Hydrological modelling of the 2021 mega-flood in the east of Belgium

C. Dessers[1], P. Archambeau[1], S. Erpicum[1], & B. Dewals[1], M. Pirotton[1]

[1] Research Group of Hydraulics in Environmental and Civil Engineering (HECE), University of Liège, Liège, Belgium.

Corresponding author: Christophe Dessers (C.Dessers@uliege.be)

1. Introduction

The July 2021 flood was one of the most devastating and extreme flood events ever recorded in western Europe, causing more than 220 fatalities and total damage of over 32 billion euros, making it one of the five costliest disasters in Europe (Mohr, et al., 2022). In Belgium, 39 people died during the flood, 24 of them in the Vesdre valley alone. In general, it was estimated that more than 160 buildings were swept away, more than 200 were destroyed and more than 3000 were partially destroyed.

The meteorological drivers of this extreme event were provoked by the low-pressure cyclone 'Bernd', whose propagation was slowed down by an anticyclone over eastern Europe, causing an almost stationary rain all over the impacted region (Dietze, et al., 2022; Mohr, et al., 2022; Ludwig, et al., 2023; Zeimetz, et al., 2021; Dewals, Erpicum, Pirotton, & Archambeau, 2021). As such extreme events are likely to occur again (Rajczak & Schär, 2017; Bertola, Viglione, Lun, Hall, & Blöschl, 2020; Hosseinzadehtalaei, Tabari, & Willems, 2018; Thirel, et al., 2021; Tradowsky, et al., 2023), it is of utmost importance to analyse them in depth for guiding the design and implementation of measures aimed at reducing fatalities and economic losses (de Goër de Herve & Pot, 2024).

In Germany, hydrological simulations were performed to statistically assess the frequency of the event (Ludwig, et al., 2023) and put it into the perspective of climate change. When corrections on the quantitative precipitation estimates (QPE) –which initially underestimated the rainfall quantities– were applied to the same region, the effect of the rainfall data quality as well as the impact of the model type (conceptual lumped vs gridded physically based model) were shown to play a significant role on the hydrological models' outcome and uncertainty (Saadi, et al., 2023). Vorogushyn, et al. (2024) underlined how worst this event could have been if the spatial distribution of the rain had been shifted. For Germany, analyses were conducted about river bank changes, landslides, scour, clogging effects by debris and water levels (Wolf, et al., 2024). For the Vesdre catchment in Belgium, a descriptive analysis of the surface and groundwater flow, together with environmental impacts, was detailed by Polrot (2021). A new approach to improve classification of building and content losses was proposed by Rodríguez, et al. (2025). A masterplan focused on urban planning has been prepared to guide the reconstruction of the Vesdre valley (Barcelloni-Corte, Bianchet, Privot, Schelings, & Teller, 2022).

The design of risk reduction measures should build on a sound understanding of the hydrological processes underpinning extreme events such as the 2021 flood. A fast hydrological assessment of the event was conducted by Dewals, et al. (2021) and Zeimetz, et al. (2021). They carried out a first hydrological analysis to reconstruct the peaks of the hydrographs by combining the use of surfaces ratios between close gauging stations, the Rational method, a conversion of water levels left by high-water marks to flow rates with Manning-Strickler formula and reconstruction of the outflow from the dams. In consequence of these dramatic events, a particular need and call was made by the Walloon region to investigate this event in the Vesdre Valley (Barcelloni-Corte, Bianchet, Privot, Schelings, & Teller, 2022). Goergen et al. (2025) proposed a hydrological modelling approach for the same flood event as considered here, but the description of their methodology fails to highlight how hydraulic structures such as large reservoirs were incorporated in the modelling, while their effect is known to be significant. In this context, the present study provides the first detailed hydrological analysis of the most impacted Belgian catchments during the July 2021 floods.

Here, we present a framework as well as a new gridded and computationally efficient hydrological model–with increasing complexity– to allow an in-depth hydrological analysis of the two most severely impacted catchments in Belgium during the July 2021 floods, namely the Vesdre and the Amblève catchments. In the former, most of the gauging stations were damaged, washed away or destroyed during the event, making it impossible to retrieve time series of observations covering the

entire flood wave and particularly the flood peak. In the latter, almost all gauging stations were operational during the event.

Therefore, two different approaches were adopted. For the Vesdre catchment, a computationally efficient, gridded, runoff model allowed the hindcasting of the event and evaluation of alternate scenarios. For the Amblève catchment, an improved version of the gridded model was compared to two conceptual models: GR4H (Perrin, Michel, & Andréassian, 2003) and VHM (Willems, 2014). The models were implemented in a modular environment, WOLFHydro, developed at the University of Liège. The influence of the product used for rainfall data was investigated in both catchments. The transferability of the runoff model from one catchment to another (from the Vesdre to the Amblève) was also tested, as well as the impact of calibrating on different events (from recent historical floods to July 2021 and reciprocally).

The modelling work presented here contributes to addressing the following research questions:

1. In the Vesdre catchment, which were the river discharges during the July 2021 flood, and is it possible to reconstruct the event's hydrographs based on the few available data?
2. For such extreme events, can a runoff model calibrated on one catchment (here the Vesdre) be directly transferred to a neighbour's catchment (Amblève)?
3. Does a semi-distributed approach calibrated by considering observations at several intermediate gauging stations reproduce well July 2021 event (Amblève catchment)? At the same time, is there a difference between gridded and lumped models?
4. Can such a model be directly applied to previous historical floods? The other way round, can a model trained on historical floods predict what happened during the extreme event of July 2021?

A description of the two considered catchments is given in Section 2. Section 3 introduces the datasets available for each catchment, the respective modelling approaches, the simulation environment WOLFHydro, the analysed scenarios and the optimisation procedure. The results are presented and discussed in Section 4.

2. Case study

2.1. Vesdre catchment

The Vesdre river has its springs in the High Fens, flows along a narrow and deep valley until discharging into the Ourthe river, which is one of the main tributaries of the Meuse river. The main tributaries of the Vesdre river, from upstream to downstream, are the Helle, the Wayai and the Hoegne rivers. As shown in Figure 1, two dams with approximately the same storage capacity (25 hm^3) are present next to the High Fens, in the upper part of the catchment: the Vesdre and the Gileppe dams (Bruwier, Erpicum, Pirotton, Archambeau, & Dewals, 2015). Tunnels deviate a part of the Helle and the Soor rivers to supply respectively the Vesdre and the Gileppe dams (Bruwier, Erpicum, Pirotton, Archambeau, & Dewals, 2015). Both tunnels are usually open and collect most of the time a maximum flux of 20 m^3/s. The main characteristics of the catchment are summarised in Table 1. The High Fens are considered as landuse of type 'water'. A more comprehensive description of the catchment is provided by Barcelloni-Corte, et al. (2022).

The Vesdre was poorly monitored during the 2021 event due to the extreme nature of the event and most of the sensors were washed away or damaged, creating a need to reconstruct the flow to fill this lack of data.

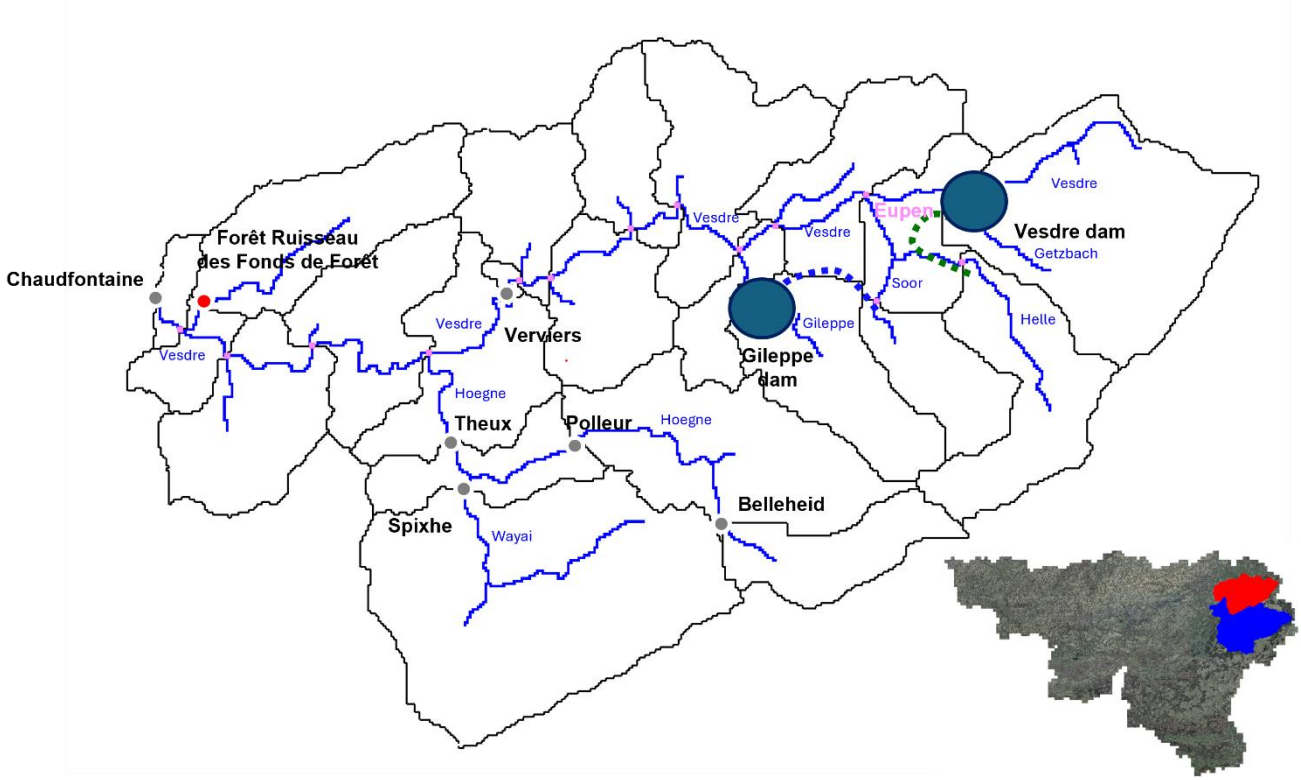


Figure 1 Subdivision of the Vesdre catchment in sub-catchments. The dams are represented as blue points, the red dots represent the gauging stations, while the grey points are all stations available but not used in the current study. The blue and the green dashed lines are respectively the deviation of the Soor and the deviation of the Helle rivers. On the smaller map of Wallonia, the Vesdre catchment is depicted in red.

2.2. Amblève catchment

The Amblève is also a tributary of the Outhe river, located in the upland Ardennes, and has a drained surface of about 1100 km² (Table 1). Its main tributaries, from upstream to downstream, are the Warche, Salm, Roannay and Lienne rivers.

Two dams are situated in series on the Warche river, upstream of Malmedy (Figure 2). The first one (Bütgenbach reservoir, 11 hm³) releases water into the reservoir of Robertville (8 hm³) located more than 5 km downstream. A pumped-storage facility is located at Coo, downstream the town Trois-Ponts (at the confluence of the Salm and the Amblève rivers); but as it operates in a closed loop system, it has a negligible effect on the discharge of the river. A water derivation for a run-of-river hydropower plant (Lorcé, located downstream the confluence with the Lienne river) was neglected in our modelling. It is estimated that the maximum discharge abstracted at the intake is no more than 25 to 30 m³/s (Benitez, Dierckx, Matondo, & Ovidio, 2016). On the Salm river, a small lake (Doyard lake, near Vielsalm) is also neglected in the present work.

The Amblève catchment was well documented during the 2021 event. It contains a dozen of stations well scattered all around the catchment, which were almost all functional during the event, offering a reference to understand the different hydrological mechanisms at the headwater parts or downstream near the outlet of the catchment. More details about the available observations are provided in Section 3.2.

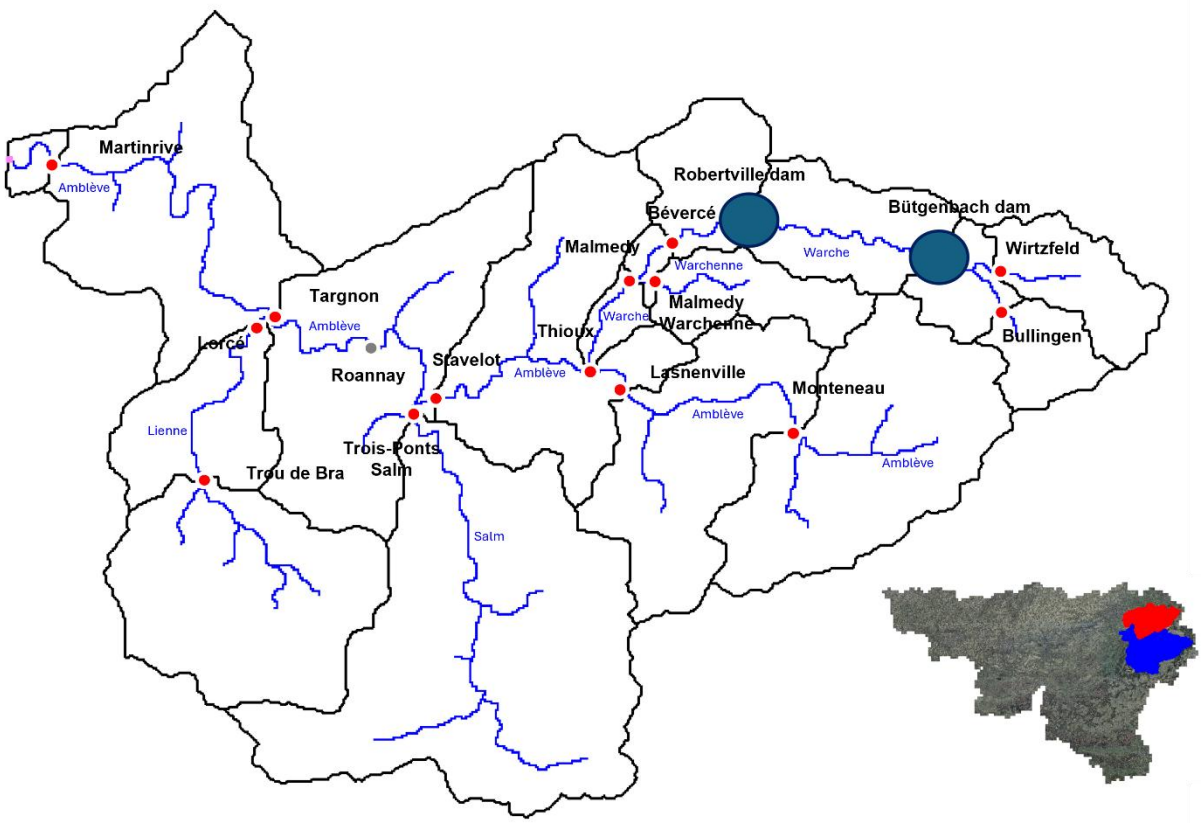


Figure 2 Subdivision of the Amblève catchments with all the gauging stations. The dams are represented as blue points, the red dots represent the gauging stations used for calibration and the grey one, a station not considered in the current study (due to its poor reliability). On the smaller map of Wallonia, the Amblève catchment is depicted in blue.

Catchment		*Vesdre*	*Amblève*
Surface [km²]		685	1072
Perimeter [km]		170	236
Average slope [%]		7.5	8.9
Gravelius coefficient [-]		1.83	2.03
Land use [%]	Forest	41	43
	Meadows	35	43
	Agriculture	3	4
	Urbans	15	9
	Rivers	1	1
	Water	5	0

Table 1 Main characteristics of the Vesdre and Amblève catchments.

3. Data and methods

The following paragraphs detail the data available and the different steps of the modelling strategy, including the selected models and the optimisation procedures, as also summarised in Table 2.

133

Catchment	Strategy	Calibration	Evaluation
Vesdre	Hindcast	Type: A1 Rain: Radclim V1 Simulation: [13-07-2021;17-07-2021] Calibration: [13-07-2021;17-07-2021] Models: Runoff	Rain: - Radclim V1 - Radclim V2 - Radflood21 - RadQPE Simulation: [13-07-2021;17-07-2021] Evaluation: [13-07-2021;17-07-2021] Models: Runoff
Amblève	Vesdre →Amblève	X	Rain: - Radclim V1 - Radclim V2 - Radflood21 - RadQPE Simulation: [13-07-2021;17-07-2021] Evaluation: [13-07-2021;17-07-2021] Models: Runoff
	Event	Type: B1 Rain: Radflood21 Simulation: [12-07-2021;17-07-2021] Calibration: [13-07-2021;15-07-2021] Models: VHM, GR4H, 2 layers	Rain: Radflood21 Simulation: [12-07-2021;17-07-2021] Evaluation: [13-07-2021;15-07-2021] Models: VHM, GR4H, 2 layers
	2021 → Historical	Type: B1 Rain: Radclim V2 Simulation: [01-07-2021; 31-07-2021]	Rain: SPW Simulation: - [14-12-1991; 01-02-1991] - [29-11-1993; 25-01-1994] - [25-12-1994; 07-02-1995] - [03-09-1998; 27-09-1998] - [16-01-2002; 06-03-2002]

Catchment	Strategy	Calibration	Evaluation
			- [01-01-2011; 23-01-2011]
		Calibration: [13-07-2021; 28-07-2021]	Evaluation: - [16-12-1991; 29-12-1991] - [06-12-1993; 10-01-1994] - [16-01-1995; 07-02-1995] - [12-09-1998; 22-09-1998] - [20-01-2002; 06-02-2002] - [10-02-2002; 04-03-2002] - [05-01-2011; 18-01-2011]
		Models: VHM, GR4H, 2 layers	Models: VHM, GR4H, 2 layers
	Historical → 2021	Type: B2 Rain: SPW Simulation: - [14-12-1991; 01-02-1991] - [29-11-1993; 25-01-1994] - [25-12-1994; 07-02-1995] - [03-09-1998; 27-09-1998] - [16-01-2002; 06-03-2002] - [01-01-2011; 23-01-2011] Calibration: - [16-12-1991; 29-12-1991] - [06-12-1993; 10-01-1994] - [16-01-1995; 07-02-1995] - [12-09-1998; 22-09-1998] - [20-01-2002; 06-02-2002] - [10-02-2002; 04-03-2002] - [05-01-2011; 18-01-2011] Models: VHM, GR4H, 2 layers	Rain: Radflood21 Simulation: [12-07-2021;17-07-2021] Evaluation: [13-07-2021;15-07-2021] Models: VHM, GR4H, 2 layers

Table 2 Summary of the considered scenarios, together with the selected strategies for model calibration and evaluation. The types of rainfall input data are presented in Section 3.1, the models are introduced in Section 3.4, and the calibration methods are detailed in Section 3.6.

3.1. Input data

The rainfall data applied in the present work come from two sources: the RMI (Royal Meteorological Institute of Belgium, 2022) and the Service Public de Wallonie (SPW) (2025). The former provides rainfall estimates distributed on a grid with a spatial resolution of 1 km. Several versions of the product are available, incorporating various corrections. As the simulation grid is finer and not aligned with the rain grid, the latter is aggregated proportionally to its surface to fit in the former. Data provided by SPW are based on measurements at rain gauges, some of them being operational since 1982. This local information was spatially distributed using the Voronoï algorithm (Thiessen polygons) (Voronoi, 1908), and by updating the polygons depending on the period of functionality of each station.

The rainfall data products considered in the modelling are the following:

- Radflood21 (version 27/06/2024): a RMI post-event radar-based quantitative precipitation estimate made especially for the July 2021 floods (Journée, Goudenhoofdt, Vannitsem, & Delobbe, 2023). It provides time series of gridded data, with a 5-minute timestep from the 11 to 18 July 2021.
- Radclim (version 2021): a RMI post-event radar-based product containing a previous version of the rainfall estimation. It provides time series of gridded data, with a 10-minute timestep from the 11 to 18 July 2021. In the following, this dataset is referred to as 'Radclim V1'.
- Radclim (version 27/06/2024): a RMI post-event radar-based product providing a 5 minutes timestep gridded rainfall data for an extended period of time: from the 1 June 2021 until 31 July 2021. This dataset is called hereafter 'Radclim V2'.
- RadQPE: a RMI real-time and radar-based generated quantitative precipitation estimate (Goudenhoofdt & Delobbe, 2012). It provides time series of gridded data with a 5-minute timestep.
- SPW rain stations: source point measurements collected at the rain gauging stations. They provide time series with a time step of 1 hour (also available with a 5-minute time step but not corrected). The operationality and the number of stations vary with time. However, at least one station has always provided a measurement since 1982.

RMI data were used for all the simulations of the 2021 flood, while SPW data were considered for previous floods. The product RADQPE does not cover the whole period of interest for historical floods (between 1991 and 2011). That is the reason why SPW data were preferred for previous floods.

For the 2021 event, the cumulative rainfall volumes derived from the different datasets are presented in Figure 3. Among these, RADQPE consistently reports substantially lower rainfall amounts than the other products. Peak rainfall values in Radclim V1 are very close to those in Radflood21 and Radclim V2, although the spatial distribution differs considerably. The most pronounced divergence occurs in the downstream part of the Vesdre catchment, where Radflood21 and Radclim V2 estimate lower rainfall amounts than Radclim V1. A smaller but notable difference appears in part of the Amblève catchment, where the highest rainfall zone extends in Radflood21 and Radclim V2 but not in Radclim V1.

For the 2021 event, the cumulated volume according to the various rainfall datasets is shown in Figure 3. Rain quantities are much lower in RADQPE than in the other products. Peak values in Radclim V1 are very close to the peaks in Radflood21 and Radclim V2, but a considerable difference in the spatial distribution can be observed. The most noticeable difference is situated in the downstream part of the Vesdre catchment, where the rain quantities are lower according to Radflood21 and Radclim V2, as

well as in a small part of the Amblève catchment where the region of highest rainfall quantities extends in the datasets Radflood21 and Radclim V2, but not in the dataset Radclim V1.

In Radflood21, the maximum cumulated volume in a cell (between the 12th of July until the 17th of July included) reaches 293 mm with a spatial average of 175 mm for the Vesdre catchment. For the Amblève catchment, these values are respectively 210 mm and 120 mm. In Radclim V1, the local peak of rainfall volume is 2% lower in the Vesdre catchment and 9% lower in the Amblève catchment when compared to Radflood21. For Radclim V2, the same differences are limited to 0.6% higher and 0.5% lower , respectively, indicating only minor deviations. In contrast, RadQPE shows substantial differences compared to Radflood21, with a local peak 39% lower in the Vesdre catchment and 28% lower in the Amblève catchment. When compared to Radflood21, the spatial average of rainfall quantities in the Vesdre and Amblève catchments are 5% higher and 2.5% lower for Radclim V1, 0.5% higher and 2.5% higher for Radclim V2, and 24% lower and 11% lower for RadQPE.

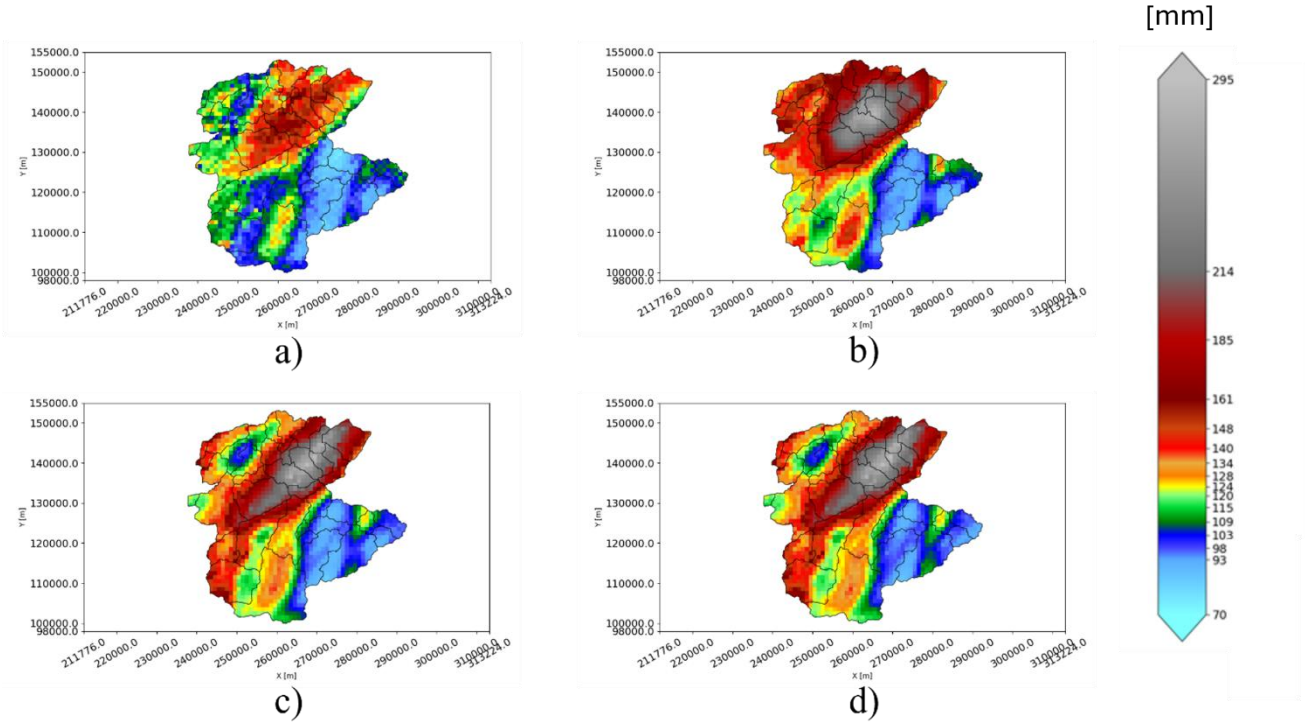


Figure 3 Cumulated rain *in [mm]* from 12/07/2021 to 17/07/2021 on the Vesdre and Amblève catchments. The black contours delineate the sub-catchments: (a) RadQPE, (b) Radclim V1, (c) Radflood21, (d) Radclim V2

The rainfall measurements at the rain gauging stations display the characteristics of a 2-day rain. Indeed, most of the rainfall quantities were measured within 2 days between the 13/07 and 16/07, depending on the sub-catchment. When compared with statistical rains at some stations, the quantities observed were almost twice greater than a 200-year return period rain. At Jalhay and Ternell, it fell in 2 days as much as a 10-day volume of a return period of 200 years. The rain amounts in the Amblève catchment were lower, but the latest version of the radar-based model demonstrates the complexity of the event and the importance of taking into account its spatial distribution.

