Catchment response to intense rainfall: Evaluating modelling hypotheses

Paul C. Astagneau, François Bourgin, Vazken Andréassian, Charles Perrin

Received: 24 February 2022 | Revised: 25 July 2022 | Accepted: 31 July 2022

DOI: 10.1002/hyp.14676

WILEY

TOWARDS MORE CREDIBLE MODELS IN CATCHMENT HYDROLOGY TO ENHANCE HYDROLOGICAL PROCESS UNDERSTANDING

Catchment response to intense rainfall: Evaluating modelling hypotheses

Paul C. Astagneau | François Bourgin | Vazken Andréassian | Charles Perrin

Université Paris-Saclay, INRAE, HYCAR research unit, Antony, France

Correspondence
Paul C. Astagneau, Université Paris-Saclay, INRAE, HYCAR Research Unit, Antony, France.
Email: paul.astagneau@inrae.fr

Funding information
Ministère de la Transition écologique et Solidaire; French Ministry of Environment, Grant/Award Number: DGPR/SNRH/SCHAPI

Abstract

Floods resulting from high-intensity rainfall events are known to be difficult to simulate. The catchment response to such events is very heterogeneous due to complex combinations of hydrological processes at fine temporal and spatial scales. The goal of this study is to find a way to inform the structure of a hydrological model on the variability in catchment response to rainfall events. To that end, rainfall intensity was used as a proxy for the activation of fast heterogeneous runoff processes. We developed three hypotheses that increase the versatility of a hydrological model by modifying the volume and temporal distribution of effective rainfall when high-intensity rainfall events occur. These modifications were implemented within the GR5H lumped rainfall–runoff model. The different model versions were run on 229 French catchments where 10 652 flood events were selected. Model performance was assessed considering five groups of catchments, and model performance was also evaluated based on three event characteristics. Results showed that introducing a dynamic dependency of fluxes to rainfall intensities at the hourly time step helps to improve the simulation of floods, especially on Mediterranean catchment areas. Generic values of the additional parameters are proposed to limit the increase in calibration complexity.

1 | INTRODUCTION

1.1 | High-intensity rainfall events and heterogeneous processes

Floods caused by high-intensity rainfall events are difficult to predict. In temperate climates, they occur most frequently in the summer or at the beginning of autumn, when soils are relatively dry, causing non-linear streamflow responses (Chappell et al., 2017). It is common to read that infiltration-excess runoff processes (or Hortonian runoff; Horton, 1933) play a major role in the generation of these events (e.g. Lana-Renault et al., 2007; Latron & Gallart, 2008; Manus et al., 2009). However, some studies report saturation mechanisms (e.g. Braud et al., 2014; Estrany et al., 2010) and contributions of subsurface storages to streamflow (e.g. Chappell et al., 2017; Hugenschmidt et al., 2014) to be active even for very intense rainfall events in summer. Overall, there is an interplay of varied flood-generating processes that are scale-dependent (e.g. Blöschl & Sivapalan, 1995; McDonnell et al., 2021), such as local soil saturation (Garambois et al., 2014) and threshold in the connectivity between catchment storages and the stream network through lateral pipe flow processes from the hillslopes (e.g. Uchida et al., 2005; van Meerveld & McDonnell, 2006). Last, these floods are often characterized by short response times to rainfall events making them even harder to predict since fewer data are available for parameter estimation and evaluation of simulations (e.g. Perrin et al., 2007).

1.2 | On the need to improve model structures

Hydrological models have poorer performance on catchments where these floods occur, especially when there is a strong seasonality of antecedent soil moisture conditions (e.g. McMillan et al., 2016; Melsen et al., 2018). This is mainly due to the non-linearity of the processes involved in the catchment response that hydrological models—which are simplifications in time, space, and processes (e.g. Hrachowitz & Clark, 2017)—struggle to characterize. For example, Astagneau, Bourgin, et al. (2021) showed that the GR5H model (Ficchì et al., 2019; Le Moine, 2008) tends to underestimate flood volumes in summer especially when high-intensity rainfall events occur. Recent advances in hydrological modelling have mostly been achieved through the development of better parameterization procedures (e.g. de Lavenne et al., 2019; Mizukami et al., 2019; Moussa & Chahinian, 2009; Pool et al., 2021), the use of additional data to streamflow (e.g. Bouaziz et al., 2021; Nijzink et al., 2018; Rakovec et al., 2016), or the development of small headwater catchment models guided by an improved understanding of underlying hydrological processes, e.g., through isotopic measurements (e.g. Birkel et al., 2014; Kuppel et al., 2018). However, there is still room for model structural improvements and development of other perceptual models at the meso- (100–1000 km^2) and regional scales (>10 000 km^2; for example, Fenicia & McDonnell, 2022; Fenicia et al., 2022), since model structure imperfections are one of the main sources of uncertainty in streamflow prediction (e.g. Clark et al., 2008; Knoben et al., 2020). In particular, there is a need for modelling the contribution of highly heterogeneous processes taking place at fine spatiotemporal scales to river flows at larger scales (Blöschl et al., 2019).

1.3 | Model improvements from better diagnostics

In recent years, model improvement has also been achieved through the use of better evaluation and diagnostic frameworks. In this respect, several studies used hydrological signatures in order to assess model "realism," adequacy, and performance and therefore go beyond the simple evaluation of models by aggregated statistics (e.g. Gupta et al., 2008; McMillan et al., 2017; Yilmaz et al., 2008). For example, de Boer-Euser et al. (2017) compared the performance of 11 models on the Meuse basin and found differences in their ability to reproduce several signatures (e.g. signatures describing flashy dynamics) while yielding similar overall performance. Euser et al. (2013) proposed a multi-objective evaluation framework to assess the ability of a hydrological model to simulate different hydrological signatures emphasizing different characteristics of the hydrograph and therefore different processes. Another way to evaluate models is to look for dominant parameters for a given catchment in order to identify simulated dominant processes and then link the results with hydroclimatic conditions (e.g. Herman et al., 2013). These frameworks enable multi-hypothesis testing (e.g. Blöschl, 2017; Clark et al., 2011) and therefore allow us to evaluate the value of making a model more complex for a given catchment (e.g. Hrachowitz et al., 2014). However, identifying situations where the modification of a model component improves streamflow simulation and model robustness remains a challenging task (van Esse et al., 2013).

1.4 | Fast catchment response in models

When computing runoff production, modellers commonly define several flow components. This can be achieved by setting a partitioning coefficient from a single production store outflow, which can be a fixed parameter as in the GR models (Perrin et al., 2003) or a calibrated parameter as in the IHACRES model (Jakeman & Hornberger, 1993). Flow partitioning can also be applied to multiple runoff production store outflows, such as in the MORDOR model (Garavaglia et al., 2017; Garçon, 1996). In some versions of the HBV model (Bergström, 1976; Parajka et al., 2007; Seibert & Vis, 2012) and in the FLEX-Topo model (Savenije, 2010), a "very-fast" runoff component is calculated from the upper routing store when it reaches a certain threshold.

Some hydrological models explicitly simulate infiltration-excess overland flow. Their formulation is often based on the Green–Ampt infiltration equation (Green & Ampt, 1911) or one of its extensions (Beven, 2021) to estimate the soil infiltration capacity rate. Then, if the intensity rate is higher than the infiltration capacity rate, ponding starts and surface runoff is generated. This is the case for some versions of TOPMODEL (Beven et al., 2021; Beven & Kirby, 1979) where soil infiltration is based on an extension of the Green–Ampt equation (Beven, 1984). Some hydrological models, like the SWAT model (Arnold et al., 2012), include the SCS-CN method (SCS, 1956) to estimate surface runoff. Several studies have focused on modifying the SCS-CN formulation to improve streamflow simulation (e.g. Mishra et al., 2004; Verma et al., 2020, 2021). For example, Pang et al. (2020) modified the SCS-CN to take slopes and precipitation intensity into account. They found that their new formulation performed better than their reference model in simulating floods of the Chao River watershed. In the JULES land-surface model (Best et al., 2011), a local surface runoff component is calculated when the throughfall rate becomes higher than the rate of maximum infiltration. Largeron et al. (2018) showed that calculating a variable maximum infiltration rate was not sufficient to overcome model deficiencies in the case of flooding from intense rainfall. Beven (2021) argues that while we can make local predictions of infiltration-excess runoff, it is not possible to estimate its contribution to the flood hydrograph at larger scales. One way to deal with this issue is to increase model temporal and spatial resolutions with "exhaustive" process description, but these models suffer from overparameterization problems (e.g. Andréassian et al., 2012; Beven, 1993), especially when no other data than streamflow are available for model validation.

Several conceptual rainfall–runoff models use dedicated functions to simulate more runoff when high-intensity rainfall events occur, even when the soil moisture content is low. For example, the MOD-SUR model (Ledoux et al., 1989) has a parameter controlling the threshold at which the production function generates runoff for low

antecedent soil moisture conditions. This parameter depends on the soil type. Another attempt at simulating infiltration-excess runoff in a conceptual way was made by Willems (2014). The conceptual infiltration equation was changed to simulate more runoff in the case of high rainfall intensities and wetness conditions without, however, making the surface runoff coefficient dependent on rainfall intensities. Recently, Peredo et al. (2022) introduced a function in the semi-distributed version of the GR5H model (Lobligeois et al., 2014) to generate more effective rainfall in the case of high-intensity rainfall events under low antecedent soil moisture conditions, and obtained some improvements in the simulation of autumn floods of the Aude River without degrading the simulation of winter floods.

1.5 | Scope of the paper

We need to improve the ability of hydrological models to characterize the catchment response to intense rainfall, especially under low antecedent soil moisture conditions and at scales where field work experiments are not always available for model validation. We have seen that several issues arise with the current perceptual models applied at the catchment scale when only streamflow time series are available.

The objective of this study is to evaluate the benefit of making a better use of rainfall intensity to inform the structure of a hydrological model on catchment response to intense rainfall events. Three questions arise from our objective:

1. Which fluxes of a lumped rainfall–runoff model should vary with rainfall intensity at the event and catchment scale?
2. Can we identify catchments and hydro-meteorological conditions for which rainfall intensity rate functions significantly improve flood simulation?
3. Is there enough information in rainfall and streamflow time series to derive stable-in-time parameter values associated with the new modelling functions?

To address these questions, we developed a large number of modelling hypotheses aimed at dynamically modifying the storages and fluxes of a lumped rainfall–runoff model using rainfall intensity rates. Three of these hypothesis are discussed in detail in this article. We first present the data used to evaluate our hypotheses. Then, we introduce the methodological basis behind our hypotheses and the framework to evaluate performance and adequacy in regard to our objectives. The main results are then presented, analysed, and discussed considering seasonality and catchment hydroclimatic conditions before summarizing the main conclusions of our research.

2 | DATA AND METHODS

2.1 | Catchment set and hydroclimatic data

Our catchment set originates from the work of Ficchì et al. (2016) and Astagneau, Bourgin, et al. (2021). It consists of 229 catchments across metropolitan France, representing a variety of hydroclimatic conditions, where human activities and snow processes have limited impacts on river flows. The hydroclimatic and morphological characteristics of this dataset can be found in Table 1 of Astagneau, Bourgin, et al. (2021). Some of these characteristics are reported in Table 1 and in Figures 1 and A1. Rainfall inputs were retrieved from the Comephore reanalysis (Tabary et al., 2012), which provides data at a 1-km resolution, and are aggregated at the catchment scale. Potential evaporation (PE) daily time series (TS) were calculated with the Oudin et al. (2005) formula using daily temperature TS from the SAFRAN product (Delaigue et al., 2020; Vidal et al., 2010). They were then disaggregated at the hourly time step by applying a parabolic distribution between 6:00 AM and 7:00 PM (UTC). Instantaneous streamflow TS were retrieved from the French hydrometric database (Leleu et al., 2014) and interpolated at the hourly time step. A total of 19 years of hourly meteorological TS and between 8 and 19 years of hourly streamflow TS were available for these catchments. Two independent sub-periods of 9.5 years each were defined. The first period (P1) runs from 1 January 2000 to 30 June 2009 and the second period (P2) runs from 1 July 2009 to 31 December 2018. In both sub-periods, an automated procedure was used to select flood events. Peak flows higher than the 95th streamflow quantile were selected. Then, the start of an event was defined as the time at which streamflow exceeds 20% of the flood peak. The end of an event was defined as the time at which streamflow is inferior to 25% of the flood peak. This selection led to a set of 10 652 flood events. Corresponding rainfall events were selected by shifting the streamflow event time window by the catchment lag time (T_L), which was estimated as the time shift that maximizes the cross-correlation between streamflow and rainfall TS. Each event was visually inspected to mitigate errors arising from the automated procedure.

