Interreg
Euregio Meuse-Rhine
EM Flood Resilience

EUROPEAN UNION
European Regional Development Fund

Evaluation of discharge extremes in the Meuse river and her tributaries

The EMfloodResilience project is being carried out within the context of Interreg V-A Euregio MeuseRhine and is 90% funded from the European Regional Development Fund."

Evaluation of discharge extremes in the Meuse river and her tributaries

Client	Rijkswaterstaat, Water, Verkeer en Leefomgeving
Contact	R. Slomp and N.R. van der Sleen
Reference	-
Keywords	

Document control	
Version	1.0
Date	01-12-2023
Project nr.	11208719-000
Document ID	11208719-000-ZWS-0002
Pages	161
Classification	
Status	final

Author(s)		
Anaïs Couasnon **Laurène Bouaziz** **Alessia Riveros** **Ruben Imhoff**		

Summary

In July 2021, extreme discharges caused extensive flooding in many tributaries of the Meuse, driven by an intense precipitation event. At these locations, this event was the highest observed on records and much higher than previous extremes observed. This event also happened in summer, which makes the event even more rare. Frequency analysis based on observed time series in the basin have a high degree of uncertainty to estimate the return period of such an event, because the observed time series are only a few decades long. To estimate large return periods of extreme discharge, the GRADE method was developed in the Netherlands. This method statistically extrapolates observed weather time series using a statistical weather generator that temporally resamples the time series and generate much longer weather time series. The resulting weather time series are used with a hydrological models to generate very long synthetic time series of discharge events. This can extend record lengths but cannot generate more extreme daily rainfall events than the one observed and outliers have a large impact on the resampling scheme, becoming too frequently resampled. Finally, it cannot generate time series at subdaily time scales because the method is based on daily weather time series, while the peak discharge event may be subject to shorter time scale processes. Even though some of the statistical uncertainty is reduced by this method by increasing the length of the record, it does not generate the physics of extreme events not observed previously, such as extreme summer events.

This report investigates improvement on the GRADE approach. Whereas GRADE uses daily observed weather variables, here we base the generation of synthetic discharges on a physically-based climate model, which results in meteorological time series at full spatial and temporal resolution for the Meuse basin. Furthermore, we capture the physical processes leading to extreme discharges through a physically-based hydrological model. In this way, the time series of meteorological variables representing the current climate are extended to a much longer record length of 1,040 years, generating physically plausible weather systems that can lead to extreme discharge but have not been captured by observations.

This study was made possible by the Interreg Euregio Meuse-Rhine program. It is part of the EMfloodResilience project managed by the Regional Water Authority of Limburg (Waterschap Limburg) in the Netherlands. The project includes stakeholders from the Meuse basin, the partners in the EMfloodResilience project, among others the Vlaamse Waterweg (Belgium) and the Service Public de Wallonie (Belgium). These stakeholders collectively reflected on new avenues for robust extreme discharge estimation methods, as the one suggested here.

The two goals of this study are:

- to set-up a physically based and gridded hydrological model for the Meuse basin, including the main river and its tributaries, which is improved based on feedback gathered by the different stakeholders and evaluated with historical observations at the daily and the hourly timestep;
- to run the developed and improved hydrological model of the Meuse basin with an ensemble of weather data from the RACMO model, developed by KNMI to generate a long timeseries of synthetic discharge data, , assess the behaviour of extreme discharges over different locations in the basin that are of interest to the stakeholders, and assess the behaviour of extremes over different seasons.

The report includes all the requirements listed for task D.T5.4.2 from work package 5, and listed in Table 1-1 in section 1.3.1 of this report.

The physically-based modelling chain is a collaboration between the Royal Netherlands Meteorological Institute (KNMI), the Dutch Ministry of Infrastructure and Water Management (Rijkswaterstaat) and Deltares (depicted in Figure 0-1). Long-term synthetic meteorological time series are provided by KNMI. Subsequently, the meteorological forcing is translated into extreme discharges through the use of the wflow model, a distributed hydrological model.

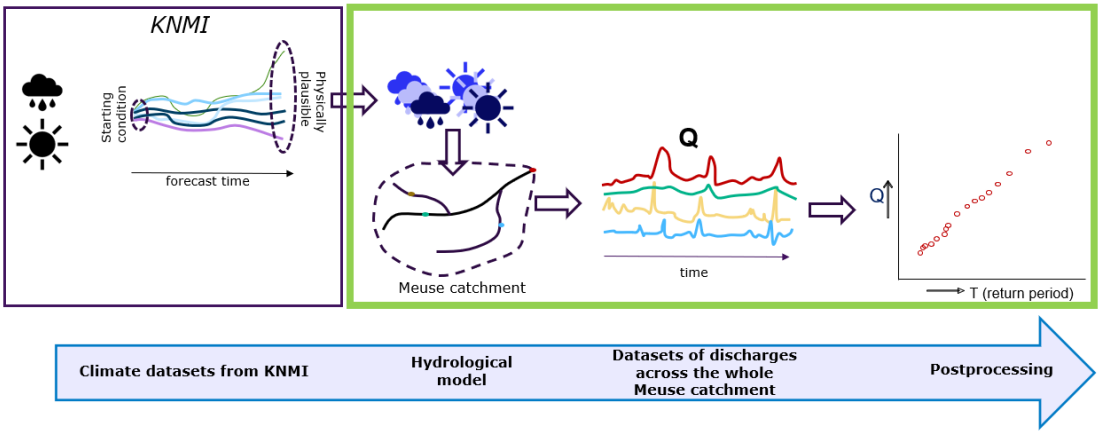


Figure 0-1 Schematic overview of general approach The scope of this report is shown with a green rectangle.

The physically-based modelling chain is applied for the Meuse basin. Based on discussions with the partners, several catchments of interest have been selected for extreme frequency analysis. They include stations along the Meuse, Sambre, Vesdre, Ourthe, Viroin, Semois, Ambleve, Lesse, Geul and Rur (or Roer in Dutch) rivers .

To accomplish the first goal, a wflow hydrological model has been set-up at the hourly and daily time step and improved throughout the project with discussions and feedback from the stakeholders. The main improvements include a manual calibration of parameters of the daily model, the implementation of a new lateral routing module, an automatic calibration of parameters of the hourly model and a further optimized implementation of the Geul and the Rur catchments. The model performs well at the daily and hourly time step across the selected locations. Nevertheless, it is particularly challenging to check the validity of the short duration discharge extremes considering the relatively limited amount of observations of such events and the various mechanisms present. An hourly time step however, is preferred to capture extremes in the smaller catchments and tributaries since it better represents processes in the fast-responding tributaries of the Meuse basin.

To accomplish the second goal, the model has been run with the synthetic meteorological RACMO dataset at the hourly and daily timestep for 1,040 years. The annual maxima from hydrological years (Oct-Sep) were retrieved from these continuous time series, and an extreme value analysis is performed by creating an empirical distribution function, and fitting the Generalized Extreme Value (GEV) and Gumbel distribution functions. Increasing the record length up to 1,040 years allows to reduce the statistical uncertainty of the parameters of the distribution. We show that the shape parameter of the GEV distribution can be incoherent (negative and positive, meaning an upward or downward curvature of the tail of the distribution) when estimated from relatively short record length (~65 years), both at the daily and hourly time step. Instead, the long record length reduces the variability of the shape parameter and leads to its convergence. Given the total length of 1,040 years using in this study, the estimated return periods are robust up to approximately the 340-year return period.

The extreme value analysis also revealed that, while most extremes occur in wintertime, the most extreme discharge events can occur in summer instead of in winter. Especially in small and steep catchments, the tail is dominated by the presence of summer events. This suggests adapting the extreme value analysis methodology to account for these different flood mechanisms by splitting summer and winter extremes. We show that failure to do so can results in a strong overestimation of the return periods, for example for the Rur at Stah.

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, 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.

We conclude that the modelling chain is successfully applied for the Meuse basin. Return periods obtained are in line with current estimates from tributaries of the Meuse both at the daily and hourly time steps. For example, for the Meuse at St Pieter, the 100-year return level obtained for the daily discharge is 3,330 m^3/s. This is in line with previous estimates of 3,220 m^3/s using GRADE (Hegnauer et al., 2014).This is also the case for smaller catchment size, such as the Meuse at Goncourt, where similar return level estimates are found at the daily time step or for the Viroin at Treignes at the hourly time step. However, the results show some bias in some areas where further optimization of the hydrological model is still needed. For example, at the hourly time step, the model results are underestimating return levels at the Meuse at Goncourt.

We recommend further optimization the model at the hourly time steps since it resulted in improvements for both the hourly and daily time scales. For flood mitigation and design, return levels from return periods much higher than 100 years are often needed, such as in the Netherlands. Given the length of the synthetic discharge of 1,040 years, we estimate our results to be reliable up to a return period of 340 years approximately. Therefore, this modelling chain may be expanded to longer synthetic time series to more reliably estimate higher return levels. This could be done for example by using other climate ensemble datasets such as the SEAS5 dataset or in combination with a statistical weather generator or other spatial resampling techniques.

For decision makers, our study provides important insights:

- We find other summer events of the same or worse magnitude than observed in July 2021 in small and medium sized catchment (up to 2,500 km^2 approximately). This means that having a physically-based modelling chain is of clear added value for extreme value analysis. This cannot be achieved by the use of a conventional statistical weather generator with short observational input timeseries, because it lacks the representation of the different physical mechanisms leading to extremes. The extreme events found in this study may be used to plan flood response (e.g. by simulating possible consequences) . For larger catchments, such as the Meuse at St Pieter, while the RACMO dataset contain rainfall events similar or worse than the July 2021 event, the total synthetic record length of 1,040 years is not long enough to assess its impact on summer extremes with confidence for large catchments.
- Using observed time series of discharge to estimate extreme summer events does not lead to robust return period estimates because these time series are often not long often to have observed extreme summer discharge events. The winter and summer discharge extremes should be addressed separately for return level estimation when summer events dominate the most extreme discharge events.

- This is because they result from very different weather systems, and therefore statistically belong to different sampling populations. If this is the case, the approach of using annual maxima irrespective of the season is ill-posed and not applicable anymore. This has strong implications for design because return periods of a given discharge value may turn out lower when summer and winter extremes are considered separately.
- Summer events such as the July 2021 event, are present in the current climate variability. The modelling chain can be used to estimate how these events would be influenced by external drivers such as climate change.

Verkorte samenvatting

In juli 2021 heeft een hoogwater plaatsgevonden in het Maasstroomgebied. Dit hoogwater was zeer uitzonderlijk van aard omdat het a) zeer extreem was en b) in de zomer plaatsvond. In deze studie is een modelinstrumentarium ontwikkeld om tot een betere schatting te komen van de herhalingstijd van heel zeldzame hoogwaters (zoals die van juli 2021) in het Maasstroomgebied en de achterliggende extreme waardeverdeling van hoogwaters. De methode is gebaseerd op de eerder ontwikkelde methode “Generator of Rainfall and Discharge Extremes” (GRADE) en heeft als voordeel dat veel langere afvoerreeksen gebruikt kunnen worden om zo een veel rijkere statistische benadering van extreme afvoercondities te bepalen. Het nieuwe instrumentarium heeft een volledig fysisch gebaseerde basis. Het bestaat uit lange fysisch gebaseerde synthetische weerreeksen uit het klimaatmodel RACMO van het KNMI en een gekalibreerd fysisch gebaseerd gedistribueerd hydrologisch Wflow_SBM model voor het Maasstroomgebied in de Wflow modelsoftware van Deltares. Het instrumentarium kan volledig nieuwe synoptische weerssystemen en hun impacts op hoogwaterstatistiek simuleren en biedt hiermee een basis voor een robuustere afvoerstatistiek en mogelijkheden om de impact van klimaatverandering op afvoerextremen te simuleren.

Het instrumentarium is als onderdeel van deze studie gebruikt om 1.040 jaar (16 ensemble members van elk 65 jaar) aan synthetische afvoeren te genereren. Dit is gedaan door 1.040 jaar synthetisch weer te simuleren met RACMO en deze reeksen door te rekenen tot rivierafvoer met het hydrologische model Wflow_SBM. Op belangrijke plaatsen in het stroomgebied, die gekozen zijn in overleg met de projectpartners, zijn afvoerreeksen geëxporteerd uit het model. Deze zijn vergeleken met meetreeksen en verder statistisch geanalyseerd met focus op de hoge extreme afvoerwaarden. Hierbij zijn zowel empirische als geparameteriseerde (Gumbel en Generalized Extreme Value - GEV) distributiefuncties van de jaarextremen afgeleid.

Deze analyse laat zien dat de parameters van deze distributiefunctie met minder onzekerheid geschat kunnen worden dan met alleen geobserveerde reeksen. Met name de “vorm”-parameter in de GEV-distributie is zeer onzeker met korte reeksen, maar convergeert wanneer de lange reeksen, zoals gegenereerd in dit project, worden gebruikt. De lange reeksen laten ook zien dat, alhoewel de meeste hoogwaters in de winter plaatsvinden, in bepaalde gevallen de meest extreme hoogwaters in de zomer plaatsvinden. Dit is met name het geval in relatief kleine en sterk hellende stroomgebieden. Voor deze gevallen moet overwogen worden om de extreme waardeverdeling te splitsen voor zomer- en winterextremen.

Afhankelijk van de kenmerken van het stroomgebied, heeft de gekozen tijdstap van het hydrologisch model aanzienlijke invloed op extreme herhalingstijden. Gemiddeld heeft dit geleid tot een verschil van 20% voor de uurlijkse tijdstap in vergelijking met de dagelijkse tijdstap. Over het algemeen zijn kleinere zijrivieren van de Maas gevoeliger voor de tijdstap van het model. Dit is te verwachten omdat deze stroomgebieden korte reactietijden hebben, van minder dan een dag.

Met de lange reeksen van 1.040 jaar kunnen extremen tot een herhalingstijd van ca. 340 jaar goed geschat worden. Daarboven blijft de onzekerheid van de schattingen nog hoog. De verwachting is dat de onzekerheid kan worden verkleind door de lengte van de reeks nog verder te vergroten.

Dit kan belangrijk zijn voor ontwerpvragen die vaak, zoals in Nederland, gebaseerd zijn op afvoeren met zeer extreme herhalingstijden. Zomerhoogwaters, vergelijkbaar met die van juli 2021 worden al teruggevonden in de reeks. De neerslagsom over het gehele Maasstroomgebied van de gebeurtenis van juli 2021 is echter dusdanig zeldzaam, dat een nog langere reeks nodig is om hiervan de kans met meer zekerheid vast te stellen.

Kurze Zusammenfassung

Im Juli 2021 kam es im Einzugsgebiet der Maas zu einem Hochwasser. Dieses Hochwasser war von außergewöhnlicher Natur, da es a) sehr extrem war und b) im Sommer stattfand. In dieser Studie wurde ein Modellierungsinstrumentarium entwickelt, um die Wiederkehrzeit sehr seltener Hochwasser (wie das vom Juli 2021) im Einzugsgebiet der Maas und die zugrunde liegende Extremwertverteilung von Hochwassern besser abschätzen zu können. Die Methode basiert auf der eher entwickelten Methode "Generator of Rainfall and Discharge Extremes" (GRADE) und hat den Vorteil, dass viel längere Abflussreihen verwendet werden können, um eine viel reichhaltigere statistische Annäherung an extreme Abflussbedingungen zu bestimmen. Das neue Instrumentarium hat eine vollständige physikalische Grundlage. Es besteht aus langen physikalisch basierten synthetischen Wetterreihen aus dem Klimamodell RACMO des KNMI und einem kalibrierten physikalisch basierten hydrologischen Wflow_SBM Modell für das Maaseinzugsgebiet in der Modellsoftware Wflow von Deltares. Die Werkzeuge können völlig neue synoptische Wettererscheinungen und deren Auswirkungen auf die Hochwasserstatistiken simulieren und bieten eine Grundlage für robustere Abflussstatistiken und Möglichkeiten zur Simulation der Auswirkungen des Klimawandels auf Abflussextreme.

Die Instrumente wurden im Rahmen dieser Studie verwendet, um 1.040 Jahre (16 Ensemblemitglieder zu je 65 Jahren) synthetischer Abflüsse zu erzeugen. Dazu wurden 1.040 Jahre synthetischen Wetters mit RACMO simuliert und diese Reihen mit dem hydrologischen Modell Wflow_SBM in Abflüsse umgerechnet. An Schlüsselstellen im Einzugsgebiet, die in Absprache mit den Projektpartnern ausgewählt wurden, wurden Abflussreihen aus dem Modell exportiert. Diese wurden mit Messreihen verglichen und weiter statistisch ausgewertet, wobei der Schwerpunkt auf den hohen extremen Abflusswerten lag. Dabei wurden sowohl empirische als auch parametrisierte (Gumbel und Generalised Extreme Value - GEV) Verteilungsfunktionen für die Jahresextremen abgeleitet.

Diese Analyse zeigt, dass die Parameter der Verteilungsfunktionen mit geringerer Unsicherheit geschätzt werden können als bei Verwendung von ausschließlich beobachteten Reihen. Insbesondere der "Form"-Parameter in der GEV-Verteilung ist bei kurzen Reihen sehr unsicher, konvergiert aber, wenn die langen Reihen, wie sie in diesem Projekt erzeugt wurden, verwendet werden. Die langen Reihen zeigen auch, dass die meisten Hochwasser im Winter auftreten, in besonderen Fällen die treten die extremsten Hochwasser jedoch im Sommer auf. Dies ist insbesondere in relativ kleinen und stark geneigten Einzugsgebieten der Fall. In diesen Fällen sollte eine Aufteilung der Extremwertverteilung für Sommer- und Winterextreme in Betracht gezogen werden.

Abhängig von den Merkmalen des Einzugsgebiets hat der gewählte Zeitschritt des hydrologischen Modells erhebliche Auswirkungen auf die Wiederkehrzeiten der Extremereignisse. Im Durchschnitt ergab sich ein Unterschied von 20 % für den stündlichen Zeitschritt im Vergleich zum täglichen Zeitschritt. Im Allgemeinen reagieren die kleineren Nebenflüsse der Maas empfindlicher auf den Zeitschritt des Modells. Dies ist zu erwarten, da diese Einzugsgebiete kurze Reaktionszeiten von weniger als einem Tag haben.

Mit den langen Zeitreihen von 1.040 Jahren können Extreme bis zu einer Wiederkehrzeit von etwa 340 Jahren gut abgeschätzt werden. Darüber bleibt die Unsicherheit der Schätzungen immer noch hoch. Es wird erwartet, dass sich die Unsicherheit durch eine noch größere Länge der Reihen verringern lässt. Dies kann für Bemessungsfragen wichtig sein, die oft, wie in den Niederlanden, auf Abflüssen mit sehr großen Wiederkehrzeiten beruhen. Sommerliche Hochwasser vergleichbar mit dem vom Juli 2021 sind in der Zeitreihe enthalten.

Die Niederschlagssumme über das gesamte Maaseinzugsgebiet des Ereignisses von Juli 2021 ist jedoch so selten, dass eine noch längere Reihe erforderlich ist, um die Wahrscheinlichkeit dieses Ereignisses mit größerer Sicherheit zu bestimmen.

Résumé abrégé

La crue de l'inondation juillet 2021 dans le bassin versant de la Meuse a été très exceptionnelle en raison a) de son caractère extrême et b) du fait qu'elle s'est produite en été. Dans cette étude, un framework de modélisation a été développée pour parvenir à une meilleure estimation des périodes de retour des hautes eaux très rares (comme celle de juillet 2021) et de la distribution de celles-ci dans le bassin versant de la Meuse. La méthode est basée sur la méthode Generator of Rainfall and Discharge Extremes (GRADE) développée précédemment et présente l'avantage de pouvoir utiliser des séries de débit beaucoup plus longues pour faire des analyses statistiques des débits extrêmes. Le nouveau framework a une base entièrement physique. Il se compose de longues séries météorologiques synthétiques à base physique, issues du modèle climatique RACMO de l'Institut royal météorologique des Pays-Bas (KNMI), et d'un modèle hydrologique distribué wflow_sbm calé et à base physique pour le bassin versant de la Meuse dans le logiciel de modélisation Wflow de Deltares. Le framework simule des systèmes météorologiques synoptiques complètement nouveaux et leurs impacts sur la formation des hautes eaux, permettant ainsi de calculer des statistiques des débits extrêmes plus robustes. Il permet aussi de simuler l'impact du changement climatique sur les débits extrêmes.

L'instrumentation a été utilisée dans le cadre de cette étude pour générer 1040 années (16 membres d'ensemble de 65 ans chacun) de débits synthétiques. Pour ce faire, 1040 années de conditions météorologiques synthétiques ont été simulées par RACMO qui ont servies pour simuler des débits fluviaux en tournant le modèle hydrologique wflow_sbm. Des séries de débits simulées ont été exportées du modèle à des endroits clés du bassin versant, choisis en concertation avec les partenaires du projet. Elles ont été comparées aux séries mesurées et ont fait l'objet d'une analyse statistique plus poussée, l'accent étant mis sur les valeurs des débits extrêmes les plus élevés. Au cours de ce processus, des modèles fréquentiels paramétrés (loi de Gumbel et loi des extrêmes généralisées GEV) ont été ajustés aux valeurs extrêmes annuelles.

Cette analyse montre que les paramètres de cette fonction de distribution peuvent être estimés avec moins d'incertitude qu'en utilisant uniquement des séries observées. En particulier, le paramètre de forme de la distribution GEV est très incertain avec des séries courtes, mais converge lorsque de longues séries, telles que celles générées dans ce projet, sont utilisées. Les longues séries montrent également que, bien que la plupart des hautes eaux se produisent en hiver, elles peuvent aussi se produire en été, notamment dans les bassins versants relativement petits et aux pentes prononcées. Le cas échéant, il faudrait envisager de séparer la distribution des valeurs extrêmes d'été de ceux d'hiver.

En fonction des caractéristiques du bassin versant, le pas de temps choisi pour le modèle hydrologique a un impact significatif sur les périodes de retour extrêmes. En moyenne, il en résulte une différence de 20 % pour le pas de temps horaire par rapport au pas de temps journalier. En général, les affluents plus petits sont plus sensibles au pas de temps du modèle, ce qui est attendu car ces bassins versants ont des temps de réponse plus courts, généralement inférieurs à un jour.

Avec la longue série de 1040 ans, les débits extrêmes jusqu'à un temps de récurrence d'environ 340 ans peuvent être bien estimés. Au-delà, l'incertitude des estimations reste élevée. On s'attend à ce que l'incertitude puisse être réduite en prolongeant encore la longueur de la série. Cela est important pour les questions de conception qui sont souvent, comme aux Pays-Bas, basées sur des débits avec des périodes de retour très élevées.

Des crues estivales similaires à celles de juillet 2021 sont déjà présentes dans les séries. Cependant, le nombre des précipitations sur l'ensemble du bassin versant de la Meuse menant à ces évènements est si rare qu'une série encore plus longue serait nécessaire pour déterminer leur probabilité de retour avec plus de certitude.

Contents

1 Introduction

1.1 Background

In July 2021, a quasi-stationary low-pressure system caused significant flooding in central Europe with devastating impacts in many countries such as Germany, Belgium, and the Netherlands in the Rhine and Meuse catchments. Among the many reasons described that set this event apart, studies mentioned the interaction between atmospheric, hydrological, and morphological processes with local processes and landscapes at different spatial and temporal scales (Lehmkuhl et al., 2022; Ludwig et al., 2023; Mohr et al., 2023). For example, a series of rainfall events in the weeks before the flood contributed to increased soil moisture conditions and lowered infiltration capacity, exacerbating the impacts of the short-lived and intense rainfall amounts received between July 13 and 16 (Mohr et al., 2023).

Understanding the frequency of these rare events is of utmost importance for decisions makers (Dewals et al., 2021). Unreliable estimation of flood return levels may result in a too low design of protection infrastructure, potentially unexpected structural failure during extreme discharge events (Mohr et al., 2023). It can also have a direct impact on flood hazard maps delineation and subsequent decision making for flood mitigation, for example for determining areas labelled as flood-prone (Dewals et al., 2021).

Reliably estimating large return period discharge values, for example larger than the 100-year return period, is challenging when based on observed time series, as demonstrated by the July 2021 flood event (Vorogushyn et al., 2022). In the severely impacted areas, the magnitude of this event by far surpassed any of the largest magnitude observed in the continuous time series used to perform the discharge frequency analysis. Estimates of return periods are classically based on statistical extrapolation using extreme value models and prone to large uncertainties. For example, initial estimates of the return period of the peak flow at Altenahr gauge on the river Ahr (Rhine basin) resulted in a return period exceeding 100 million years. However, this estimate was judged non representative since similar discharges have been observed in the past centuries (for example the 1804 and 1910 floods). However, these historical floods were not used in the extreme frequency analysis which was based on the last 74 years of data. Including these additional events resulted in a return period for the July 2021 event between 2,600 and 58,700 years (90% confidence interval) (Vorogushyn et al., 2022). Including the July 2021 event also almost doubled the estimate of the 100-year discharge compared to previous calculations (Ludwig et al., 2023).

Relying solely on observed time series not only limits the number but also the type of rare events observed. In locations with multiple flood mechanisms, not properly accounting for these different flood mechanisms can heavily impact flood hazard estimates (Couasnon et al., 2022; Hoshino et al., 2022). In the Meuse basin, summer flood events are less frequent than winter flood events and are often not present in the observed time series. The presence of different flood mechanisms can point towards a non-identically distributed extreme discharge population, or mixed distribution, for which fitting a single extreme value distribution function is invalid.

Another limitation of relying only on observed time series is that estimates are only available at the measurement locations. This can be circumvented by applying a hydrological model resulting in either longer time series for a location and/or discharge information at other locations where no observations are available. However, understanding the model performance, especially at capturing extremes, becomes key.

Limited model performance may add an additional model uncertainty besides the statistical uncertainty (Hoshino et al., 2022).

In the Netherlands, a statistical weather generator is currently used to extend rainfall time series at a daily time step for Borgharen to the equivalent of 50,000 years of current climate. The synthetic weather time series are subsequently used as input in a hydrological model to generate 50,000 years of flows, which are then used to establish more robust statistics. The entire method is referred to as GRADE (Generator of Rainfall And Discharge Extremes,Hegnauer et al., 2014). However, this method cannot be directly applied for the whole Meuse catchment because processes at smaller tributary level have a much shorter time scale and therefore require sub-daily simulation time steps, for example hourly, to capture the dynamics of extreme discharges. Another limitation of the current weather generator is that it does not capture physics to extrapolate extreme events but is based on the statistical resampling of observed time series. The statistical extrapolation leads to an unrealistic asymptotic behaviour for large return periods and especially for short accumulation times, which are of particular relevance for tributaries. In the 50,000 years of GRADE, no summer event reached the level of July 2021. This is because extreme summer events have seldom appeared in the observed precipitation time series used. Therefore, even though some of the statistical uncertainty is reduced by increasing the length of the record, it does not generate the physics of extreme events not observed, such as extreme summer events.

An alternative is to generate synthetic weather observations from a physically-based climate model to extend the time series. This may capture the physical processes leading to extreme rainfall and discharge events. If used with a gridded hydrological model at a sub-daily time step, this can subsequently provide estimates of extreme discharge along all catchments of the Meuse, including fast responding tributaries. In a parallel study, such synthetic weather series were generated by downscaling of a global climate model using the RACMO regional climate model covering Europe (van Voorst and van den Brink, 2023). Several present-day representative ensemble members were created, by using historical greenhouse gas emissions and by running from different initial conditions. In this way, coherent gridded time series of meteorological variables representing the current climate can be extended to more than a 1,000 years, a much longer record length than observations. Another advantage compared to observations is that it separates the internal variability from the current climate to other external drivers such as climate change.

The Interreg Euregio Meuse-Rhine program made it possible to carry out this study. It is part of the EMfloodResilience project managed by the Regional Water Authority of Limburg (Waterschap Limburg) in the Netherlands. The project includes a lot of stakeholders from the Meuse basin, the partners in the EU Flood resilience Program EMfloodResilience project, among others the Vlaamse Waterweg (Belgium) and the Service Public de Wallonie (Belgium), to collectively reflect on possible new avenues for robust extreme discharge estimation methods, such as the one suggested here.

1.2 Objectives

The overarching objective of this study is to derive a physically-based modelling chain to estimate extreme discharges across the whole Meuse catchment, including its tributaries.

To reach this objective, the following two goals are defined:

- set-up a hydrological model for the Meuse, which is improved based on feedback gathered by the different stakeholders, partners in this Interreg EMfloodResilience project, and evaluated with historical observations at the daily and the hourly timestep.

This is a necessary step to evaluate the model uncertainty under observed historical conditions.

- to run the developed and improved hydrological model of the Meuse basin with an ensemble of weather data from the RACMO model, developed by KNMI (van Voorst and van den Brink, 2023) to generate a long timeseries of synthetic discharge data, assess the behaviour of extreme discharges over different locations in the basin that are of interest to the stakeholders, and assess behaviour of extremes over different seasons.

Figure 1-1 schematizes the approach to fulfil these objectives and derive extreme discharge return values for the Meuse catchment and her tributaries. First, the generation of the synthetic meteorological variables is performed by KNMI using the regional climate model RACMO. This task is performed by KNMI under work package WP5, task D.T5.4.1 and is described in van Voorst and van den Brink (2023) and denoted by a black box in Figure 1-1. At the end of this step, 16 ensemble members of 65 calendar years from 1950-2014 are generated, leading to 1,040 years of current climate variables (precipitation, temperature and potential evaporation). These series are referred to as the RACMO dataset and are available at the daily and hourly time step.

The green box in Figure 1-1 represents the work presented in this report for work package WP5, task D.T5.4.2, and consists of three main steps:

- The precipitation, temperature and potential evaporation from the RACMO dataset are used as forcing to run the spatially distributed hydrological model wflow for the Meuse basin at the daily and hourly time step.
- This results in 1 ,040 years of synthetic continuous discharge time series for the Meuse and her tributaries, at the daily and hourly time step.
- The synthetic discharge time series are analysed to extract discharge peaks and to estimate daily and hourly discharge return levels for the Meuse river and her tributaries.

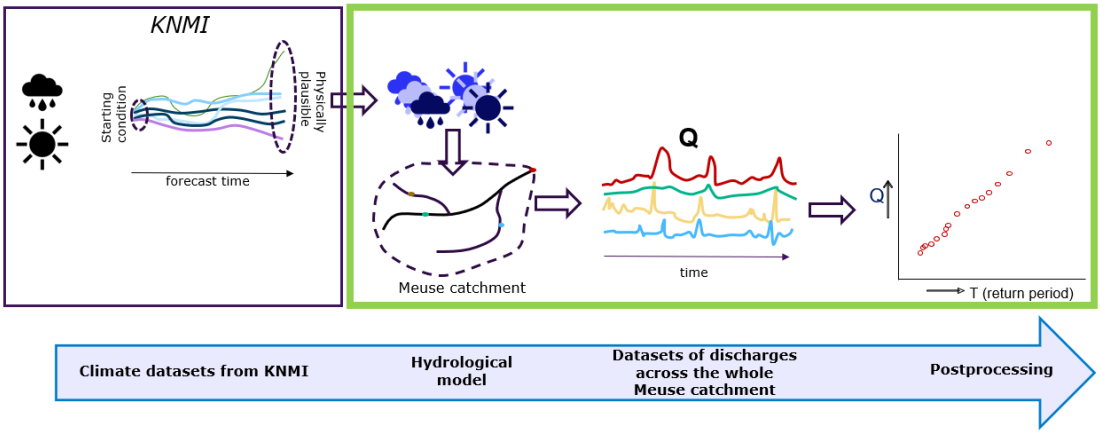


Figure 1-1 Schematic overview of general approach. The scope of this report is shown with a green rectangle.

The selection of the specific Meuse tributaries and case study locations is based on stakeholder involvement and further described in section 2. These locations reflect areas of specific importance for flood assessment and the design of flood defences along the Meuse river and her tributaries. In total, 15 locations were selected, spread across France, Belgium, Germany and the Netherlands and encompassing catchment areas varying from 147 km^2 for the Rur river at Monschau to 21,233 km^2 for the Meuse at Borgharen.

This study provides a better understanding, evaluation and improvement of the hydrological model for the Meuse basin both at a daily and hourly time step.

It also provides estimates of extreme discharge return levels at these time steps for the selected case studies where return level estimates may already exist. This can be beneficial to complement local information or inform on future development for this work.

1.3 Structure of the report

Chapter 2 introduces the study area, the Meuse catchment and her tributaries, and a description of the specific case study area selected.

Chapter 3 provides a description of the forcing dataset used in this study as well as the observed discharge time series and current discharge statistics available at the selected case study areas.

Chapter 4 presents information about the hydrological model. First, a description of the model is given in Sect. 4.1. The improvements that were made to the model after feedback from the partners and stakeholders are described in Sect.4.2. The model evaluation results using historical observed data at the daily and the hourly timestep is presented in Sect. 4.3.

The model developed in Chapter 4 is used to perform the extreme discharge frequency analysis shown in Chapter 5. A description of relevant literature and previous existing work using synthetic discharge data for the Meuse catchment is described in Sect. 5.1. The exact workflow to obtain the discharge extremes and fit the extreme value distribution is presented in Sect. 5.2. Sect. 5.3 explains some methodological choices and theory of extreme value distribution fit. The obtained discharge return levels at the daily and hourly time steps are presented in Sect. 5.4 and more specifically reflecting on the influence of record length, flood seasonality, and the selected timestep in the obtained return level estimates.

Chapter 6 provides an overview of conclusions and some recommendations for future work based on the most important findings of this study.

1.3.1 Requirements of the Interreg-Meuse project

The requirements for task D.T5.4.2 have been addressed in the report as follows:

Table 1-1: List of requirements for task D.T5.4.2 and corresponding sections

Requirement	Covered in sections
Information about the model	Sect. 4.1
Improvements to the model and feedback from partners	Sect. 4.2
Results from modelling flood events, based on historical events and synthetic events	Sect. 4.3 and Sect. 5.4
Overview of conclusions and literature	Sect. 6 and 7
Dutch summary	page 8
Disclaimer on the funding source	page 2

On top of these requirements, a French and German summary were added (pages 10 and 12). We also added a table with the 1, 10, 100, 1,000 year return periods for the Meuse basin and her tributaries at the selected case study locations in Sect. 5.4.4 to align with task D.T5.4.1.

2 Study area

The Meuse basin upstream of St Pieter, at the border between Belgium and the Netherlands, covers an area of approximately 21,300 km^2 in France and Belgium and can be divided in three main zones. The first zone is the French Southern part of the basin and is characterized by thick soil layers, broad valleys bottoms and gentle slopes, underlain by sedimentary consolidated rock from the Middle and Late Jurassic. The second zone includes thin soils on relatively impermeable Cambrian metamorphic rock and Early Devonian sandstone that dominate the steeper and relatively high Ardennes Massif in Belgium. Finally, the third zone is on the West bank of the Meuse in Wallonia, characterized by porous chalk layers with deep groundwater systems (L. Bouaziz, 2021). Elevation in the basin ranges between 50 and 700 m above mean sea level. Land use in the basin consists of 35% forest, 32% agriculture, 21% pasture and 9% urban areas (European Environment Agency, 2018).

The Meuse is a rain-fed river with relatively short response times. Streamflow in the tributaries can rise quickly (sub-daily timescale) during floods due to the steep slopes and impermeable soils of the Ardennes. The strong streamflow seasonality with low summer and high winter flows reflects the seasonality of potential evaporation, as precipitation is relatively uniformly distributed throughout the year. At St Pieter, observed streamflow has been found to vary between 20 m^3 s^{-1} to approximately 3,300 m^3 s^{-1} (maximum hourly observed), with a mean annual flow of around 250 m^3 s^{-1}.

The selection of specific points of interests shown in this report are based on discussions with the partners and stakeholders involved in the workshops held on December 1st, 2022 and June 30th, 2023. They involved inputs from Établissement Public d'Aménagement de la Meuse et de ses Affluent (EPAMA) for the French catchments, Service Public de Wallonie (SPW) for the Belgian catchments, Wasserverband Eifel-Rur for the Rur (Roer in Dutch) catchment, Waterschap Limburg for the Geul catchment. Moreover, Borgharen, now St Pieter[1], was mentioned as a key measurement site by Rijkswaterstaat for the Netherlands. Table 2-1 lists all locations deemed of importance by the involved stakeholders. The specific catchments for each location are shown in Figure 2-1 and Figure A-1.

For the Netherlands (for Rijkswaterstaat and the four Regional Water Authorities along the Meuse river) and Belgium (for Vlaamse Waterweg), the daily Meuse discharges at St Pieter are particularly important for the assessment and design of flood defences along the Meuse river (Chbab, 2017; Schweckendiek & Slomp, 2018; Slomp et al., 2016). St Pieter is the main station for the design criteria for the Netherlands and Belgium and a key station for operational forecasting purposes. The contributions from the main four tributaries in the Netherlands, the Rur, de Geul, the Niers and the Dommel river, are also of importance. For our partners in Belgium and Germany information on the tributaries themselves was considered particularly valuable. For these smaller catchments, hourly data are needed to represent extreme discharge dynamics.

[1] The measurement location was moved from Borgharen to St Pieter. It is important for the Assessment and Design of flood defenses in the Netherlands and Belgium (Flanders).

Table 2-1: Identified catchments of interests from the partners and stakeholders (ordered from small to large catchment areas)

River	Location	Country	Catchment area (km^2)	Results shown in
Rur	Monschau	Germany	147	Main report
Geul	Hommerich	The Netherlands	151	Appendix
Geul	Meerssen	The Netherlands	338	Main report
Meuse	Goncourt	France	364	Main report
Viroin	Treigne	Belgium	548	Appendix
Vesdre	Chaudfontaine	Belgium	683	Main report
Ambleve	Martinrive	Belgium	1,068	Appendix
Semois	Membre Pont	Belgium	1,226	Appendix
Lesse	Gendron	Belgium	1,286	Appendix
Ourthe	Tabreux	Belgium	1,607	Main report
Rur	Stah	Germany	2,152	Appendix
Meuse	St Mihiel	France	2,551	Appendix
Sambre	Salzinne	Belgium	2,842	Main report
Meuse	Chooz	France	10,120	Main report
Meuse	St Pieter	The Netherlands	21,233	Main report

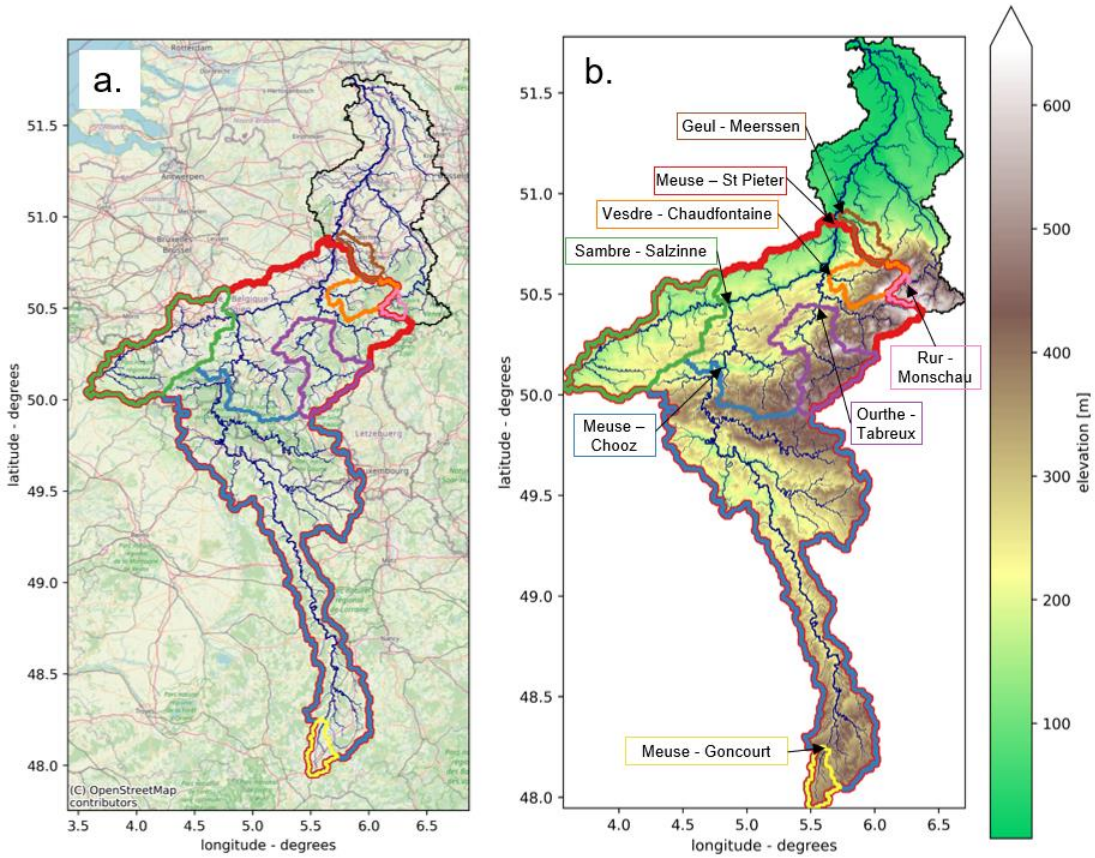


Figure 2-1: Selected catchments for the Meuse river and its tributaries, with background map (a) and topography (b).

3 Hydrometeorological data collection

This chapter describes the hydrometeorological data collected for this study and the discharge data. The description of the initial hydrological model for the Meuse can be found in (L. Bouaziz, 2020) and (L. Bouaziz & Buitink, 2022) and in section 4.1.3 for the static datasets. First, an overview of the application of the hydrometeorological datasets for this study is provided in Section 3.1. Section 3.2 and 3.3 describe the meteorological and discharge time series datasets in more detail.

3.1 Overview of precipitation and discharge datasets and their application

Table 3-1 presents an overview of the hydrometeorological data used in this study and specific application.

Table 3-1: Overview of the hydrometeorological data used in this study

Dataset name	Source	Used for
E-OBS	ECA&D	Climate forcing data used for historical model evaluation at the daily time step (Chapter 4)
GenRE	Bouaziz et al. (2020)	Climate forcing data used for historical model evaluation at the hourly time step (Chapter 4)
RACMO	KNMI	Climate forcing data used to obtain the synthetic discharge time series (Chapter 5)
Observed streamflow	SPW, EPAMA, Waterschap Limburg, LANUV NRW, Rijkswaterstaat	-Validation and calibration of the wflow model (Chapter 4) -Calculation of empirical return periods (Chapter 5) -General indication of the July 2021 discharge peak recorded or estimated (Chapter 5)
Reported statistics	SPW, EPAMA, Waterschap Limburg, LANUV-NRW, Rijkswaterstaat	Comparison with discharge frequency analysis obtained from synthetic events sets (Chapter 5)

3.2 Meteorological datasets

3.2.1 E-OBS

Daily historical meteorological data for the Meuse basin is available from the E-OBS gridded dataset (Cornes et al., 2018). The E-OBS dataset (v25.0e) is comprised of daily grids of precipitation, temperature and radiation for the period 1980-2021 at a resolution of 25 km^2, all used as forcing for the historical model evaluation of the hydrological model. The data are based on station data collected by the European Climate Assessment & Dataset (ECA&D) initiative. The resolution of the temperature grids is downscaled to the wflow model grid (~1km^2 resolution) using the digital elevation model and a fixed lapse rate of 0.0065°C m^{-1}. Potential evaporation is estimated using the Makkink formula (Hooghart & Lablans, 1988).

3.2.2 GenRE

The genRE dataset, described in Bouaziz et al. (2020), consists of hourly historical meteorological data for the Meuse basin upstream of St Pieter for the period 01-01-2005 to 31-12-2017, which is based on the interpolation of station data from the Dutch operational forecasting system RWSoS. It comprises hourly precipitation, temperature and potential evaporation calculated with the Makkink formula (Hooghart & Lablans, 1988).

The genRE interpolation is done using a climatological climate grid based on E-OBS dataset, following the method presented in van Osnabrugge et al. (2017).

3.2.3 RACMO

We refer the reader to van Voorst and van den Brink (2023) for a detailed description of the synthetic meteorological time series obtained from the regional climate model RACMO. In line with the report from van Voorst and van den Brink (2023), we refer to this dataset in this report as “the RACMO dataset”. Note that this dataset is also referred to as the RACMO’23 dataset in other reports. As summarized in van Voorst and van den Brink (2023), the RACMO dataset can be described as follows:

> “The GCM and RCM used in this study are respectively EC-EARTH3 (Döscher et al., 2022) and RACMO (van Meijgaard et al., 2008). In the context of the KNMI-23 climate scenarios (van Dorland et al., 2023) an ensemble of 16 climate simulations were generated with EC-EARTH3 with a resolution of 80x80 km2 and then dynamically downscaled to a 12x12 km2 grid covering Europe using RACMO. Simulations of different ensembles are required in order to reduce the contribution of natural variability in the climate change signal. For the time period of 1950 – 2120, 16 ensemble members are generated that correspond to ssp1.26, ssp2.45 and ssp5.85. Up to 2014, the 16 members are identical for every ssp scenario, and based on historical greenhouse gas emissions. After 2014, the 16 members will start deviating from the members of the other greenhouse gas scenarios.
>
> The primary objective of this research is to analyse extreme discharges with synthetic data in the present climate. Therefore, the meteorological model outputs from 1950 to 2014 for all 16 ensemble members have been used as the source data in this study, providing a dataset spanning a total of 1,040 years for all relevant meteorological variables. Precipitation data is accessible at an hourly time scale, whereas the other meteorological variables are available at 3-hourly time scale.”

3.3 Discharge datasets

3.3.1 Observed discharge

The continuous time series of discharge have been downloaded from the respective platforms of the different partners or shared by email. Table 3-2 provides some brief description of the source of each time series used at the selected catchment of interests.

Some of the partners had specific web portal to access the discharge information. Data from EPAMA was downloaded from the HydroPortail (https://hydro.eaufrance.fr/) web portal, for the Rur catchment from the Hochwasserportal.NRW portal (https://hochwasserportal.nrw/lanuv/webpublic) of LANUV (Landesamt für Natur, Umwelt und Verbraucherschutz) NRW (Nordhrein-Westfalen). Data from SPW was downloaded from https://hydrometrie.wallonie.be/home.html, for Rijkswaterstaat from https://waterinfo.rws.nl/. Data from Waterschap Limburg was received directly from the waterboard.

Table 3-2: Description of the observed discharge time series obtained for the catchment of interests.

River	Location	Source – station number
Rur	Monschau	LANUV NRW - 2821530000200
Geul	Hommerich	Waterschap Limburg - 10.Q.30
Geul	Meerssen	Waterschap Limburg – 10.Q.36
Meuse	Goncourt	EPAMA - B022 0010 01
Viroin	Treigne	SPW - 9021
Vesdre	Chaudfontaine	SPW - 6228SVC2
Ambleve	Martinrive	SPW - 6621
Semois	Membre Pont	SPW - 9434
Lesse	Gendron	SPW - 8221
Ourthe	Tabreux	SPW - 5921
Rur	Stah	LANUV NRW - 2829100000100
Meuse	St Mihiel	EPAMA - B222 0010 01
Sambre	Salzinne	SPW - 73290002
Meuse	Chooz	EPAMA - B720 0000 01
Meuse	St Pieter	Waterschap Limburg, Rijkswaterstaat and waterinfo.be (see next paragraph)

In this report, we refer to the Meuse at St Pieter to represent the Meuse discharge from the “unsplitted” Meuse, i.e. the combined discharge from the Meuse at St Pieter and the discharge flowing through the Albert Kanaal at Kanne. The data for the Albert Kanaal at Kanne is downloaded from waterinfo.be.

The discharge time series between July 1st and July 31st 2021 is extracted and further analysed to report an approximate estimate of the maximum peak discharge recorded or estimated during July 2021. The resulting time series, when available, are shown in Figure A-2 and Figure A-3 in Appendix. There are considerable uncertainties related to these estimates as most gauges broke or stopped recording during the event. For the Geul at Meerssen, van der Veen (2021) provided an estimate of 88 m^3/s for the maximum hourly discharge reached. Based on these figures, we decided not to use the maximum daily peak discharge for the Geul at Meerssen (deemed completely unreliable). 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).We use the largest value reported of 354 m^3/s as a rough estimate of the hourly discharge peak. 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). No hourly peak time series are available for the Rur at Monschau. When available, the maximum daily and hourly discharge peak value is stored and shown for reference in Chapter 5 for the frequency analysis. Note that these are shown for reference only as a first order estimate and can still deviate from the truly occurred maximum flow.

3.3.2 Reported discharge return levels

Reported discharge return levels have been downloaded from the partners' website or received per email at locations where those calculations have been made. Table 3-3 provides a summary of the reported statistics present at the catchment of interests.

Table 3-3: Description of official discharge return levels present at the catchment of interest

River – Location	Source	Return periods (years)	Extreme Value Family	Temporal resolution	Hydrological year
Rur – Monschau	-	-	-	-	-
Geul - Hommerich	Waterschap Limburg	10,25,50	unknown	Hourly or higher	unknown
Geul – Meerssen	Waterschap Limburg	1, 10, 25,	unknown	Hourly or higher	unknown
Meuse – Goncourt	EPAMA	2, 5, 10, 20, 50	Gumbel	Daily	01/09 – 31/08
Viroin – Treignes	SPW	25, 50, 75, 100	Exponential	Hourly	01/10 – 30/09
Vesdre – Chaudfontaine	SPW	25, 50, 75, 100	Weibull	Hourly	01/10 – 30/09
Ambleve – Martinrive	SPW	25, 50, 75, 100	Lognormal	Hourly	01/10 – 30/09
Semois-Membre Pont	SPW	25, 50, 75, 100	Gamma	Hourly	01/10 – 30/09
Lesse – Gendron	SPW	25, 50, 75, 100	Lognormal 2	Hourly	01/10 – 30/09
Ourthe – Tabreux	SPW	25, 50, 75, 100	Lognormal	Hourly	01/10 -30/09
Rur – Stah	Waterschap Limburg	1, 10, 25, 50, 100	unknown	Hourly or higher	unknown
Meuse – St Mihiel	EPAMA	2, 5, 10, 20, 50	Gumbel	Daily and Hourly	01/09 – 31/08
Sambre – Salzinne	SPW	25, 50, 75, 100	General Pareto	Hourly	01/10 – 30/09
Meuse – Chooz	EPAMA	2, 5, 10, 20, 50	Gumbel	Daily	01/09 – 31/08
Meuse – St Pieter	Hegnauer et al. (2014)	5,10,50,100, 250,500, 1250,4000, 10000, 100000	Gumbel	Daily	01/01-31/12

For SPW stations, the fitting of the frequency model is either made using the maximum likelihood or the method of moments. For SPW and EPAMA stations, the L-moment method is used.

4 Hydrological modelling

This chapter describes the hydrological modelling work. First, we present the distributed hydrological model wflow_sbm which is used in this study. We then describe previous work with the wflow_sbm model in the Meuse basin, as this forms the starting point of the hydrological model used in this study. Subsequently, we describe the improvements that were made to the model following discussions with the stakeholders during the two EM Flood Resilience workshops held in December 2022 and June 2023. Finally, we present the model performance of the hourly and daily model by comparing the modelled streamflow with observations.