As only major flood events are analysed in this study, evaporation and evapotranspiration were neglected. Due to the summer period, a snow model is out of scope for July 2021 and not considered for winter historical floods. To take into account in the infiltration models the spatial variability of the soil, a vectorial land use map was extracted from “Occupation et utilisation du sol en Wallonie –

COSW 2007" (available on WalOnMap (Géoportail de la Wallonie)) and adapted according to Table 5. The catchment was delimited thanks to a digital terrain model (DTM) also provided on WalOnMap.

3.2. River flow 'measurements'

While almost no gauging station in the Vesdre catchment captured the peak discharge (Table 3), the event in the Amblève catchment was well recorded at almost all the gauging stations (Table 4). The gauging stations are operated by two different departments of the river authority SPW (Hydrométrie, 2025) and they have a different sampling frequency. The stations with an ID starting with an "L" provide data with a timestep of 10 minutes, while the others use a 5-minute time step.

ID	Name	Time step	Quality
L7150	Verviers	10 min	Low
6526	Belleheid	5 min	Incomplete
6517	Polleur	5 min	Incomplete
L6790	Spixhe	10 min	Incomplete
	Theux		Incomplete
L7600	Ruisseau des Fonds de Forêt	10 min	Complete
6228	Chaudfontaine	5 min	Incomplete
-	Vesdre dam	1 min	Complete
-	Gileppe dam	5 min	Complete

Table 3 List of the stations and measurements considered for the Vesdre catchment and some of their characteristics

For the Vesdre catchment, the only complete and reliable time series are at 'Ruisseau des Fonds de Forêt' (L7600), which is on a small tributary of the Vesdre, and the water levels in the Gileppe and Vesdre reservoirs (Table 3).

Some additional information about the digitalisation of the water height was provided by the SPW, and these results can be compared with a flood modelling of the event (WalOnMap)

ID	Name	Time step	Quality
6981	Bullingen	5 min	Complete
6971	Wirtzfeld	5 min	Complete
6991	Malmedy-Warchenne	5 min	Complete

6933	Malmedy	5 min	Complete
L5580	Thioux	10 min	Complete
L7280	Monteneau	10 min	Complete
6753	Lasnenville	5 min	Complete
L6070	Trois-Ponts	10 min	Complete
6732	Stavelot	5 min	Complete
L7290	Trou-de-Bra	10 min	Complete
L6240	Lorcé	10 min	Complete
6671	Targnon	5 min	Complete
6621	Martinrive	5 min	Complete
6946	Bévercé	5 min	Complete

Table 4 List of the stations and measurements considered for the Amblève catchment and some of their characteristics

3.3. Influence of dams and reservoirs

For the Vesdre dam, the availability of sluice gates operations and water level measurements allowed the outlet and inlet discharges of the reservoir to be evaluated externally. The corresponding outlet hydrograph was prescribed as a boundary condition in the modelling and was also used for models' calibration. Although some tributary stations exist in the dams' reservoirs, their data was either unavailable or of poor quality for the extreme flood of 2021 and would require further investigation.

3.4. Numerical models

3.4.1. Hydrological models

Event-based hydrological simulations were carried out. Given the scarcity of observations in the Vesdre catchment, a simple runoff model (as detailed in Appendix B) was used to hindcast the river hydrographs, which were not measured. To investigate the main driving factors of extreme floods and to assess the predictability of such events based on knowledge of the past, additional lumped models were also used in the Amblève catchment, for the sake of comparison and validation: VHM (Willems, 2014) and GR4H (Perrin, Michel, & Andréassian, 2003). These two models were implemented in WOLFHydro following the structure described in Figure 4. The flux fractions in the VHM model were determined using their exponential formulation.

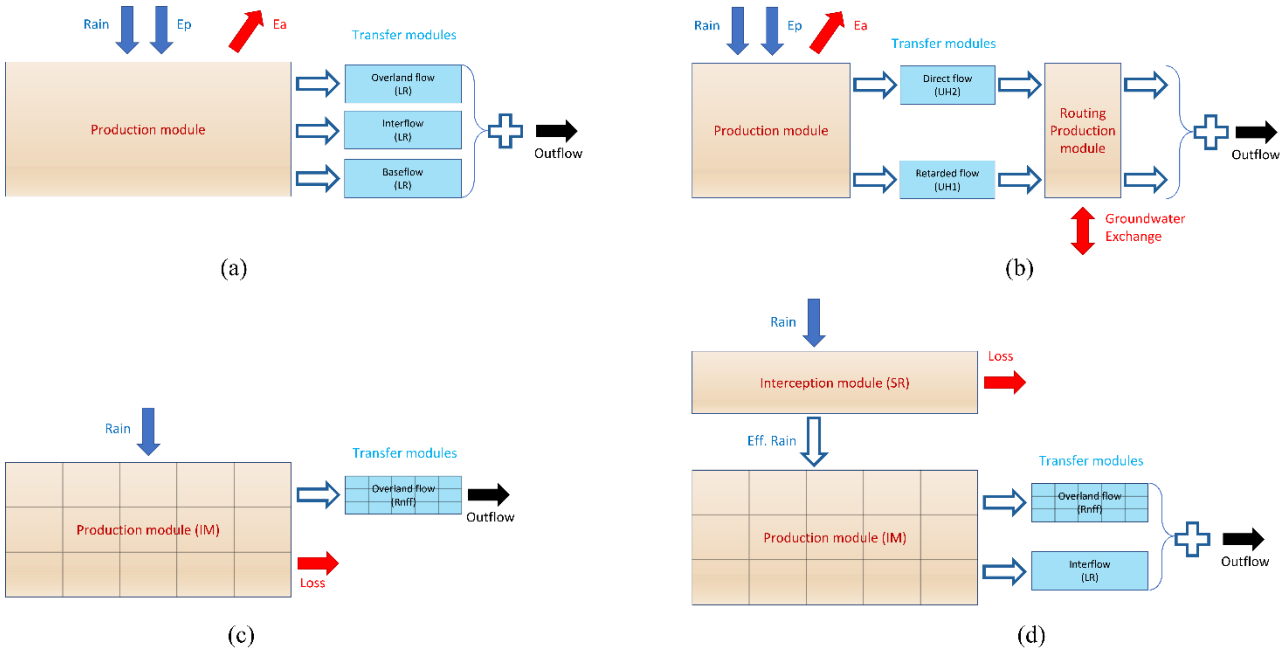


Figure 4 Conceptual representation of the models used: (a) VHM, (b) GR4H, (c) simple runoff model, (d) 2-layer model. The solid arrows represent the forcing input (in blue) and outputs (in black and in red). The other arrows represent the internal exchanges within the model and between modules. The production modules are depicted as orange boxes, while blue boxes represent transfer modules. A *box* showing a regular grid represents a gridded model, whereas its absence indicates a lumped model.

Additionally, an in-house developed two-layer model was also applied in the Amblève catchment (Figure 4d). This model can be regarded as an extended version of the simple runoff model (Figure 4c), taking into account interception and fast groundwater flow (as described in detail in Appendix B). As the model is gridded, it enables considering the spatial variability of rainfall, which is important in the case of the 2021 flood as emphasised in Section 4.1.

All the models used here were implemented in the package WOLFHydro of the modelling system WOLF (Archambeau, Erpicum, Dewals, & Pirotton, 2024) developed at the University of Liège. WOLFHydro enables combining in a modular way different types of hydrological models, possibly with different spatial representations (lumped, semi-distributed, gridded). It allows adjusting the level of complexity in the representation of a catchment while remaining in the same modelling environment (Dessers, Archambeau, Dewals, Erpicum, & Pirotton, 2024), hence also ensuring that exactly the same processing of input data is applied (e.g. topography corrections, catchment delineation, river definition, mesh size, etc.) and the same calibration strategy. This structure is well suited to addressing the research questions outlined in the introduction (Section 1).

3.4.2. Hydraulic model

The hydraulic model implemented in WOLF software (Archambeau, Erpicum, Dewals, & Pirotton, 2024) was employed. It solves the shallow-water equations using a finite volume method with regularly sized cells (Cartesian grid) and an explicit Runge-Kutta time integration scheme (RK22). Manning-Sticker friction formula is used. The model was used in numerous previous research on flood modelling (Kitsikoudis, et al., 2020; Mustafa, et al., 2018; Stilmant, Pirotton, Archambeau, Erpicum, & Dewals, 2018; Detrembleur, et al., 2015; Bruwier, Erpicum, Pirotton, Archambeau, & Dewals, 2015).

Although not central to this study, this hydraulic model contributes to a direct validation tool restricted to the Vesdre catchment. Its primary utility lies in comparing the effects of simulated hydrological hydrographs on flood map footprints (one of the few data sources available for the Vesdre during the event), in assessing the plausibility of the runoff model, and in supporting the calibration of the runoff transfer module (Rnff, Appendix A.4.1).

3.5.Simulation configuration

A semi-distributed approach was adopted to divide the catchments into sub-catchments, with outlets defined at each gauging station and the confluences of tributaries along the Vesdre river, and at each gauging station only along the Amblève river. For the gridded components of the simple runoff model and the two-layer model, a computation mesh with a resolution of 100 m x 100 m was used. The same mesh was used for preprocessing of input data such as DTM (e.g. for topographic corrections and delineation of sub-catchments), landuse maps and rainfall data (e.g. assignment of rainfall quantities to individual sub-catchments), regardless of the model used.

In WOLFHydro, a cell is classified as a 'river' when its local catchment size exceeds a threshold. In all simulations, this threshold was set at 15 km².

The timestep of the simulations was set to match the timestep of the rainfall data. The spatial component of the runoff coefficient c_x (Appendix A.3.2), is given by the values in Table 5.

Slopes (s)	**s< 4 %**	**4 %≤ s < 8 %**	**8 %≤ s <16 %**	**s>16 %**
Forests	0.4 (0.7)	0.5 (0.8)	0.55 (0.85)	0.6 (0.9)
Meadows	0.1 (0.3)	0.2 (0.4)	0.3 (0.5)	0.4 (0.6)
Agriculture	0.5 (0.7)	0.6 (0.8)	0.65 (0.85)	0.7 (0.9)
Urbans	0.4 (0.8)	0.45 (0.85)	0.5 (0.9)	0.55 (0.95)
Rivers	1	1	1	1
Water	1	1	1	1

Table 5 *The time-dependent part of the runoff coefficients (C_t)* depending on the slope and the land use. The values displayed are those selected for the 2-layer model, while the values in brackets are those selected for the runoff model.

3.6.Optimisation procedure

For both watersheds, the chosen optimisation algorithm was the Simulated Annealing, as proposed by Corana, et al (1987), with the Nash–Sutcliffe efficiency coefficient (Nash & Sutcliffe, 1970)–also referred to as NSE– as an objective function.

To address the research questions of interest here and accommodate the type and quantity of available data, the optimisation strategies were adapted accordingly. Consequently, three different strategies are applied. For the sake of simplicity, a name is assigned to each of them, i.e. the strategy with the 'A' letter is referring to Vesdre, while the letter 'B' pertains to the Amblève watershed. In the same vein,

and without being restricted to any type of rain or simulation intervals, the three types of calibration are summarised in Table *6* and defined as follows:

- A: the calibration strategy for July 2021 on the Vesdre watershed to generate a hindcast.
- B1: the calibration strategy for models trained with July 2021 data on the Amblève watershed.
- B2: the calibration strategy for models trained with historical floods data on the Amblève watershed.

	Strategy A	Strategy B1	Strategy B2
Objective function	N/A	NSE	NSE
Optimisation method	Visual inspection	Simulated annealing	Simulated annealing
Parameters initial guess	N/A	Three random selections	Three random selections
Application in this study	Vesdre catchment	Amblève catchment	Amblève catchment
Influence of sub-catchments upstream of reservoirs	Yes	Release from reservoir prescribed	Flow rates prescribed at station "Thioux"
Calibration period	2021 flood event	2021 flood event	Six flood events between 1991 and 2011
Calibration data	• Time series of pool level in reservoirs, reservoir inflows • partial time series of flow rates, • 2D hydrodynamic model outcomes	Time series of flow rates at gauging stations	Time series of flow rates at gauging stations
Sequence of calibration	1. Sub-catchments upstream of reservoirs 2. All other sub-catchments	Sequential calibration from upstream (headwaters) towards downstream (parameters and hydrographs of upstream sub-catchments being frozen)	Sequential calibration from upstream (headwaters) towards downstream (parameters and hydrographs of upstream sub-catchments being frozen)
Time transfer between sub-catchments	• First estimation by visual inspection • Adaptation of the Froude river parameters to average 2D simulations time transfers	• VHM & GR4H : lag time as an additional parameter for lumped model • In-house gridded model: given directly by the time transfer computed by the model	• VHM & GR4H : lag time as an additional parameter for lumped model • In-house gridded model : given directly by the time transfer computed by the model
Initial conditions	Model structure ensures no influence of initial conditions on internal variables	Initial conditions of the internal variables: sensitivity analysis / initial height of linear reservoirs in transfer module: estimated from data	Initial conditions of the internal variables: sensitivity analysis / initial height of linear reservoirs in transfer module: estimated from data

306 Table 6 Summary of the optimisation procedures for each strategy

307

308

3.6.1. Vesdre: July 2021 calibration (strategy A)

For the Vesdre catchment, as most of the measurements at gauging stations were unusable (Table 3) due to a lack of data at the peak of the flood waves or a poor reliability of the available data, the hydrograph at the inlet of the Vesdre reservoir and the evolution of the reservoir level were used for a manual calibration of the simple runoff model. Following this, the whole set of identified parameter values was transferred to all other sub-catchments. Once these optimal parameters are applied to all the sub-domains, a visual inspection was carried out to fine-tune the parameters. This was done using the only complete hydrograph available (stream 'Ruisseau des Fonds de Forêt') and rising limbs of the incomplete hydrographs at other stations.

A preliminary application of a 2D shallow water hydrodynamic model was used to compute the flood propagation in the main valley, using the hydrographs generated by the hydrological model as a source term. This simulation provided a first approximation of the propagation time for the upstream hydrographs. These estimates helped to calibrate the parameters of the runoff transfer module (Appendix A.4.1), particularly the Froude river parameters ($Fr_{\min}^{\text{riv}}$, $Fr_{\max}^{\text{riv}}$) and the reference flow ($\tilde{q}_{\text{ref}}$) so that the propagation times in the hydrological model agree as much as possible with those of the 2D hydrodynamic model.

3.6.2. Amblève: July 2021 calibration (strategy B1)

For the Amblève catchment, the strategy favoured here was a semi-distributed optimisation proceeding from upstream to downstream. Once the optimal set of parameters was determined for an upper sub-catchment, the corresponding outflow hydrograph could be generated for that sub-domain, and this hydrograph was then considered as fixed when calibrating the parameters for sub-catchments situated further downstream.

The hydrographs corresponding to the sub-catchments were propagated downstream using a 'lag time' parameter, assuming that the shape of the hydrographs is not altered over the course of their propagation and that the propagation speed is independent of the flow intensity. This is similar to a linear kinematic wave propagation, whose celerity is constant and equal to the ratio of the river distance between two points and the 'lag time' parameter.

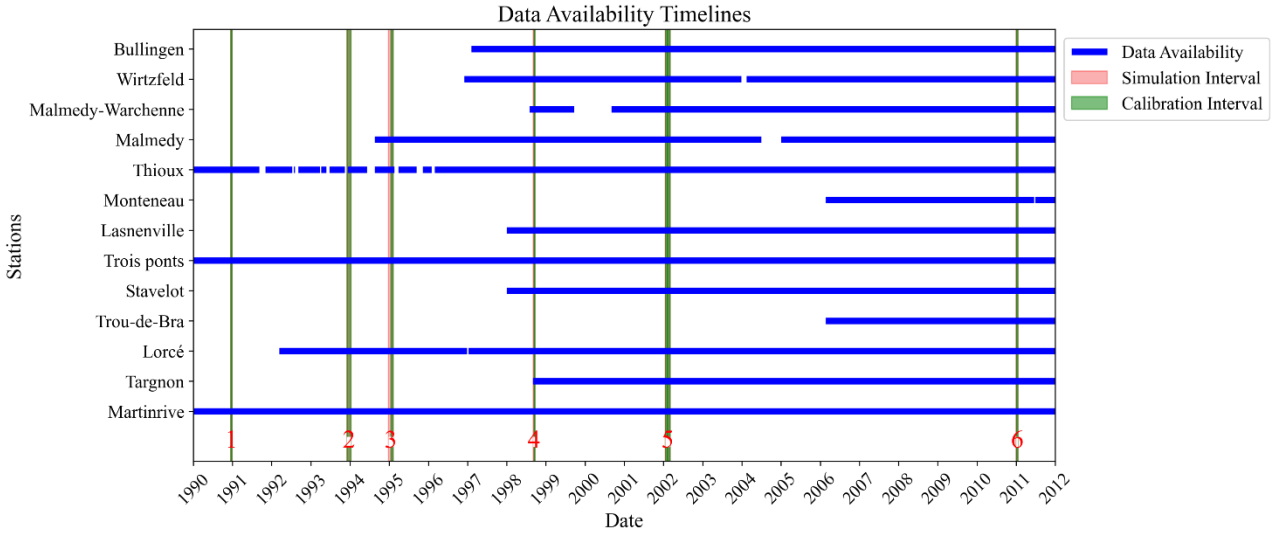
The calibration intervals were adjusted from one sub-catchment to the other depending on the availability of data within the initial calibration interval. Instead of interpolating missing data points, missing data intervals were simply excluded from the calibration intervals. It means that the objective function was not evaluated at this time step.

The calibrations were started from random initial guess values of the parameters, within a specified search window. Once the calibration algorithm converged, the optimisation was relaunched starting from the best set of parameters from the previous iteration. Three calibrations were started from random initial values, each followed by one additional calibration starting from the best parameters initially obtained. The best overall parameters were taken as those leading to the highest NSE.

A sensitivity analysis was carried out to evaluate the influence of the initial conditions of the internal variables, such as the initial water level in the production reservoirs (VHM and GR4H). As for the initial height in the linear reservoir used in the transfer modules, it was assumed that only the interflow or the second flow layer was contributing to the outlet before the event began. By using the linear reservoir equation (Appendix A.4.2), it is possible to determine the initial water level of the reservoir to fit the estimated flow just before the event.

At the dams, an outlet flow is difficult to reconstruct reliably, making it arguable to force it directly in the optimisation and simulations. Moreover, the following stations, from upstream to downstream — Bévercé, Malmedy, Thioux, Stavelot, Targnon, and Martinrive—are influenced by dam operations. Nonetheless, the gauging station Bévercé (6946), located just downstream of the dam, provides time series during the 2021 flood and is used to force the hydrograph downstream and directly replace the effect of the dam.

3.6.3. Amblève: historical floods calibration (strategy B2)



6 Discontinuous Simulation periods :

1. [14/12/1991; 29/12/1991] → Q_{max}^{Mart} = 374 m³/s
2. [29/11/1993; 10/01/1994] → Q_{max}^{Mart} = 297 m³/s
3. [25/12/1994; 07/02/1995] → Q_{max}^{Mart} = 242 m³/s
4. [03/09/1998; 27/09/1998] → Q_{max}^{Mart} = 293 m³/s
5. [16/01/2002; 04/03/2002] → Q_{max}^{Mart} = 275 m³/s
6. [01/01/2011; 18/01/2011] → Q_{max}^{Mart} = 287 m³/s

7 Calibration periods :

- [16/12/1991; 29/12/1991]
- [06/12/1993; 10/01/1994]
- [16/01/1995; 07/02/1995]
- [12/09/1998; 22/09/1998]
- [20/01/2002; 06/02/2002]
- [10/02/2002; 04/03/2002]
- [05/01/2011; 18/01/2011]

Figure 5 Representation of the data available per station, as well as the simulation and calibration intervals chosen.

To take into account historical floods, a preliminary screening was conducted to identify the most relevant events for comparing the July 2021 flood with previous events. For this, the candidates should correspond to major flood events over the last three decades but also share similarities with the studied event. Specifically, the chosen events should have several rain-free days preceding the event, followed by a constantly increasing measured hydrograph leading up to the peak, and then a recession curve, with minimal subsequent rainfall. Six events meeting most of these criteria were selected. In contrast with the July 2021 floods, almost all of these events occurred during winter, while only one taking place in summer. This seasonal distribution reflects the region's typical flood patterns, where most floods occur in winter. The maximum discharge of each event in Martinrive is described in Figure 5, and they are ranging from 242 m³/s up to 374 m³/s, while a flow rate of 661 m³/s was reached in July 2021 at the same station.

Conducting an automatic optimisation on such extensive time series and numerous events presents several challenges and implies some constraints.

First, like the July 2021 event, a flow out of the dams is difficult to reconstruct reliably. Nonetheless, the gauging station Thioux (L5580), located just downstream of the dam, provides complete time series during all the calibration periods selected. Even though it compels to renounce calibrating the model

with this station and the one in Malmedy, the complete measurements at Thioux make it an excellent choice for forcing the hydrographs and calibrate the downstream stations mentioned previously.

Second, the availability of data at the gauging stations varies significantly. The upstream-to-downstream calibration carried out for July 2021 (B1) was still applied, but calibration intervals without available data were excluded, resulting in varying the numbers of calibration intervals across stations. It means that some sub-catchment were not calibrated with the same level of quality. Nonetheless, once the optimal set of parameters was obtained, time series for all the simulation periods were generated and forced as inputs of the downstream elements. This strategy implies that the number of calibration intervals in a sub-catchment is independent of the calibration intervals of the upstream elements.

Finally, calibrating hydrological models over such extensive simulation intervals, particularly for gridded models, can be time-consuming. Moreover, as the focus is on event simulations, the periods between the flood events are unnecessary. That is why, discontinuous simulations are carried out to remove non-essential intervals in times series, while still providing event simulations, and allowing the use of a consistent objective function across all extreme events.

