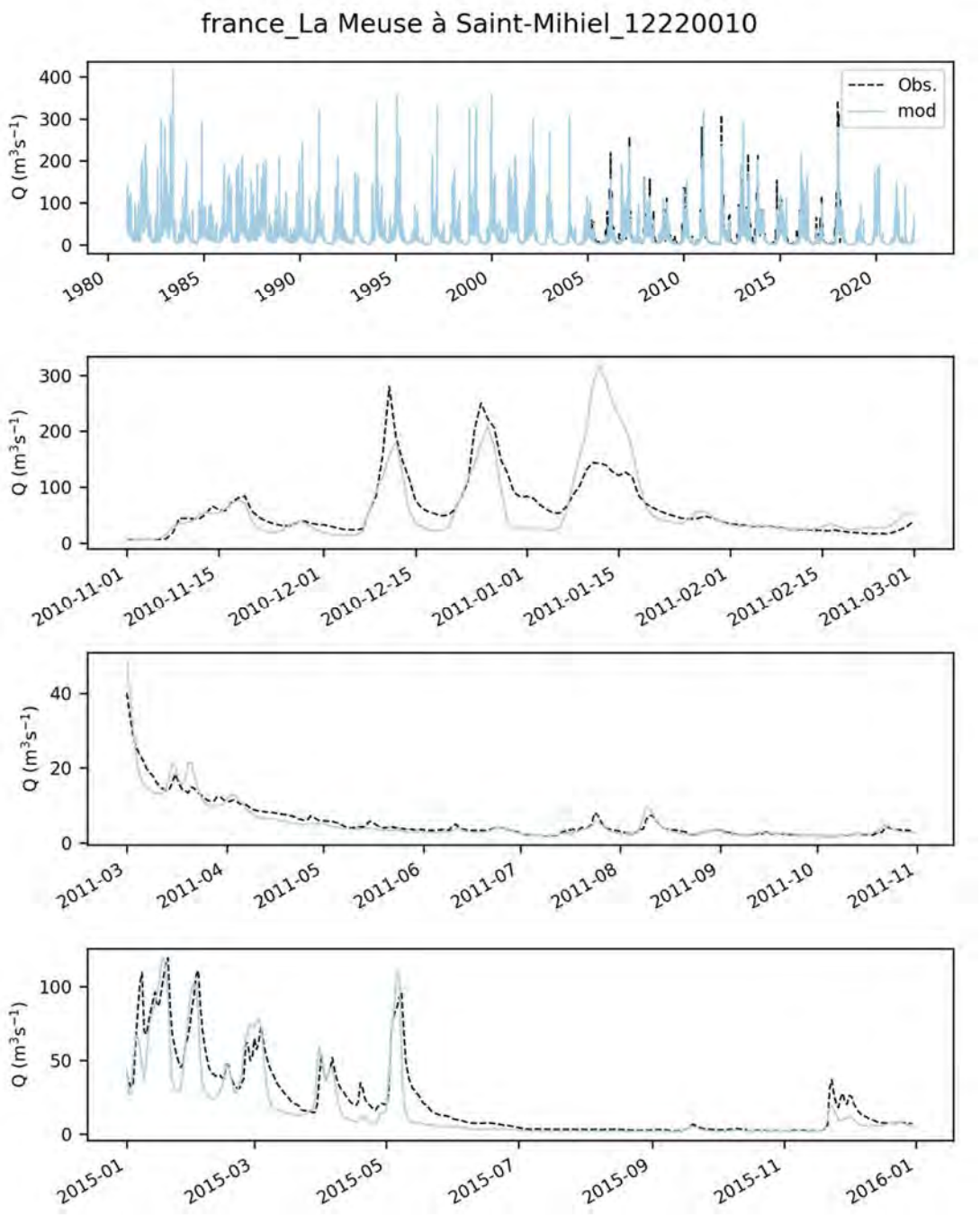


Figure B-13 Modelled (blue) and observed (black) hydrographs for the Meuse at St Mihiel

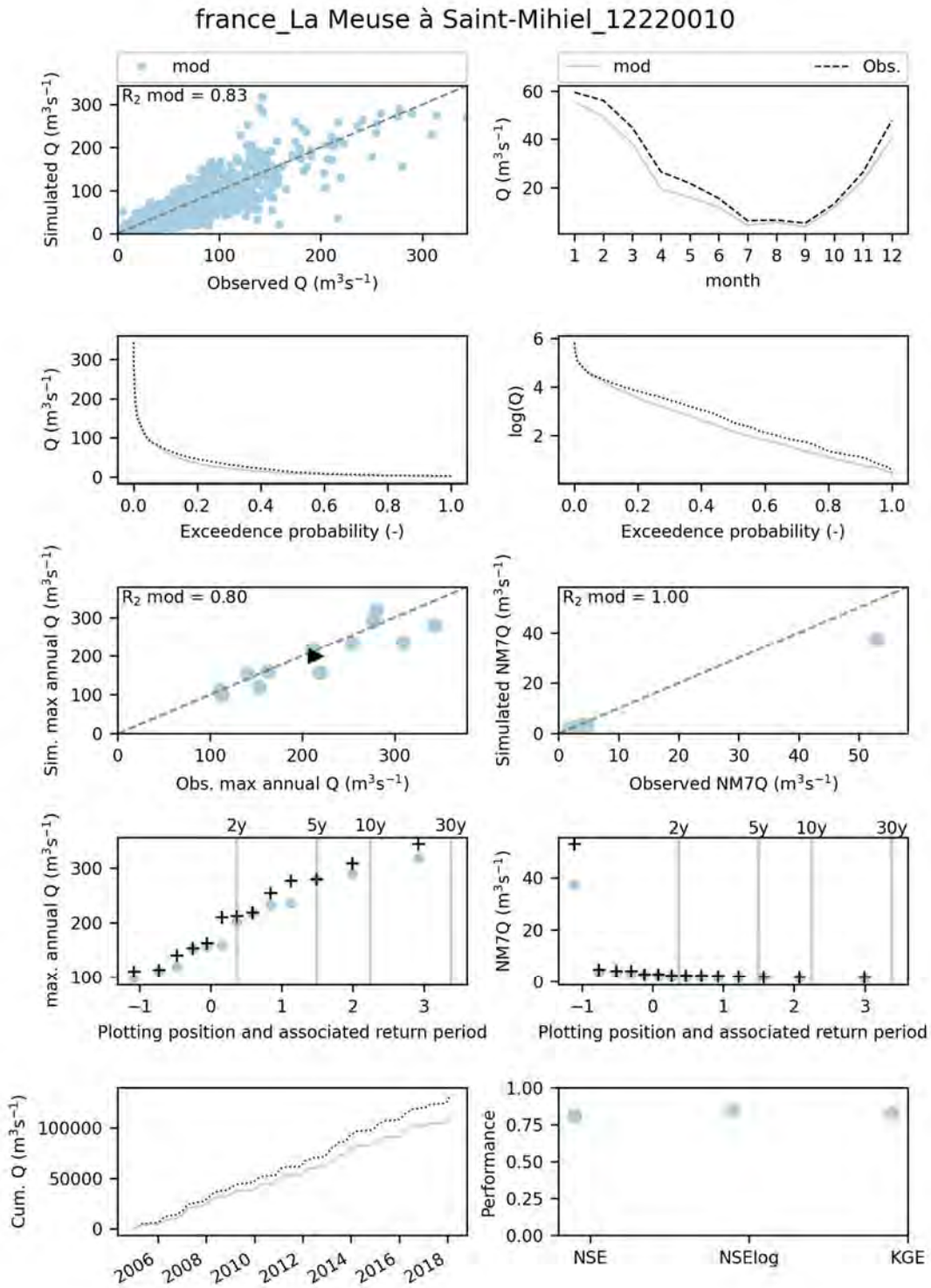


Figure B-14 Modelled (blue) and observed (black) signatures for the Meuse at St Mihiel

B.2 Hourly

For the hourly timestep model, the model performance if shown for the following additional stations:

- Geul at Hommerich
- Viroin Treigne
- Ambleve at Martinrive
- Semois at Membre Pont
- Lesse at Gendron
- Meuse at St-Mihiel

The main findings in terms of visual evaluation of model performance and performance indicators are summarised in the tables below.

Table B-3 Summary of hourly model performance based on a visual inspection of the signature plots

Station	Catchment area [km2]	Low flows	Average flows (mean monthly and cumulative flows)	High flows
Geul at Hommerich	151	Underestimation by the model	Good (slight overestimation of monthly flows between Jan-May and of the cumulative flows)	Underestimation by the model
Viroin at Treigne	548	Good but underestimation of min. annual flows with highest return periods	Slight underestimation of winter monthly flows and cumulative flows	Good but underestimation of max. annual flows with highest return periods
Ambleve at Martinrive	1068	Good	Slight underestimation of mean monthly flows and cumulative flows	Underestimation of max. annual flows with lowest return period and vice versa for higher return periods
Semois at Membre-Pont	1226	Overestimation of min. annual flows	Good despite slight underestimation of mean monthly flows and cumulative flows	Good
Lesse at Gendron	1286	Good	Good, slight underestimation by the model	Good
Meuse at St-Mihiel	2551	Good	Underestimation of mean monthly flows and cumulative flows	Underestimion of max. annual flows

Table B-4 Summary of hourly model performance based on the scores of performance indicators

	NSE	KGE	NSElog	NM7Q	MAXQ
Geul at Hommerich	0.56	0.74	0.7	0.13	-0.8
Viroin at Treignes	0.75	0.68	0.82	0.38	0.38
Ambleve at Martinrive	0.74	0.74	0.81	0.88	0.03
Semois at Membre-Pont	0.85	0.79	0.88	0.41	0.75
Lesse at Gendron	0.75	0.79	0.85	0.79	0.75
Meuse at Saint-Mihiel	0.79	0.76	0.86	0.74	0.45

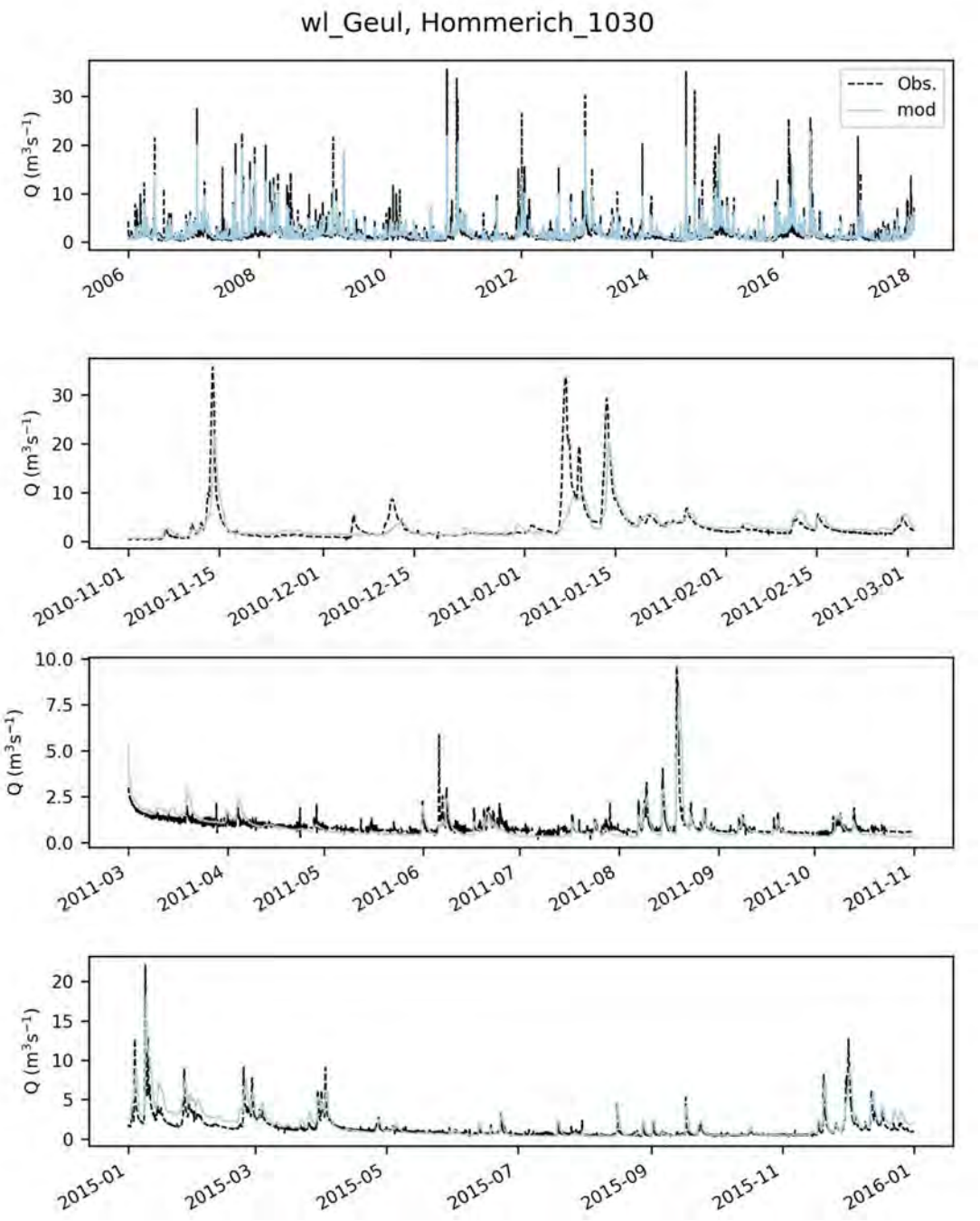


Figure B-15 Modelled (blue) and observed (black) hydrographs for the Geul at Hommerich