4.1 The wflow_sbm Meuse model

4.1.1 The wflow framework

The distributed hydrological modelling software wflow is a free and open source hydrological modelling framework developed by Deltares. The wflow framework is designed to perform distributed hydrological simulations using GIS raster data (Van Verseveld et al., 2022). The model calculates hydrological state and flux variables at any given point in the model at a given time step, based on physical parameters and meteorological input data (precipitation, temperature, and potential evaporation). Over the last 10 years, the wflow framework has been successfully applied worldwide to evaluate flood hazards, droughts and the impact of climate and land use change on hydrological resources.

Wflow is a framework which includes different hydrological concepts, including the wflow_sbm model concept and the wflow_flextopo concept. Within the wflow_sbm concept, alternative options exist to compute the river and land flow routing, either based on the kinematic wave or the local inertial approximation (see further details on the lateral routing in Sect. 4.1.2.2).

From 2021, the wflow code is distributed under the MIT license (https://github.com/Deltares/Wflow.jl). Wflow is also available as a compiled executable maintained and distributed by Deltares. The wflow computational engine is built in the Julia programming language, which is a high-performance computing language. The wflow framework does not include a graphical user interface. Documentation is available online (https://deltares.github.io/Wflow.jl/dev/) and through a detailed discussion paper (Van Verseveld et al., 2022).

4.1.2 The wflow_sbm model concept

4.1.2.1 Vertical processes

The wflow_sbm model concept is the most widely used concept within the wflow framework and it is also used in this study. It belongs to the so-called physically-based class of hydrological models and is based on the lumped Topog_SBM model, which was developed by Vertessy et al. (1999). The model describes the most relevant hydrological processes, including glacier and snow processes, interception from the canopy, transpiration, soil and open water evaporation, infiltration into the soil considering the fraction of (un)paved area, a representation of the saturated and unsaturated store with exchanges between both through recharge and capillary rise. When the soil is saturated or when precipitation exceeds the infiltration capacity, overland flow is generated. Exfiltration from the saturated store occurs when the saturated store reaches the surface.

Deep groundwater losses from the saturated store (leakages) can be included in the model. The current application of the wflow_sbm model for the Meuse includes all these processes, except for the glacier module. The main processes are schematized in Figure 4-1 and a detailed description of the equations is available in van Verseveld et al. (2022).

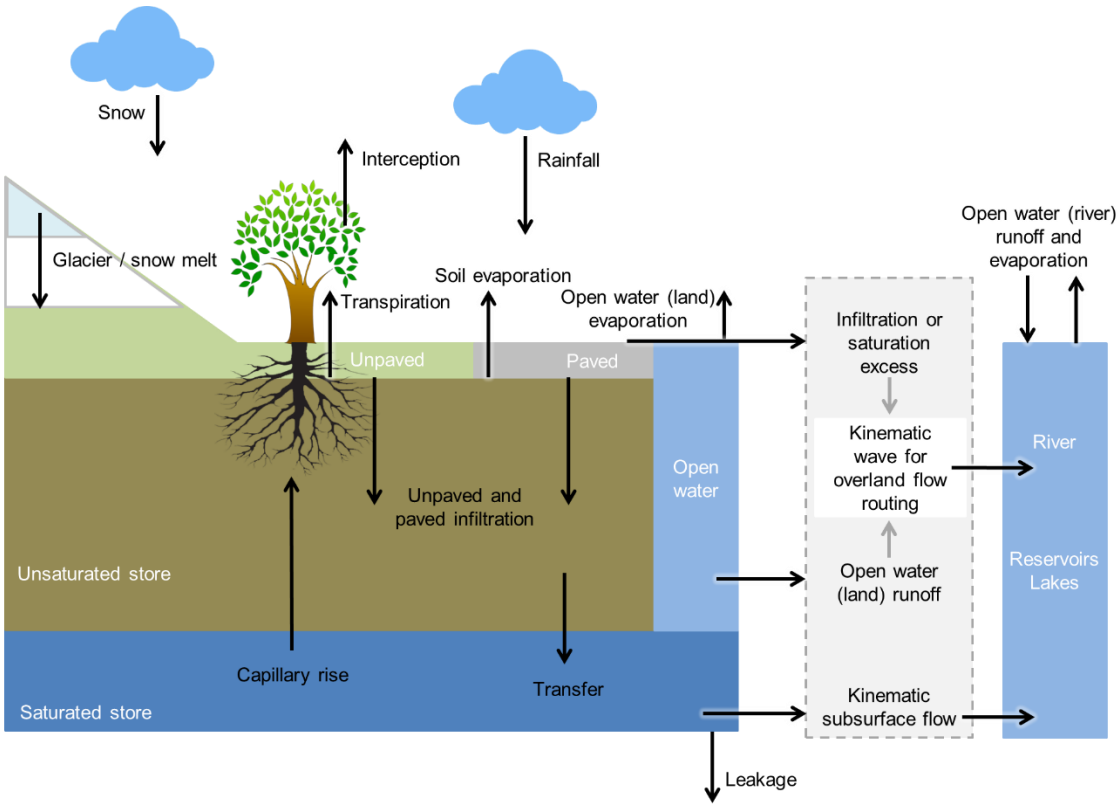


Figure 4-1 Schematic representation of the wflow_sbm concept

4.1.2.2 Lateral processes

Water in the river, the subsurface and on land is transported downslope through the catchment along the river network. The kinematic-wave approach is used for lateral subsurface flow. Several options are possible for the lateral routing of river and overland flow, including the kinematic wave or local inertial approximation, with the possibility of accounting for river overbank flow and floodplain routing. The four options for the lateral routing of river and overland flow are summarized in Table 4-1 and are detailed below. The current wflow code does not allow to mix multiple lateral routing concepts in a single model schematization. In the next section, we show how the choice of lateral routing concept influences the hydrological response in the Meuse basin, which includes both areas with relatively steep slopes and areas with large floodplains.

Table 4-1 Overview of different routing concepts available in the wflow modelling framework.

Routing concept	Features	Typical application region
Kinematic wave (river routing and overland flow)	Flow driven by topography No backwater effects No flood inundation	Areas with relatively steep slopes
Local inertial (river routing) + kinematic wave (overland flow)	Flow driven by hydraulic gradient Backwater effects No flood inundation	Areas with limited slopes and without large floodplains
Local inertial (river routing, 2D floodplain routing and overland flow)	Flow driven by hydraulic gradient Backwater effects Flood inundation in 2D	Areas with limited slopes and floodplains
Local inertial (river and 1D floodplain routing) + kinematic wave (overland flow)	Flow driven by hydraulic gradient Backwater effects Flood inundation in 1D	Areas with limited slopes and floodplains

Kinematic wave

The kinematic wave equation assumes that topography controls the water flow by assuming the energy gradient is always equal to the bottom gradient. The kinematic wave approach is a simplification of the Saint-Venant equations. As the flow propagation is only in the downstream direction as opposed to both upstream and downstream directions in the Saint-Venant equations.

Due to the assumed static nature of the energy slope, the kinematic wave approach is only described by the continuity equation and a uniform-flow equation. The underlying assumption is that the friction slope equals the bed slope. The local inertia, convective inertia and pressure difference in time terms of the shallow-water equation are assumed to be negligible. The discharge is assumed to be a function of the vertical stream depth only. The kinematic wave does not attenuate as it propagates downstream. Therefore, backwater effects are not considered. Flood waves on steep slopes are adequately described by the kinematic wave model because the assumption of the energy gradient is close to reality. However, in low relief areas, the use of the kinematic wave approximation may be difficult due to the need for a monotonically decreasing riverbed elevation and the presence of strongly diffusive flow regimes (Neal et al., 2012).

As the kinematic wave routing has the fastest computational runtime across the different routing options and as it was the first implementation of lateral routing, it has most often been applied.

Local inertial river routing and kinematic wave for overland flow (no inundation in the floodplains)

The local inertial approximation of the shallow water equation only neglects the convective inertia term in the Saint-Venant equation (De Almeida & Bates, 2013). In contrast to the kinematic wave, where the flow is driven by topography, it is the time and space variable hydraulic gradient that controls the flow in the local inertial approximation. Attenuation and backwater effects are possible.

The local inertial approximation of the shallow water approximation is recommended for less steep areas. To prevent instability in steep areas, flow velocities are reduced to have Froude numbers below or equal to 1, as is done in (Adams et al., 2017).

Local inertial 1D2D river routing (with inundation in the floodplains)

To simulate river hydraulics and floodplain inundation, the routing in wflow is extended with a local inertial approximation of the shallow water equation, 1D sub-grid for the river channel, and 2D for the floodplain, similar to (Neal et al., 2012).

The representation of river hydraulics and floodplain inundation should lead to a more realistic representation of flow routing and floodplain inundation processes, especially in relatively flat areas, and therefore more realistic peak flow simulations. However, this concept is the most computationally demanding, and results in longer model run times compared to the other two options. The 1D2D wflow_sbm model can be applied at each model resolution. Model resolutions of ~100m will give therefore more accurate representations of the floodplain geometry as compared to model resolutions of ~1000m, which are more typical resolutions for basin scale modelling.

Local inertial river and floodplain routing (1d schematization of the floodplains)
As an alternative to the 1D2D approach, another approach was implemented in the local inertial routing of the wflow code early 2023, which consists of representing the floodplains in 1D to speed up calculation times and improve the representation of the geometry of the floodplains in the model schematization. The 1D floodplain schematization is based on provided flood volumes as a function of flood depth (per flood depth interval) for each river cell. Wflow calculates from these flood volumes a rectangular floodplain profile for each flood depth interval. Routing is done separately for the river channel and floodplain.

4.1.2.3 Lakes and reservoirs

Wflow includes a simple representation of reservoirs and lakes as part of the river network. 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 simulates the reservoir behavior by keeping the reservoir volume between the target minimum and maximum fill fraction as derived from these databases, while accounting for the (environmental) discharge demand downstream of the reservoir(s). Although this gives a good first estimate of the reservoir and lake dynamics, it does not consider the day-to-day dynamics of especially reservoir management that can consist of usage for: hydropower, drinking water, flood and drought protection. Generally, this means that water levels are adjusted based on forecast water levels and that a seasonal pattern exists where the reservoir refills during wetter (winter) seasons and net releases water during drier (summer) seasons. This also influences simulated peak and low flows.

4.1.3 Previous developments of the wflow_sbm Meuse model

A first set-up of a wflow_sbm model for the Meuse basin using global data and pedotransfer functions, as proposed by Imhoff et al. (2020), was made by Bouaziz (2020). The model was set-up in WGS84 (EPSG:4326) projection at a resolution of approximately 1 km x 1 km (0.008333°) for the Meuse basin upstream of Mook in the Netherlands (Figure 2-1). The model was developed by Deltares for Rijkswaterstaat for applications relating to operational forecasting, flood protection and climate adaptation. To clarify, the process of this model setup (described more elaborately by Bouaziz, 2020) is briefly described here.

The model was set-up using HydroMT model builder (Eilander et al., 2023), a Python toolbox developed by Deltares that prepares the input to the required resolution and data format for the wflow_sbm model. The datasets used to set-up the model are:

- The MERIT Hydro adjusted Elevations of 3 arc seconds (~90 m) database (Yamazaki et al., 2019) was used as digital elevation model (DEM) to derive topographic information such as the slope, the drainage direction, and the stream order network. This dataset has been specifically developed for hydrological applications. An upstream area threshold of 25 km^2 was applied for river initiation.

- The Soilgrids database (Hengl et al., 2017) includes soil properties information of soil, clay, silt, organic carbon content, pH, and bulk density at several depths (0cm to 200 cm below the surface). This dataset, at a resolution of 250 m x 250 m, allows us to estimate the soil related parameters of wflow_sbm (soil hydraulic conductivity, porosity, residual water content etc.) by making use of pedotransfer functions (Imhoff et al., 2020).
- The CORINE land cover dataset is a European land cover classes map (European Environment Agency, 2018) with a spatial resolution of 500 m x 500 m. The landcover information is used in wflow_sbm to make an estimation of land cover related parameters to represent interception processes or the surface roughness. A look-up table based on literature values relates land cover classes to parameter values. The look-up table is available in the HydroMT tool (https://github.com/Deltares/hydromt_wflow/blob/main/hydromt_wflow/data/lulc/corine_mapping.csv).
- The dataset of Lin et al. (2020) contains river width and bankfull discharge estimates at the global scale based on a machine learning algorithm in combination with satellite observations. This dataset can be used to estimate the bankfull river depth, using a power law (Andreadis et al., 2013). However, data are often lacking for small tributaries and in Bouaziz and Buitink (2022), a method was applied to extrapolate the river width and depth based on a linear relation with the upstream area for the smaller tributaries.

All maps prepared by HydroMT, either related to model schematization from the DEM (catchment delineation, stream network and characteristics, slope and flow direction) or parameters derived from soil and land use data, were first calculated at the original fine resolution datasets before being resampled and aggregated to the model resolution (ca. 1,000 m x 1,000 m).

The wflow_sbm model of the Meuse was initially set-up at the daily time step using E-OBS data as meteorological input data and the kinematic wave for river flow, overland flow and subsurface lateral routing (L. Bouaziz, 2020). The daily timestep model is sufficient to provide extreme discharge frequencies at St Pieter, which is the key station for the Netherlands and Belgium. However, for the partners and stakeholders focusing on the smaller tributaries, a model at the hourly timestep is required due to the faster response time.

In Bouaziz and Buitink (2022), the Meuse wflow_sbm model was further developed and improved through an extensive analysis of different routing concepts: kinematic wave, local inertial and local inertial with 1D2D overland flooding. In addition, Bouaziz and Buitink (2022) performed a calibration of the parameter related to the horizontal hydraulic conductivity (*KsatHorFrac*) and the snow parameters. The *KsatHorFrac* parameter is used to calculate the horizontal hydraulic conductivity from the vertical hydraulic conductivity ($K_{horizontal} = K_{vertical} * K_{satHorFrac}$). This parameter has a large impact on the partitioning between baseflow and peaks. The main idea was to improve the low flow model performance through the calibration of the *KsatHorFrac* parameter and to improve the high flow model performance through an improved representation of the routing concept. The original model with kinematic wave routing tended to overestimate maximum annual peak flows and it was hypothesized that model performance could be improved by adapting the routing concept to account for attenuation of the flood peak through overbank flow and floodplain routing. The results of Bouaziz and Buitink (2022) indeed confirm that both the simulations of the lowest and highest flows improved with the local inertial 1D2D routing concept. However, computational demand is high for the local inertial 1D2D routing concept and some artifacts occurred for smaller tributaries. Additionally, as the model was set-up at a ~1000 m x 1000 m resolution, it is likely that the floodplain volume is overestimated.
The report therefore recommends testing the (at that time, yet to be implemented) routing concept of the local inertial with 1D floodplain schematization to further improve model performance and reduce computational demand.

The wflow_sbm Meuse model, which was developed and calibrated at the daily time step for stations upstream of St Pieter at the Dutch-Belgian boarder, was also evaluated at the hourly time step, using the GenRE data as meteorological input data. Overall model performance was relatively similar between the daily and the hourly timestep model, except for the streamflow performance in the Belgian Ardennes, which showed an overestimation of maximum and minimum annual flows (L. Bouaziz & Buitink, 2022). For the Meuse at St Pieter, the daily and the hourly timestep models showed high performance indicators (Nash-Sutcliffe efficiency of the daily flow and logarithm of the daily flows and Kling-Gupta efficiency of above 0.75).

When local reservoir information is available or when operation rules are known, this can be included in wflow_sbm to better mimic the reservoir behavior. However, for most reservoirs in the Meuse (there are no natural lakes in the model), sufficient local data was not present in this study to do so. Appendix section 7C gives an exploration of including more local information to better simulate the reservoir behavior of the reservoirs Eupen and La Gileppe in the Vesdre basin. Nevertheless, the reservoir operation rules are known for the river Rur and can be added to the model. Hartgring (2023) tested this for the Rur by including a stage-discharge relationship, relating the reservoir outflow to the simulated reservoir levels, following predefined reservoir operation rules as used by the Wasserverband Eifel-Rur. This approach for the reservoirs of the Rur tributary has been added to the wflow_sbm Meuse model in this study. All other reservoirs are modelled according to the aforementioned standard approach in wflow_sbm.

The reservoirs included in the wflow_sbm Meuse model are:

- The reservoirs of the Eau d'Heure, Plate Taille and Val Joly in the Sambre catchment
- The reservoirs of the Gileppe, Eupen and Butgenbach in the Vesdre and Ambleve catchments
- The reservoirs of the Olef, the Rur (which combines the URfttalsperre and Vortalsperre as one single reservoir) and the Wehebach in the Rur catchment.

This model was presented at the first workshop of the Interreg EM Flood Resilience project, in December 2022, as being the starting point of the hydrological modelling work for the current project. At the workshop, we presented the wflow model, the general approach, and objectives of the project. Together with the stakeholders, we discussed that further improvements of the model within the Interreg project could focus on testing an additional lateral routing concept with 1D schematization of the floodplains and an improved calibration of the model at the hourly time step. In addition, we discussed that the model performance for the Geul and the Rur rivers could be improved by integrating the optimized models developed in the master theses of two Delft University of Technology students (Hartgring, 2023; Klein, 2022).

4.2 Improvements to the wflow_sbm Meuse model

After the first workshop of the Interreg EM Flood Resilience project in December 2022, and based on the discussions and feedback received by the partners and stakeholders, we further improved the wflow_sbm Meuse hydrological model. This was done within the scope of this project.

Improvements were made in the following order:

- Additional manual calibration step of the daily model to improve the overall performance of the model (mainly in terms of mean monthly flows and cumulative flows at the station of St Pieter at the Dutch-Belgian boarder)

- Testing the performance of the wflow_sbm with local inertial approximation with 1D floodplain schematization for the lateral routing of river + floodplain flow
- Automatic calibration of the hourly model to improve the overall performance of the model at the hourly time step. This was especially important considering the overestimation of maximum and minimum peak flows in the Ardennes.
- Integrating an optimized model for the Geul and the Rur catchments as developed in the master theses work of two Delft University of Technology (TU Delft) students into the larger Meuse model.

It is important to note that the first three steps were performed for the Meuse basin upstream of St Pieter (as direct follow-up of the previous developments done in the Rijkswaterstaat projects and because the hourly genRE dataset only covers the basin area upstream of St Pieter). The optimization of the Geul and the Rur catchment (which are downstream of St Pieter) are therefore seen as separate steps.

Table 4-2 Descriptive requirements to improve the model and action undertaken to implement this requirement in chronological order

Requirement from discussed feedback / own insights	**Action to meet this requirement**
Improve the overall performance (mean monthly flows and cumulative flows) of the daily model at St Pieter and for the tributaries upstream of St Pieter (to start with long-time scale improvements before going into the hourly time scale)	Manual calibration of the daily model
Increase computational efficiency and model run times and improve floodplain geometry schematization by implementing and testing the 1D floodplain approach instead of the 1D2D floodplain approach of the local inertial routing.	Implement the 1D floodplain schematization in the wflow code Test the model performance with the 1D floodplain local inertial routing.
Automatic calibration of the hourly model for the Meuse upstream of St Pieter to improve the model performance in the Belgian Ardennes for the hourly time step.	Automatic calibration of the hourly model for a selection of parameters. Test model performance with the new calibration for the hourly and daily timestep model and compare with observations.
Integrate the calibration effort made by two students of the TU Delft for the Geul and the Rur catchments in the larger Meuse model. It is good to note that the Geul and Rur catchments are not included in the previous steps as the first two steps were direct follow-ups of previous GRADE work and the hourly forcing genRE data only includes the Meuse area upstream of St Pieter.	Implement the calibrated Geul and Rur models in the larger Meuse model.

Each of these steps (as also summarized in Table 4-2) are described in more detail in the sections below.

4.2.1 Manual calibration of the daily model

The daily wflow_sbm model of the Meuse as delivered in Bouaziz and Buitink (2022) has a high overall performance. However, mean monthly flows from October to March and cumulative flows show an overestimation compared to observations for the larger basin of the Meuse at St Pieter. The overestimation of the streamflow cannot be solved through a different routing concept or through the calibration of the *KsatHorFrac* parameter as these parameters do not control the longer-term water balance within the model. The main parameter to modify the partitioning between streamflow and evaporation is the rooting depth as it controls the amount of transpiration from the vegetation at timescales in the order of several months.

The rooting depth in wflow_sbm is typically estimated through a look-up table which links a land use class to a specific rooting depth. Alternatively, estimates of the rooting depth at the catchment scale can be obtained using observed hydrometeorological data, assuming that the ecosystem controls its root zone through hydrometeorological constraints in a method called the mass curve technique (L. J. E. Bouaziz et al., 2022a; de Boer-Euser et al., 2016; Wang-Erlandsson et al., 2016). The main assumption behind the method is that vegetation optimizes its root zone water storage so that it is sufficiently large to overcome typical dry spells, much like human size dams to sustain droughts (Gao et al., 2023). Instead of relying on the look-up table approach for the rooting depth, we updated the model with the climate-based method for the rooting depth estimates (using the genRe dataset and observed streamflow).

The sensitivity of two other parameters, the maximum leakage term, and the soil thickness, was also evaluated in the manual calibration of the daily wflow_sbm model to further improve the baseflow model performance. The initial estimate of the soil thickness is provided by the Soilgrids dataset (Hengl et al., 2017). The dataset is restricted to a maximum soil thickness of 2000 mm. However, the French part of the basin is known to have thick soils, which are predominantly rock deposits from the Jura (de Wit, 2008). During this period, northwest Europe was below sea level and thick packages of limestone and calcareous sand and claystone were formed. The baseflow performance of the model in combination with the adapted rooting depth estimates substantially improved when the soil thickness in the French part of the basin (and in several limestone or chalk underlain subcatchments in Belgium) was multiplied by a factor two.

In several catchments of the Meuse basin, mainly underlain by karstic rock, deep groundwater losses may occur, affecting the overall water balance (L. Bouaziz et al., 2018). The *MaxLeakage* parameter within wflow_sbm can be used to explicitly account for these losses, it represents the amount of water leaking out of the domain from the saturated store. A first estimate of the leakage term is done by evaluating the observed long-term water balance components (evaporation index $E_A/P = (P - Q_{\mathrm{obs}})/P$ and dryness index E_P/P, where P is precipitation, Q_{obs} is observed streamflow, E_P is potential evaporation and E_A is actual evaporation) of the catchments of the Meuse using observed hydrometeorological data.

The changes applied to these three parameters positively affected the water balance and resulted in an improved model performance in reproducing the cumulative streamflow and the monthly discharge regime for the Meuse at St Pieter and the tributaries.

4.2.2 Routing

Early 2023, a new feature was implemented in the wflow code in the context of the current EM Flood Resilience Interreg project, which is the possibility to represent floodplains in 1D in the local inertial approximation lateral routing concept. This new routing concept considerably reduces the run times of the model (approximately 3x faster than the local inertial with 1D2D routing). Moreover, it results in similar results as the 1D2D local inertial routing concept (which had the best results so far), but in a substantial peak attenuation compared to kinematic wave and local inertial without floodplain routing, as shown for the Meuse at St Pieter in Figure 4-2 (top left panel). The choice of the routing concept mainly influences the maximum annual peak flows, while mean monthly flows and lowest flows are almost unaffected, see Figure 4-2 for the Meuse at St Pieter.

In steeper catchments, such as the Ourthe at Tabreux in the Belgian Ardennes (Figure 4-3), the difference between routing concepts for maximum annual peak flows, minimum flows and monthly flows is much less than in the larger and flatter Meuse basin at Chooz (Figure 4-4). This is explained by the fact that flood routing in steeper areas is adequately described by the kinematic wave. Moreover, maximum annual flows with the highest return periods are mostly affected because then a relatively large fraction of the flow conveyance is through the floodplain compared to the streamflow for lower return periods.

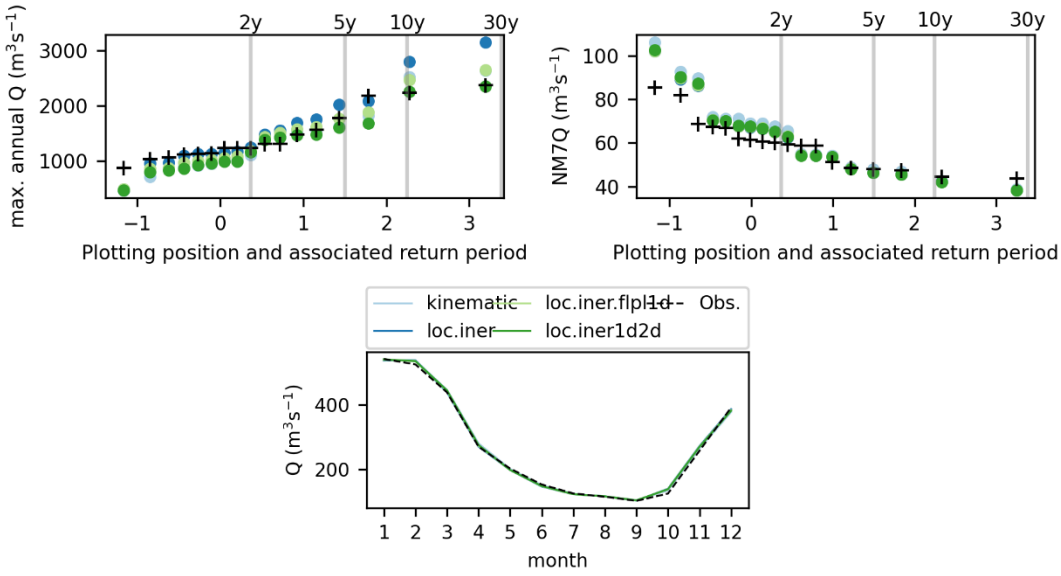


Figure 4-2 For the Meuse at St Pieter: Difference in model performance (top-left: maximum annual peaks, top-right: minimum annual 7-days discharge NM7Q, bottom: mean monthly flow regime) for the model run with four different options for the lateral routing (kinematic wave, local inertial, local inertial with 1D floodplains and local inertial with 1D2D floodplains). In the legend, the following abbreviations are used: kinematic for kinematic wave (light blue), loc.iner for local inertial without floodplain flow (dark blue), loc.iner.1d2d for local inertial with 1D2D floodplains (dark green) and loc.iner.flp1d for local inertial with 1D floodplains (light green).

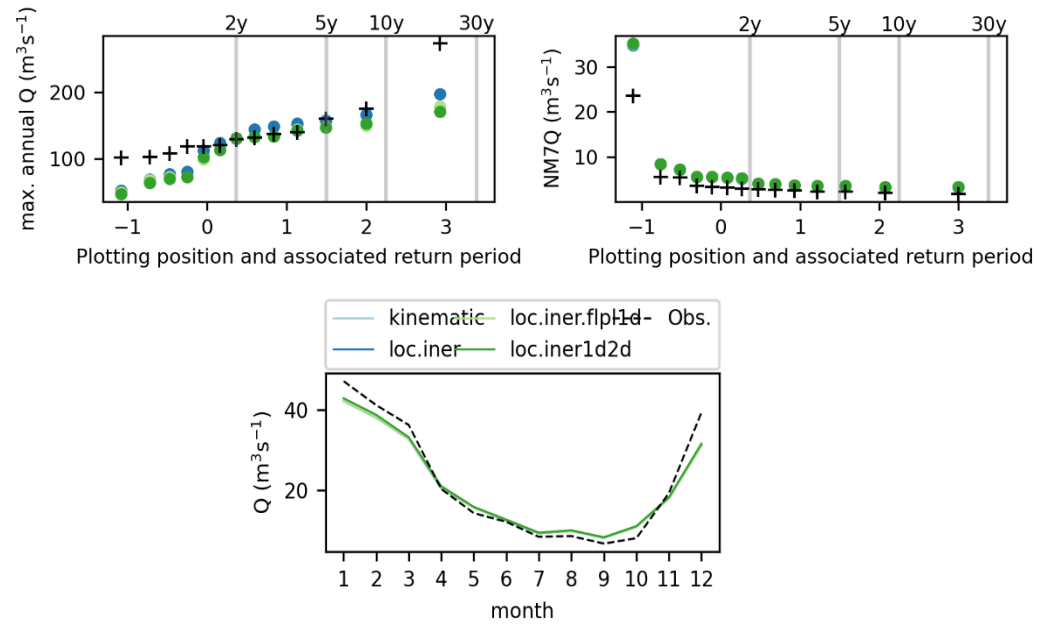


Figure 4-3 For the Ourthe at Tabreux: Difference in model performance (top-left: maximum annual peaks, top-right: minimum annual 7-days discharge NM7Q, and bottom: mean monthly flow regime) for the model run with four different options for the lateral routing (kinematic wave, local inertial, local inertial with 1D floodplains and local inertial with 1D2D floodplains). In the legend, the following abbreviations are used: kinematic for kinematic wave (light blue), loc.iner for local inertial without floodplain flow (dark blue), loc.iner.1d2d for local inertial with 1D2D floodplains (dark green) and loc.iner.flp1d for local inertial with 1D floodplains (light green).

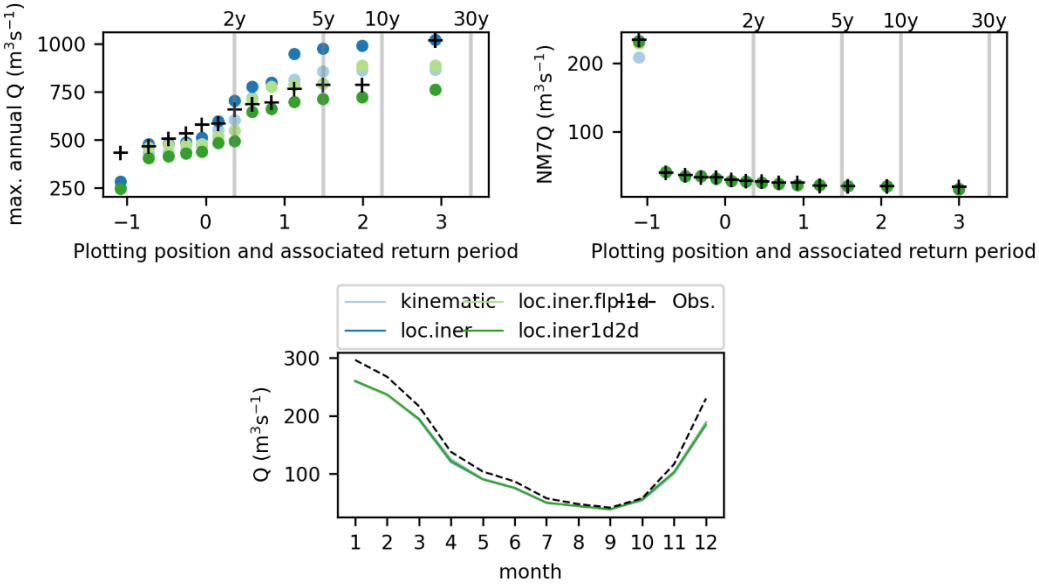


Figure 4-4 For the Meuse at Chooz: Difference in model performance (top-left: maximum annual peaks, top-right: minimum annual 7-days discharge NM7Q, and bottom: mean monthly flow regime) for the model run with four different options for the lateral routing (kinematic wave, local inertial, local inertial with 1D floodplains and local inertial with 1D2D floodplains). In the legend, the following abbreviations are used: kinematic for kinematic wave (light blue), loc.iner for local inertial without floodplain flow (dark blue), loc.iner.1d2d for local inertial with 1D2D floodplains (dark green) and loc.iner.flp1d for local inertial with 1D floodplains (light green).

Based on these results, it was decided to use the local inertial routing with 1D floodplain schematization in the remainder of the improvements since it leads to significantly faster runs and still a reasonable model performance. Therefore, in the subsequent section, the calibration of the model at the hourly timestep is performed with this routing scheme.

4.2.3 Automatic calibration of the hourly model

To further improve the performance of the model at the hourly timestep, a more extensive automatic calibration at the sub-basin scale is performed using the genRE dataset as meteorological forcing data. The calibration period is set to 2005-01-01 to 2011-12-31, using the year 2005 as spin-up time. This time period represents the first half of the full length of the genRE dataset (2005-2017). The model is subsequently evaluated for the full record length of the genRE dataset, so from 2006-01-01 to 2017-12-31 (therefore also including the calibration period).

A selection of sensitive model parameters is changed according to a predefined set of values (Table 4-3), which is based on expert knowledge and previous experience with the Meuse model. Due to the relatively long run time of the model, it is not possible to perform an extensive Monte Carlo calibration with a very large number of samples. All combinations of the predefined parameter variations result in 1134 combinations of parameter sets.

The selection of calibration parameters is based on the sensitivity of the parameter on specific aspects of model performance:

- The rooting depth controls the partitioning of precipitation to drainage and evaporation. In general, a larger rooting depth should result in more water to transpiration. Although the rooting depth has been estimated using observed hydrometeorological data (in the manual calibration described in Sect. 4.2.1), there is still uncertainty in our estimates and a variation around that estimate could potentially result in improved parameter estimation with better model performance.
- The *KsatHorFrac* parameter is used to calculate the horizontal hydraulic conductivity from the vertical hydraulic conductivity ($K_{horizontal} = K_{vertical} * K_{satHorFrac}$). This parameter has a large impact on the partitioning between baseflow and peaks. Low values of *KsatHorFrac* result in a flashy hydrograph, while high values result in a more damped hydrograph dominated by groundwater processes. The vertical hydraulic conductivity is estimated using Soilgrids data and pedotransfer functions, but no pedotransfer function exists for the $K_{satHorFrac}$ parameter, nor for the $K_{horizontal}$, it is therefore a logical parameter to further calibrate.
- The soil thickness has a substantial impact on the baseflow, with thicker soils resulting in higher baseflows. Increasing the soil thickness in the French part of the basin resulted in an improved model performance (in the manual calibration described in Sect. 4.2.1). In contrast, the reduction of the soil thickness in the Belgian Ardennes may result in an improved performance for the currently overestimated lowest flows.
- Deep groundwater losses are important in areas with karstic geology where the water balance may be affected by deep groundwater losses (L. Bouaziz et al., 2018). The importance of this parameter was also found in the manual calibration (see Sect. 4.2.1) and a further evaluation is performed in this automatic calibration step.
- Interception controls how much water infiltrates in the soil. It is interesting to note that the interception module of the wflow_sbm model differs between the daily and the sub-daily timestep. While the Gash module is used for the daily model, the Rutter module is used for the hourly timestep (for more details on the equations, see van Verseveld et al. (2022)), which results in lower interception rates. This is why the *Swood* parameter, which represents the storage on the woody part of the vegetation is also included in the calibration.

- Surface roughness controls the shape of the hydrograph, as the surface roughness increases, the hydrograph is more smoothed. The influence of the surface roughness on model performance is also evaluated in this automatic calibration step.

For each of the 1134 runs, the model performance is evaluated for 14 French stations and 80 stations from the Service Public de Wallonie, using the following performance indicators:

- Nash-Sutcliffe Efficiency (Nash & Sutcliffe, 1970) of daily streamflow (NSE Q)
- Nash-Sutcliffe Efficiency of the logarithm of daily streamflow (NSE logQ)
- Nash-Sutcliffe Efficiency of mean monthly streamflow (NSE monthly regime)
- Nash-Sutcliffe Efficiency of cumulative streamflow (NSE cumulative)
- Nash-Sutcliffe Efficiency of minimum annual flows over a moving average of 7 days (NSE NM7Q)

The Euclidean distance which combines each of these performance indicators was subsequently calculated to summarize the performance in a single value to select the best performing parameter set in a multi-objective calibration.

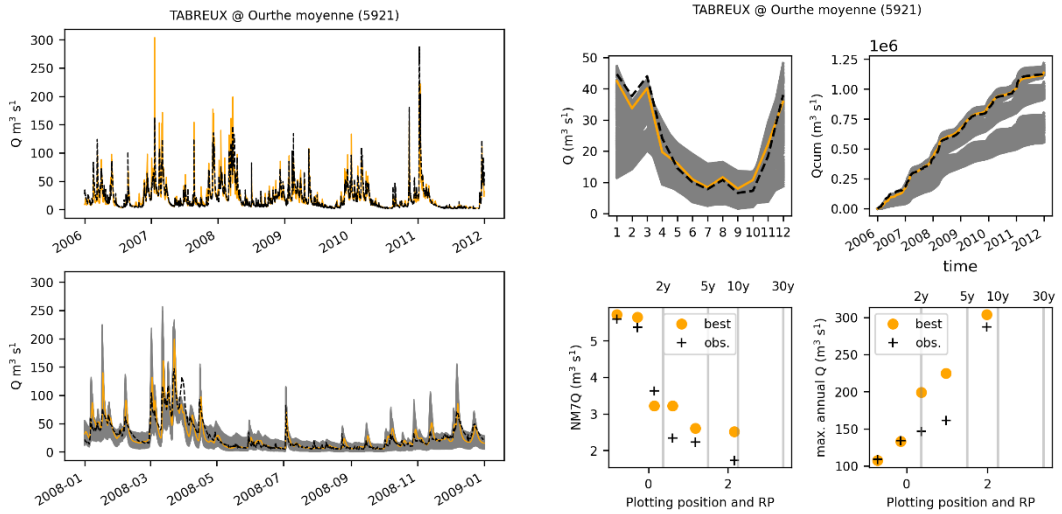
We initially also tested including a specific indicator for the annual maxima. However, as the calibration period is relatively short, we found that when this indicator was included, the chosen optimal parameter set did not seem to be the right visual choice in terms of cumulative flows and mean monthly flows for several catchments. Therefore, it was decided to leave the indicator out of the optimization to ensure a good overall performance of the model for low, average and high streamflow regimes.

Table 4-3 Parameter change table for the automatic calibration of the hourly wflow_sbm model

Parameter	**Initial estimate**	**Multiplication factor or offset or fixed value**
Soil thickness [mm]	Soil thickness from the Soilgrids database	Multiplication factors [-] applied to the Soilgrids map: 0.5, 1.0, 2.0
Multiplication factor to determine the horizontal conductivity from the vertical conductivity ($K_{horizontal}$ = $K_{vertical}$ * $K_{satHorFrac}$) [-]	Map with KsatHorFrac value per catchments based on the calibration performed in Bouaziz and Buitink (2022)	Multiplication factors [-] applied to initial map: 0.3, 0.5, 0.7, 1.0, 1.5, 2.0, 3.0.
Rooting depth [mm]	Estimates based on hydrometeorological data at the subcatchment scale, based on Bouaziz et al., (2022b)	Multiplication factors [-] applied to initial map: 0.8, 1.0, 1.2
Deep groundwater losses / leakage [mm d^{-1}]	Uniform value of 0 mm d^{-1}	Fixed values tested [mm d^{-1}] 0.0, 0.2, 0.6
Storage on the woody part of the vegetation within the interception module [mm]	Estimate based on the land use map and look-up table	Offset [mm] applied of : 0.0, 2.0

Parameter	Initial estimate	Multiplication factor or offset or fixed value
Manning roughness of land and river routing [s $m^{-1/3}$]	Estimate based on the land use map and look-up table for the land routing Estimate based on streamorder for the river routing (same multiplication factor was applied simultaneously for both land and river)	Multiplication factors [-]: 0.7, 1.0, 1.5

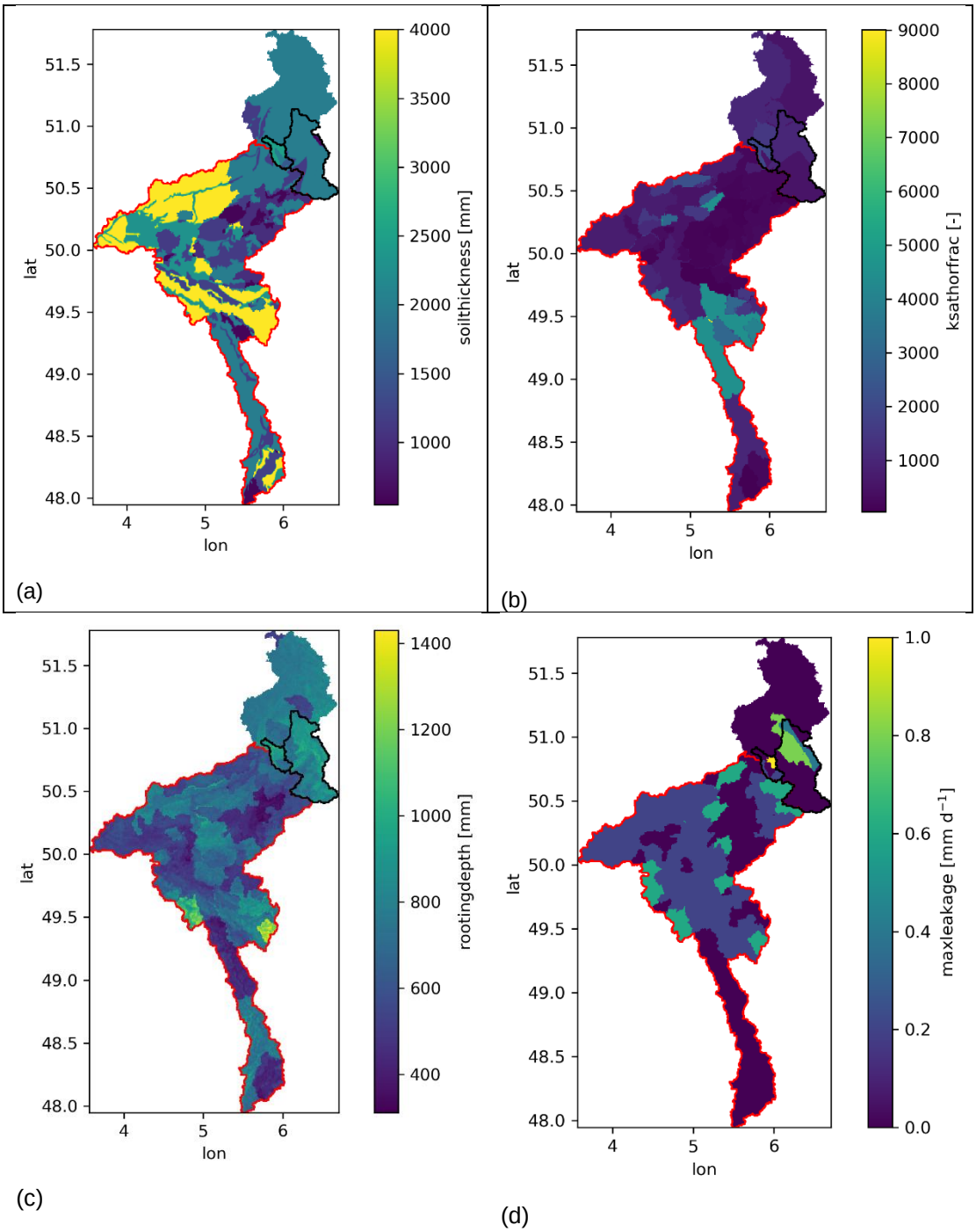
An illustrative example of the calibration results is shown in Figure 4-5 for the catchment of the Ourthe at Tabreux. All other stations were also visually inspected.

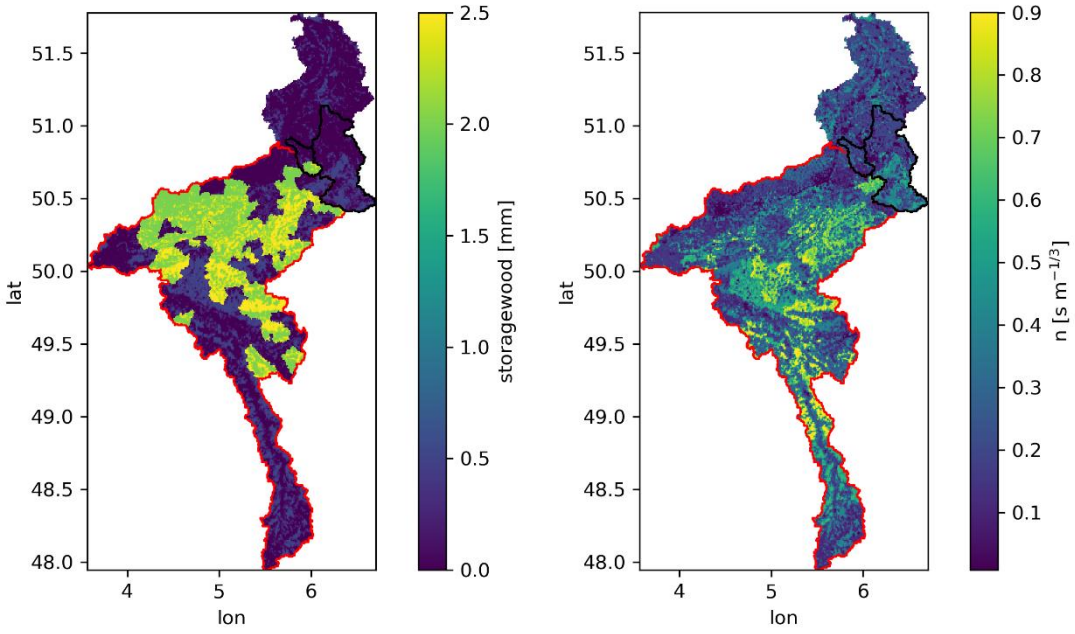


(a) (b)

Figure 4-5 Calibration results for the station of the Ourthe at Tabreux. (a) The model performance is evaluated through a visual inspection of the hydrograph (upper panel show the full calibration period and lower panel zooms in an illustrative year). (b) In addition, several signatures are shown, top left of panel (b): the modelled and observed mean monthly flows, top right: cumulative flows, bottom left and right: minimum and maximum annual flows are plotted. The grey lines show the full range of parameter sets, the black line shows the observations, and the orange line shows the best performing parameter set based on the defined selection criteria.

The resulting parameter maps after calibration are shown in Figure 4-6.





(e) (f)

Figure 4-6 Maps resulting from the calibration (a) soil thickness [mm], (b) ksathorfrac [-], (c) rooting depth [mm], (d) maximum leakage [mm d^{-1}], (e) wooded storage part of the vegetation Swood [mm] and (f) Manning n roughness [s m$^{-1/3}$]. The red border shows the calibrated area of the Meuse upstream of St Pieter. The black polygons show the Geul and the Rur parameter values after implementing the Geul and the Rur model as described in the next section.

The performance of the model at the daily and hourly timestep after the automatic calibration are shown in Sect. 4.3.

4.2.4 Geul and Rur models

For the Geul and the Rur catchments, we implement the extensive calibration work done by two master students of the Delft University of Technology. They extensively tested and developed a wflow_sbm model for the Geul (Klein, 2022) and for the Rur (Hartgring, 2023). Their calibration effort was implemented in the larger Meuse model.

Geul

The thesis work of Klein (2022) aims to understand the hydrological response of the Geul catchment to the extreme event of July 2021. She performed an extensive data analysis of precipitation, streamflow and groundwater and set up a model to reproduce the event and understand the contributions from the different tributaries and test the influence of antecedent moisture conditions. She adjusted three parameters in the model to be able to reproduce the floods and the responses of the subcatchments.

The following parameters are estimated and calibrated in the thesis of Klein (2022):

- The soil thickness (*SoilThickness*) is updated based on the International Hydrogeological Map of Europe (IHME, (Günther, 2023)) with values of 1.5 m, 2 m, and 2.5 m.

- The maximum leakage (*MaxLeakage*) in the Geul is considered per subcatchment and an estimate is obtained using the long-term water balance equation:

$$\Delta S/\Delta t = P - E - Q + Q_{\text{leakage}} = 0$$

This resulted in values of 0, 0.2 and 1 mm day^{-1} throughout the catchment.

- The parameter *KsatHorFrac* which links the vertical flow to the horizontal flow was calibrated from January 2020 to June 2020 based on the NSE. The short calibration period is due to the lack of data availability.

These changes in parameters are applied only to the Geul catchment within the larger Meuse model.

Rur/Roer
In his thesis work, Hartgring focuses on the prediction of the July 2021 event for the Rur catchment. Important aspects of his thesis include the improvement of the schematization of reservoir operations in the Rur catchment, setting-up and calibration of a wflow model and a hydrodynamic model.

Based on the work of Hartgring (2023), we implemented the following changes for the parameters within the Rur subcatchment of the larger Meuse model:

- Three reservoirs are simulated as lakes in the Rur model: the Rurtalsperre (combined with the Urfttalsperre and Vortalsperre as one single reservoir), the Oleftalsperre and the Wehebachtalsperre. The lake outflow function and the lake storage function parameters are updated accordingly. It is interesting to note that Hartgring improved the wflow_sbm code of the reservoir/lake schematization allowing to implement more detailed operational rules to better represent reservoir operations. This highlights the added value of having an open-source modelling framework, which can be improved based on user needs.
- The *KsatVer* parameter, describing the vertical saturated hydraulic conductivity, is updated using local information from the BK50 soil map.
- The *KsatHorFrac* parameter is calibrated for the period of 01-10-2011 to 30-09-2020 resulting in values of 1000 and 500 instead of the uncalibrated 1000 model value throughout the whole catchment.
- The wflow river width and the bankfull river depth are updated using the cross-sectional data from the Protection Measures against Inundation Decision Support (ProMaIDes, https://promaides.h2.de/promaides/) hydrodynamic model.

Lignite mining lowers the groundwater table affecting the water balance. To account for this, the maximum leakage (*MaxLeakage*) parameter was derived from the Rurscholle groundwater model resulting in values of 0.4 and 0.8 mm day^{-1} throughout the catchment This corresponds to two areas of 300-350 Mm3 yearly leakage which is similar to the 250 Mm3 reported by the mining company.

4.3 Historical model evaluation

The model performance is evaluated for the model at the daily and the hourly time step for a selection of stations on the Meuse river and its tributaries across the riparian countries, including:

- Rur at Monschau (Germany) – observations only available for the daily model
- 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)

These stations were selected based on discussions with the partners and stakeholders: EPAMA for the French stations, Service Public de Wallonie for the Belgian stations Wasserverband Eiffel-Rur for the German stations and Rijkswaterstaat and Waterschap Limburg for the Dutch stations.

Several additional stations are shown in the Appendix after discussions with the stakeholders:

- Geul at Hommerich (Netherlands)
- Viroin Treigne (Belgium)
- Ambleve at Martinrive (Belgium)
- Semois at Membre Pont (Belgium)
- Lesse at Gendron (Belgium)
- Rur at Stah (Germany) – observations only available for the daily model
- Meuse at St-Mihiel (France)

For each station, we perform a visual inspection of the hydrograph and of several signatures of the hydrograph, including:

- observed versus modelled daily streamflow,
- observed versus modelled mean monthly streamflow,
- observed versus modelled flow duration curves of daily Q,
- observed versus modelled flow duration curves of daily log(Q),
- observed versus modelled maximum annual streamflow,
- observed versus modelled minimum 7-days annual streamflow,
- observed versus modelled plotting positions of maximum annual streamflow,
- observed versus modelled plotting positions of minimum 7-days annual streamflow,
- observed versus modelled cumulative flow,
- NSE, NSE logQ and KGE performance indicators.

This enables us to have an overview of model performance for several aspects, including low, average, and high flows.