4. Results and discussion

4.1. Vesdre catchment

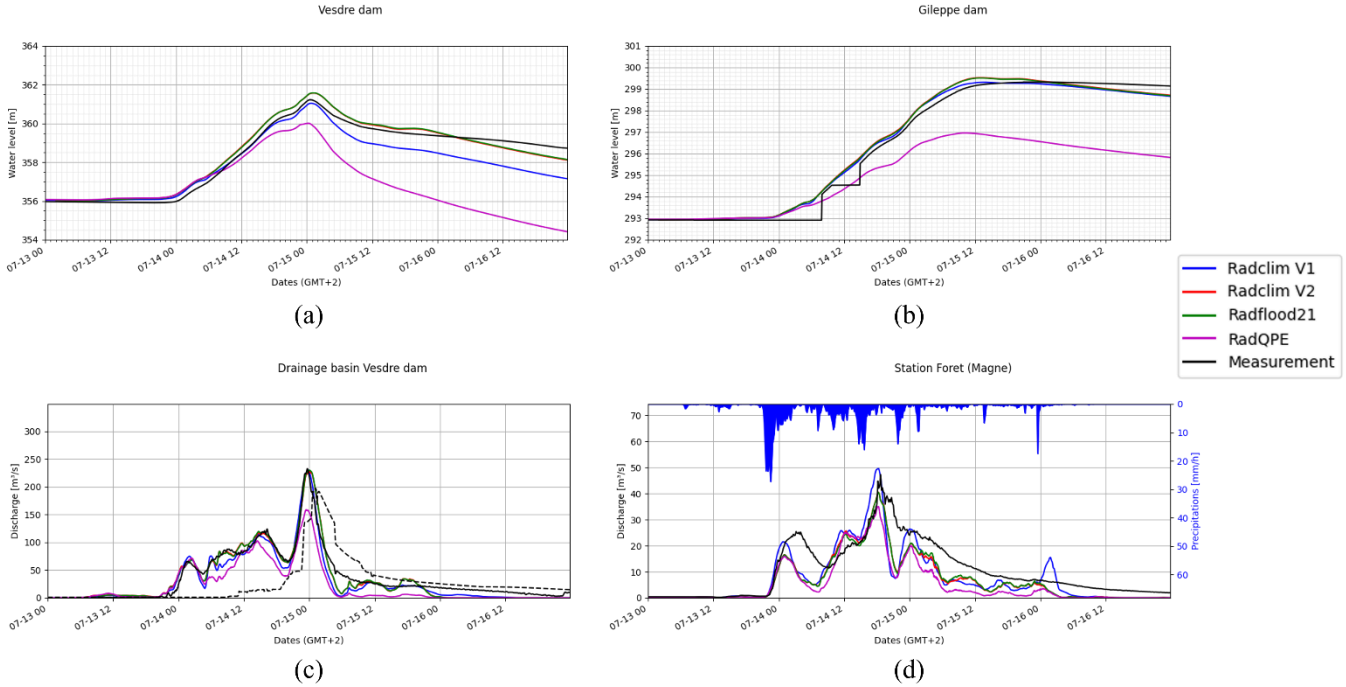


Figure 6 Examples of observations and computations based on various products for precipitation data. The water levels in the reservoirs are displayed in (a) for the Vesdre dam and in (b) for the Gileppe dam. In (c), is represented the inlet of the Vesdre dam in solid line and outlet in dashed line. The hydrograph in (d) is the one at the station Station Forêt Magne.

4.1.1. Validation against hydrological observations

Upstream of the reservoirs

When the runoff hydrological model is applied upstream of the reservoirs located in the Vesdre catchment, forced with Radclim V1 data, a manual calibration enables obtaining a relative error at the peaks of the hydrographs of about 1%, where it occurs the 15th of July at about 23h40 (GMT+2) according to the inlets reconstruction and at 23h50 (GMT+2) for the calibrated model (Figure 6c). As for the water levels, the error made is less than 0.1%, which represents an underestimation of the water levels less than 0.01 metre for the Gileppe and 0.1 metre for the Vesdre dam (Figure 6a and Figure 6b).

Downstream of the reservoirs

Furthermore, when the simulation is applied to the whole catchment, the estimations seem to be consistent with the partially reliable measurement and with the station downstream in 'Ruisseau des Fonds de Forêt' (Figure 6d). At this station, the peak coincides perfectly with the one measured (relative error negligible and much less than 1%) and estimated it on the 14th at 18h20, while measured at 18h40. Even though the peaks are, in general, well captured by the model, as well as the rising limb of the hydrographs, it fails to represent correctly the recession limb, which decreases too fast in the computations. It is not surprising, as the model structure enables only representing the runoff flow, while the recession limb of a hydrograph is primarily controlled by subsurface flow. It was also observed that there is a lag time of 4 hours ($\tilde{\delta}$ parameter in Appendix A.3.2) between the moment intense rain is falling and the moment the infiltration rate in the soil seem to evolve.

4.1.2. Discussion on the influence of precipitation data

Using different rainfall input data alters notably the results of the hydrological model. As highlighted in Section 3.1, the RadQPE input data predicts significantly lower total rainfall volumes compared to the other rainfall data. This discrepancy is perceptible in the computed water levels in the reservoirs (Figure 6a and Figure 6b) and in the estimated hydrographs both at the Vesdre reservoir inlet (Figure 6c) and across the entire catchment (Figure 6d).

The limited differences between Radclim V2 and Radflood21 inputs lead to minimal influence on the respective hydrographs. On the other hand, the Radclim V1 model tends to predict higher discharge volumes, especially in the downstream areas of the Vesdre catchment. Despite this, it engenders lower volume in regions surrounding Belleheid and the two reservoirs.

Although the hydrological models were initially calibrated using Radclim V1 input data for the dam, Radflood21 and Radclim V2 appear to offer even more accurate estimations of the inlet hydrographs for the Vesdre dam. The three models tend to overestimate the first section of the rising limb of the hydrographs. However, while Radclim V2 and Radflood21 align closely with the second part of the rising limb, Radclim V1 underestimates this portion slightly (Figure 6c). This underestimation compensates for the earlier overestimations, resulting in a more accurate estimation of the water level variations in the reservoir.

4.1.3. Interpretation of the results in terms of extreme nature of the event, and link with runoff coefficients

When these different results are analysed and compared with the statistical flow values established before the 2021 event, it was found that the peak flow estimated in Eupen is equal to $Q^E_{\max} \gtrsim 4\, Q^E_{100}$, where Q^E_{100} represents the flow in Eupen with a return period of 100 years. If the same comparison is

applied in Chaudfontaine, it indicates a peak flow of $Q^C_{max} \lesssim 3\, Q^C_{100}$, further emphasising the exceptional nature of this flood event.

Plausibility of obtained runoff coefficients (and values supporting the selected model structure)

This hindcast displays a global runoff coefficient near the dams to reach around 67%, while it varies between 55% and 60% in the main tributary (Hoëgne river). Further downstream in the catchment, the runoff coefficient decreases to approximately 40%. These runoff coefficients are linked to the volume of rain fallen locally, which is coherent with the observation made in Section 3.1 and in Section 4.1.2. The results highlight the extreme character of the event and importance of spatial influence of the rainfall in the runoff generation across different parts of the catchment, with higher runoff near the dams and lower values in the downstream areas. The high runoff coefficients support the initial hypothesis to only consider the runoff flows.

4.1.4. Indirect validation by examining hydrodynamic modelling outcomes

The computed hydrographs, obtained with Radflood21, were used to force a GPU 2D hydraulic model of the Vesdre river and a portion of the Hoegne tributary using the WOLF software (Archambeau, Erpicum, Dewals, & Pirotton, 2024). This involved applying the hydrographs generated by the hydrological model at Theux and those from the dams as boundary conditions. The hydrographs from each internal point along the Vesdre were injected into the river's 2D model as lateral flow at their respective locations. The results of the simulations were then compared to data collected by the SPW.

In Verviers, at first glance, the comparison between the simulated water heights and SPW observations appeared to show a poor fit, as seen in Figure 7. However, the SPW distributes spatially the water height without taking into account the presence of buildings, while the hydraulic model explicitly resolves the flow around them. When the buildings are excluded from the comparison, the flood footprint between the simulated and observed results becomes much closer. This trend is even more obvious in Chaudfontaine, where both footprints coincide almost perfectly, as observed in Figure 8.

Although these good matches need to be nuanced, since these footprints may be influenced more by the underlying topography than the intensity of the flow itself. Further analysis is required to thoroughly assess the quality and accuracy of these numerical simulations.

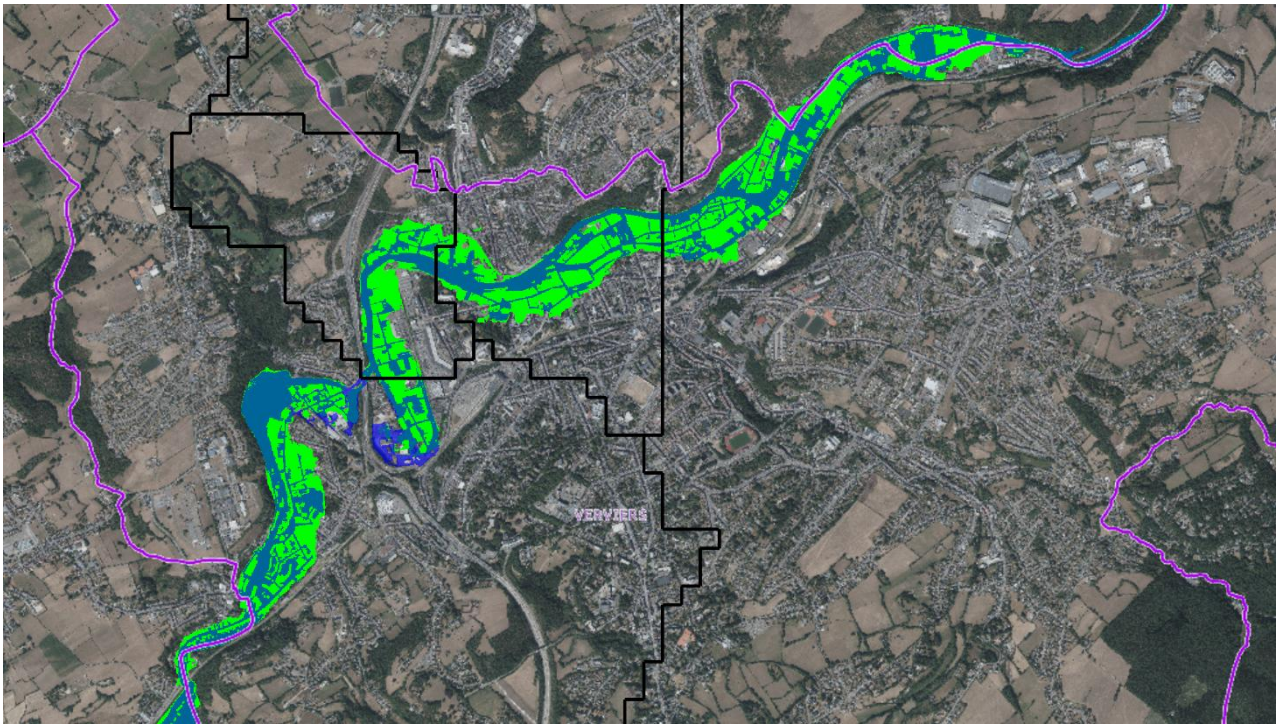


Figure 7 Comparison of the footprint of the floods in Verviers between the WOLF 2D hydraulic modelling (in transparent deep blue) and the digitalisation of the SPW (in green). When the blue colour appears lighter, it indicates an overlap with the green colour. The black contours are the delimitation of the sub-catchment and the purple ones are the municipality's borders.

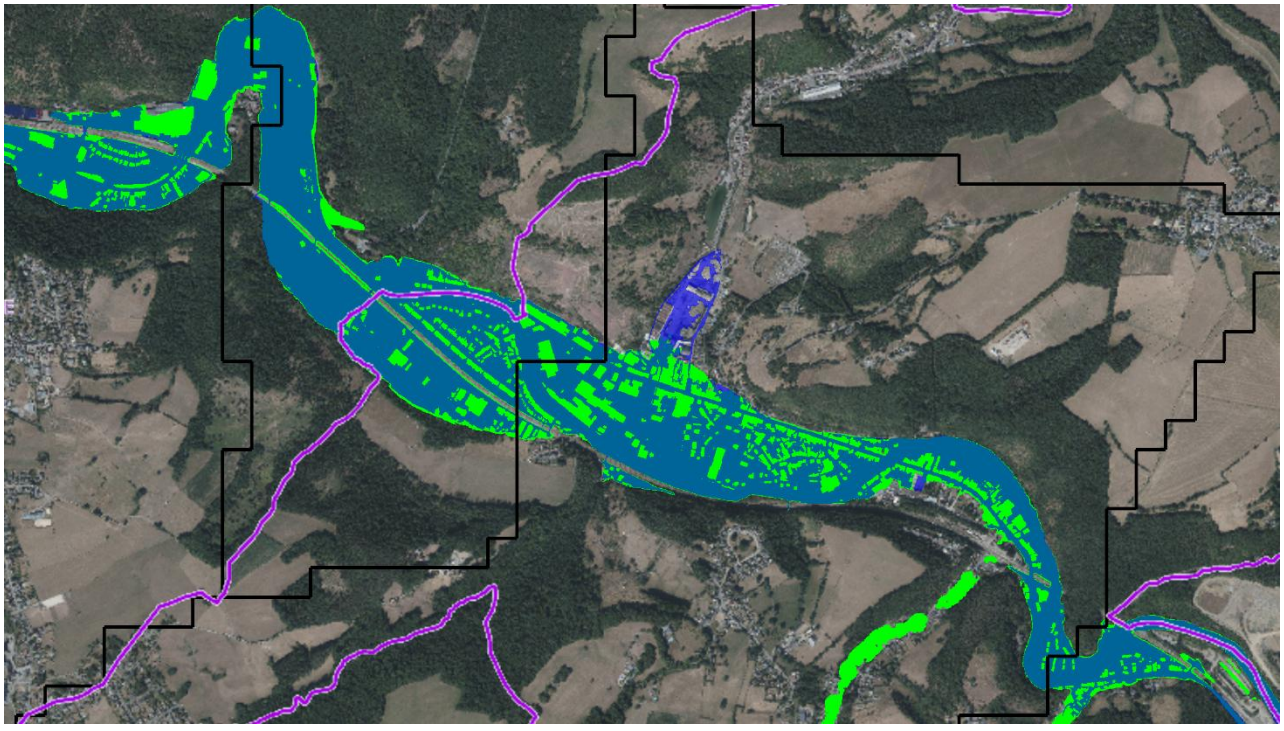


Figure 8 Comparison of the footprint of the floods in Chaudfontaine between the WOLF 2D hydraulic modelling (in transparent deep blue) and the digitalisation of the SPW (in green). When the blue colour appears lighter, it indicates an overlap with the green colour. The black contours are the delimitation of the sub-catchment and the purple ones are the municipality's borders.

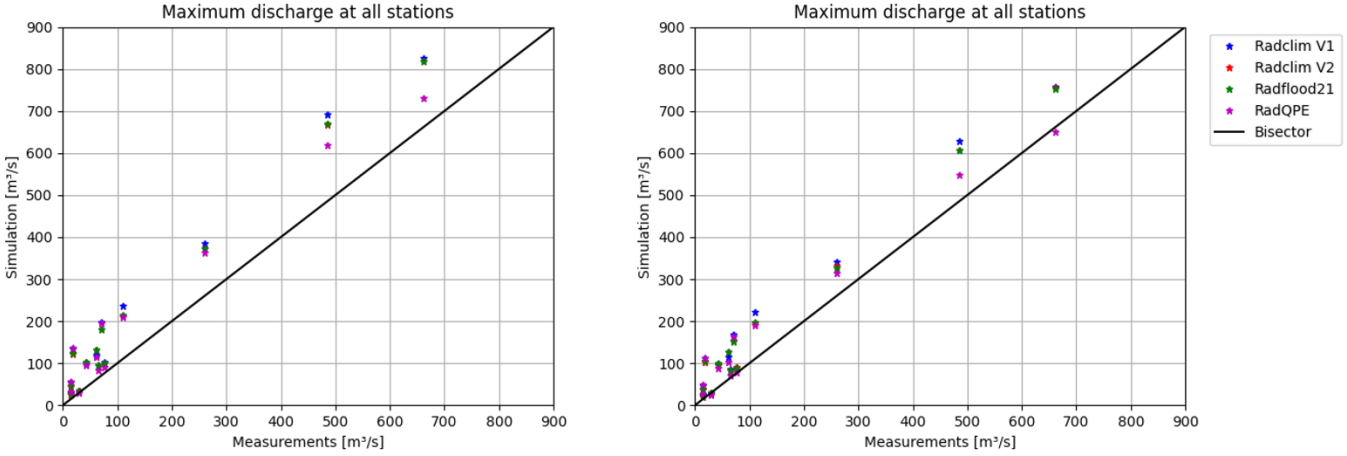
4.2. Amblève catchment

4.2.1. Transfer of the runoff model calibrated for the Vesdre catchment

For the gridded single-layer runoff model, two different approaches were employed to transfer the parameters from the Vesdre catchment to the Amblève catchment.

1. The first approach replicated the strategy in the Vesdre valley by leveraging the transferability of the IM module parameters in accordance with Horton's law, as described by Equation (A.21) . This method assumes that the maximum runoff capacity of any soil will be reached regardless of the rain intensity. With this hypothesis, the runoff response essentially relies on the land use of the catchment and the temporal progression will be identical for all catchments.

2. The second approach applied the IM module parameters calibrated with the most reliable data, i.e. from the Vesdre dam catchment, and transferred these parameters unaltered to all sub-catchments. This method ties the runoff coefficient to the rain volume within each sub-domain with respect to a maximum soil capacity considered to be constant throughout the entire catchment.



(a) (b)

Figure 9 Computed vs. observed peak discharges based on the first (a) and the second (b) approaches for transferring to the Amblève catchment the *gridded* single-layer runoff model calibrated for the Vesdre catchment

As shown in Figure 9a, the first approach for model transfer leads to considerable overestimations of the peak discharges in all sub-catchments, with an average relative error reaching 140%, mainly increased by the strong overestimations in the headwater catchments. This suggests that the first approach for model transferability is not appropriate here, because the rainfall volume was much lower in some parts of the Amblève catchment than in the Vesdre catchment. The upper part of the Amblève catchment experienced about half the volume of rainfall recorded in the Vesdre catchment between the 13/07 and the 17/07 (ratio Ternell–289 mm–and Bütgenbach–124 mm–stations is ≈ 2.3). The effects of this uneven distribution of rainfall are not properly captured by the first tested approach for model transfer, leading to inaccurate results (Figure 9a). Additionally, the assumption that every sub-catchment would reach its maximum runoff capacity appears to be inappropriate in this context.

When the second approach for model transfer is used, strong overestimations of peak flows are still observed in Figure 9b for the headwater catchments (from 87 to 500% of relative error at the peak). In contrast, lower relative errors (between 15% and 30%) are obtained at gauging stations located further

downstream. For instance, at the gauging station Martinrive, the computed peak discharge aligns closely with the observed values (relative error between − 2% and 15% on the peak discharge, and a NSE between 0.49 and 0.52), irrespective of the considered rainfall dataset. At this station and for all rainfall estimations, the rising limb is overestimated and this model—due to its inherent characteristics—systematically leads to a rapid decline in the falling limb of the hydrographs (Appendix B). Despite these model features, in the sub-catchment of Malmedy-Warchenne (a headwater stream), the model performs well (NSE from 0.57 to 0.70), with accurate predictions of the peak discharge (relative error between less than <1% and 13%), suggesting that the single-layer runoff model is suitable in certain types of sub-catchments and for this specific event.

Changing the rainfall dataset used for forcing the model has a limited influence on the computed hydrographs, except in the case of the product RadQPE according to which the cumulative rainfall volumes are about 20% lower than in the other datasets (e.g. Radclim V2). Consequently, the model forced with RadQPE produces hydrograph peaks that are closer to the observations, as the underestimation of rainfall inputs partially compensates for the overall overestimation introduced by the runoff model.

The results indicate the necessity of adopting more elaborate hydrological models that incorporate subsurface flow, in order to accurately reproduce the 2021 flood event in the Amblève catchment.

4.2.2. Models' calibration based on the 2021 event

The results presented in Section 4.1.1 hint at the influence of spatially-varying properties of the sub-catchments. That is the reason why the model parameters were identified per sub-catchment, progressing from upstream towards downstream.

The obtained Nash–Sutcliffe coefficients (NSE) obtained for each sub-catchment are summarised in Figure 10. They range between 0.89 and 0.99. When the NSE is averaged over all sub-catchments, the following values are obtained: 0.966 for the two-layer gridded model, 0.952 for model VHM and 0.962 for model GR4H. Hence, the three models exhibit a strong ability to reproduce the observed data.

The performance of the two-layer model and of the model VHM is comparatively lower in sub-catchments situated upstream of dams, such as Bullingen and Wirtzfeld, where complex hydrological dynamics are challenging to capture. At gauging stations located in the downstream part of the catchment (Targnon and Martinrive), water deficits are observed in Figure 12. This may be attributed either to inter-catchment groundwater exchanges, or to errors arising from extrapolation of the rating curves for such an extreme event, or even to the effect of flood plains on the flow, which were more present in that section of the river. Consequently, the calibration of the Two-layer model and of model VHM leads to parameter values enhancing runoff, but this outcome may result from unresolved hydrological processes and/or inaccuracies in streamflow data.

In contrast, the model GR4H performs well for predicting peak discharges at downstream stations, as shown in Figure 12, thanks to external flow adjustments being incorporated into the model. This feature enables simulating inter-catchment exchanges, as well as compensates for errors in rainfall or flow volume measurements. Interestingly, the accuracy of this model tends to be particularly high in sub-catchments not well simulated by the other two models, but GR4H shows a lower accuracy in sub-catchments where the other two models perform remarkably well (Figure 10).

Similar additional tests (not shown here), allowed to conclude that the results of the optimisation appear mostly insensitive to the sampling frequency (5 minutes, 10 minutes, or 1 hour) of the

streamflow data used for model calibration. Similarly, the source of rainfall data (Radclim V1, Radclim V2 or Radflood21) has a limited influence on the optimisation results, except when the product RadQPE is utilised. The use of RadQPE rainfall data leads to underestimations of the peak discharges, particularly in the downstream part of the catchment (at gauging stations Targnon and Martinrive).

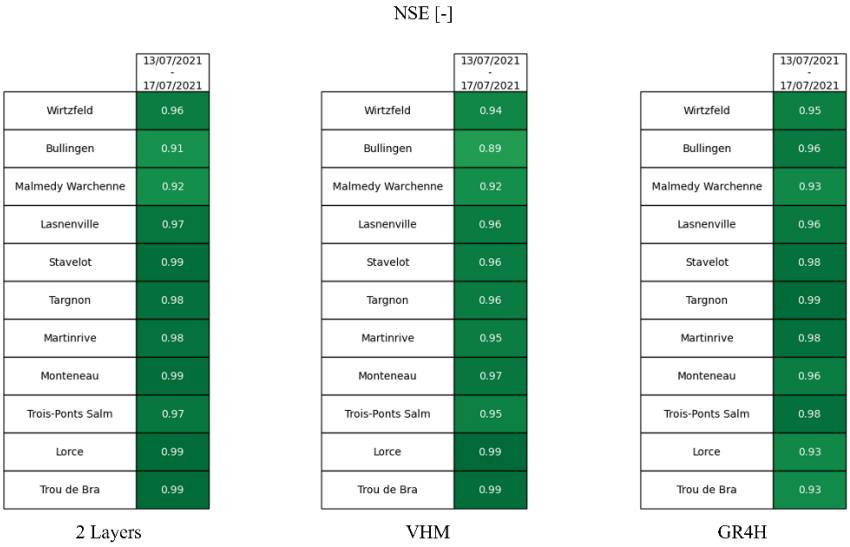


Figure 10 Nash–Sutcliffe efficiency coefficient (NSE) of the hydrological models (2 layers, VHM, GR4H) during 2021 floods *on their calibration interval*.

4.2.3. Application of models from historical floods to the 2021 event

Calibrating the three considered models (Two-layer model, VHM and GR4H) based on historical floods leads to the model performance metrics provided in Figure *11*. For the events used for the calibration, the three models tend to underestimate the observed peak discharges in most cases.

The two-layer model performs generally well. Specifically, it accurately captures the two-peak flood event of 2002 (Figure *F.12* and Figure *F.14* in Appendix), as reflected in the relatively high NSE values obtained for this event. The VHM model consistently performs well for all the considered floods and at all gauging stations (except for a single event, at a single station: Lorcé). Observations at stations such as Malmedy-Warchenne and other higher-discharge gauging points downstream are particularly well represented by the model VHM. In contrast, the model GR4H exhibits poorer performance, failing to capture the dynamics of floods involving several peaks (Figure *F.13* in Appendix). Except for two flood events (1991 and 1993), GR4H fails to reproduce the flood dynamics at the stations Trois-Ponts and Lorcé (Figure F.1 and Figure *F.3* in Appendix). It is also unable to reproduce the streamflow in the headwater sub-catchments for the 2002 event (Figure *F.7* in Appendix).

The 1998 event is the only summer flood present in the sample. At several stations (e.g. Malmedy-Warchenne, Lasneville, etc.), the performance of the models is lower than it is for other events (all winter events). For instance, the streamflow is substantially overestimated by all models at Lorcé station, which is not the case for the other events. At the station Trois-Ponts, only the model VHM captures with fair accuracy the observed hydrograph (Figure *F.1*, Figure *F.3*, Figure *F.5*, Figure *F.7*, Figure *F.11* and Figure *F.15* in Appendix), whereas the two other models strongly overestimate the peak discharge at this location. This comparatively poorer performance of the models for this event hints either at issues in the spatial rainfall data or at the influence of hydrological processes typical of summer floods, which are underrepresented in the calibration process.