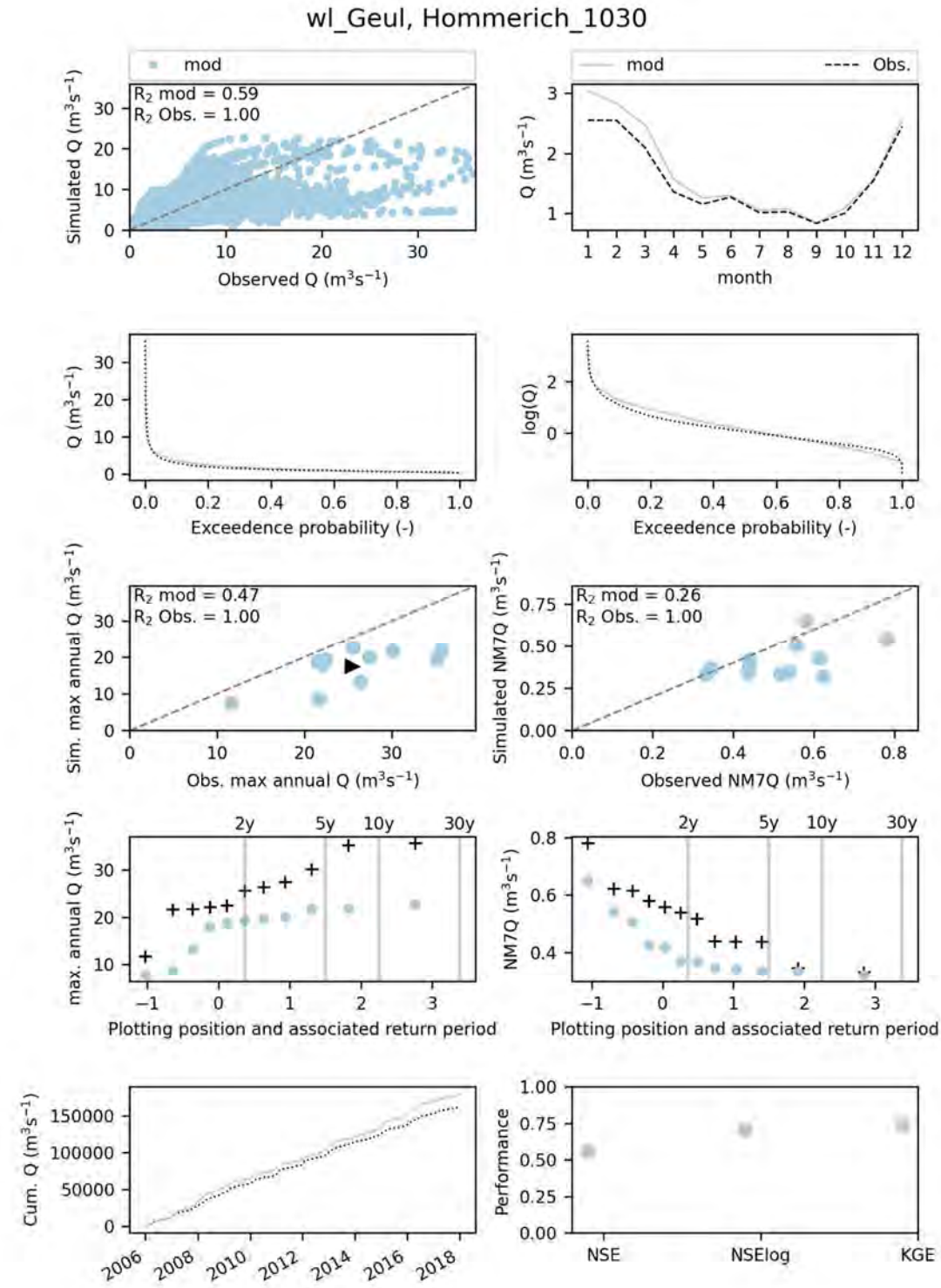


Figure B-16 Modelled (blue) and observed (black) signatures for the Geul at Hommerich

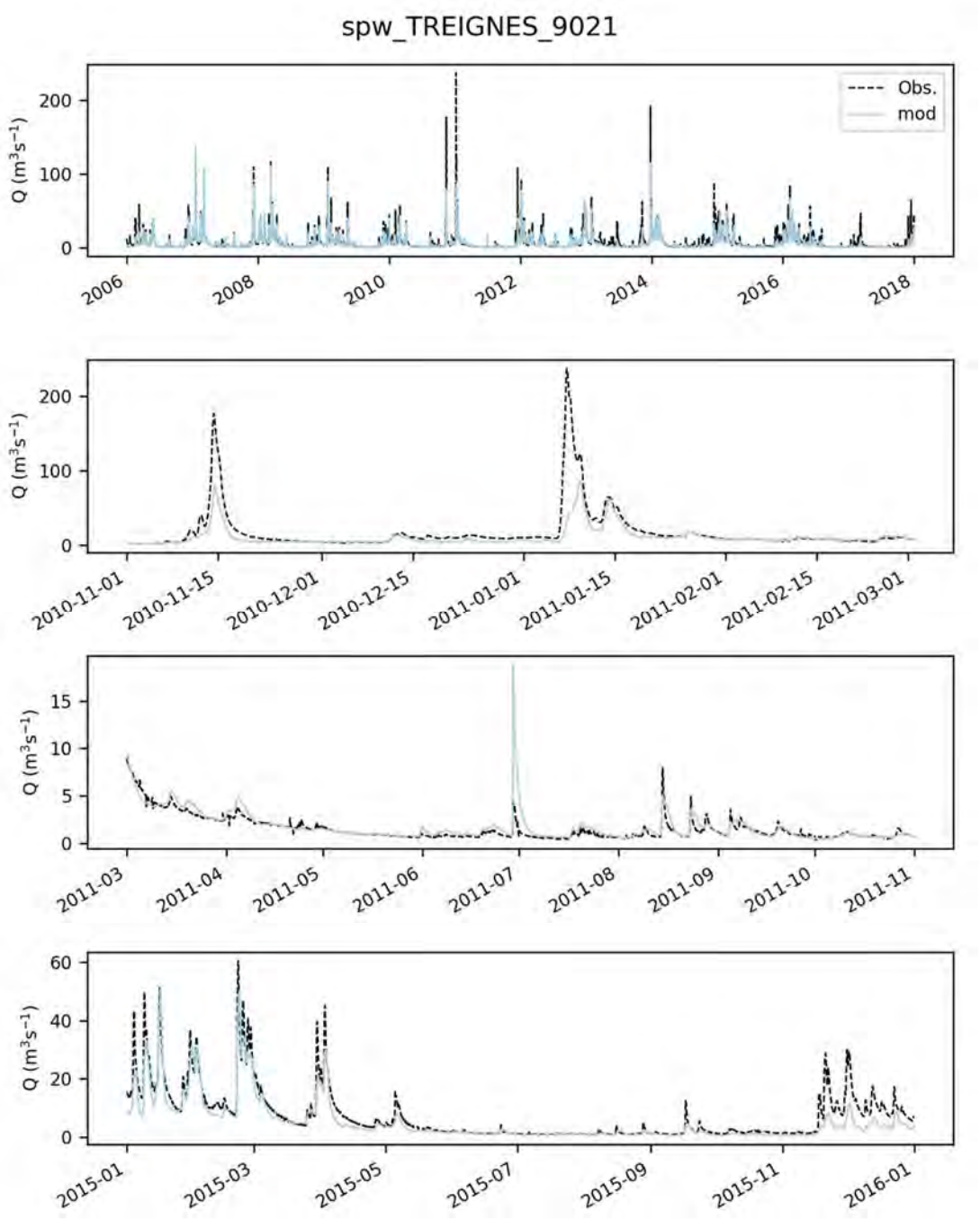


Figure B-17 Modelled (blue) and observed (black) hydrographs for the Viroin at Treigne

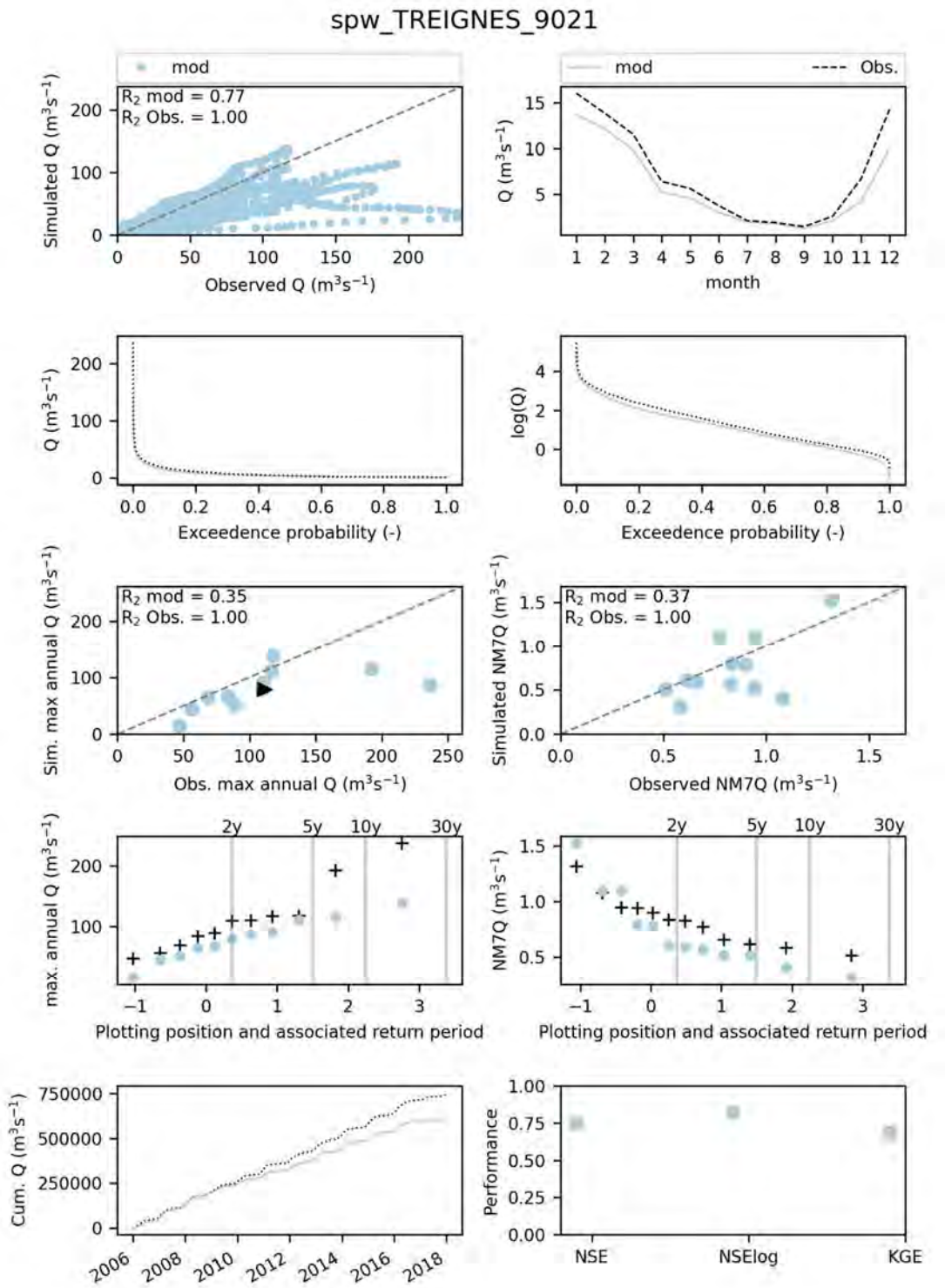


Figure B-18 Modelled (blue) and observed (black) signatures for the Viroin at Treigne

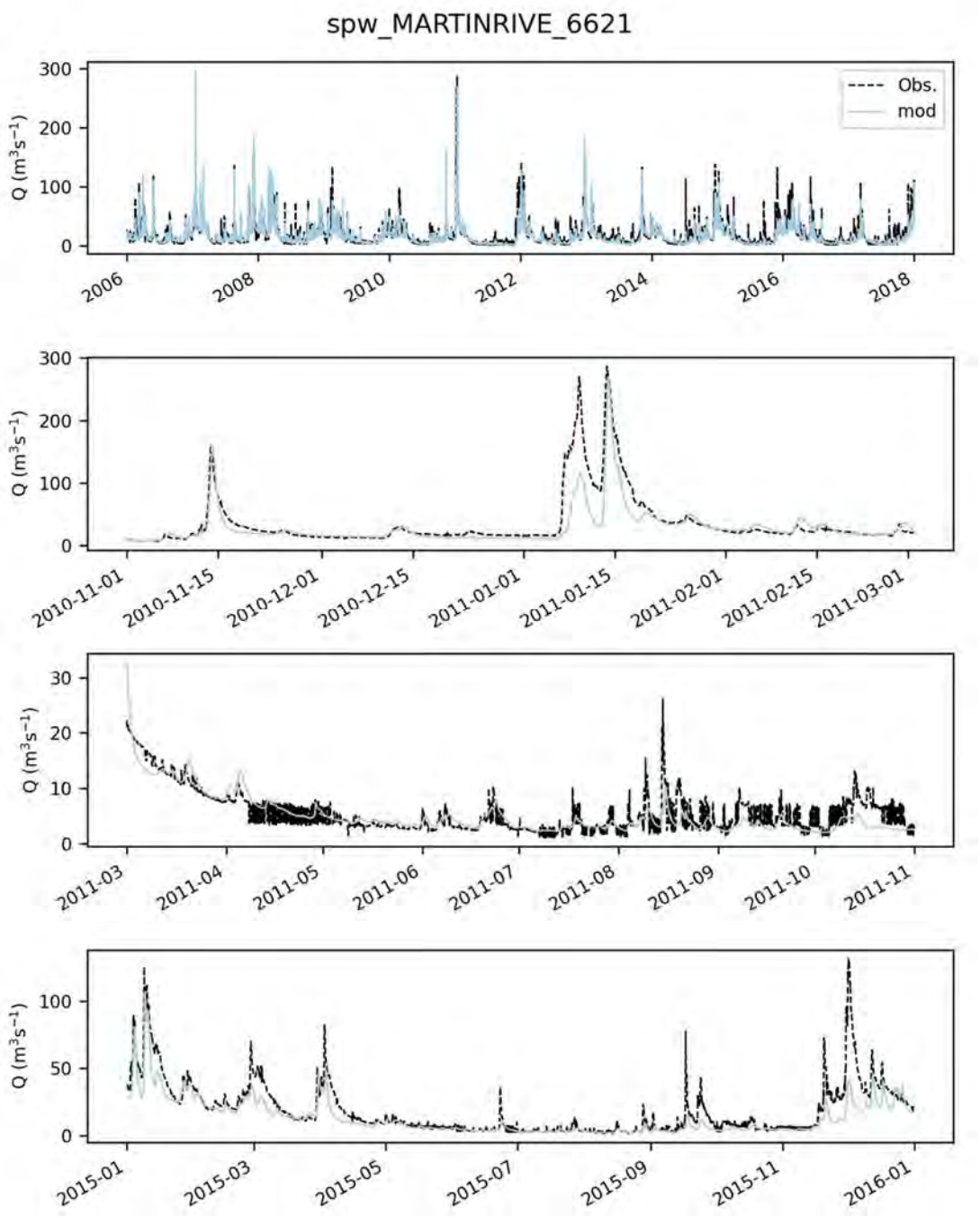


Figure B-19 Modelled (blue) and observed (black) hydrographs for the Ambleve at Martinrive