2.2 | Catchment clustering

Aiming to investigate the links between hydroclimatic properties and flood-generating processes, we divided our catchment set into five groups by performing a clustering procedure. We used four indicators based on the selected rainfall and streamflow events and five indicators based on catchment hydroclimatic properties (see Table 2).

Two indices were calculated to characterize rainfall events: the mean rainfall intensity, *intP*, and the spatial variability in cumulative rainfall, *spatP* (or "spatial coefficient of variation in precipitation event volume"; Tarasova et al., 2020). They were calculated as follows:

$$\mathrm{int}P = \frac{1}{n_t}\sum_{t=1}^{n_t}\left[\mu(P_{x,t})_{1\le x\le n_x}\right], \tag{1}$$

$$\mathrm{spat}P = \frac{\sigma\left(\sum_{t=1}^{t=n_t} P_{x,t}\right)_{1\le x\le n_x}}{\sum_{t=1}^{n_t}\left[\mu(P_{x,t})_{1\le x\le n_x}\right]} \tag{2}$$

in which $P_{x,t}$ is the rainfall of grid cell x at time t, n_t is the duration of the rainfall event (in time steps), n_x is the number of grid cells of the

Characteristic	Min	Q25	Median	Mean	Q75	Max
Area (km^2)	3.54	164.1	354.1	680.3	772.3	7918
Mean flow (Q_m) (mm/year)	35	262	349	437	524	1398
Mean annual temperature (°C)	8.2	9.8	10.4	10.6	11.1	14.3
Mean annual precipitation (P_m) (mm)	651	818	937	990	1097	2108

TABLE 1 Basic characteristics of the catchment dataset. Modified from Astagneau, Bourgin, et al. (2021)

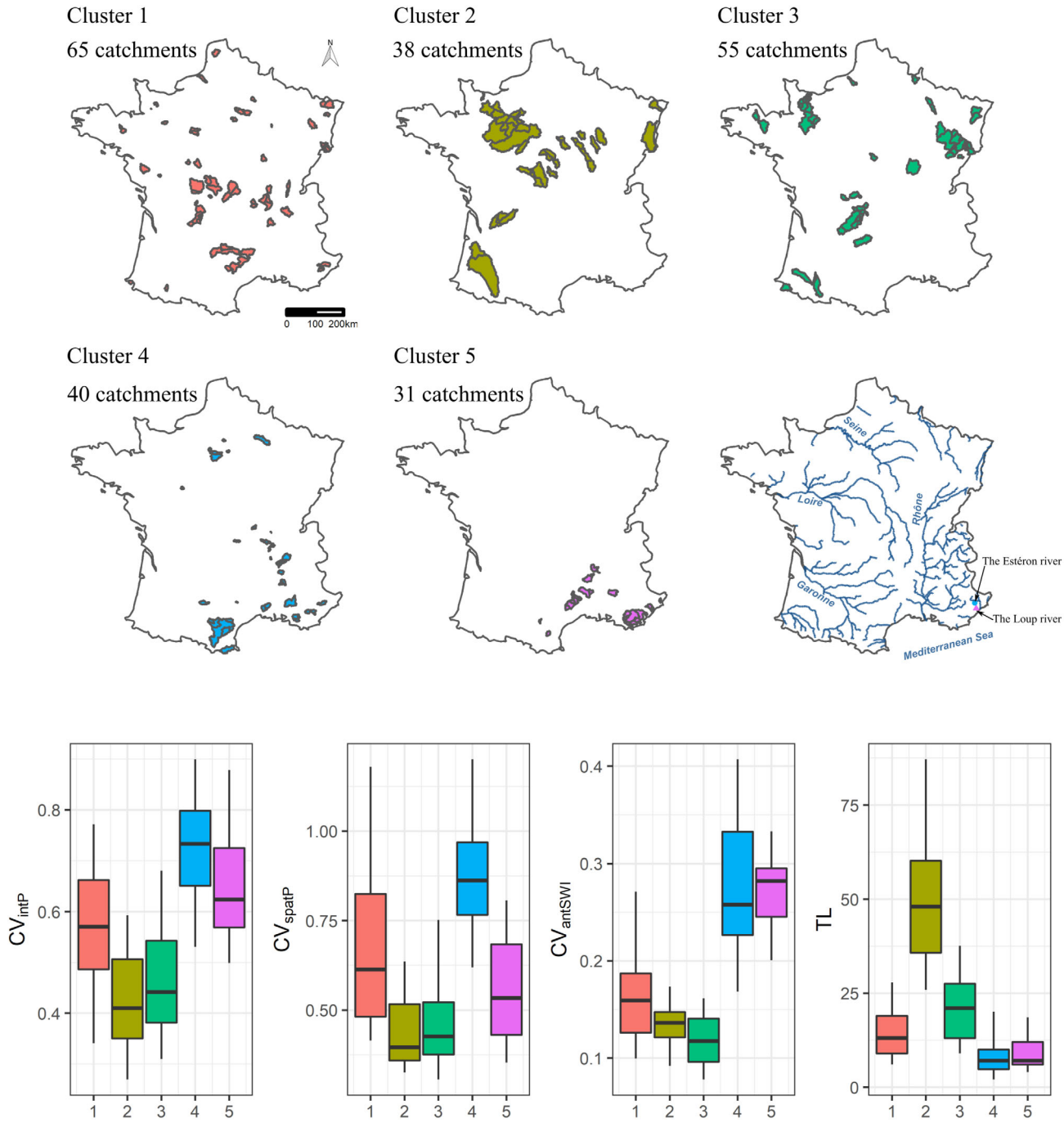

FIGURE 1 Location of the five hydroclimatic groups of catchments and distributions of four characteristics in these groups. The distributions are presented between the 5th and 95th percentiles

catchment (based on a 100-m resolution flow direction grid), μ is the mean, and σ is the standard deviation. Large values of *intP* indicate that there was a large amount of rainfall in a short period. Large values of *spatP* indicate that a large proportion of the rainfall event occurred in a small portion of the catchment. Antecedent soil moisture conditions can have a significant influence on flood-generating processes

TABLE 2 Event and catchment characteristics used for catchment clustering

Characteristic	Type	Definition/reference	Abbreviation
Variability in event rainfall intensity (–)	Event-based	Equation (1)	CV_{intP}
Variability in event cumulative rainfall spatial variability (–)	Event-based	Equation (2)	CV_{spatP}
Variability in event antecedent soil wetness condition (–)	Event-based	Coustau et al. (2015)	CV_{antSWI}
Variability in event duration (–)	Event-based	–	$CV_{duration}$
Rainfall–runoff lag time (h)	Hydroclimatic	Ficchì (2017)	TL
Wetness index (–)	Hydroclimatic	P_m/PE_m	W_i
Runoff coefficient (–)	Hydroclimatic	Q_m/P_m	R_c
Magnitude of high flows (–)	Hydroclimatic	Q_{99}/Q_m	Q_{mag}
Flow autocorrelation at 24 h (–)	Hydroclimatic	–	$R_{Q\_24\ h}$
Magnitude of intense rainfall (–)	Hydroclimatic	P_{99}/P_m; Ficchì (2017)	P_{mag}

Note: Q_{99} is the 99th streamflow percentile. P_{99} is the 99th rainfall percentile. Q_m is the mean streamflow. P_m is the mean rainfall. PE_m is the mean potential evapotranspiration.

(Berghuijs et al., 2014; Blöschl et al., 2013). We used the soil wetness index (SWI) calculated by the ISBA surface model (Coustau et al., 2015; Thirel et al., 2010a, 2010b). The SWI was aggregated at the catchment scale and retrieved at the start of rainfall events to attribute one value to each flood event.

Coefficients of variability CVc (i.e. the ratio of standard deviation to mean value of a given event characteristic c) were calculated to associate event characteristics with catchment properties. In other words, for each flood characteristic, a value of its variability between events was attributed to each catchment.

We used unsupervised random forest (RF; Breiman, 2001) and hierarchical clustering algorithms available in the R environment (Liaw & Wiener, 2002; R Core Team, 2021) to classify the catchments according to the aforementioned characteristics (Table 2). RF is a powerful tool for detecting non-linear dependencies between variables and has been used in many hydrological studies (e.g. Saadi et al., 2019; Stein et al., 2021). More details about the use of random forest algorithms for water science applications can be found in Tyralis et al. (2019).

The clustering procedure led to five groups of catchments (Figures 1 and A1). We give here a few comments on their main characteristics:

- Group 1: they have lower values of rainfall and flow magnitude and lower values of variability in flood antecedent wetness conditions than the catchments of groups 4 and 5 but higher values than the catchments of groups 1 and 2.
- Group 2: they have the slowest flow dynamics with the highest rainfall–runoff lag times and the highest values of flow autocorrelation at 24 h.
- Group 3: they have slow flow dynamics but higher rainfall–runoff coefficients and higher values of wetness index than the catchments of group 2.
- Group 4: they are characterized by very different hydroclimatic conditions than the catchments of groups 1 to 3. They have faster flow response to rainfall, higher variability in event rainfall intensity, and higher variability in event antecedent wetness conditions, which indicates a strong flow seasonality. They are mostly located in the Rhône basin and the Mediterranean area. Seven catchments are tributaries of the Seine River. They share most of the characteristics of the other catchments of group 4 except for the magnitude of rainfall, which falls in the lower part of the distribution.
- Group 5: they have lower variability in cumulative rainfall spatial variability. It means that the rainfall fields associated with their flood events can be spatially highly variable, but, if so, this variability does not change much between events. They have higher values of streamflow magnitude than the catchments of group 4. They are all located in the Rhône basin and the Mediterranean area.

2.3 | Hydrological modelling

We made three modelling hypotheses on the possible dependence of the storages and fluxes of the GR5H lumped hydrological model on rainfall intensity rates (Figure 2), starting with a "reference" model version. All versions include an explicit interception store as proposed by Ficchì et al. (2019). The GR5H model has been used in several hydrological modelling studies and usually yields high streamflow performance over large sets of catchments (e.g. Astagneau, Bourgin, et al., 2021; Ficchì et al., 2016, 2019). Its structure may, however, experience difficulties to simulate streamflow response to specific hydrological events, in particular the high-intensity rainfall events.

We first introduce the relevant specificities of the reference model version and then the rationales behind the three modelling hypotheses.

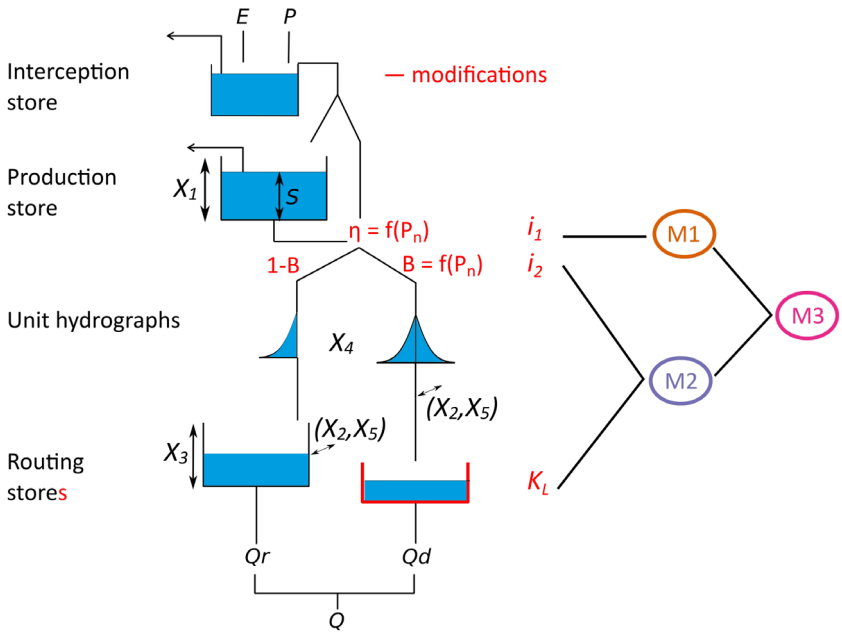

FIGURE 2 Diagram of the modifications (in red) made to the original GR5H model (in black)

2.3.1 | Reference model

The GR5H rainfall–runoff model used in this study consists of an interception store, a production store, two unit hydrographs (UHs), and a routing store. A net rainfall flux (P_n) is calculated from the interception store and either fills the production store or becomes effective rainfall (P_r). The instantaneous production rate (or the fraction of net rainfall that becomes effective rainfall) is calculated as follows:

$$\eta = \left(\frac{S}{X_1}\right)^2 = s^2 = 1 - \frac{P_s}{P_n} = \frac{P_r}{P_n} \quad (3)$$

with S the level of the production store (mm), X_1 the production store capacity (mm), s the production store filling rate, and P_s (mm/h) the rainfall infiltrating (and remaining in) the production store (see Astagneau, Bourgin, et al., 2021, for the complete set of equations of GR5H).