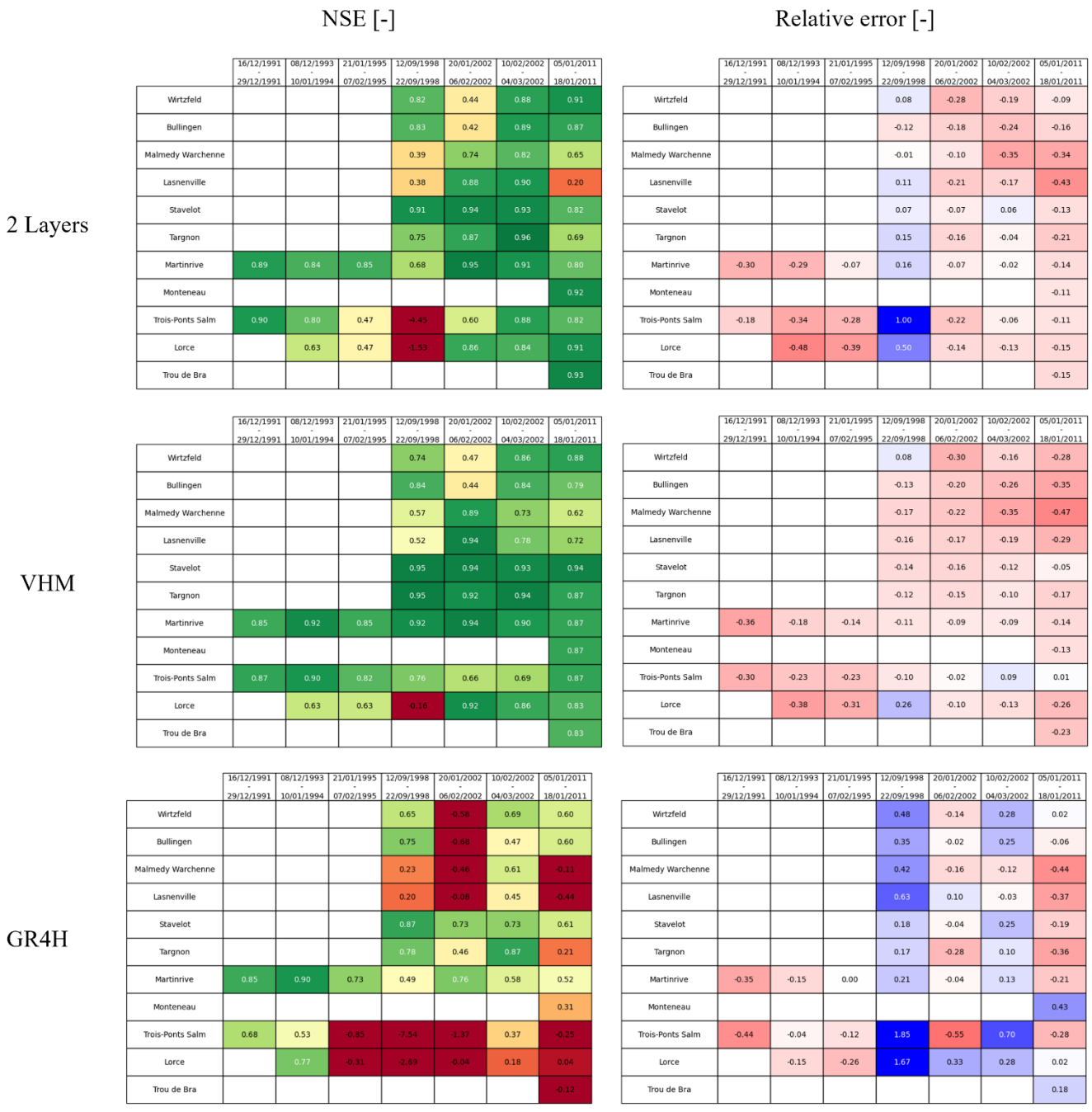


Figure 11 Performance of the considered models on their training data (historical floods). The left tables give the NSE at each station and the right table display the relative error on the peak discharge, with positive values indicating an overestimation of the model. *Cell colours provide a visual indication of performance: in the NSE tables, values below 0 (worse than using the mean discharge as a predictor) are shown in dark red, whereas an NSE of 1 (perfect agreement) is displayed in dark green. In the table of peak discharge errors, strong underestimations are highlighted in dark green and strong overestimations in dark blue.*

Applying the optimal parameters derived from historical floods to the July 2021 event results in significant underestimations of the peak discharges, as shown in Figure 12 and Figure 13. Among the models, the VHM model achieves an average NSE over all the stations of 0.72, while the Two-layer model reaches 0.68, and the GR4H model performs poorly, with an average NSE of − 0.02. The GR4H model's dynamic response is particularly inadequate (Figure *G.3* in Appendix), while the VHM and Two-layer models exhibit correct dynamics but consistently face issues with peak intensity at all locations (Figure *G.1*, Figure *G.2*, Figure *G.3*, Figure *G.4* in Appendix).

The poor performance of GR4H across different sub-catchments may be attributed to its original hourly simulation timestep. This timestep is insufficient for sub-catchments with concentration times of less than one hour. This limitation arises from the unit hydrographs containing only one or two elements, reducing the model's accuracy in smaller sub-catchments.

More specifically, all models notably fail to capture the hydrograph dynamics at Trou de Bra, likely due to scarcity of available data. Nonetheless, complete time series of available data for all the past events in the Trois-Ponts station do not necessarily induce a significant improvement in the quality of the results. Both the VHM and the Two-layer models predict the hydrograph in Wirtzfeld accurately, but none of the models effectively represents the hydrograph at Malmedy-Warchenne.

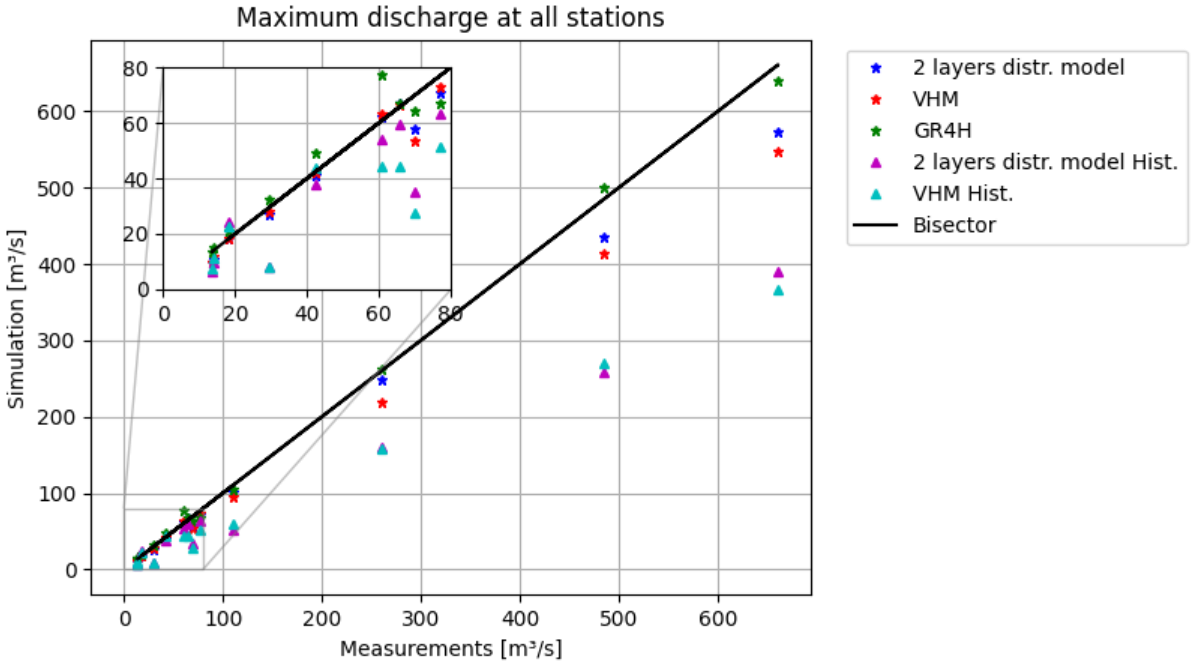


Figure 12 Hydrographs peaks (from models calibrated on historical floods) compared with measurement peaks at all stations for the 2021 event.

A striking point in the NSE values is the substantial difference in GR4H's performance between the gauging stations Martinrive and Targnon, which are relatively close. This difference may stem from the fact that the GR4H model was used here with time steps different from the hourly time step for which it was originally designed. At the gauging station Martinrive, the inter-catchment exchange feature of the model artificially adds a large water flux, so that the computed flood wave volume ends up being more than twice the rainfall volume. Using the same model parameters, an hourly simulation would result in less exchange flux, reducing the computed peak discharge from 480 m³/s to 344 m³/s. This highlights the sensitivity of GR4H to changes in timestep, although this phenomenon is most prominent in Martinrive and observed to a lesser extent in other sub-catchments.

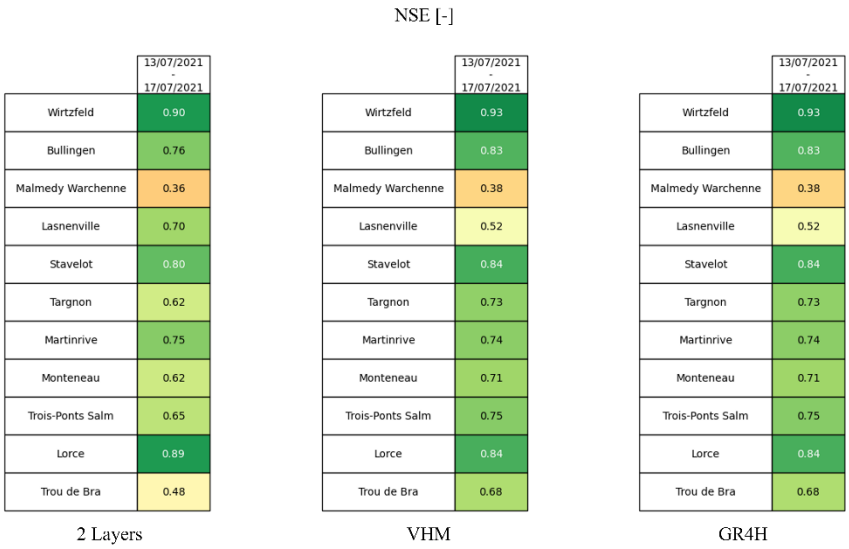


Figure 13 Nash–Sutcliffe efficiency coefficient (NSE) of the hydrological models (2 layers, VHM, GR4H) calibrated during historical floods and evaluated during 2021 floods.

The study indicates that models calibrated using historical flood events are generally inadequate for accurately predicting peak intensities, especially in downstream areas where the river is broader. The influence of the summer 1998 flood likely contributed to the reduced peak estimates for the July 2021 event, underscoring the limitations of relying solely on historical data for event-based simulation calibration and maybe emphasising the unpredictable nature of the 2021 flood.

4.2.4. Application of models from the 2021 event to historical floods

In this section, the models trained on the July 2021 flood event were applied to previous historical floods, with the results presented in Figure 14. The analysis shows model performance that strongly varies with the station and the flood event. As concluded in the previous section, the models that perform the best are VHM and the 2 layers models, while GR4H fails to represent most of the historical floods. At first sight, it can also be seen that models generally tend to overestimate more the measured hydrographs.

For instance, all models tend to overestimate the late summer flood of 1998, with notably poor performance in terms of Nash-Sutcliffe Efficiency (NSE). In many cases, the models overestimated the discharge by almost twice the measured values across nearly all computational domains. This overestimation is even more pronounced with the GR4H model.

However, the two-layer model performed well at stations located downstream, where larger surface areas were drained, showing better predictions compared to other models. Conversely, the Malmedy-Warchenne sub-catchment was poorly represented by all models, with significant overestimation of discharge. Among the models, VHM showed relatively better performance in capturing the 2011 flood, but it still struggled with some other events.

The models also failed to predict floods accurately at Trou de Bra and Lorcé, with all models providing suboptimal results. Specifically for the Trois-Ponts Salm sub-catchment, none of the models succeeded in predicting the floods between 1995 and 2003. The two-layer model consistently failed to capture most floods except for the 1991 event, where it excelled with an NSE of 0.95. The other two models performed better for the 1991 flood as well, and they managed to modestly predict the 1993-1994 and 2011 floods. This might suggest a similarity between the July 2021 and the 1991 flood in terms of hydrological conditions. Nonetheless, the Trois-Ponts Salm, and possibly Trou de Bra and Lorcé sub-

basins appear to be hydrologically complex, with multiple processes not accounted for in the selected models. Among all the models, VHM comes closest to predicting the behaviour of these aforementioned sub-catchments.

The two-layer model also captured the 2002 double flood event well, but overall, models calibrated on the July 2021 event demonstrated a tendency to predict the 1991 flood more accurately, while posing challenges for prediction, as the performance is not characterised by consistent overestimation or underestimation, but rather by varying inaccuracies across different events and locations

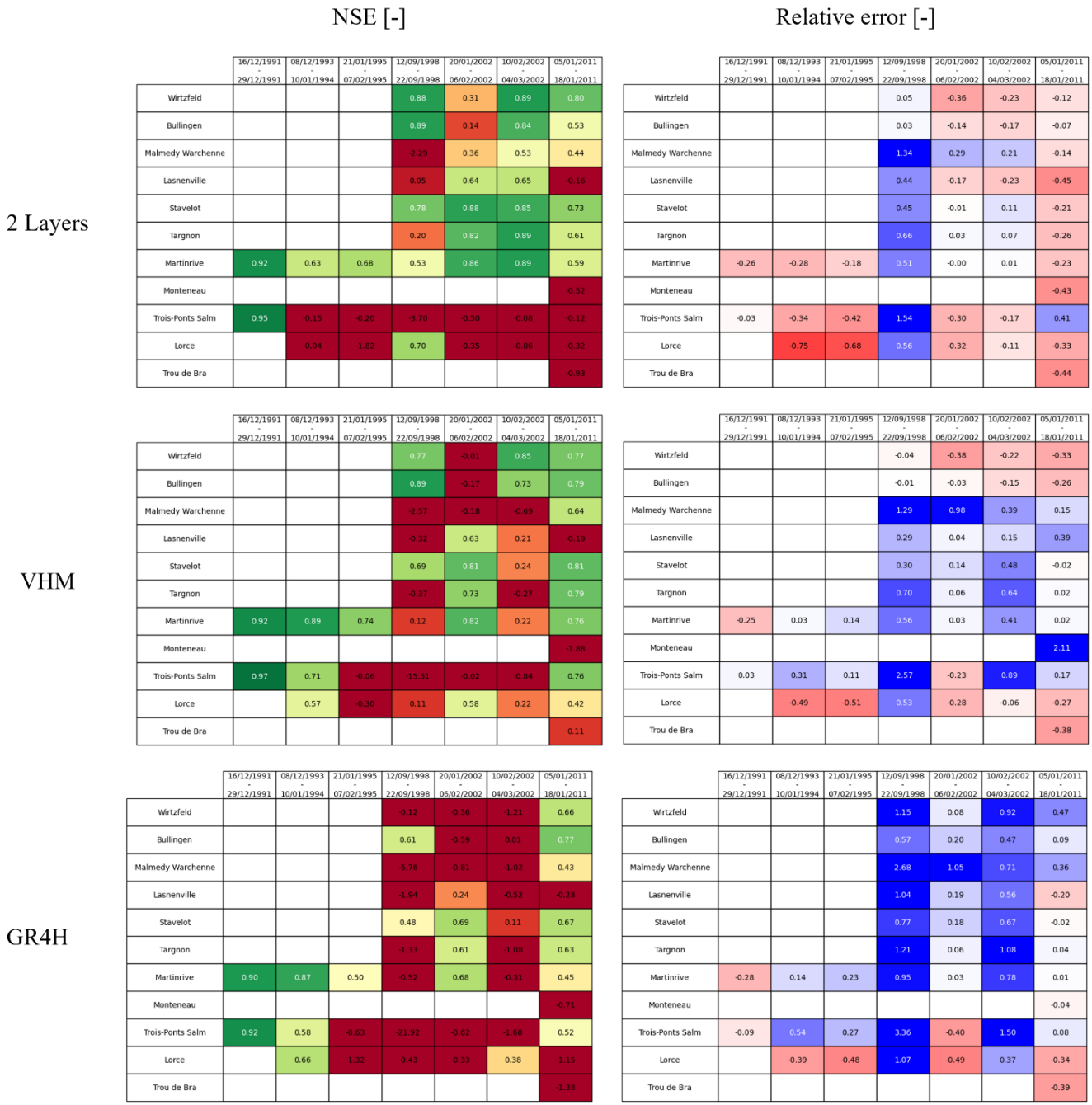


Figure 14 Tables summarising the performance of the models trained on July 2021 event and tested on historical floods. The left tables give the NSE at each station and the right table display the *relative error* on the peak discharge , with positive values indicating an overestimation of the model. Cell colours provide a visual indication of performance: in the NSE tables, values below 0 (worse than using the mean discharge as a predictor) are shown in dark red, whereas an NSE of 1 (perfect

665 agreement) is displayed in dark green. In the table of peak discharge errors, strong underestimations
666 are highlighted in dark green and strong overestimations in dark blue.

5. Conclusion

This study presents hydrological modelling of two catchments (Vesdre and Amblève, in Belgium) affected by the 2021 mega-flood in Europe. Because of considerable differences in available observation data between the two catchments, distinct modelling strategies were utilised to simulate the 2021 event. In all simulations, the simulation time step was chosen equal to the time step of the input rain data, and the presence of dams in the two catchments was explicitly considered in the modelling.

In the first catchment (Vesdre), a physically-based and gridded single-layer runoff model was selected to facilitate parameters transferability between sub-catchments. A good agreement could be obtained between the simulation outcomes and the scarce available observations, mostly pool level evolutions in reservoirs. By comparing several products for spatial radar-based precipitation data, it was shown that using more realistic rain data leads to substantial improvements in the results, even without model recalibration. The hydrological modelling results were used to force a 2D hydrodynamic model of the river and floodplains. The computed inundation extent matches well field observations.

In the second catchment (Amblève), observed rainfall quantities were considerably lower than in the first one. Therefore, a simple transfer of the runoff model calibrated for the first catchment does not lead to satisfactory results and inadequate. Several other models were tested, including a two-layer version of the gridded model and two lumped conceptual models (VHM and GR4H). Each model was calibrated based on a semi-distributed optimisation. Thanks to their internal structure based on two layers or reservoirs, the three models enable representing flow characterised by two distinct time scales. The three models performed well in reproducing the 2021 event at the 11 gauging stations available in this catchment. Using a more advanced product for spatial rainfall data further enhances the model's performance and leads to more physically sound model parameters, such as consistent mean runoff coefficients during the event. This observation was particularly relevant in the downstream part of the catchment.

The performance of the models between flood events was tested in two ways: either calibrating the models on historical flood events (period from 1991 to 2011) and testing them on the 2021 extreme flood, or the other way round. In the former case, the peak discharges of the 2021 event were strongly underestimated by the three models. In the latter case, the achieved performance was model-dependent. The model VHM tends to overestimate the peaks, whereas the two-layer gridded model slightly underestimates them. The peak discharge of the 1998 summer flood is overestimated by the three models. In this study, the model GR4H model shows poor performance when applied out of its calibration interval, particularly for small sub-catchments. This behaviour may be attributed to the time step used, which differs from that for which the model was originally designed. It may also reflect inherent limitations of the model, which is not well suited for event-based simulations and thus tends to overfit around flood events. Longer continuous simulations should be conducted to verify these conclusions.

The quality of data on rainfall quantities and their spatial distribution was shown to play a substantial role in the accuracy of the computed hydrographs, particularly at the sub-catchment level, where spatially heterogeneous model performances were observed when models were applied for historical floods, for which the rainfall spatial distribution was scarcely known. This work also puts in perspective the performance of real time generated rainfall data in their ability to be used in extreme flood event.

The study underscores the challenges of predicting a mega-flood such as that of July 2021 using only historical flood records of much lower intensity. This point is reinforced by the highly heterogeneous model performance across events, with no single model (lumped or gridded) consistently prevailing in all the cases. These findings call for further investigation into the physical processes (e.g. evapotranspiration, snowmelt, soil moisture, etc.) that control the hydrological response of each sub-catchment, as well as the role of initial catchment conditions. Snow processes, which were neglected in this study, likely played an important role in the performance of historical floods, most of which occurred in winter when snow influence is non-negligible in the Ardennes.

Overall, the study reinforces the importance of combining multiple rainfall datasets and hydrological models, implemented in the same framework, to construct envelope curves for dramatic, statistically rare events of this kind.

References

Apel, H., Vorogushyn, S., & Merz, B. (2022, September). Brief communication: Impact forecasting could substantially improve the emergency management of deadly floods: case study July 2021 floods in Germany. *Natural Hazards and Earth System Sciences, 22*(9), 3005-3014. doi:10.5194/nhess-22-3005-2022

Archambeau, P., Erpicum, S., Dewals, B., & Pirotton, M. (2024). WOLF software. *WOLF software*. Récupéré sur https://wolf.hece.uliege.be/

Barcelloni-Corte, M., Bianchet, B., Privot, J., Schelings, C., & Teller, J. (2022, October). Schéma stratégique multidisciplinaire du bassin versant de la Vesdre: Diagnostic approfondi et multithématique. *Schéma stratégique multidisciplinaire du bassin versant de la Vesdre: Diagnostic approfondi et multithématique*.

Benitez, J.-P., Dierckx, A., Matondo, B. N., & Ovidio, M. (2016). *Monitoring des passes à poissons de Lorcé et suivi de la dévalaison des saumons par télémétrie.* Tech. rep., Université de Liège, Laboratoire de Démographie des Poissons et d'Hydroécologie (LDPH).

Bertola, M., Viglione, A., Lun, D., Hall, J., & Blöschl, G. (2020, April). Flood trends in Europe: Are changes in small and big floods different? *Hydrology and Earth System Sciences, 24*(4), 1805-1822. doi:10.5194/hess-24-1805-2020

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

Chow, V. T. (1964). *Applied Hydrology* (éd. 1st). New York: McGraw-Hill.

Corana, A., Marchesi, M., Martini, C., & Ridella, S. (1987, September). Minimizing multimodal functions of continuous variables with the "simulated annealing" algorithm—Corrigenda for this article is available here. *ACM Trans. Math. Softw., 13*, 262–280. doi:10.1145/29380.29864

de Goër de Herve, M., & Pot, W. D. (2024, May). When, at what speed, and how? Resilient transformation of the Vesdre river basin (Belgium) following the 2021 floods. *Environmental Sciences Europe, 36*(1), 105. doi:10.1186/s12302-024-00928-3

Dessers, C., Archambeau, P., Dewals, B., Erpicum, S., & Pirotton, M. (2024, April). WOLFHydro: a modular framework for multi-model hydrological simulations. *EGU General Assembly 2024.* Vienna. doi:10.5194/egusphere-egu24-19450

Detrembleur, S., Stilmant, F., Dewals, B., Erpicum, S., Archambeau, P., & Pirotton, M. (2015). Impacts of climate change on future flood damage on the river Meuse, with a distributed uncertainty analysis. *Natural Hazards, 77*. doi:10.1007/s11069-015-1661-6

Dewals, B., Erpicum, S., Pirotton, M., & Archambeau, P. (2021). July 2021 extreme floods in the Belgian part of the Meuse basin. 104-107.

Dietze, M., Bell, R., Ozturk, U., Cook, K. L., Andermann, C., Beer, A. R., . . . Thieken, A. H. (2022, June). More than heavy rain turning into fast-flowing water- A landscape perspective on the 2021 Eifel floods. *Natural Hazards and Earth System Sciences, 22*(6), 1845-1856. doi:10.5194/nhess-22-1845-2022

764 Géoportail de la Wallonie. (s.d.). Occupation et utilisation du sol en Wallonie – COSW 2007 - Série -
765 Donnée historique. *Occupation et utilisation du sol en Wallonie – COSW 2007 - Série -*
766 *Donnée historique*.

767 Goergen, K., Belleflamme, A., Hammoudeh, S., Vanderborght, J., & Kollet, S. (2025). The July
768 2021 flood event in the Eifel-Ardennes mountains as simulated by the high-resolution
769 integrated hydrologic model ParFlow. *The July 2021 flood event in the Eifel-Ardennes*
770 *mountains as simulated by the high-resolution integrated hydrologic model ParFlow, 7*.
771 Frontiers Research Foundation.