4.3.1 Daily

The results of the model performance at the daily timestep for the different stations are shown in Figure 4-7 to Figure 4-22. The main findings per station are summarised in Table 4-4. 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 4-4 Summary of daily 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
Rur at Monschau	147	overestimation of lowest flows	underestimation of winter monthly flows and cumulative flows	underestimation of high flows
Geul at Meerssen	338	good	overestimation of winter flows and cumulative flows	good
Meuse at Goncourt	364	good	Good (but slight overestimation of cumulative flows)	Good (but underestimation for the highest flows)
Vesdre at Chaudfontaine	683	underestimation of low flows	underestimation of monthly flows and cumulative flows	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	underestimation of winter flows and cumulative flows	good
Meuse at St Pieter	21233	good	good	good

Table 4-5 Summary of daily model performance based on a scores of the performance indicators

	NSE	KGE	NSElog	NM7Q ([2])	MAXQ ([2])
Rur at Monschau	0.69	0.59	0.74	-19.48	-0.09
Geul at Meerssen	0.33	0.54	0.28	0.39	-0.01
Meuse at Goncourt	0.75	0.85	0.77	-3.28	0.63
Vesdre at Chaudfontaine	0.72	0.68	0.72	-0.17	0.3
Ourthe at Tabreux	0.85	0.87	0.88	0.95	0.91
Sambre at Salzinnes	0.82	0.79	0.79	0.82	0.7
Meuse at Chooz	0.84	0.75	0.85	0.85	0.42
Meuse at St Pieter	0.91	0.96	0.91	0.93	0.87

[2] 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.

The following figures show the visual results for the different locations of interest. The figures show the hydrograph with a number of zoomed in interesting periods, as well as aggregated views such as the duration curve, climatological behaviour and behaviour of extreme (low and high) values. The visual inspection shows that the 2011 event seems to be underestimated by the model in all catchments shown below.

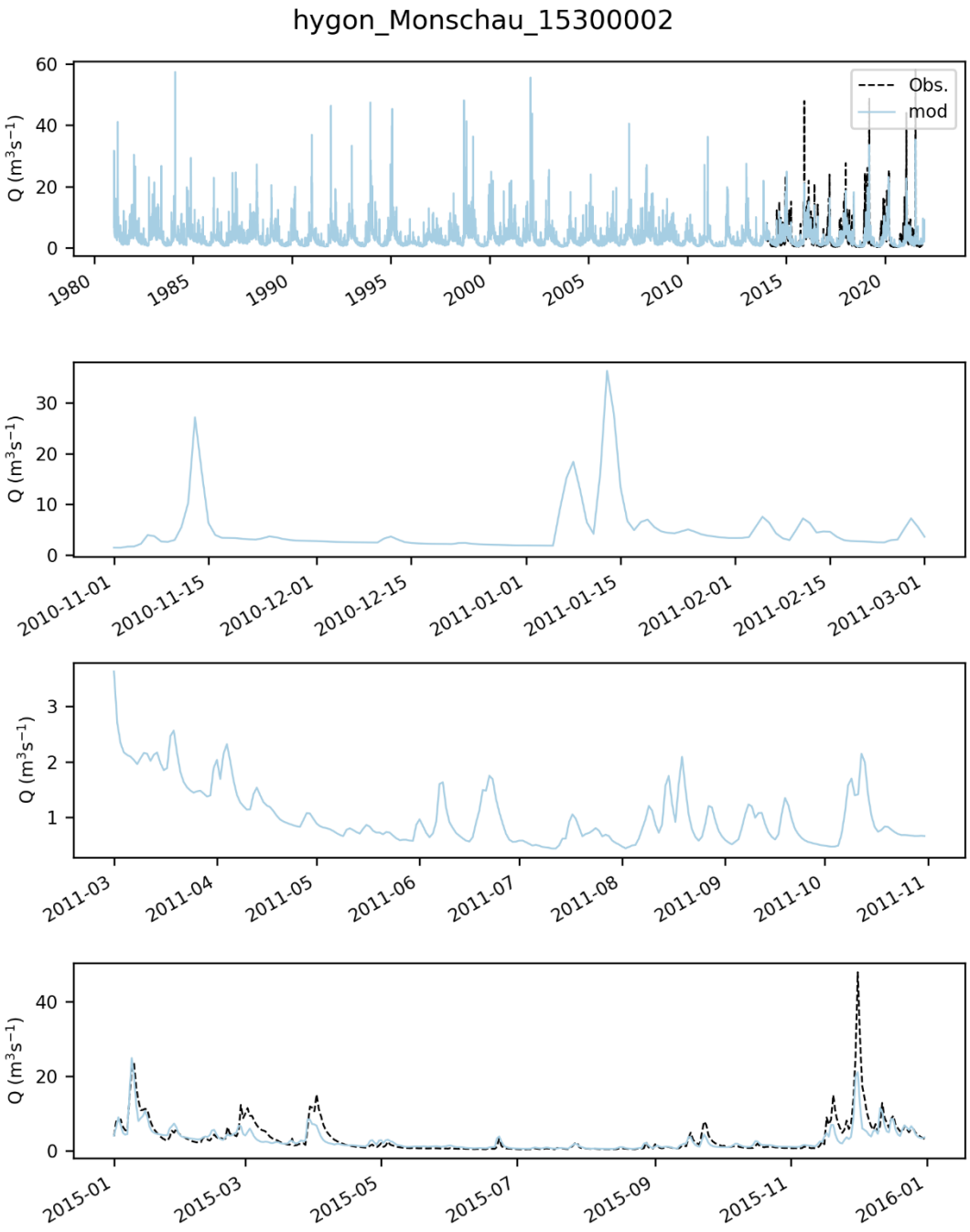


Figure 4-7 Modelled (blue) and observed (black) hydrographs for the Rur at Monschau.

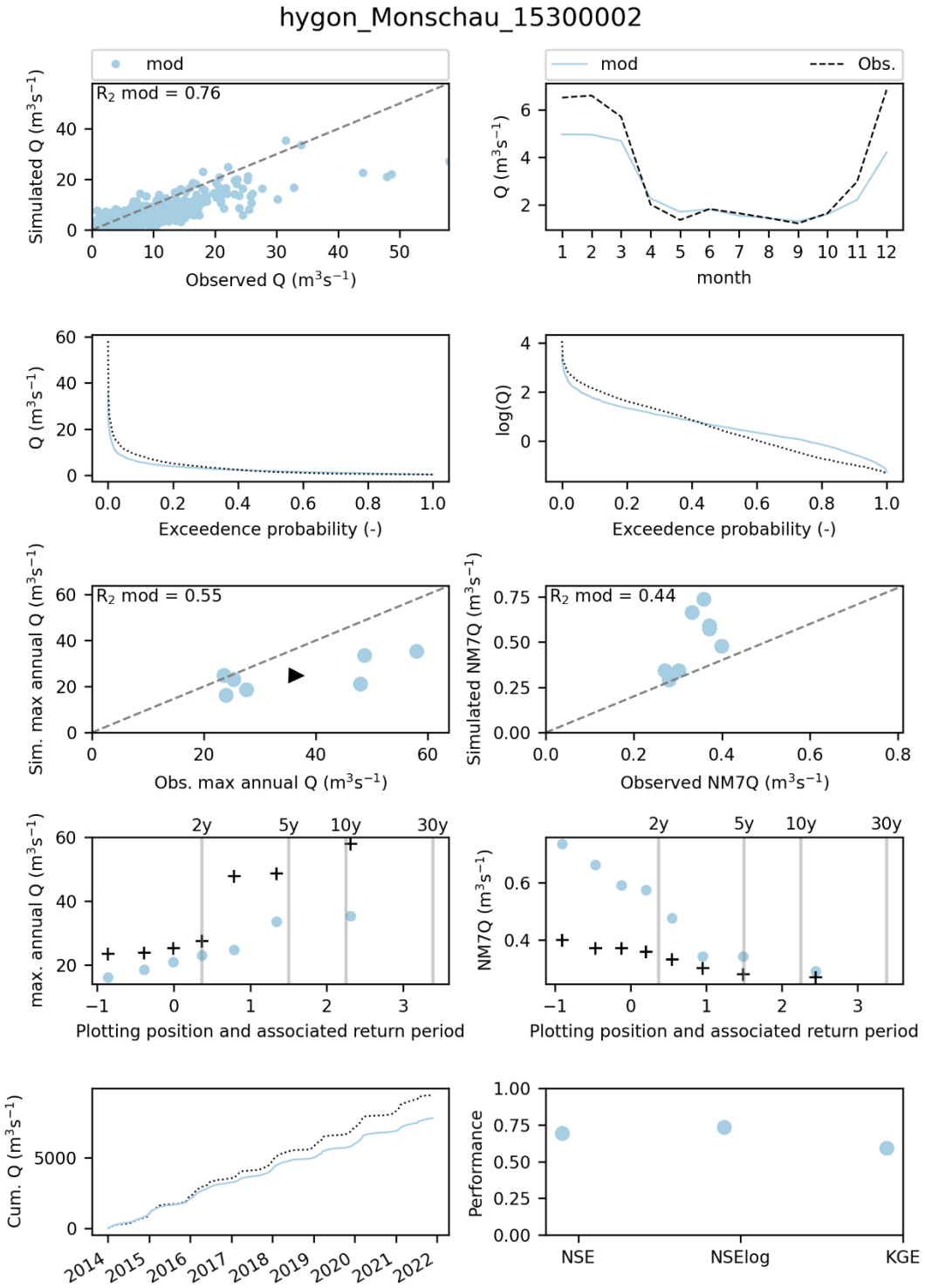


Figure 4-8 Modelled (blue) and observed (black) signatures for the Rur at Monschau.

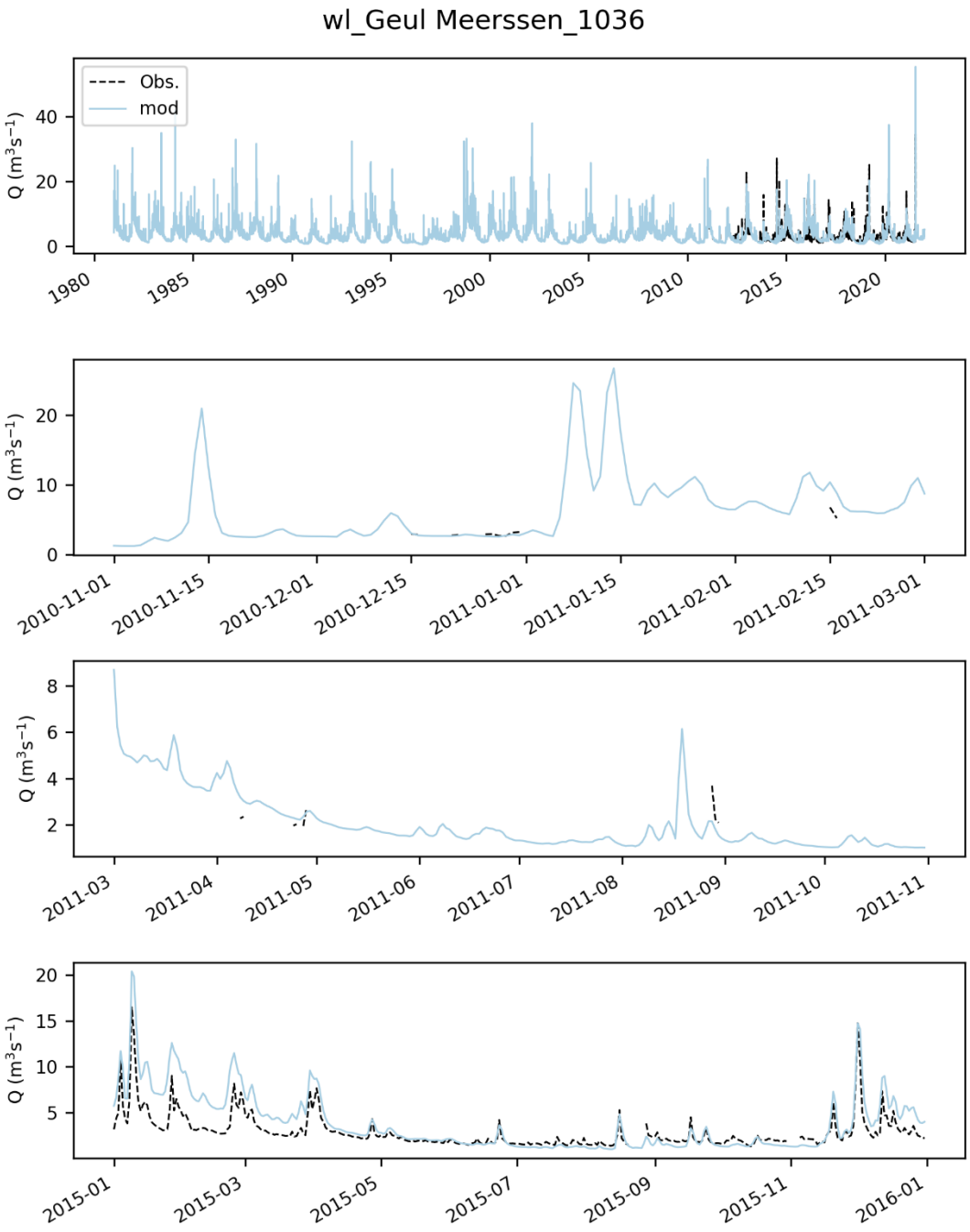


Figure 4-9 Modelled (blue) and observed (black) hydrographs for the Geul at Meerssen.

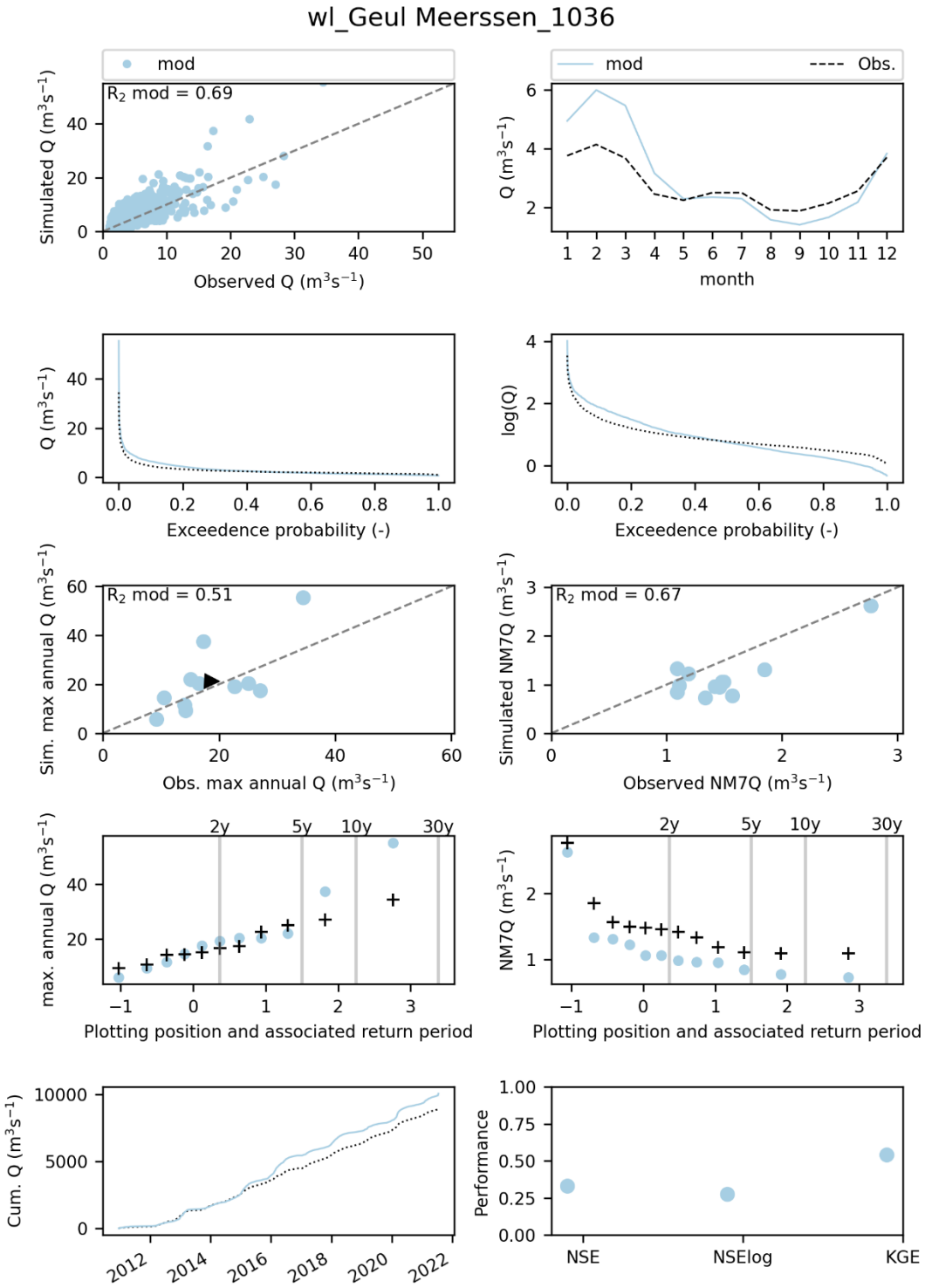


Figure 4-10 Modelled (blue) and observed (black) signatures for the Geul at Meerssen.

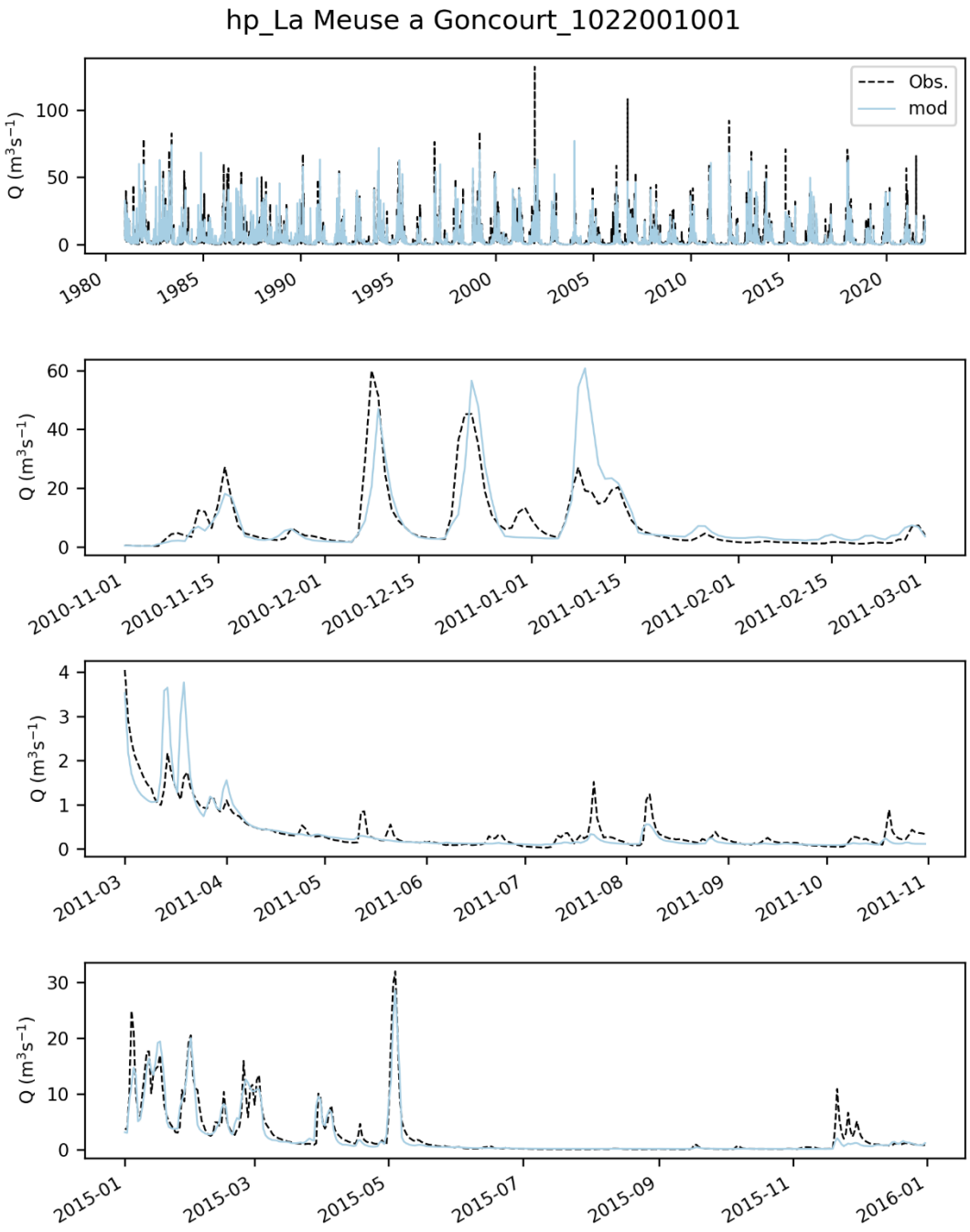


Figure 4-11 Modelled (blue) and observed (black) hydrographs for the Meuse at Goncourt.

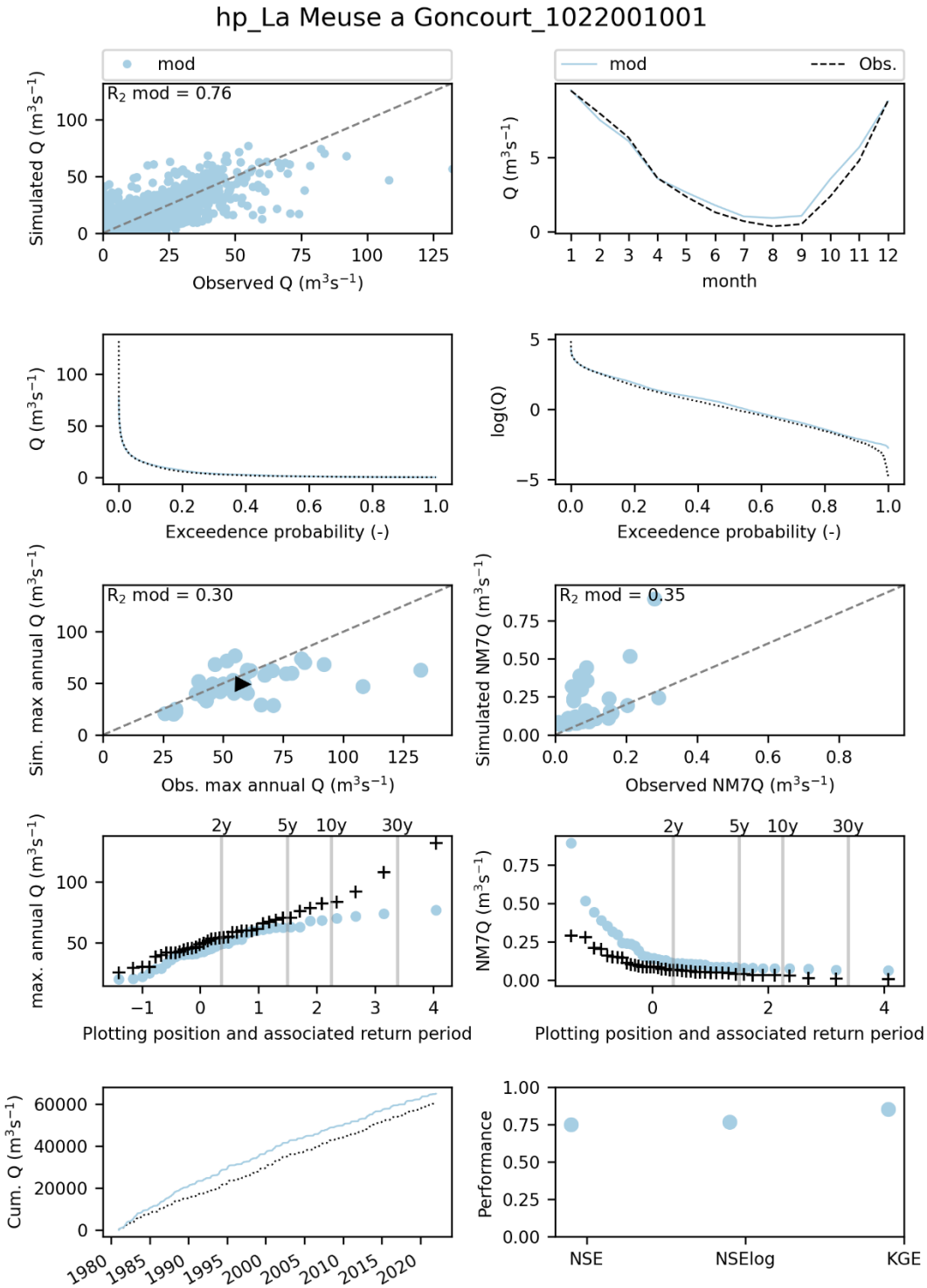


Figure 4-12 Modelled (blue) and observed (black) signatures for the Meuse at Goncourt.

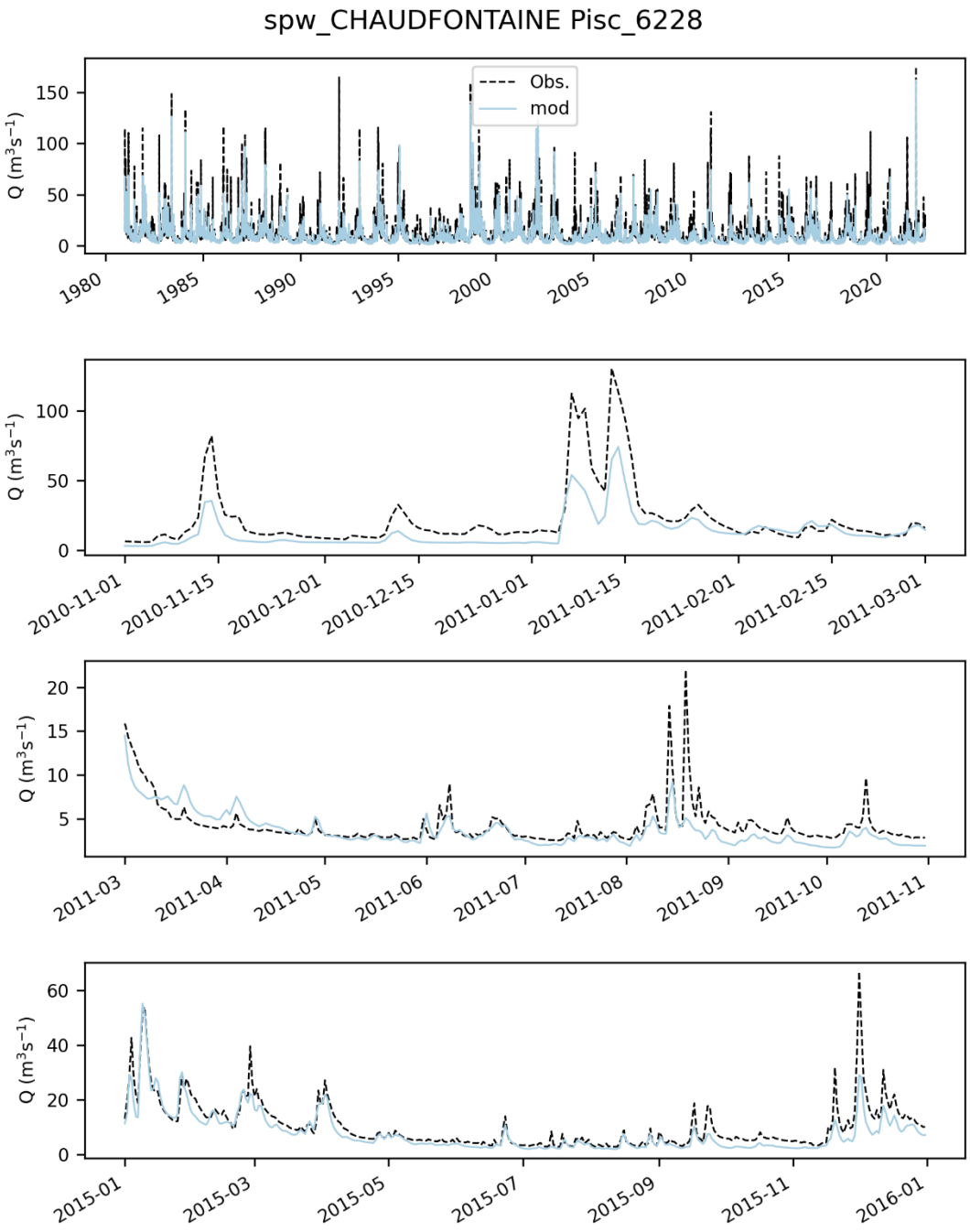


Figure 4-13 Modelled (blue) and observed (black) hydrographs for the Vesdre at Chaudfontaine.

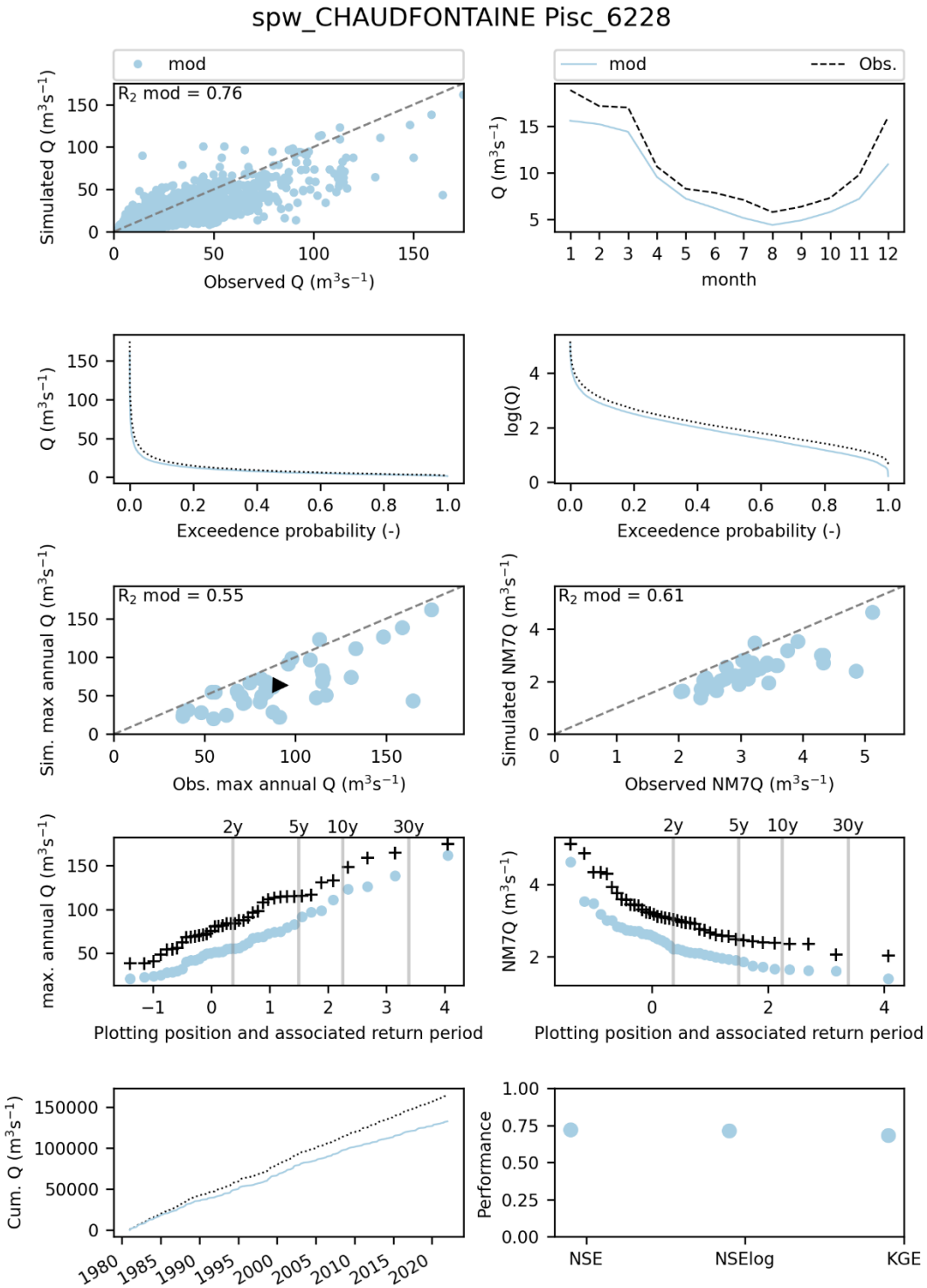


Figure 4-14 Modelled (blue) and observed (black) signatures for the Vesdre at Chaudfontaine.

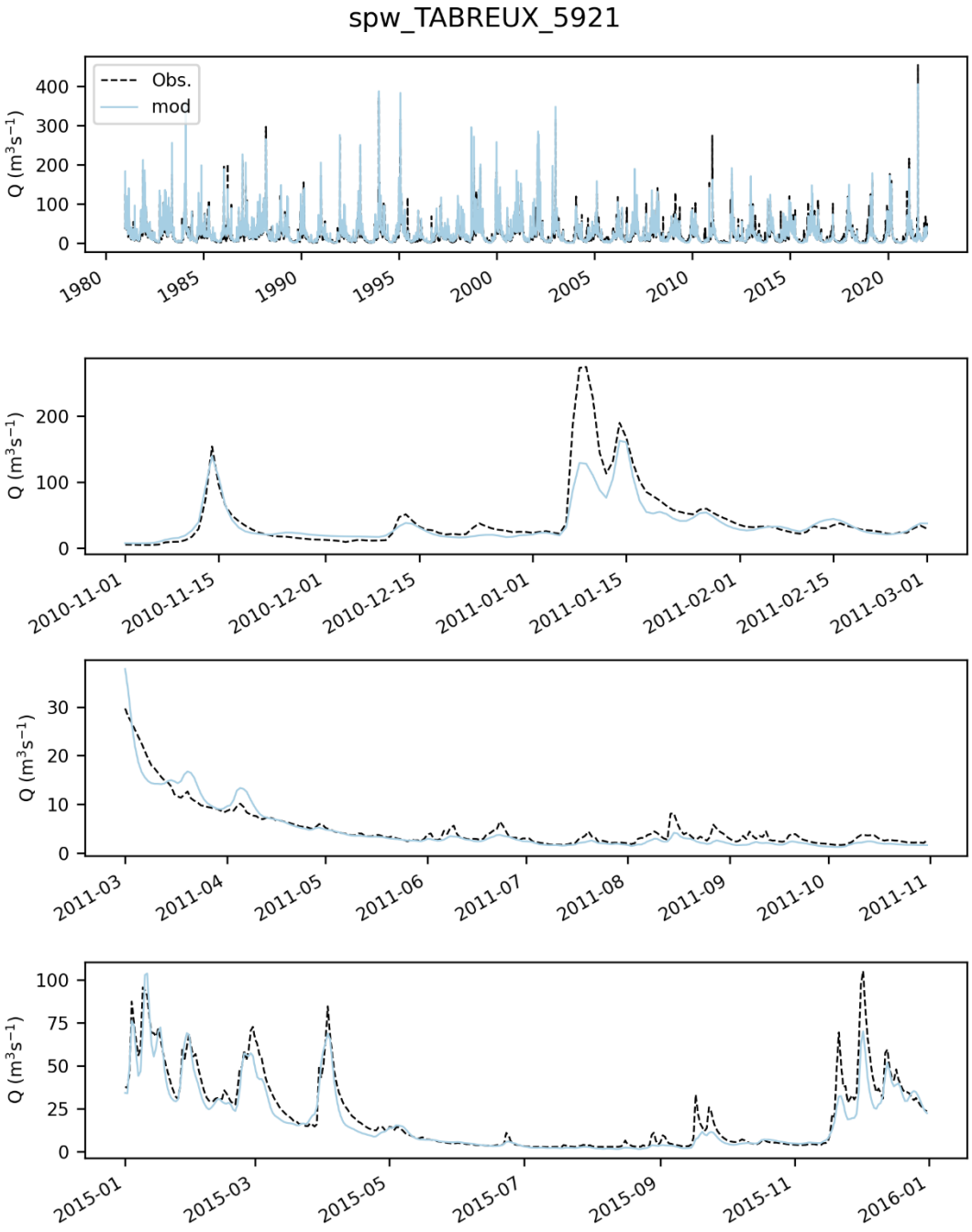


Figure 4-15 Modelled (blue) and observed (black) hydrographs for the Ourthe at Tabreux.

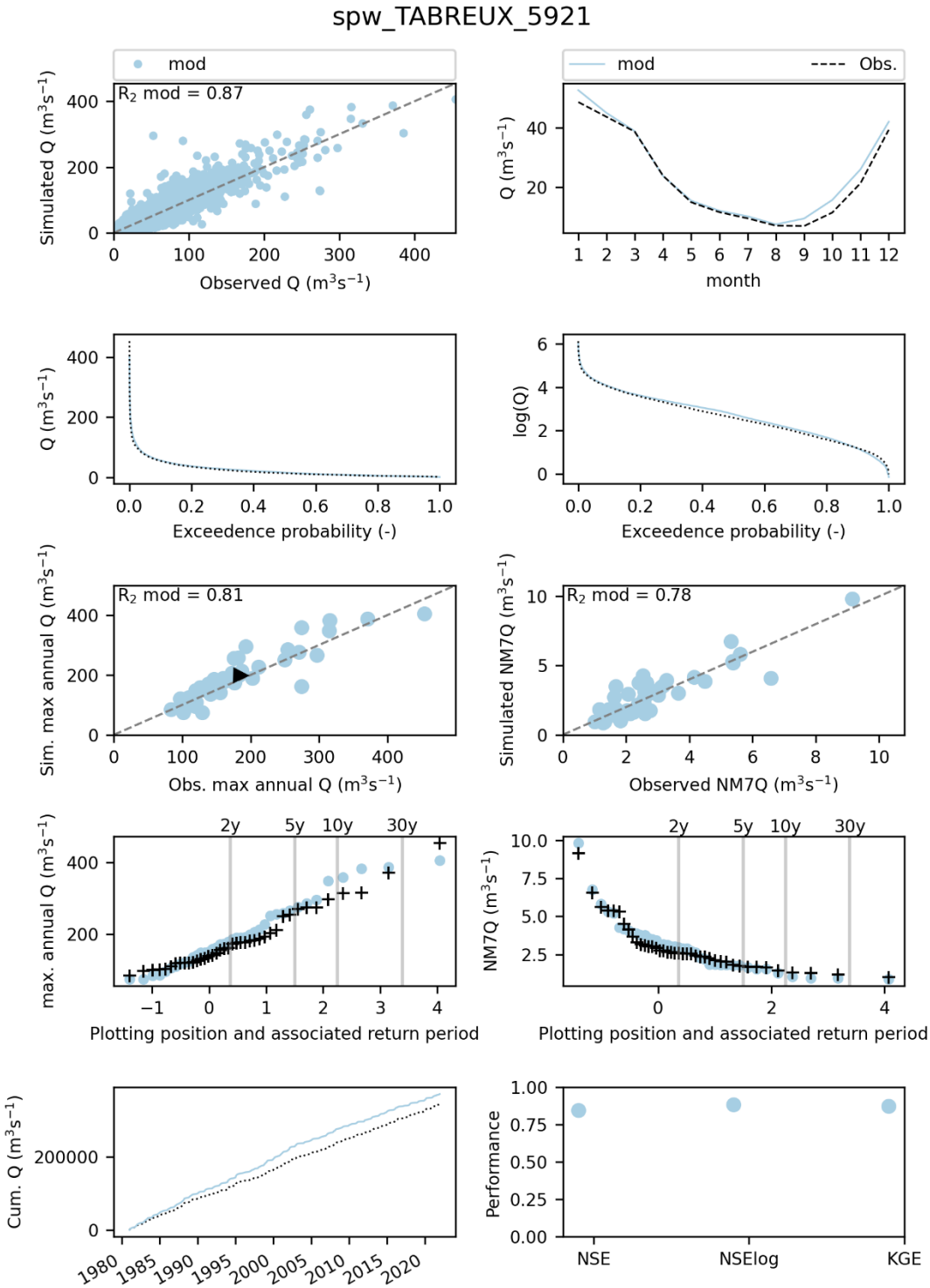


Figure 4-16 Modelled (blue) and observed (black) signatures for the Ourthe at Tabreux.

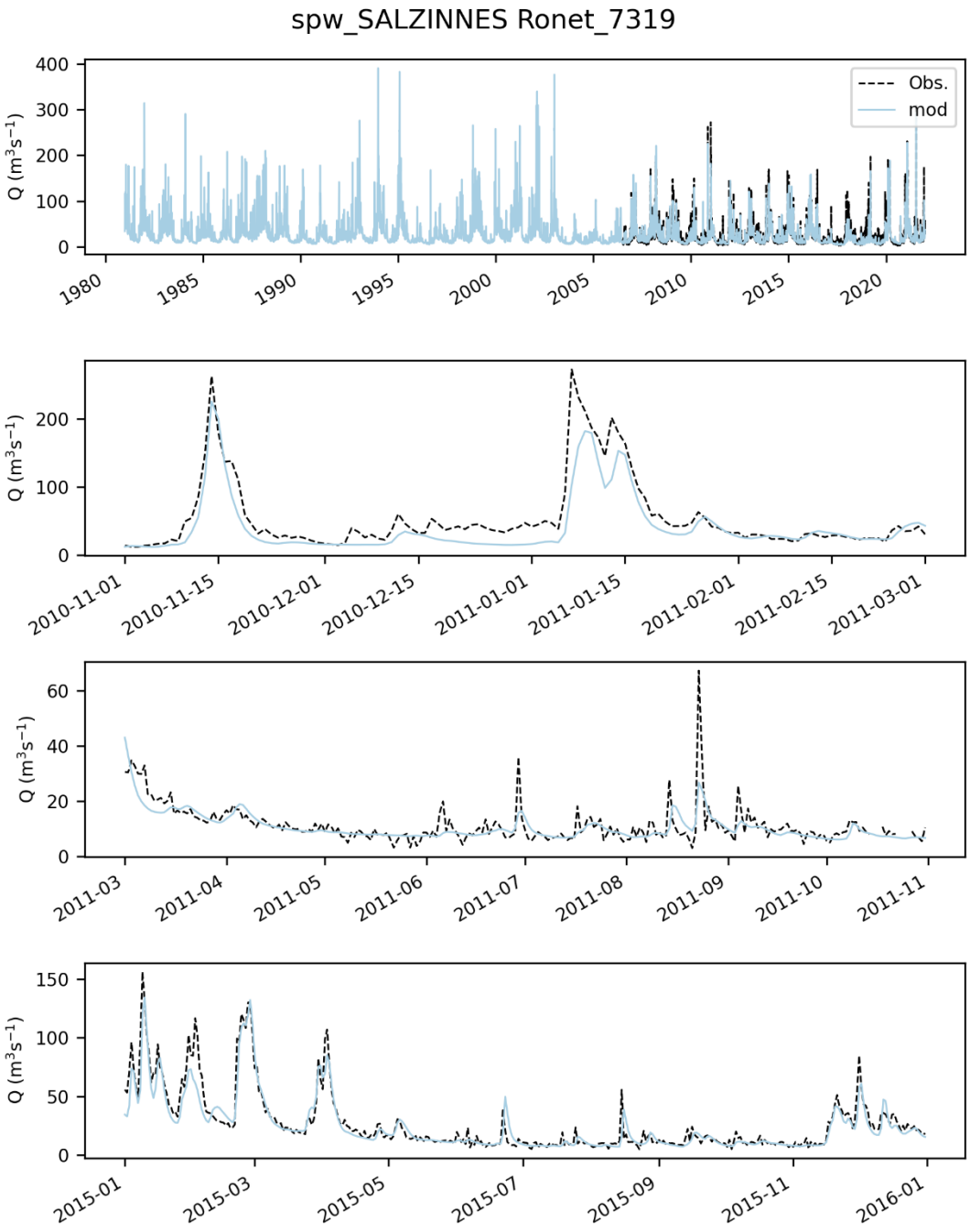


Figure 4-17 Modelled (blue) and observed (black) hydrographs for the Sambre at Salzinnes.

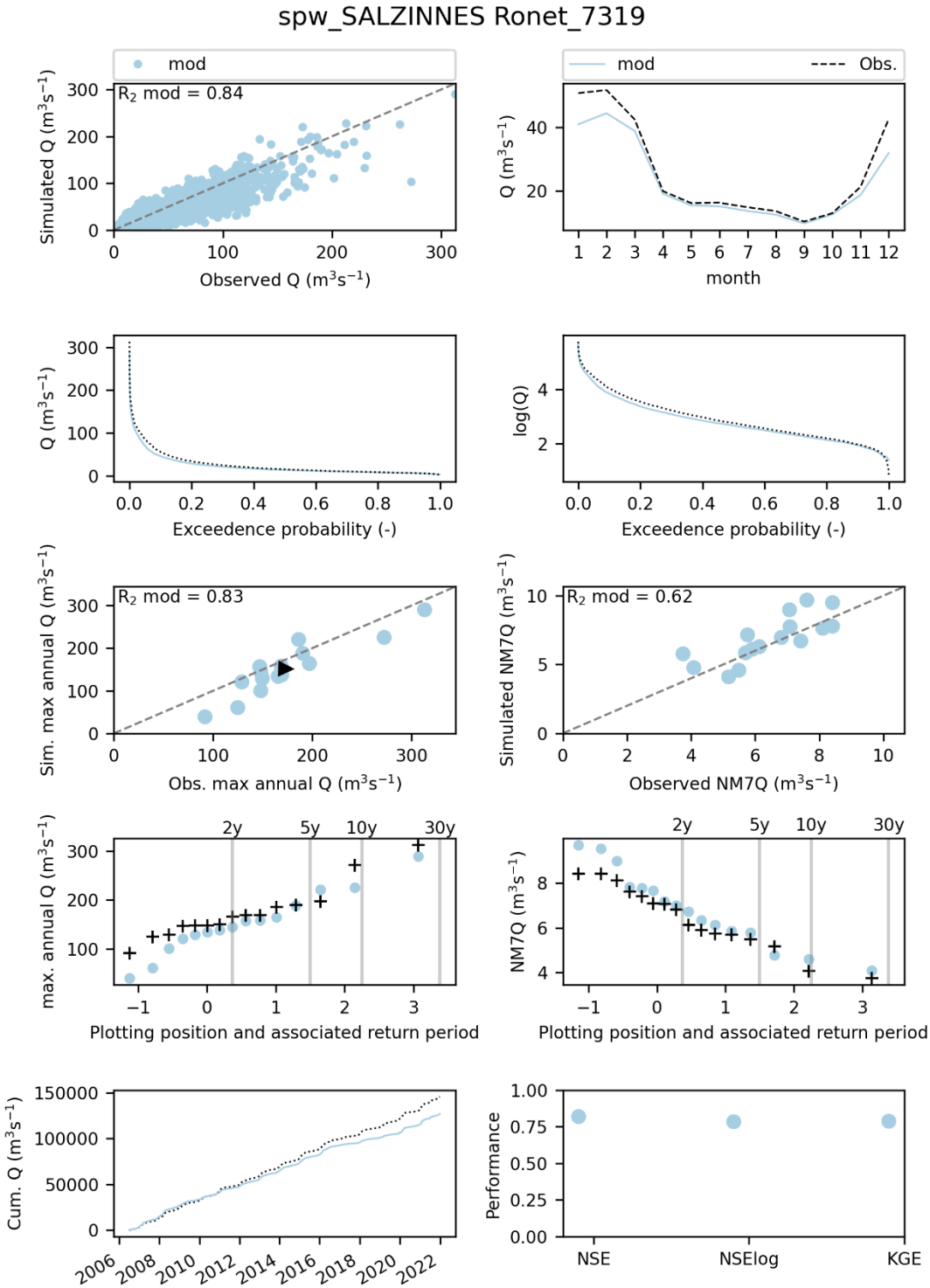


Figure 4-18 Modelled (blue) and observed (black) signatures for the Sambre at Salzinnes.

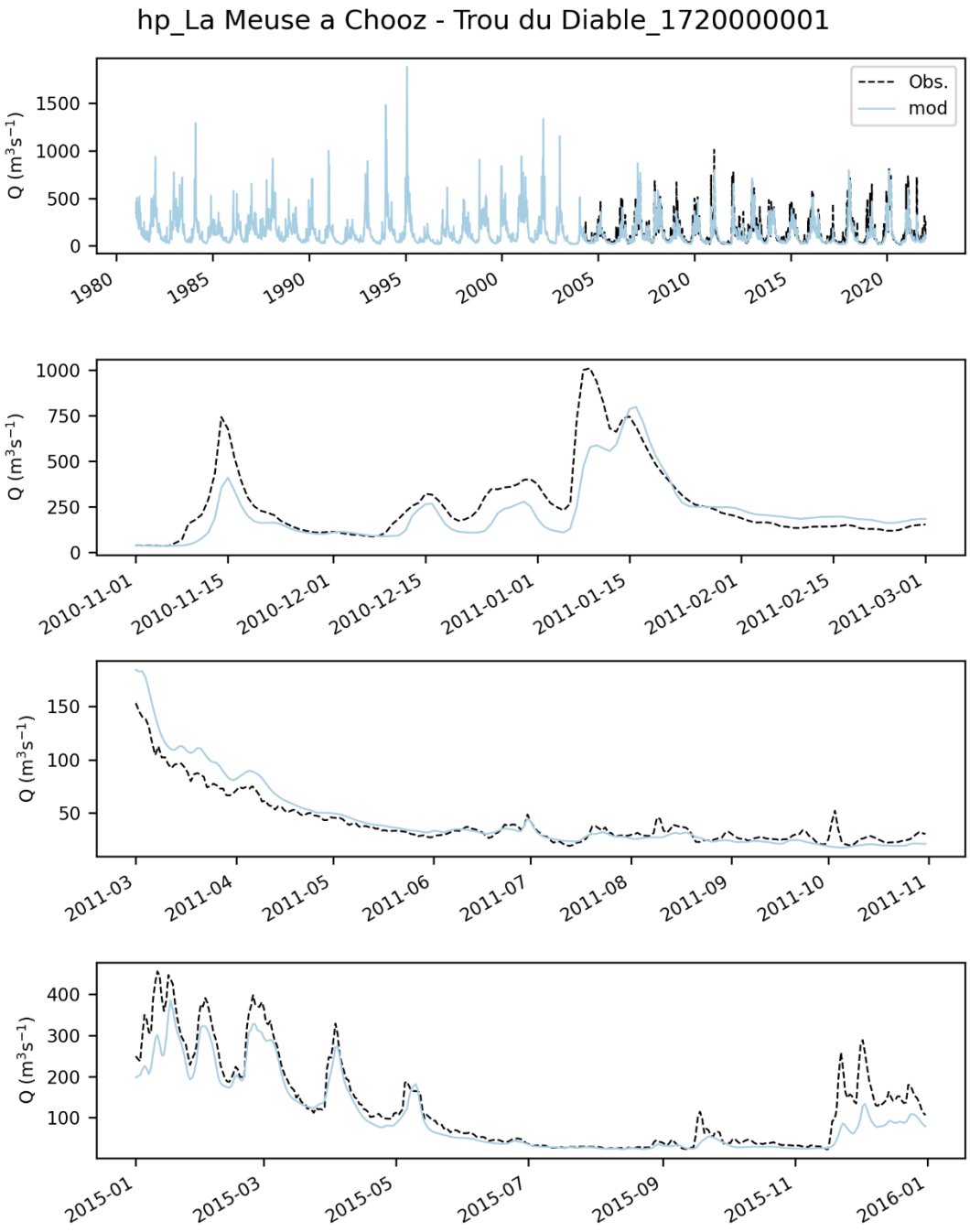


Figure 4-19 Modelled (blue) and observed (black) hydrographs for the Meuse at Chooz.