A percolation flux calculated from the production store is then added to P_r. A total of 10% (fixed partitioning coefficient $B = 0.1$) of P_r goes to a direct branch where it is routed to the catchment outlet by a symmetric UH. The remaining 90% ($1 - B = 0.9$) of P_r goes to a second branch where it is routed to the catchment outlet by an asymmetric UH and a routing store. The flow exiting the routing store (Q_r) is added to the flow exiting the symmetric UH (Q_d) to form the final simulated flow.

2.3.2 | Modelling hypotheses

- *Volume hypothesis*: The first hypothesis is based on the assumption that, at the scale of a rainfall event, when intensities are high and soil moisture is low, the reference model underestimates the effective rainfall volume (part of net rainfall that does not infiltrate the production store).
- *Temporal distribution hypothesis*: The second hypothesis is based on the assumption that there is enough simulated effective rainfall, but that the temporal distribution of effective rainfall should be modified when rainfall intensities are high to allow more of the effective rainfall to reach the outlet during the event.
- The third hypothesis is a combination of the first and second hypotheses.

The concept of "enough effective rainfall" can be seen as fundamentally conceptual: It is only used here to explicate our modelling hypotheses. In addition, the effective rainfall volume and its temporal distribution are closely related concepts and their separation depends to some extent on the chosen time window.

Our starting points to construct the first modelling hypothesis are the formulations proposed by Ficchì (2017) and Peredo et al. (2022). They tested a modification of the effective rainfall rate calculation within the production store of the GR5H model. Equation 3 implies that P_r depends on P_n. However, the instantaneous production rate does not: It only depends on the reservoir filling rate. Therefore, when the production store level is low, the production rate becomes small even for large values of rainfall intensity that will mostly fill the production store. The time required for the production store to reach a level high enough to produce more effective rainfall might be too long in the case of fast flood events under low antecedent soil moisture conditions. To cope with this problem, Ficchì (2017) proposed making the instantaneous production rate dependent on the net rainfall intensity using the following expression (also used but in a different context by Saadi, 2020):

$$\eta' = (1 - \gamma).s^2 + \gamma, \quad (4)$$

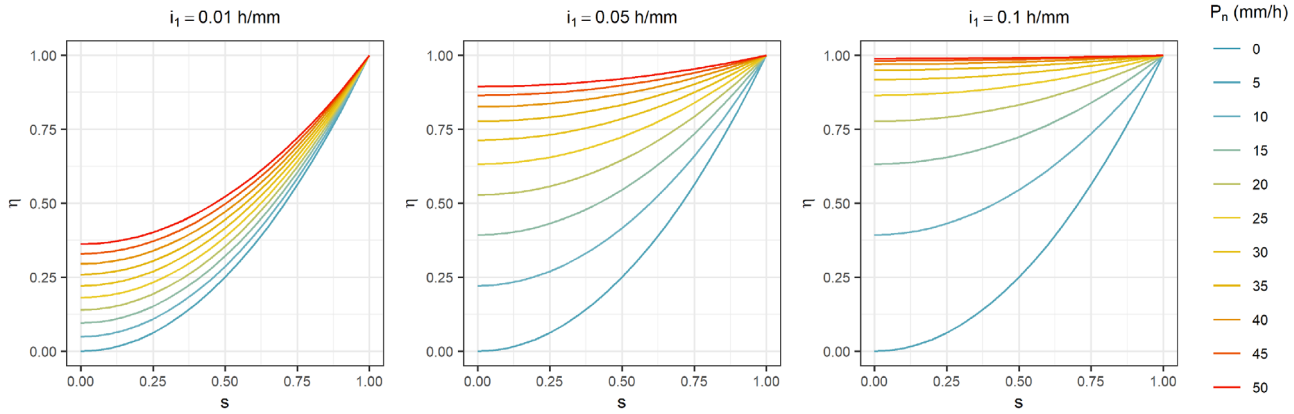

FIGURE 3 Variation in production rate given a production store level and a rainfall intensity rate (first modelling hypothesis)

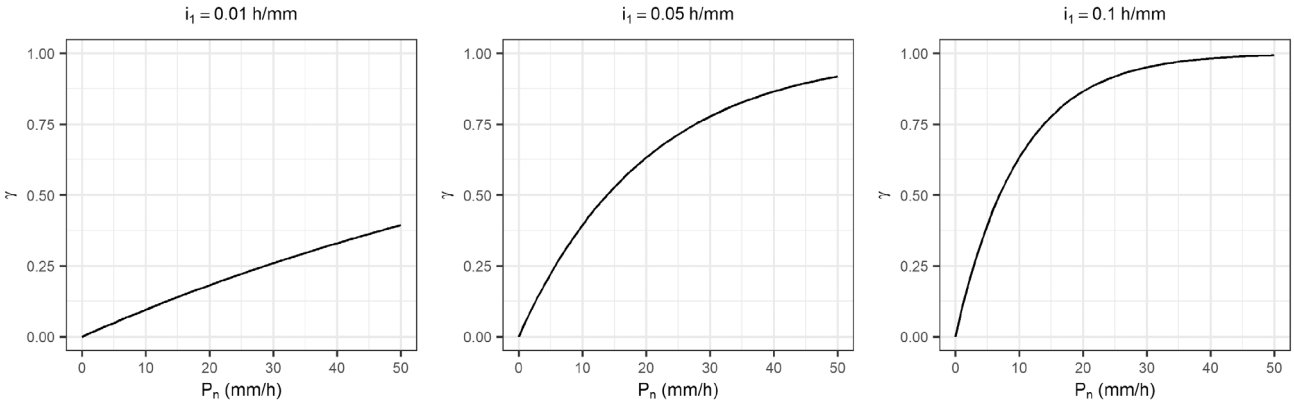

FIGURE 4 Variation in production rate (first modelling hypothesis) when the production store is theoretically empty ($\eta'[s=0]=\gamma$)

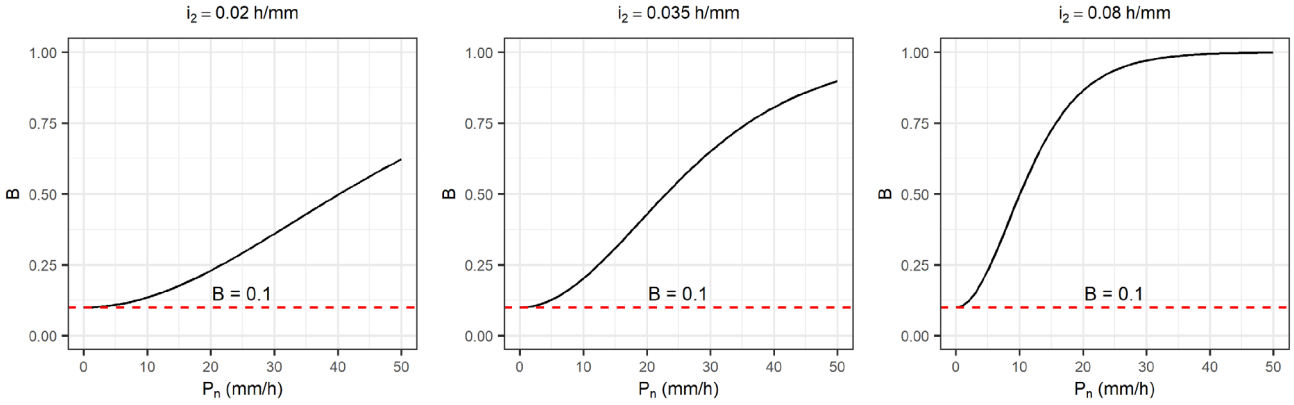

FIGURE 5 Variation in effective rainfall partitioning with rainfall intensities (second modelling hypothesis)

$$\gamma = 1 - \exp(-P_n \times i_1) \tag{5}$$

with i_1 an additional parameter in h/mm and γ a rainfall intensity coefficient. This formulation enables the production store to produce more effective rainfall when its level is low but when rainfall intensities are high (Figures 3 and 4). When the production store level is high or when rainfall intensities are small, the production rate tends toward the reference GR5H calculation (Equation 3). The function is disabled when $i_1 = 0$.

For the second modelling hypothesis, we made the effective rainfall partitioning between the two branches of the routing function vary with rainfall intensities: The assumption is that for larger rainfall intensities, faster processes occur in the catchment. Therefore, the model needs to route the effective rainfall faster than usual by diverting a higher fraction of P_r to the direct branch (see Andrieu et al., 2021, for an application with geomorphological instantaneous unit hydrographs). Note that this dynamic effective rainfall partitioning into two branches (direct and indirect) is conceptual, which means

TABLE 3 List of the four model versions tested

Modification	Equation(s)	Abbreviation	No. of free param.
None	(3)	M0 (Benchmark GR5H)	5
Production store	(4) and (5)	M_1-P_r	6
Partitioning coefficient and direct branch	(6), (7), and (8)	M_2-B	7
Production store, partitioning coefficient and direct branch	(4), (5), (6), (7), and (8)	M_3-P_r-B	8

Note: P_r is the effective rainfall. B is the partitioning coefficient.

that it aims to reproduce the catchment behaviour rather than any specific local physical process.

The partitioning coefficient B is calculated as follows:

$$B = 0.9 \times \tanh\left(\frac{\sum_{j=t_1}^{t} P_n(j).i_2}{t_1 = t - 2.X_4 + 1}\right)^2 + 0.1 \quad (6)$$

with X_4 the UH half-time constant and i_2 an additional parameter in h/mm. When i_2 equals 0, B equals 0.1. This is a way to disable the function when rainfall intensities are low or when there is no rainfall at all. The shape of the intensity function differs from the first modelling hypothesis because B only depends on P_n here. We made the assumption that when intensities are low, B should not vary from its original value. B should also gradually increase with rainfall rates (Figure 5). A rain smoothing function was introduced to place less emphasis on 1-h rainfall values that can be subject to large uncertainties especially in the case of large rainfall intensities. Therefore B varies with the net rainfall intensity of the current time step and also of the previous time steps.

The shape of the unit hydrograph used in the direct branch UH was modified (exponent changed from 1.25 to 2.5) because of the resulting hydrograph shape when B increases (not shown here; for more details, see Figure 9.11 of Le Moine, 2008). This change was applied for all three modelling hypotheses. A linear store was added after the UH of the direct branch to improve the transfer of effective rainfall. The outflow of the linear store is calculated as follows (discrete formulation; see Appendix C of Le Moine, 2008):

$$Q_{d,\Delta t} = K_L \times L_t \quad (7)$$

with $Qd_{\Delta t}$ the specific volume (mm) exiting the linear store between time $t-1$ and time t, $K_L \in [0;1]$ the dimensionless linear store coefficient, and L_t the updated store level.

$$L_t = L_{t-1} + Q_{dHU,\Delta t} \quad (8)$$

with $Q_{dHU,\Delta t}$ the specific volume (mm) exiting the direct UH between time $t-1$ and time t and entering the linear store.

Table 3 summarizes the four model versions used in this study. Overall, the proposed modifications increasingly modify model functioning as rainfall intensity increases, and are neutral in the case of low rainfall intensity (and in the case of high soil moisture for the volume hypothesis). In total, approximately 100 versions were tested but are not detailed here for the sake of brevity. They mostly consisted in testing different shapes of the intensity functions, adding a dependency of the partitioning coefficient to the production store filling rate, routing the supplementary effective rainfall by the direct-branch, and changing the shapes of the UHs.