772 Goudenhoofdt, E., & Delobbe, L. (2012, June). Radar-based statistics of point and areal rainfall.
773 *Proceedings of the 7th European Conference on Radar in Meteorology and Hydrology*
774 *(ERAD).* Toulouse.

775 Hosseinzadehtalaei, P., Tabari, H., & Willems, P. (2018, February). Precipitation intensity–duration–
776 frequency curves for central Belgium with an ensemble of EURO-CORDEX simulations, and
777 associated uncertainties. *Atmospheric Research, 200*, 1-12.
778 doi:10.1016/j.atmosres.2017.09.015

779 Hydrométrie. (2025). L'hydrométrie en Wallonie. *L'hydrométrie en Wallonie*. SPW Environement.

780 Javaux, M., & Dubois, D. (2016, October). *Notice technique de l'Atlas cartographique et des Fiches*
781 *descriptives développés dans le cadre du Contrat de gestion des risques d'inondation liés aux*
782 *crues et ruissellement en Brabant wallon.* techreport, Université Catholique de Louvain
783 (UCL), Earth & Life Institute.

784 Journée, M., Goudenhoofdt, E., Vannitsem, S., & Delobbe, L. (2023, September). Quantitative
785 rainfall analysis of the 2021 mid-July flood event in Belgium. *Hydrology and Earth System*
786 *Sciences, 27*(17), 3169-3189. doi:10.5194/hess-27-3169-2023

787 Keßels, J., Wolf, S., Römer, W., Dörwald, L., Schulte, P., & Lehmkuhl, F. (2024, December).
788 Enormous headward and gully erosion in a floodplain area reclaimed for open-cast lignite
789 mining during the July 2021 flood in the Inde River valley (Western Germany).
790 *Environmental Sciences Europe, 36*(1). doi:10.1186/s12302-024-00997-4

791 Kitsikoudis, V., Becker, B., Huismans, Y., Archambeau, P., Erpicum, S., Pirotton, M., & Dewals, B.
792 (2020). Discrepancies in flood modelling approaches in transboundary river systems: legacy
793 of the past or well-grounded choices? *Water Resources Management, 34*.
794 doi:10.1007/s11269-020-02621-5

795 Lehmkuhl, F., Schüttrumpf, H., Schwarzbauer, J., Brüll, C., Dietze, M., Letmathe, P., . . . Hollert, H.
796 (2022, December). Assessment of the 2021 summer flood in Central Europe. *Assessment of*
797 *the 2021 summer flood in Central Europe, 34(1)*. Springer. doi:10.1186/s12302-022-00685-1

798 Ludwig, P., Ehmele, F., Franca, M. J., Mohr, S., Caldas-Alvarez, A., Daniell, J. E., . . . Wisotzky, C.
799 (2023, April). A multi-disciplinary analysis of the exceptional flood event of July 2021 in
800 central Europe - Part 2: Historical context and relation to climate change. *Natural Hazards*
801 *and Earth System Sciences, 23*(4), 1287-1311. doi:10.5194/nhess-23-1287-2023

802 Mohr, S., Ehret, U., Kunz, M., Ludwig, P., Caldas-Alvarez, A., Daniell, J. E., . . . Wisotzky, C.
803 (2022, May). A multi-disciplinary analysis of the exceptional flood event of July 2021 in
804 central Europe. Part 1: Event description and analysis. *A multi-disciplinary analysis of the*
805 *exceptional flood event of July 2021 in central Europe. Part 1: Event description and*
806 *analysis*. doi:10.5194/nhess-2022-137

Mustafa, A. M., Bruwier, M., Archambeau, P., Erpicum, S., Pirotton, M., Dewals, B., & Teller, J. (2018). Effects of spatial planning on future flood risks in urban environments. *Journal of Environmental Management, 225*. doi:10.1016/j.jenvman.2018.07.090

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

Notebaert, B., Houbrechts, G., Verstraeten, G., Broothaerts, N., Haeckx, J., Reynders, M., . . . Poesen, J. (2011). Fluvial architecture of Belgian river systems in contrasting environments: Implications for reconstructing the sedimentation history. *Fluvial architecture of Belgian river systems in contrasting environments: Implications for reconstructing the sedimentation history, 90(1)*, 31-50. Stichting Netherlands Journal of Geosciences. doi:10.1017/s0016774600000652

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

Polrot, F. (2021, December). Numéro spécial Décembre 2021. *Karst et Inondations, Eco Karst*, 18-21. Récupéré sur https://www.cwepss.org/download/ecoKarst/ecokarst126.pdf

Pot, W., de Ridder, Y., & Dewulf, A. (2024, December). Avoiding future surprises after acute shocks: long-term flood risk lessons catalysed by the 2021 summer flood in the Netherlands. *Environmental Sciences Europe, 36*(1). doi:10.1186/s12302-024-00960-3

Rajczak, J., & Schär, C. (2017, October). Projections of Future Precipitation Extremes Over Europe: A Multimodel Assessment of Climate Simulations. *Journal of Geophysical Research: Atmospheres, 122*(20), 10,773-10,800. doi:10.1002/2017JD027176

Re, M. (2022). Hurricanes, cold waves, tornadoes: Weather disasters in USA dominate natural disaster losses in 2021. *Hurricanes, cold waves, tornadoes: Weather disasters in USA dominate natural disaster losses in 2021*.

Rodríguez, D. C., Cools, M., Roucour, S., Archambeau, P., Molinari, D., Scorzini, A. R., . . . Dewals, B. (2025, January). Can macro- or meso-scale coping capacity variables improve the classification of building flood losses? *Natural Hazards*. doi:10.1007/s11069-025-07123-4

Royal Meteorological Institute of Belgium. (2022, October). Climatological Data and Observations. *Climatological Data and Observations*.

Saadi, M., Furusho-Percot, C., Belleflamme, A., Chen, J. Y., Trömel, S., & Kollet, S. (2023, January). How uncertain are precipitation and peak flow estimates for the July 2021 flooding event? *Natural Hazards and Earth System Sciences, 23*(1), 159-177. doi:10.5194/nhess-23-159-2023

Stilmant, F., Pirotton, M., Archambeau, P., Erpicum, S., & Dewals, B. (2018). Hydraulic determination of dam releases to generate warning waves in a mountain stream: performance of an analytical kinematic wave model. *Journal of Hydraulic Engineering, 144*. doi:10.1061/(ASCE)HY.1943-7900.0001425

Thapa, B. R., Ishidaira, H., Pandey, V. P., & Shakya, N. M. (2017). A multi-model approach for analyzing water balance dynamics in Kathmandu Valley, Nepal. *Journal of Hydrology: Regional Studies, 9*, 149-162. doi:10.1016/j.ejrh.2016.12.080

Thirel, G., Collet, L., Rousset, F., Delaigue, O., François, D., Gailhard, J., . . . Wagner, J.-P. (2021, April). *Projet CHIMERE 21. CHIers - Meuse : Evolution du RégimE hydrologique au 21e siècle.* Tech. rep., INRAE UR HYCAR ; Météo-France ; EDF-DTG ; Université de Lorraine (Metz) ; INRAE UR-Riverly ; Dreal Grand-Est. Récupéré sur https://hal.inrae.fr/hal-03206168

Thompson, V., Coumou, D., Galfi, V. M., Happé, T., Kew, S., Pinto, I., . . . van der Wiel, K. (2024, October). Changing dynamics of Western European summertime cut-off lows: A case study of the July 2021 flood event. *Atmospheric Science Letters*. doi:10.1002/asl.1260

Tradowsky, J. S., Philip, S. Y., Kreienkamp, F., Kew, S. F., Lorenz, P., Arrighi, J., . . . Wanders, N. (2023, July). Attribution of the heavy rainfall events leading to severe flooding in Western Europe during July 2021. *Climatic Change, 176*(7). doi:10.1007/s10584-023-03502-7

Vorogushyn, S., Han, L., Apel, H., Nguyen, V. D., Guse, B., Guan, X., . . . Merz, B. (2024, June). It could have been much worse: spatial counterfactuals of the July 2021 flood in the Ahr valley, Germany. *It could have been much worse: spatial counterfactuals of the July 2021 flood in the Ahr valley, Germany*. doi:10.5194/nhess-2024-97

Voronoi, G. (1908). Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les paralléllòèdres primitifs. *Journal für die reine und angewandte Mathematik, 134*, 198-287. Récupéré sur http://eudml.org/doc/149291

WalOnMap. (s.d.). Géoportail de la Wallonie. *Géoportail de la Wallonie*.

Wiesehahn, L. T., & Kaifie, A. (2024, December). The impact of the 2021 flood on the outpatient care in the North Rhine region, Germany: a cross-sectional study. *BMC Public Health, 24*(1). doi:10.1186/s12889-023-17279-y

Willems, P. (2014). Parsimonious rainfall-runoff model construction supported by time series processing and validation of hydrological extremes - Part 1: Step-wise model-structure identification and calibration approach. *Journal of Hydrology, 510*, 578-590. doi:10.1016/j.jhydrol.2014.01.017

Willems, P., Mora, D., Vansteenkiste, T., Taye, M. T., & Steenbergen, N. V. (2014). Parsimonious rainfall-runoff model construction supported by time series processing and validation of hydrological extremes - Part 2: Intercomparison of models and calibration approaches. *Journal of Hydrology, 510*(February), 591-609. doi:10.1016/j.jhydrol.2014.01.028

Wolf, S., Burghardt, L., Stark, N., Ahlswede, A., Gardner, M., Lemnitzer, A., & Schüttrumpf, H. (2024, March). Morphologic interaction of bridges during high-energy flood events by the example of the flood of mid-July 2021 in the Ahr Valley, Germany. *Morphologic interaction of bridges during high-energy flood events by the example of the flood of mid-July 2021 in the Ahr Valley, Germany*. doi:10.5194/egusphere-egu24-12931

Zeimetz, F., Launay, M., Bourqui, P., Calixte, E., Fallon, C., & Teller, J. (2021, October). *Analyse indépendante sur la gestion des voies hydrauliques lors des intempéries de la semaine du 12 juillet 2021.* Tech. rep., Stucky SA.

Zhao, L., Wu, J., Zhang, Q., & Wu, F. (2014). Runoff, erosion and sediment particle size from smooth and rough soil surfaces under steady rainfall-runoff conditions. *Acta Agriculturae Scandinavica Section B: Soil and Plant Science, 64*(7), 623-632. doi:10.1080/09064710.2014.949297

Appendices

Appendix A: Modules and models description

This section aims to introduce the overall philosophy of WOLFHydro and provide a detailed description of the modules that is used throughout this study.

Appendix A.1: Conventions definition

In this article, all variables written in lower case are referred to a local value (i.e. in a chosen position or mesh), whereas capital letters represent a value valid for the whole catchment or sub-catchment studied, written $\mathcal{D}$. Consequently, if any local variable q evaluated in the position $\mathbf{x}$ as $q(\mathbf{x},t)$, a catchment variable $Q(t)$ is be defined in the following way:

$$Q(t) = \int_{\mathcal{D}} q(\mathbf{x},t)\, dS \tag{A.1}$$

In all these models, it is also important to separate internal variables to parameters to calibrate and identify them clearly. For the sake of example, then any parameter represented generally λ will be noted $\tilde{\lambda}$.

Appendix A.2: Global organisation

This model has been developed within the hydrological component of the software WOLF (Archambeau, Erpicum, Dewals, & Pirotton, 2024), called WOLFHydro. This flexible hydrological tool is organised into modules, which can be conceptualised as building blocks to combine with the aim to represent the catchment to study. This catchment can be taken as a zero-dimensional (0D) system or it can be spatially parcelled out into sub-catchments or evaluation points, which can be organised in a topology network.

In the following section, the various modules necessary to build the new models are described. Two primary module types are used in this present work: the production module and the transfer module. In **Error! Reference source not found.**, a schematic presents a representation of WOLFHydro, illustrating the general construction of a hydrological model inside the program.

The input data (e.g., rainfall, temperature, evaporation, etc.) to provide to the program can either be lumped or distributed on the study domain. This precipitation is then processed through one or more production modules, which aim to represent the physical mechanisms of the soil to distribute flow to transfer modules. The latter is usually organised into superimposed layers to represent the different flow characteristic times and their type of way to be transformed. While this is an idealised framework, it captures the fundamental organisation of most of the hydrological model's, particularly the ones that were selected for this study.

Appendix A.3: Production modules

In the WOLFHydro model, a production module is defined as a module which accepts several inputs and generates multiple outputs. In the present work, the inputs consist solely of precipitation, while the outputs include flow with different temporal characteristics (runoff and possibly fast groundwater

flow) and losses. The primary purpose of this module is to characterise the evolution of soil properties. It comprises three different processes, which are explained in the following subsections.

Appendix A.3.1: Storage reservoir (SR)

This reservoir-spillway with 2 parameters ($\tilde{H}_S$, $\tilde{T}_S$) aims to evaluate the proportion X_p of effective rain to apply to the whole studied domain. To do so, the input rain $p(x,t)$ is necessary to obtain the lumped rain of the whole catchment:

$$P(t) = \int_{\mathcal{D}} p(\mathbf{x},t)\, dS\,, \tag{A.2}$$

as well as a reservoir with the following two parameters:

- $\tilde{H}_S$: the maximum height in this reservoir,
- $\tilde{T}_S$: the characteristic time needed to empty the filled reservoir.

This reservoir is composed of two outlets:

- A net rain flow $P_{\text{net}}(t)$ to transfer to the rest of the model,
- A loss flow $Q_S(t)$, here fixed to be constant with time and space.

The loss flow is determined by the formula

$$Q_S = \frac{\tilde{H}_S}{\tilde{T}_S}\,. \tag{A.3}$$

This model contains a state variable S , which represents the evolution of the height in the reservoir.

The update equation of this variable is

$$\frac{dS}{dt} = P_{\text{raw}} - Q_S\,, \tag{A.4}$$

and the discreet version of this equation in Euler implicit scheme

$$S(t_n) = S(t_{n-1}) + [P_{\text{raw}}(t_n) - Q_S]\,\Delta t\,. \tag{A.5}$$

When this reservoir is filled ($S(t_n) > H_s$), all additional water will pour out and be considered as net rain $P_{\text{net}}(t)$.

Consequently,

$$P_{\text{net}}(t_n) = \begin{cases} (S - \tilde{H}_s)/\Delta t & \text{, if S}(t_n) > \tilde{H}_s \\ 0 & \text{, if S}(t_n) \le \tilde{H}_s \end{cases}. \tag{A.6}$$

From this net rain, the effective rain fraction everywhere in the catchment noted X_p is defined as followed:

$$X_p(t) = \frac{P_{\text{net}}(t)}{P(t)}\,. \tag{A.7}$$

In this model, the losses are considered as uniform in the whole studied catchment and are given in each point by the formula

$$q_s(t) = (1 - X_p(t))\ p(\mathbf{x}, t)\,. \tag{A.8}$$

Appendix A.3.2: Infiltration model (IM)

In this in-house runoff model containing 5 parameters ($\tilde{F}_0$, $\tilde{F}_c$, $\tilde{U}_{max}$, $\tilde{T}$, $\tilde{\delta}$), the spatial variability x and the temporal one t of the runoff coefficient c_r are decoupled, such as

$$c_r(x,t) = c_x(x)\, C_t(t) \quad , c_r \in [0;1]\,. \tag{A.9}$$

With $c_x(x)$ which is determined by the runoff coefficient matrix, dependent on land use and the slope, and $C_t(t)$ that is determined by the relation

$$C_t(t) = (1 - f_i(t))\,, \tag{A.10}$$

where $f_i(t)$ is an infiltration function evolving with time.

In this study, a function inspired by the Horton's law (Chow, 1964) is applied

$$f_i(u(t),t) = \tilde{F}_c + (\tilde{F}_0 - \tilde{F}_c)\exp\left(-\frac{U(t)}{\tilde{U}_{max}}\right), \tag{A.11}$$

with $\tilde{F}_c$ the parameter that gives the asymptotic value of infiltration $\tilde{F}_0$ gives the initial infiltration and $\tilde{U}_{max}$ the maximum capacity of the soil. The state variable $U(t)$, is determined in the following way

$$\begin{aligned} U(t) &= P_{\text{net}}(t) * \Lambda, \\ &= \int_{\mathbb{R}} P(t-\tau)\ \Lambda(\tau) d\tau, \end{aligned} \tag{A.12}$$

with $\tilde{T}$ the period of time while rain is taken into account, $\tilde{\delta}$ the delay parameter that corresponds to the time it requires before the rain affects the soil properties, and $\Lambda(t)$ the function defined as

$$\Lambda(t) = \begin{cases} 1 & \text{, if } \tilde{\delta} \le t \le \tilde{T} + \tilde{\delta} \\ 0 & \text{, otherwise} \end{cases}. \tag{A.13}$$

One can also reformulate this function as follows

$$f_i(u(t),t) = (\tilde{F}_0 - \Delta F) + \Delta F \exp\left(-\frac{U(t)}{\tilde{U}_{\max}}\right), \tag{A.14}$$

with $\Delta F = F_0 - F_c$.

Or else,

$$f_i(u(t),t) = (\tilde{F}_0 - \Delta F) + \phi(U(t),t)\,, \tag{A.15}$$

with

$$\phi(U(t),t) = \Delta F \exp\left(-\frac{U(t)}{\tilde{U}_{\max}}\right). \tag{A.16}$$

Consequently, to analyse the infiltration function, the derivation is applied

$$\frac{\partial \phi}{\partial t} = -\frac{\frac{dU}{dt}}{\tilde{U}_{\max}} \phi, \tag{A.17}$$

and then

$$\frac{\partial f_i}{\partial t} = -\frac{\frac{dU}{dt}}{\tilde{U}_{\max}} \phi(t). \tag{A.18}$$

Moreover, the derivative of the soil moisture is given by the formula

$$\frac{dU}{dt} = \frac{dP_{\text{net}}(t)}{dt} * \Lambda. \tag{A.19}$$

Finally,

$$\frac{\partial f_i}{\partial t} = -\frac{\frac{dP_{\text{net}}(t)}{dt} * \Lambda}{\tilde{U}_{\max}} \phi(t). \tag{A.20}$$

Therefore, it is possible to conclude that the variation of the infiltration does not depend on the initial condition U_0 of the soil moisture $U(t)$.

An alternative form of the Equation (A.11) in the infiltration model (IM) can be reformulated as :

$$f_i(u(t),t) = \tilde{F}_c + (\tilde{F}_0 - \tilde{F}_c) \exp\left(-\tilde{\kappa} \frac{U(t)}{V_{max}}\right), \tag{A.21}$$

with $V_{\max} = \max(U(t))$ and $\tilde{U}_{max} = \frac{V_{\max}}{\tilde{\kappa}}$.

Thus, a parallel can be made with the Horton's formula (Chow, 1964) if the Equation (A.21) is multiplied by the rain and the expression $U(t)/V_{max}$ is replaced by the time. This indicates that each type of soil will reach their maximum potential infiltration at the end of the simulation.

Appendix A.3.3: Separation of the rain in different layers

The different portions of flows use the two models introduced in the previous sections. Thus, the surface flow in each point q_{OF} is computed by the formula

$$q_{\text{of}} = x_{\text{of}}\ p(\mathbf{x},t), \tag{A.22}$$

with x_{of} the portion of rain that turns to runoff flow. This latter is determined by

$$x_{\text{of}} = c_r \; X_p \,. \tag{A.23}$$

The fraction of flow to transfer to the interflow x_{if} is given by

$$x_{\text{if}} = (1 - c_r)\, X_p \,, \tag{A.24}$$

and then,

$$q_{\text{if}} = x_{\text{if}} \; p(\mathbf{x}, t) \,. \tag{A.25}$$

Appendix A.4: Transfer modules

Modules composed of one input and one output, and as its name hints, it is usually employed to define the mathematical formulations chosen to propagate the different types of flow it is supposed to model. In this analysis, this type of module uses the mean of unit hydrographs, praised for their ease and fast computation, to represent a lag function to apply to all flows, which will depend on their time characteristics. These unit hydrographs can be constructed with a gridded method (Appendix A.4.1) or with decreasing exponential (Appendix A.4.2).

Appendix A.4.1: Runoff transfer module (Rnff)

The original model, presented in this section, proposes a new physically based and gridded way to compute runoff velocities in a fast way. This gridded model has recourse to a dimensionless number adapted to surface flow to evaluate the water speed transferred from this mesh towards the neighbour following the steepest slope. The rain or fraction of rain received by one or several neighbours can be determined by the runoff coefficient, as presented in the infiltration module.

As a reminder, the Froude number, noted Fr is a significant dimensionless number in the study of open-channel flow and its value indicates particularly the flow regime. In a rectangular section, the Froude number is defined as

$$Fr = \frac{u}{\sqrt{gh}} \,, \tag{A.26}$$

where u is the velocity of the flux ([L/T]), g the gravitational field of the Earth ([L/T²]), and h the mean water height ([L]). The Froude number can be seen as a comparison between the inertia of the flux and the gravitational field, or also the comparison between the mean flow speed and the speed of the gravity wave.

In hydrological simulations, the speed and the water height are often unknowns in a problem. The local slope is, for their part, the only data that can be collected a priori. That's the reason why, the equations, introduced hereunder, are mainly based on the slopes to distinguish a Froude number value in each mesh i. These meshes are defined with a length Δx ([L]) along the flux direction towards the flux is transferred. A distinction is made between 'river' and 'basin' meshes: to be a river element, the sum of the surfaces of all its upstream elements should be greater than a 'convergence threshold' defined by the user. For the sake of simplicity, the symbol η means that an element can either be 'river' or 'basin'.

In this model, 5 parameters are considered to characterise the runoff:

- $Fr_{\min}^{\text{riv}}$ the minimum Froude number in all 'river' elements [-],

- $Fr_{\max}^{\text{riv}}$, the maximum Froude number in all 'river' elements [-],
- $Fr_{\min}^{\text{basin}}$, the minimum Froude number in all ‘basin’ elements [-],
- $Fr_{\max}^{\text{bassin}}$, the maximum Froude number in all 'basin' elements [-],
- $\tilde{q}_{\text{ref}}$, the reference flow [L³/T].

Starting from the definition of the Froude number, the flow q_i evaluated in mesh i is given by

$$q_i = Fr_i \sqrt{g\, h_i}\; h_i\, l_c, \tag{A.27}$$

with l_c the characteristic width of the flow. Isolating the water height h_i, one obtains the expression

$$h_i = \left(\frac{q_i}{Fr_i \sqrt{g}} \right)^{2/3} \left(\frac{1}{l_c} \right)^{2/3}. \tag{A.28}$$

To find a solution to this expression, three unknows remains. For the first one, namely q_i, it is linked to a reference flow $\tilde{q}_{\text{ref}}$, which is a parameter of the model.

The second unknown is the Froude number evaluated locally. To estimate it, it was assumed that this number only depends on the local slope s_i and that its value evolves linearly within an interval delimited by the two parameters $Fr_{\min}^{\eta}$ and $Fr_{\max}^{\eta}$.

$$Fr_{i,\alpha} = Fr_{\min}^{\eta} + \omega_{i,\alpha} \left(Fr_{\max}^{\eta} - Fr_{\min}^{\eta} \right). \tag{A.29}$$

In uniform flow, speed is a function of the square root of the local slope i.e. $u_i \propto \sqrt{s_i}$. Therefore, the coefficient $\omega_{i,\alpha}$ is defined as the ratio

$$\omega_{i,\alpha} = \frac{\left(s_i\right)^{\alpha} - \left(s_{\min}^{\eta}\right)^{\alpha}}{\left(s_{\max}^{\eta}\right)^{\alpha} - \left(s_{\min}^{\eta}\right)^{\alpha}}, \tag{A.30}$$

such the coefficient $\omega_{i,0.5}$ is obtained with this formula a gives

$$\omega_{i,0.5} = \frac{\sqrt{s_i} - \sqrt{s_{\min}^{\eta}}}{\sqrt{s_{\max}^{\eta}} - \sqrt{s_{\min}^{\eta}}}. \tag{A.31}$$

Finally, the last unknown is the width of the cross section l_c, and the way it is calculated will depend on the type of cell to consider. In a river network, it was observed empirically that the width of a river was linked to the size of the catchment with the ratio of Froude number evaluated at the outlet (Fr_{out}) with two different weights of the slopes (0.5 and 1.0), times a factor 0.8 . As for a basin mesh, the width varies as a function of the ratio between the surface drained by the current mesh A_i (alternatively called convergence) and the total surface of the catchment A_{tot}.