Deltares
Royal Netherlands Meteorological Institute Ministry of Infrastructure and the Environment
Rijkswaterstaat Ministry of Infrastructure and Water Management
Deltares

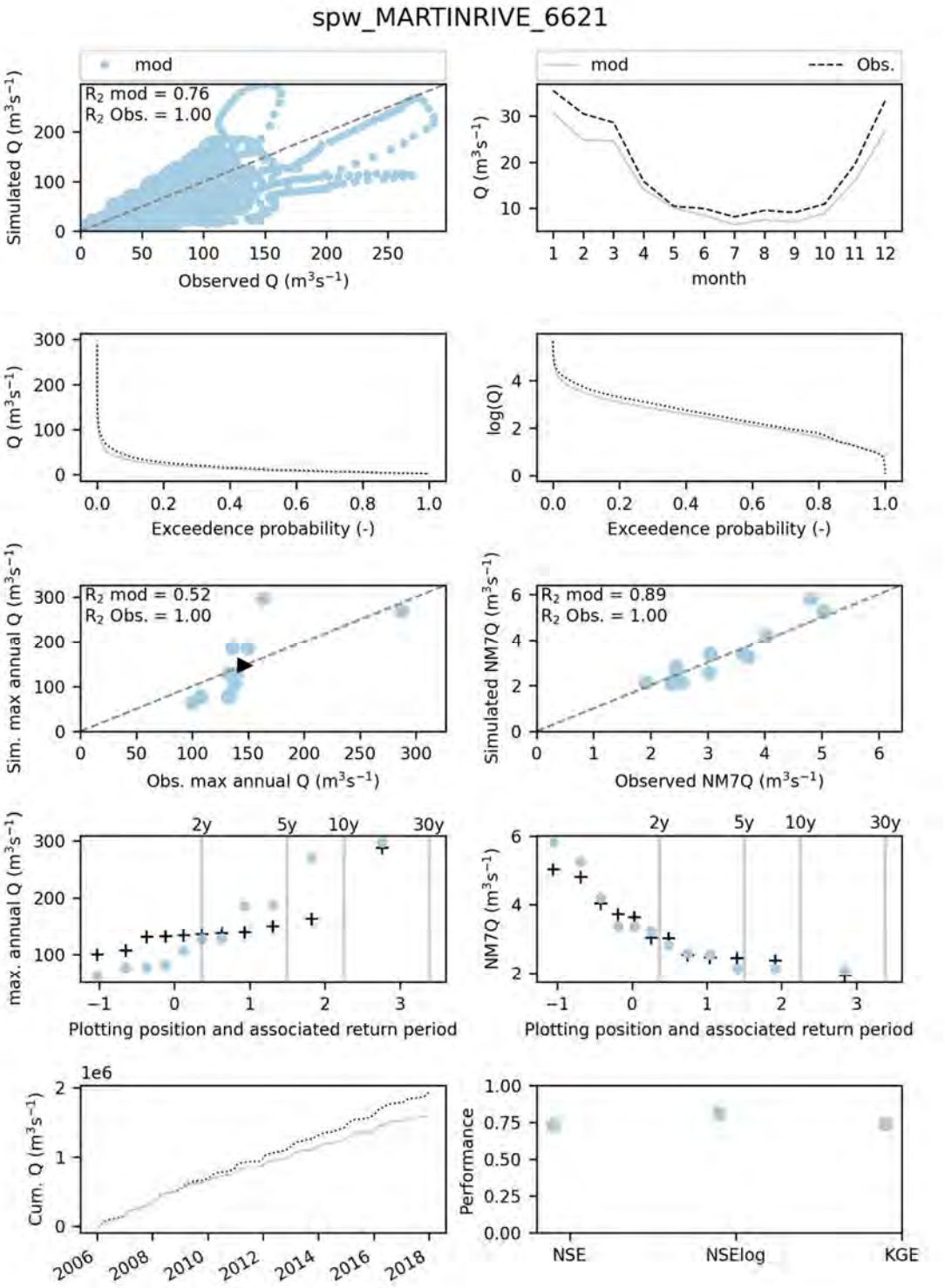


Figure B-20 Modelled (blue) and observed (black) signatures for the Ambleve at Martinrive

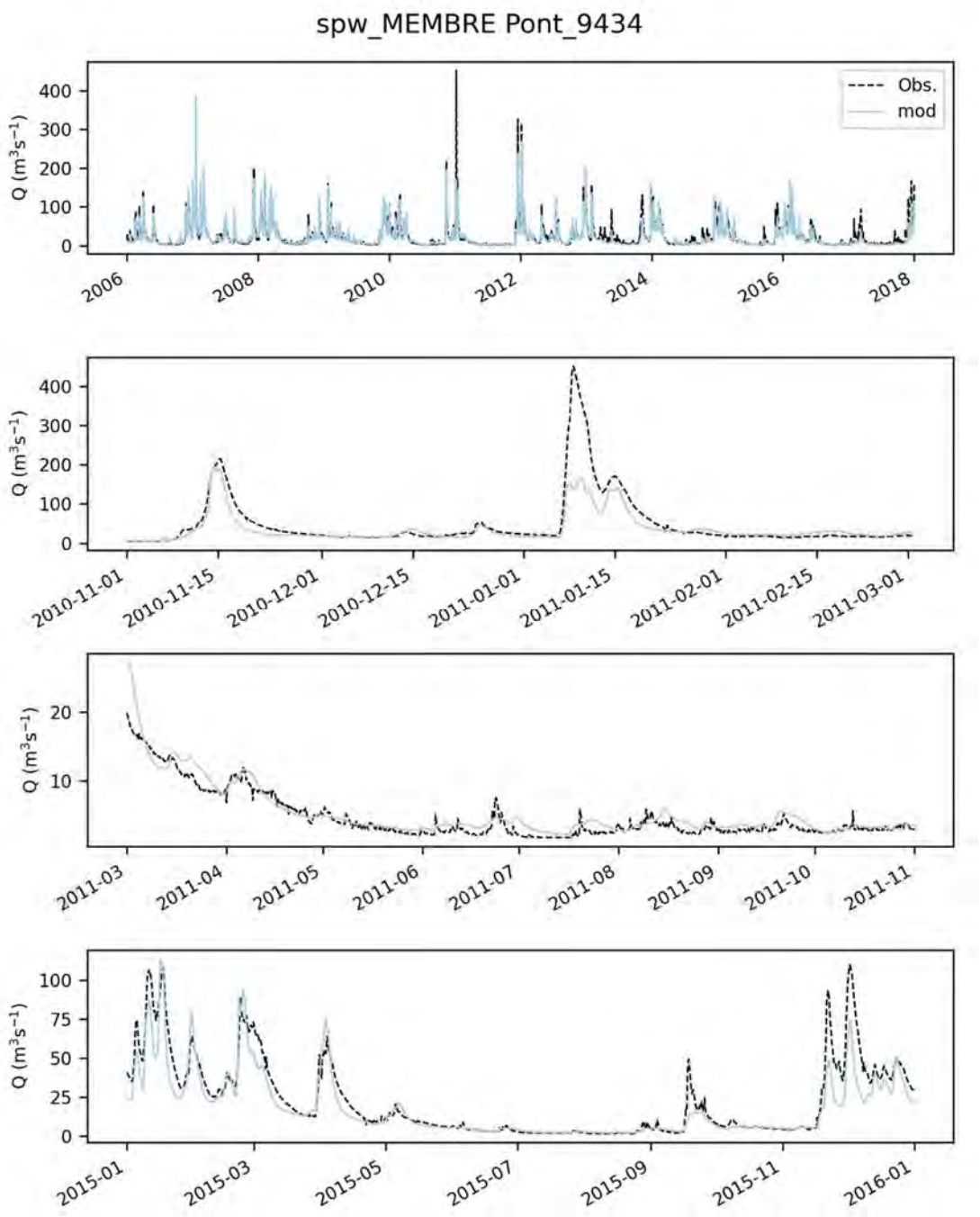


Figure B-21 Modelled (blue) and observed (black) hydrographs for the Semois at Membre-Pont

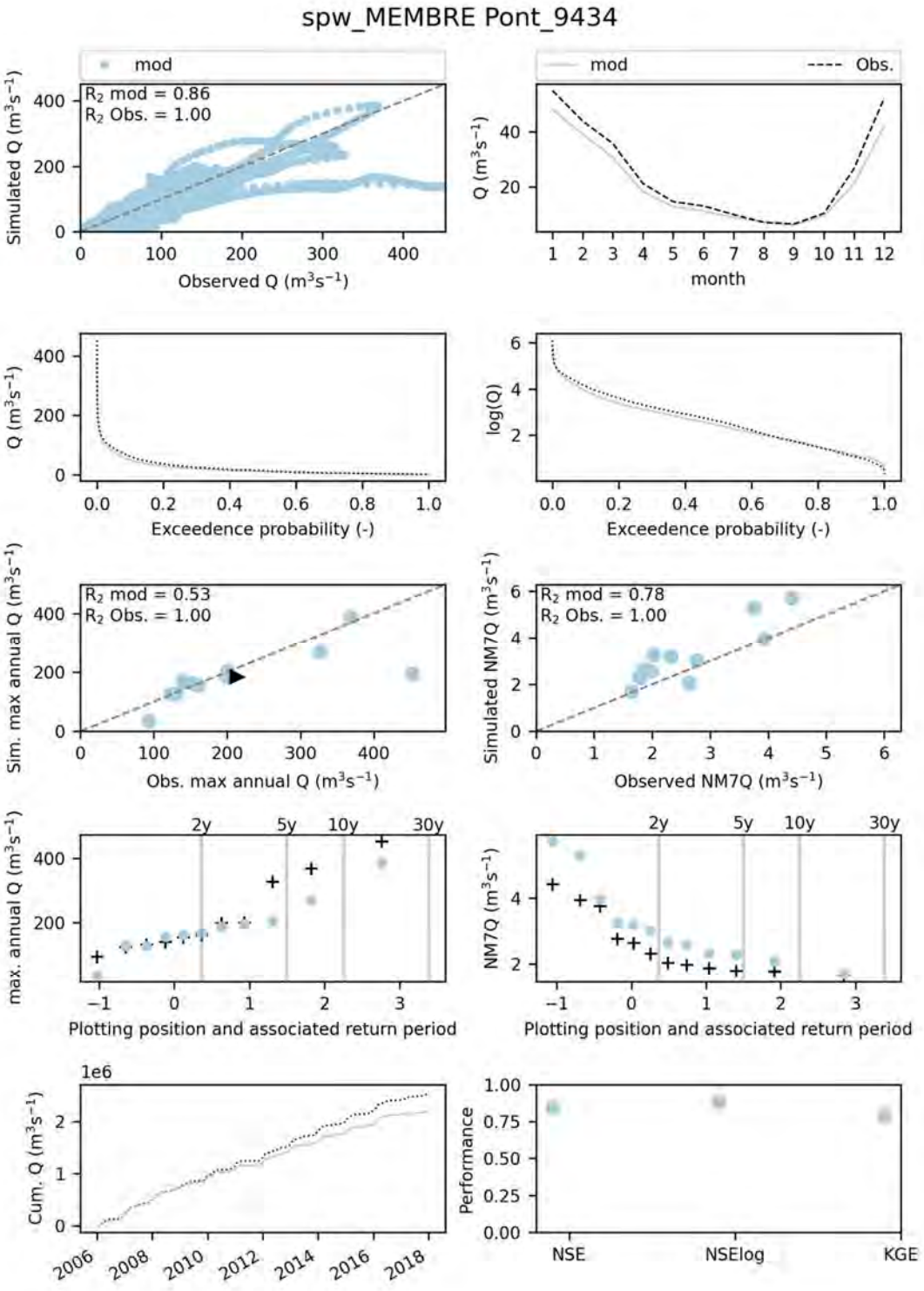


Figure B-22 Modelled (blue) and observed (black) signatures for the Semois at Membre-Pont

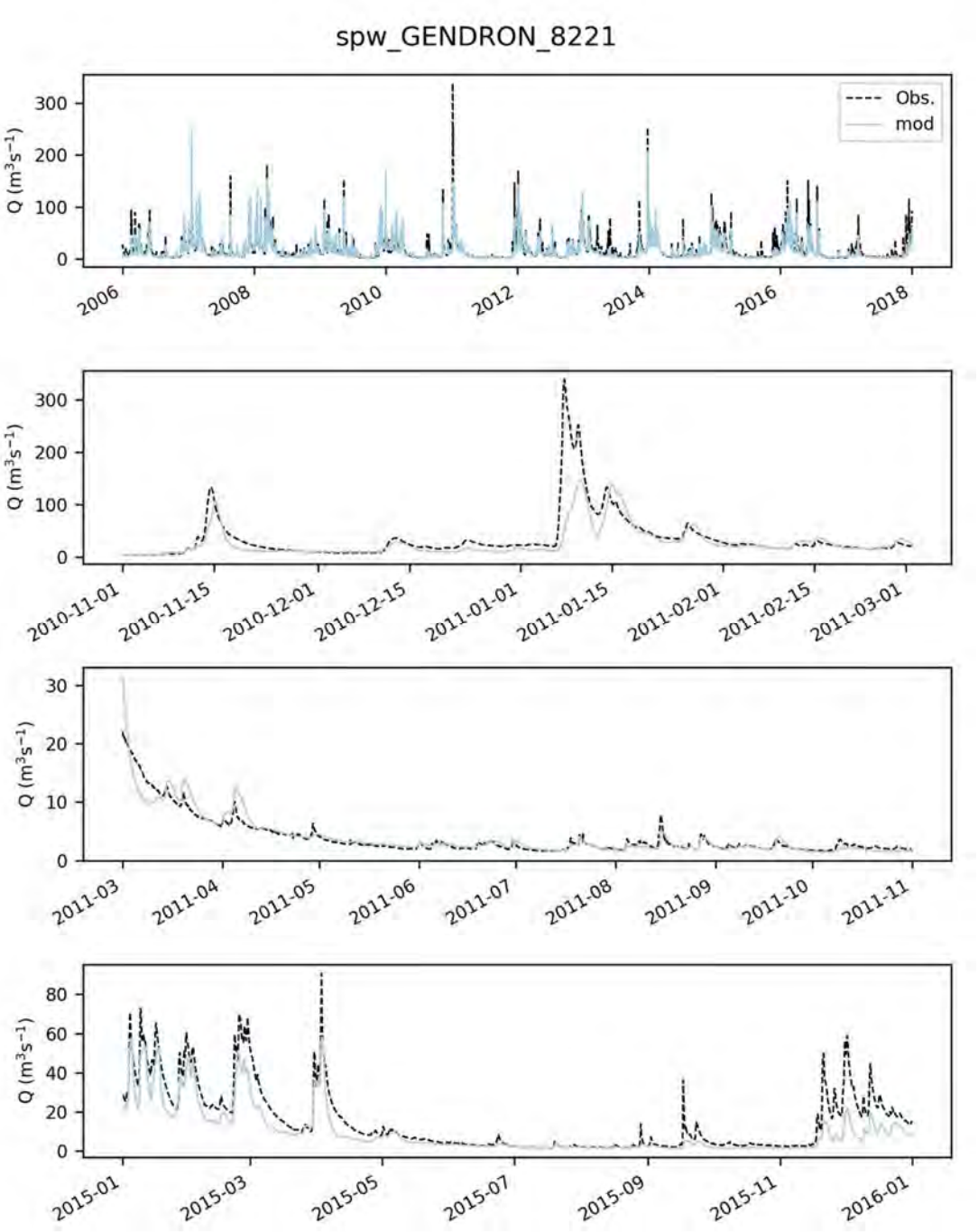


Figure B-23 Modelled (blue) and observed (black) hydrographs for the Lesse at Gendron