2.4 | Calibration and evaluation

We estimated the free parameters of the four model versions (see Table 3) following the method used in Astagneau, Bourgin, et al. (2021): The models were calibrated for each catchment on two independent sub-periods P1 and P2 (see Section 2.1) and the EGD (exhaustive gridding discretization) calibration algorithm (Perrin et al., 2008) was applied considering a warm-up period of 2 years to initialize the states of the models. The EGD method proceeds in two steps: First, a screening of the parameter space is performed to find a starting point from 3^n parameter sets (with n the number of parameters); second, a local-gradient-based optimization is run to find the parameter set that gives the highest performance. The Kling–Gupta efficiency criterion (KGE; Gupta et al., 2008) was used as objective function. The GR5H model was run and modified in the R environment (Astagneau, Thirel, et al., 2021; R Core Team, 2021) with the airGR package (Coron et al., 2017, 2020).

Model performance was first evaluated following the evaluation framework of Astagneau, Bourgin, et al. (2021): The overall performance in simulating the entire hydrograph is first assessed using the KGE index; then, the performance of the models in simulating floods is evaluated by calculating two performance metrics on each flood event independently – a volume error criterion (or event relative bias, β) and the Nash–Sutcliffe efficiency criterion (NSE; Nash & Sutcliffe, 1970). The catchment mean flow was used as the benchmark for the calculation of NSE on events. The NSE was calculated to cover more properties of the flood hydrograph than the volume efficiency criterion alone. We calculated a bounded version of these criteria to enable a better comparison between catchments and between events (Mathevet et al., 2006). Both bounded criteria take values between -1 and 1. Negative values of β indicate underestimation of the flood volume, positive values indicate overestimation. Negative values of the bounded NSE criterion indicate that the catchment mean flow is a better predictor than the model in simulating a flood event. When the NSE value tends toward 1, the observed and

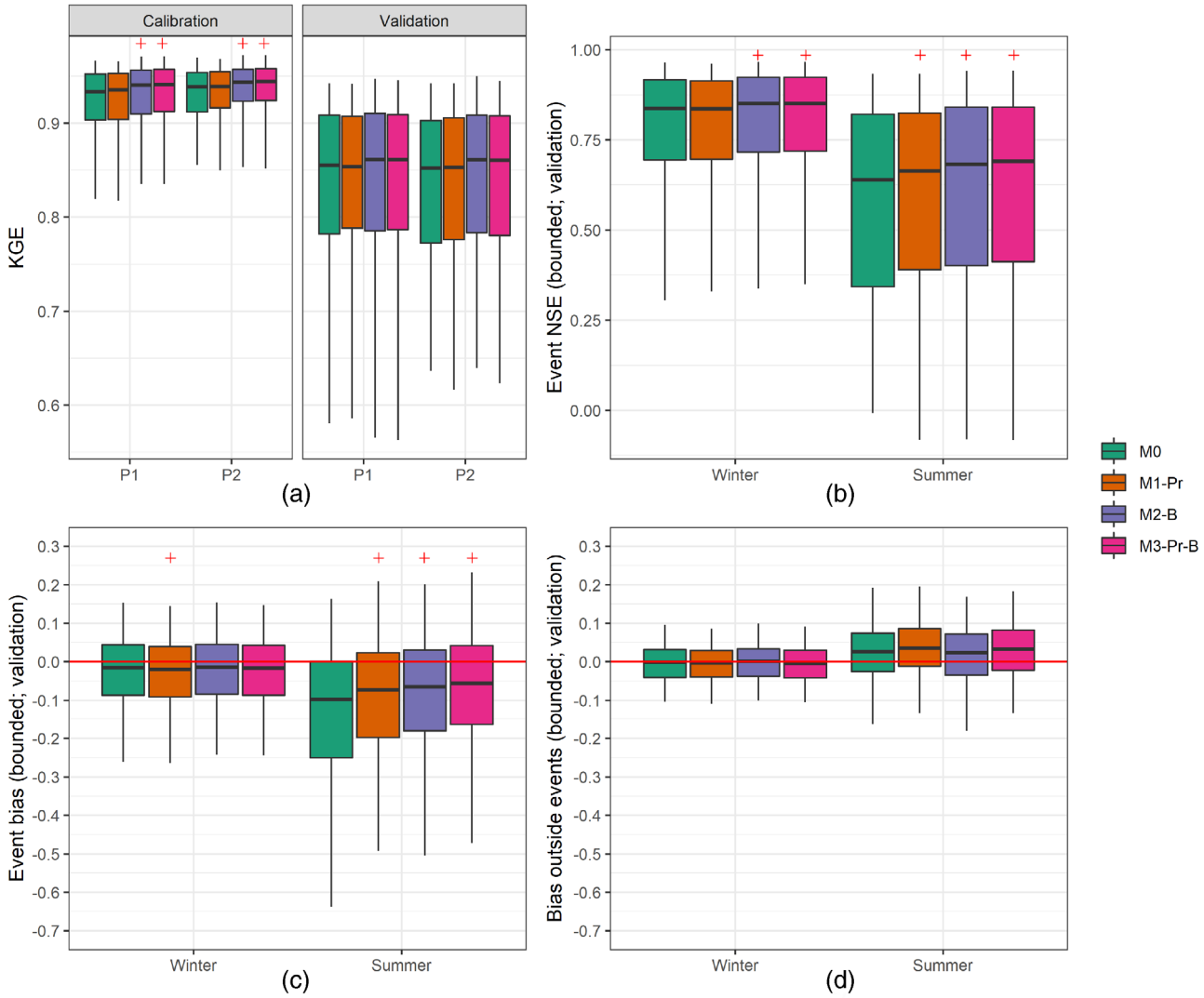

FIGURE 6 Overall performance (one value for each catchment) (a), performance in simulating 8290 winter events and 2362 summer events (b,c), performance outside flood periods (d) for the three modelling hypotheses compared to the reference model (cross-validation values). The red crosses indicate a significant change from the reference model. The distributions are presented between the 5th and 95th percentiles

simulated flood hydrographs tend to match perfectly. This bounded criterion enables a better graphical representation of the model results because values of the original NSE criterion that are lower than −1 draw the distribution toward large negative values. This transformation equals zero when NSE is zero. β was also calculated on the streamflow time series without the observed and simulated flood events to investigate possible compensations between high-flow and low-flow periods. We considered two seasons for the evaluation of flood performance. Flood peaks occurring between May and October were labelled as summer floods. Flood peaks occurring between November and April were labelled as winter floods. Flood simulations were then analysed according to the five groups of catchments defined in Section 2.2 and three event characteristics: mean rainfall intensity, spatial variability in cumulative rainfall, and antecedent soil moisture conditions (SWI). All the results are presented using cross-validation simulations, which means that the criteria were calculated on P1 while model simulations were generated with the parameter sets estimated on P2 and vice versa. As the GR5H model was used in a continuous mode, the simulations of the flood events that we have evaluated were not run separately. They were extracted from the two continuous time series (P1 and P2) of model simulations (cross-validation values). We performed the Wilcoxon rank test (Wilcoxon, 1945) to detect statistically significant changes in performance between the three modelling hypotheses and the reference model (at significance level 0.05). The resulting p values are reported in the text.

3 | RESULTS

3.1 | Overview of model performance

First, Figure 6 shows the distribution of model performance over the 229 catchments and 10 652 flood events. In calibration mode, for

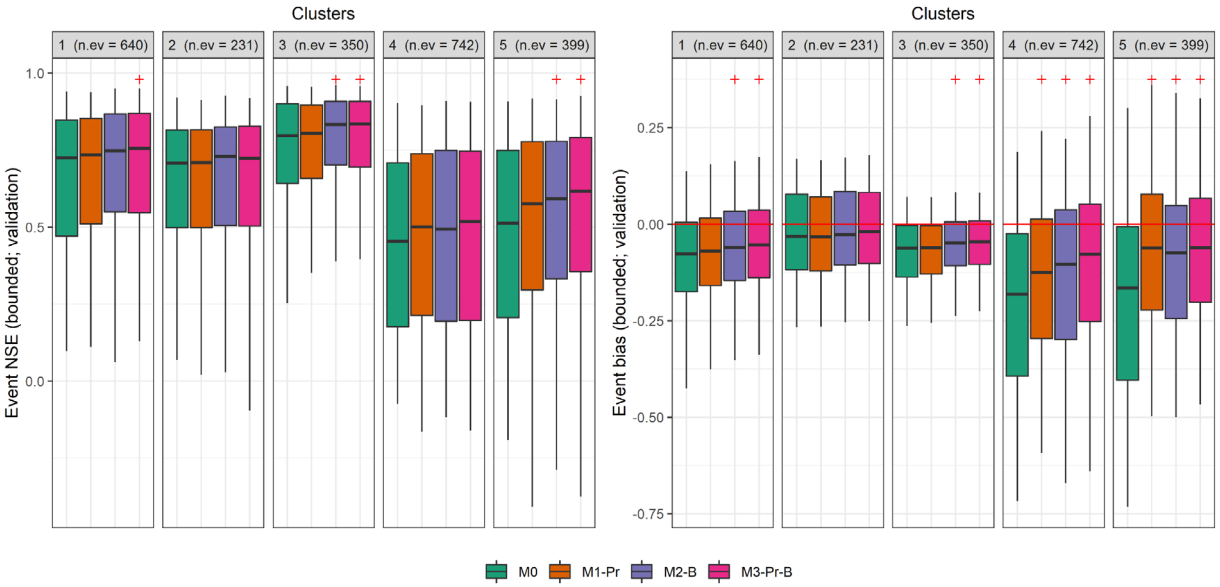

FIGURE 7 Distribution of flood performance (cross-validation values) over 2362 summer floods divided into five groups of catchments. The red crosses indicate a significant change from the reference model. The distributions are presented between the 5th and 95th percentiles

both sub-periods, all three model versions show an improvement over the reference version M0 (p values of 0.39, 1.8e−3 and 6.0e−4 in the order of model numbering): they have higher median values and a smaller dispersion of the KGE index than the reference model (6a). M3-Pr-B exhibits the highest KGE values and M2-B the second highest (see Table 3 for naming). In validation mode, all three models have similar KGE values to M0. M2-B and M3-Pr-B have higher median KGE values than M0 for both sub-periods (p values of 0.19 and 0.18 respectively).

Event performance as assessed by the NSE index increases in summer for all model versions (p values of 0.03, 7.0e-5 and 7.2e-6 in the order of model numbering) with M3-Pr-B having the highest median NSE values and M2-B the second highest (6b). In winter, M1-Pr performs equally to M0 (p value of 0.50). The median NSE values of M2-B and M3-Pr-B are higher than the median NSE of M0 (p values of 1.3e−7 and 7.1e−9, respectively). In terms of volume error performance in summer, M3-Pr-B has the median value closest to zero and the smallest dispersion (6c; p value <2.2e−16). Both M1-Pr and M2-B have median values closer to zero and smaller dispersion of volume error than M0 (p values of 1.0e−9 and 1.2e−14, respectively), with M2-B having the smallest bias of the two. In winter, the volume error of M2-B and M3-Pr-B is close to the reference model (p values of 0.80 and 0.32, respectively). There is a slight degradation of the volume error of M1-Pr toward negative values compared to M0 (p value of 0.01). The distributions of the volume error values calculated outside the periods of flood events show that M1-Pr and M3-Pr-B tend to produce larger overestimations of streamflow in summer (p values of 0.08 and 0.36, respectively) than M0 and M2-B (6d). In winter, the median values of this criterion are close to the reference model with a slight degradation toward negative values for M1-Pr and M3-Pr-B (p values of 0.40 and 0.50, respectively). M2-B exhibits a similar distribution to M0 (p value of 0.63). This overview of model performance showed that the proposed model versions have similar overall performance but that simulation of summer floods is improved both in terms of NSE and volume error. Simulation of winter floods is not degraded and is even improved for M2-B and M3-Pr-B. The volume hypothesis leads to an overestimation of volume outside periods of flood events. Overall, the largest improvements in terms of flood performance compared to the reference model are achieved by the combination of the two modelling hypotheses (M3-Pr-B).