1066 $$l_c = \begin{cases} \dfrac{A_{\text{tot}}}{\sqrt{A_i}/0.8}\dfrac{Fr_{\text{out},0.5}}{Fr_{\text{out},1}} & \text{, river} \\ \dfrac{A_{\text{tot}}}{A_i/\Delta x} & \text{, bassin} \end{cases} . \quad \text{(A.32)}$$

1067 Since then, the formula becomes

1068 $$h_i = \left(\frac{\tilde{q}_{\text{ref}}}{Fr_{i,0.5}\sqrt{g}} \right)^{2/3} \left(\frac{1}{l_c} \right)^{2/3} . \quad \text{(A.33)}$$

1069 Whatever the type of cell considered, the speed is given by

1070 $$u_i = Fr_{i,0.5}\sqrt{gh_i} . \quad \text{(A.34)}$$

1071 Finally, the time t_i [T] necessary to transfer the flux from one cell to another is computed with

1072 $$t_i = \Delta x / u_i . \quad \text{(A.35)}$$

1073

1074 ### Appendix A.4.2: Linear reservoir (LR)

1075 This module contains 1 parameter ($\tilde{K}$) and it follows the equation

1076 $$\tilde{K}\frac{dy}{dt} = x - y , \quad \text{(A.36)}$$

1077 and

1078 $$q = y / \tilde{K} , \quad \text{(A.37)}$$

1079 with $\tilde{K}$ the recession time [T], which represents the constant decrease of the baseflow with time and
1080 q the outlet flow. The variables x and y are respectively the inlets and outlets of the reservoir. This
1081 function can be solved in different numerical schemes, such as Euler explicit or implicit with Newton-
1082 Raphson algorithm.

1083 # Appendix B: In-house models

1084 The in-house modules are combined, as depicted in Figure 4 and detailed in Table 2 to form the
1085 following two models:

1086 - *Runoff model*: This is the simplest model, gridded and physically motivated, using a 1 layer
1087 model. It uses as a production module the Infiltration module (IM) from Appendix A.3.2 with
1088 the formulation from Equation (A.21) (with $\tilde{\kappa}$ parameter instead of $\tilde{U}_{max}$) and the Runoff
1089 transfer module from Appendix A.4.1. For the IM module, the parameter $\tilde{T}$ was kept equal to
1090 the simulation total time duration. The x_{if} component of the Equation (A.24) is therefore
1091 treated as loss.
1092 - *2 layers model*: It includes the Storage reservoir (SR) production module from Appendix A.3.1
1093 to generate effective rainfall before reaching the IM production module with formulation from
1094 Equation (A.14) (with $\tilde{U}_{max}$ parameter). Additionally, a linear reservoir (LR) transfer module is

applied on the second layer to represent the fast groundwater flow. In the IM module, the parameter $\tilde{T}$ was not fixed anymore and left free for optimisation. Yet, the parameter $\tilde{\delta}$ was fixed to zero. The x_{if} fraction is here sent to the interflow LR module of the second layer.

		Vesdre	Amblève		
Model		Runoff	2 layers	VHM	GR4H
Production		Gridded [IM]	Gridded [SR+IM]	Lumped	Lumped
Transfer	**Layer 1 (Qof)**	Gridded [Rnff]	Gridded [Rnff]	Lumped	Lumped
	Layer 2 (Qif)	X	Lumped [LR]	Lumped	Lumped
	Layer 3 (Qbf)	X	X	(Lumped)	X
Number of parameters		**9**	**12**	**11**	**4**

Table B.1 Summary of all the hydrological models selected and their characteristics. The Qof, Qif and Qbf refer respectively to the runoff or overland flow, the interflow and the baseflow.

Appendix C Land use conversion

The land uses were adapted and simplified to fit to the ones used in the Adali project (Javaux & Dubois, 2016)(another model used in WOLFHydro, not used in this article). They consist in 6 different types:

- Forest
- Meadows
- Agriculture
- Urban
- Rivers
- Water

In COSW 2007 data, it is organised in several levels, increasingly specific as the level increases.The first one almost contains all these characteristics, but the meadows. That's the reason why the second layer was also considered. Additionally, as these maps are given by polygons, it is possible that some regions are not contained in any polygon, or that the first layers explicitly refer to this polygon as 'no data'. In these cases, a default land use 'Urban' was attributed. For all other data the following conversion was applied:

COSW land use	**WOLFHydro conversion**
Residential areas	Urban area

Economic, service, infrastructure and communication areas	Urban area
Mines, landfills and abandoned sites	Urban area
Artificial green areas	Meadows
Other artificial surfaces	Urban area
Arable land	Agriculture
Permanent crops	Agriculture
Grassland	Meadows
Abandoned agricultural land	Agriculture
Forests	Forest
Shrub and/or herbaceous vegetation associations	Meadows
Inland wetlands	Water
Inland waters	Rivers

Table 7 Conversion of land uses provided by the second layer by COSW 2007 map towards WOLFHydro land uses selected.

Appendix D: Hydrographs Vesdre → Amblève

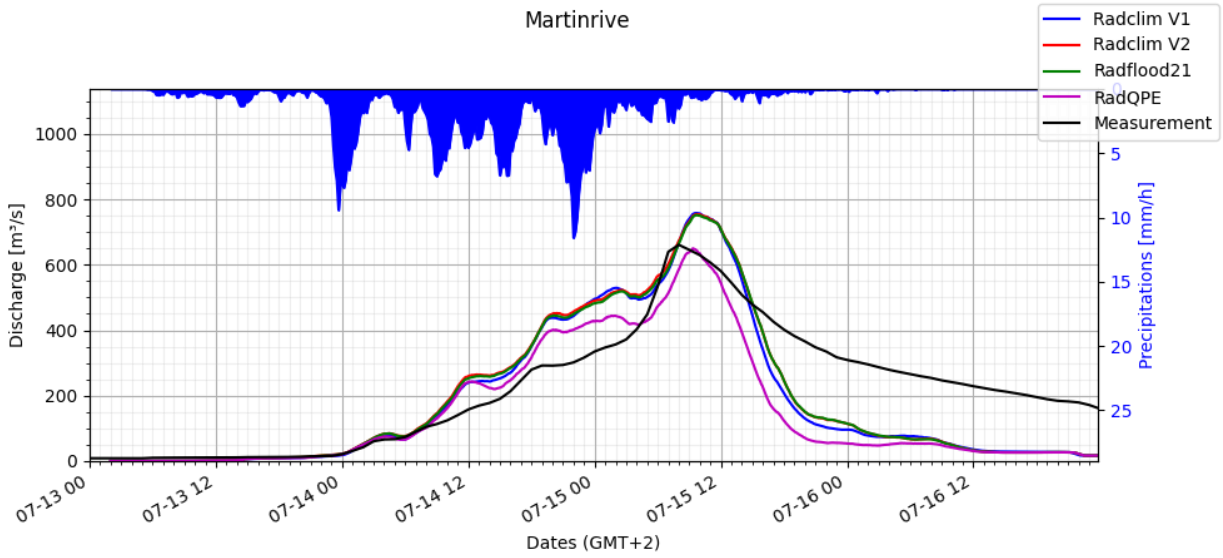


Figure D.1 Hydrographs of the Runoff model with different rainfall input data at Martinrive

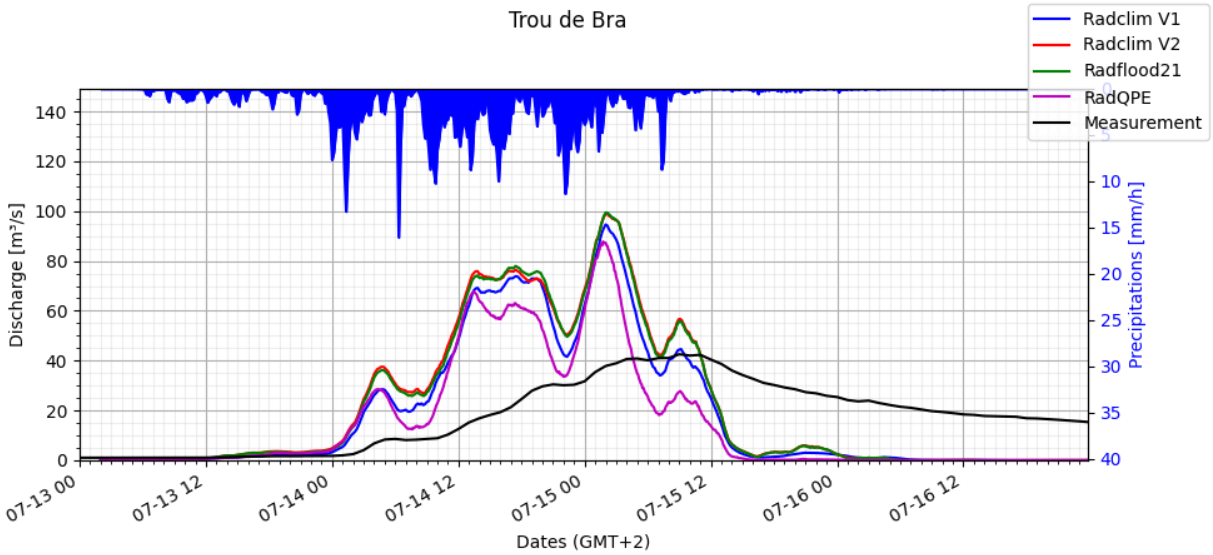


Figure D.2 Hydrographs of the Runoff model with different rainfall input data at Trou de Bra

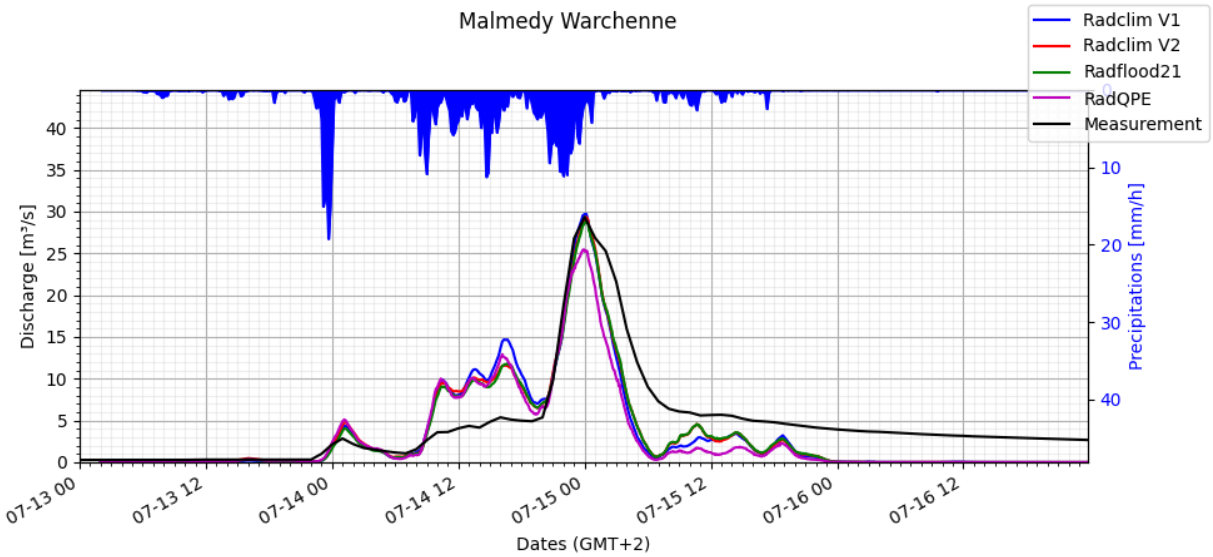


Figure D.3 Hydrographs of the Runoff model with different rainfall input data at Malmedy Warchenne

Appendix E: hydrographs Amblève event of July 2021

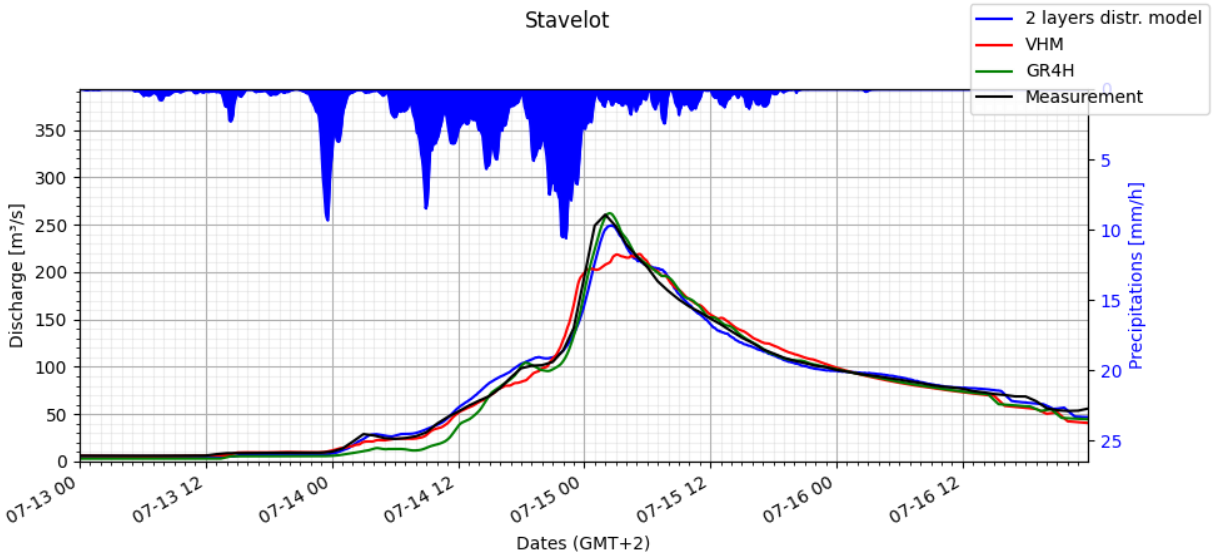


Figure E.1 Hydrographs of the models calibrated on July 2021 flood at Stavelot

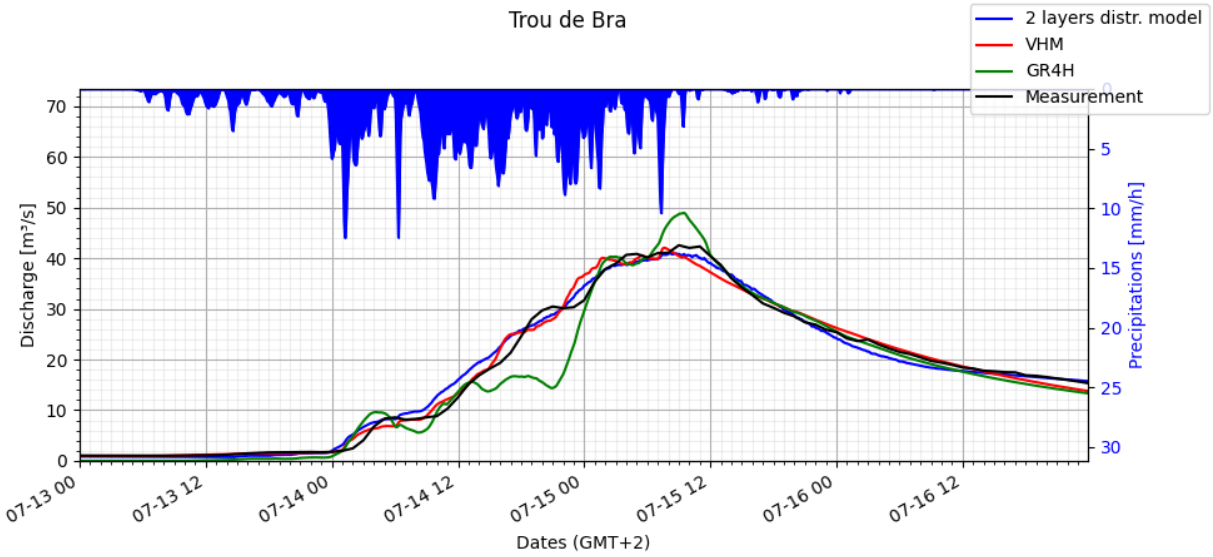


Figure E.2 Hydrographs of the models calibrated on July 2021 flood at Trou de Bra

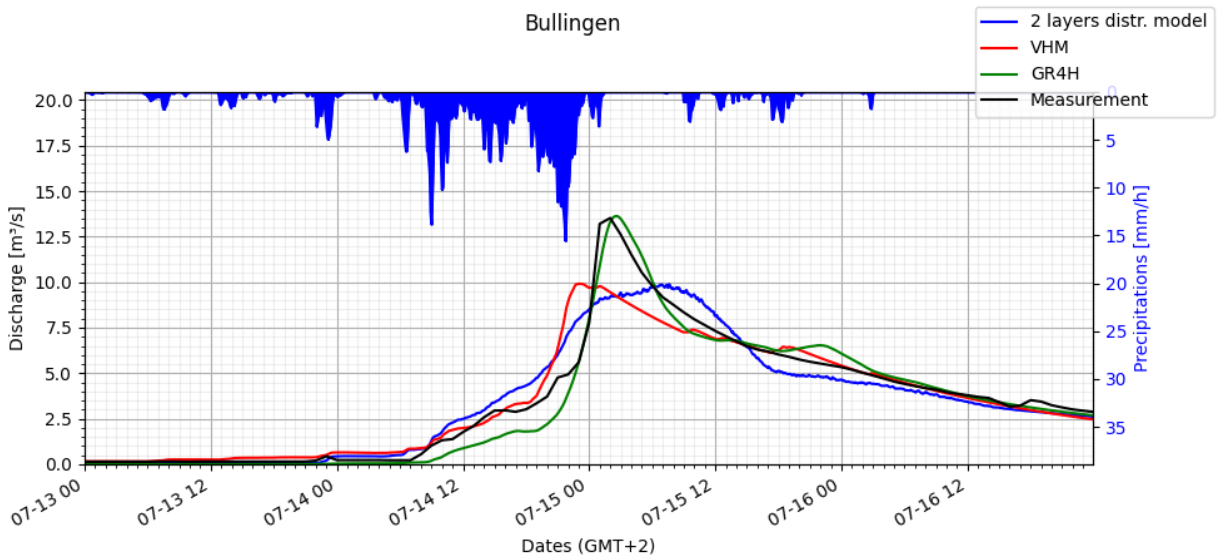


Figure E.3 Hydrographs of the models calibrated on July 2021 flood at Bullingen

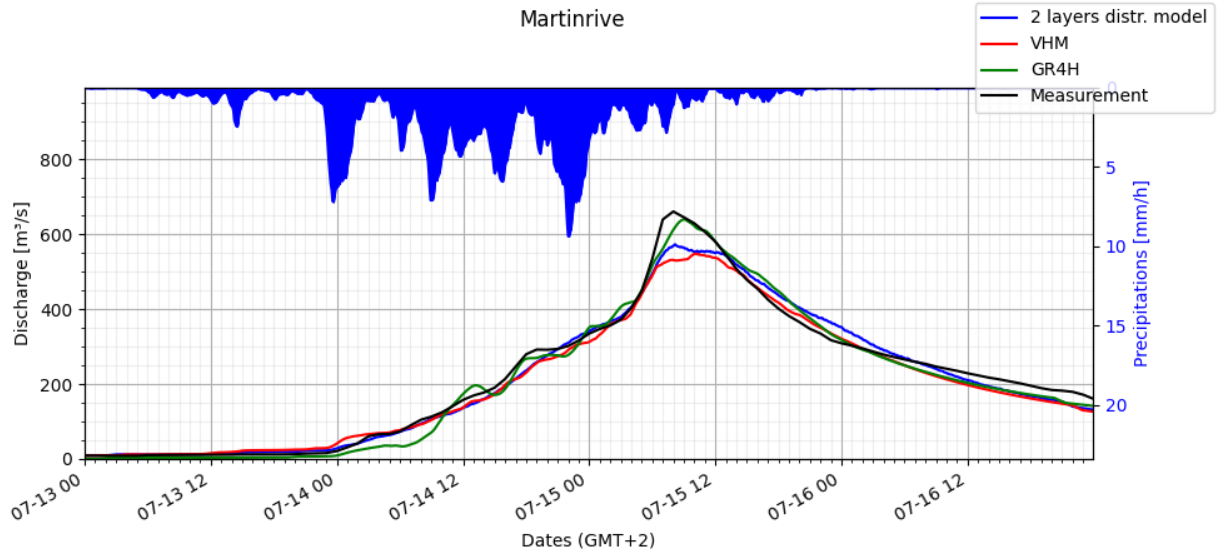


Figure E.4 Hydrographs of the models calibrated on July 2021 flood at Martinrive

Appendix F: Hydrographs Amblève 2021 → Historical

Appendix F.1 : 1991

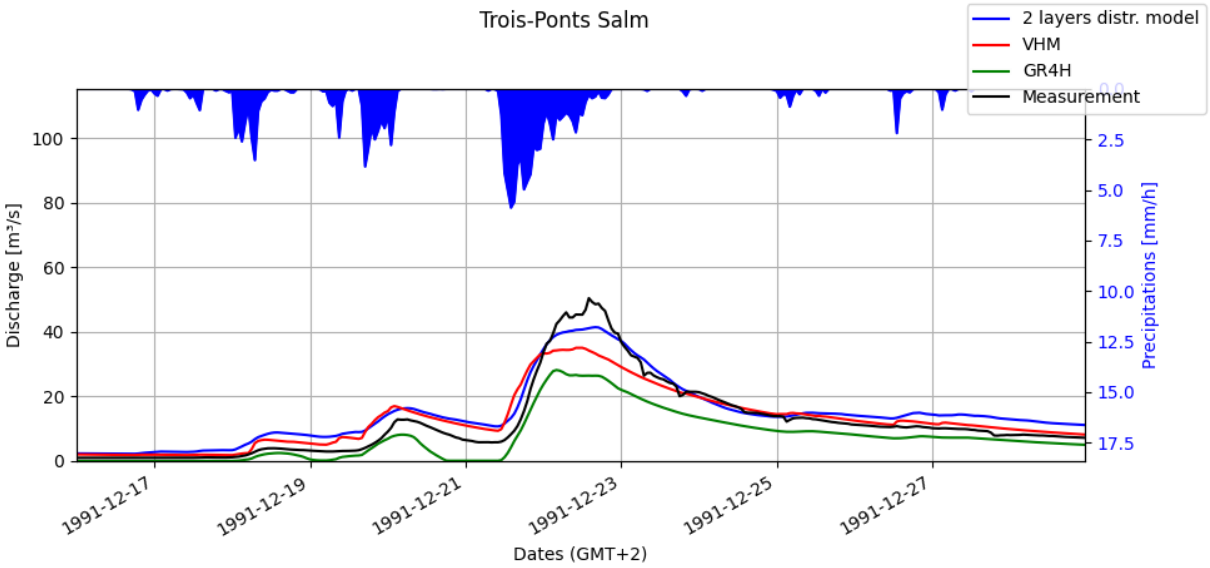


Figure F.1 Hydrographs of the models calibrated with July 2021 and applied to historical floods at Trois-Ponts

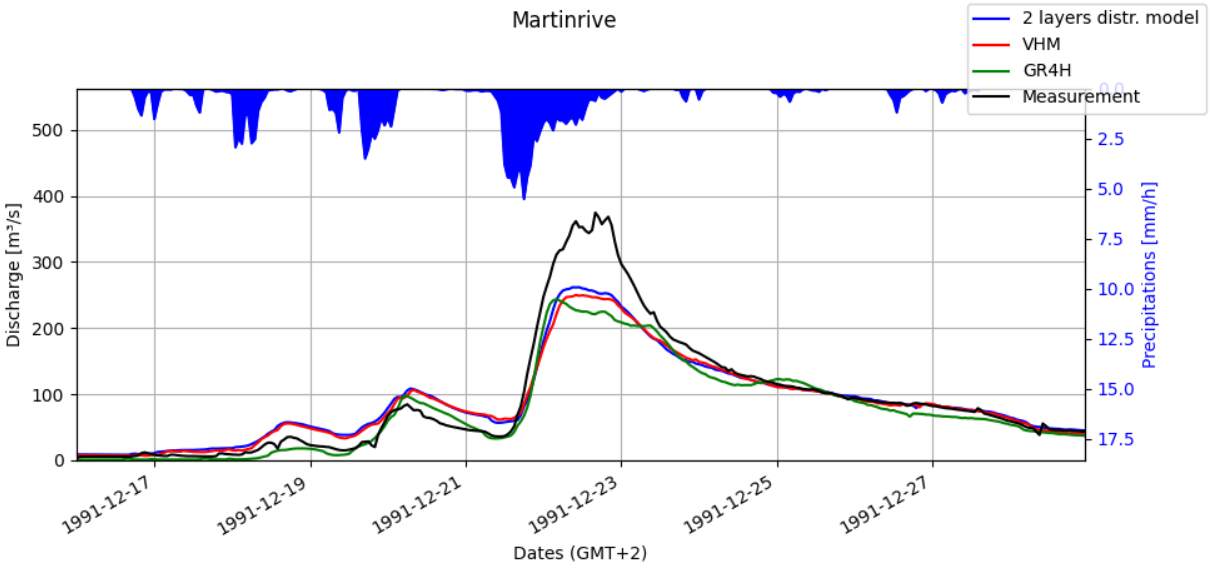


Figure F.2 Hydrographs of the models calibrated on July 2021 and applied to historical floods at Martinrive