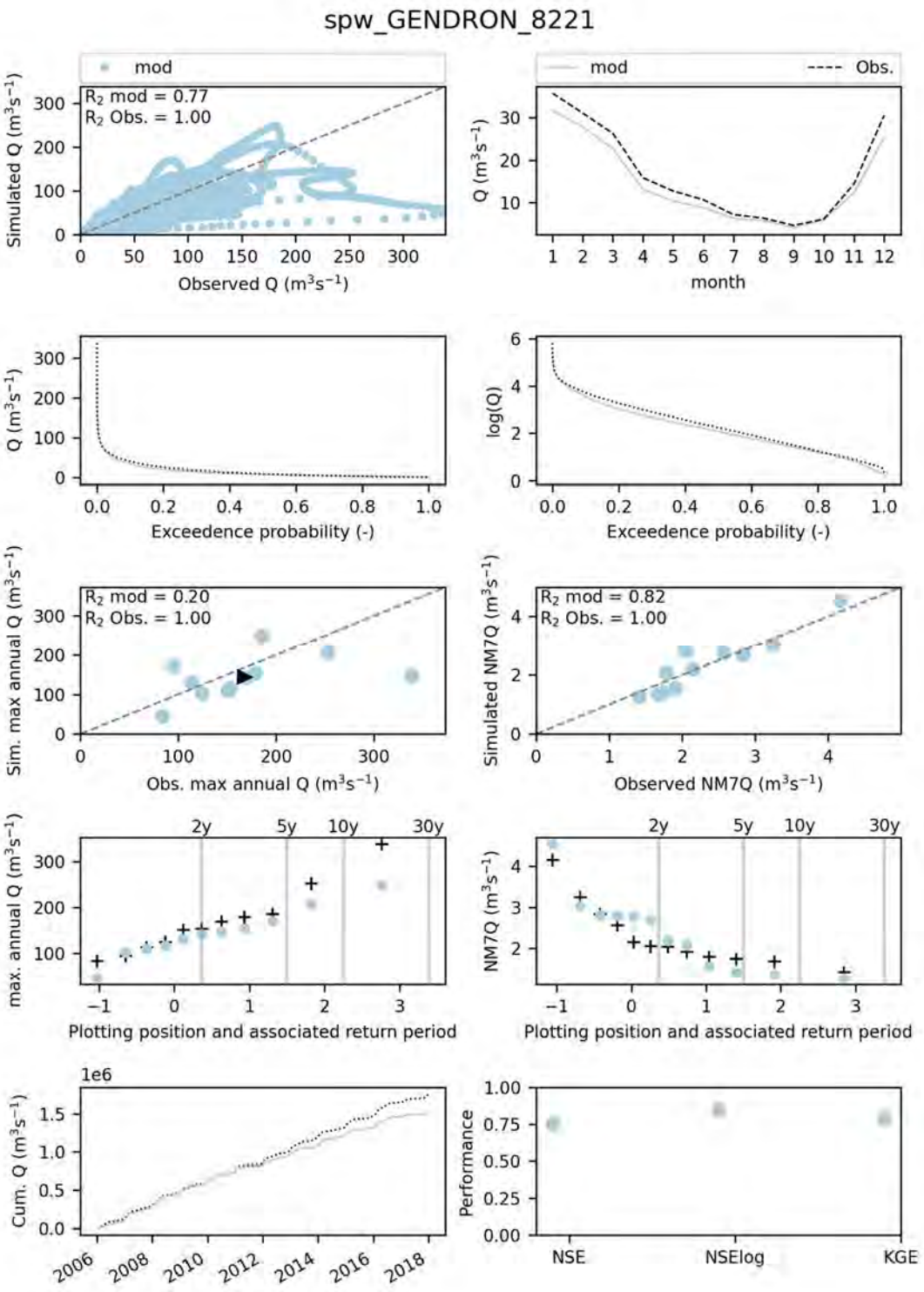


Figure B-24 Modelled (blue) and observed (black) signatures for the Lesse at Gendron

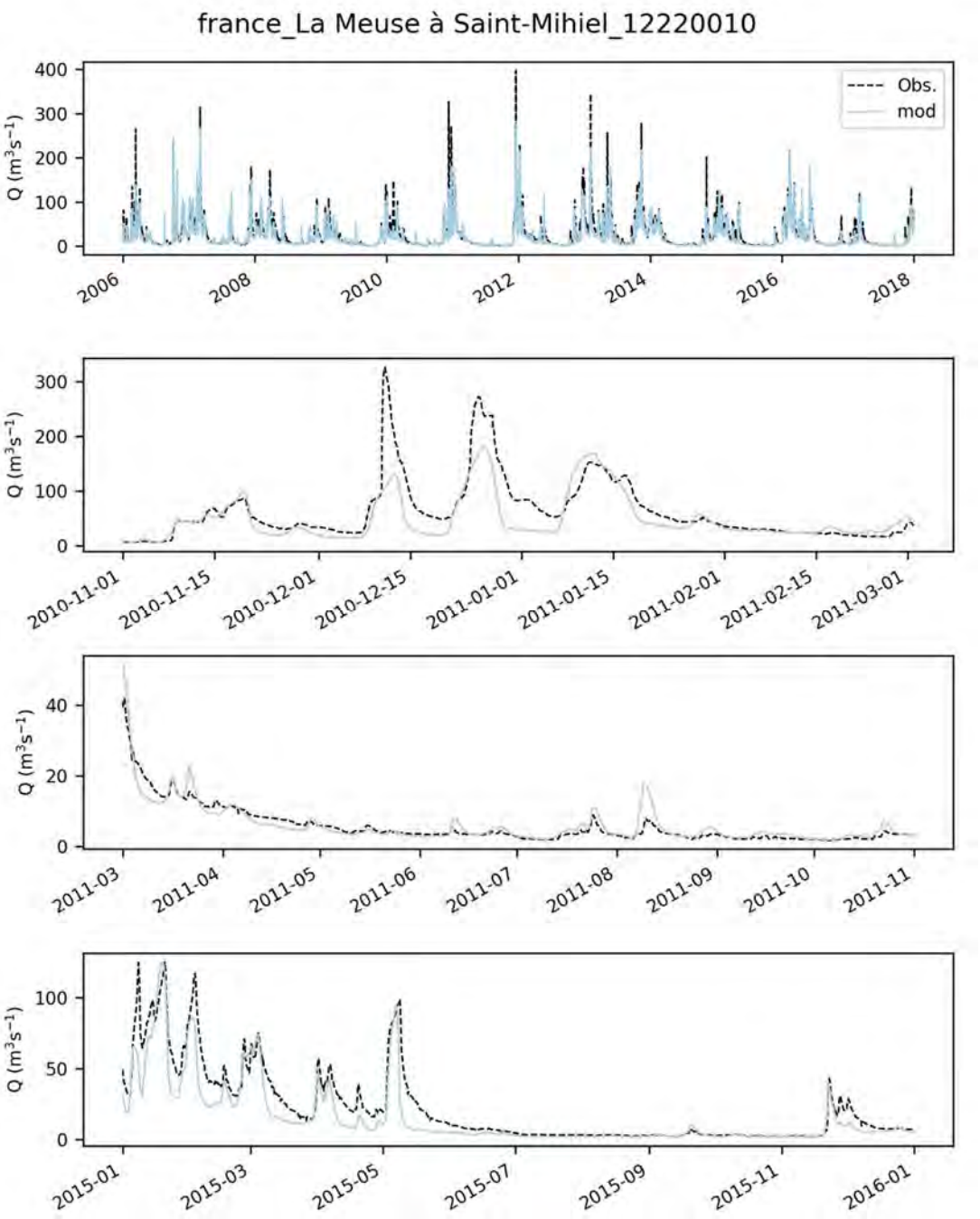


Figure B-25 Modelled (blue) and observed (black) hydrographs for the Meuse at St Mihiel

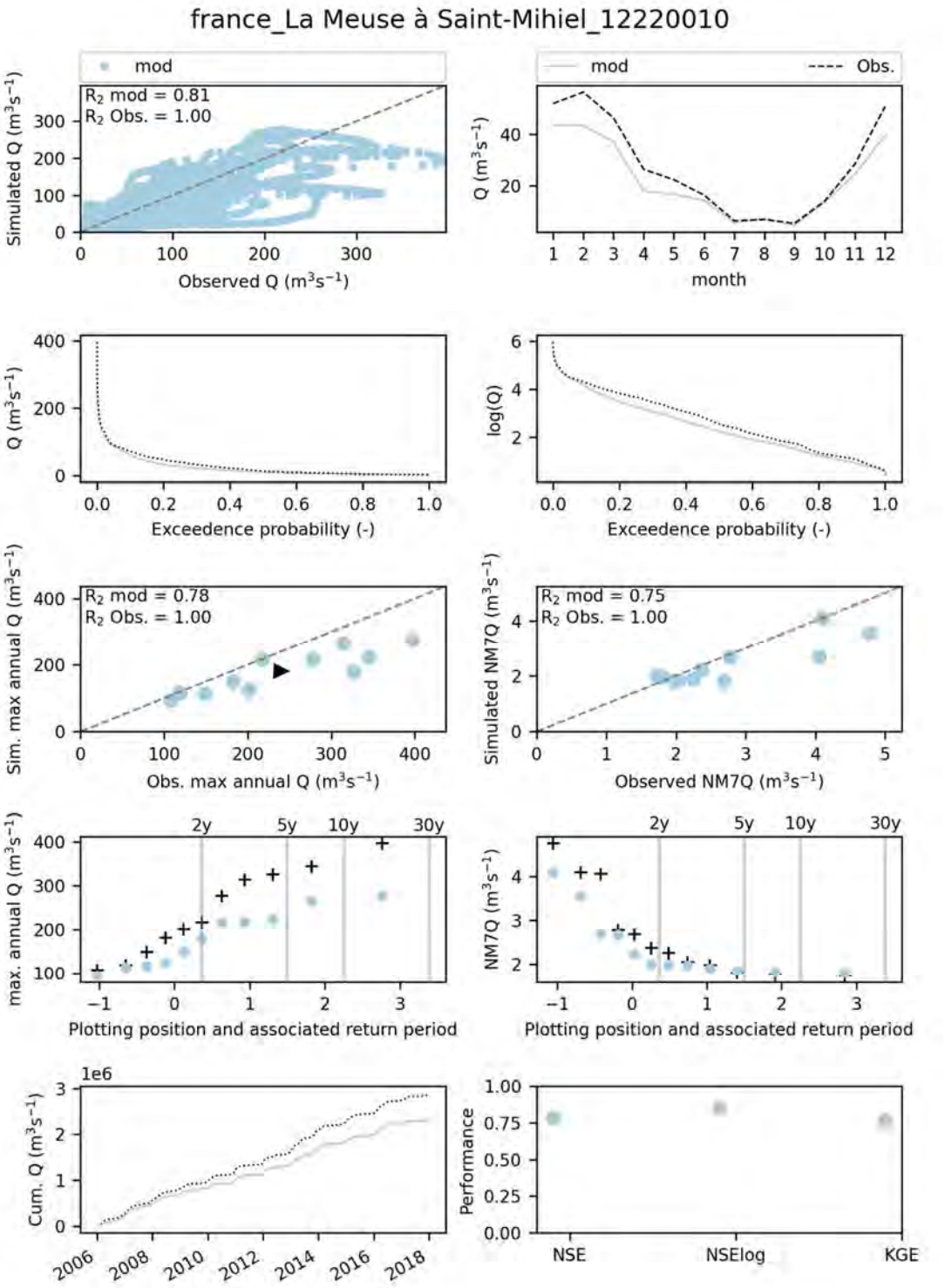


Figure B-26 Modelled (blue) and observed (black) signatures for the Meuse at St Mihiel

C Reservoir schematization in wflow_sbm

C.1 Introduction

In some tributaries of the Meuse, in particular the Vesdre and Rur, reservoirs play an important role in amplifying or reducing the discharge magnitude and timing, particularly during low and high flow conditions. Wflow_sbm has a reservoir and natural lake module, and (location) parameters for both modules can be automatically derived with the HydroMT tool (Eilander et al., 2023). Currently, the location and properties of the reservoirs and lakes are retrieved from the hydroLAKES database (Messager et al., 2016) and the Global Reservoir and Dam Database (GRanD, (Lehner et al., 2011)). Wflow_sbm then attempts to mimic the reservoir behaviour by keeping the reservoir volume between the target minimum and maximum fill fraction as derived from these databases, taking into account the (environmental) discharge demand downstream of the reservoir(s).

To further optimize the modelling of reservoirs when operation rules are known, local data can be used to adjust these parameters. However, the best way to do so, and whether the current reservoir module in wflow_sbm is sufficient to do so, requires testing. In theory, a better simulation of the reservoir operations will also manifest itself in better capturing high and low discharge situations in the model simulations.