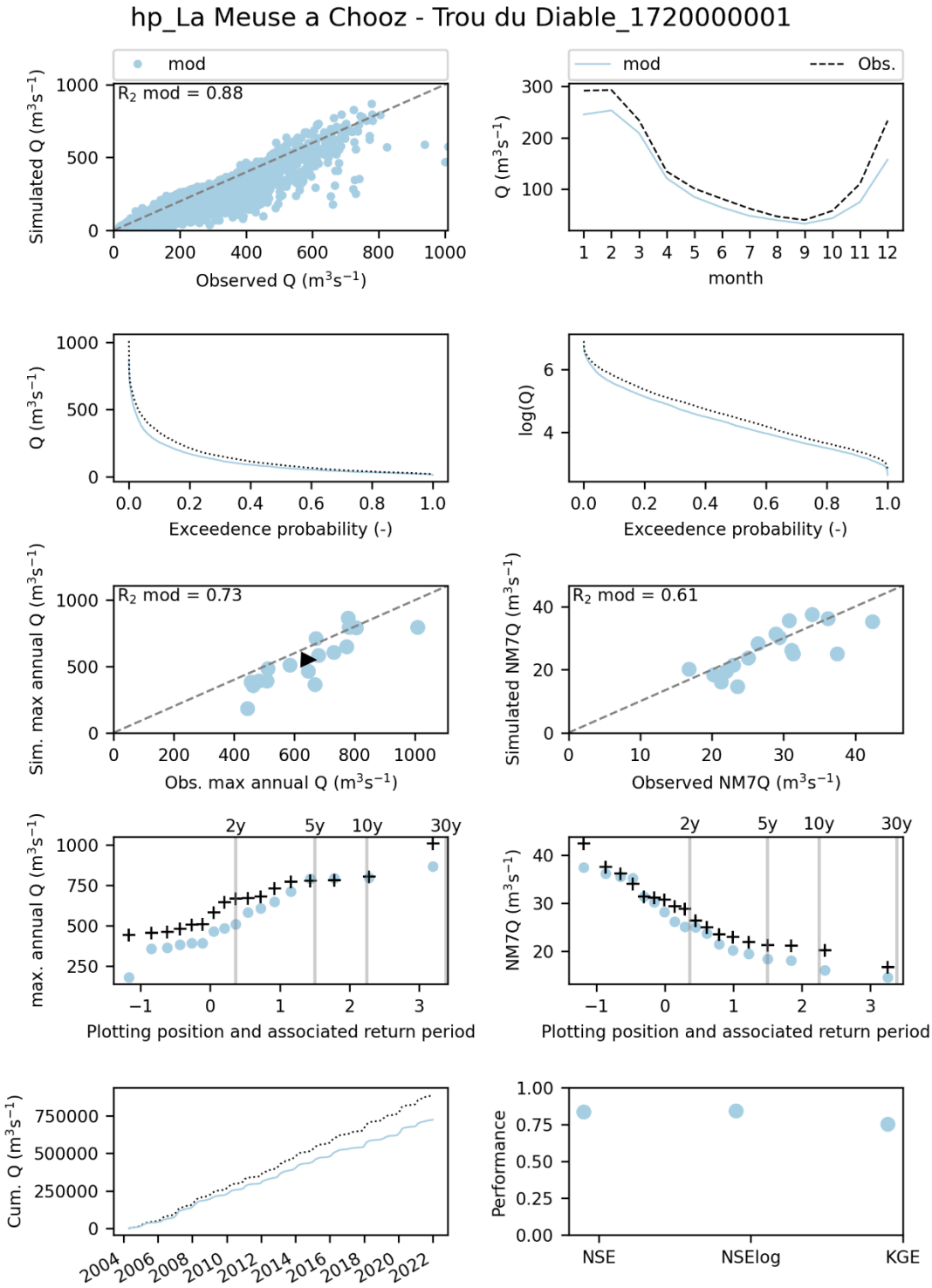


Figure 4-20 Modelled (blue) and observed (black) signatures for the Meuse at Chooz.

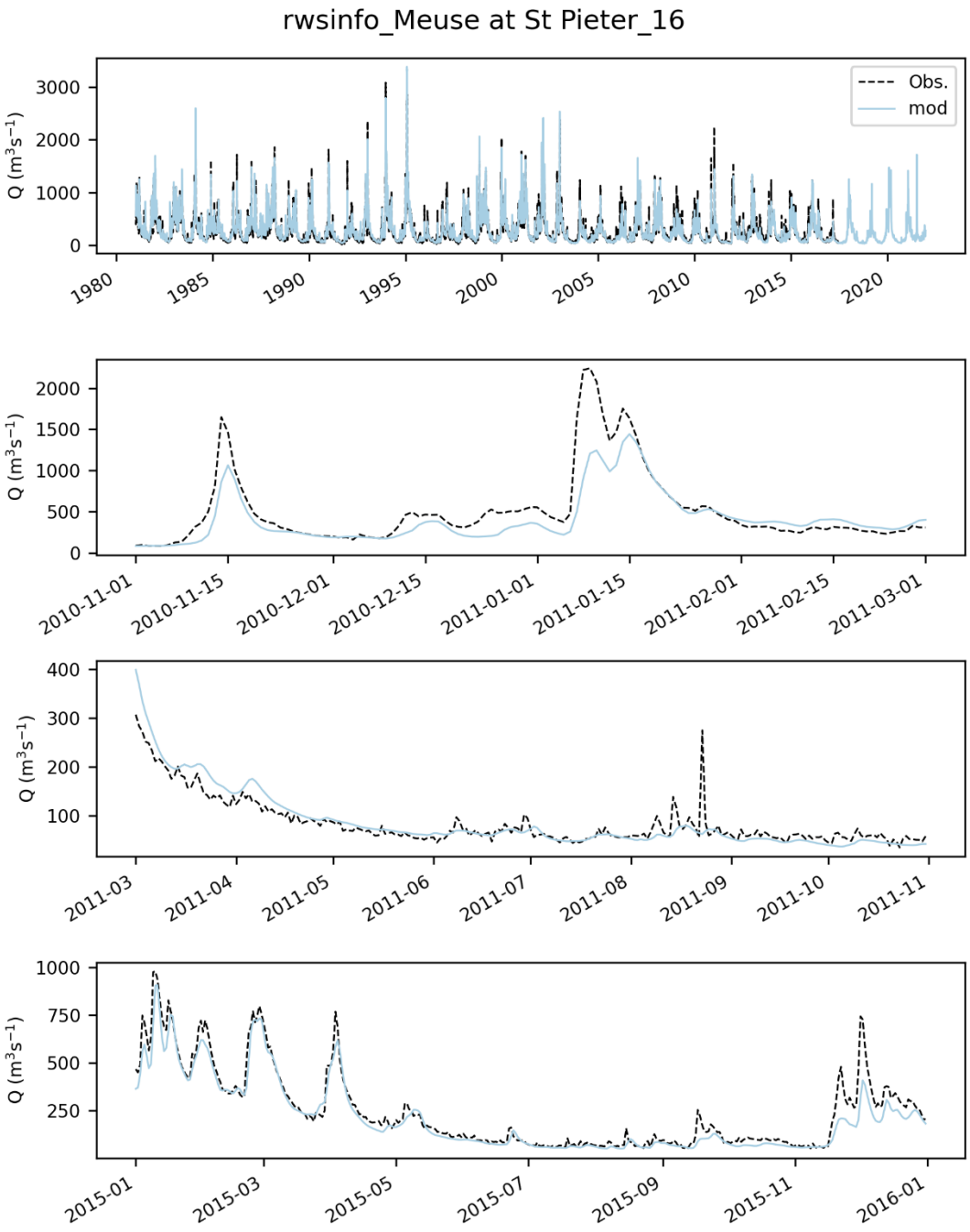


Figure 4-21 Modelled (blue) and observed (black) hydrographs for the Meuse at St Pieter.

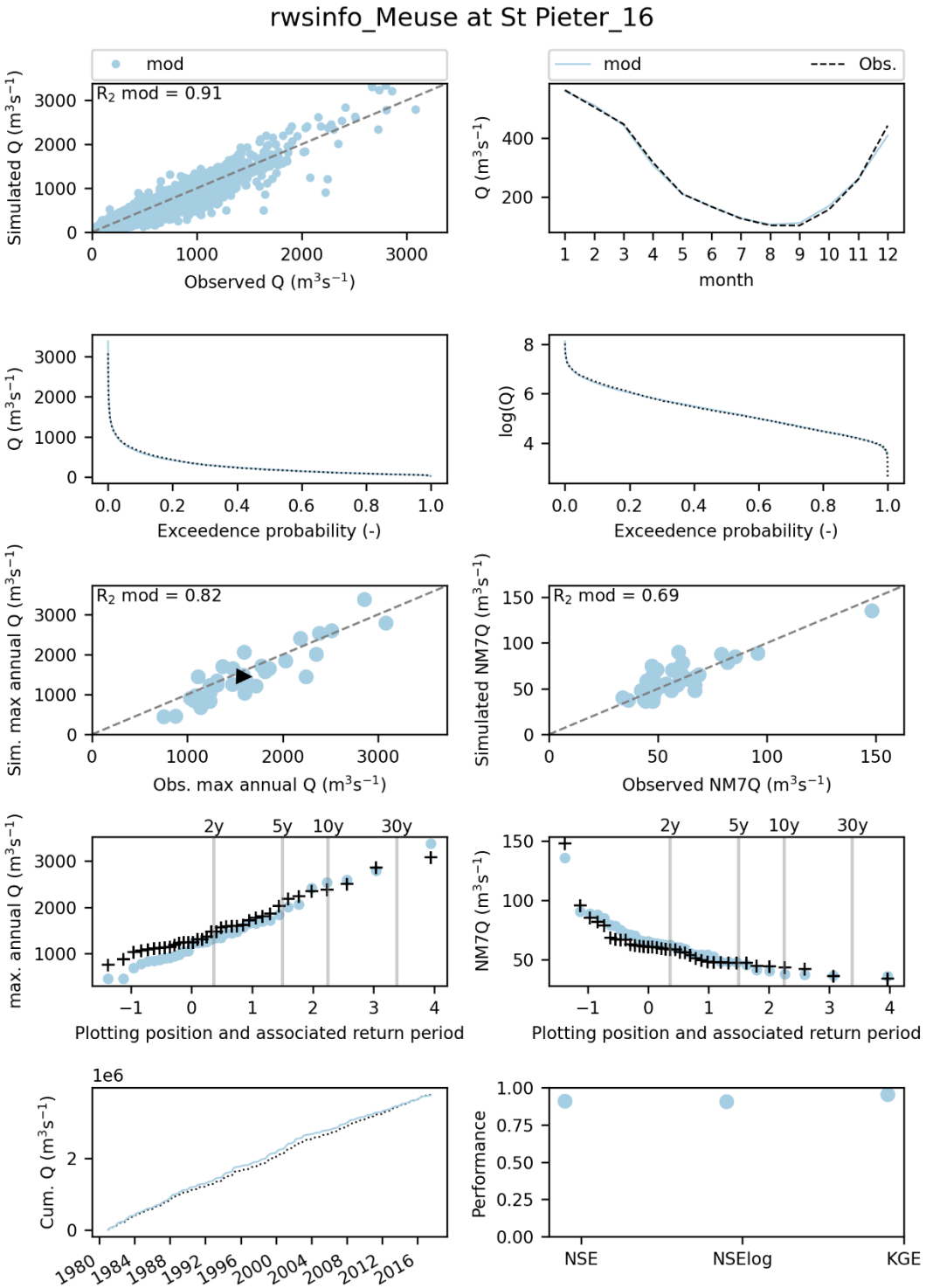


Figure 4-22 Modelled (blue) and observed (black) signatures for the Meuse at St Pieter.

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.

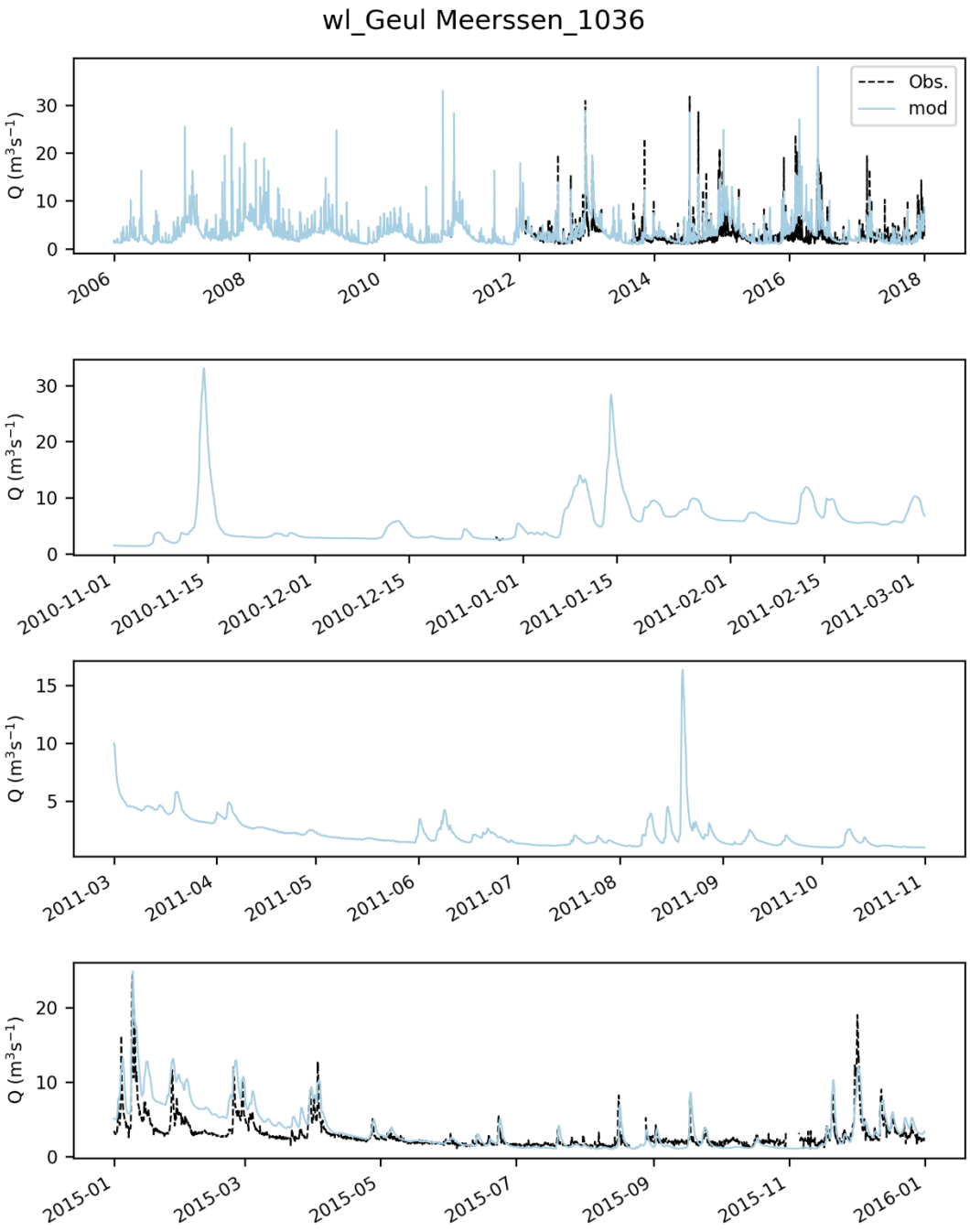


Figure 4-23 Modelled (blue) and observed (black) hydrographs for the Geul at Meerssen.

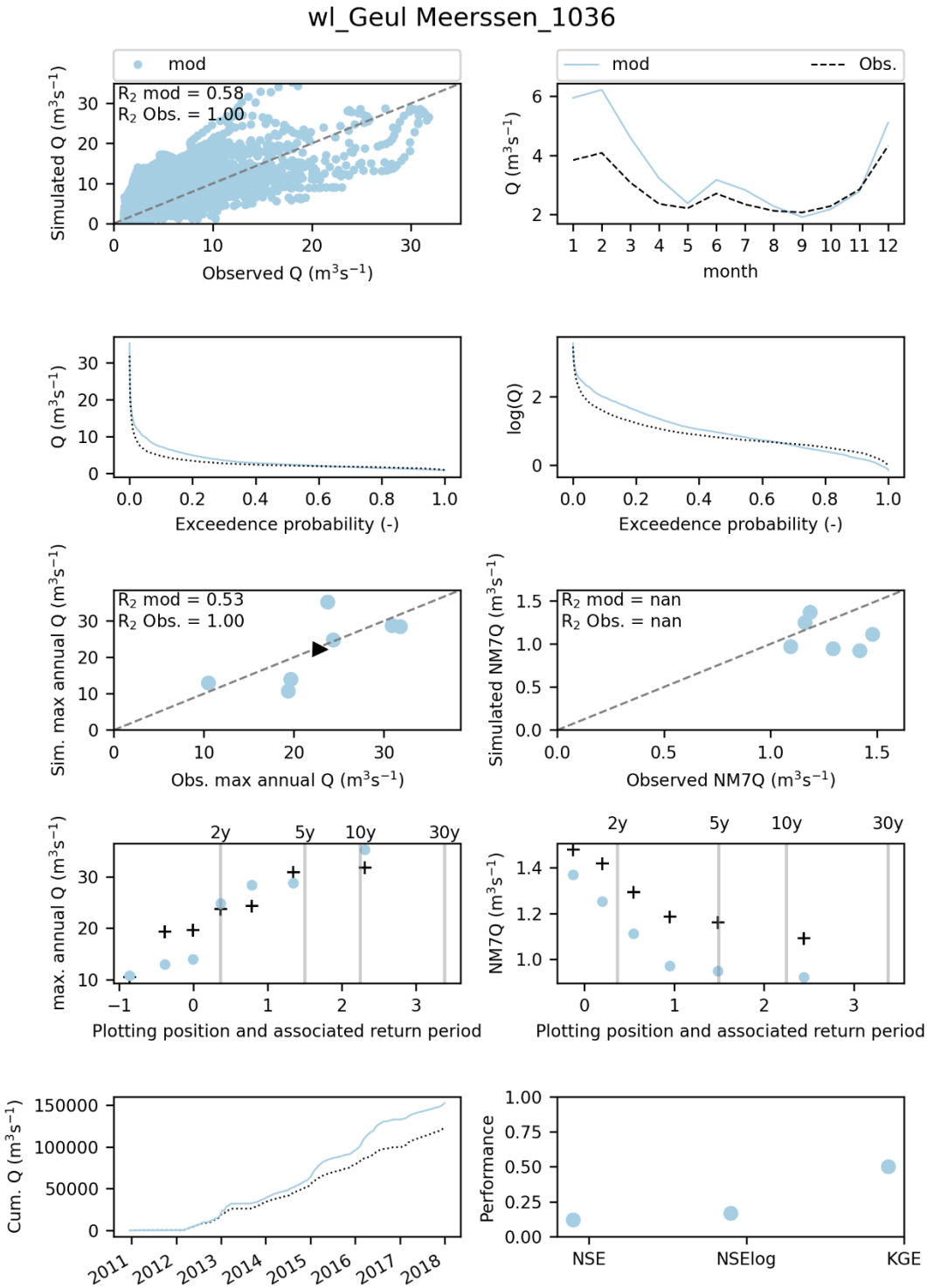


Figure 4-24 Modelled (blue) and observed (black) signatures for the Geul at Meerssen.

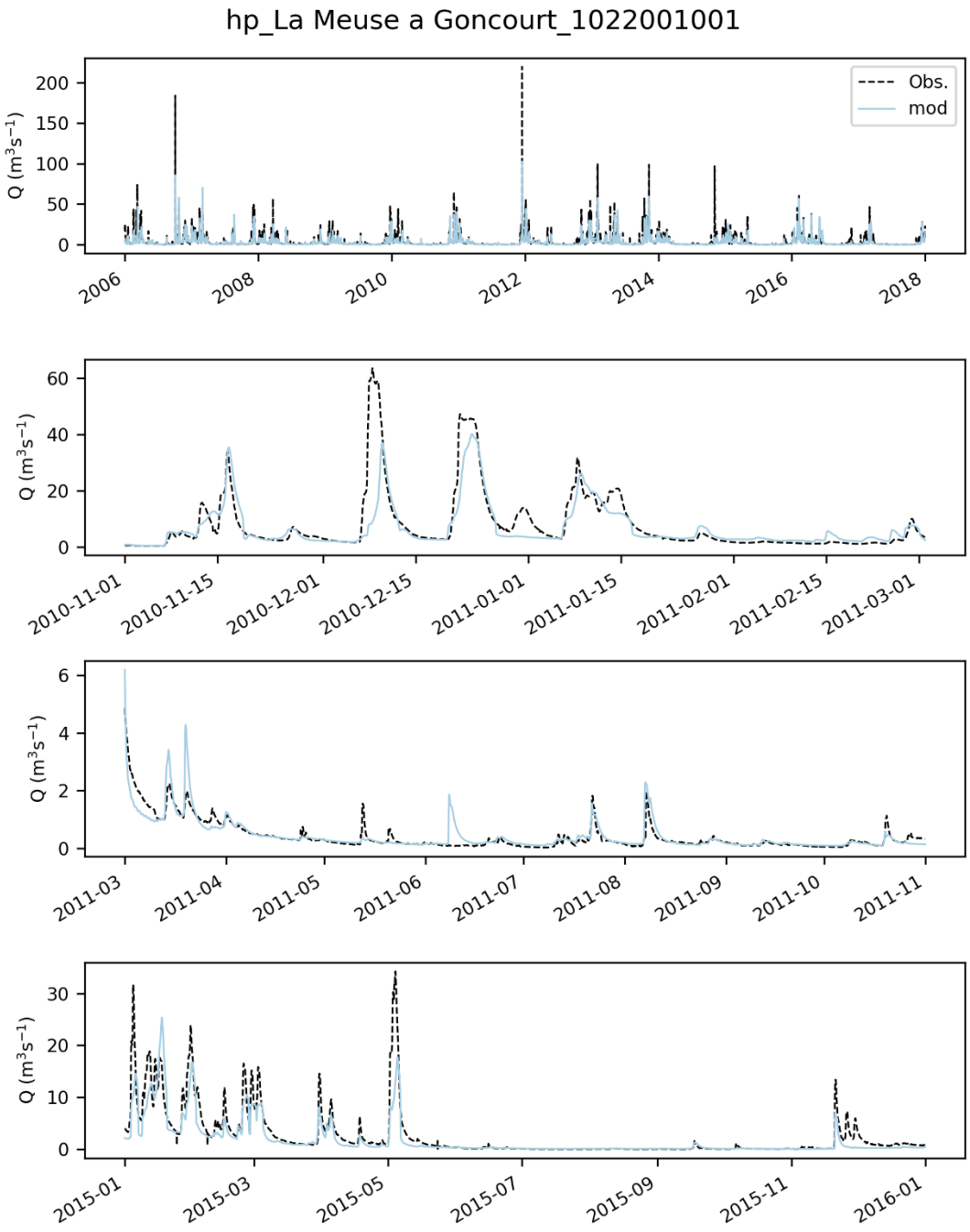


Figure 4-25 Modelled (blue) and observed (black) hydrographs for the Meuse at Goncourt.

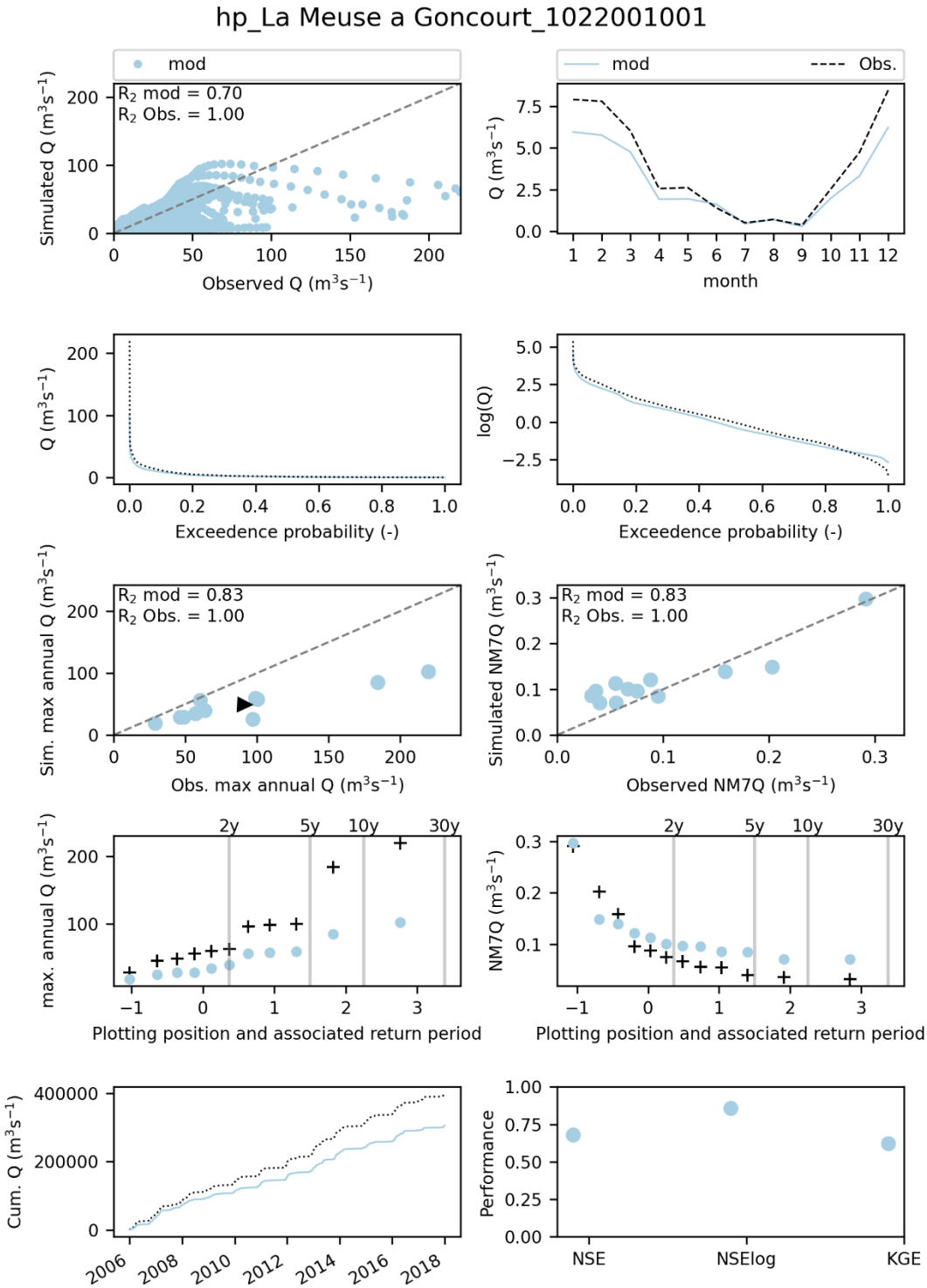


Figure 4-26 Modelled (blue) and observed (black) signatures for the Meuse at Goncourt.

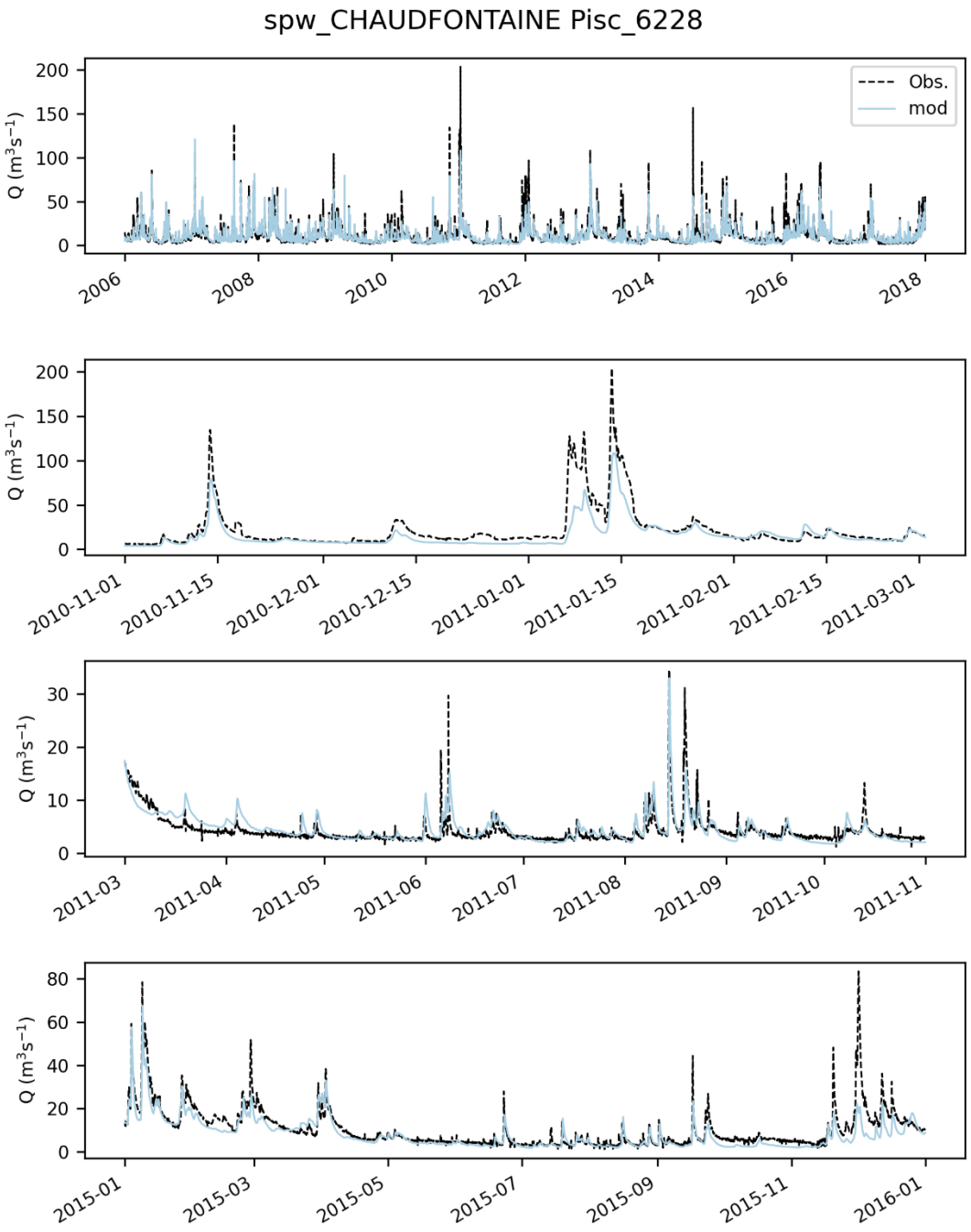


Figure 4-27 Modelled (blue) and observed (black) hydrographs for the Vesdre at Chaudfontaine.

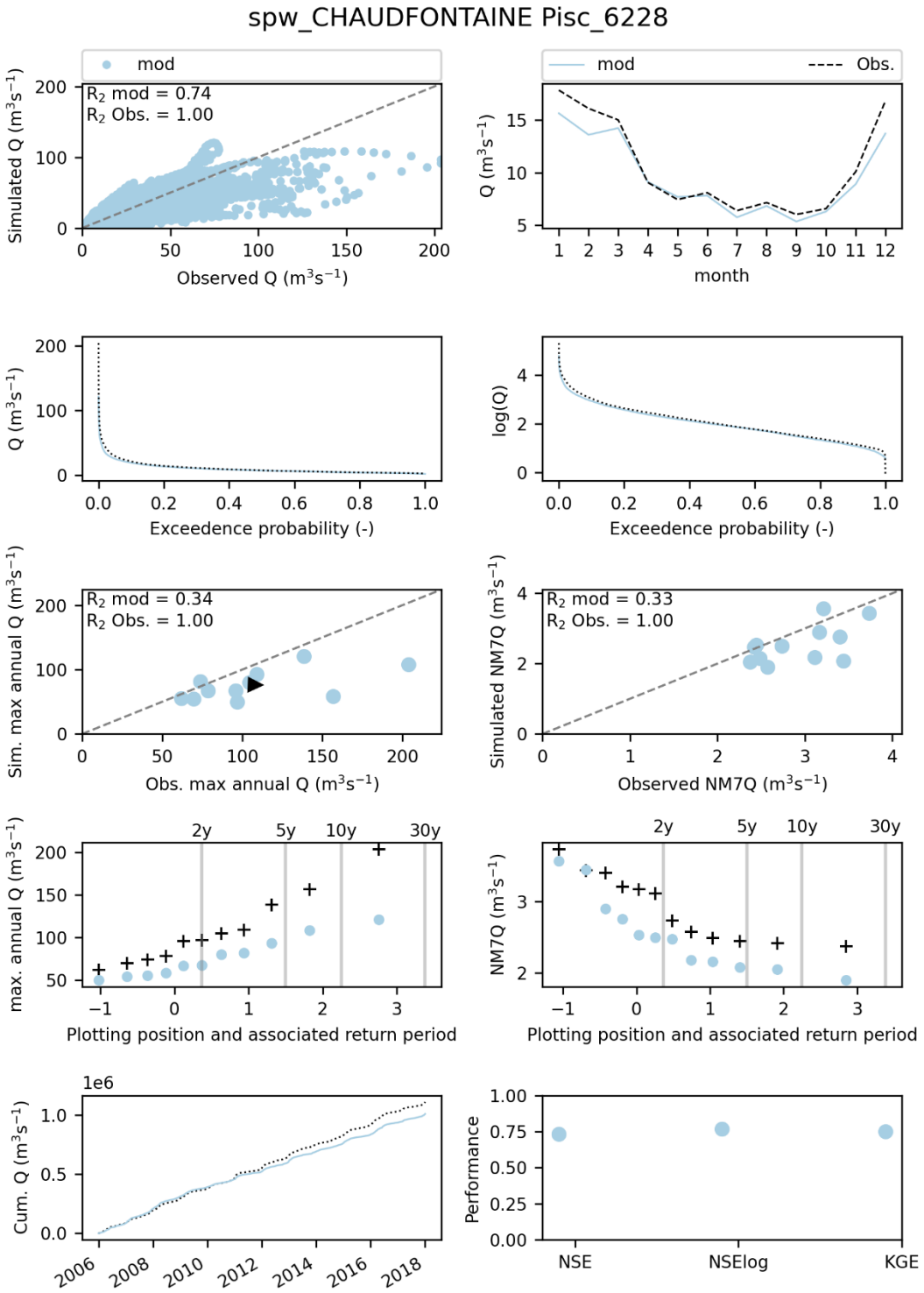


Figure 4-28 Modelled (blue) and observed (black) signatures for the Vesdre at Chaudfontaine.

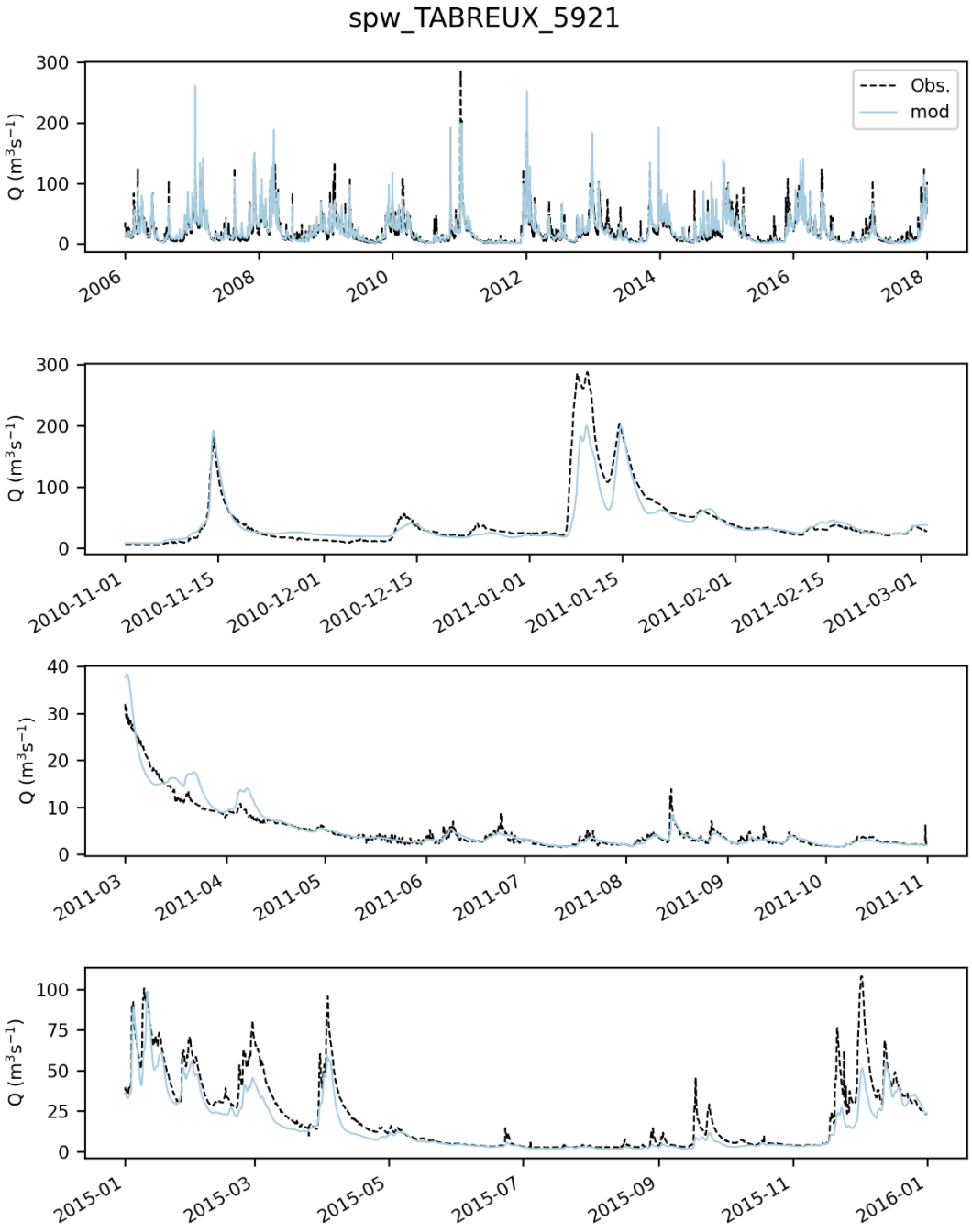


Figure 4-29 Modelled (blue) and observed (black) hydrographs for the Ourthe at Tabreux.

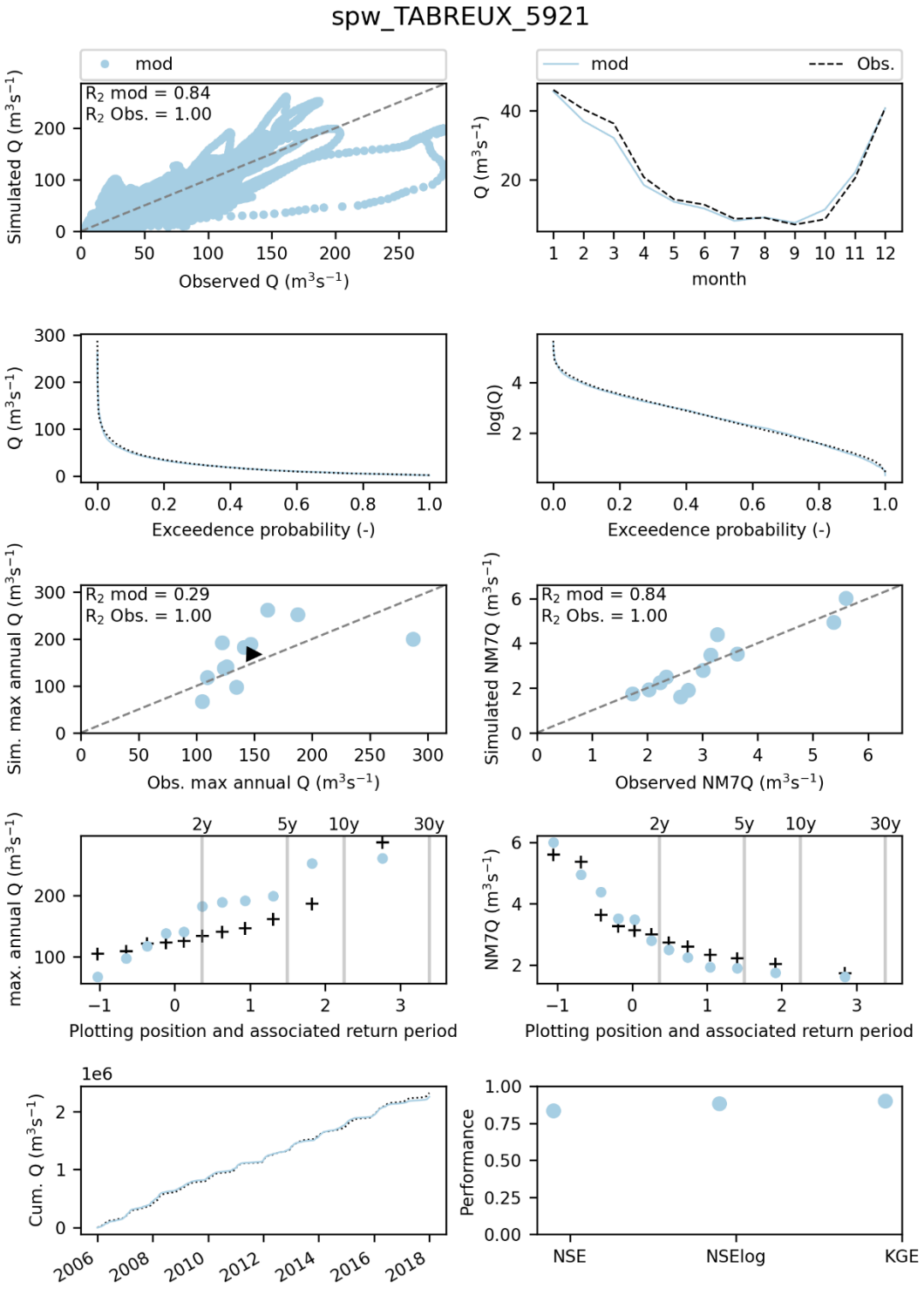


Figure 4-30 Modelled (blue) and observed (black) signatures for the Ourthe at Tabreux.

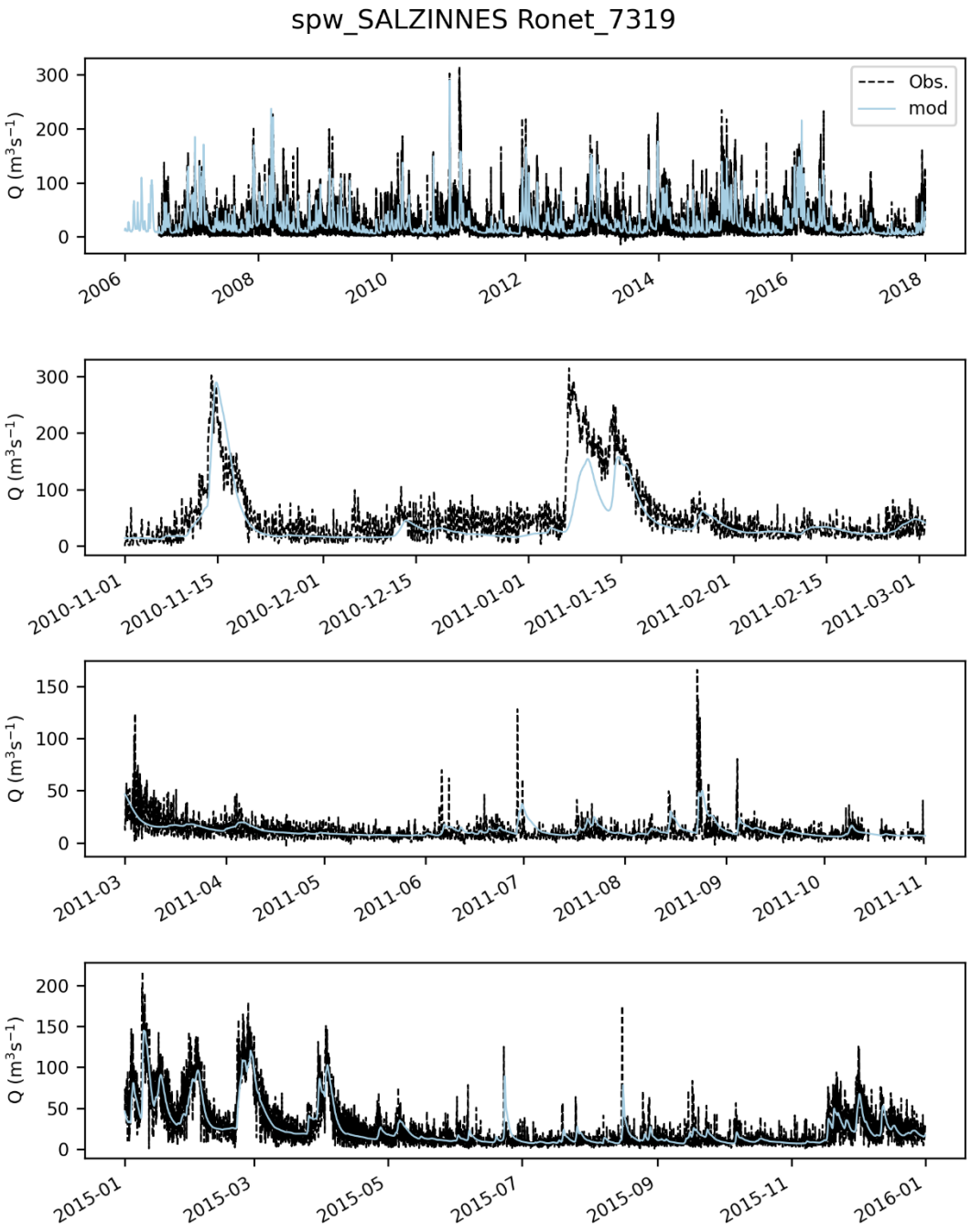


Figure 4-31 Modelled (blue) and observed (black) hydrographs for the Sambre at Salzinnes. The strong short-term variations in the observed series are caused by upstream weir operations.

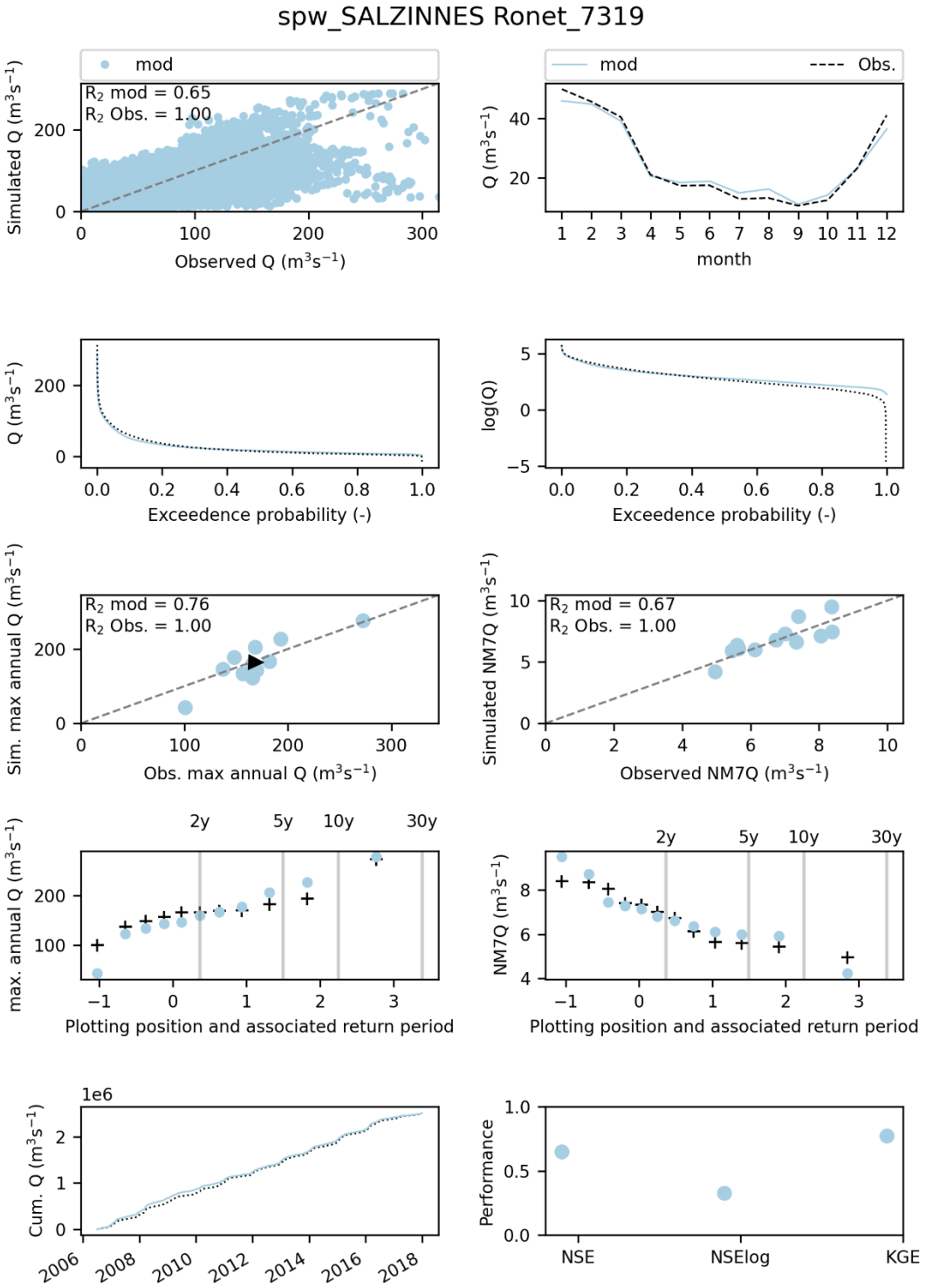


Figure 4-32 Modelled (blue) and observed (black) signatures for the Sambre at Salzinnes. For the Sambre at Salzinnes, we show the maximum annual flow using a rolling mean of 1 day to account for variations due to upstream weir operations.