Appendix F.2: 1993-1994

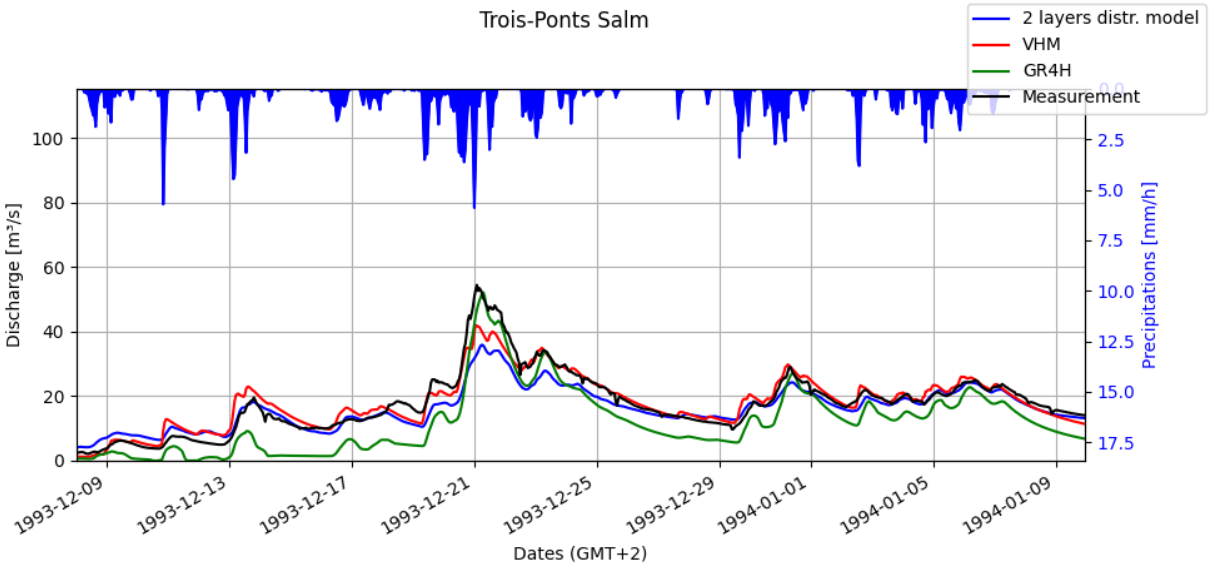


Figure F.3 Hydrographs of the models calibrated with July 2021 and applied to historical floods at Trois-Ponts

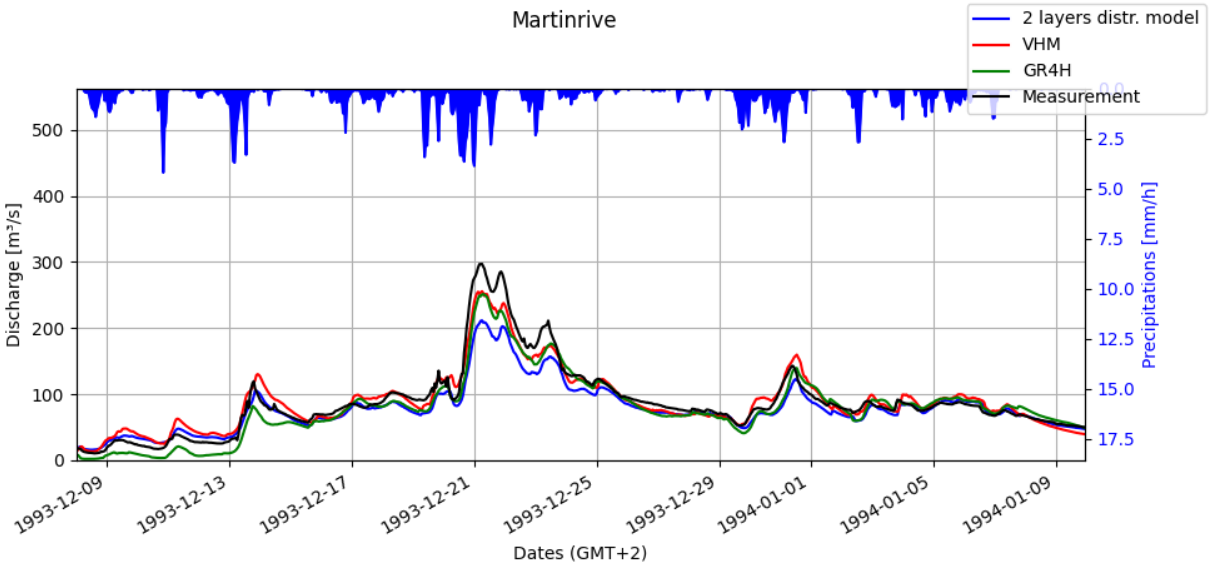


Figure F.4 Hydrographs of the models calibrated on July 2021 and applied to historical floods at Martinrive

Appendix F.3 : 1995

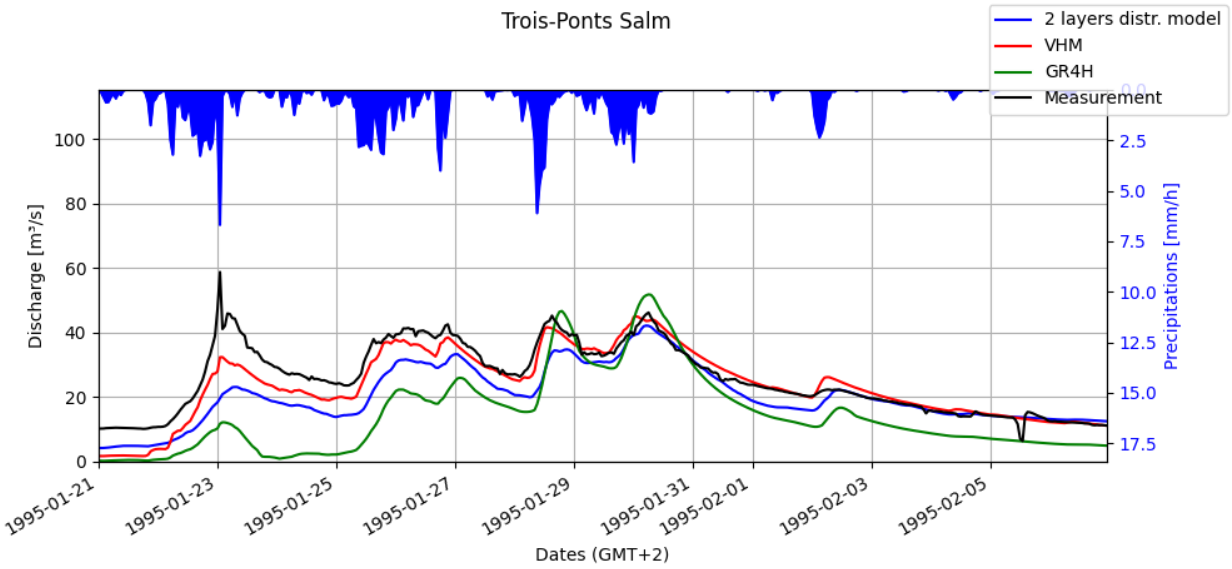


Figure F.5 Hydrographs of the models calibrated with July 2021 and applied to historical floods at Trois-Ponts

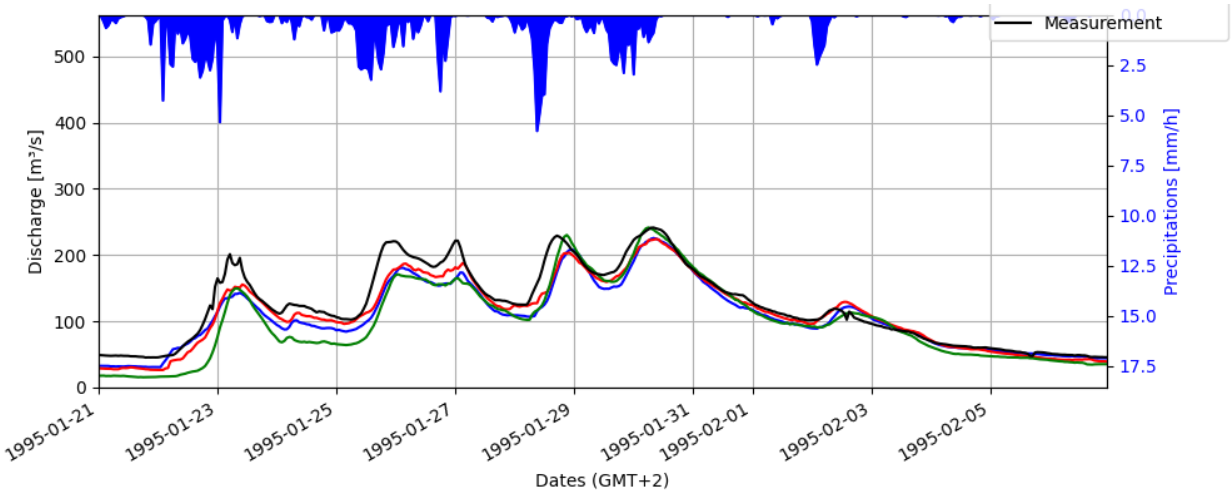


Figure F.6 Hydrographs of the models calibrated with July 2021 and applied to historical floods at Martinrive

Appendix F.4: 1998

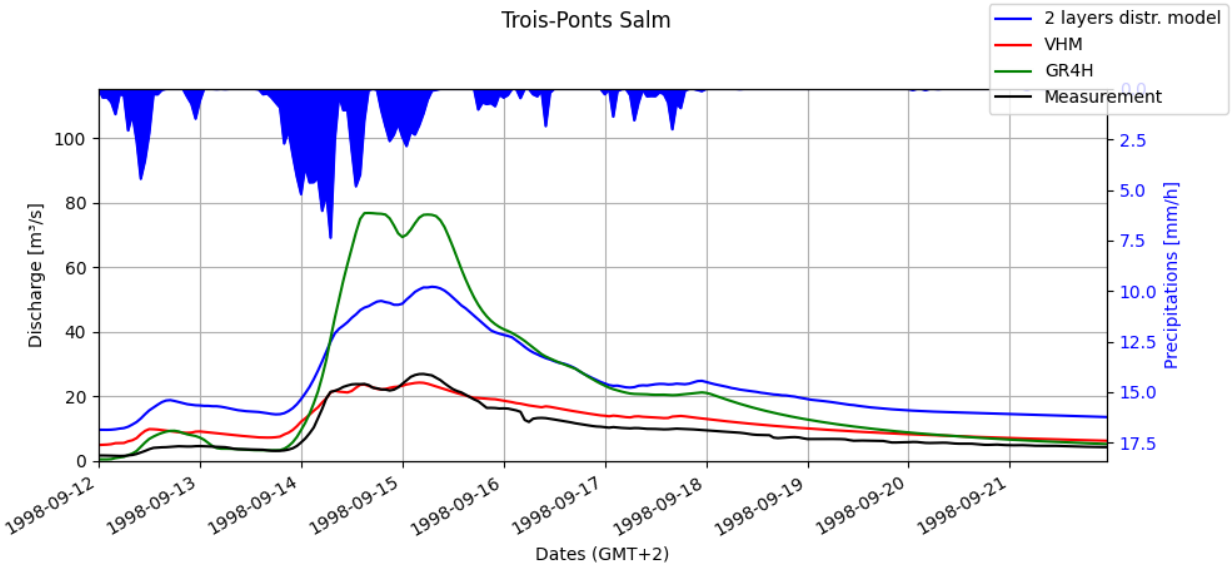


Figure F.7 Hydrographs of the models calibrated with July 2021 and applied to historical floods at Trois-Ponts

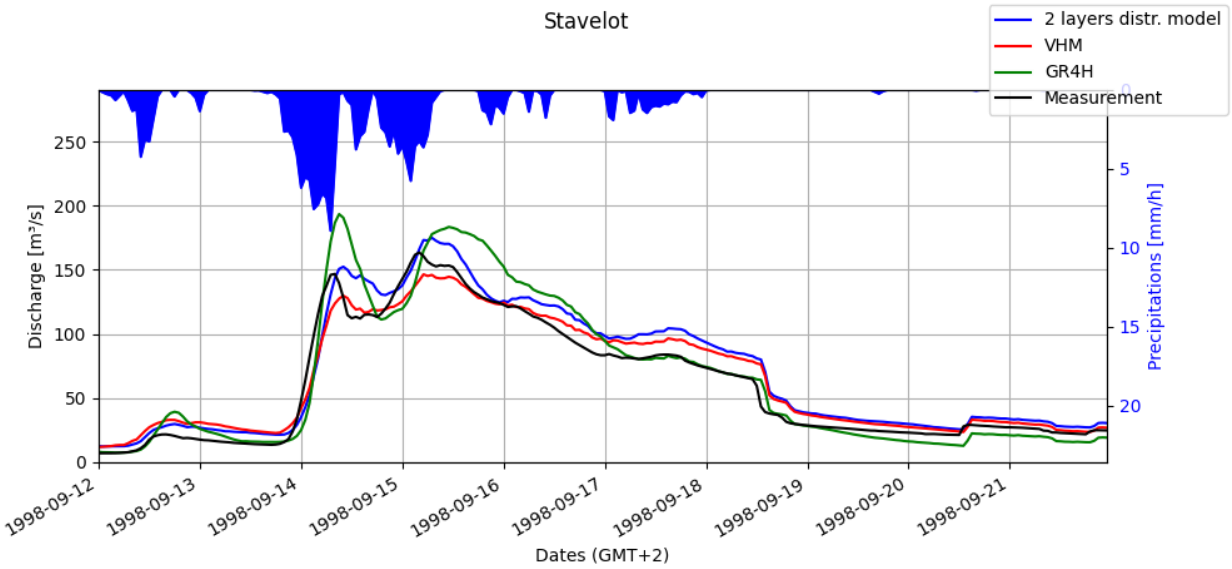


Figure F.8 Hydrographs of the models calibrated with July 2021 and applied to historical floods at Stavelot

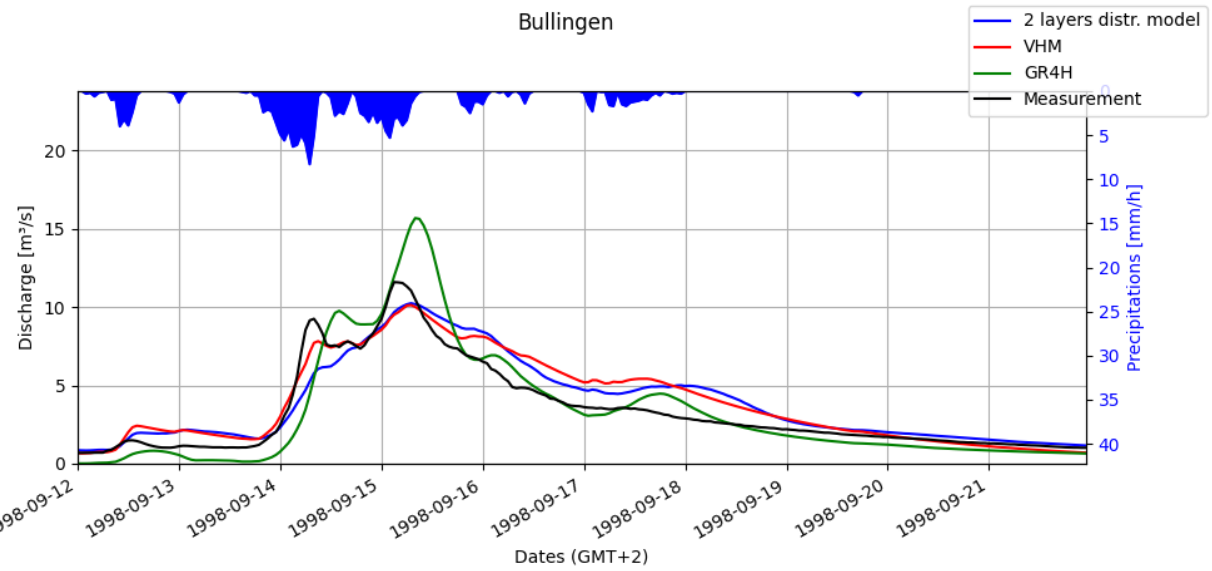


Figure F.9 Hydrographs of the models calibrated with July 2021 and applied to historical floods at Bullingen

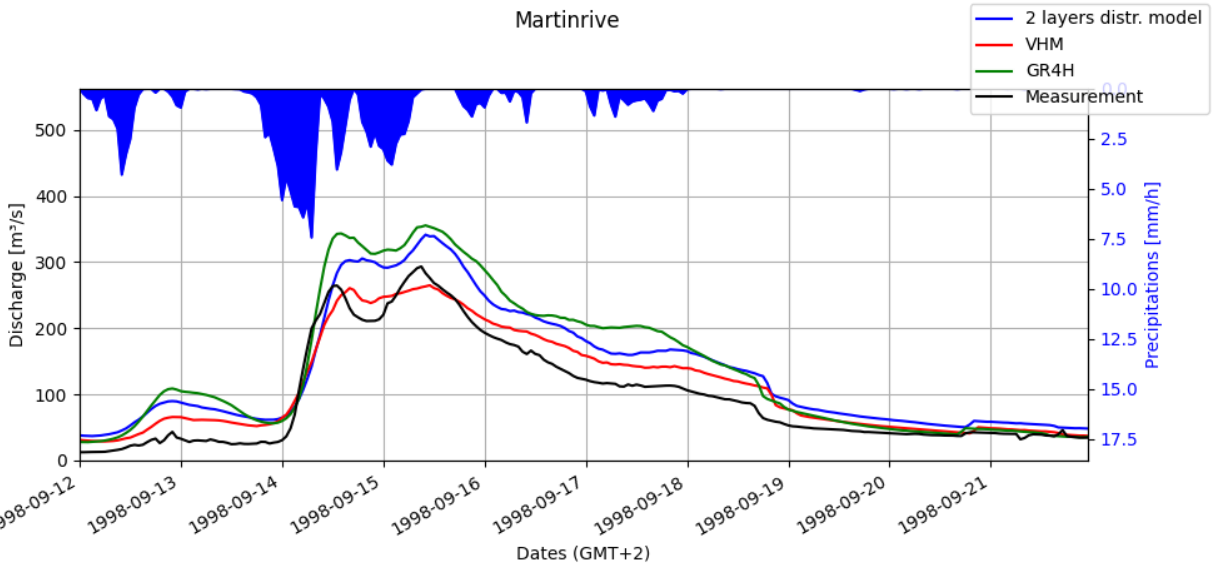


Figure F.10 Hydrographs of the models calibrated with July 2021 and applied to historical floods at Martinrive

Appendix F.5 : 2002

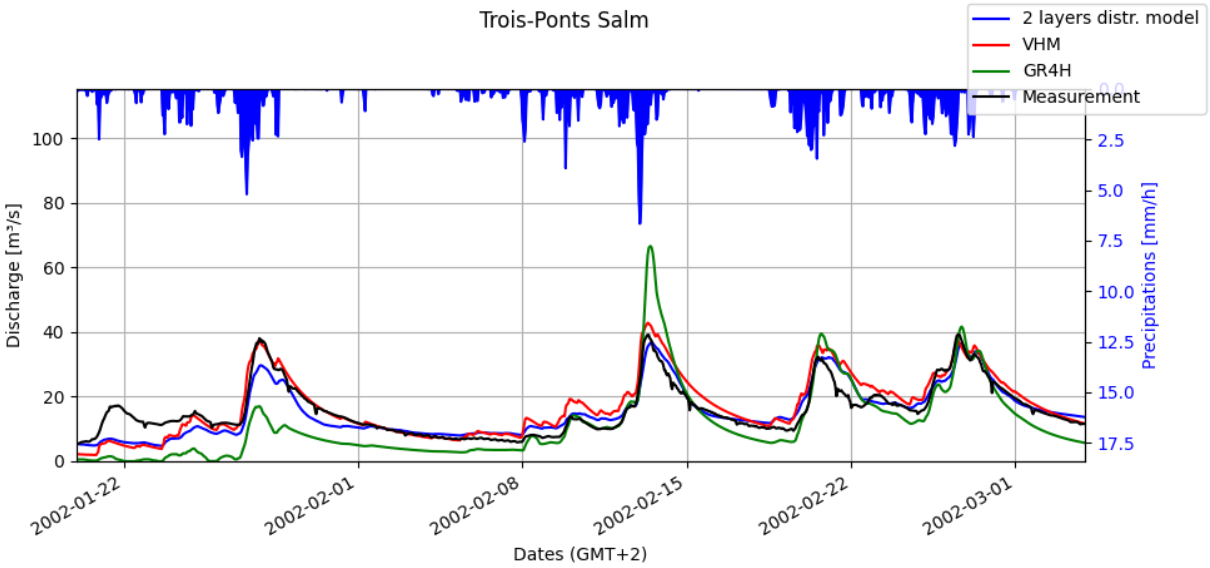


Figure F.11 Hydrographs of the models calibrated with July 2021 and applied to historical floods at Trois-Ponts

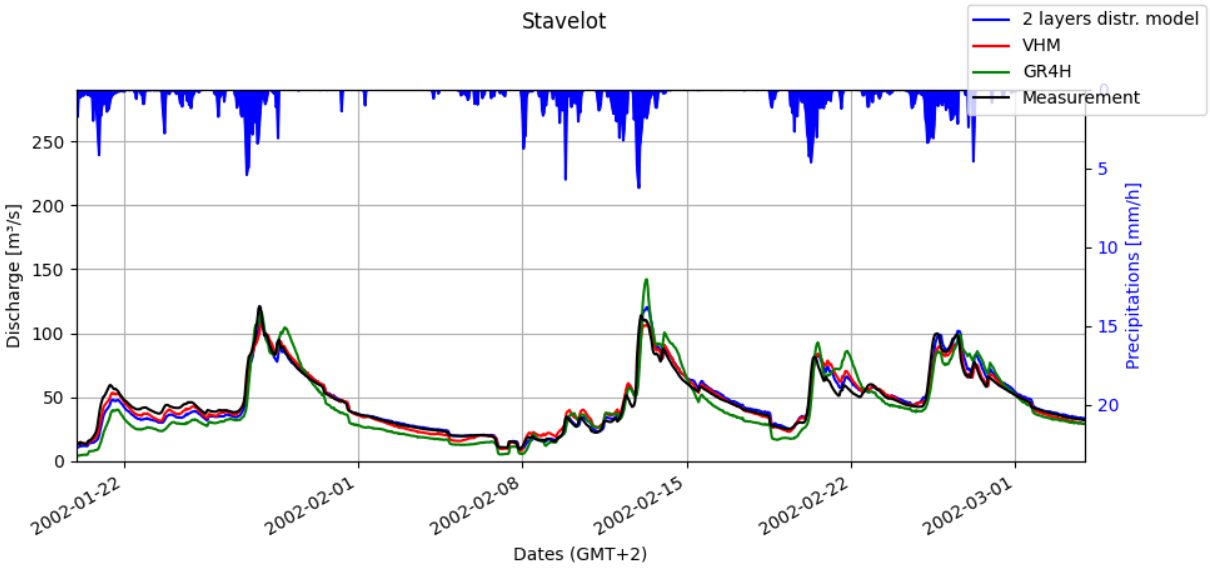
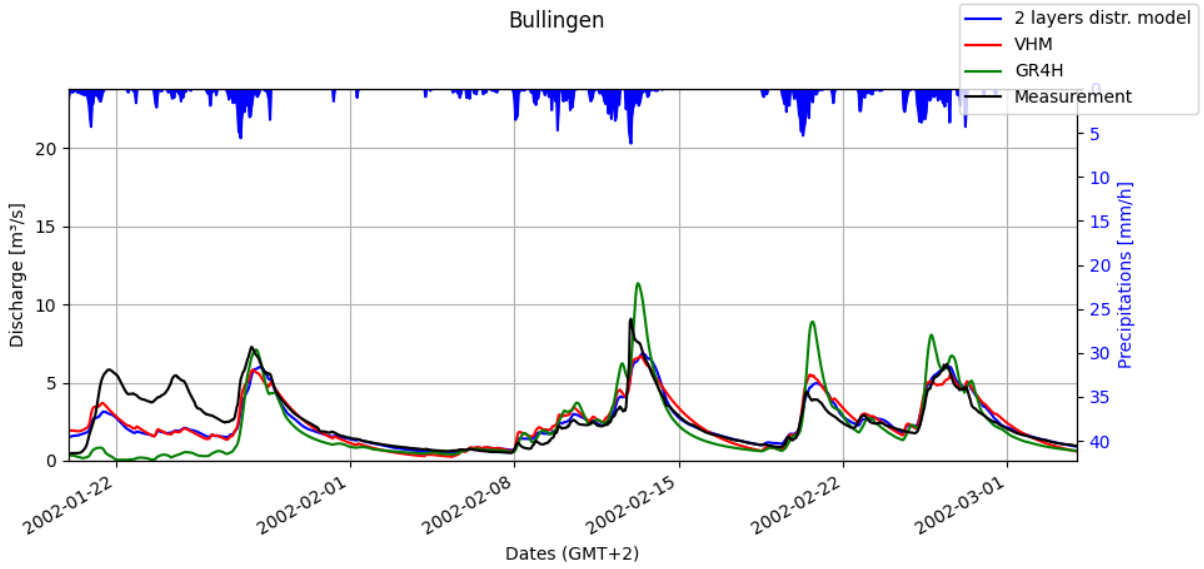