In this appendix section, an exploratory test of the current wflow_sbm reservoir approach is performed using five different approaches and local data for the Vesdre catchment. This test is meant to identify the effect different approaches have on the resulting simulated discharge for the Vesdre River. The Vesdre has two reservoirs, Eupen and La Gileppe (Figure C-1), which contribute significantly to the Vesdre discharge (Bruwier et al., 2015). Although not all reservoir information and operation rules are known about both reservoirs, there should be sufficient information to add local information to the reservoir modelling approach for these reservoirs.

Figure C-1 Schematic overview of the Vesdre catchment (delineation in orange) and the two reservoir locations. Background map from OpenStreetMap.org.

C.2 Tested reservoir modelling methods

C.2.1 Overview of the tested methods

Table C-1Table C-1 summarizes the tested methods. The subsequent sections give more details on the approaches in each method.

Table C-1: Overview and description of the tested reservoir modelling methods.

Step	Approach	Information used from
1	Implementation of reservoir information from Bruwier et al. (2015). Comparison with original HydroMT derivation approach.	Bruwier et al. (2015)
2	Implementation of reservoir operation information from Stucky report (Stucky, 2021) into time-variable wflow_sbm reservoir parameters for the Eupen reservoir.	Stucky (2021) Bruwier et al. (2015)
3	Same as step 2, but reservoir information from Eupen also used for La Gileppe reservoir.	Stucky (2021) Bruwier et al. (2015)
4	Same as step 3 but drinking water supply constraint added as MaxLeakage term in wflow_sbm instead of as environmental flow demand in the reservoir.	Stucky (2021) Bruwier et al. (2015)
5	Same as step 4, but Eupen reservoir information added as a natural lake module using a storage-discharge relationship table.	Stucky (2021) Bruwier et al. (2015)

C.2.2 Description of each method

C.2.2.1. Step 1 – Translation of reservoir information from Bruwier et al. (2015) to wflow_sbm reservoir parameters

HydroMT automatically derives reservoir parameters. For the reservoirs of Eupen and La Gileppe, these parameters are listed in *Table* C-2. Although HydroMT gets the area and volume of both reservoirs correct as compared to the information in Bruwier et al. (2015), more specific reservoir information and operation rules are not known and are based on the estimates from the database. Bruwier et al. (2015) have published information and some operation rules for both reservoirs. From this paper and in particular Table 1 in this paper, the maximum release and reservoir demand can be determined. The demand is a combination of the environmental flow requirement of 0.04 $m^3 s^{-1}$ and the drinking water supply demand of 0.6944 $m^3 s^{-1}$ for Eupen and 0.3472 $m^3 s^{-1}$ for La Gileppe. The parameters TargetMinFrac and TargetFullFrac are not directly derivable from the paper but can be estimated with the information given.

Table C-2: wflow_sbm reservoir parameters as derived with HydroMT using the hydroLAKES database (Messager et al., 2016) *and the Global Reservoir and Dam Database (GRanD, Lehner et al., 2011). The two right columns indicate the reservoir parameters as derived using the information from Bruwier et al. (2015). TargetMinFrac and TargetFullFrac indicate the minimum and maximum fraction of the reservoir volume that is within the set optimal window.*

Parameter	HydroMT		Bruwier et al. (2015)	
	Eupen	La Gileppe	Eupen	La Gileppe
Area (m^2)	915617.8	838538.5	915617.8	838538.5
Max. volume (m^3)	2.5e7	2.7e7	2.5e7	2.7e7
Max. release ($m^3 s^{-1}$)	3.9	1.8	4.5	1.8
Demand ($m^3 s^{-1}$)	0.5	0.2	0.73	0.39
TargetMinFrac	0.6	0.2	0.53	0.5
TargetFullFrac	1	1	0.86	0.88

Table C-3: TargetMinFrac and TargetFullFrac estimation for Eupen and La Gileppe reservoirs based on Bruwier et al. (2015) information.

	Eupen	La Gileppe
Total height reservoir	18.0 m	16.0 m
Min. pool level for drinking water	343 m AMSL	284 m AMSL
Max. safety level	361 m AMSL	300 m AMSL
Target max. level	358.5 m AMSL	298 m AMSL
Target min. level	352.5 m AMSL	292 m AMSL
TargetFullFrac (wflow_sbm parameter)	0.861	0.875
TargetMinFrac (wflow_sbm parameter)	0.528	0.500

This estimation consists of two steps: (1) the determination of the total water level height capacity of the reservoir and (2) the determination of the Targetfullfrac and Targetminfrac parametrs based on set minimum and maximum target levels. Especially step 1 is quite uncertain, as these specific values are dependent on the bathymetry, which not publicly available. In the first step, the total, practically available, water level height between min. and max. filled is estimated (Table C-3) using the minimum pool level for drinking water and the max safety level of both reservoirs, as indicated in Table 1 of Bruwier et al. (2015). The TargetFullFrac and TargetMinFrac parameters can then be derived with the target max. and min water levels, which are the dashed blacked lines in Figure 3 of Bruwier et al. (2015). With this information, the wflow_sbm parameters are calculated as follows:

$$Targetfullfrac = \frac{total\ height\ reservoir - (max.safety\ level - target\ max.level)}{total\ height\ reservoir}$$

$$Targetminfrac = \frac{total\ height\ reservoir - (max.safety\ level - target\ min.level)}{total\ height\ reservoir}$$

C.2.2.2. Step 2 – Implementation of time-variable reservoir parameters for Eupen reservoir

In response to the July 2021 floods, Stucky (2021) has conducted a study into, among others, the Vesdre catchment and its reservoirs. From this report, the basic reservoir operation rules of the Eupen reservoir are known, see Figure C-2. Although these reservoir operation rules are not always followed, it gives a good starting point to make the TargetMinFrac and TargetMaxFrac parameters of the Vesdre model time dependent and thus closer to reality given a particular day of the year. To do so, Cote B1 (the dark green line) and Cote P (the dark blue line) in Figure C-2 are regarded as the TargetFullFrac and TargetMinFrac, respectively. The parameters TargetFullFrac and TargetMinFrac are determined from the volumes (right y-axis of the figure) as a fraction of the maximum volume of the reservoir. As only a few volumes are indicated in Figure C-2, specific values for a point on one of the lines are determined by linear interpolation between the two values. The resulting parameter values are visualized in Figure C-3.

C.2.2.3. Step 3 – Implementation of time-variable reservoir parameters for both reservoirs

The information used for step 2 is not available for the La Gileppe reservoir. However, as a test, the TargetFullFrac and TargetMinFrac parameters of Eupen (in step 2) are also used for La Gileppe to make them time dependent for testing.

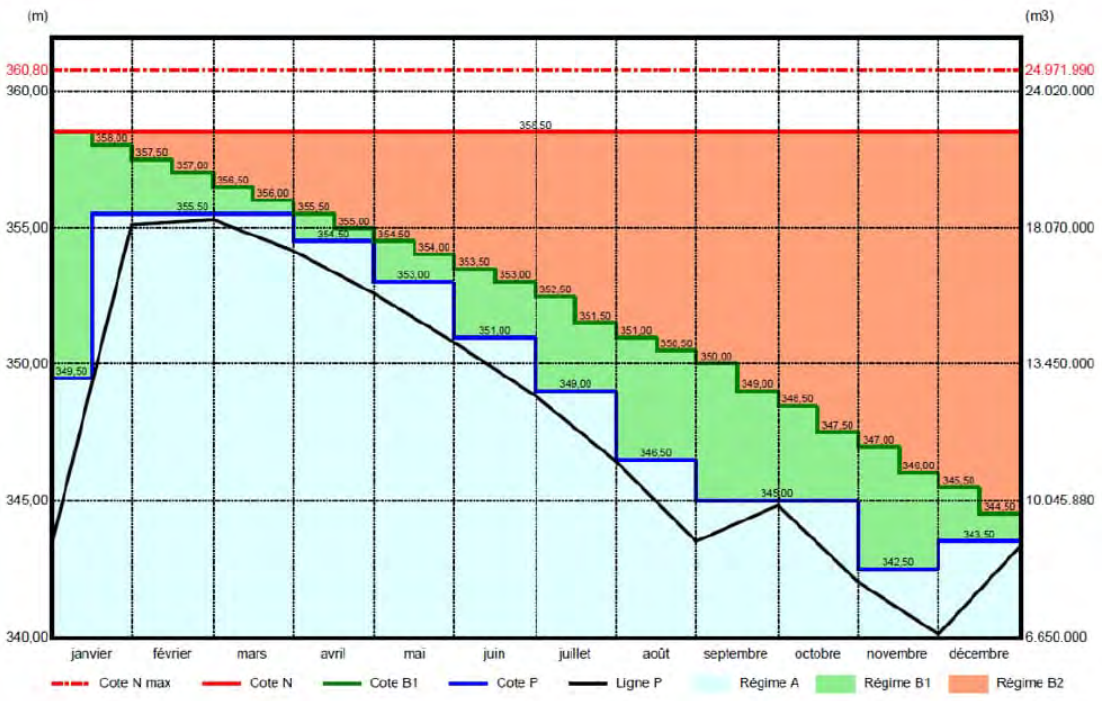


Figure C-2: Basic reservoir operation rules for the Eupen reservoir (Stucky, 2021).

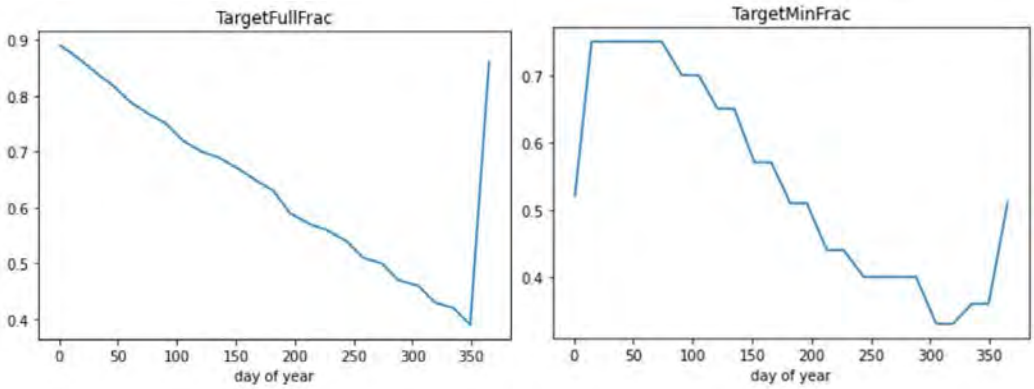


Figure C-3: TargetMinFrac and TargetFullFrac estimates for the Eupen reservoir.

C.2.2.4. Step 4 – Drinking water supply constraint as leakage term instead of as reservoir parameter

Part of the reservoir parameters is the water demand (in $m^3\ s^{-1}$) from the downstream region. Both reservoirs have a base environmental flow requirement (0.04 $m^3\ s^{-1}$) to maintain sufficient water in the downstream part of the Vesdre River during dry periods. In addition, both reservoirs are used for drinking water supply and to a lesser extent for hydropower. The drinking water supply has a demand as well, which corresponds to 0.6944 and 0.3472 $m^3\ s^{-1}$ for Eupen and La Gileppe, respectively. This is part of the reservoir demand now, but this water does not leave the reservoir through the Vesdre River. Instead, it ends up as drinking water and may be returned as discharge from waste-water treatment plants at a later stage (not taken into account in the model at this moment). A simple first modelling approach would be to consider the drinking water demand as a loss term, a leakage, instead of a reservoir demand. Hence, in this step, we continue with the setup of step 3 and lower the reservoir demand to the environmental flow requirement (0.04 $m^3\ s^{-1}$) and add a catchment-wide MaxLeakage of 0.13 mm d^{-1} to the model. This leakage is the sum of both drinking water supply demands divided by the surface area of the Vesdre catchment (685 km^2).

C.2.2.5. Step 5 – Eupen reservoir added as natural lake

Finally, Hartgring (2023) mentioned that reservoirs can also be modelled with a so-called storage-discharge (S-Q) relationship, which are frequently used as a guideline for reservoir operators to optimally operate the reservoir. If available, such an S-Q relationship can help to mimic the reservoir operation rules even better. Hartgring (2023) added the option to use an S-Q relationship to the lake module of wflow_sbm and tested this for the reservoirs of the Rur River. Hence, this requires modelling a reservoir with wflow_sbm's lake module. In contrast to the study by Hartgring (2023), there is no direct information about the S-Q relationship for the Vesdre reservoirs, but some information for the Eupen reservoir can be obtained from Stucky (2021) and Bruwier et al. (2015).

To test this, the Eupen reservoir is changed to a lake in the model. LakeMaxVolume is set to the maximum volume of the reservoir (Bruwier et al., 2015) and a table is added to the model containing the lake S-Q relationship for every day of the year. Wflow interpolates between these values to find the volume and corresponding discharge for a given time.

Table C-4: Approximation of TargetMinFrac and TargetMaxFrac for the Eupen reservoir as a function of time as estimated from the Stucky (2021) report (in particular Fig. 5-45 in this report).

Date	Target max level	Target min level	Target max volume	Target min volume	TargetFullFrac	TargetMinFrac
	(m)		(10^6 m3)			
2022-01-01	358.5	349.5	22.24	13.11	0.89	0.52
2022-01-15	358	355.5	21.64	18.67	0.87	0.75
2022-02-01	357.5	355.5	21.05	18.67	0.84	0.75
2022-02-14	357	355.5	20.45	18.67	0.82	0.75
2022-03-01	356.5	355.5	19.86	18.67	0.79	0.75
2022-03-15	356	355.5	19.26	18.67	0.77	0.75
2022-04-01	355.5	354.5	18.67	17.61	0.75	0.70
2022-04-15	355	354.5	18.07	17.61	0.72	0.70
2022-05-01	354.5	353	17.61	16.22	0.70	0.65
2022-05-15	354	353	17.15	16.22	0.69	0.65
2022-06-01	353.5	351	16.68	14.37	0.67	0.57
2022-06-15	353	351	16.22	14.37	0.65	0.57
2022-07-01	352.5	349	15.76	12.77	0.63	0.51
2022-07-15	351.5	349	14.84	12.77	0.59	0.51
2022-08-01	351	346.5	14.37	11.07	0.57	0.44
2022-08-15	350.5	346.5	13.91	11.07	0.56	0.44
2022-09-01	350	345	13.45	10.05	0.54	0.40
2022-09-15	349	345	12.77	10.05	0.51	0.40
2022-10-01	348.5	345	12.43	10.05	0.50	0.40
2022-10-15	347.5	345	11.75	10.05	0.47	0.40
2022-11-01	347	342.5	11.41	8.35	0.46	0.33
2022-11-15	346	342.5	10.73	8.35	0.43	0.33
2022-12-01	345.5	343.5	10.39	9.03	0.42	0.36
2022-12-15	344.5	343.5	9.71	9.03	0.39	0.36

The discharge that corresponds a given lake volume is based on the time-dependent operating rules (Figure C-2 and Table C-4) and the following discharge corresponding to these rules:

- Volume below Cote P (dark blue line): environmental flow demand (0.04 $m^3 s^{-1}$; Bruwier et al., 2015).
- Volume above Cote B1 (dark green line) and within regime B2 (orange shaded area): maximum flow used (4.5 $m^3 s^{-1}$; Stucky, 2021; Bruwier et al., 2021).
- Volume in between Cote P and Cote B1: linear interpolation between the previous two points and corresponding discharges.
- Volume above maximum safety level (dashed red line): spillway is used, $Q_{in} = Q_{out}$.
- Volume between Cote N (dark red level, set maximum level) and maximum safety level: linear interpolation between maximum flow (4.5 $m^3 s^{-1}$) and maximum flow when the additional outlets are opened (based on angle of opening, max. discharge of 70 $m^3 s^{-1}$; Stucky, 2021).