3.2 | Variation in performance within hydroclimatic groups

We refine the analysis by looking at model performance in simulating summer floods considering each group of catchments identified in Section 2.2. Figure 7 shows that summer flood performance increases for all groups of catchments. The smallest increase in summer flood performance is for catchments of group 2 (for the NSE distributions, p values of 0.92, 0.52, and 0.46 in the order of model numbering). The largest increases in summer flood performance are found for groups 4 and 5 where the three modelling hypotheses perform better than the reference model, especially in terms of event bias. For these two groups, the reference model greatly underestimates flood volumes in summer and yields the lowest flood performance compared to the other three groups. The M3-Pr-B model has the highest flood performance, especially for catchments of group 5 (for the NSE distributions, p value of 7.4e-4).

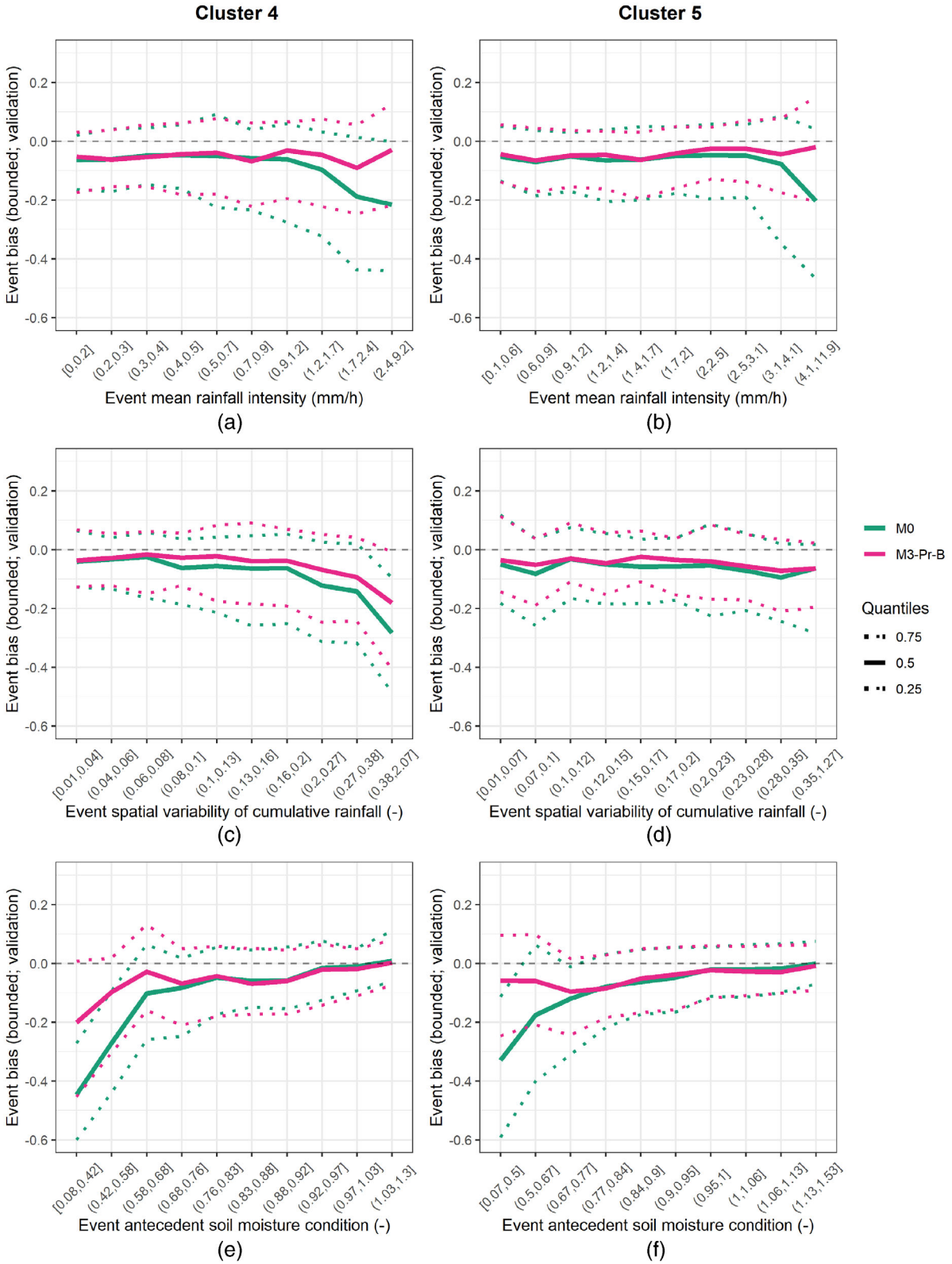

FIGURE 8 Variation in event performance (cross-validation values) against three event characteristics for two groups of catchments. Floods were divided into 10 quantile classes per characteristic and cluster group. A total of 1932 events for cluster 4 and 1508 events for cluster 5 are considered here

The volume hypothesis (M1-Pr) performs better in terms of event bias than the temporal distribution hypothesis (M2-B) for catchments of group 5 (p values of 3.5e−9 and 9.0e−7 respectively). For the other four groups, M2-B has higher flood performance than M1-Pr. We will only focus on M3-Pr-B and catchments of groups 4 and 5 in the analyses that follow.

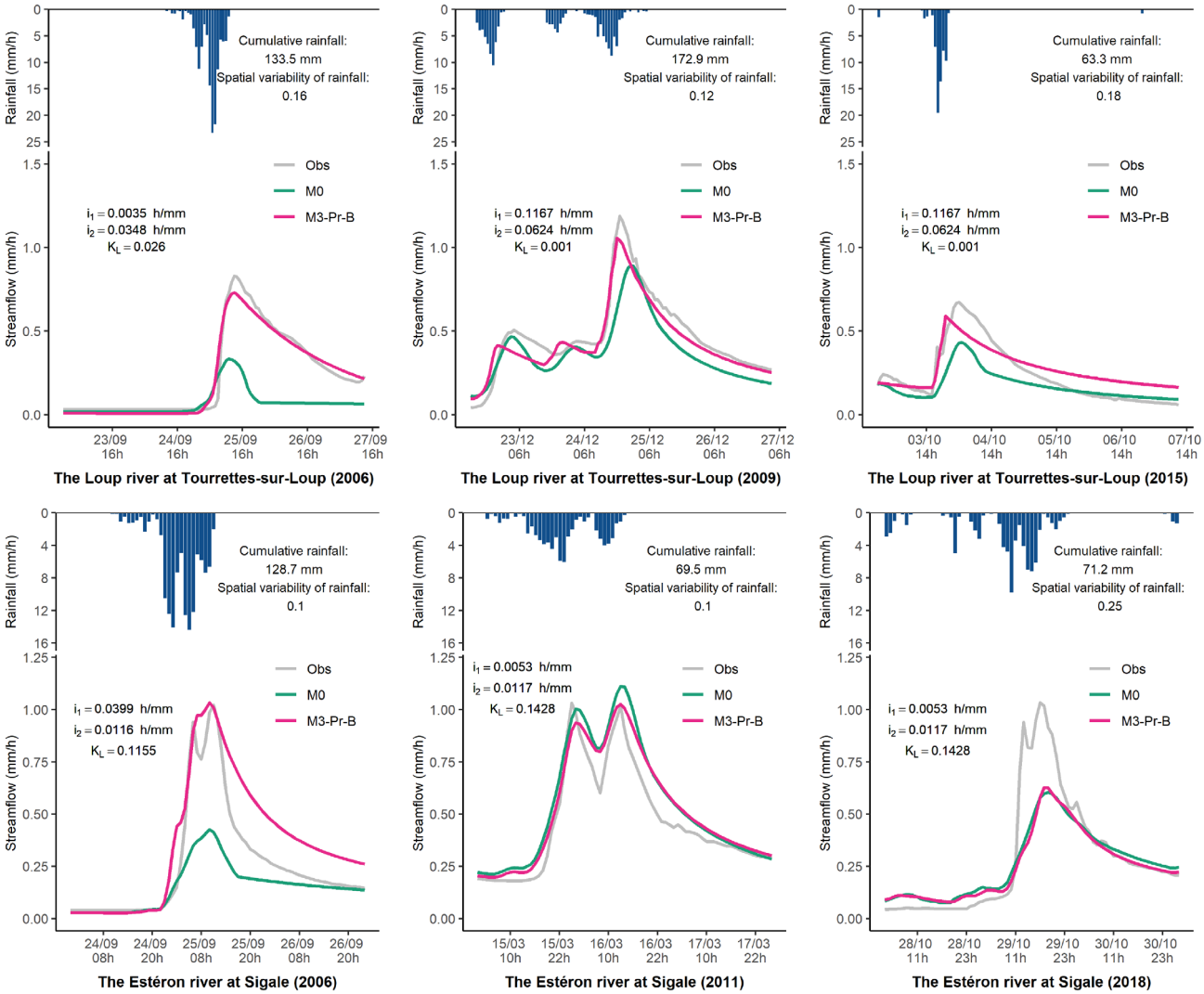

FIGURE 9 Simulated (M0 and M3-Pr-B) and observed hydrographs of six floods that occurred in the Loup River at Tourrettes-sur-Loup and the Estéron River at Sigale (cross-validation values)

3.3 | Performance and flood characteristics

Considering all floods occurring in the catchments of groups 4 and 5, we investigate the relationships between bias in simulating floods and three event characteristics (defined in Section 2.2) for the reference model and the third modelling hypothesis (Figure 8). Figure 8a, b show that the reference model underestimates flood volumes when there are large rainfall intensities. For these events, the event bias values of M3-Pr-B are less dispersed and the median is closer to 0.

For catchments of group 4, the event bias values of the reference model tend to be lower with increasing spatial variability in cumulative rainfall (Figure 8c). The event bias obtained with the new model version follows the same trend but with less dispersed values and a median value closer to 0. The decrease in performance with increasing rainfall spatial variability is less pronounced for catchments of group 5 (Figure 8d). The reference model clearly underestimates flood volumes for both groups of catchments under low antecedent soil moisture conditions. Under these conditions, M3-Pr-B performs better, especially for floods of group 5.

Figure 9 shows the simulation of six flood hydrographs for two different Mediterranean catchments with typical streamflow responses to intense rainfall. The flood hydrographs presented on Figure 9 were retrieved from two different sub-periods (validation periods P1 and P2), which means that the parameter values are not necessarily the same between the events (two different calibration periods were used). The Loup River at Tourrettes-sur-Loup (206 km^2, group 5) and the Estéron River at Sigale (262 km^2, group 4) are located in southern France (see Figure 1). The Estéron River is a tributary of the Var River, and the Loup River flows directly into the Mediterranean Sea. These catchments are characterized by high seasonal variations in streamflow and steep slopes. Severe floods can occur, usually in winter or in the beginning of autumn and sometimes due to intense convective rainfall events. These examples were selected to illustrate our results and complement the presentation of the scores with boxplots. They cannot be considered as representative of the

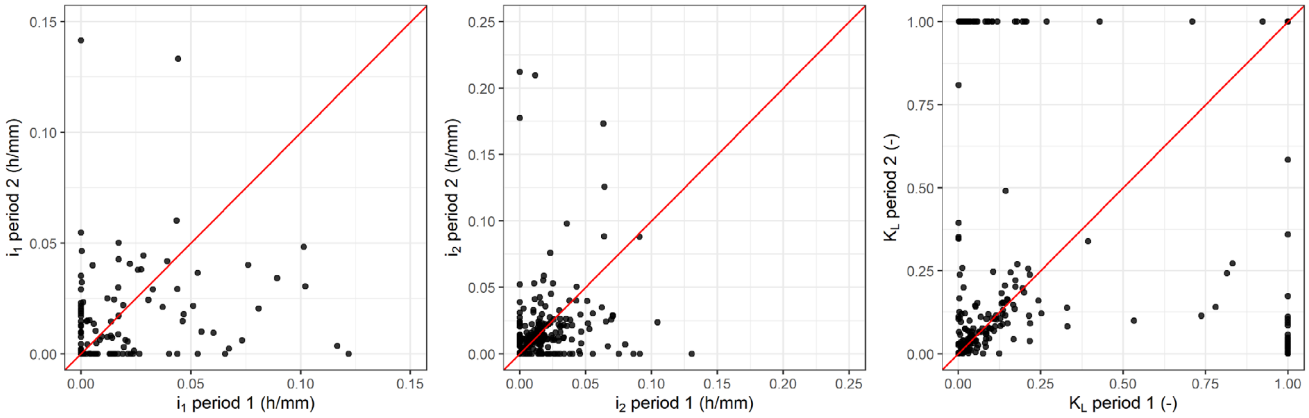

FIGURE 10 Variation in additional M3-Pr-B free parameters between calibration periods

10 652 events of our dataset in the sense of exploiting the outputs of a clustering procedure.