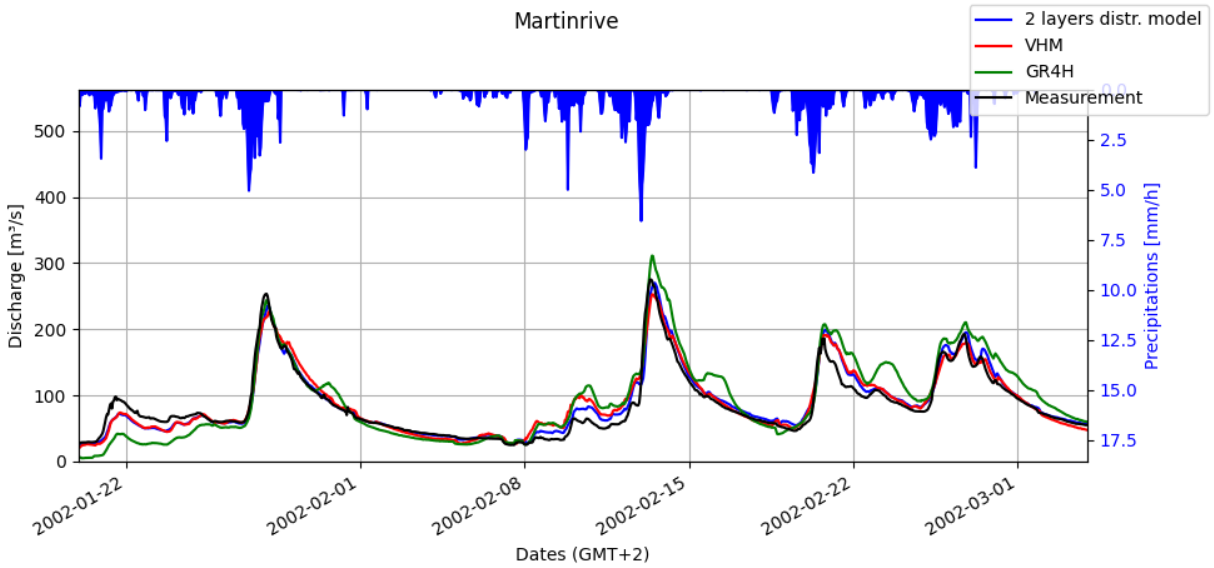
Figure F.12 Hydrographs of the models calibrated with July 2021 and applied to historical floods at Stavelot



1211

1212 Figure F.13 Hydrographs of the models calibrated with July 2021 and applied to historical floods at
1213 Bullingen

1214



1215

1216 Figure F.14 Hydrographs of the models calibrated with July 2021 and applied to historical floods at
1217 Martinrive

1218

Appendix F.6 : 2011

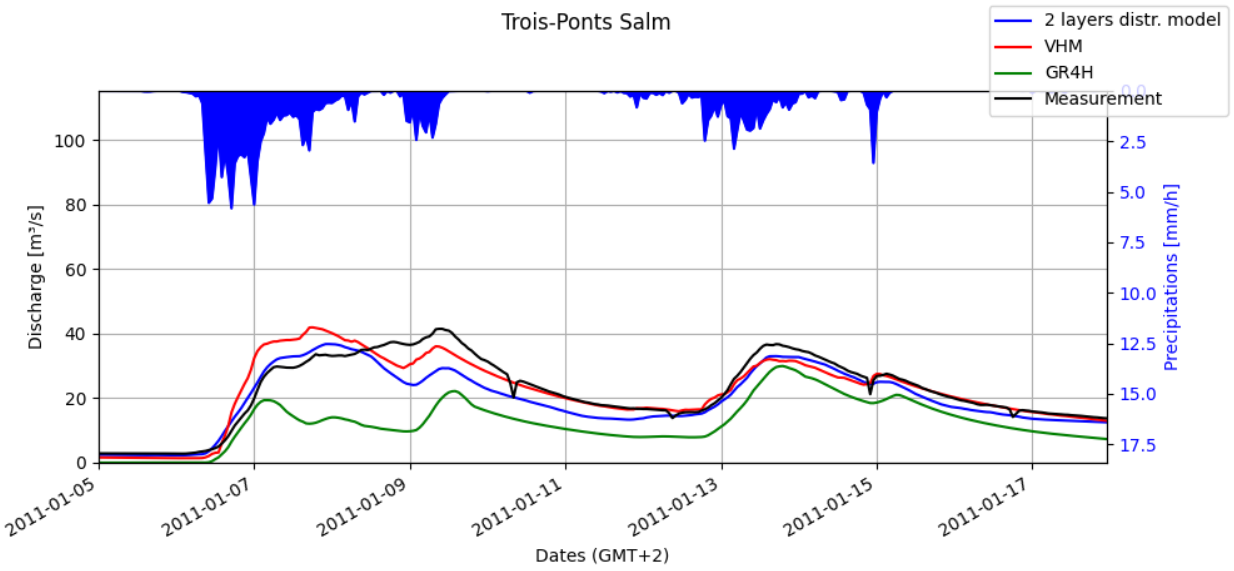


Figure F.15 Hydrographs of the models calibrated with July 2021 and applied to historical floods at Stavelot

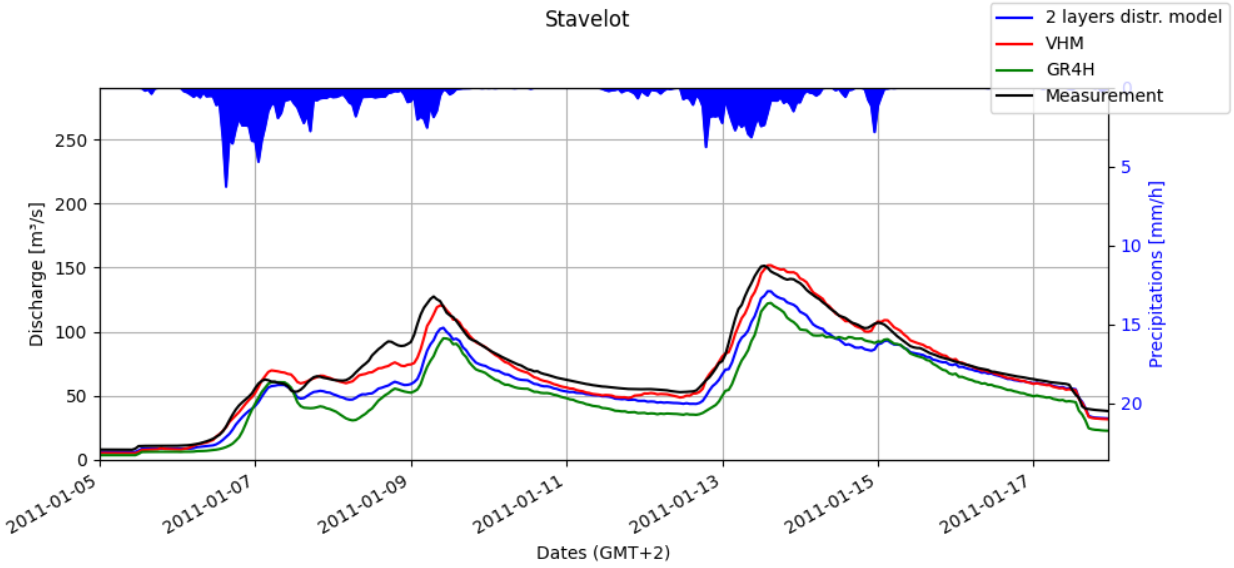


Figure F.16 Hydrographs of the models calibrated with July 2021 and applied to historical floods at Stavelot

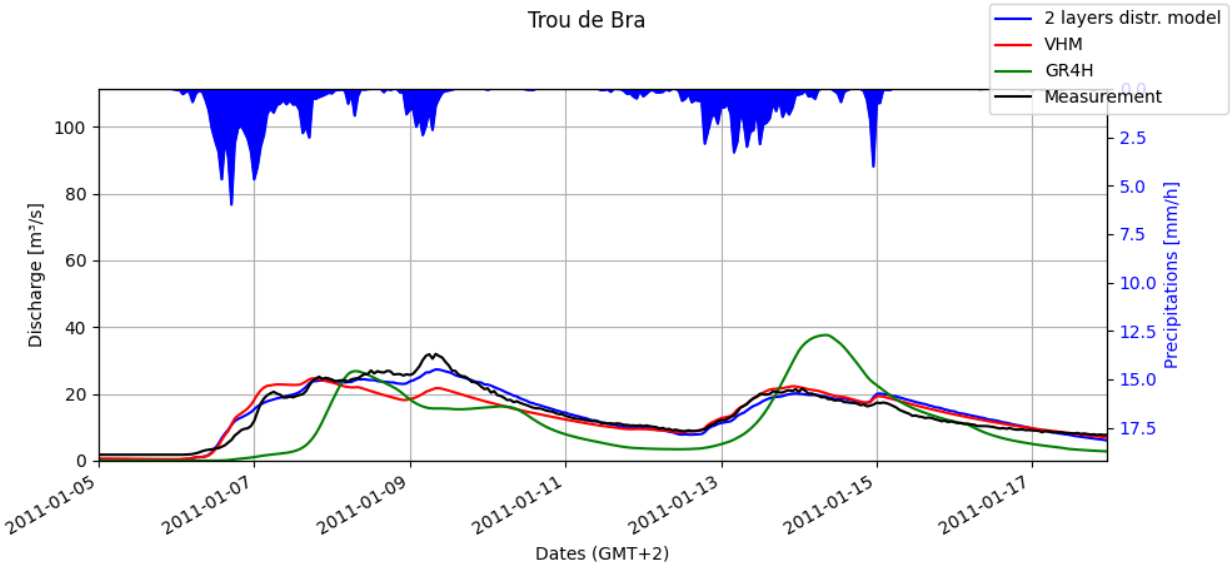


Figure F.17 Hydrographs of the models calibrated with July 2021 and applied to historical floods at Trou de Bra

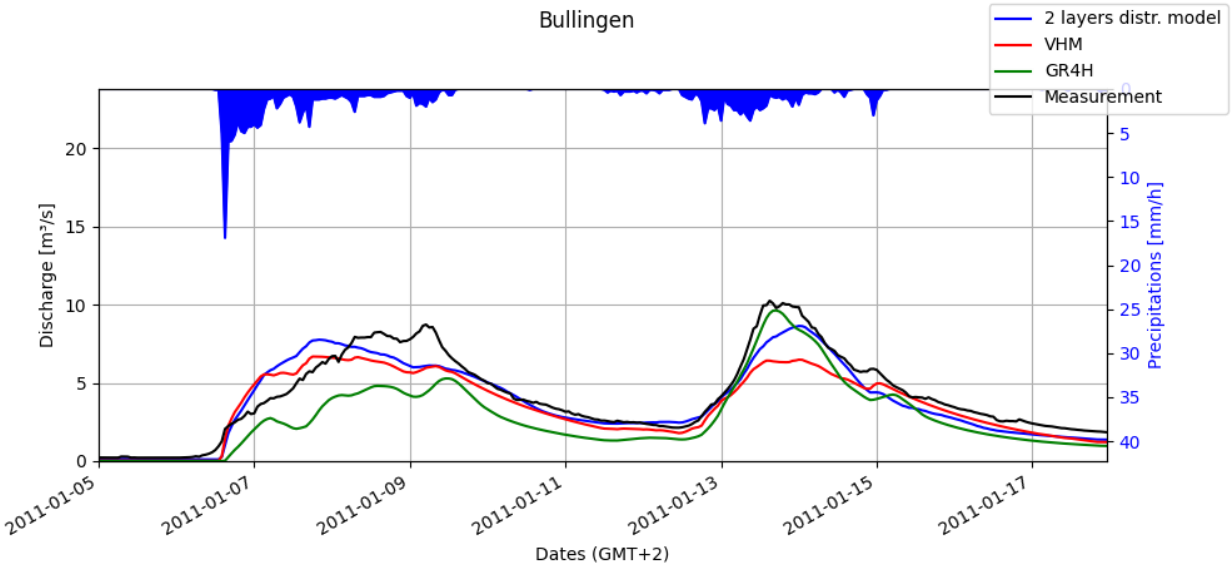


Figure F.18 Hydrographs of the models calibrated with July 2021 and applied to historical floods at Bullingen

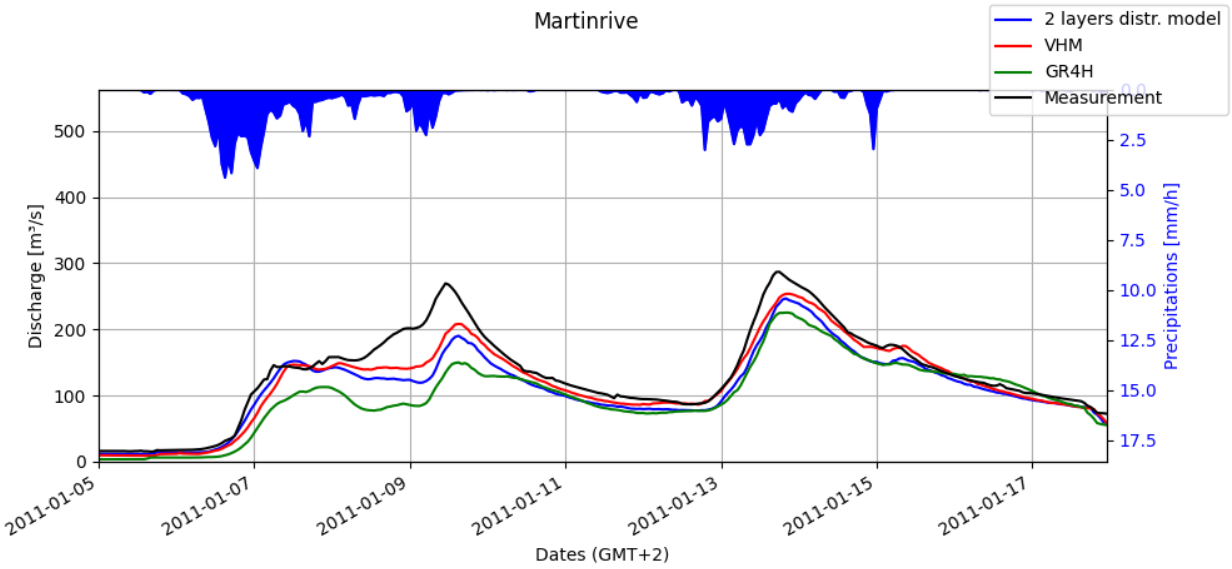


Figure F.19 Hydrographs of the models calibrated with July 2021 and applied to historical floods at Martinrive

Appendix G: Hydrographs Amblève Historical → 2021

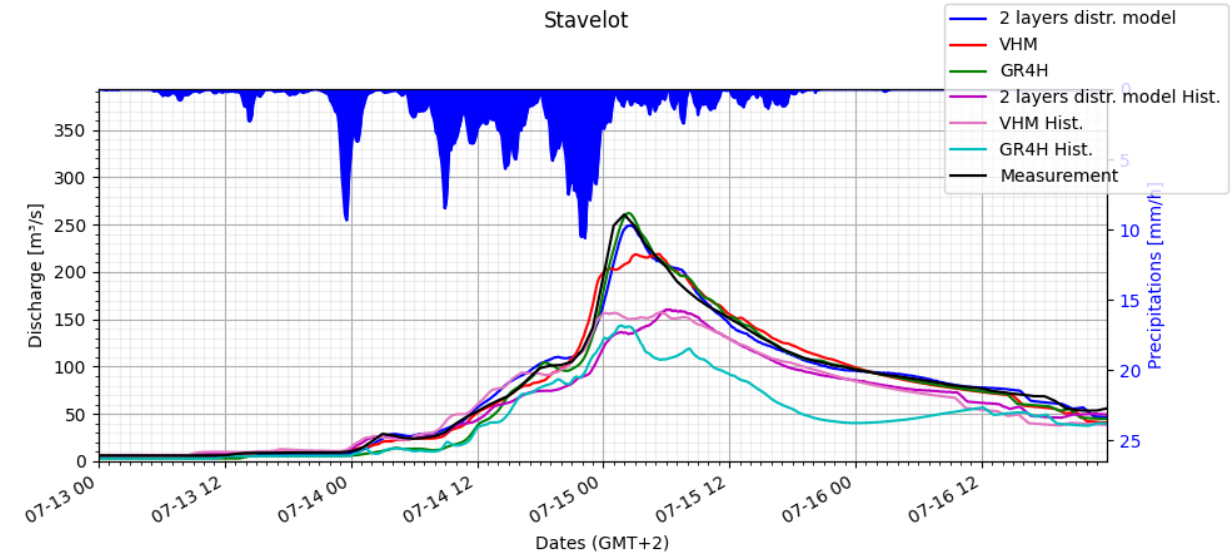


Figure G.1 Hydrographs of the models calibrated with historical floods and applied on July 2021 at Stavelot

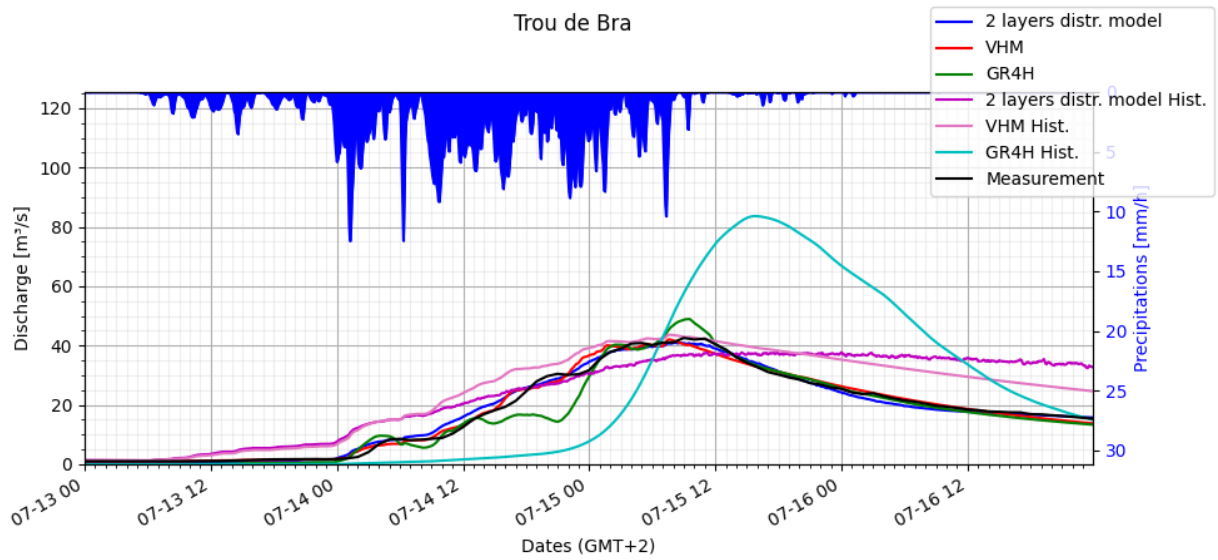


Figure G.2 Hydrographs of the models calibrated with historical floods and applied on July 2021 at Trou de Bra

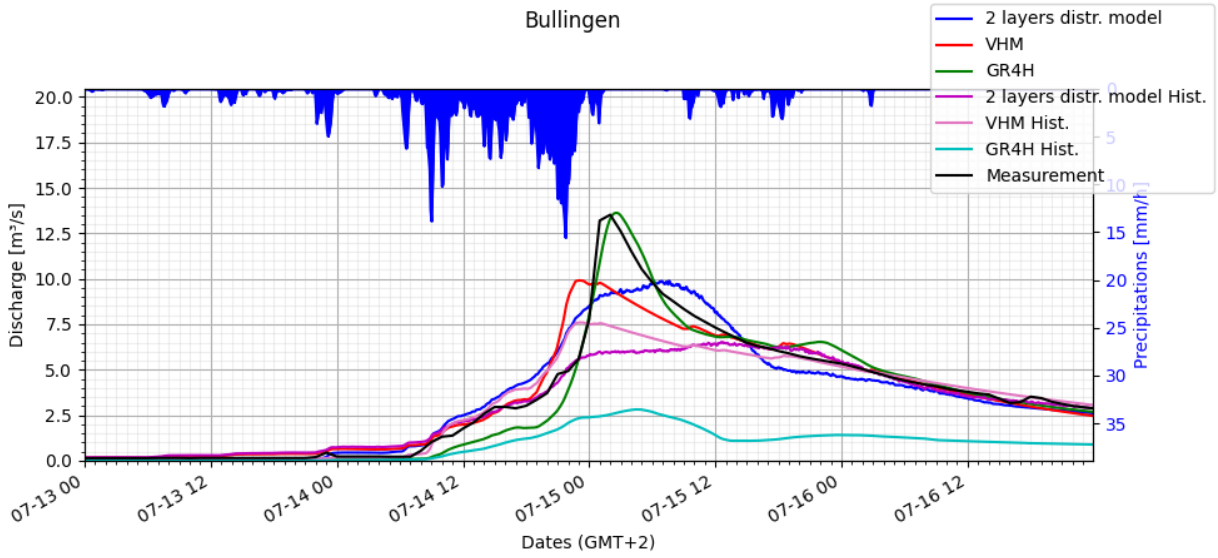
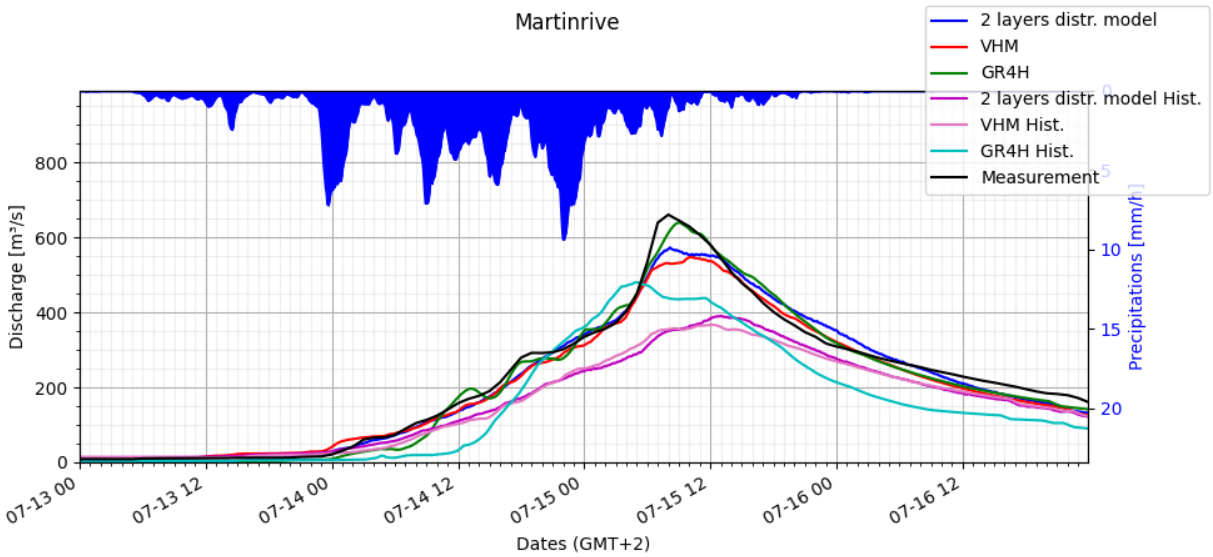


Figure G.3 Hydrographs of the models calibrated with historical floods and applied on July 2021 at Bullingen



1257

1258 Figure G.4 Hydrographs of the models calibrated with historical floods and applied on July 2021 at
1259 Martinrive

1260