C.3 Results

C.3.1 Step 1 – Translation of reservoir information from Bruwier et al. (2015) to wflow_sbm reservoir parameters

The difference between the original HydroMT derivation of the wflow_sbm model for the Vesdre and the derivation described in step 1 is quite small (Figure C-4 and Figure C-5). The implementation of step 1, which includes the information from Bruwier et al. (2015), results in somewhat lower discharge peaks, which are overestimated with the wflow_sbm model derived from HydroMT. This results in a somewhat higher, though still negative, NSE value for the approach of step 1.

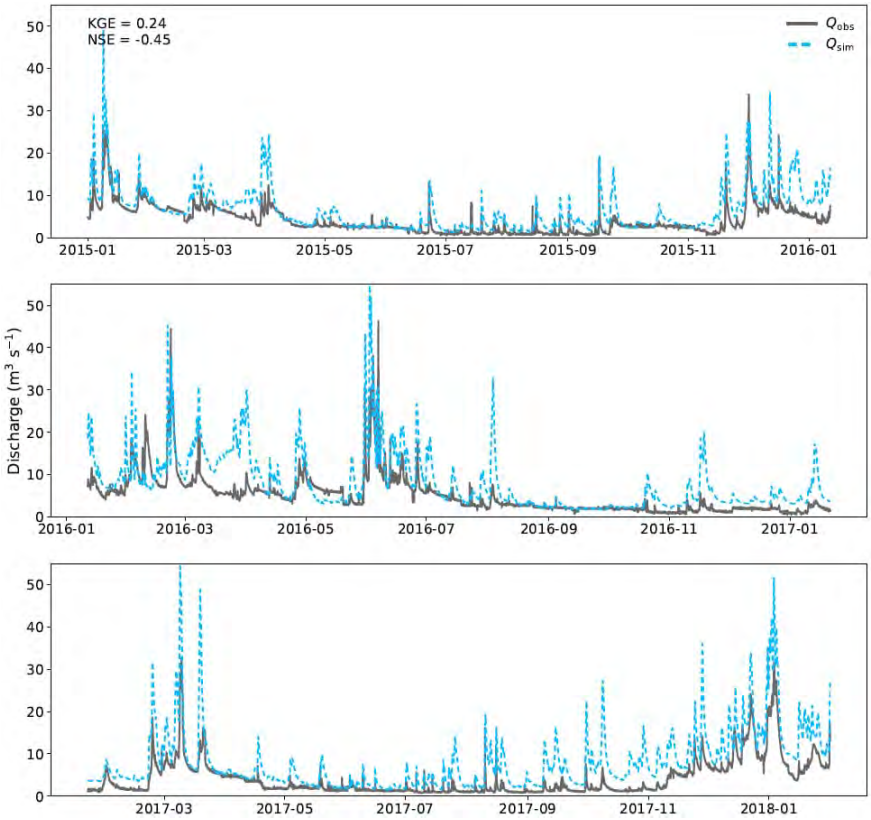


Figure C-4: Simulated versus observed discharge for the Vesdre at gauge location Verviers, which is just downstream of the two reservoirs. Simulations with the original model setup.

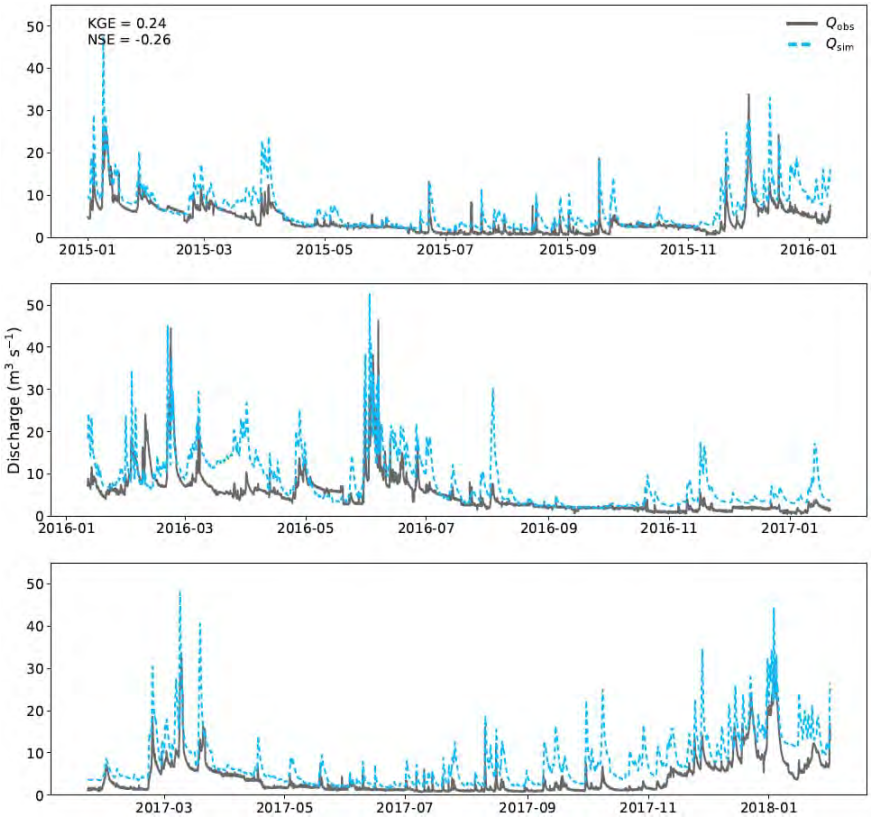


Figure C-5: Simulated versus observed discharge for the Vesdre at gauge location Verviers, which is just downstream of the two reservoirs. Simulations with the approach of step 1.

C.3.2 Step 2 – Implementation of time-variable reservoir parameters for Eupen reservoir

The implementation of a time-variable reservoir parameter results in better discharge estimates during winter (Figure C-7) when the reservoir is filling up, which is better resembled in the simulated reservoir volumes with this method (Figure C-6). However, the approach also results in discharges at Verviers that are systematically too high during summer. This results from the tendency of the reservoir module of wflow_sbm to have a volume close to TargetFullFrac, resulting in quick releases of short duration (a clear on-off behaviour) during the gradual decrease of the target levels from spring until fall. This is an unrealistic behaviour. As a result of the worse performance during low-flow conditions and better performance during winter, the KGE and NSE values remain similar to the previous simulations.

C.3.3 Step 3 – Implementation of time-variable reservoir parameters for both reservoirs

The implementation of the time-variable reservoir parameters for both reservoirs gives similar results to the results of step 2 (results not shown here). In short, the winter discharges are closer to the references than in step 2 and the previous simulations, but discharges during low-flow conditions are systematically overestimated (even more than in step 2).

Figure C-6: Simulated reservoir volume for the Eupen reservoir for the period 2008 until 2018. The grey line indicates the simulations with the original HydroMT derivation, the blue dashed line indicates the simulations of step 1 and the orange dashed line indicates the simulations of step 2.

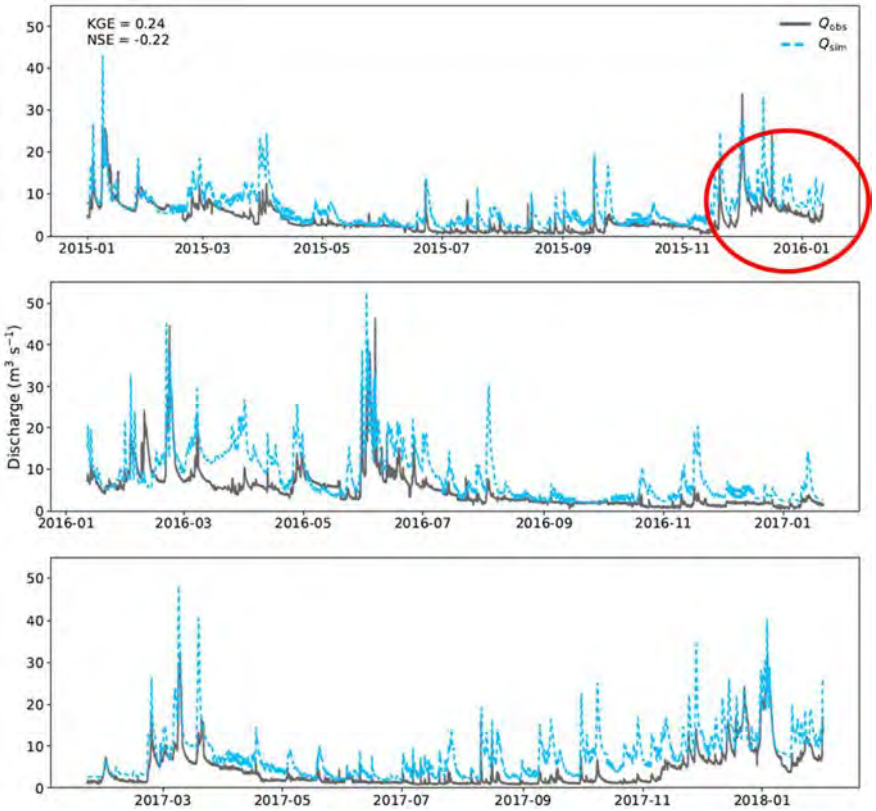


Figure C-7: Simulated versus observed discharge for the Vesdre at gauge location Verviers, which is just downstream of the two reservoirs. Simulations with the approach of step 2.

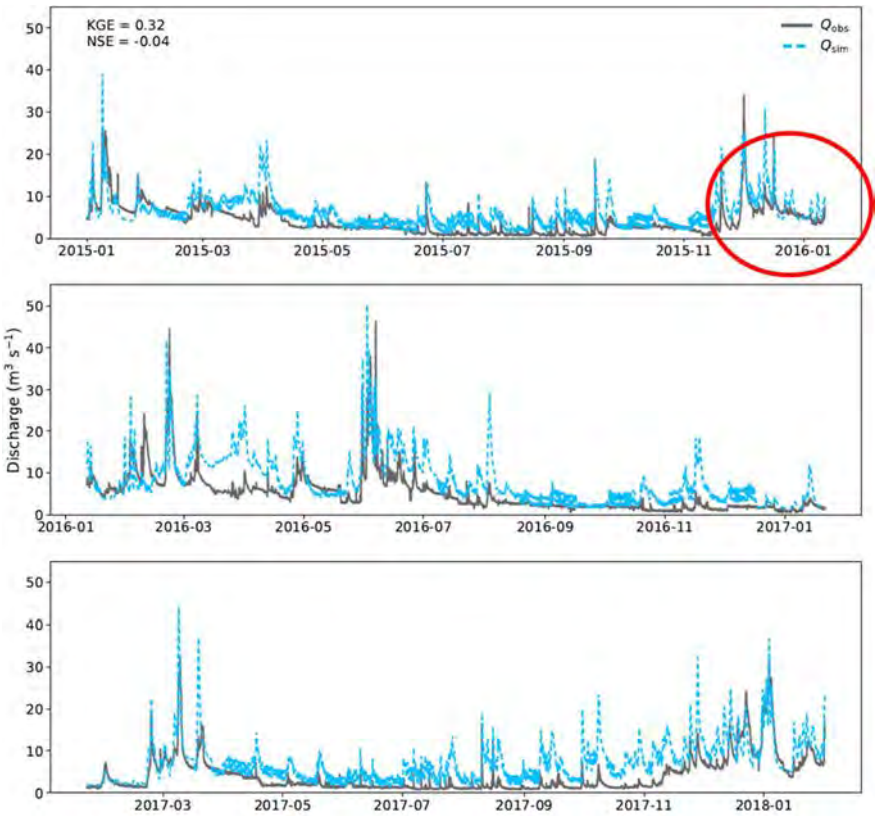


Figure C-8: Simulated versus observed discharge for the Vesdre at gauge location Verviers, which is just downstream of the two reservoirs. Simulations with the approach of step 4.

C.3.4 **Step 4 – Drin**king water supply constraint as leakage term instead of reservoir parameter

The implementation of the drinking water supply as leakage term in wflow_sbm, on top of the implementation of step 3, gives small improvements to the simulated discharge at Verviers (Figure C-8). The simulated discharge for winter conditions is now close to the observations, while the overestimation during low-flow conditions is reduced compared to step 3. This also becomes clear from the KGE and NSE values (0.32 and -0.04, respectively, compared to 0.24 and -0.45 for the original HydroMT derivation). The simulated discharge at the outlet of the Vesdre at Chaudfontaine also improves, especially during winter and for peak discharges (not shown here). KGE and NSE values have a small increase from 0.7 and 0.64 for the original HydroMT derivation to 0.73 and 0.71 for the approach in step 4.

C.3.5 **Step 5 – Eupen reservoir added as n**atural lake

Finally, the implementation of the reservoirs as a natural lake, which allows to implement a table with a fixed S-Q relationship, results in discharge simulations that are similar to the results from step 4 (Figure C-9). However, the on-off behavior that follows from the reservoir module as implemented in the previous steps (especially visible during low flows, see for instance Figure C-8) reduces, resulting in a more natural looking hydrograph.

C.4 Concluding remarks and recommendations

The tested approaches (step 1 to 5) result in small improvements in the discharge simulations for the Vesdre catchment at Verviers, just downstream of both reservoirs. A time-dependent reservoir implementation improves the results, especially in winter when the reservoir fills up.

Implementing the reservoir as a lake with a fixed storage-discharge relationship further enhances this and has as main advantage that it leads to a more natural looking hydrograph, while the original wflow_sbm reservoir module has the tendency to give a strong on-off behavior of the outflow when time-dependent reservoir parameters are implemented.