The summer flood of September 2006 that occurred in the Loup River followed an intense rainfall event with intensities of up to 23 mm/h but evenly distributed over the entire catchment area. For this event, the reference model clearly underestimated the peaks and the overall flood volume. The M3-Pr-B on the other hand managed to reproduce the flood volume and the flood peak better, although it is still underestimated. The flood of December 2009 is associated with a larger amount of, but less intense, rainfall. Both models are able to reproduce the hydrograph shape. In the beginning of October 2015, a smaller flood event (in terms of peak and volume) was followed by a short and intense rainfall event. The spatial variability in this rainfall event is slightly larger than the other two events and the total amount of rainfall is almost three times smaller. For this event, the reference model failed to simulate the catchment response, whereas the M3-Pr-B model was able to reproduce the flood peak, volume, and timing better, with a slight overestimation of the recession volume.

The flood of September 2006 that occurred in the Estéron River followed a large and intense rainfall event with intensities of up to 14 mm/h mainly distributed over the entire catchment area. The reference model response to this event was clearly limited compared to the catchment response. The new model version, on the other hand, managed to reproduce the peak and timing of the flood but overestimated the recession volume. The March 2011 flood was associated with a smaller amount of rainfall and smaller intensities but was reproduced well by both models. In October 2018, a rainfall event with cumulative rainfall in the same order of magnitude as the winter flood of 2011, but highly variable in space, led to a fast and intense catchment response. Both models greatly underestimated the flood volume and peak of this event.

3.4 | Additional model parameters

The most complex model that we have introduced in this study is M3-Pr-B, which has one additional reservoir and three additional free parameters compared to M0. These parameters control the modification of fluxes and storages of the model with varying rainfall intensities. The additional functions we proposed were also made to deactivate the functions in the case of low intensities and, for certain values of i_1, i_2, and K_L, are equivalent to M0. Figure 10 shows that these parameters vary greatly between calibration periods, especially i_1 and to some extent i_2. For some catchments, i_1 or i_2 is null in one sub-period and above zero in the other, indicating that the intensity function is disabled for one period but not for the other (see Equations 4, 5 and 6 and the related explanations). The same behaviour is observed for K_L, with several values equal to 1 but with smaller variations between periods and for more catchments.

As the additional parameters of the third modelling hypothesis vary between calibration periods, the intensity functions are not activated in all periods. As a consequence, on some catchments both the reference model and the new modelling hypotheses are selected as the best performing models. Figure 11 shows that the same modelling hypotheses are activated in both calibration periods for 118 catchments of our dataset.

We investigated whether there are differences of parameter distributions between the five hydroclimatic groups of catchments. Results indicate that i_1 follows different distributions for groups 4 and 5 (Figure 12) with more values above zero, especially for catchments of group 5 (85% of pairs [catchment, sub-period] have i_1 values greater than zero). This means that the first modelling hypothesis is more often activated for catchments of group 5 than for the other catchments. The distributions of i_2 values are very similar between the five groups. The distributions of K_L are different between the five groups. Group 5 has the lowest values of K_L and group 2 has the highest number of values equal to 1 (no linear store on the direct branch).

4 | DISCUSSION

4.1 | On the available information content to identify additional parameter values

4.1.1 | Activation of intensity-rate functions between periods

The previous results indicate that the additional parameters of the proposed intensity functions vary between the two calibration

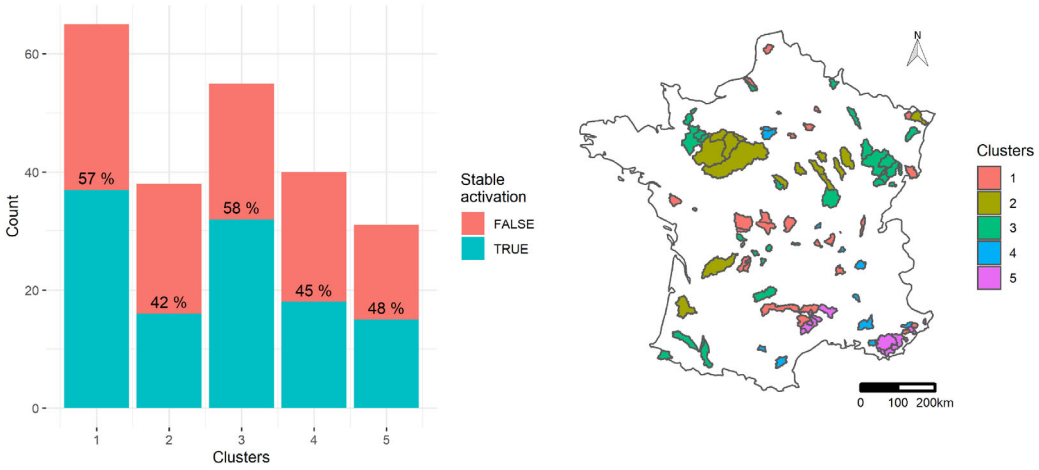

FIGURE 11 Catchments on which the intensity functions of M3-Pr-B are activated on both calibration periods

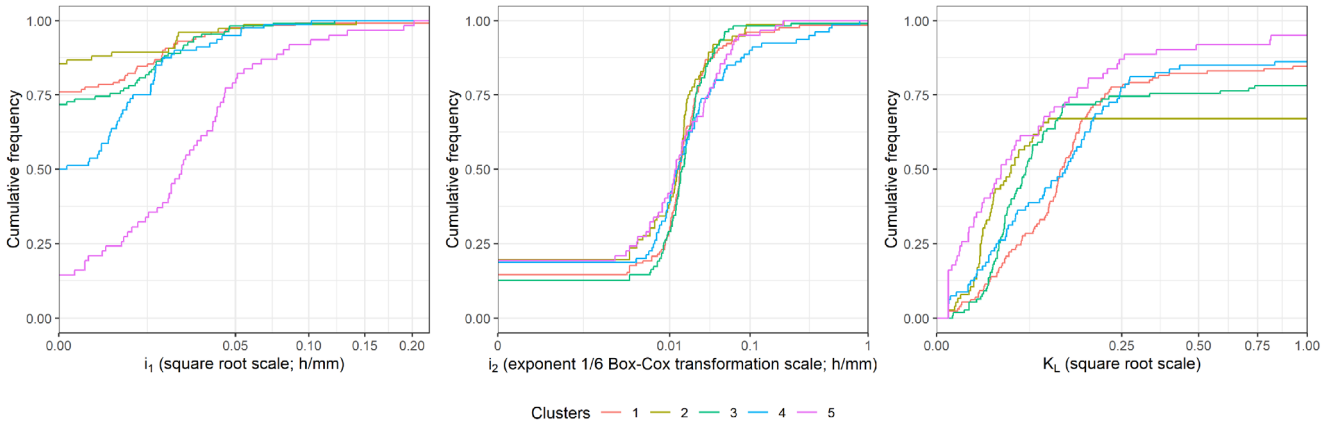

FIGURE 12 Distributions of the additional M3-Pr-B free parameters for five groups of catchments

periods, which implies some variations in the intensity function activation. In addition, the parameter distributions (Figure 12) and the performance of the different modelling hypotheses (Figure 7) demonstrate that the additional parameters and their sensitivity to the chosen metrics differ between the groups of catchments. Parameter identifiability can therefore be considered as catchment-dependent. We refer to "parameter identifiability" for issues related to the dependence of parameters to groups of catchments and to the variation of parameter values between calibration periods.

This issue was also raised by van Esse et al. (2013) who found that a more flexible model often yields equifinality of possible structures (also demonstrated here as an equifinality of parameters) for a given catchment. One reason can be that on some catchments, most of the flood events that we have selected occur in winter, which suggests that the intensity functions are activated for a very small number of data points. Peredo et al. (2022) found that a function similar to the first modelling hypothesis affects the effective rainfall volume simulated on the Aude catchment only for a very few time steps and mostly between July and December. Therefore, the estimation of the added parameters can become very uncertain, especially since there are more parameters to be estimated compared to the initial model structure. Furthermore, as suggested by Astagneau, Bourgin, et al. (2021), the KGE index calculated on the whole streamflow time series might not be sufficiently informative to evaluate model performance in simulating summer floods, especially since summer floods are often shorter than winter floods and therefore less data points are available for calculating a performance criterion. Although we used a multi-objective framework for model evaluation, only the KGE index was used for parameter estimation in this study. This parameter identifiability issue could also be related to a possible temporal variability in hydroclimatic conditions between the calibration periods. Finally, there might just not be enough information in the rainfall time series at the hourly time step to derive optimal parameter values at the event scale.

4.1.2 | Can we estimate generic parameter values?

To limit the parameter identifiability issue, one solution would be to either fix the values of the additional parameters, or prescribe them

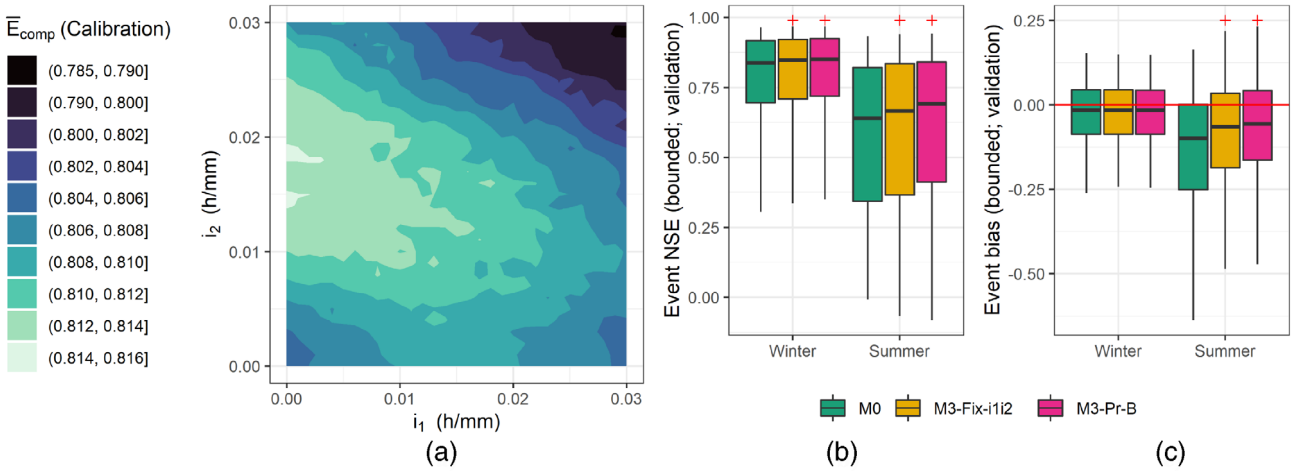

FIGURE 13 Sensitivity of the additional parameters to the composite criterion E_{comp} over the 229 catchments (a). Distribution of performance in simulating 8290 winter floods and 2362 summer floods for the third modelling hypothesis run with generic i_1 and i_2 values in validation (b,c). M3-Fix-i1i2 is the model version with fixed i_1 and i_2 values for all catchments. The red crosses indicate a significant change from the reference model. The distributions are presented between the 5th and 95th percentiles

using climatic or physical descriptors. Here, we investigated the use of generic values of i_1 and i_2 for the third modelling hypothesis (M3-Pr-B). We selected the combination (i_1, i_2)—for which X_1–X_5 and K_L were calibrated on P1 and P2 with the KGE objective function—that yielded the maximum mean value of a composite criterion in calibration over the 229 catchments of our dataset. This criterion is calculated as follows for each catchment and calibration period:

$$E_{comp} = \frac{1}{2}\left(KGE^b + \overline{NSE}^b_{ev}\right)$$
$$KGE^b = \frac{KGE}{2 - KGE} \quad (9)$$
$$\overline{NSE}^b_{ev} = \frac{1}{n_{ev}} \times \sum_{i=1}^{n_{ev}} \frac{NSE_i}{2 - NSE_i}.$$

KGE^b is a bounded version of the KGE index. $\overline{NSE}^b_{ev}$ is the average bounded NSE calculated on the events of one catchment and one calibration period. n_{ev} is the number of selected events on one catchment and one calibration period.