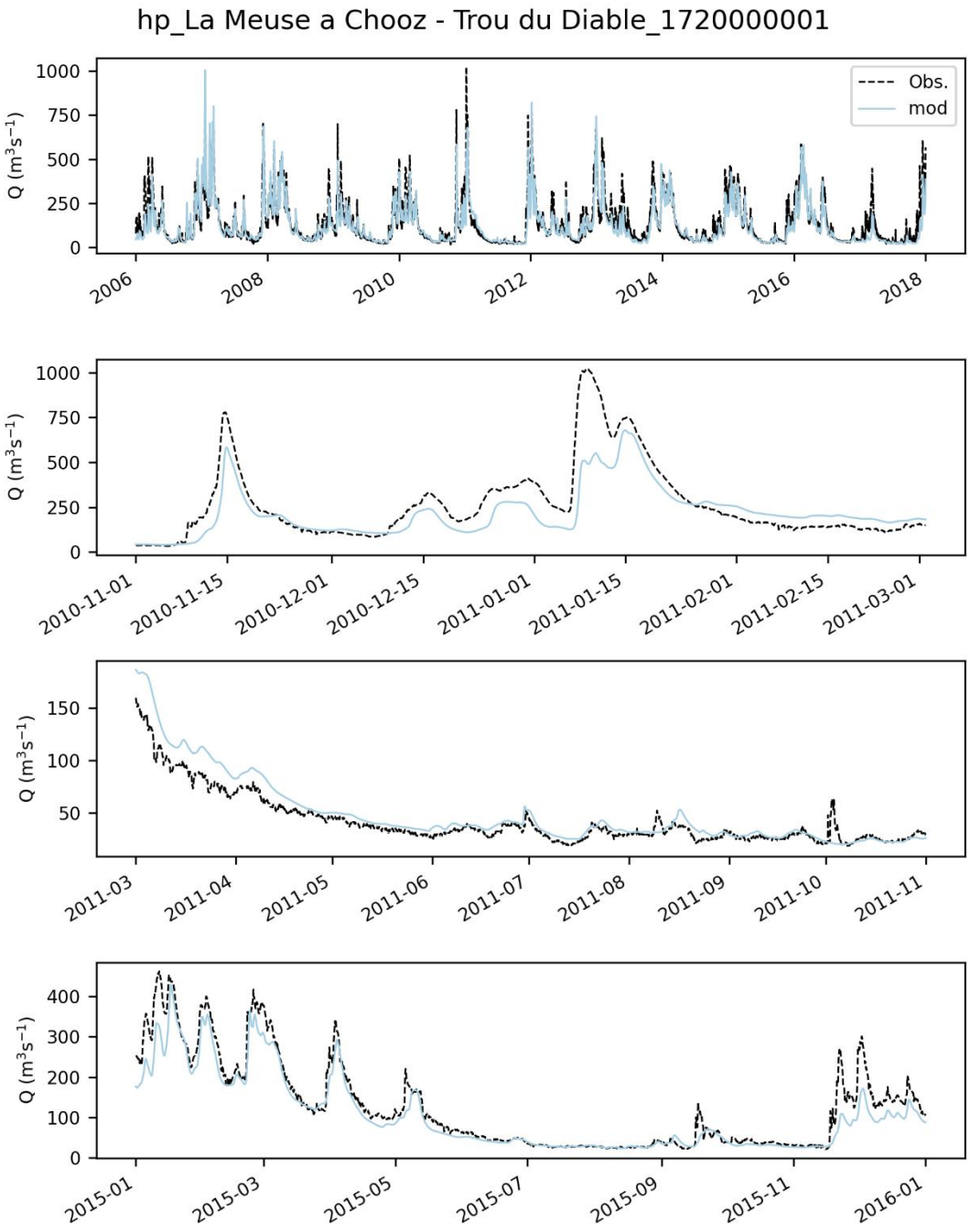


Figure 4-33 Modelled (blue) and observed (black) hydrographs for the Meuse at Chooz.

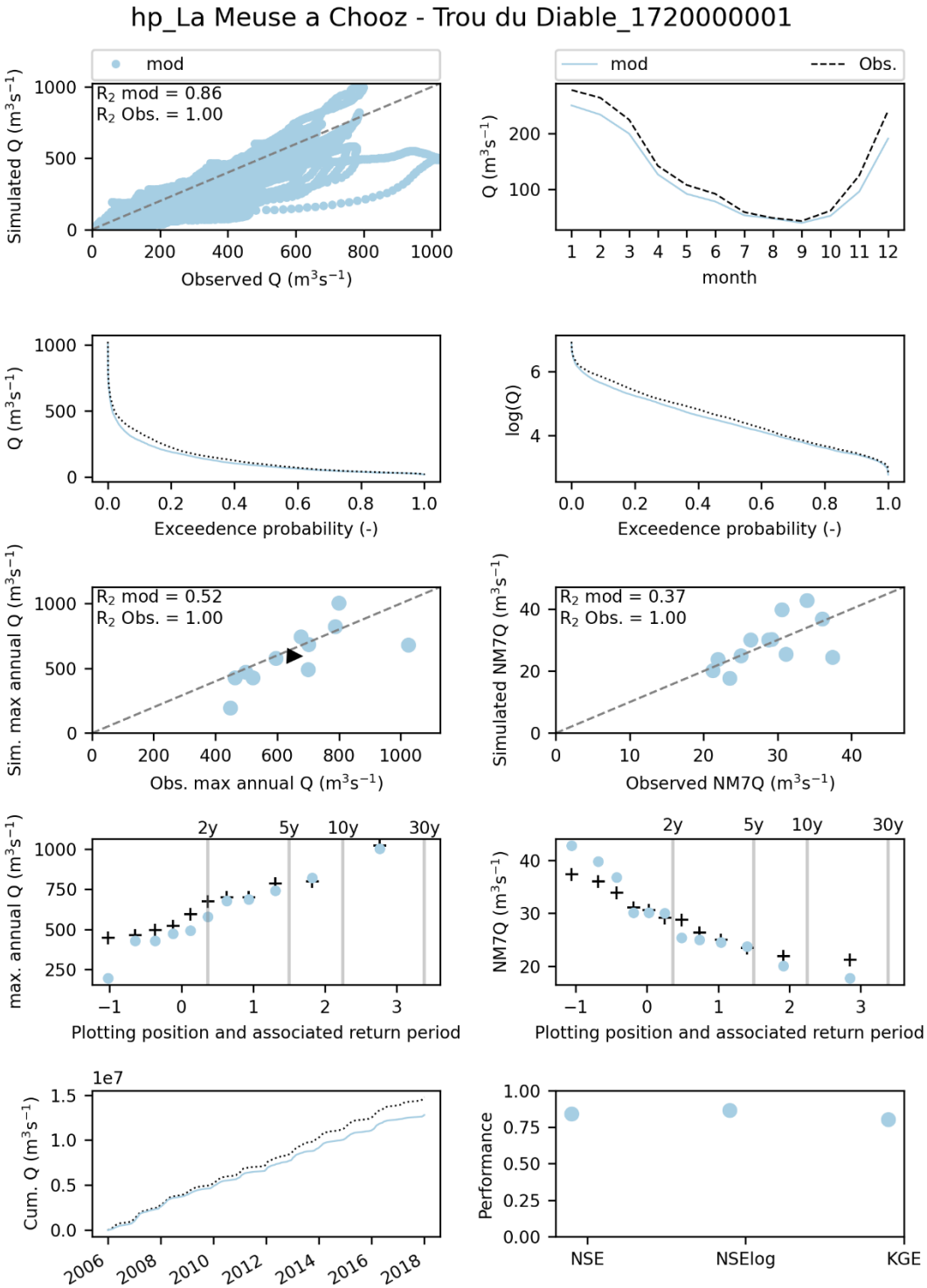


Figure 4-34 Modelled (blue) and observed (black) signatures for the Meuse at Chooz.

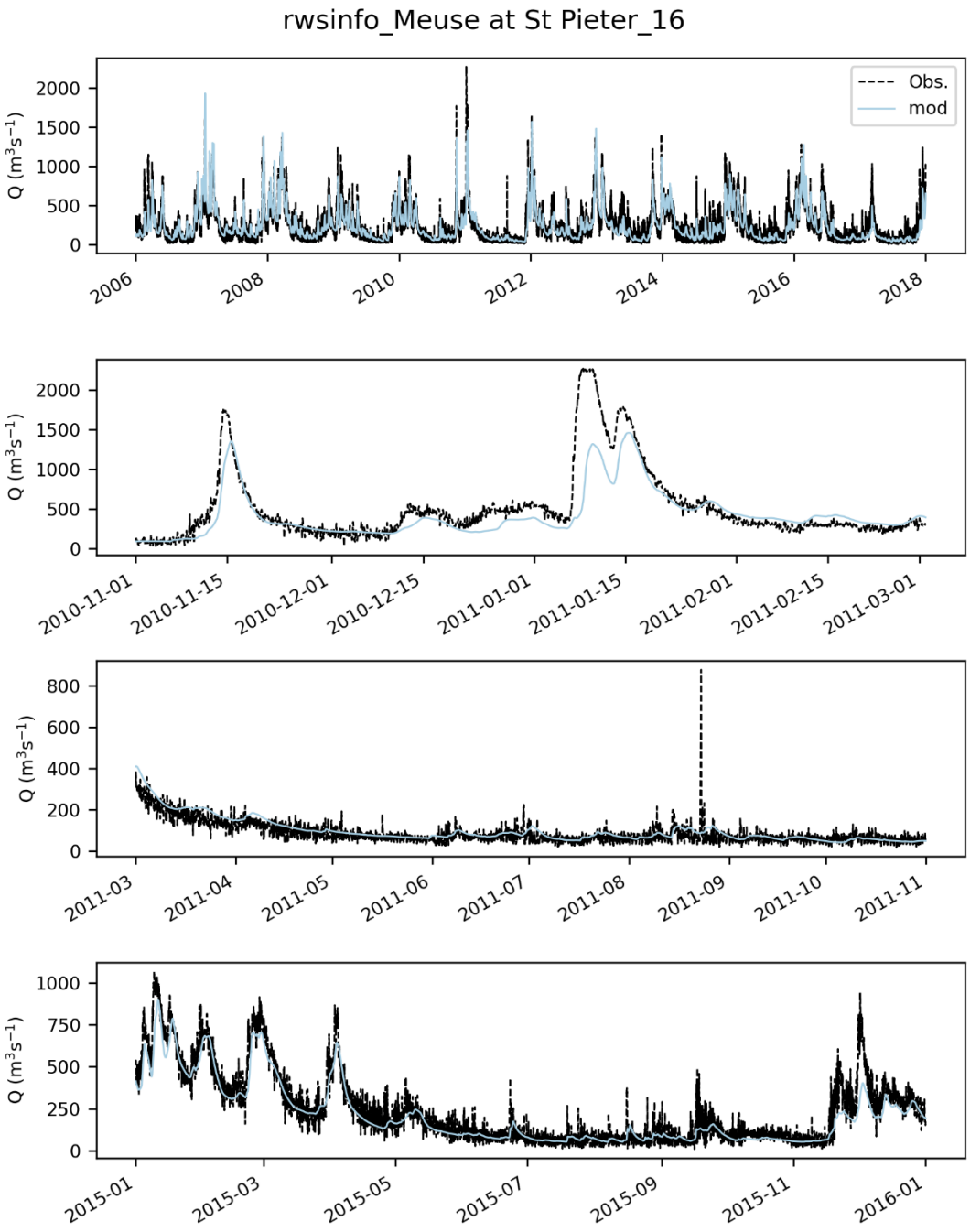


Figure 4-35 Modelled (blue) and observed (black) hydrographs for the Meuse at St Pieter. Also here, large short-duration fluctuations are caused by upstream weir operations.

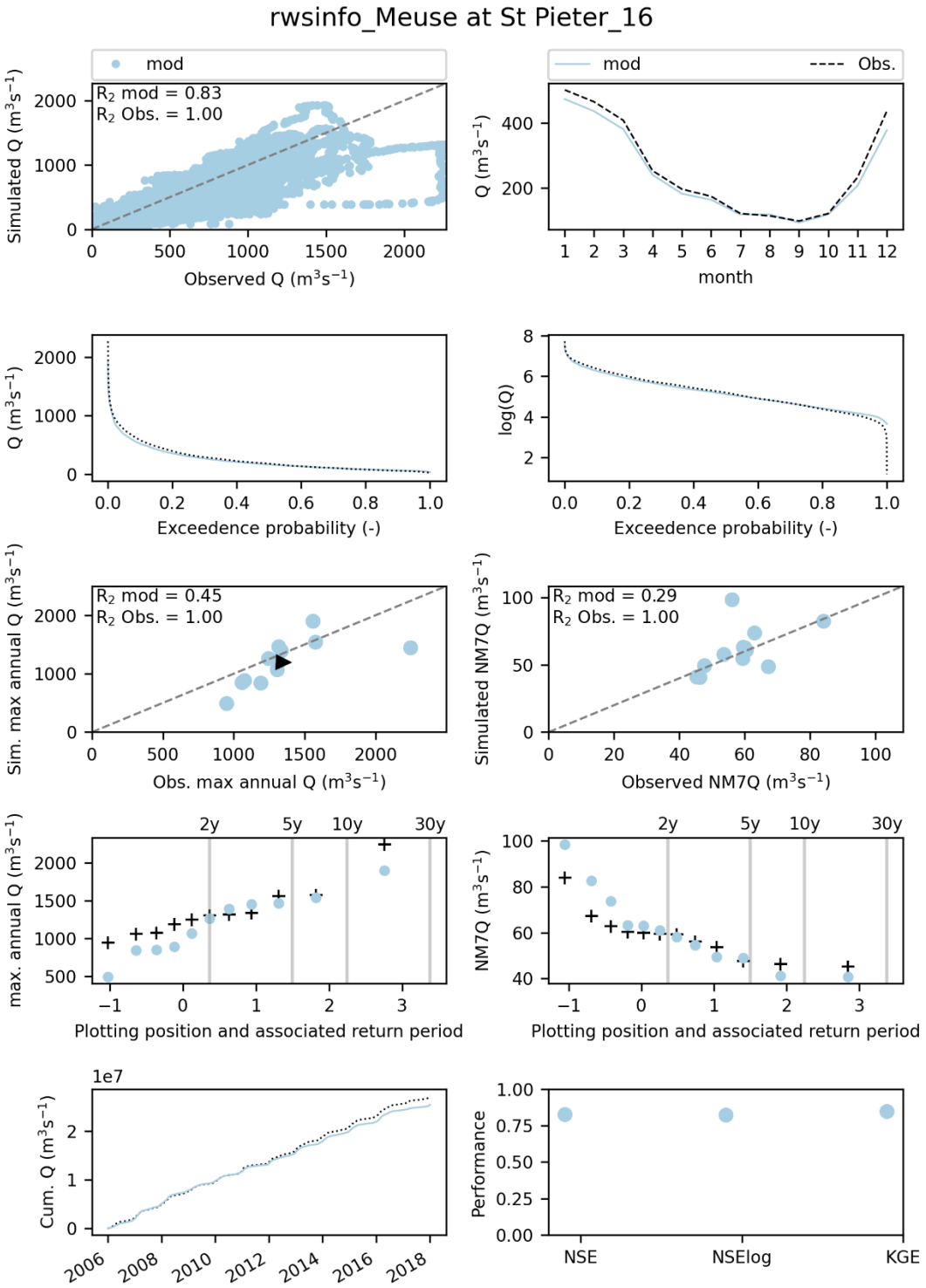


Figure 4-36 Modelled (blue) and observed (black) signatures for the Meuse at St Pieter. For the Meuse at St Pieter, we show the maximum annual flow using a rolling mean of 1 day to account for variations due to upstream weir operations.

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.

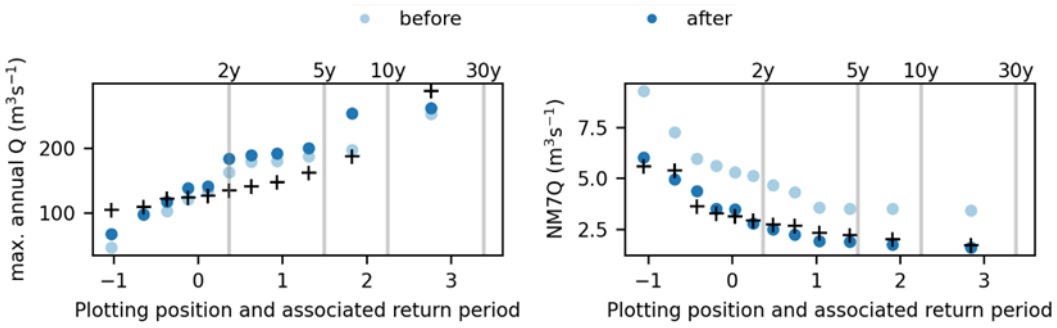


*Figure 4-37 Performance of the **hourly** model in reproducing annual maximum and minimum 7days flows (NM7Q) before and after the calibration (dark blue represents the model performance after calibration and light blue before calibration, while the black symbols show the observations) for the Ourthe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

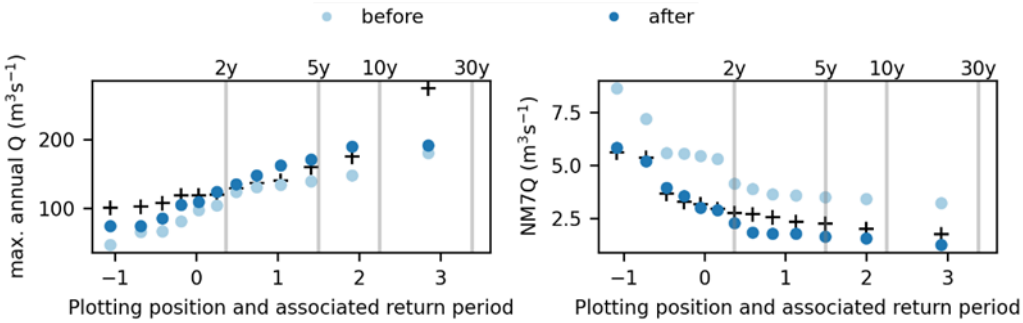


*Figure 4-38 Performance of the **daily** model in reproducing annual maximum and minimum 7days flows (NM7Q) before and after the calibration (dark blue represents the model performance after calibration and light blue before calibration, while the black symbols show the observations) for the Ourthe 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.

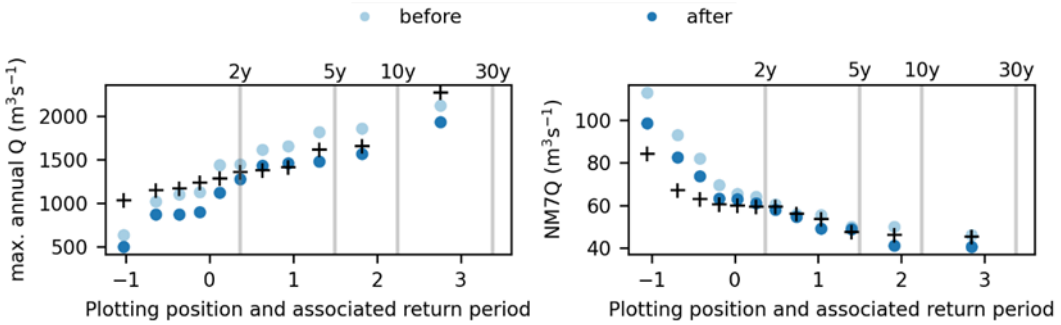


*Figure 4-39 Performance of the **hourly** model in reproducing annual maximum and minimum 7days flows (NM7Q) before and after the calibration (dark blue represents the model performance after calibration and light blue before calibration, while the black symbols show the observations) for the Meuse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

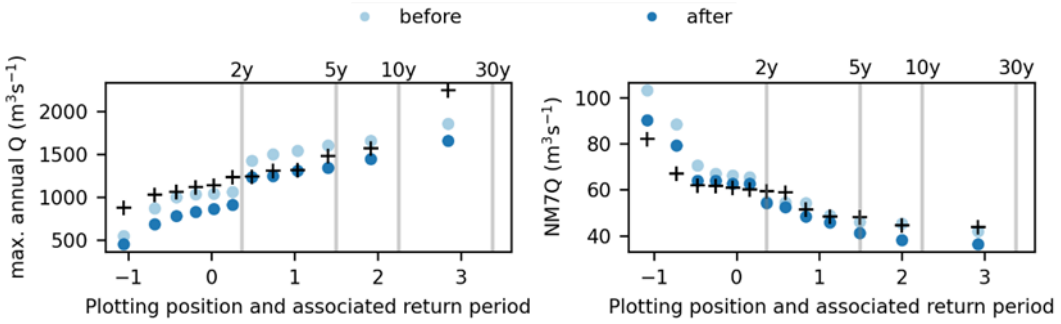


*Figure 4-40 Performance of the **daily** model in reproducing annual maximum and minimum 7days flows (NM7Q) before and after the calibration (dark blue represents the model performance after calibration and light blue before calibration, while the black symbols show the observations) for the Meuse at 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).

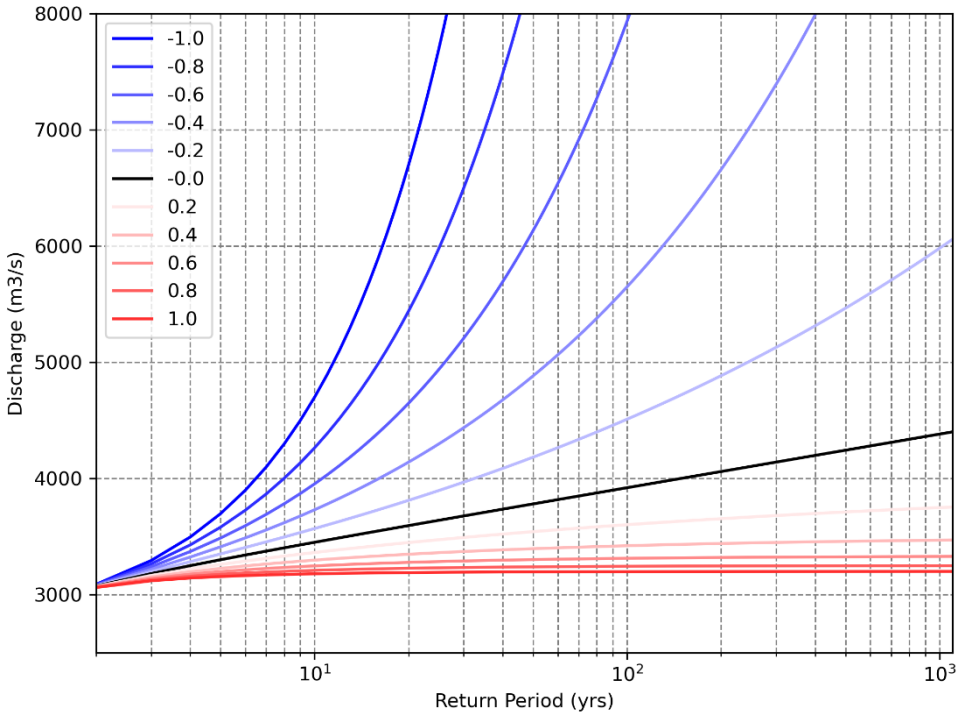


Figure 5-1: Exceedance probabilities obtained from fitting a Generalized Extreme Value (GEV) distribution with a location parameter of 3000, a scale parameter of 200, and varying shape parameters.

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.

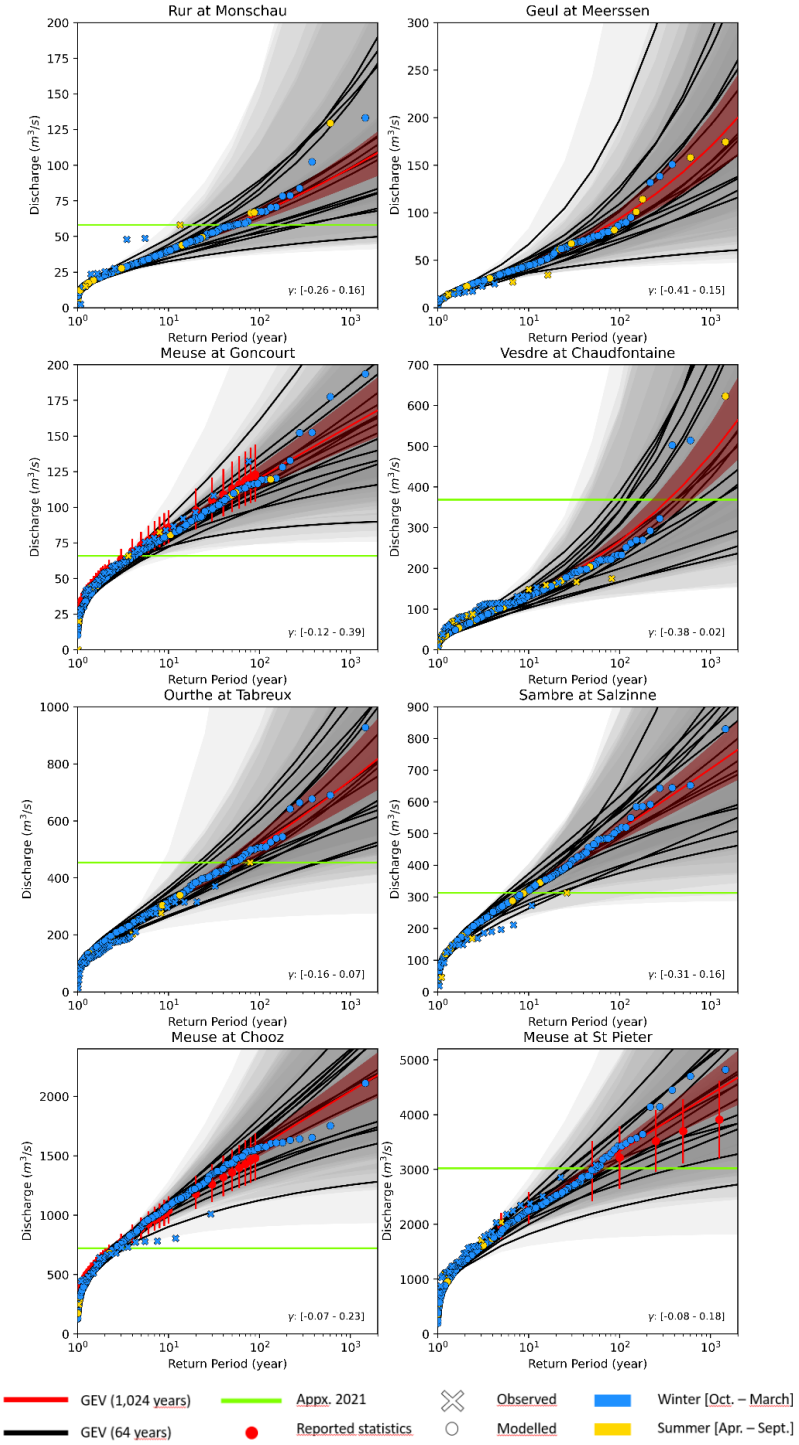


Figure 5-2: Extreme value distribution fit at selected locations across the Meuse catchment at a daily time step. Black lines represent the GEV fit obtained from one ensemble member (64 hydrological years) and the range of shape parameters is written at the bottom right of each figure. Red lines show 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.

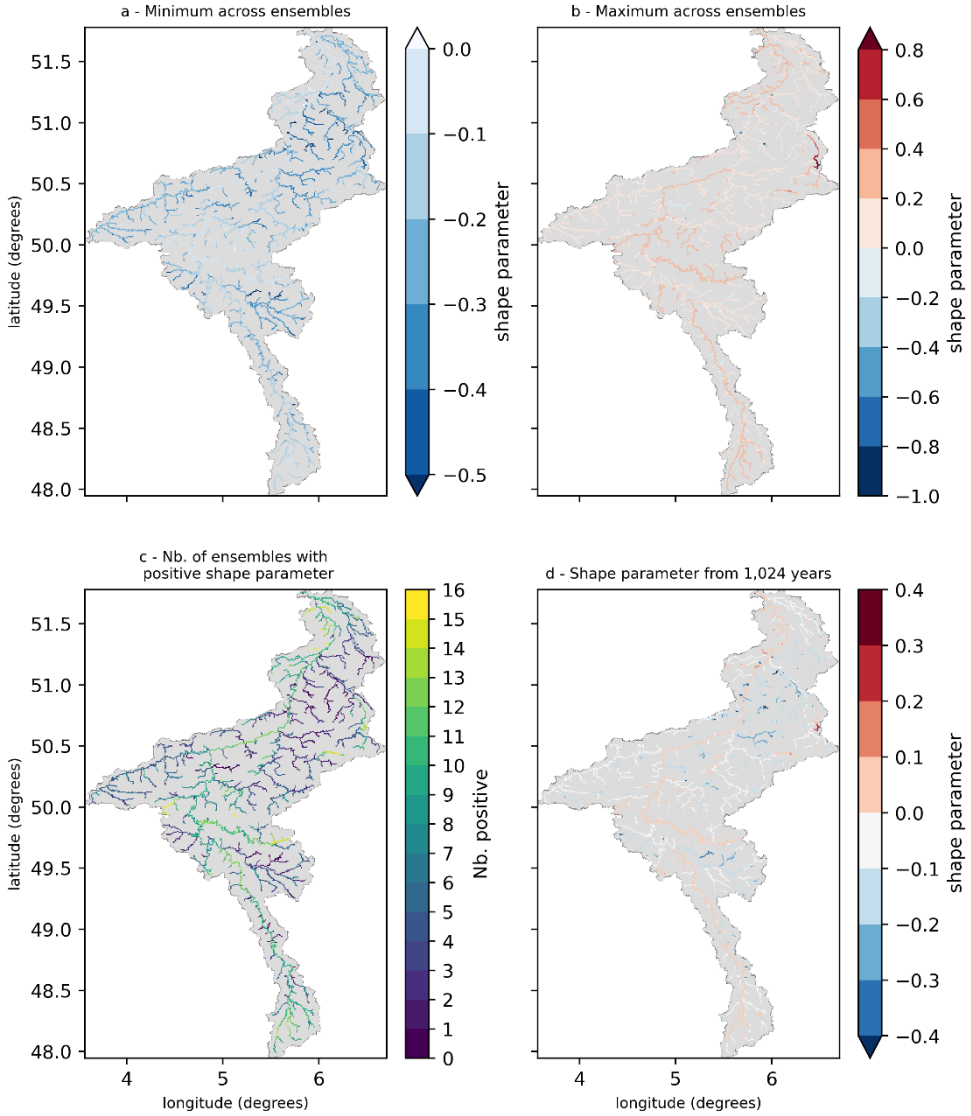


Figure 5-3: a) Minimum GEV shape parameter obtained from the 16 ensemble members for the daily runs. b) Maximum GEV shape parameter obtained from the 16 ensemble members for the daily runs. c) Number of ensembles with a positive GEV shape parameter. d) GEV shape parameter obtained by combining records 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.

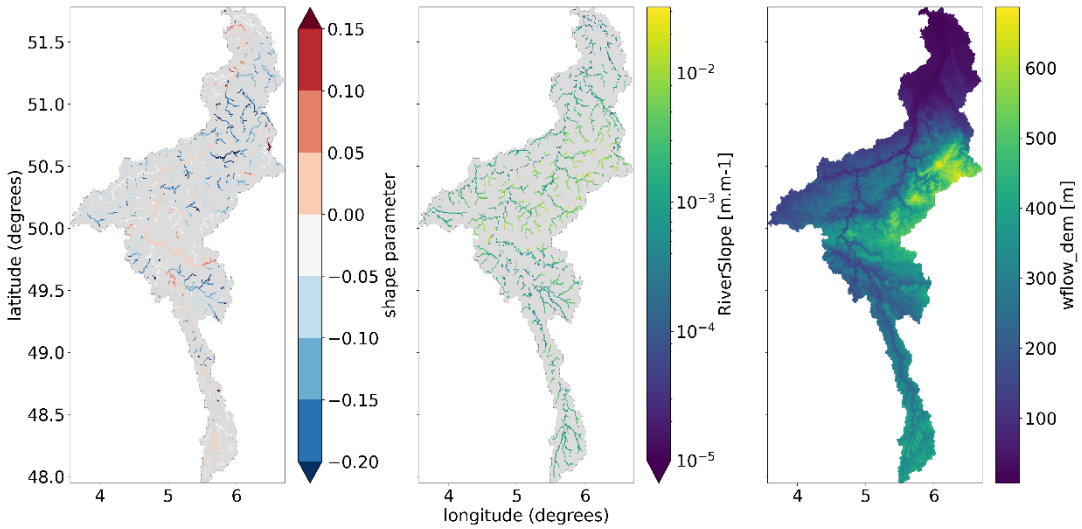


Figure 5-4: left panel: GEV shape parameter obtained by combining records from all ensembles, i.e. 1,024 hydrological years. A positive shape parameter indicates a bounded tail, a negative value a heavy tail and a value close to 0 a light tail (same as Figure 5-3d but with a different color scale). Middel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

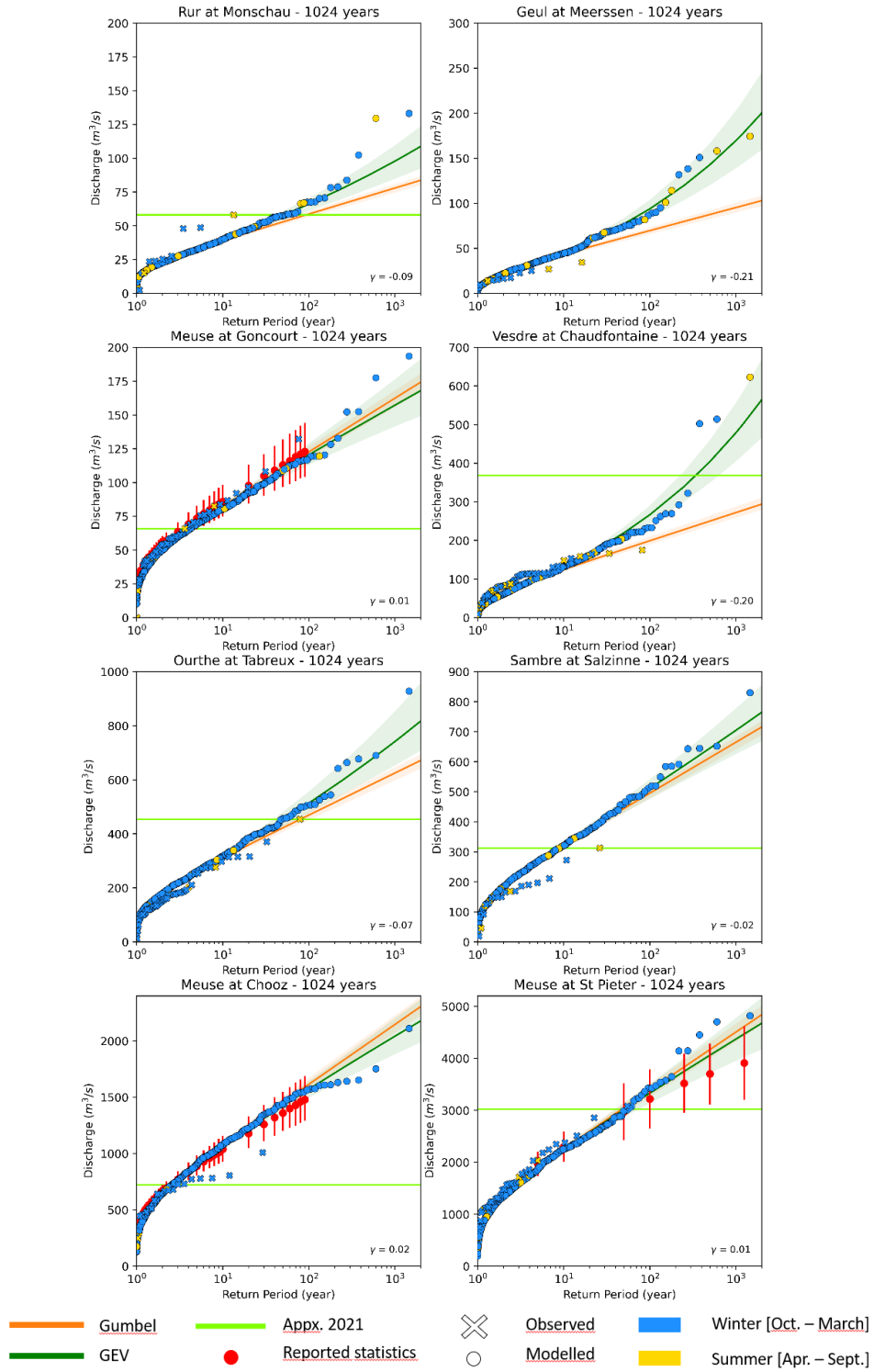


Figure 5-5: Extreme value distribution fit at selected locations across the Meuse catchment at a daily time step. Dark green lines represent the GEV fit obtained from combining all ensembles (1,024 hydrological years, shown as full dots) and blue lines the Gumbel fit. The shading represents the 95% confidence 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.

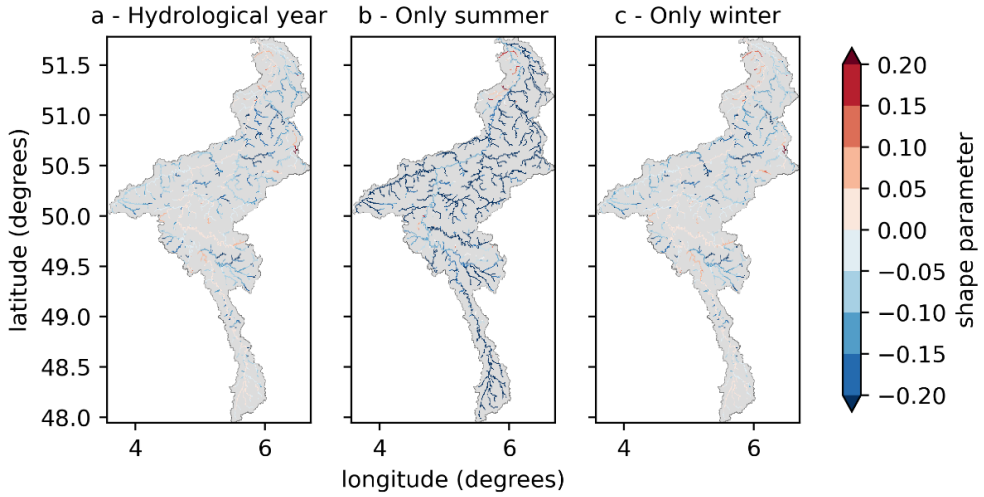
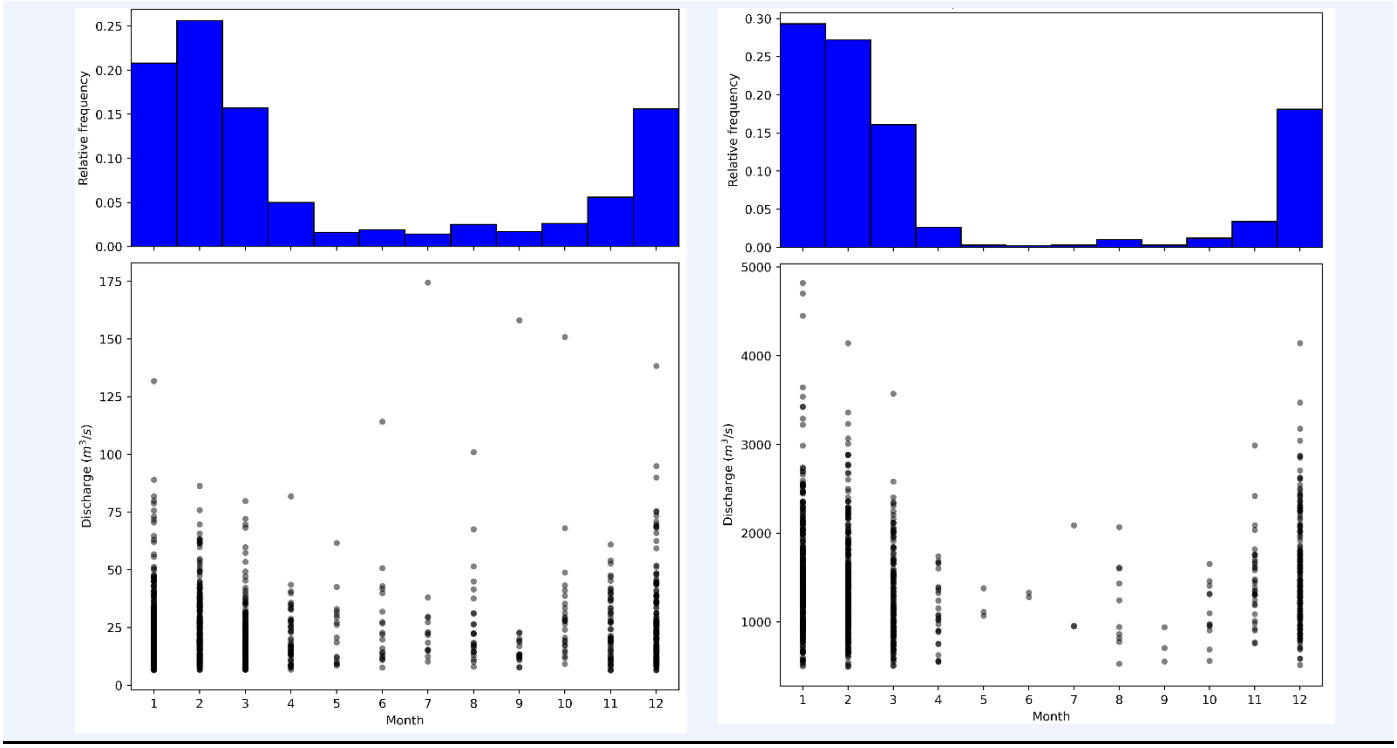


Figure 5-6: GEV shape parameter obtained from sampling discharge maxima from (a) 1,024 hydrological years, (b) summers only and (c) winters only. Summer is defined from April to September and winter from October to March.

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.

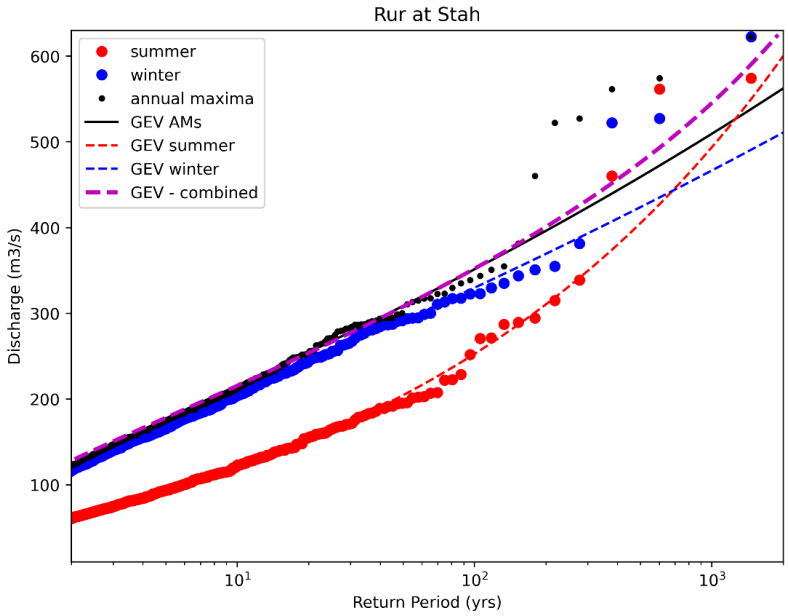


Figure 5-8: Combined GEV analysis for the Rur at Stah. Fitting a GEV model irrespective of summer and winter events leads to the black line (GEV AMs). Differentiating between winter and summer events as different populations leads to the blue and red line, respectively. These exceedance frequencies can 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.

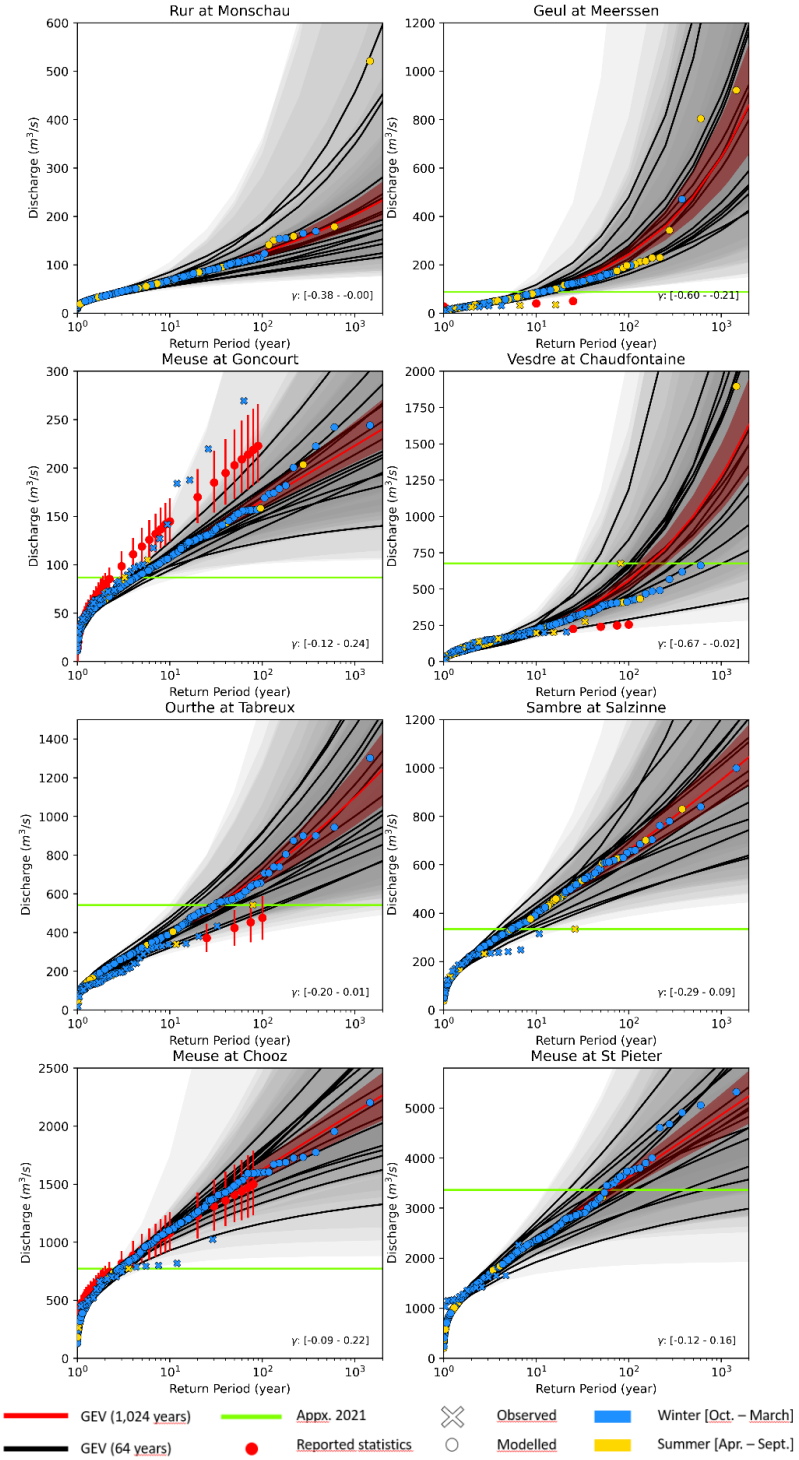


Figure 5-9:Extreme value distribution fit at selected locations across the Meuse catchment at an hourly time step. Black lines represent the GEV fit obtained from one ensemble member (64 hydrological years) and the range of shape parameters is written at the bottom right of each figure. Red lines show 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.

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).

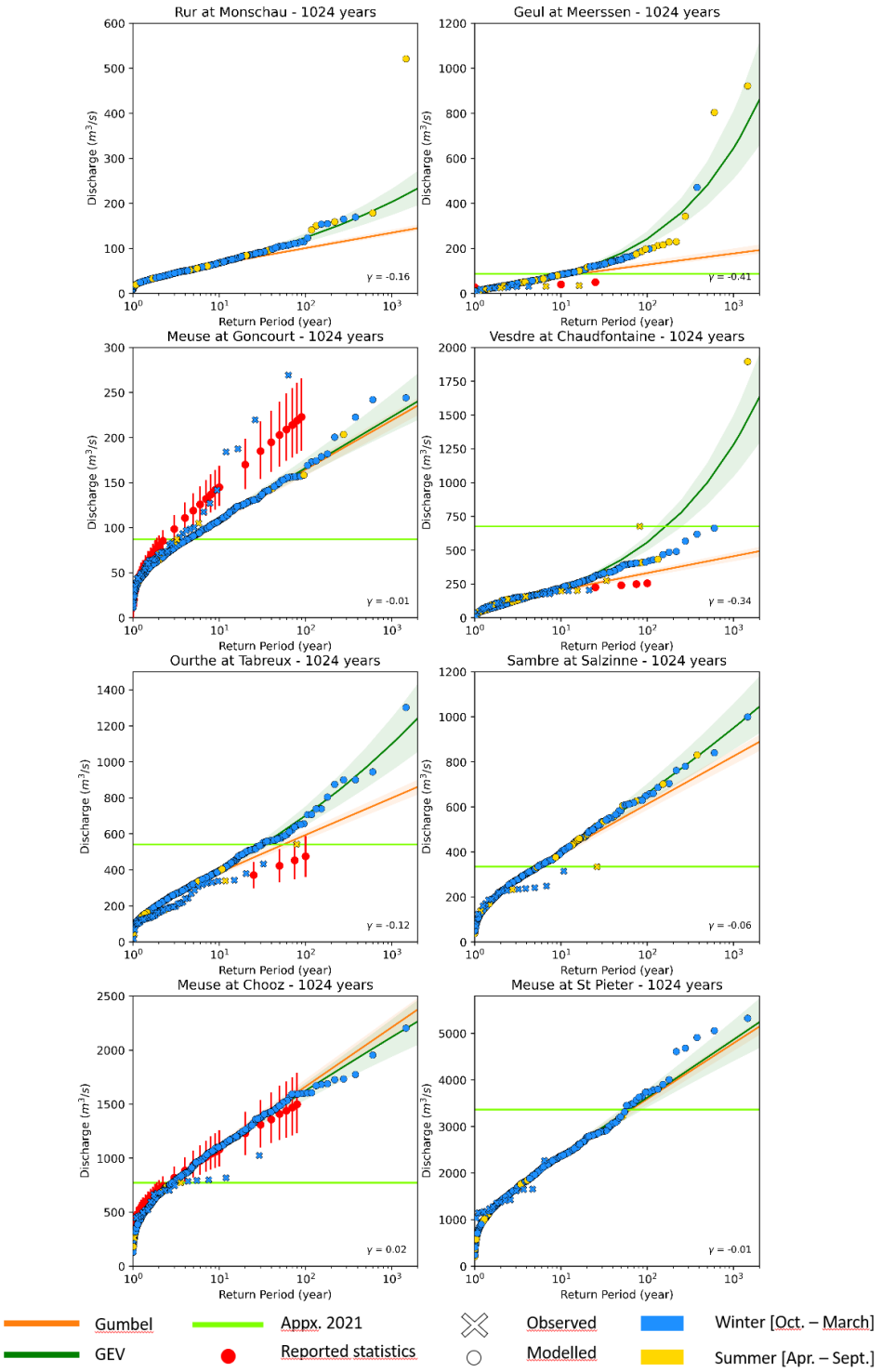


Figure 5-10: Extreme value distribution fit at selected locations across the Meuse catchment, based on hourly model results. Dark green lines represent the GEV fit obtained from combining all ensembles (1,024 hydrological years, shown as full dots) and blue lines the Gumbel fit. The shading represents 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.