A point for improvement remains the overestimation of the discharge during low-flow conditions. It is possible that the available reservoir operation rules are not perfectly resembling the operations during low-flow conditions, but this can only be assessed when reservoir volume and outflow information is available for both the Eupen and La Gileppe reservoirs. This is recommendation for further research and can help to better simulate the reservoirs of the Vesdre and the reservoirs of the Meuse in general.

Finally, based on this exploratory analysis and the results from Hartgring (2023), wflow_sbm (wflow_sbm v0.7.0 and onwards) now has the option to add a stage-discharge relationship to the reservoir module when reservoir operation rules or reservoir observations are available.

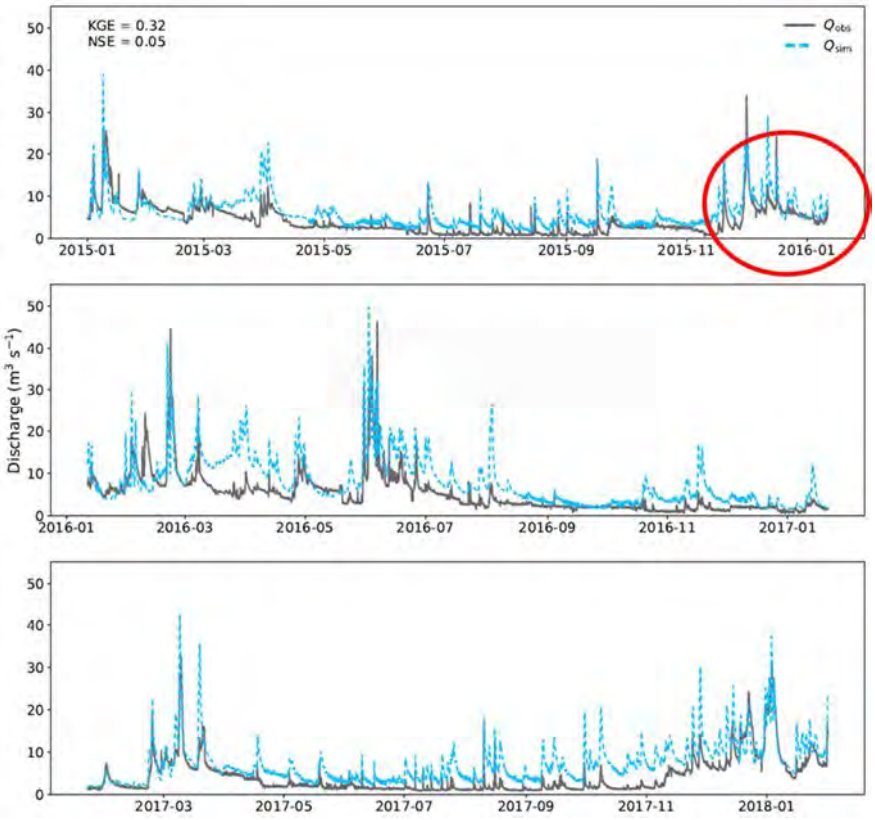


Figure C-9: Simulated versus observed discharge for the Vesdre at gauge location Verviers, which is just downstream of the two reservoirs. Simulations with the approach of step 5.

D Frequency analysis

D.1 Important tools and packages used

For sake of transparency, we list in Table D-1 below the important tools and packages used to implement the workflow described in section 5.2.

Table D-1: Important tools and packages used to obtain the synthetic discharges analysed in this report, see also the schematic in Figure 0-1

Tools / packages	Purpose	Source
snakemake	Setup of the automatic workflow to perform the runs	(Mölder et al., 2021)
cdo	Downscaling of the RACMO forcing data to the required resolution for the wflow_sbm model	(Schulzweida, 2022)
xclim	Extreme value distribution fitting for the continuous gridded synthetic daily discharge. See also scipy.stats	(Bourgault et al., 2023)
pyextremes	Extreme value distribution fitting for the continuous synthetic daily and hourly discharge at the catchments of interest. See also scipy.stats	(Bocharov, 2022)
scipy.stats	Extreme value distribution fitting	(Virtanen et al., 2020)

D.2 Frequency analysis – other selected locations

In addition to the selected locations shown in the main report, return levels from the simulated daily and hourly discharges are shown in Appendix for the following stations:

- Geul at Hommerich
- Viroin at Treignes
- Ambleve at Martinrive
- Semois at Membre Pont
- Lesse at Gendron
- Rur at Stah
- Meuse at St-Mihiel

D.2.1 Daily

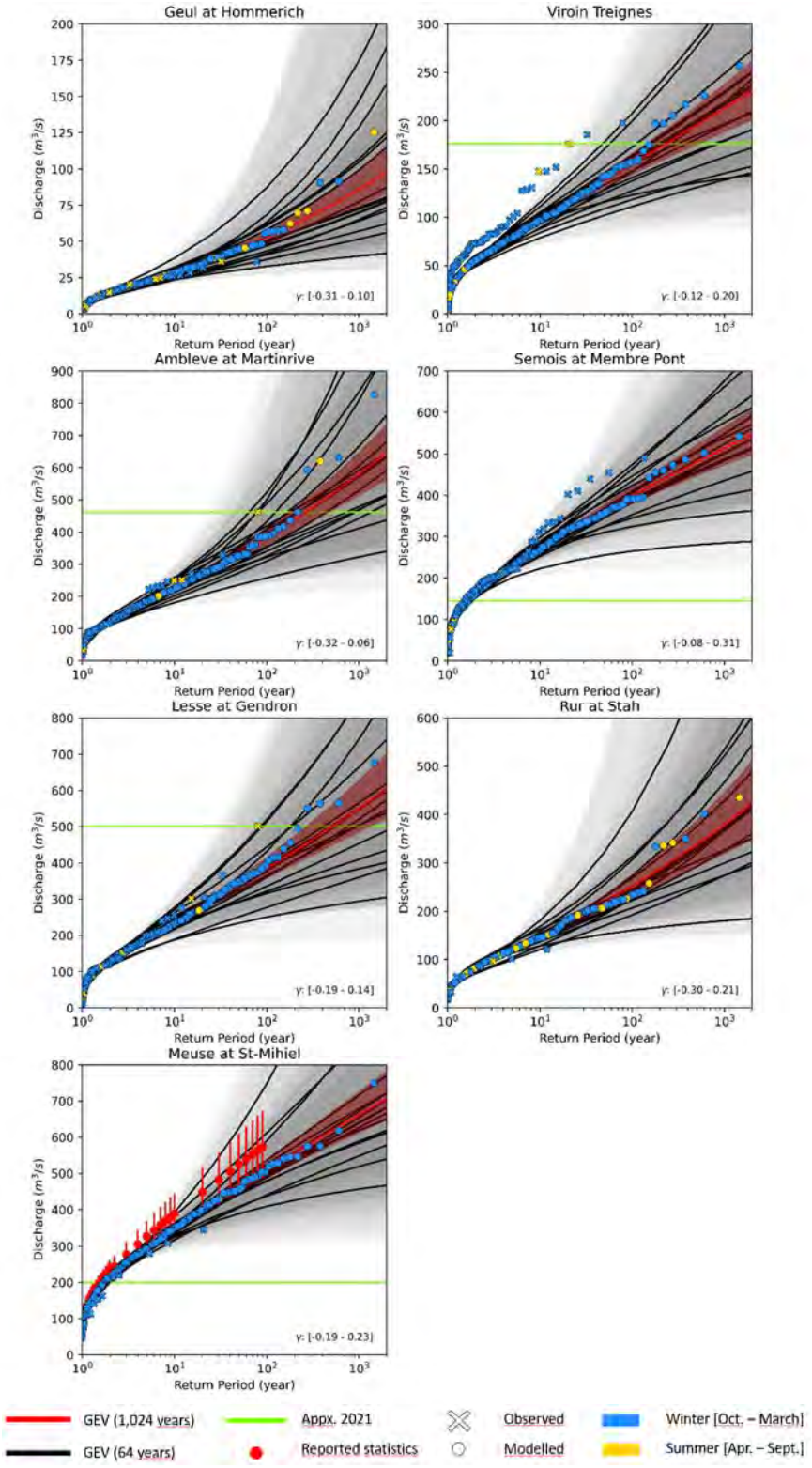


Figure D-1 Extreme value distribution fit at selected locations across the Meuse catchment. Black lines represent the GEV fit obtained from one ensemble member (64 hydrological years). Red lines show the GEV fit obtained from combining all ensembles (1,024 hydrological years, shown as full dots). The shading represents the statistical uncertainty from the distribution parameters uncertainty. Observed discharge peaks when present are shown with crosses. Blue (yellow) colors indicates whether the peak occurred in winter (summer). Recorded July 2021 level when present are shown in green. Official statistics when present and reported uncertainty are shown with red circles and vertical bars, respectively.

Deltares
Royal Netherlands Meteorological Institute Ministry of Infrastructure and the Environment
Rijkswaterstaat Ministry of Infrastructure and Water Management
Deltares

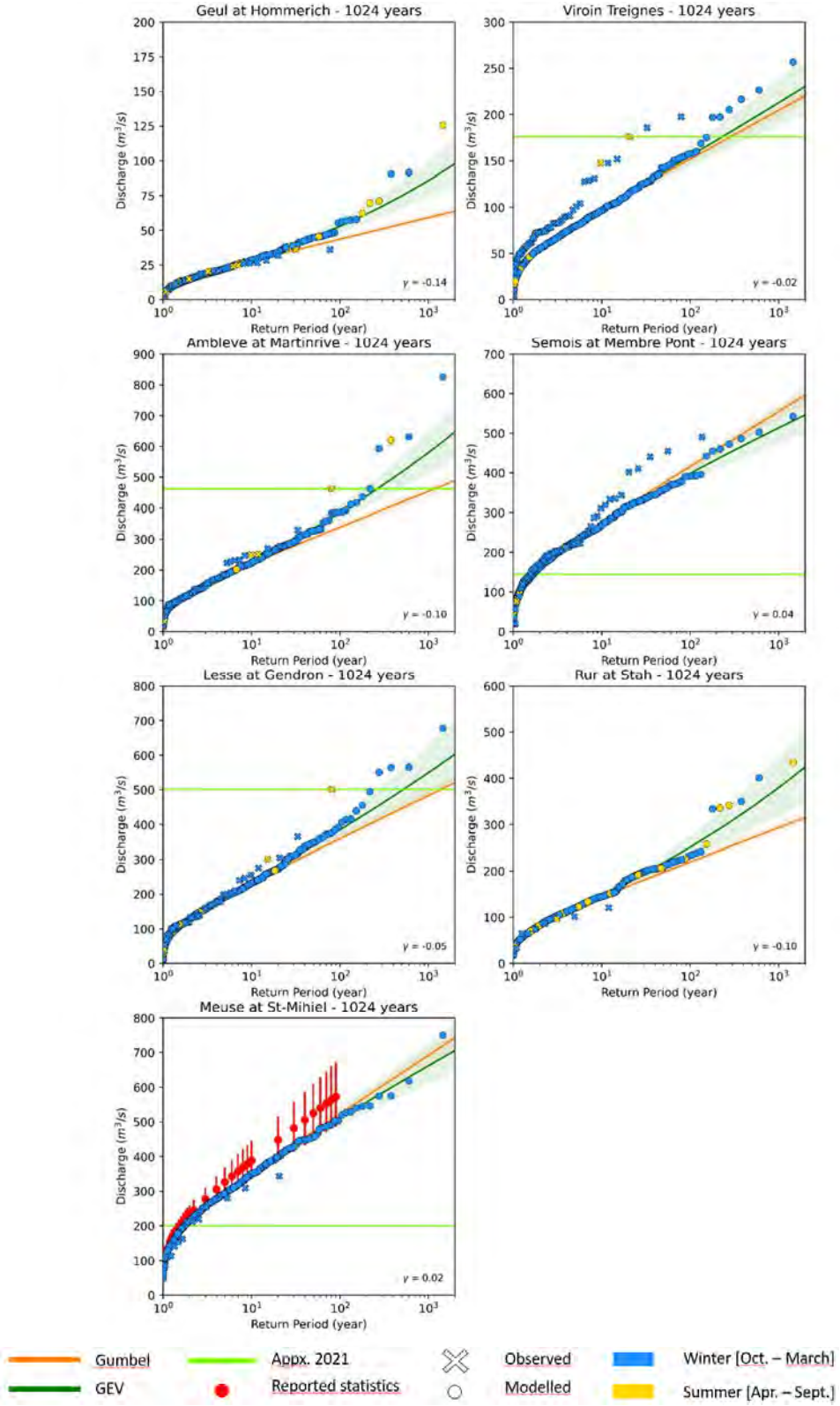


Figure D-2: Extreme value distribution fit at selected locations across the Meuse catchment. Dark green lines represent the GEV fit obtained from combining all ensembles (1,024 hydrological years, shown as full dots) and blue lines the Gumbel fit. The shading represents the statistical uncertainty from the distribution parameters uncertainty. Observed discharge peaks when present are shown with crosses. Blue (yellow) colors indicates whether the peak occurred in winter (summer). Recorded July 2021 level when present are shown in green. Official statistics when present and reported uncertainty are shown with red circles and vertical bars, respectively.