$i_1 = 0.02$ and $i_2 = 0.018$ were selected as the best generic parameter values for the intensity functions of M3-Pr-B. Figure 13a shows that i_2 (temporal distribution hypothesis) is well identified among the different parameter sets that were tested. Interestingly, an increase in i_1 (volume hypothesis) when i_2 (temporal distribution hypothesis) tends toward its optimal value degrades the composite criterion, whereas an increase in i_2 when i_1 tends toward its optimal value improves the composite criterion. This means that the temporal distribution hypothesis is better adapted to make the GR5H model simulate catchment response to different flood-generating processes across different hydroclimatic conditions and therefore improves its versatility.

The third hypothesis run with the best generic i_1 and i_2 parameters (M3-Fix-i1i2) improves the performance of the reference model M0 for winter and summer events (validation), especially in terms of bias in summer (p value of 5.3e−15). However, the performance in simulating floods in validation periods is lower than the fully calibrated version. Looking back at the performance of the fully calibrated volume hypothesis (M1-Pr; Figure 6), we can see that M3-Fix-i1i2 achieves the same level of performance as M1-Pr in simulating summer floods but M3-Fix-i1i2 seems to be more suited to simulating winter floods. Although the volume hypothesis is almost not activated ($i_1 = 0.002$) when looking for generic values of the additional parameters, the results of Section 3.2 indicate that this hypothesis is needed to simulate the summer floods of several catchments in cluster 5. However, it might be too costly to activate this function for the other catchments of our dataset where summer flood-generating processes may have different properties at the catchment scale.

Overall, the generic parameters that we have identified can be seen as starting points to improve the ability of the GR5H model to better simulate streamflow in the case of flooding from intense rainfall. Their use enables a reduction of the parameter space during the calibration process, which can be especially useful when not enough data are available or when parameter identifiability is low.

4.2 | Rainfall intensity and hypothesis testing at the catchment scale

4.2.1 | A better temporal distribution of increased runoff production

Our results show that the second modelling hypothesis (temporal distribution hypothesis) performs better than the first hypothesis (volume hypothesis) in terms of overall performance and flood performance over 229 catchments and 10 652 flood events. However, for several catchments of the Rhône River and the Mediterranean arc, the first modelling hypothesis seems to be needed to

reproduce flood volume in the case of large intensities, as indicated by the event bias values for summer events in cluster 5 (Figure 7). The more frequent positive values of the associated additional parameter (i_1; Figure 12) for the third hypothesis (combination of the first and the second hypotheses) confirm that the volume hypothesis is needed to simulate summer floods in cluster 5. For these catchments, this could mean that under low antecedent wetness conditions, the runoff coefficient is higher in the case of intense rainfall events. Therefore, improving the temporal distribution of effective rainfall is not sufficient to simulate the flood hydrograph. However, we expect the model to be able to simulate more volume at the event scale with the first hypothesis, as the water balance is affected at the event scale in that case whereas this is not the case for the second hypothesis. This increased volume in the case of large intensities results in larger flow overestimation outside flood periods (Figure 6d), which indicates that the temporal distribution of the additional volume simulated at the event scale should be improved as well.

Here, we made the assumption that the effective rainfall volume produced by the model can be separated from its temporal distribution. This is certainly not always true, as a more delayed distribution of effective rainfall would result in a lower simulated flood volume and therefore in a lower simulated runoff coefficient at the event scale. The performance of the third modelling hypothesis (e.g. Figure 6b,c) shows that the model needs both an increase in runoff production (first hypothesis) and a faster transfer at the event scale (second hypothesis). The third modelling hypothesis adapts to more heterogeneous flood-generating processes in summer while keeping its ability to simulate floods under wetter conditions.

Overall, the results indicate that more complexity is needed to simulate floods of catchments in groups 4 and 5. Conversely, the intensity functions do not seem to be needed for catchments in cluster 2. The catchments in cluster 2 are associated with the most constant hydroclimatic conditions between events (low variability in rainfall intensities, low variability in rainfall spatial variability, low variability in antecedent wetness conditions, low variability in flood duration) and with the slowest catchment response to rainfall (high time lag and high flow autocorrelation at 24 h) indicating very uniform flood-generating processes that the reference model seemed to capture well. This confirms that a robust parsimonious model can perform well under wetter and more uniform hydroclimatic conditions without unnecessary complexity (e.g., Jakeman & Hornberger, 1993), but some adaptation of its structure might be needed to simulate streamflow response to more heterogeneous processes (e.g. Knoben et al., 2020), especially at the hourly time step (Ficchì et al., 2019).

4.2.2 | Spatial variability in rainfall intensities

The analysis of performance in simulating floods in clusters 4 and 5 depending on three causative processes further indicates the value of our modelling hypotheses in a Mediterranean context. Figure 8a,b indicate that the floods following the 10% most intense rainfall events are less underestimated by the new model version. The upper part of the error distribution also indicates that there might be a small bias toward overestimation of 25% of these events. Extreme rainfall events, especially in summer, are known to be subject to large uncertainties (e.g. Ruiz-Villanueva et al., 2012; Zhang et al., 2017) in their estimation, making the interpretation of the upper part of the error distribution difficult to establish.

Overall, the results show that bias in simulating floods associated with high-intensity rainfall events under low antecedent soil moisture conditions is reduced for many catchments located in the Mediterranean area. However, we noticed that for catchments in cluster 4, and for events with large values of the spatial variability index, the new model still largely underestimates flood volumes, while this is not the case for the catchments in cluster 5. This result can first appear counter-intuitive since we used a spatially lumped model in our study. However, looking at the cluster's hydroclimatic properties (Figures 1 and A1), we can see that there is a higher variability in the rainfall spatial variability coefficient between events for catchments in cluster 4 than for catchments in cluster 5. This could mean that the lumped GR5H model is better able to reproduce flood volumes when rainfall events usually affect the same area of the catchment. This explanation is consistent with the findings of Lobligeois et al. (2014), who showed that the impact of spatial resolution of rainfall on model performance is catchment-dependent. Another explanation could be that the flood events associated with high spatial variability for catchments in cluster 4 are not associated with very intense rainfall at the catchment scale for the whole duration of the events, whereas the intensities associated with the flood events of catchments in cluster 5 in the case of large values of spatial variability are sufficient to make the new model simulate more volume. These low intensity values could also be the result of large uncertainties that are sometimes associated with short-duration rainfall events (e.g. Ruiz-Villanueva et al., 2012; Zhang et al., 2017). Overall, separating rainfall spatial variability from rainfall intensity allowed us to highlight the benefit of making better use of rainfall intensity information for flood simulation. However, many rainfall events occurring in summer are intense and highly variable in space, and the benefits of using a finer spatial resolution remains to be investigated in order to generalize the findings of some recent studies (e.g. Loritz et al., 2021; Peredo et al., 2022).

5 | CONCLUSIONS

The objective of this study was to use rainfall intensity as a proxy for the activation of fast hydrological processes in order to improve the simulation of summer floods by a conceptual lumped rainfall–runoff model. We worked at the catchment scale, with a large catchment set, and made three modelling hypotheses on the dependence of the fluxes of our model to hourly rainfall intensity rates:

- The first hypothesis consists in assuming that large rainfall intensities increase the volume of effective rainfall.
- The second hypothesis consists in assuming that large rainfall intensities induce a faster routing of effective rainfall to the catchment outlet.

- The third hypothesis combines the first two hypotheses, modifying both the volume and the routing of effective rainfall.

Our results showed that the three hypotheses increased the ability of the hydrological model to simulate summer floods, especially in terms of volume error, and particularly for tributaries of the Rhône River and catchments located in the Mediterranean area. The third hypothesis shows the highest performance in capturing the streamflow response to more heterogeneous storms in summer, while maintaining good performance in simulating winter floods. We name this new model GR5H-RI. Our results indicate that there is a clear benefit in introducing a dependency of the storages and fluxes of a lumped conceptual model on rainfall intensities (at least at the hourly time step). Our work obviously has some limitations; the lumped configuration tested here may be restrictive for some catchments, especially where the spatial organization of rainfall varies between events. Furthermore, the values of the additional parameters are difficult to estimate, because the intensity-dependent functions are only activated on very few time steps. To overcome this issue, we proposed default values for two of the additional parameters, which cause a decrease in model performance in comparison to the fully calibrated model, but still enable a better simulation of floods in comparison to the reference model. Regionalizing the parameters of the intensity-dependent function (with relationships between these parameters and catchment meteorological properties) would be another solution.

Our hope is that a better simulation of floods from intense rainfall will lead to improvements in flood forecasting, providing that our functions are compatible with the assimilation of real-time rainfall and streamflow data. There is a potential for using the proposed modelling hypotheses in a Mediterranean context, especially since an intensification of extreme rainfall events is to be expected with climate change in some parts of this region (e.g. Tramblay & Somot, 2018). Further tests will be conducted in future studies to evaluate the value of this functions to simulate and forecast extremes, for example by using post-estimations of peak discharges of extreme events that occurred in Mediterranean catchments (e.g. the 2nd October 2020 Alex storm in south-eastern France; Payrastre et al., 2022). Finally, even if this paper only discusses the application of the three above modelling hypotheses to the GR5H structure, we believe that they can be applied to other models. Depending on the specificities of each model, the first, second or third hypothesis may prove to be the most suitable.

ACKNOWLEDGEMENTS

This work was supported by the French Ministry of Environment (DGPR/SNRH/SCHAPI), which provided financial support to the PhD grant of the first author. The authors thank Météo-France and Banque HYDRO for providing the climatic and hydrological data, respectively. We would like to thank Lionel Berthet, Keith Beven, Flora Branger, Pierre Brigode, Pierre-André Garambois, Alban de Lavenne, Nicolas Le Moine, Simon Munier, Daniela Peredo, Paul Royer-Gaspard, Léonard Santos, and Félicien Zuber for remarks on, and suggestions for, this work. We thank Jérémy Verrier for his help in using a computer cluster to run our model. Finally, we thank the Associate Editor and two anonymous reviewers for their comments that helped us improve the paper.

DATA AVAILABILITY STATEMENT

The streamflow data are available from Banque HYDRO. The climatic data are not available due to third party restrictions.