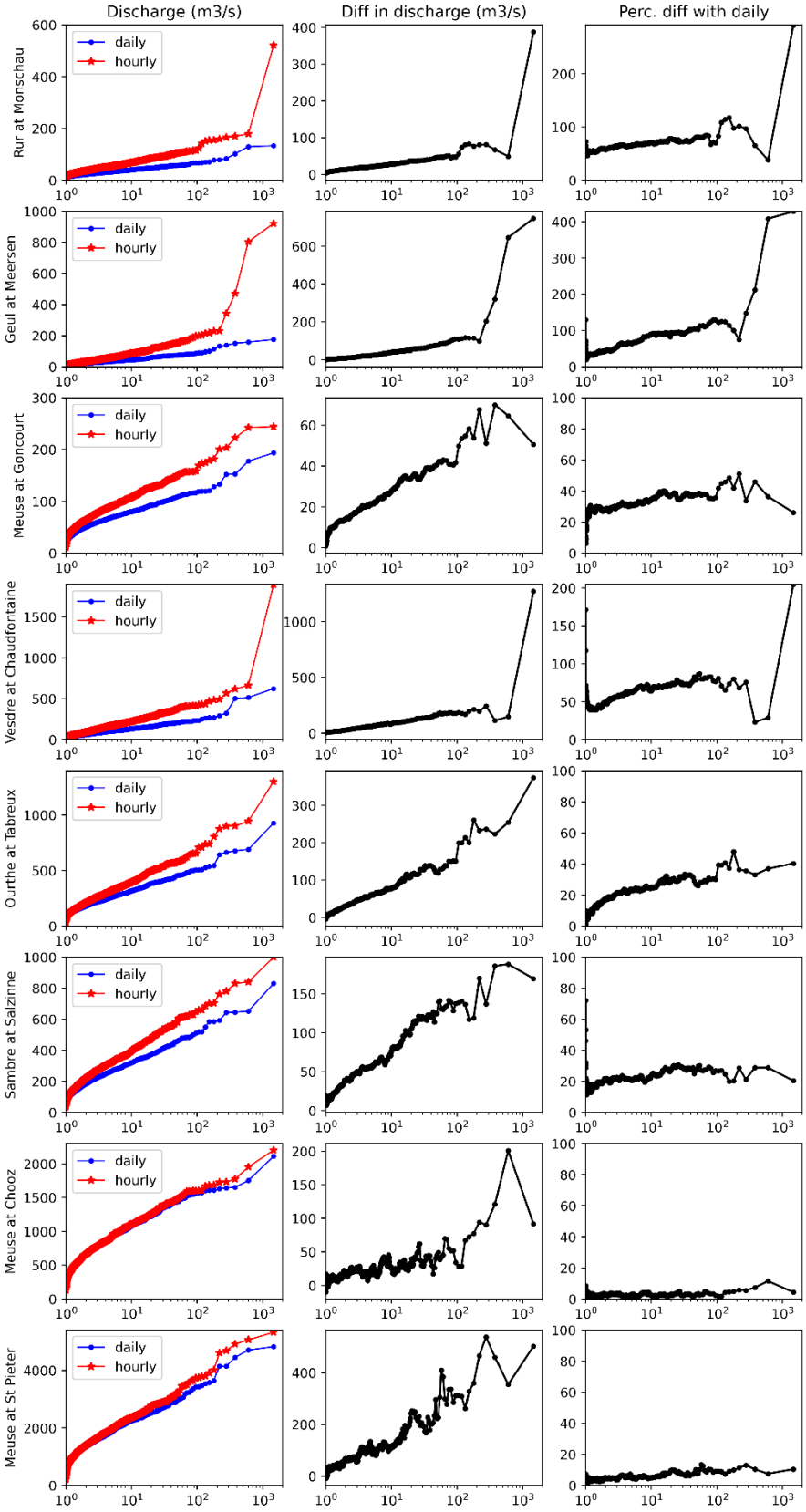


Figure 5-11. Left column - Empirical return levels obtained from 1,024 hydrological years at selected locations from the daily and hourly runs. Middle column: Difference in discharge between the two temporal resolutions. Right column: Relative percentage difference with respect to the empirical daily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

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.

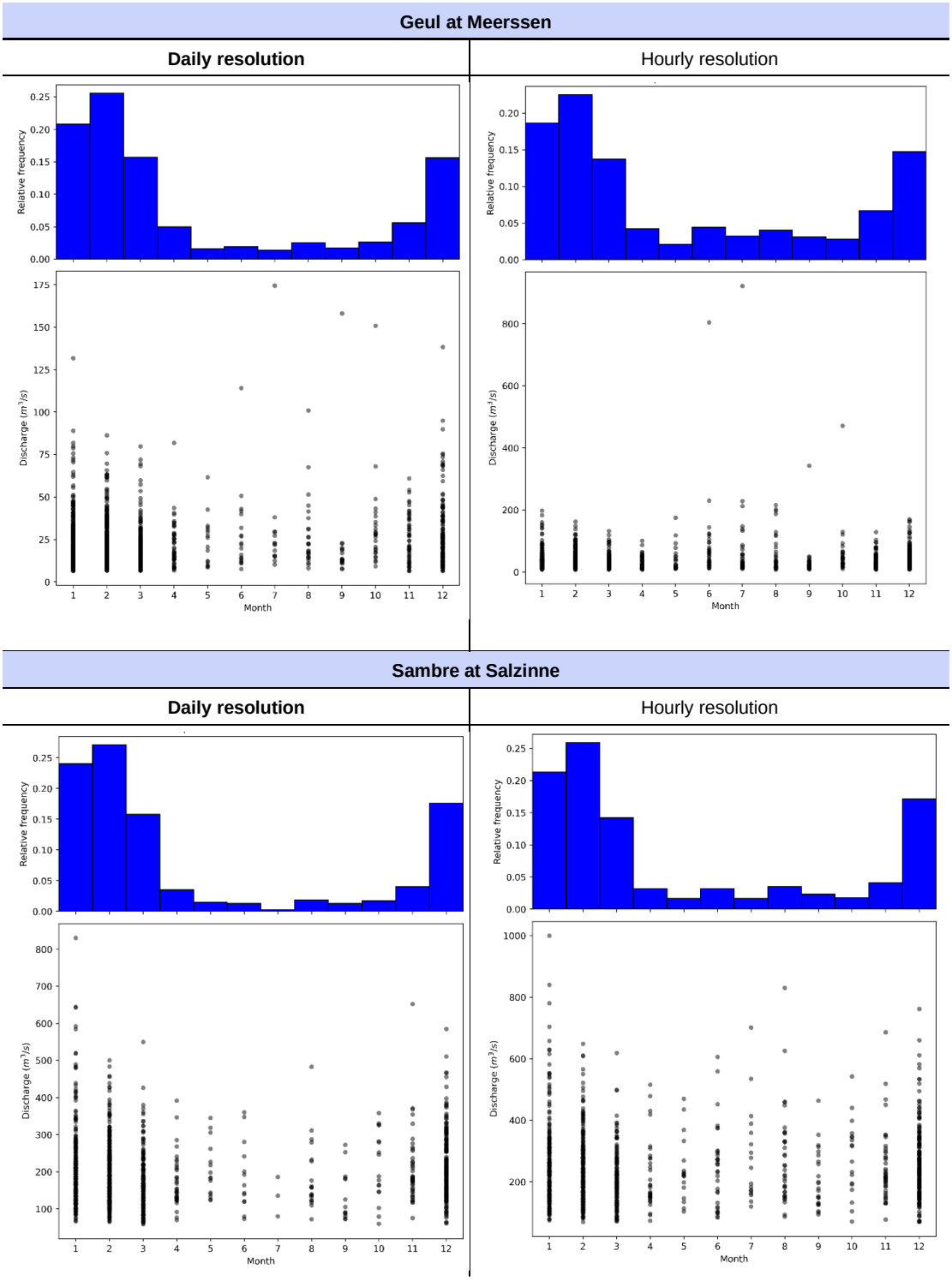


Figure 5-12: Relative frequency of the month of occurrence of the discharge extremes for the Geul at Meerssen (top) and the Sambre at Salzinne (bottom) for a daily time step (left column) and hourly time step (right column).

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

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)	
	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 essentiaI. 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

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.

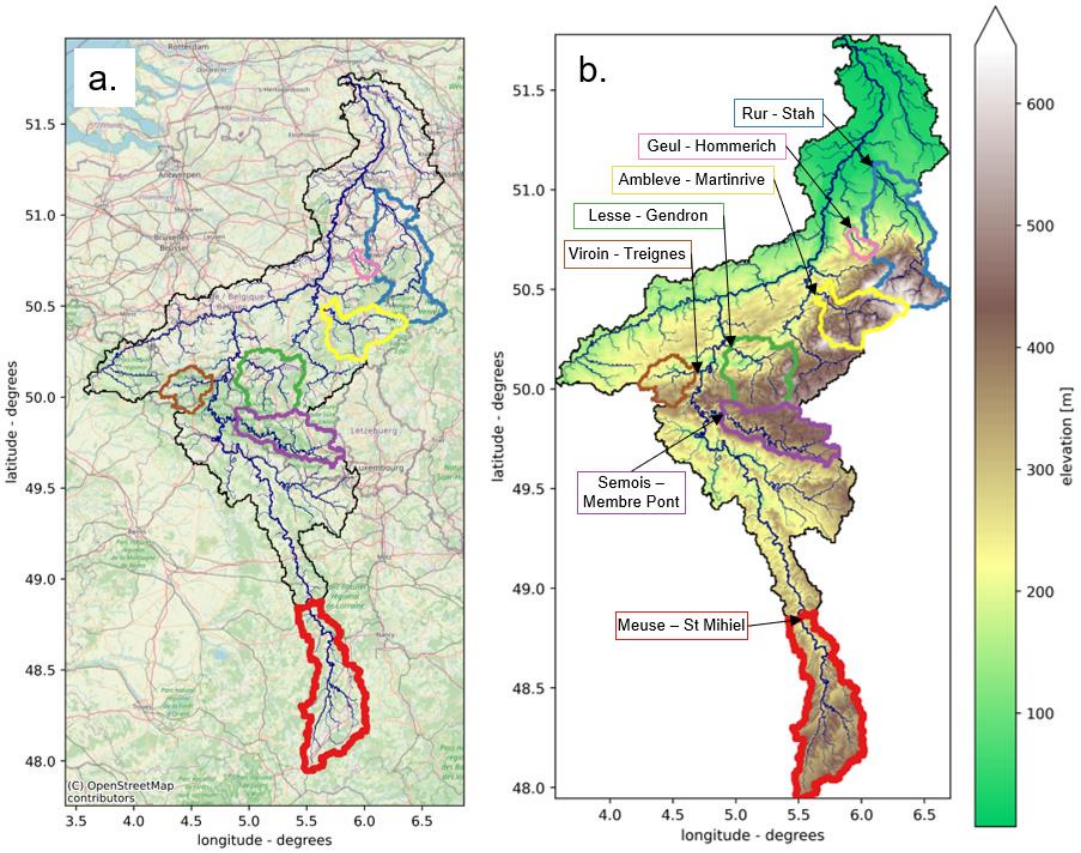


Figure A-1: Selected catchments for the Meuse river and her tributaries, with background map (a) and topography (b) presented in the appendix.

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).

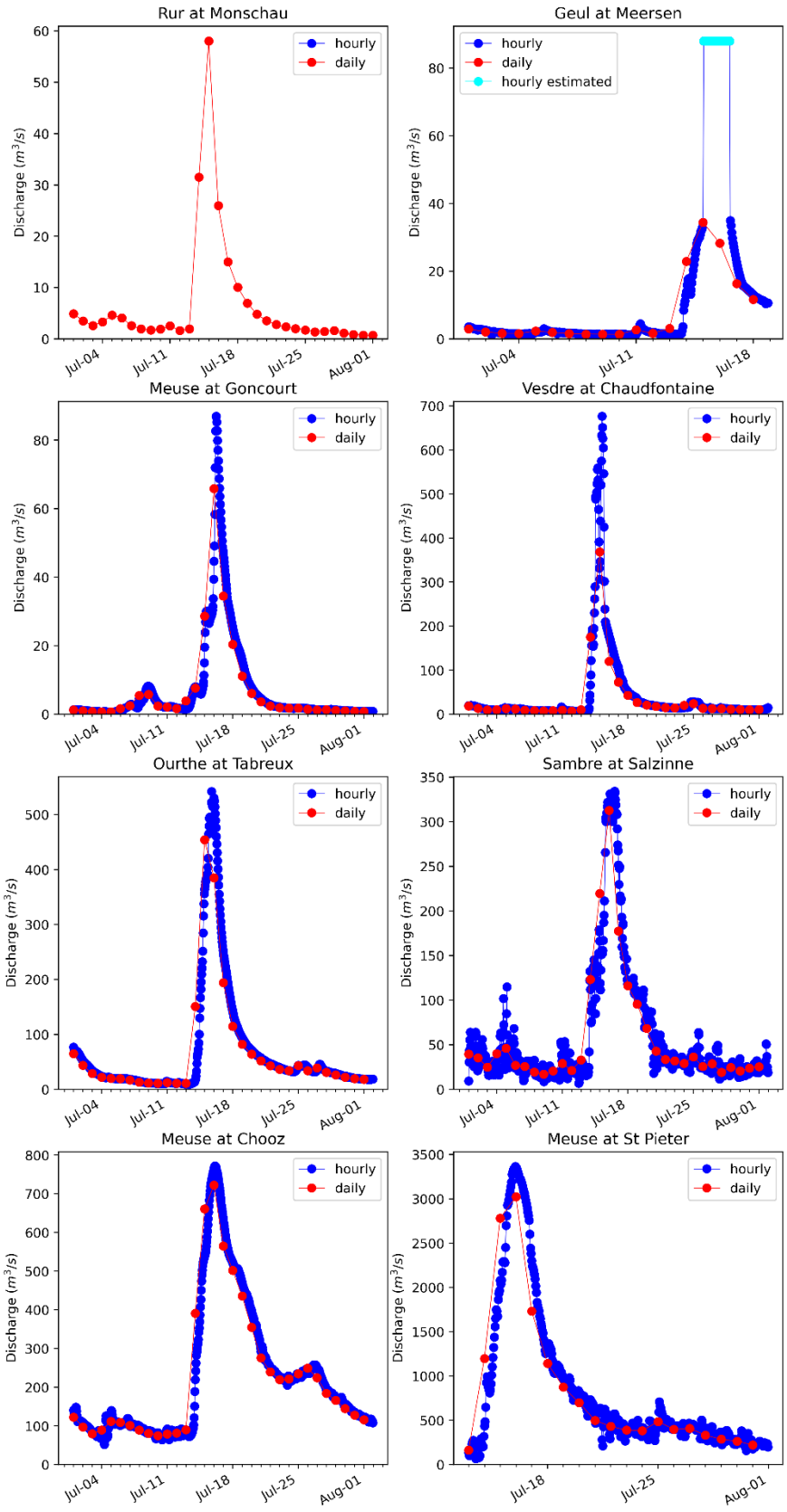


Figure A-2: Recorded or estimated discharge for the selected catchment of interests shown in the main report between July 1st and August 1st 2021.

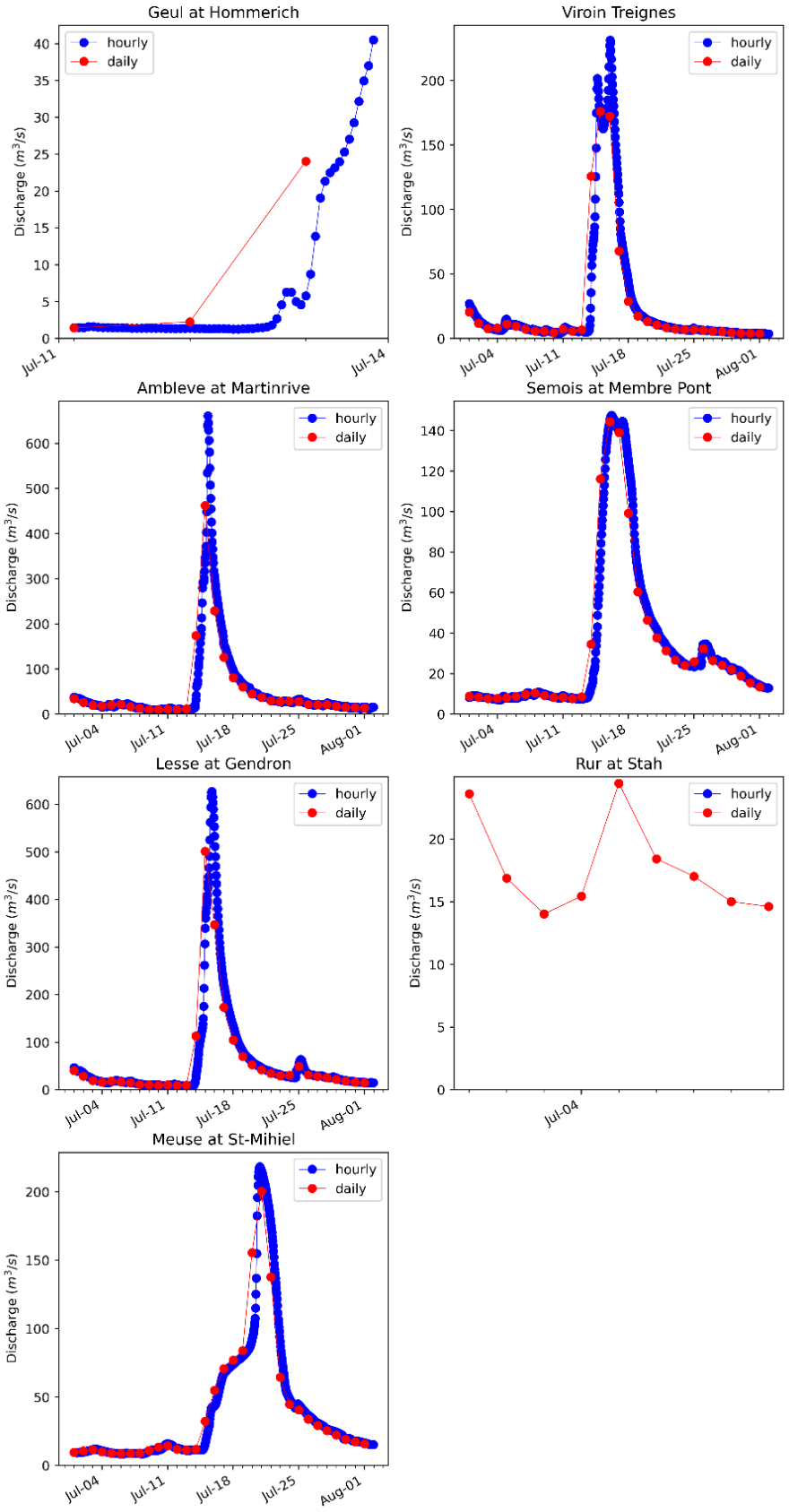


Figure A-3: Recorded or estimated discharge for the selected catchment of interests shown in the appendix, section D, between July 1st and August 1st 2021.

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

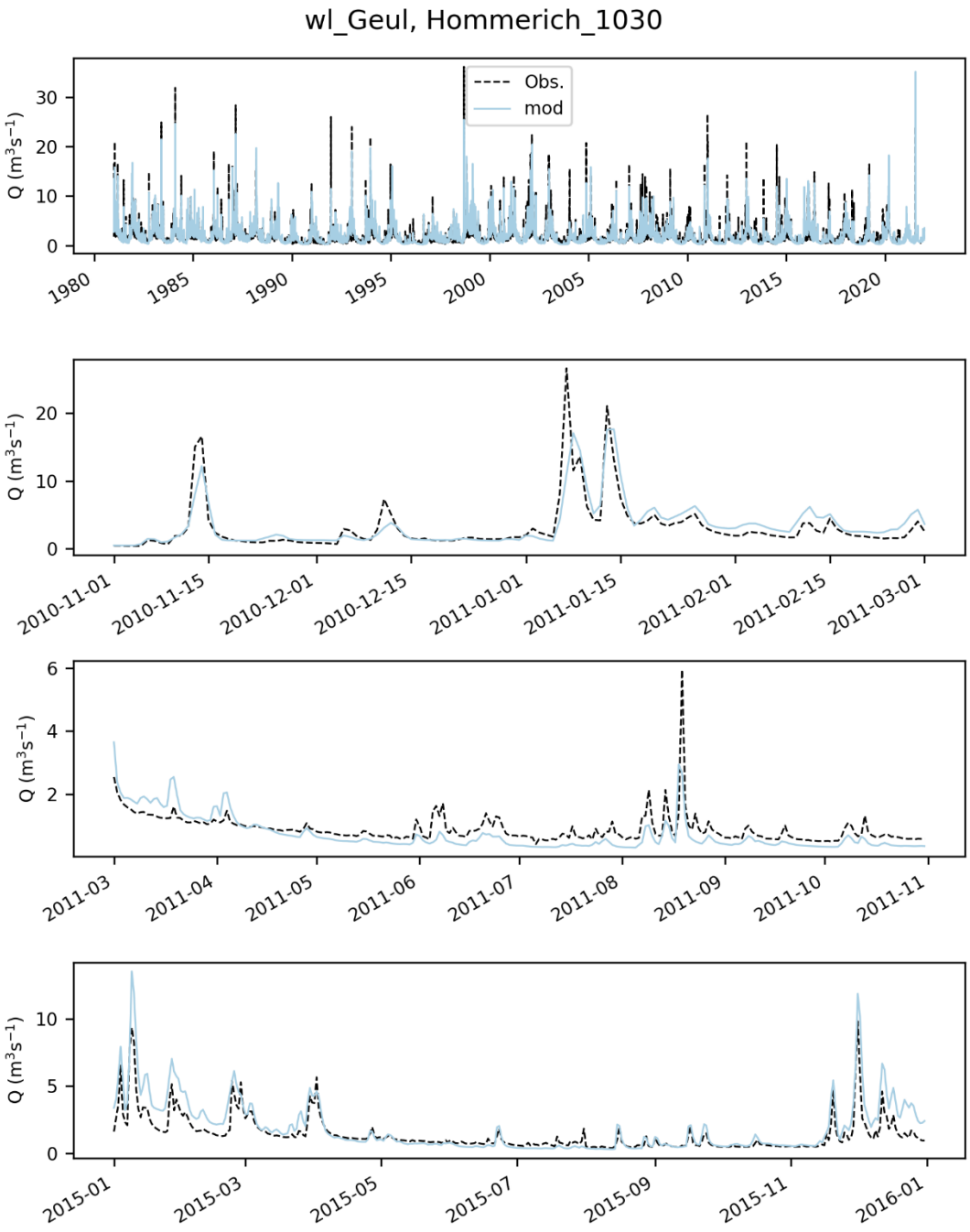


Figure B-1 Modelled (blue) and observed (black) hydrographs for the Geul at Hommerich

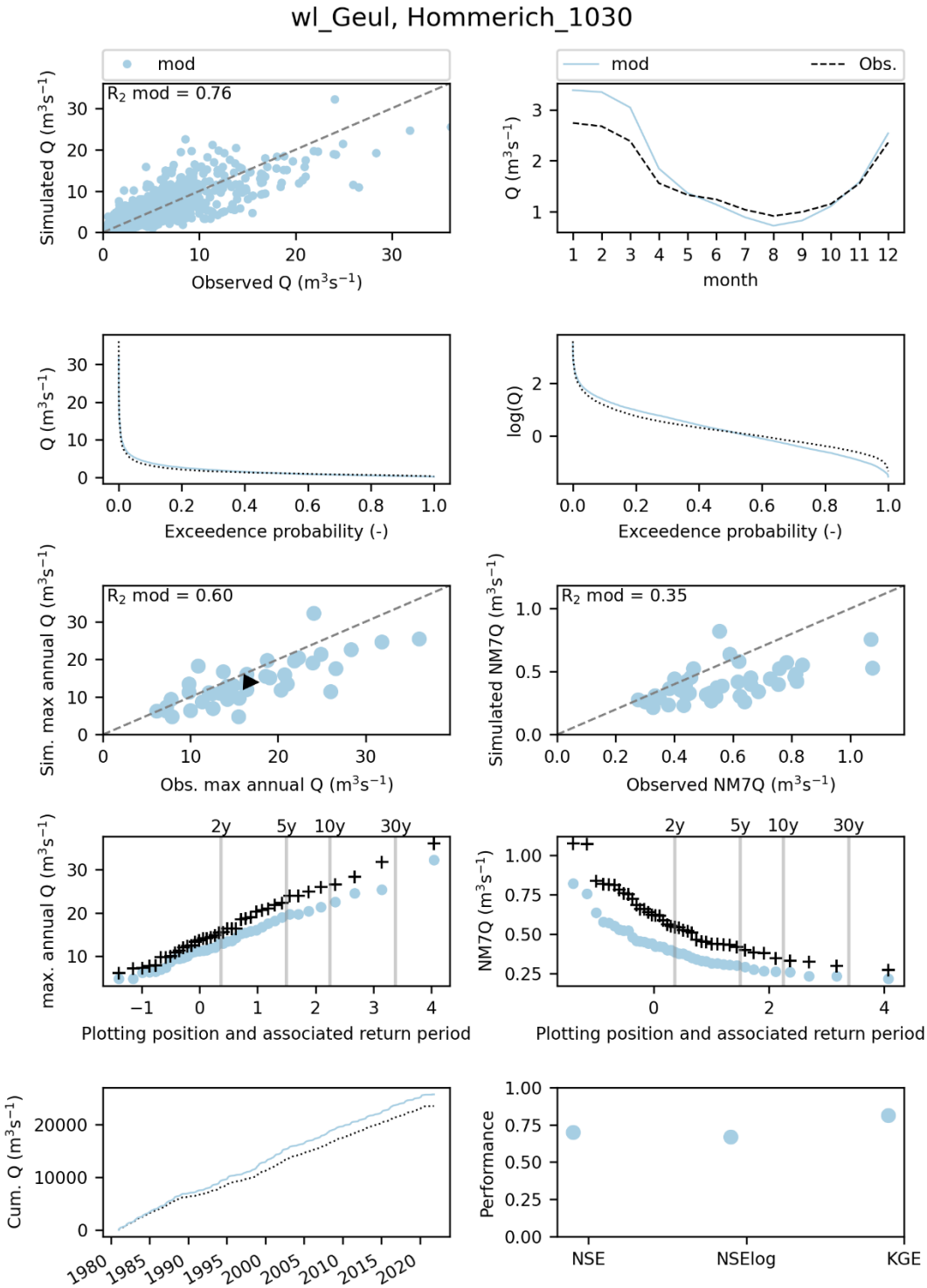


Figure B-2 Modelled (blue) and observed (black) signatures for the Geul at Hommerich

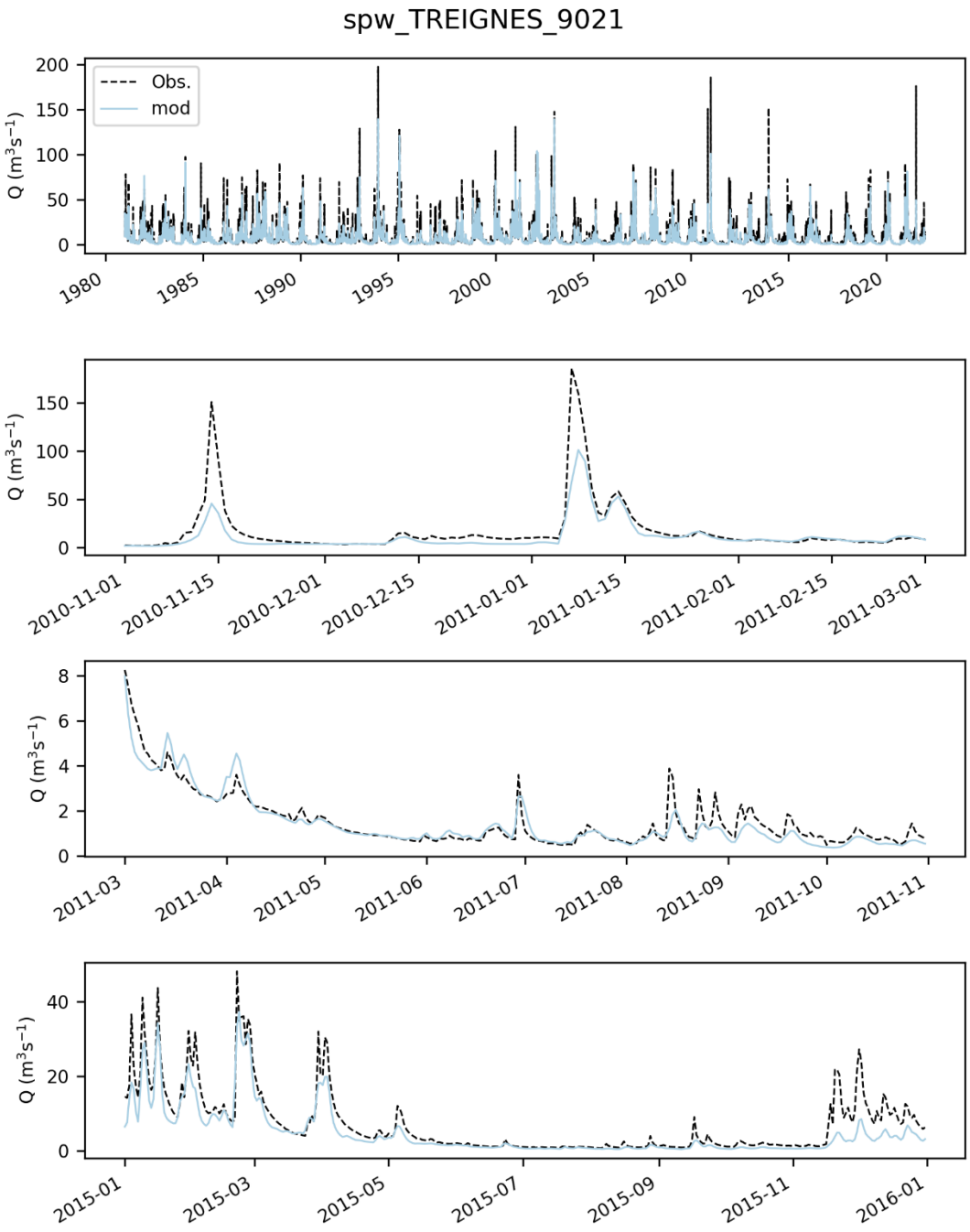


Figure B-3 Modelled (blue) and observed (black) hydrographs for the Viroin at Treignes

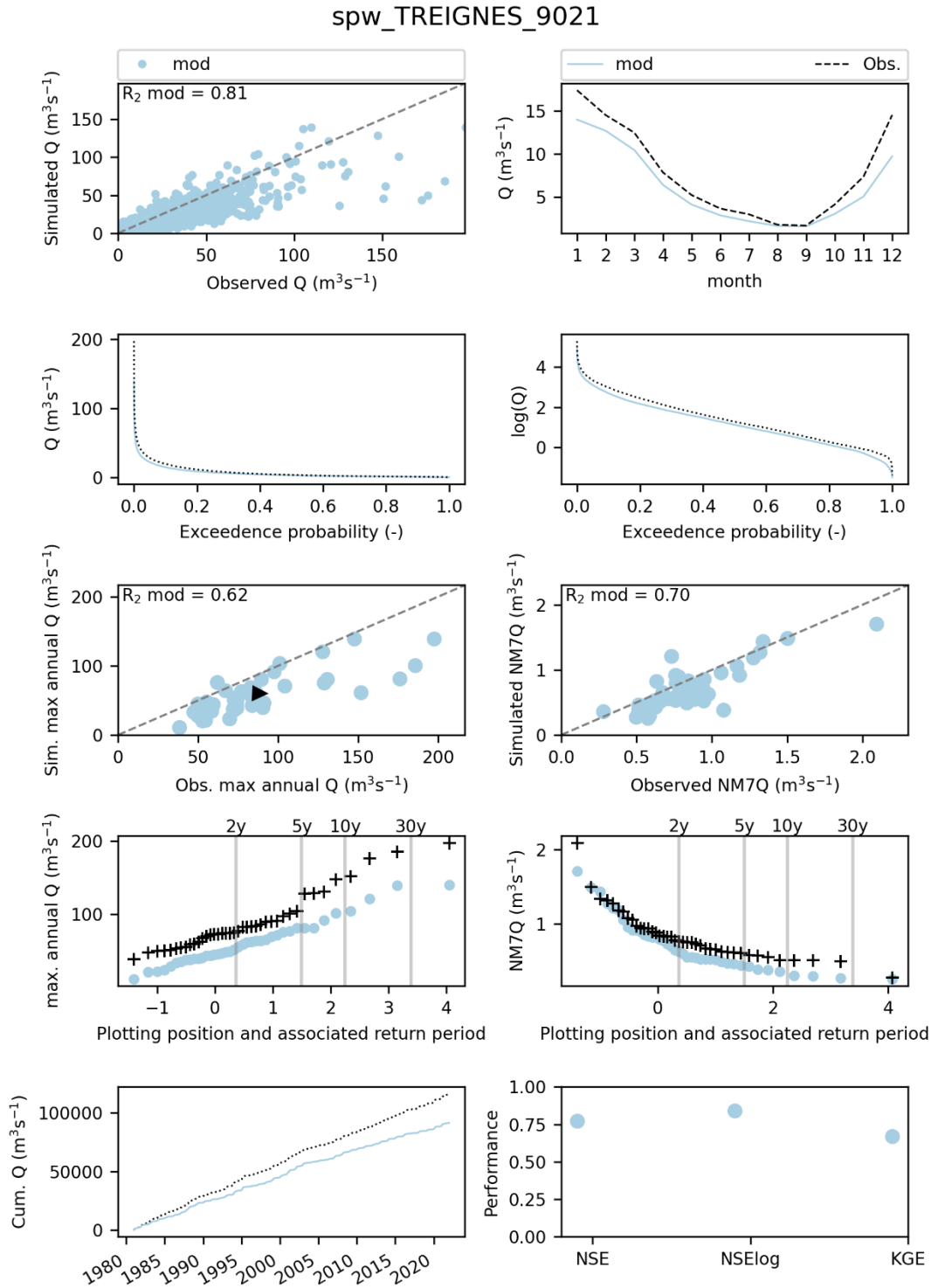


Figure B-4 Modelled (blue) and observed (black) signatures for the Viroin at Treignes

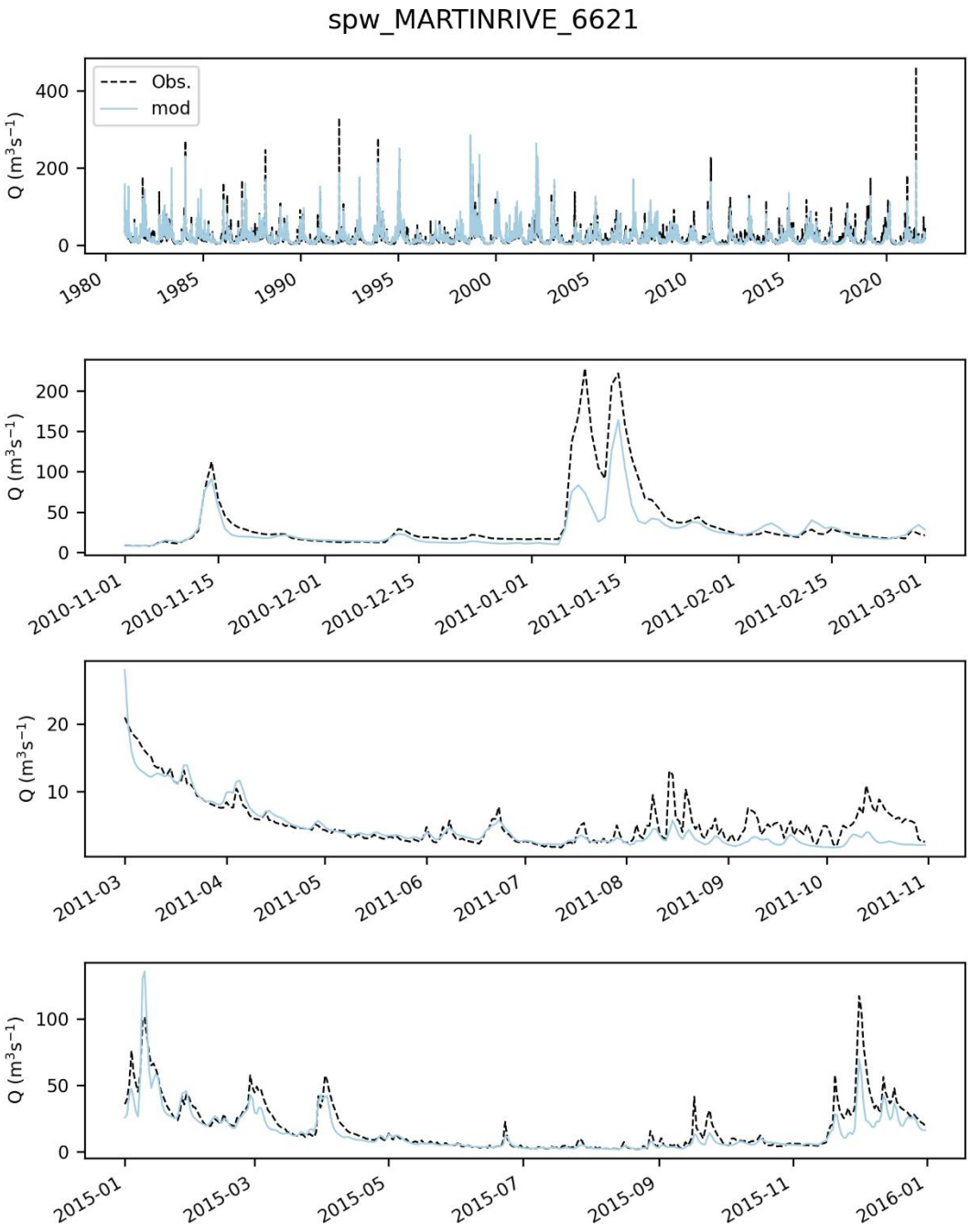


Figure B-5 Modelled (blue) and observed (black) hydrographs for the Ambleve at Martinrive

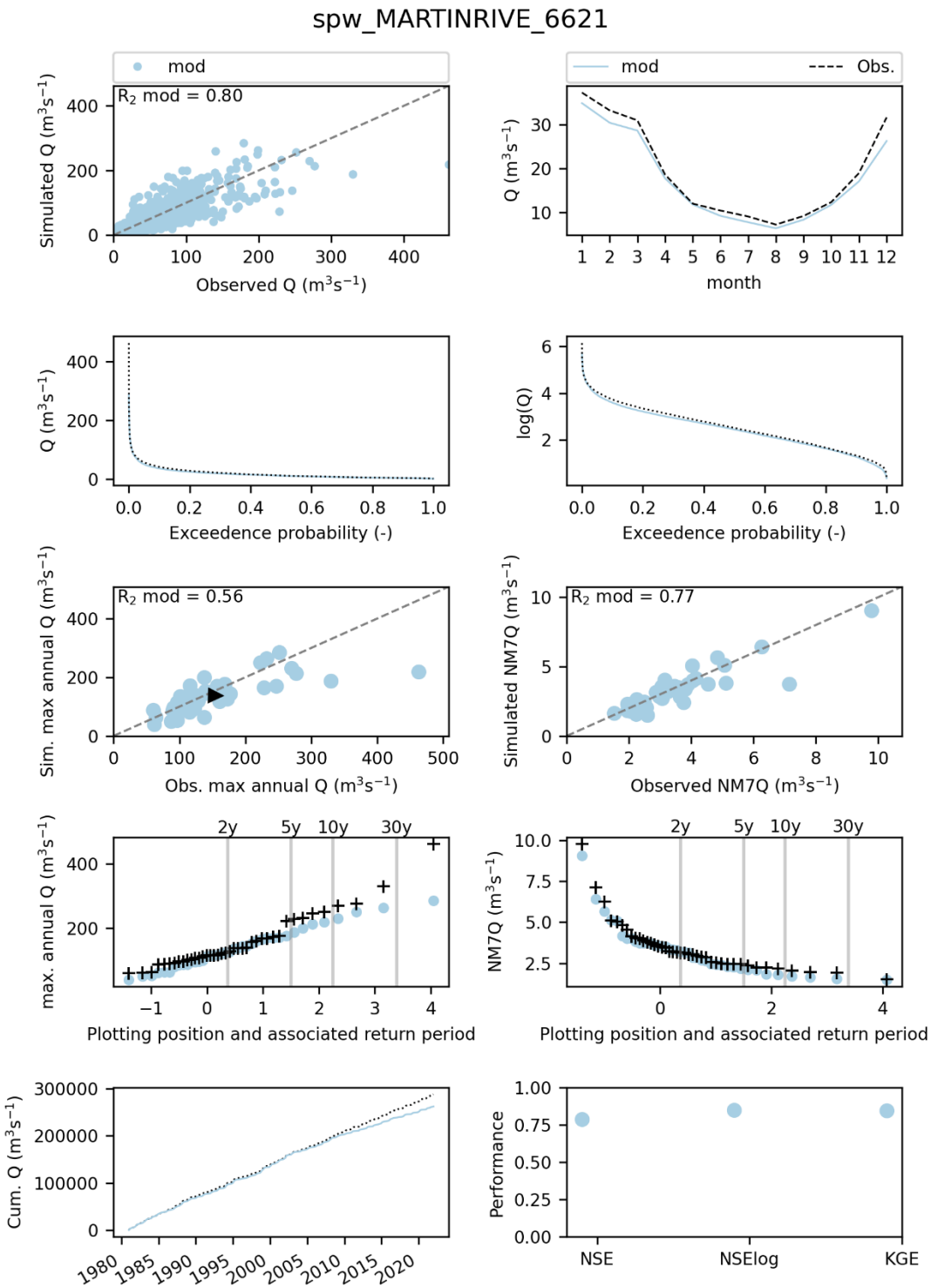


Figure B-6 Modelled (blue) and observed (black) signatures for the Ambleve at Martinrive

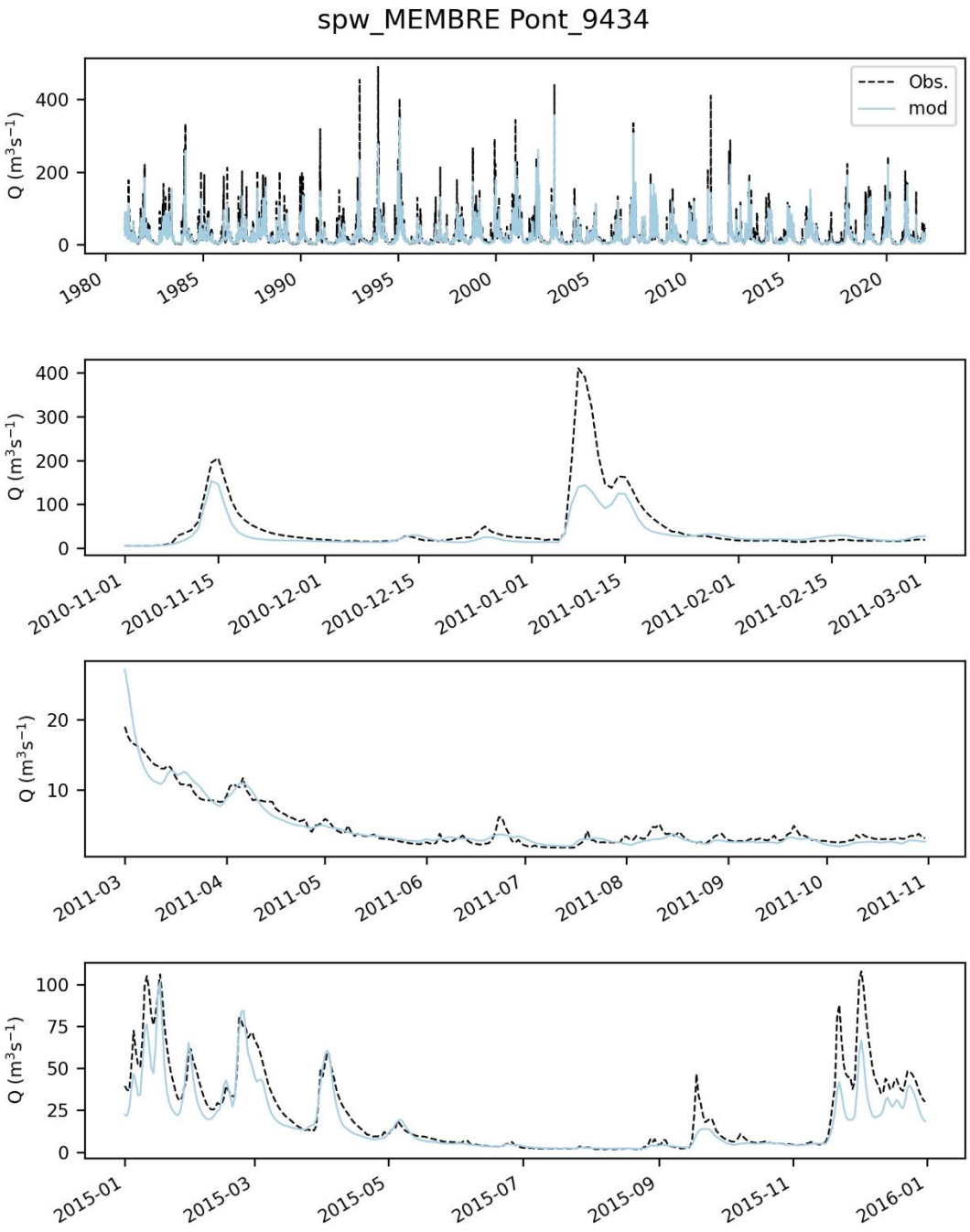


Figure B-7 Modelled (blue) and observed (black) hydrographs for the Semois at Membre-Pont

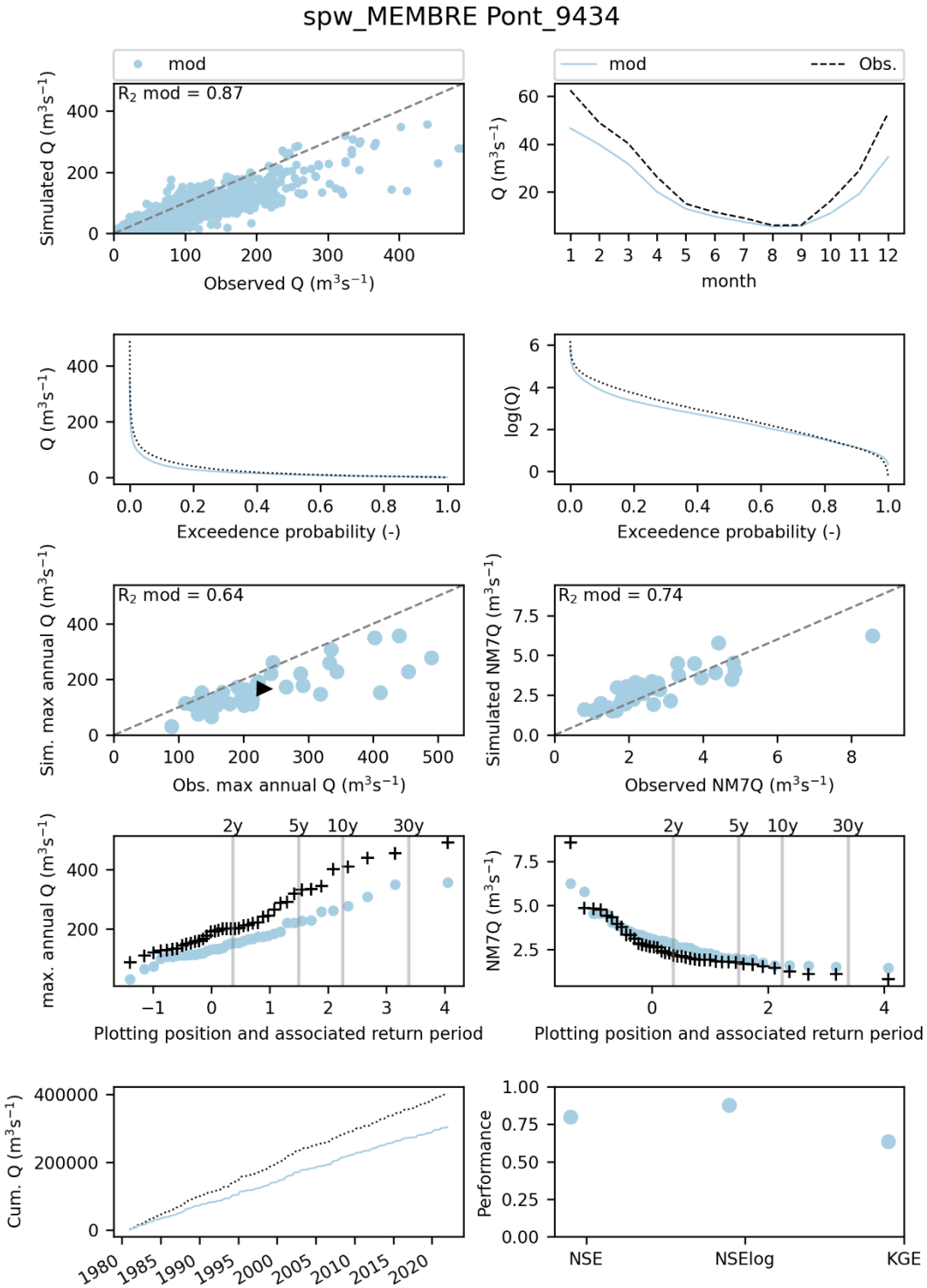


Figure B-8 Modelled (blue) and observed (black) signatures for the Semois at Membre-Pont

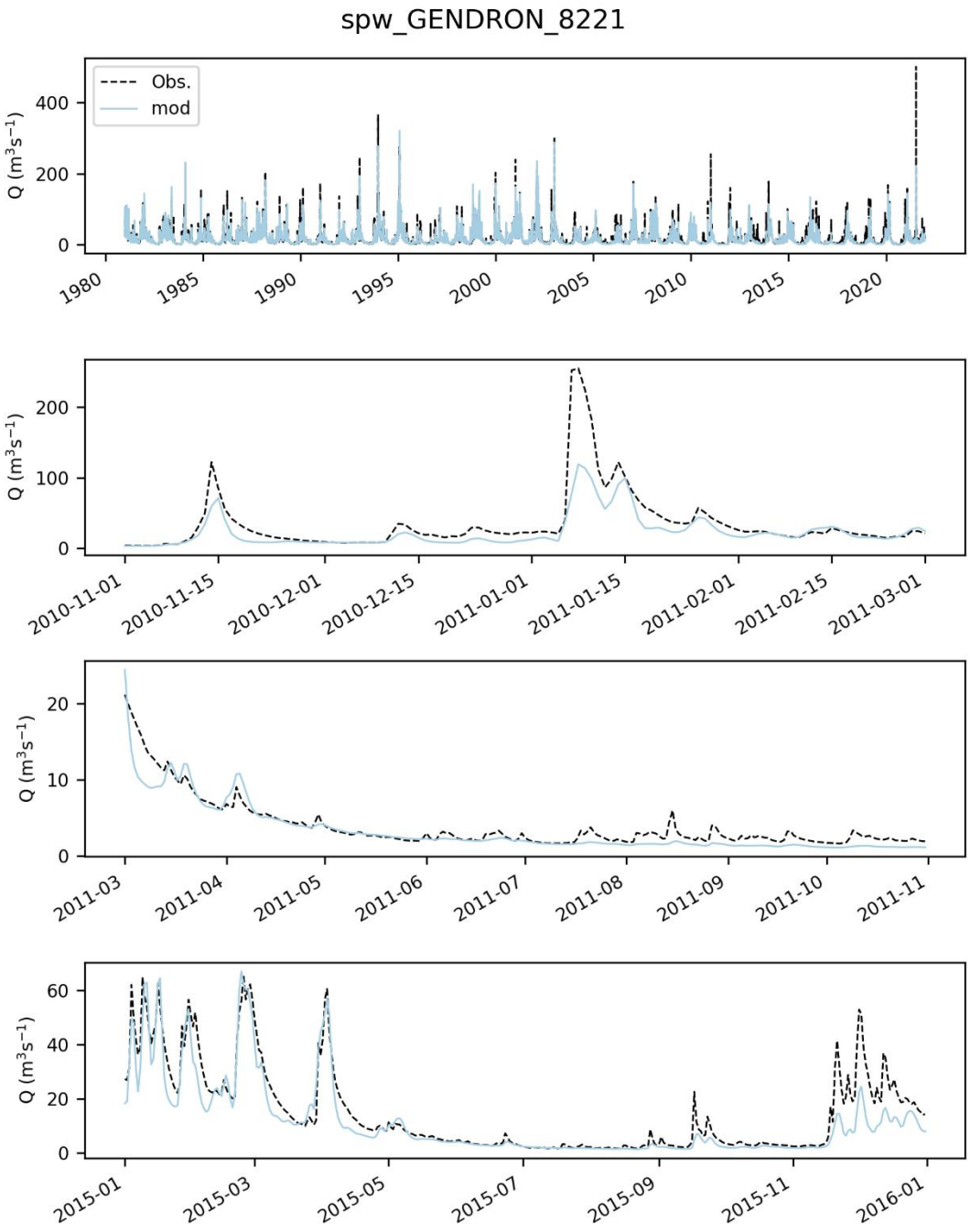


Figure B-9 Modelled (blue) and observed (black) hydrographs for the Lesse at Gendron