D.2.2 Hourly

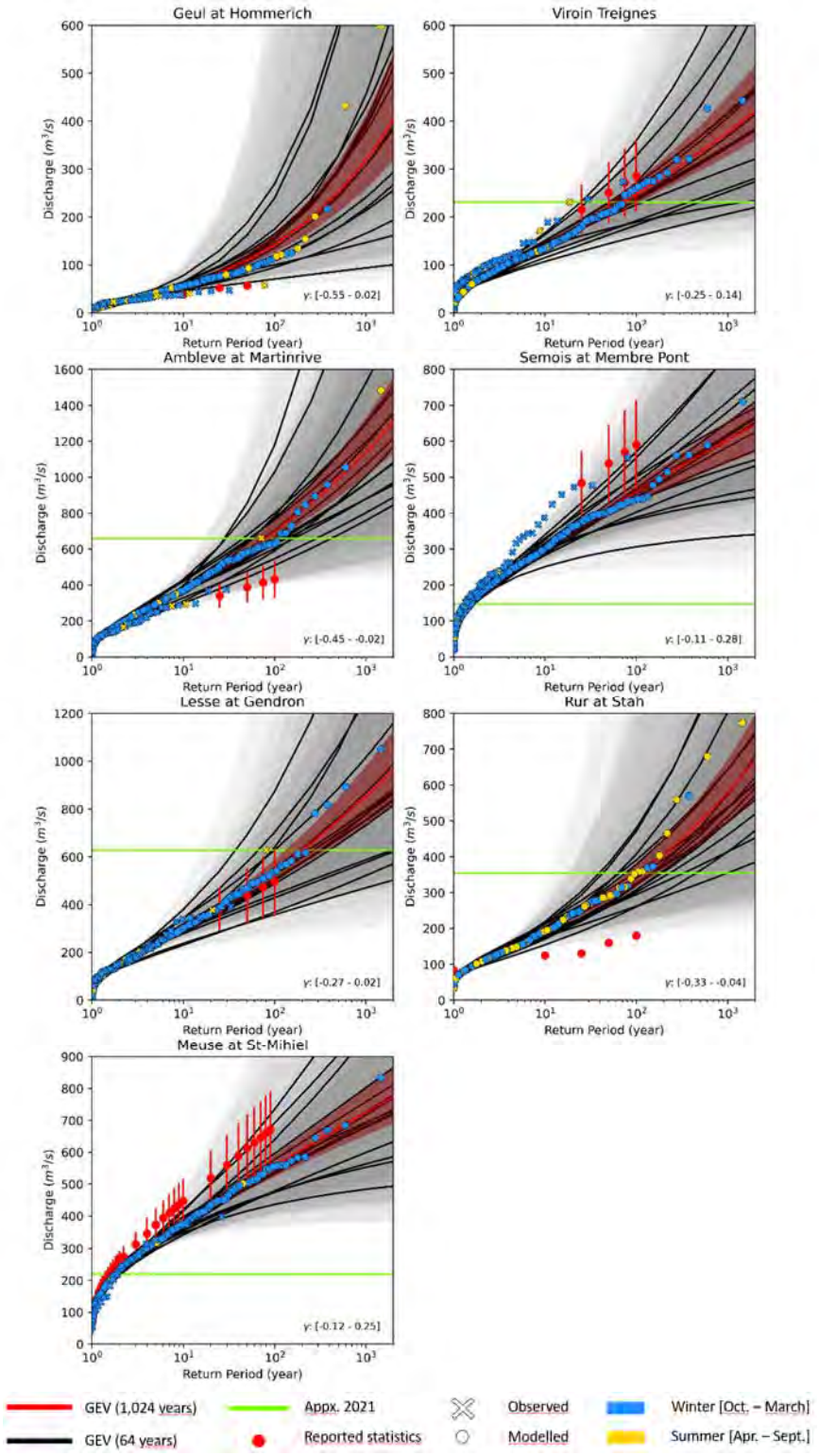


Figure D-3: Extreme value distribution fit at selected locations across the Meuse catchment. Black lines represent the GEV fit obtained from one ensemble member (64 hydrological years). Red lines show the GEV fit obtained from combining all ensembles (1,024 hydrological years, shown as full dots). The shading represents the statistical uncertainty from the distribution parameters uncertainty. Observed discharge peaks when present are shown with crosses. Blue (yellow) colors indicates whether the peak occurred in winter (summer). Recorded July 2021 level when present are shown in green. Official statistics when present and reported uncertainty are shown with red circles and vertical bars, respectively.

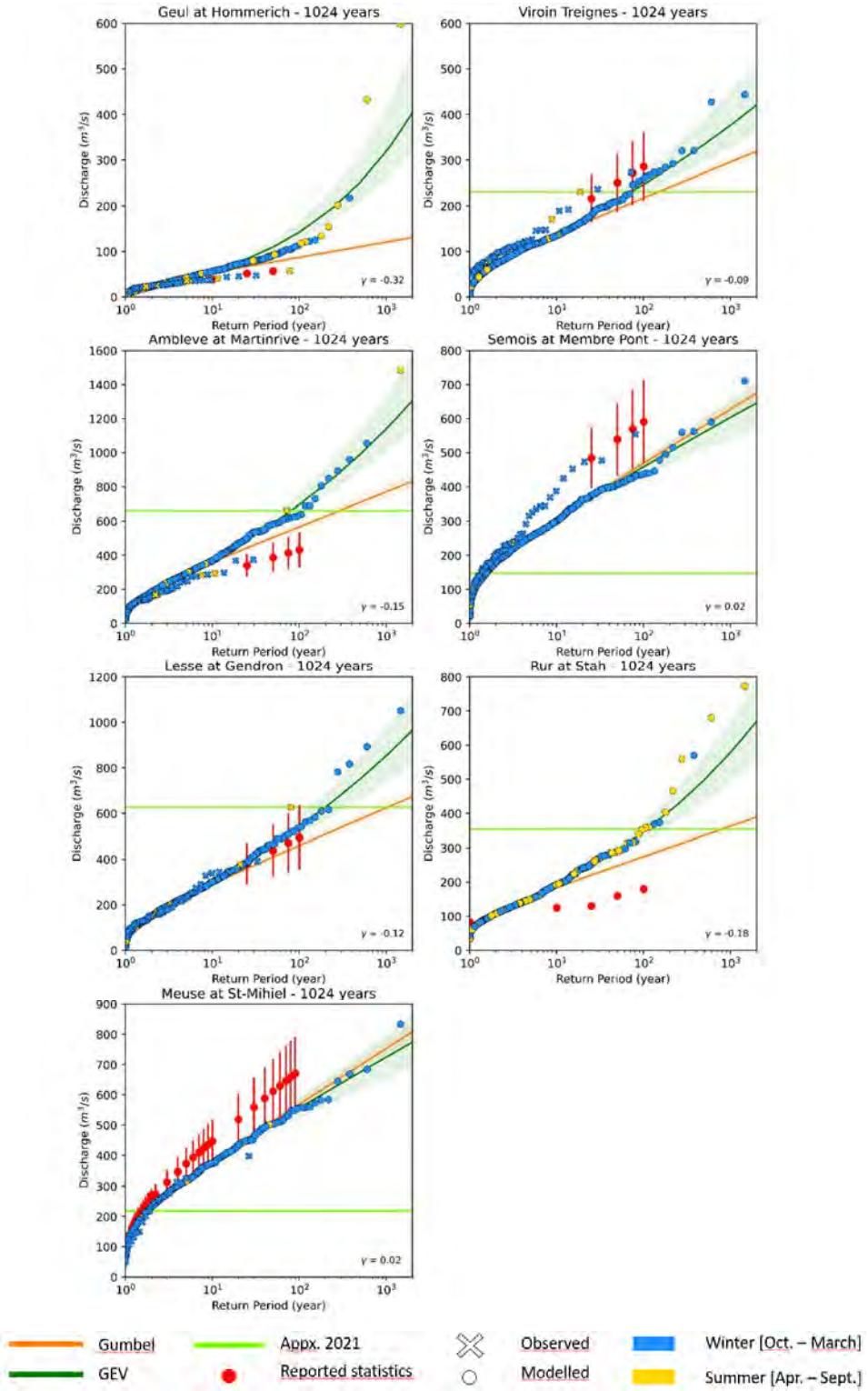


Figure D-4: Extreme value distribution fit at selected locations across the Meuse catchment. Dark green lines represent the GEV fit obtained from combining all ensembles (1,024 hydrological years, shown as full dots) and blue lines the Gumbel fit. The shading represents the statistical uncertainty from the distribution parameters uncertainty. Observed discharge peaks when present are shown with crosses. Blue (yellow) colors indicates whether the peak occurred in winter (summer). Recorded July 2021 level when present are shown in green. Official statistics when present and reported uncertainty are shown with red circles and vertical bars, respectively.

D.3 Influence of timestep

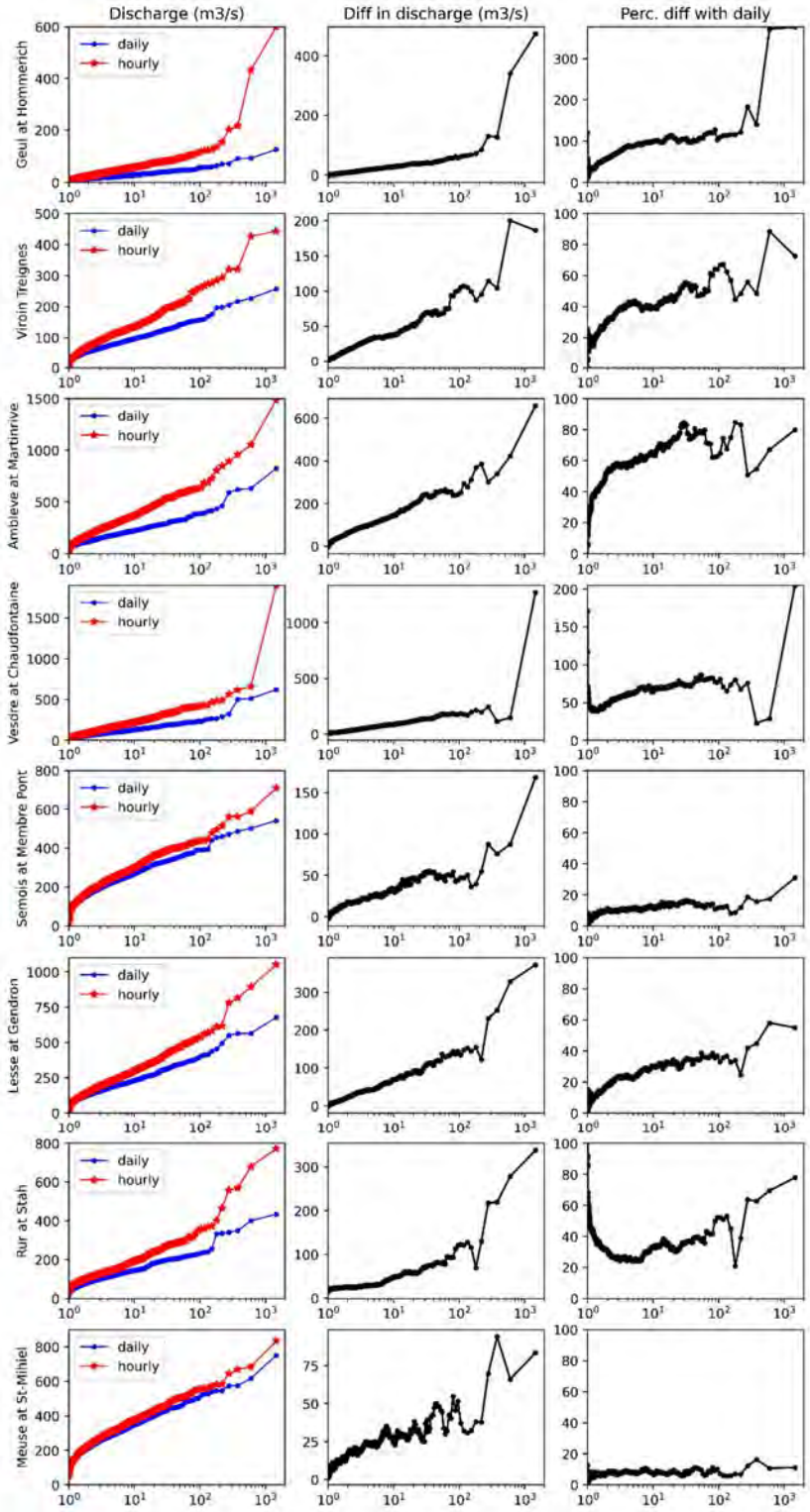


Figure D-5. Left column - Empirical return levels obtained from 1,024 hydrological years at selected locations from the daily and hourly runs. Middle column: Difference in discharge between the two temporal resolution. Right column: Relative percentage difference with respect to the empirical daily return levels.

Deltares
Royal Netherlands Meteorological Institute Ministry of Infrastructure and the Environment
Deltares
Rijkswaterstaat Ministry of Infrastructure and Water Management

D.4 Return discharge levels for given return periods of selected tributaries

Table D-2: Obtained return levels from fitting GEV and 95% confidence interval from the simulated daily discharges

Location of interest	5 year return level (m^3/s)		10 year return level (m^3/s)		100 year return level (m^3/s)		1000 year return level (m^3/s)	
	daily	95% CI	daily	95% CI	daily	95% CI	daily	95% CI
Geul at Hommerich (The Netherlands)	23	[22, 23]	29	[28, 30]	53	[48, 58]	86	[73, 101]
Viroin at Treignes (Belgium)	81	[80, 82]	98	[96, 101]	155	[146, 166]	213	[192, 238]
Ambleve at Martinrive (Belgium)	183	[180, 186]	227	[220, 233]	385	[354, 416]	579	[502, 663]
Semois at Membre-Pont (Belgium)	229	[226, 232]	271	[266, 279]	397	[376, 428]	511	[467, 579]
Lesse at Gendron (Belgium)	192	[190, 195]	236	[231, 243]	385	[362, 416]	549	[496, 629]
Rur at Stah (Germany)	120	[119, 122]	148	[145, 152]	251	[233, 270]	379	[334, 432]
Meuse at Saint-Mihiel (France)	296	[292, 299]	349	[342, 355]	510	[482, 538]	662	[601, 723]

Table D-3: Obtained return levels from fitting GEV and 95% confidence interval from the simulated hourly discharges

Location of interest	5 year return level (m^3/s)		10 year return level (m^3/s)		100 year return level (m^3/s)		1000 year return level (m^3/s)	
	daily	95% CI	daily	95% CI	daily	95% CI	daily	95% CI
Geul at Hommerich (The Netherlands)	41	[40, 42]	57	[54, 60]	142	[124, 166]	319	[250, 420]
Viroin at Treignes (Belgium)	112	[110, 113]	141	[137, 145]	247	[230,267]	376	[333, 430]
Ambleve at Martinrive (Belgium)	292	[287, 296]	374	[362, 383]	695	[636, 744]	1139	[976, 1285]
Semois at Membre-Pont (Belgium)	254	[252, 257]	305	[299, 311]	458	[438, 484]	603	[559, 661]
Lesse at Gendron (Belgium)	238	[235, 241]	302	[294, 309]	542	[505, 581]	853	[755, 962]
Rur at Stah (Germany)	154	[152, 157]	191	[186, 197]	346	[317, 381]	579	[492, 688]
Meuse at Saint-Mihiel (France)	320	[315, 322]	378	[369, 385]	556	[521, 584]	724	[650, 786]

Deltares is an independent institute for applied research in the field of water and subsurface. Throughout the world, we work on smart solutions for people, environment and society.

Deltares

www.deltares.nl