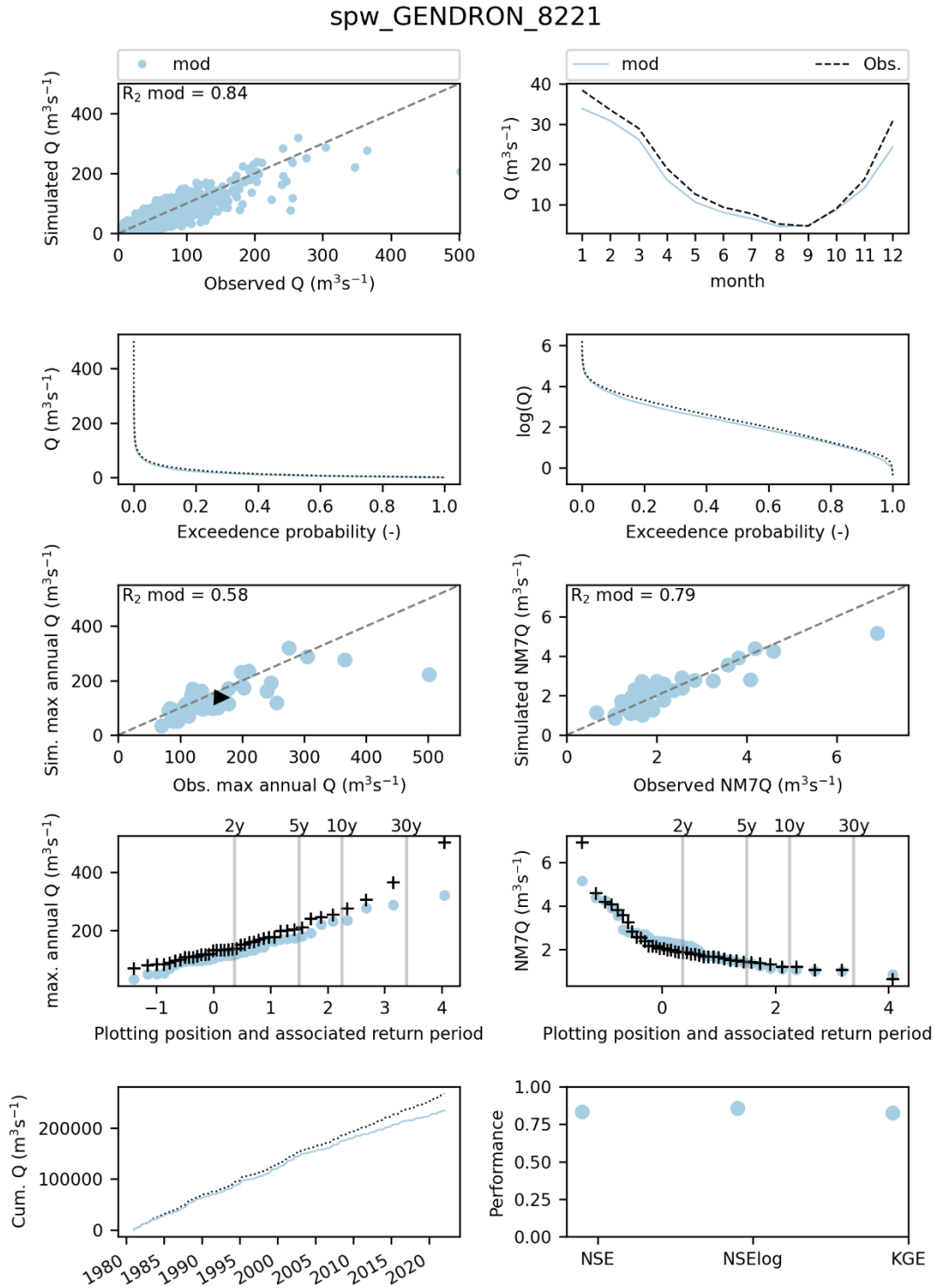


Figure B-10 Modelled (blue) and observed (black) signatures for the Lesse at Gendron

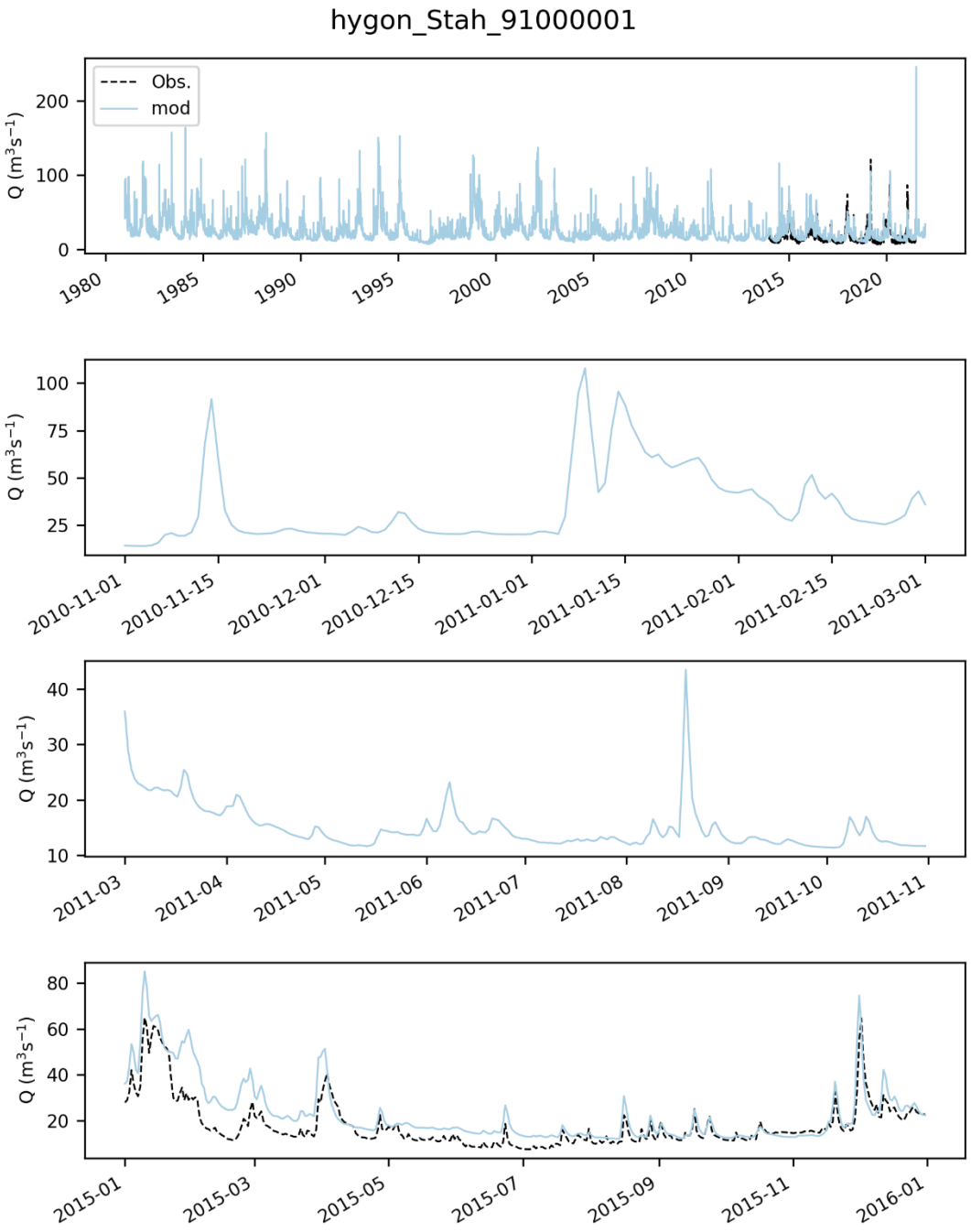


Figure B-11 Modelled (blue) and observed (black) hydrographs for the Rur at Stah

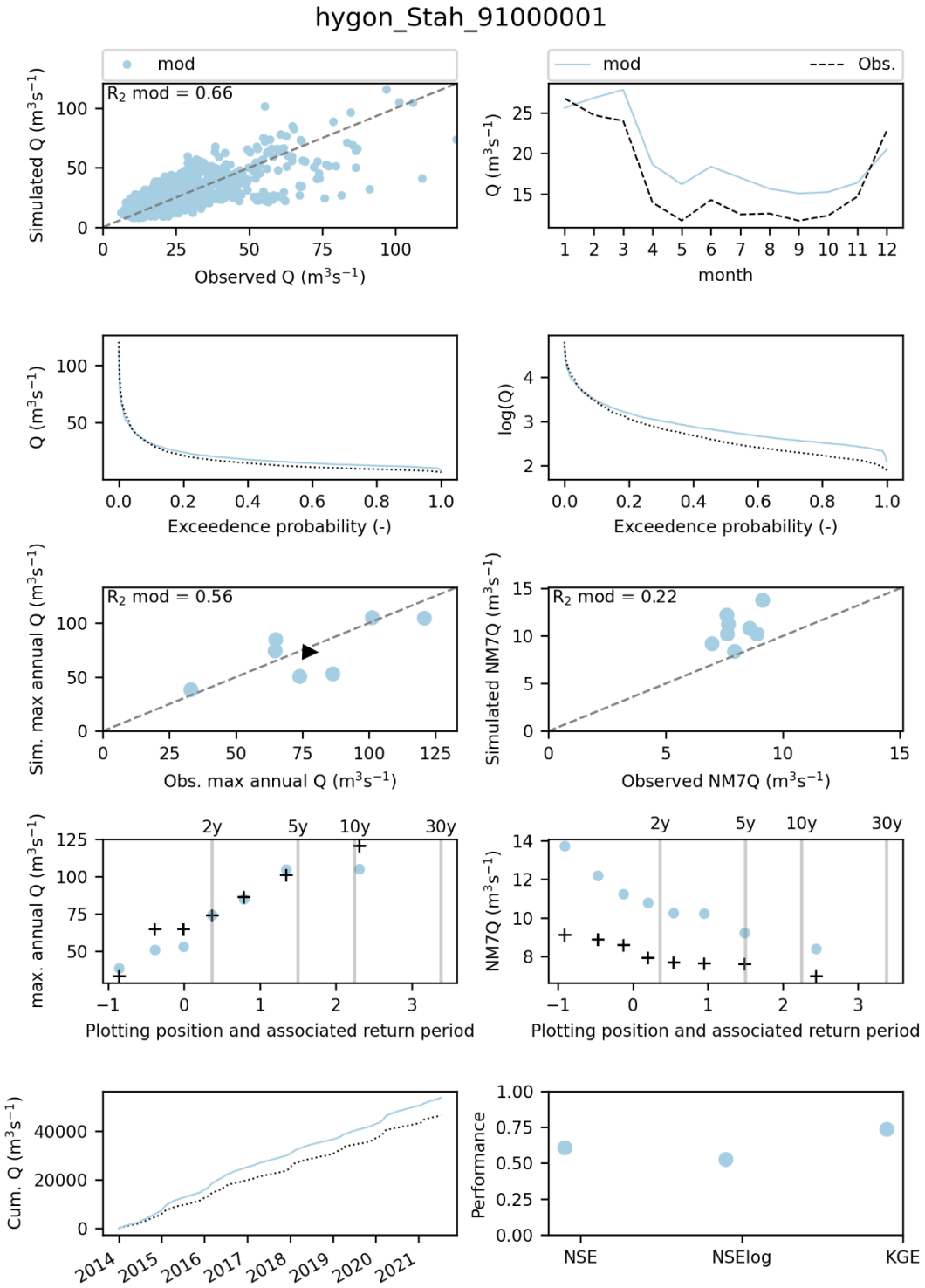


Figure B-12 Modelled (blue) and observed (black) signatures for the Rur at Stah

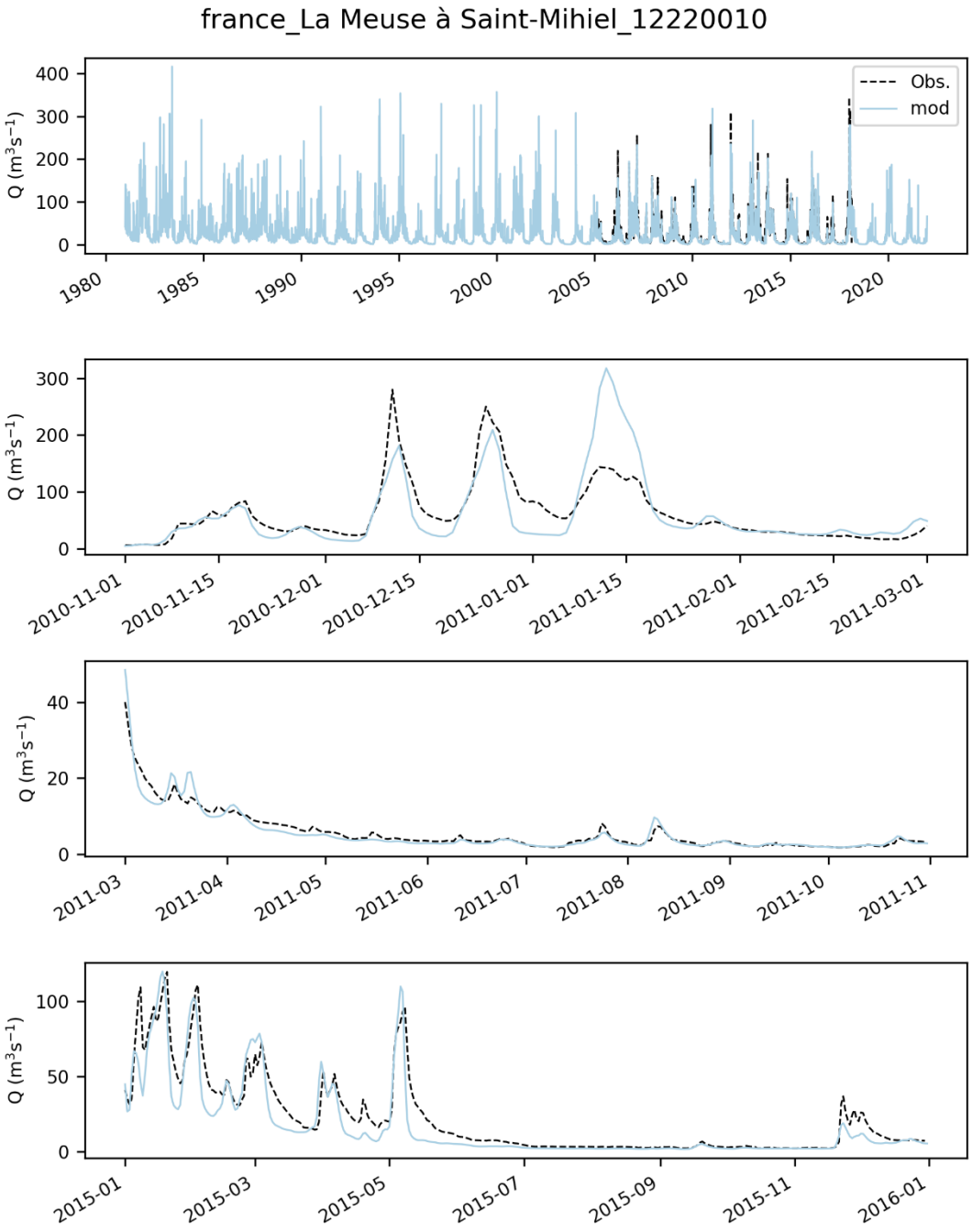


Figure B-13 Modelled (blue) and observed (black) hydrographs for the Meuse at St Mihiel

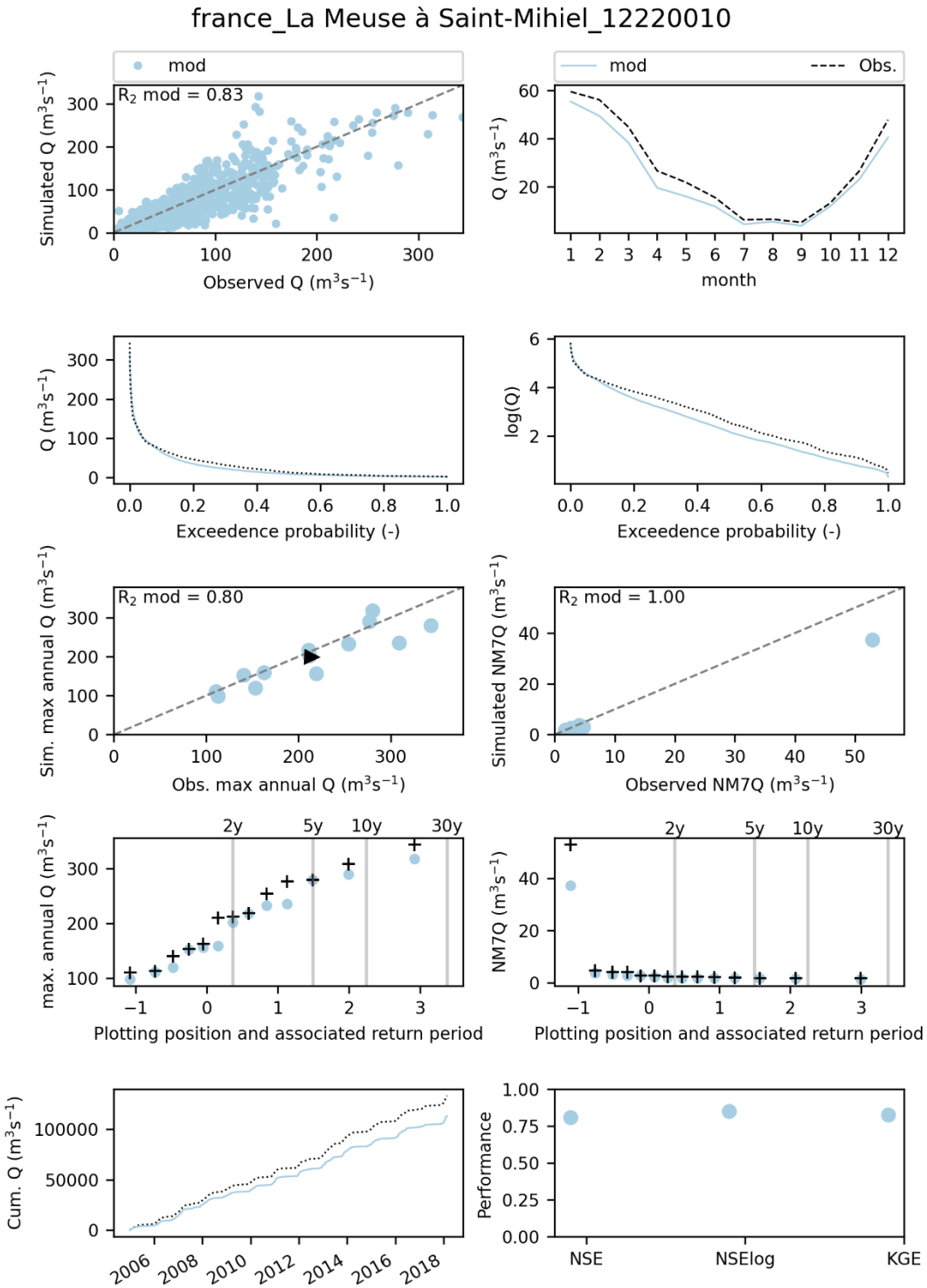


Figure B-14 Modelled (blue) and observed (black) signatures for the Meuse at St Mihiel

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

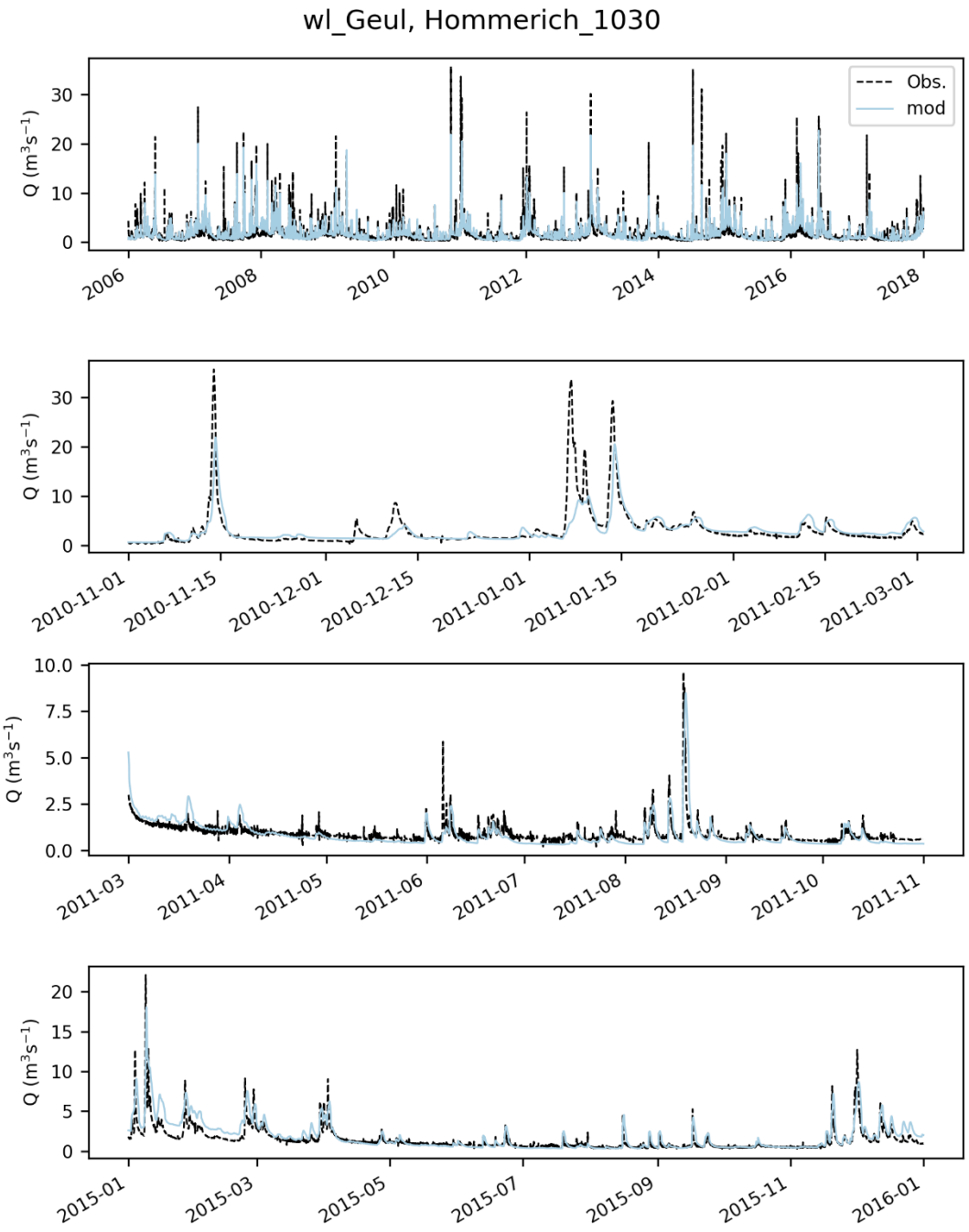


Figure B-15 Modelled (blue) and observed (black) hydrographs for the Geul at Hommerich

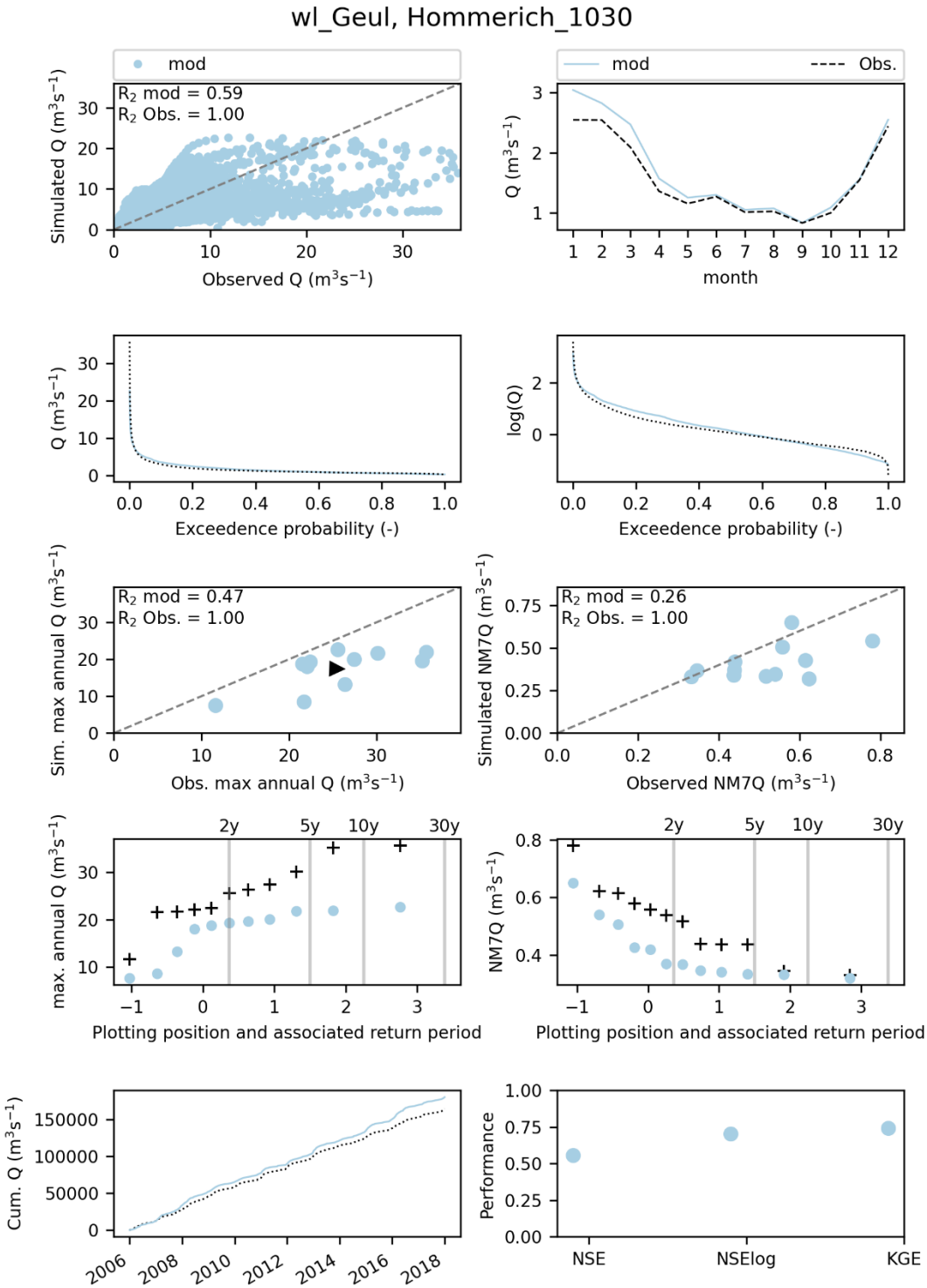


Figure B-16 Modelled (blue) and observed (black) signatures for the Geul at Hommerich

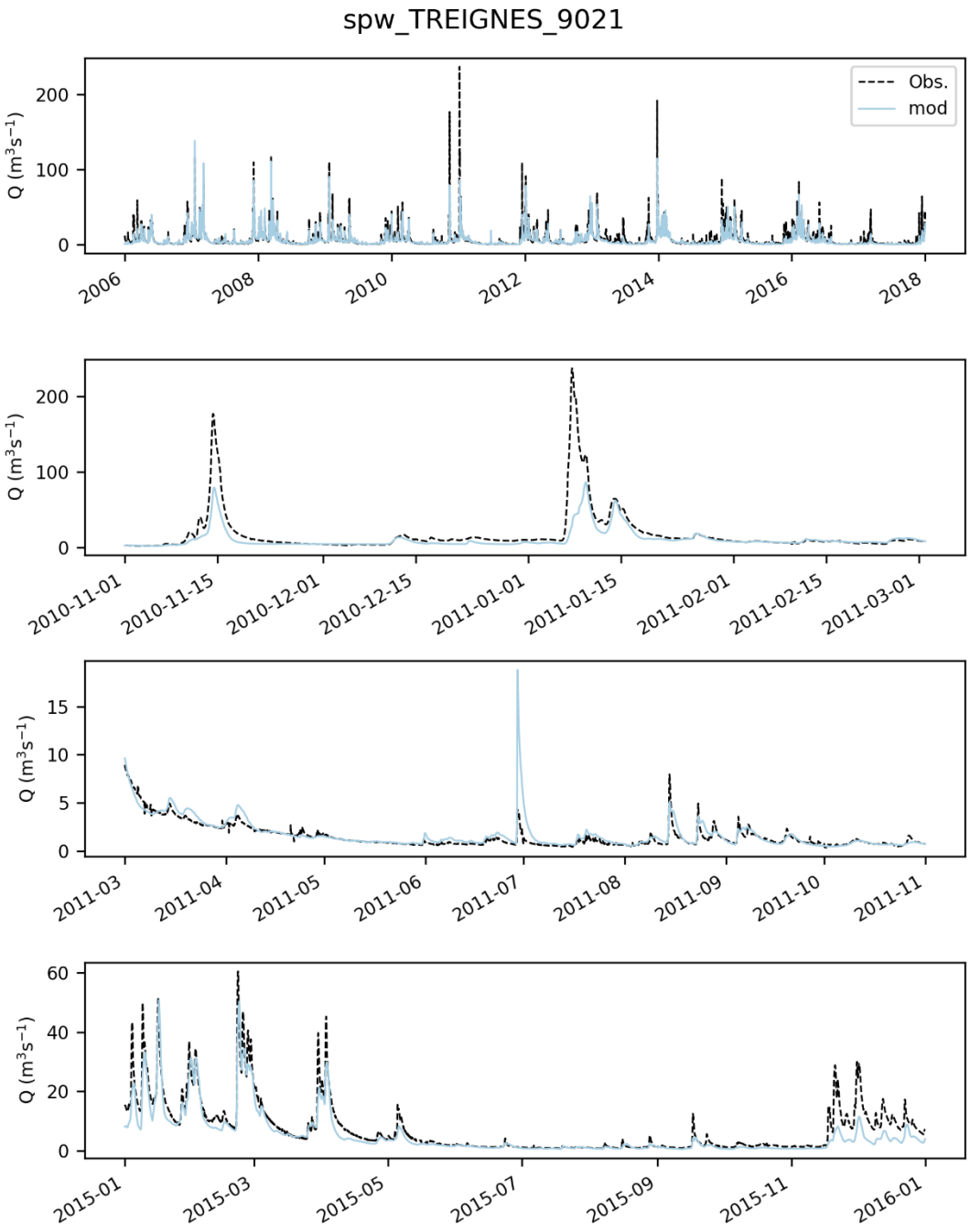


Figure B-17 Modelled (blue) and observed (black) hydrographs for the Viroin at Treigne

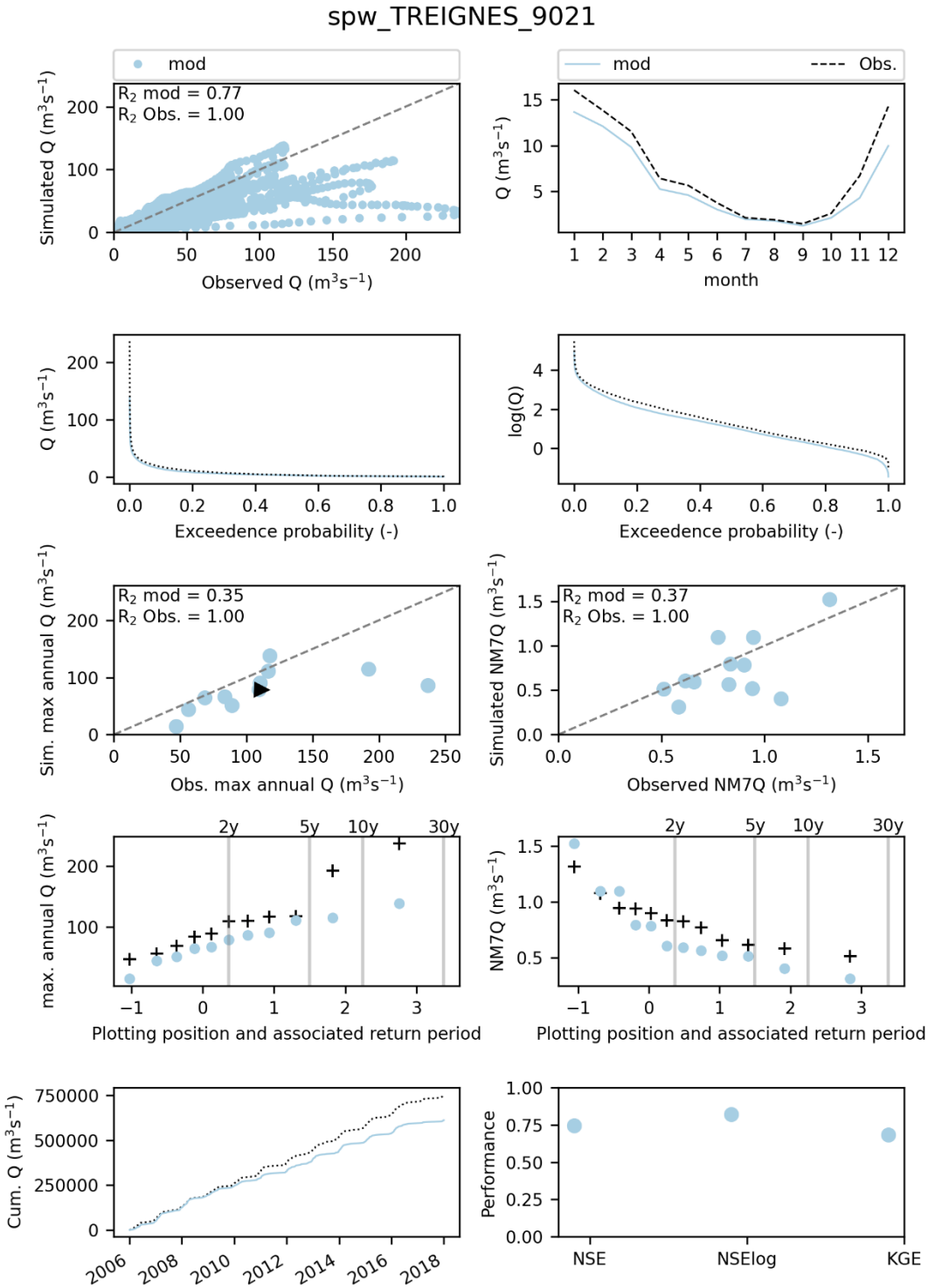


Figure B-18 Modelled (blue) and observed (black) signatures for the Viroin at Treigne

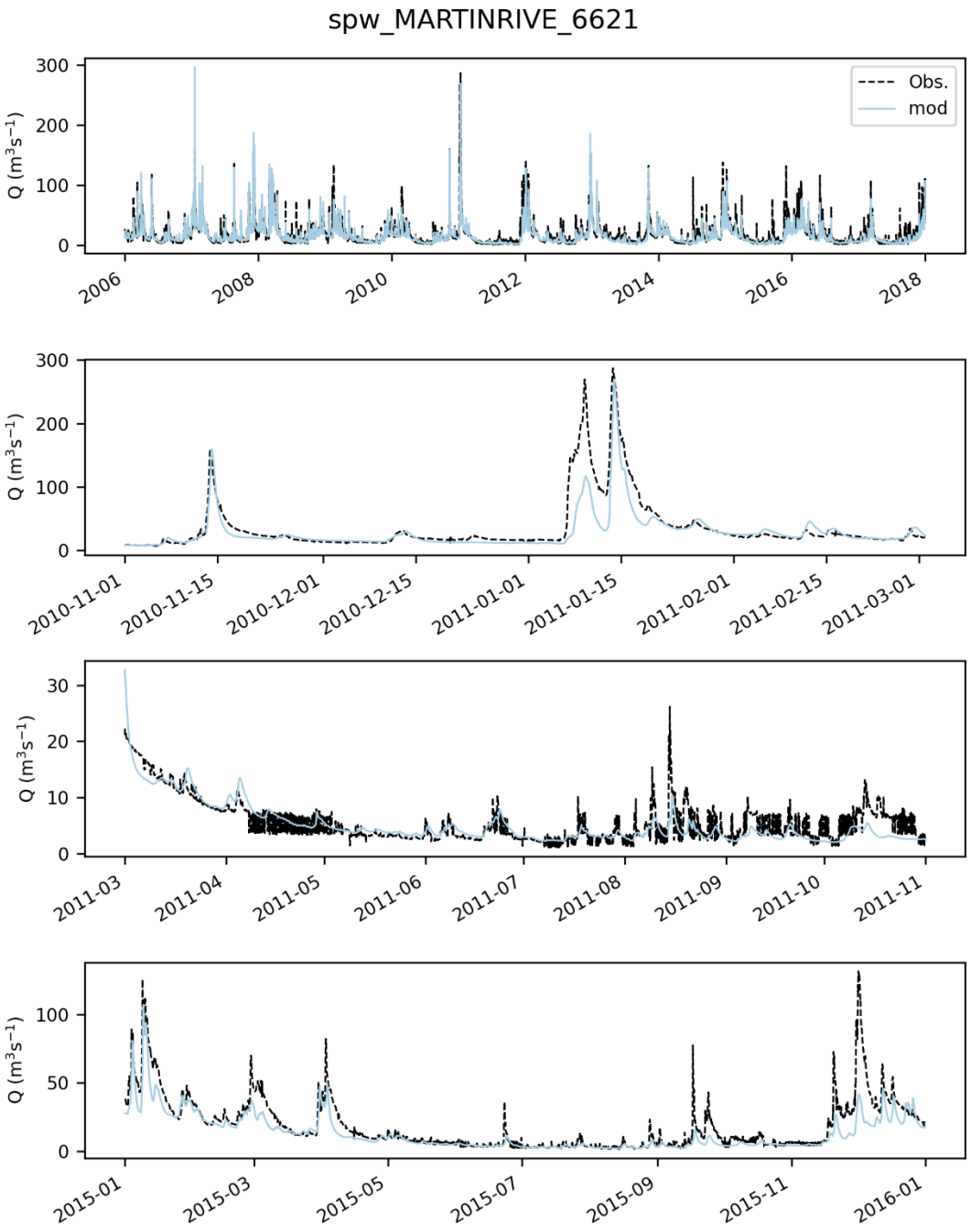


Figure B-19 Modelled (blue) and observed (black) hydrographs for the Ambleve at Martinrive

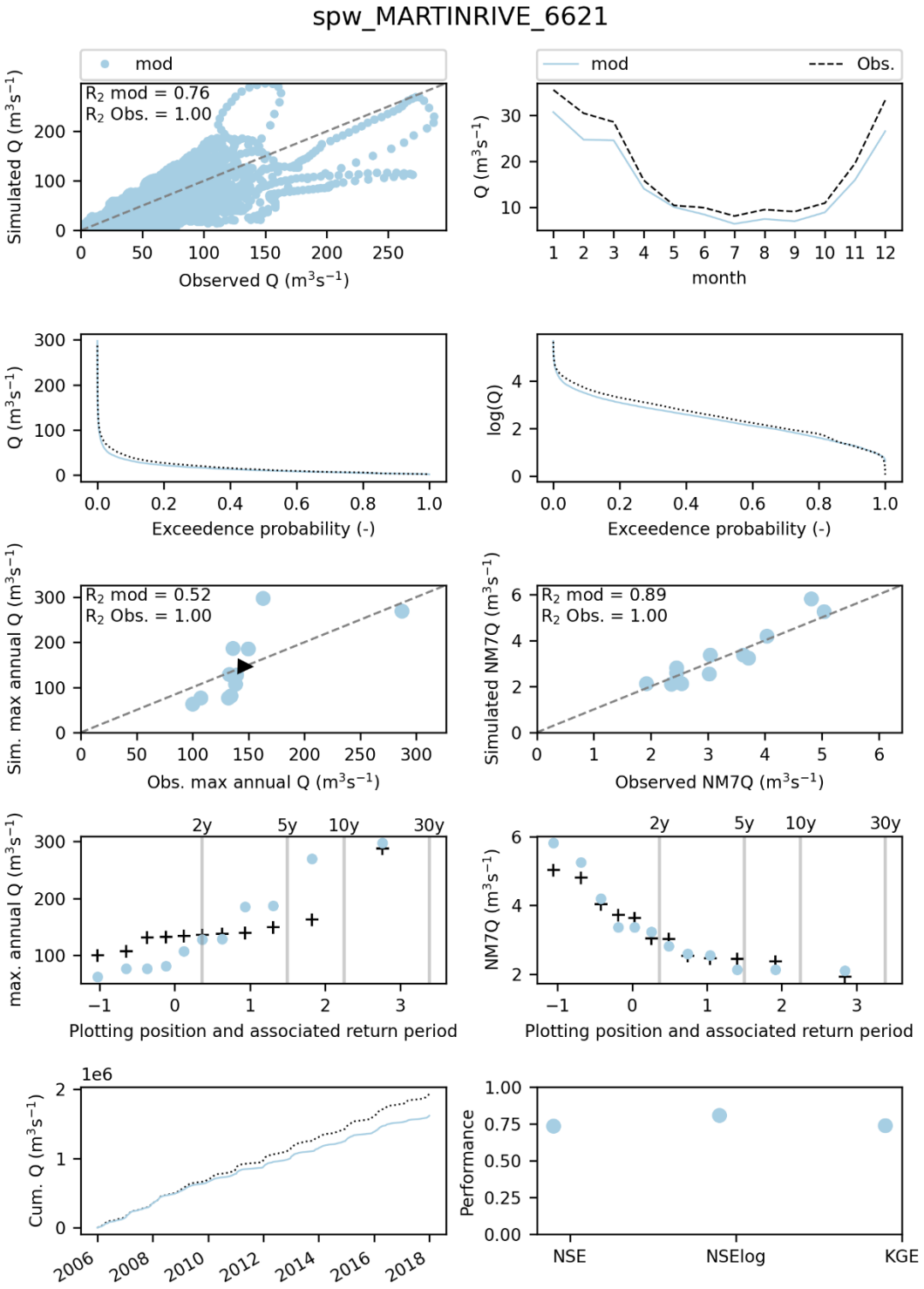


Figure B-20 Modelled (blue) and observed (black) signatures for the Ambleve at Martinrive

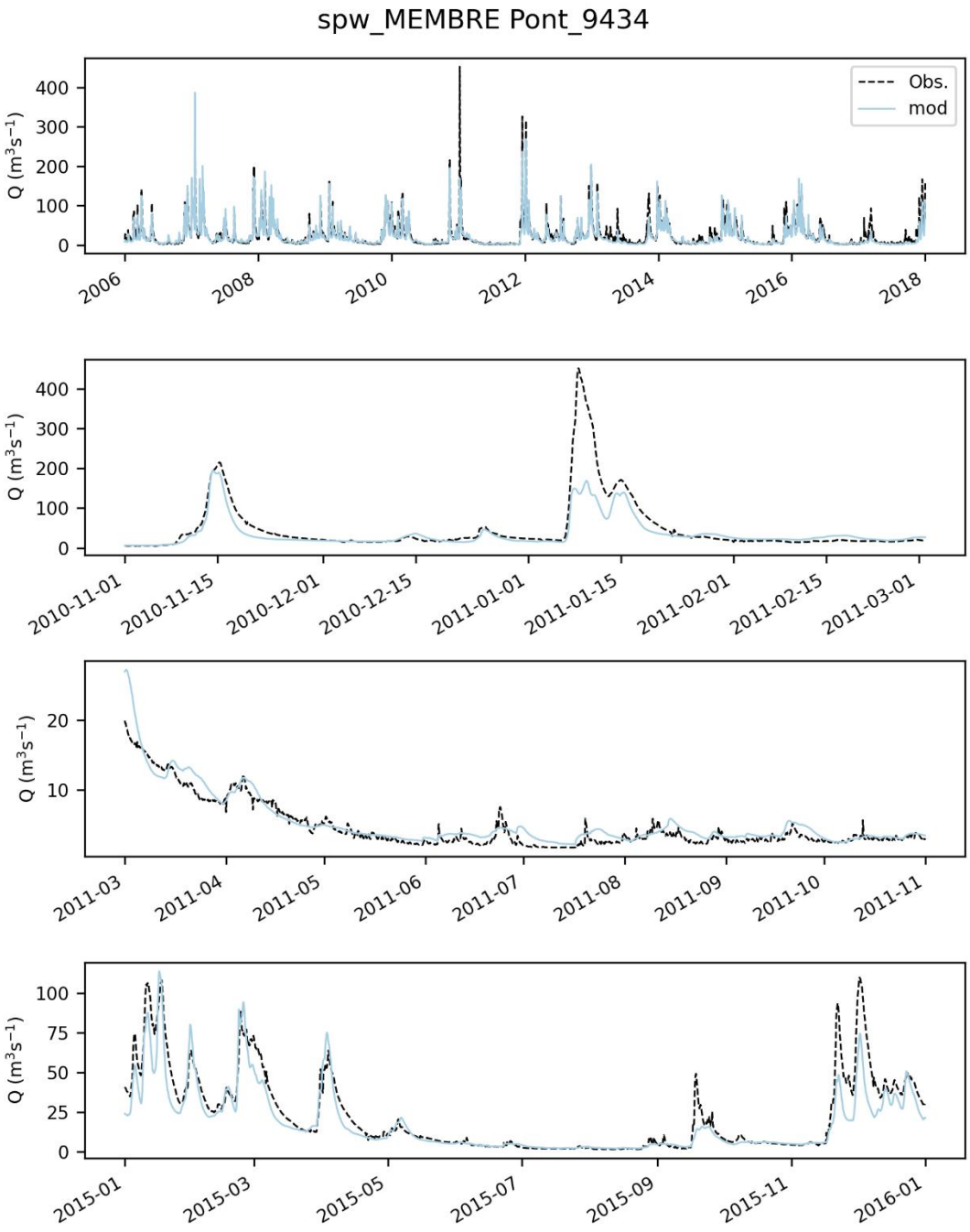


Figure B-21 Modelled (blue) and observed (black) hydrographs for the Semois at Membre-Pont

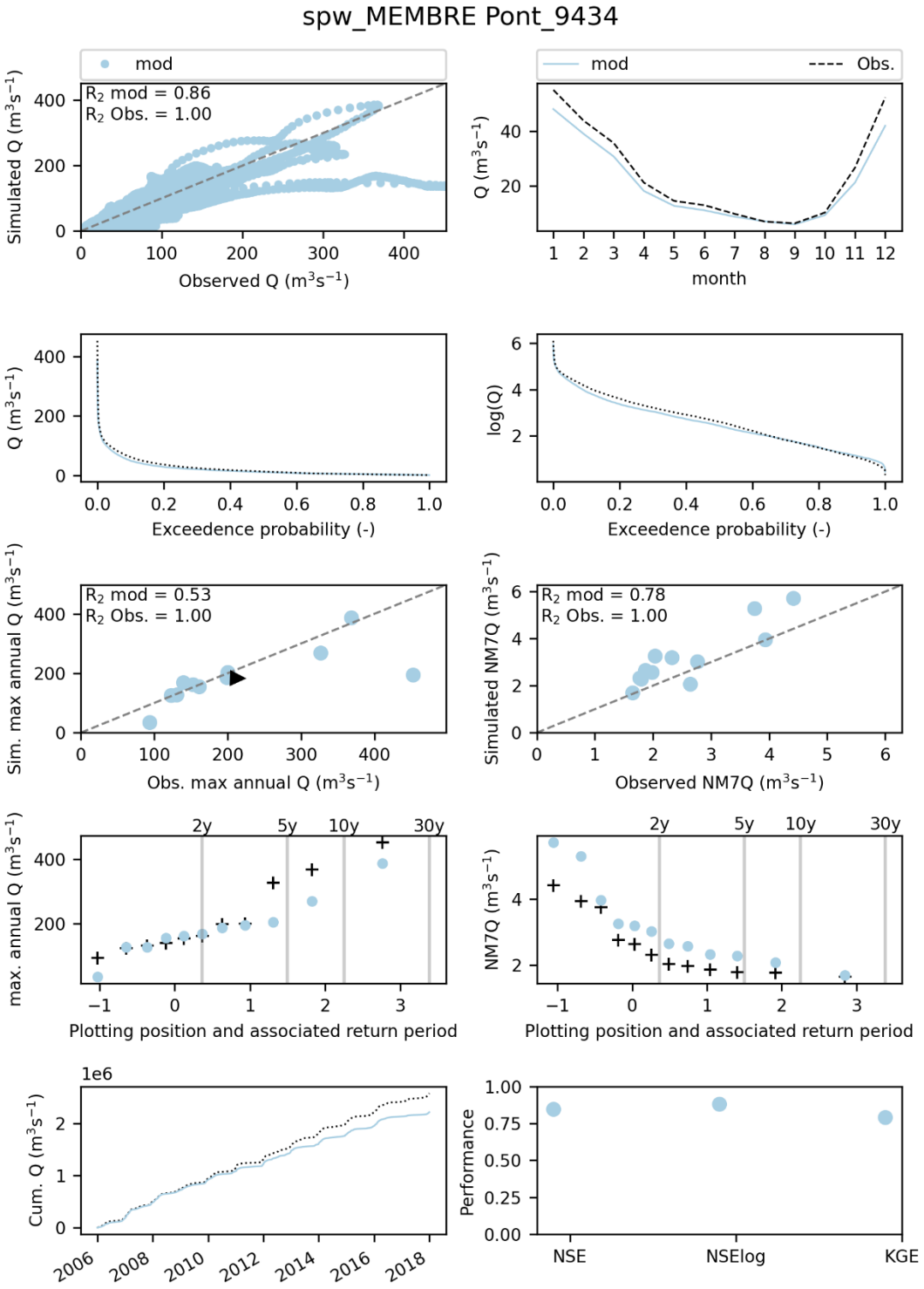


Figure B-22 Modelled (blue) and observed (black) signatures for the Semois at Membre-Pont

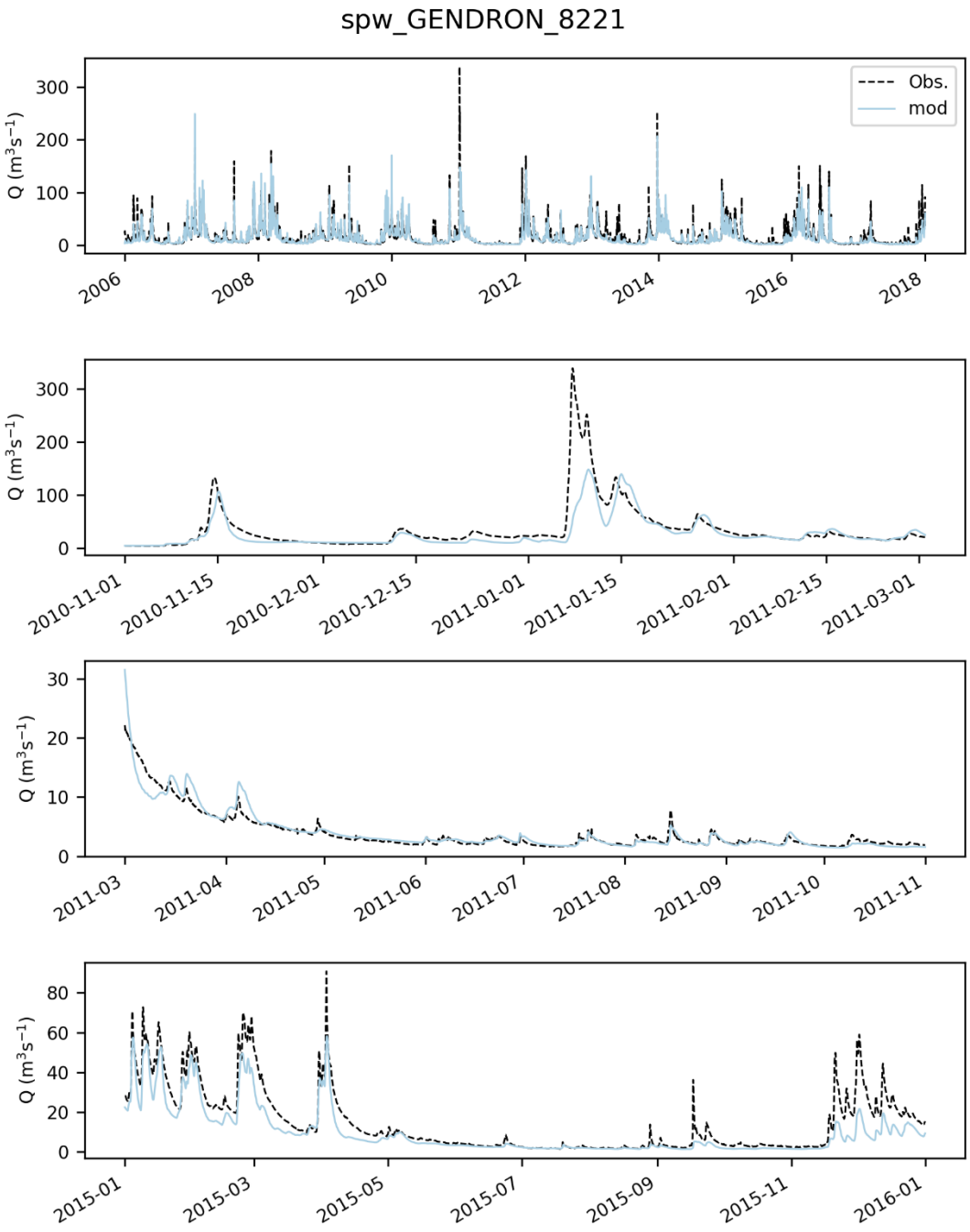


Figure B-23 Modelled (blue) and observed (black) hydrographs for the Lesse at Gendron

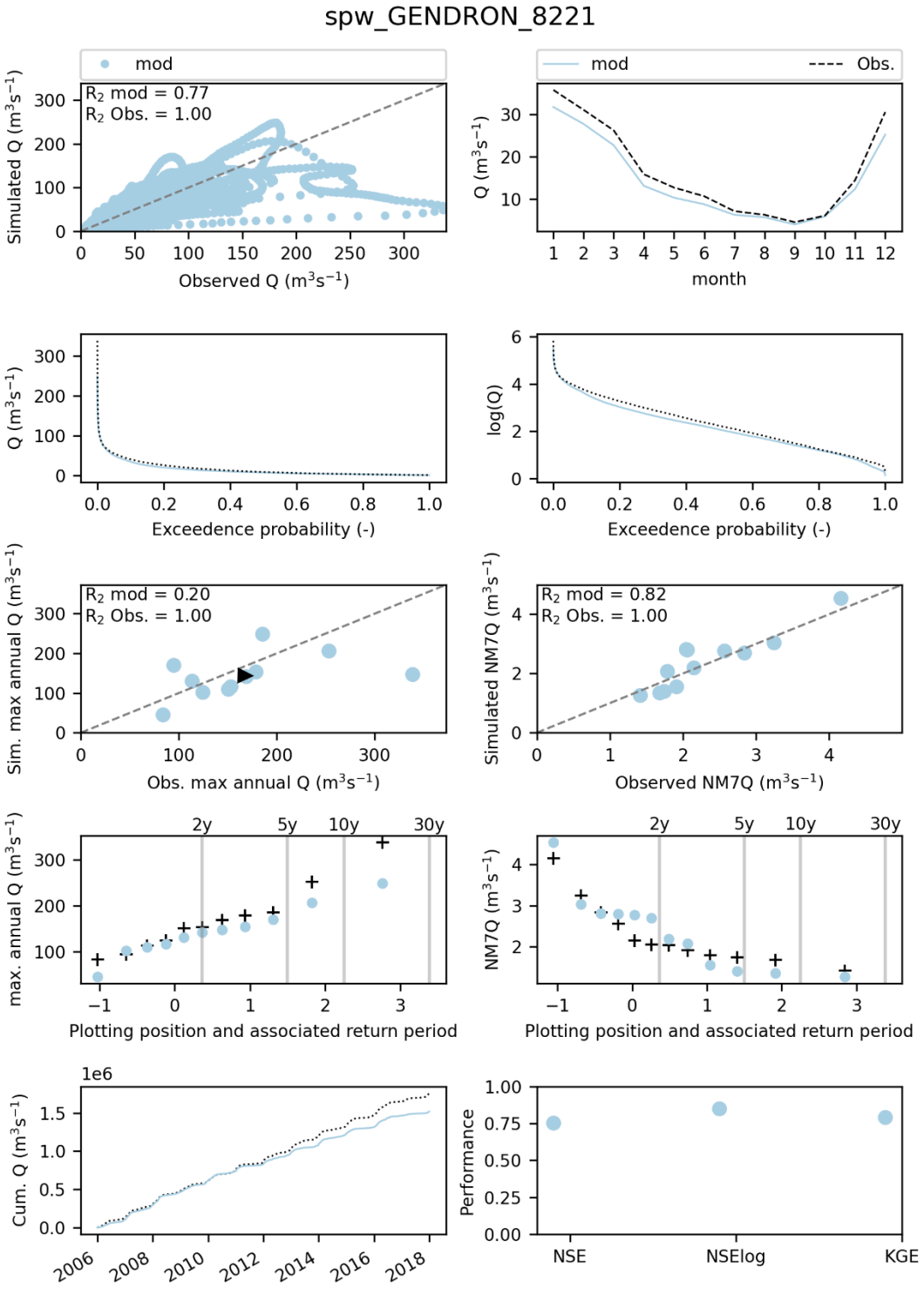


Figure B-24 Modelled (blue) and observed (black) signatures for the Lesse at Gendron

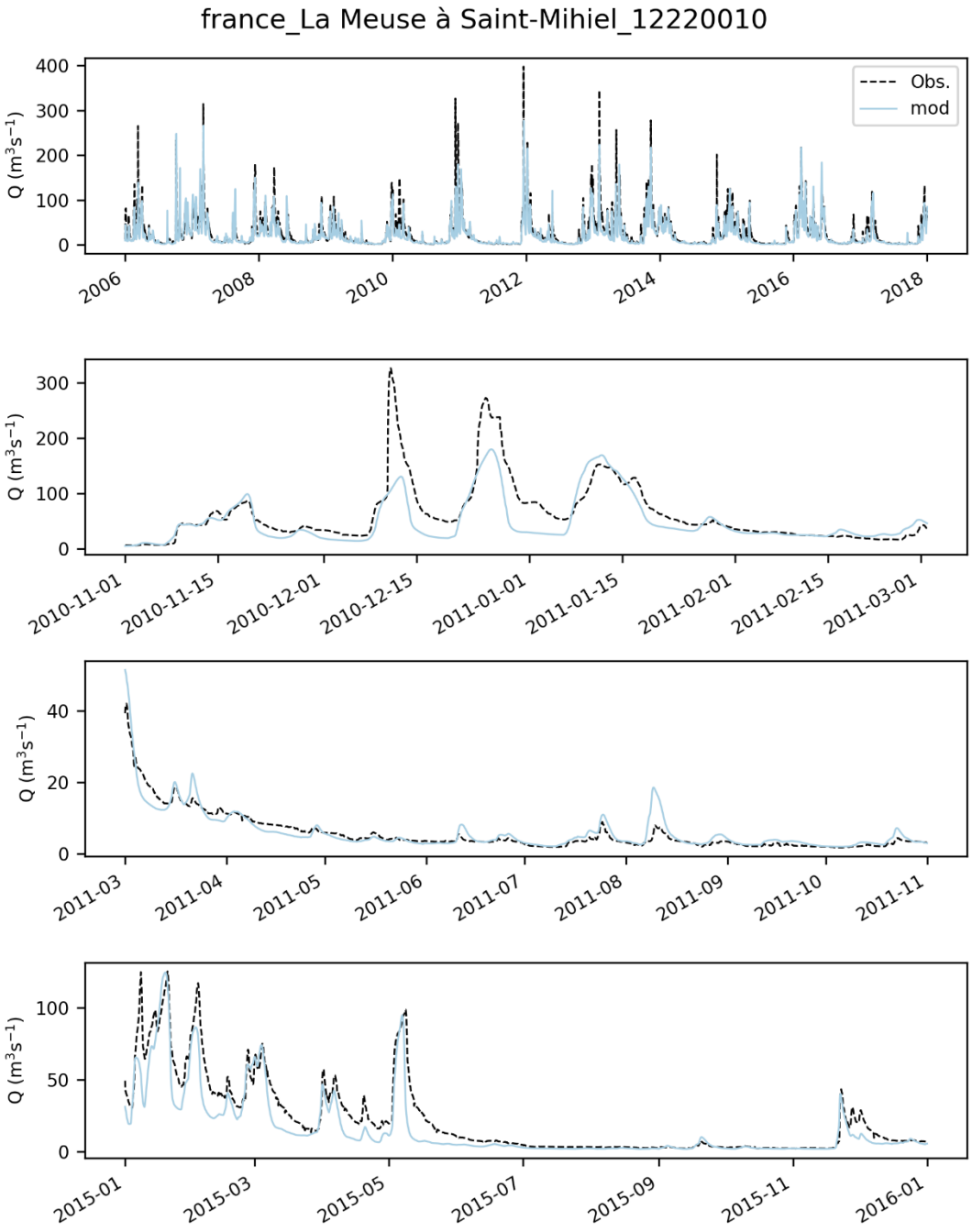


Figure B-25 Modelled (blue) and observed (black) hydrographs for the Meuse at St Mihiel

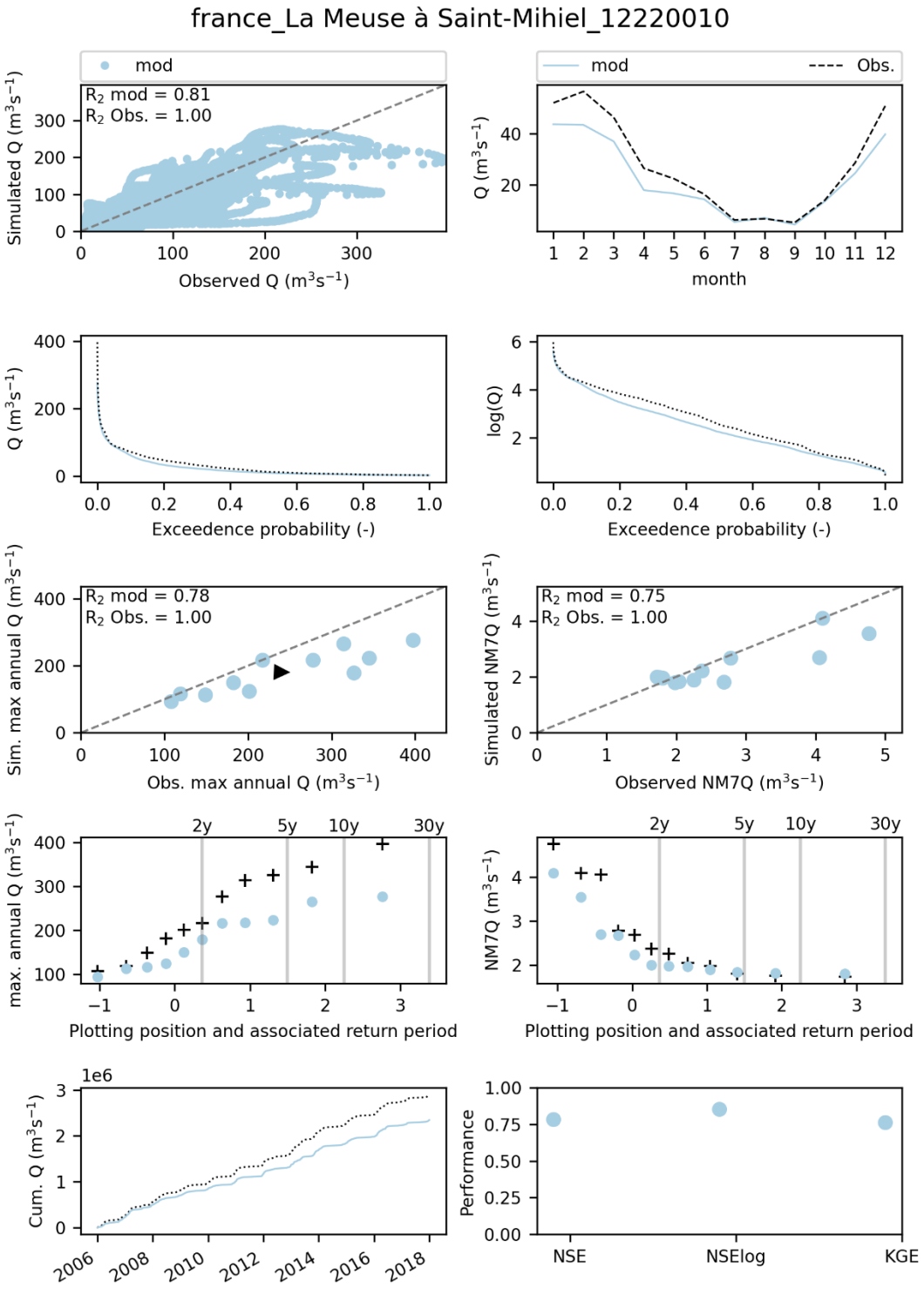


Figure B-26 Modelled (blue) and observed (black) signatures for the Meuse at St Mihiel

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.

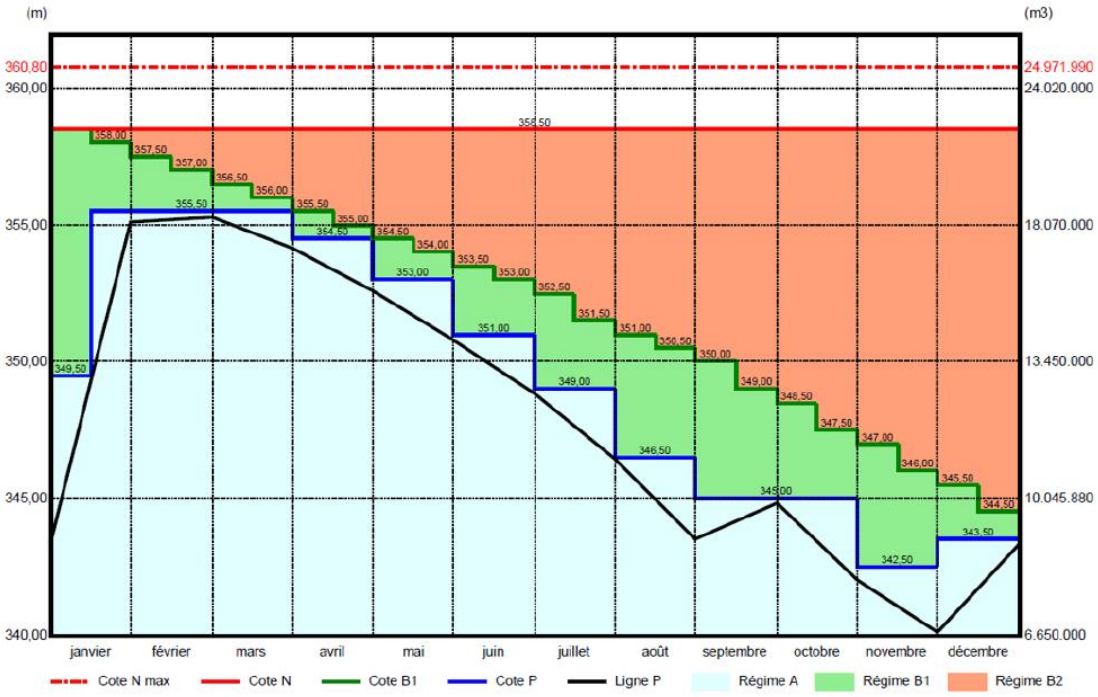


Figure C-2: Basic reservoir operation rules for the Eupen reservoir (Stucky, 2021).

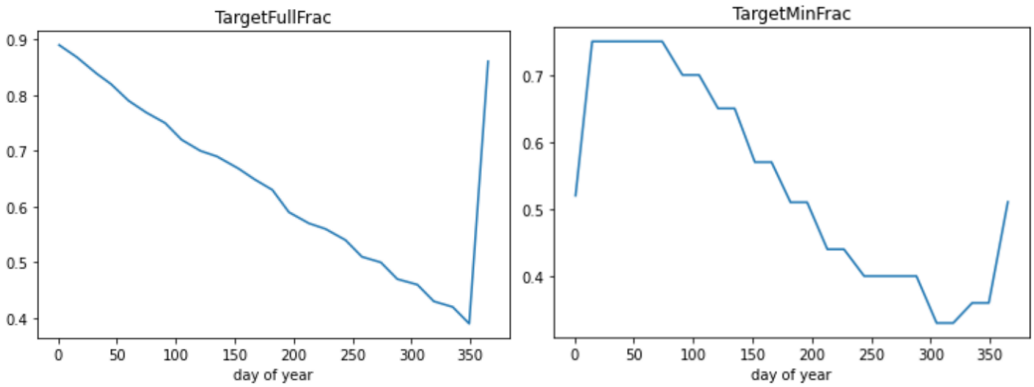


Figure C-3: TargetMinFrac and TargetFullFrac estimates for the Eupen reservoir.

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.

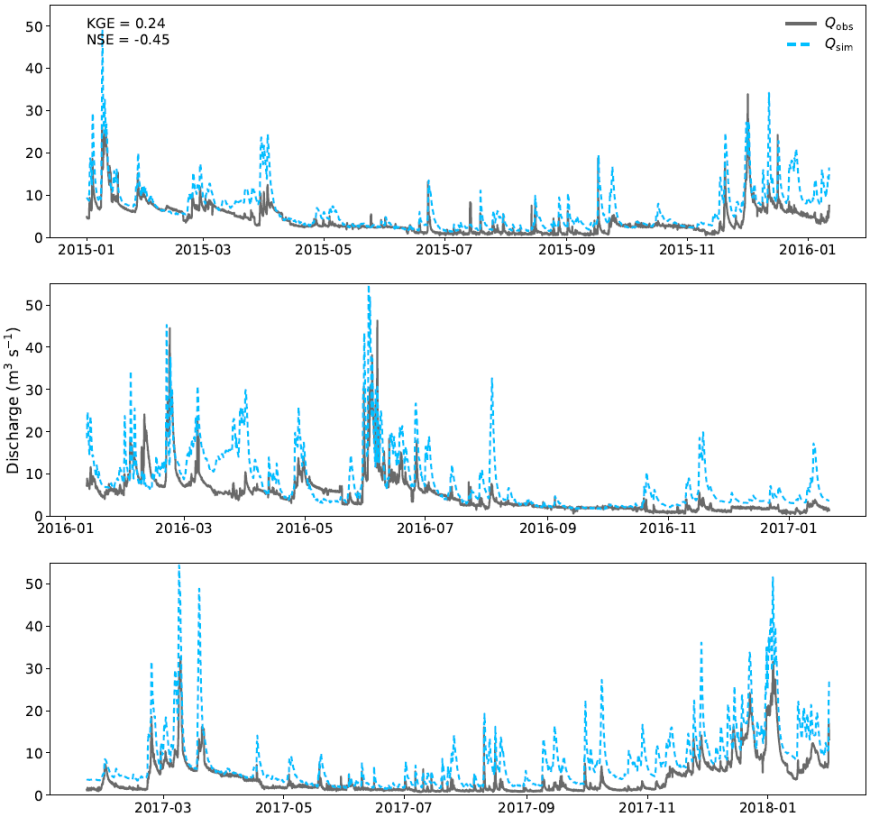


Figure C-4: Simulated versus observed discharge for the Vesdre at gauge location Verviers, which is just downstream of the two reservoirs. Simulations with the original model setup.

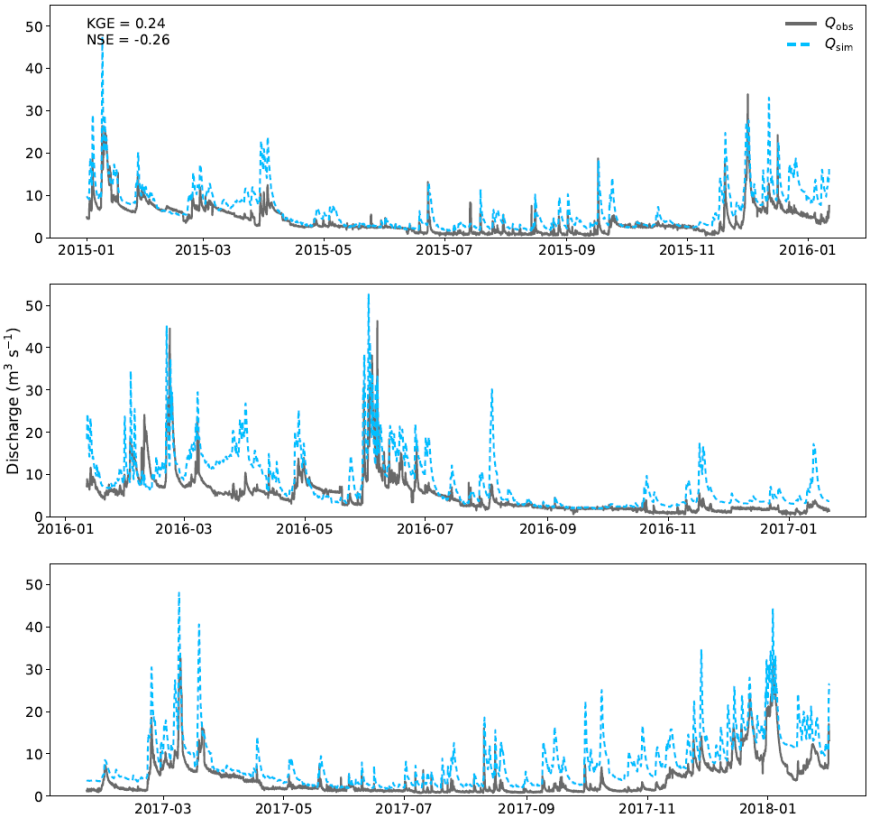


Figure C-5: Simulated versus observed discharge for the Vesdre at gauge location Verviers, which is just downstream of the two reservoirs. Simulations with the approach of step 1.

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.

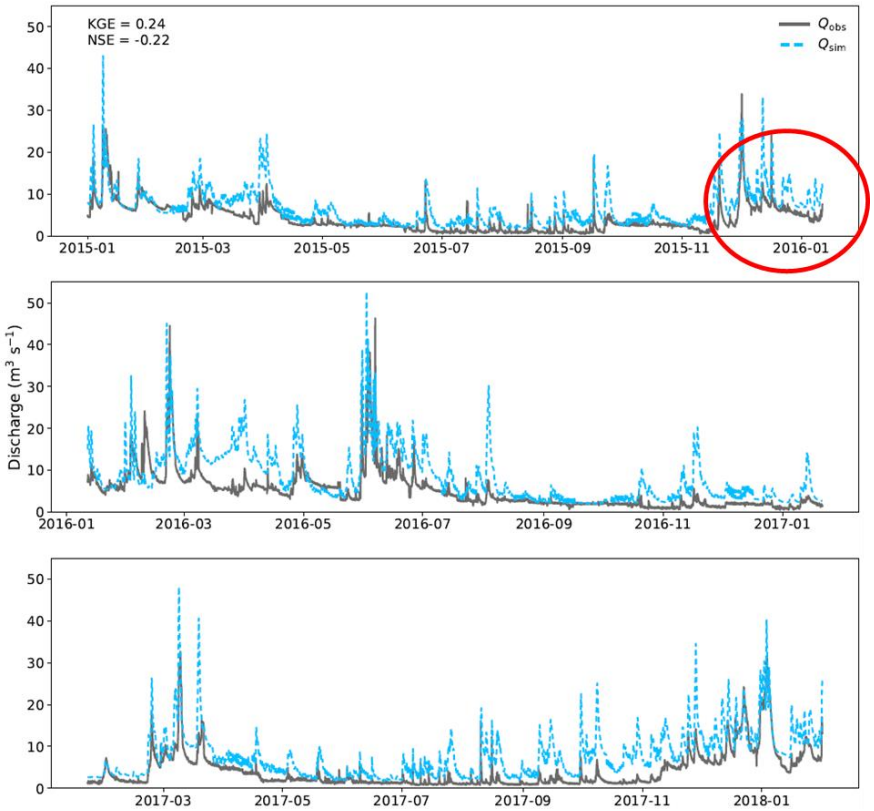


Figure C-7: Simulated versus observed discharge for the Vesdre at gauge location Verviers, which is just downstream of the two reservoirs. Simulations with the approach of step 2.

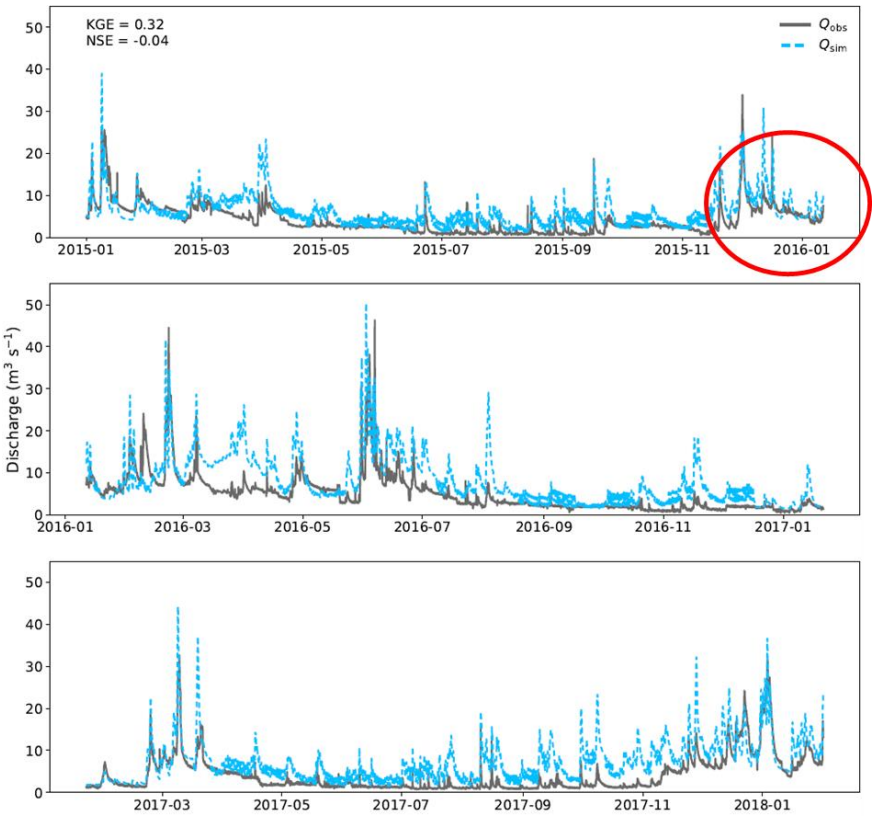


Figure C-8: Simulated versus observed discharge for the Vesdre at gauge location Verviers, which is just downstream of the two reservoirs. Simulations with the approach of step 4.

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.

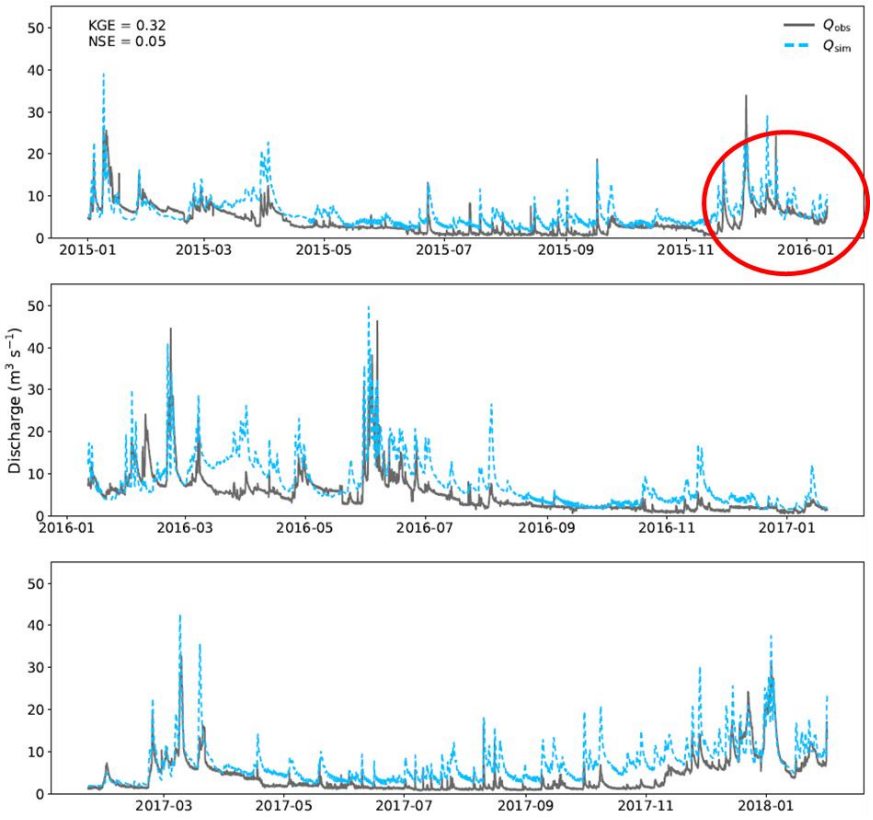


Figure C-9: Simulated versus observed discharge for the Vesdre at gauge location Verviers, which is just downstream of the two reservoirs. Simulations with the approach of step 5.

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

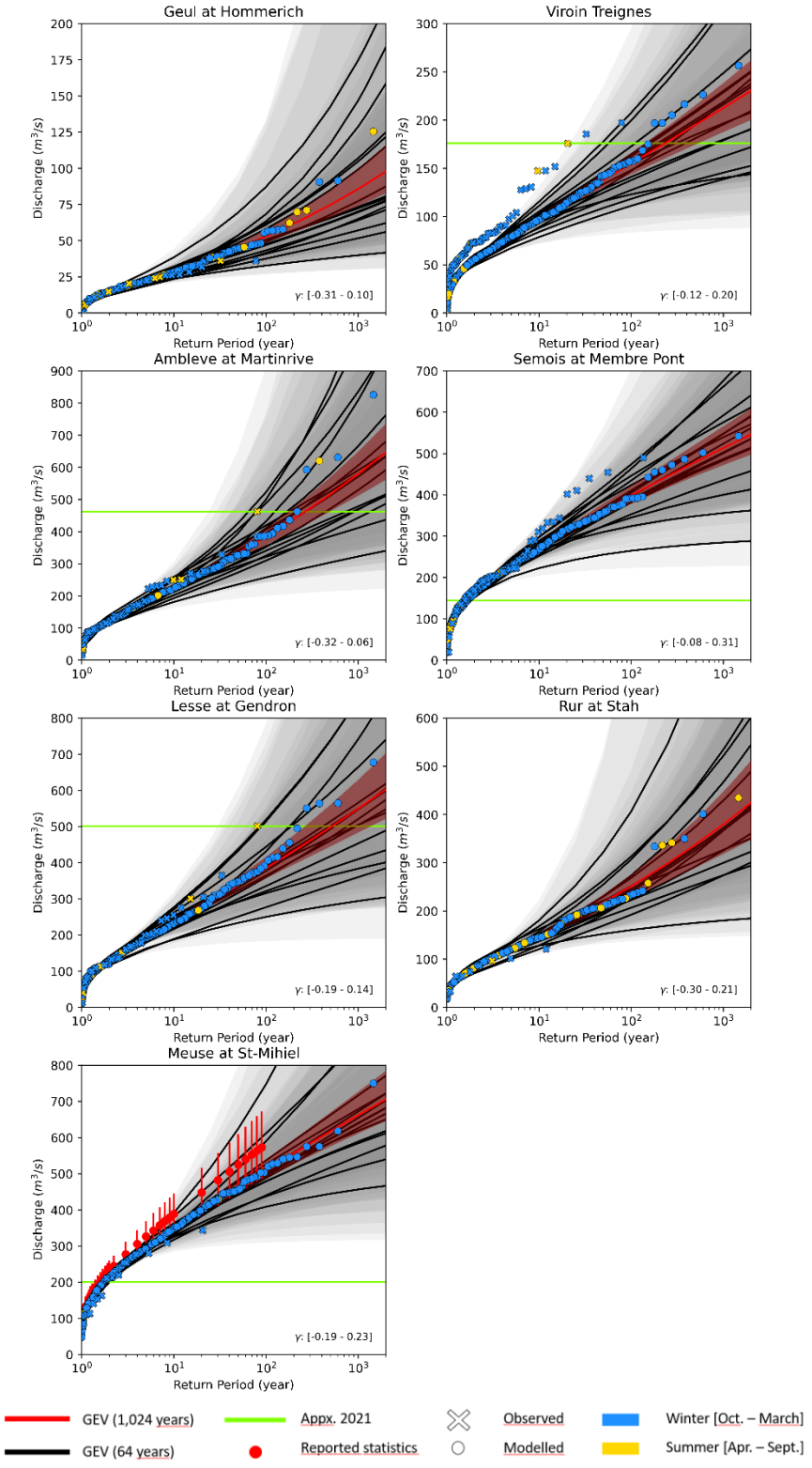


Figure D-1 Extreme value distribution fit at selected locations across the Meuse catchment. Black lines represent the GEV fit obtained from one ensemble member (64 hydrological years). Red lines show the GEV fit obtained from combining all ensembles (1,024 hydrological years, shown as full dots). The 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.

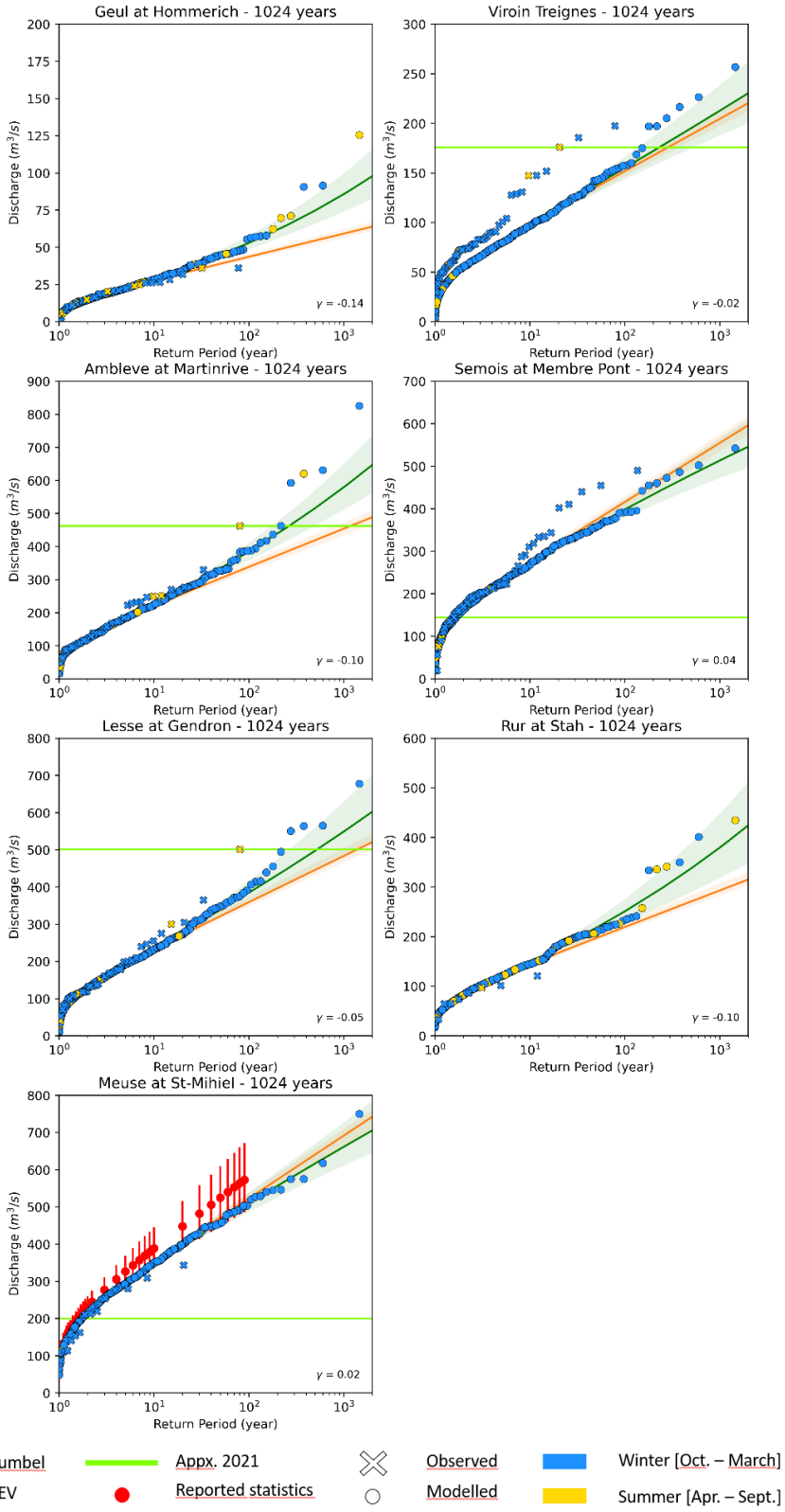


Figure D-2: Extreme value distribution fit at selected locations across the Meuse catchment. Dark green lines represent the GEV fit obtained from combining all ensembles (1,024 hydrological years, shown as full dots) and blue lines the Gumbel fit. The shading represents the statistical uncertainty from 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

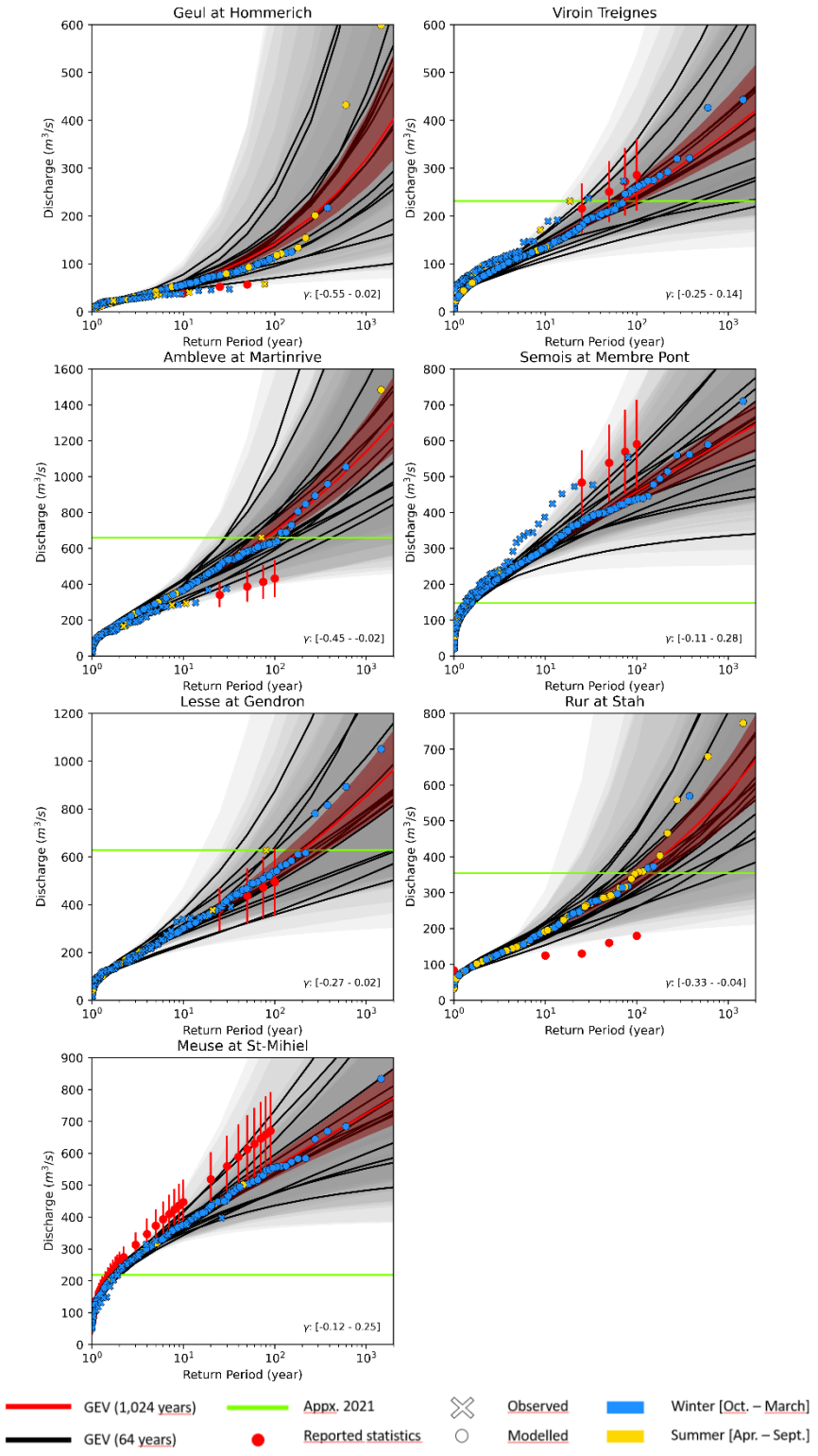


Figure D-3: Extreme value distribution fit at selected locations across the Meuse catchment. Black lines represent the GEV fit obtained from one ensemble member (64 hydrological years). Red lines show the GEV fit obtained from combining all ensembles (1,024 hydrological years, shown as full dots). The 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.

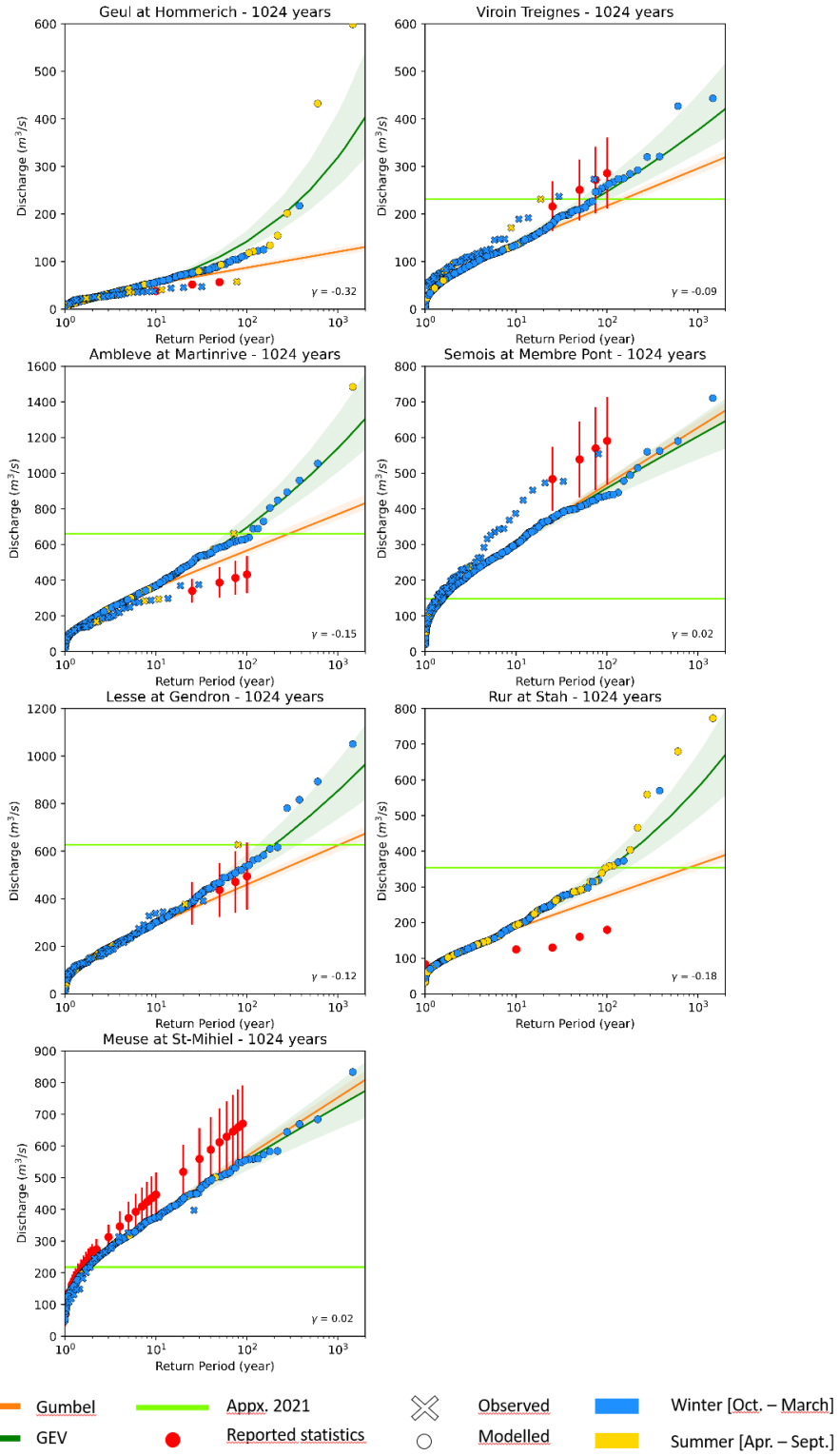


Figure D-4: Extreme value distribution fit at selected locations across the Meuse catchment. Dark green lines represent the GEV fit obtained from combining all ensembles (1,024 hydrological years, shown as full dots) and blue lines the Gumbel fit. The shading represents the statistical uncertainty from 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

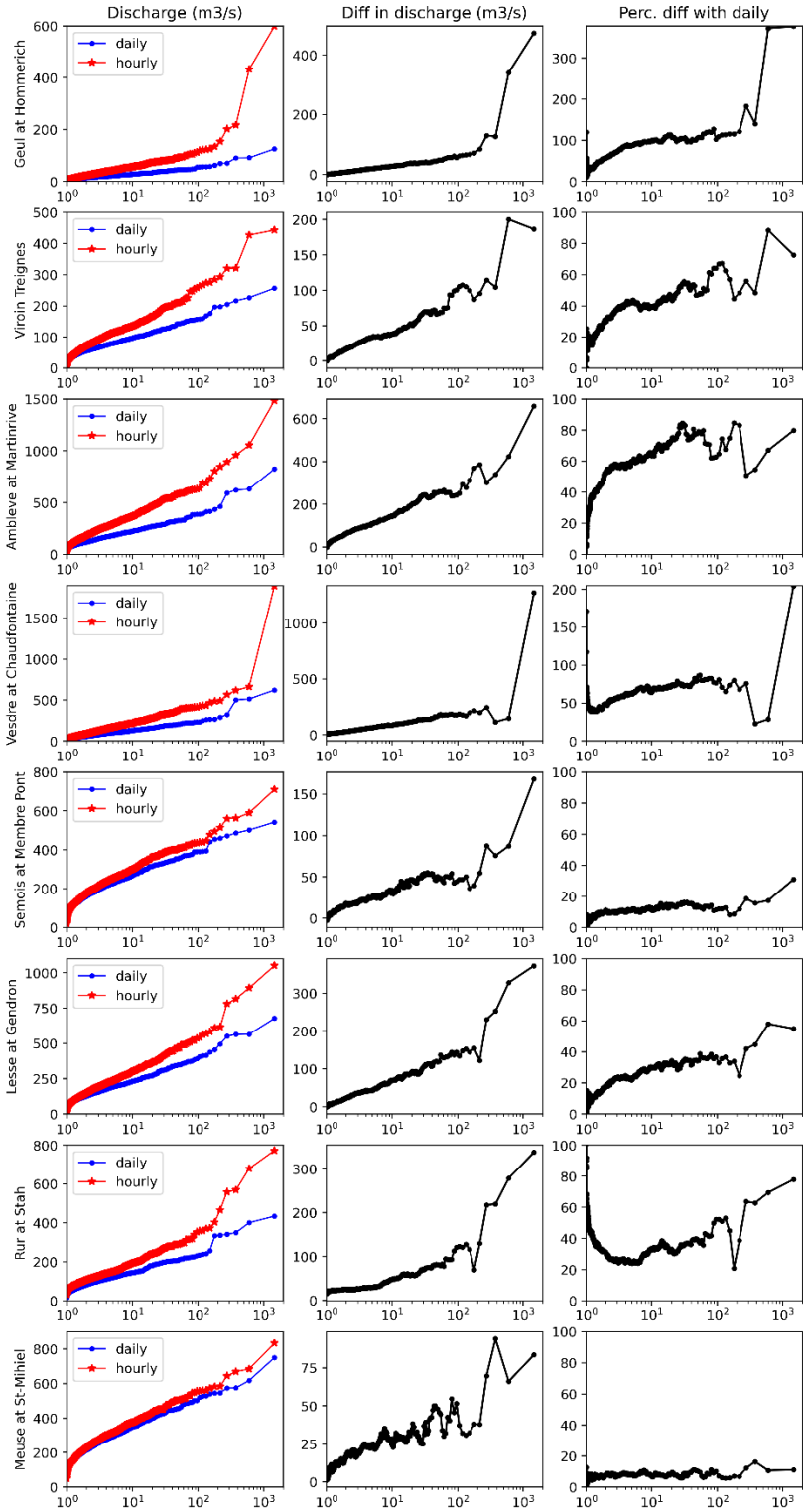


Figure D-5. Left column - Empirical return levels obtained from 1,024 hydrological years at selected locations from the daily and hourly runs. Middle column: Difference in discharge between the two temporal resolution. Right column: Relative percentage difference with respect to the empirical daily return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

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