ORCID

Paul C. Astagneau https://orcid.org/0000-0002-6688-5783
François Bourgin https://orcid.org/0000-0002-2820-7260
Vazken Andréassian https://orcid.org/0000-0001-7124-9303
Charles Perrin https://orcid.org/0000-0001-8552-1881

REFERENCES

Andréassian, V., Moine, N. L., Perrin, C., Ramos, M.-H., Oudin, L., Mathevet, T., Lerat, J., & Berthet, L. (2012). All that glitters is not gold: The case of calibrating hydrological models. *Hydrological Processes*, *26*, 2206–2210. https://doi.org/10.1002/hyp.9264

Andrieu, H., Moussa, R., & Kirstetter, P.-E. (2021). The event-specific geomorphological instantaneous unit hydrograph (e-GIUH): The basin hydrological response characteristic of a flood event. *Journal of Hydrology*, *603*, 127158. https://doi.org/10.1016/j.jhydrol.2021.127158

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., Santhi, C., Harmel, R. D., van Griensven, A., Liew, M. W. V., Kannan, N., & Jha, M. K. (2012). SWAT: Model use, calibration, and validation. *Transactions of the ASABE*, *55*(4), 1491–1508. https://doi.org/10.13031/2013.42256

Astagneau, P. C., Bourgin, F., Andréassian, V., & Perrin, C. (2021). When does a parsimonious model fail to simulate floods? Learning from the seasonality of model bias. *Hydrological Sciences Journal*, *66*(8), 1288–1305. https://doi.org/10.1080/02626667.2021.1923720

Astagneau, P. C., Thirel, G., Delaigue, O., Guillaume, J. H. A., Parajka, J., Brauer, C. C., Viglione, A., Buytaert, W., & Beven, K. J. (2021). Technical note: Hydrology modelling R packages—A unified analysis of models and practicalities from a user perspective. *Hydrology and Earth System Sciences*, *25*(7), 3937–3973. https://doi.org/10.5194/hess-25-3937-2021

Berghuijs, W. R., Sivapalan, M., Woods, R. A., & Savenije, H. H. G. (2014). Patterns of similarity of seasonal water balances: A window into streamflow variability over a range of time scales. *Water Resources Research*, *50*(7), 5638–5661. https://doi.org/10.1002/2014WR015692

Bergström, S. (1976). Development and application of a conceptual runoff model for scandinavian catchments. Technical Report RHO 7.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., & Harding, R. J. (2011). The joint UKland environment simulator (JULES), model description—Part 1: Energy and water fluxes. *Geoscientific Model Development*, *4*(3), 677–699. https://doi.org/10.5194/gmd-4-677-2011

Beven, K. (2021). The era of infiltration. *Hydrology and Earth System Sciences*, *25*(2), 851–866. https://doi.org/10.5194/hess-25-851-2021

Beven, K., & Kirby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, *24*(1), 43–69. https://doi.org/10.1080/02626667909491834

Beven, K. J. (1984). Infiltration into a class of vertically non-uniform soils. *Hydrological Sciences Journal*, *29*(4), 425–434. https://doi.org/10.1080/02626668409490960

Beven, K. J. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, *16*(1), 41–51. https://doi.org/10.1016/0309-1708(93)90028-E

Beven, K. J., Kirkby, M. J., Freer, J. E., & Lamb, R. (2021). A history of TOPMODEL. *Hydrology and Earth System Sciences*, *25*(2), 527–549. https://doi.org/10.5194/hess-25-527-2021

Birkel, C., Soulsby, C., & Tetzlaff, D. (2014). Developing a consistent process-based conceptualization of catchment functioning using measurements of internal state variables. *Water Resources Research*, *50*(4), 3481–3501. https://doi.org/10.1002/2013WR014925

Blöschl, G. (2017). Debates-hypothesis testing in hydrology: Introduction. *Water Resources Research*, *53*(3), 1767–1769. https://doi.org/10.1002/2017WR020584

Blöschl, G., Bierkens, M. F., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J. W., McDonnell, J. J., Savenije, H. H., Sivapalan, M., Stumpp, C., Toth, E., Volpi, E., Carr, G., Lupton, C., Salinas, J., Széles, B., & Viglione, A. (2019). Twenty-three unsolved problems in hydrology (UPH)—A community perspective. *Hydrological Sciences Journal*, *64*(10), 1141–1158. https://doi.org/10.1080/02626667.2019.1620507

Blöschl, G., & Sivapalan, M. (1995). Scale issues in hydrological modelling: A review. *Hydrological Processes*, *9*(3–4), 251–290. https://doi.org/10.1002/hyp.3360090305

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., & Savenije, H. (Eds.). (2013). *Runoff prediction in ungauged basins*. Cambridge University Press. https://doi.org/10.1017/CBO9781139235761

Bouaziz, L. J. E., Fenicia, F., Thirel, G., de Boer-Euser, T., Buitink, J., Brauer, C. C., Niel, J. D., Dewals, B. J., Drogue, G., Grelier, B., Melsen, L. A., Moustakas, S., Nossent, J., Pereira, F., Sprokkereef, E., Stam, J., Weerts, A. H., Willems, P., Savenije, H. H. G., & Hrachowitz, M. (2021). Behind the scenes of streamflow model performance. *Hydrology and Earth System Sciences*, *25*(2), 1069–1095. https://doi.org/10.5194/hess-25-1069-2021

Braud, I., Ayral, P.-A., Bouvier, C., Branger, F., Delrieu, G., Coz, J. L., Nord, G., Vandervaere, J.-P., Anquetin, S., Adamovic, M., Andrieu, J., Batiot, C., Boudevillain, B., Brunet, P., Carreau, J., Confoland, A., Didon-Lescot, J.-F., Domergue, J.-M., Douvinet, J., … Wijbrans, A. (2014). Multi-scale hydrometeorological observation and modelling for flash flood understanding. *Hydrology and Earth System Sciences*, *18*(9), 3733–3761. https://doi.org/10.5194/hess-18-3733-2014

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chappell, N. A., Jones, T. D., Tych, W., & Krishnaswamy, J. (2017). Role of rainstorm intensity underestimated by data-derived flood models: Emerging global evidence from subsurface-dominated watersheds. *Environmental Modelling & Software*, *88*, 1–9. https://doi.org/10.1016/j.envsoft.2016.10.009

Clark, M. P., Kavetski, D., & Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research*, *47*(9), W09301. https://doi.org/10.1029/2010WR009827

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., & Hay, L. E. (2008). Framework for understanding structural errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research*, *44*(12), W00B02. https://doi.org/10.1029/2007WR006735

Coron, L., Delaigue, O., Thirel, G., Dorchies, D., Perrin, C., and Michel, C. (2020). airGR: Suite of GR Hydrological Models for Precipitation-Runoff Modelling. R package version 1.6.9.27. doi: https://doi.org/10.15454/EX11NA

Coron, L., Thirel, G., Delaigue, O., Perrin, C., & Andréassian, V. (2017). The suite of lumped GR hydrological models in an R package. *Environmental Modelling and Software*, *94*, 166–171. https://doi.org/10.1016/j.envsoft.2017.05.002

Coustau, M., Rousset-Regimbeau, F., Thirel, G., Habets, F., Janet, B., Martin, E., de Saint-Aubin, C., & Soubeyroux, J.-M. (2015). Impact of improved meteorological forcing, profile of soil hydraulic conductivity and data assimilation on an operational hydrological ensemble forecast system over France. *Journal of Hydrology*, *525*, 781–792. https://doi.org/10.1016/j.jhydrol.2015.04.022

de Boer-Euser, T., Bouaziz, L., Niel, J. D., Brauer, C., Dewals, B., Drogue, G., Fenicia, F., Grelier, B., Nossent, J., Pereira, F., Savenije, H., Thirel, G., & Willems, P. (2017). Looking beyond general metrics for model comparison—Lessons from an international model intercomparison study. *Hydrology and Earth System Sciences*, *21*(1), 423–440. https://doi.org/10.5194/hess-21-423-2017

de Lavenne, A., Andréassian, V., Thirel, G., Ramos, M.-H., & Perrin, C. (2019). A regularization approach to improve the sequential calibration of a semidistributed hydrological model. *Water Resources Research*, *55*(11), 8821–8839. https://doi.org/10.1029/2018WR024266

Delaigue, O., Génot, B., Lebecherel, L., Brigode, P., and Bourgin, P. (2020). Database of watershed-scale hydroclimatic observations in France. https://webgr.inrae.fr/base-de-donnees.

Estrany, J., Garcia, C., & Batalla, R. J. (2010). Hydrological response of a small mediterranean agricultural catchment. *Journal of Hydrology*, *380*(1–2), 180–190. https://doi.org/10.1016/j.jhydrol.2009.10.035

Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., & Savenije, H. H. G. (2013). A framework to assess the realism of model structures using hydrological signatures. *Hydrology and Earth System Sciences*, *17*(5), 1893–1912. https://doi.org/10.5194/hess-17-1893-2013

Fenicia, F., & McDonnell, J. J. (2022). Modeling streamflow variability at the regional scale: (1) perceptual model development through signature analysis. *Journal of Hydrology*, *605*, 127287. https://doi.org/10.1016/j.jhydrol.2021.127287

Fenicia, F., Meißner, D., & McDonnell, J. (2022). Modeling streamflow variability at the regional scale: (2) development of a bespoke distributed conceptual model. *Journal of Hydrology*, *605*, 127286. https://doi.org/10.1016/j.jhydrol.2021.127286

Ficchì, A. (2017). An adaptive hydrological model for multiple time-steps: Diagnostics and improvements based on fluxes consistency [Unpublished PhD thesis], UPMC, Paris, https://webgr.inrae.fr/publications/theses

Ficchì, A., Perrin, C., & Andréassian, V. (2016). Impact of temporal resolution of inputs on hydrological model performance: An analysis based on 2400 flood events. *Journal of Hydrology*, *538*, 454–470. https://doi.org/10.1016/j.jhydrol.2016.04.016

Ficchì, A., Perrin, C., & Andréassian, V. (2019). Hydrological modelling at multiple sub-daily time steps: Model improvement via flux-matching. *Journal of Hydrology*, *575*, 1308–1327. https://doi.org/10.1016/j.jhydrol.2019.05.084

Garambois, P., Larnier, K., Roux, H., Labat, D., & Dartus, D. (2014). Analysis of flash flood-triggering rainfall for a process-oriented hydrological model. *Atmospheric Research*, *137*, 14–24. https://doi.org/10.1016/j.atmosres.2013.09.016

Garavaglia, F., Lay, M. L., Gottardi, F., Garçon, R., Gailhard, J., Paquet, E., & Mathevet, T. (2017). Impact of model structure on flow simulation and hydrological realism: From a lumped to a semi-distributed approach. *Hydrology and Earth System Sciences*, *21*(8), 3937–3952. https://doi.org/10.5194/hess-21-3937-2017

Garçon, R. (1996). Prévision opérationnelle des apports de la durance à serre-ponçon à l'aide du modèle MORDOR. bilan de l'année 1994-1995. *La Houille Blanche*, *82*(5), 71–76. https://doi.org/10.1051/lhb/1996056

Green, W. H., & Ampt, G. A. (1911). Studies on soil phyics. *The Journal of Agricultural Science*, *4*(1), 1–24. https://doi.org/10.1017/S0021859600001441

Gupta, H. V., Wagener, T., & Liu, Y. (2008). Reconciling theory with observations: Elements of a diagnostic approach to model evaluation. *Hydrological Processes*, *22*(18), 3802–3813. https://doi.org/10.1002/hyp.6989

Herman, J. D., Reed, P. M., & Wagener, T. (2013). Time-varying sensitivity analysis clarifies the effects of watershed model formulation on model behavior. *Water Resources Research*, *49*(3), 1400–1414. https://doi.org/10.1002/wrcr.20124

Horton, R. E. (1933). The role of infiltration in the hydrologic cycle. *Transactions of the American Geophysical Union*, *14*(1), 446. https://doi.org/10.1029/TR014i001p00446

Hrachowitz, M., & Clark, M. P. (2017). HESS opinions: The complementary merits of competing modelling philosophies in hydrology. *Hydrology and Earth System Sciences*, *21*(8), 3953–3973. https://doi.org/10.5194/hess-21-3953-2017

Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., Freer, J., Savenije, H. H. G., & Gascuel-Odoux, C. (2014). Process consistency in models: The importance of system signatures, expert knowledge, and process complexity. *Water Resources Research*, *50*(9), 7445–7469. https://doi.org/10.1002/2014wr015484

Hugenschmidt, C., Ingwersen, J., Sangchan, W., Sukvanachaikul, Y., Duffner, A., Uhlenbrook, S., & Streck, T. (2014). A three-component hydrograph separation based on geochemical tracers in a tropical mountainous headwater catchment in northern Thailand. *Hydrology and Earth System Sciences*, *18*(2), 525–537. https://doi.org/10.5194/hess-18-525-2014

Jakeman, A. J., & Hornberger, G. M. (1993). How much complexity is warranted in a rainfall-runoff model? *Water Resources Research*, *29*(8), 2637–2649. https://doi.org/10.1029/93WR00877

Knoben, W. J. M., Freer, J. E., Peel, M. C., Fowler, K. J. A., & Woods, R. A. (2020). A brief analysis of conceptual model structure uncertainty using 36 models and 559 catchments. *Water Resources Research*, *56*(9), e2019WR025975. https://doi.org/10.1029/2019wr025975

Kuppel, S., Tetzlaff, D., Maneta, M. P., & Soulsby, C. (2018). EcH$_2$O-iso 1.0: Water isotopes and age tracking in a process-based, distributed ecohydrological model. *Geoscientific Model Development*, *11*(7), 3045–3069. https://doi.org/10.5194/gmd-11-3045-2018

Lana-Renault, N., Latron, J., & Regüés, D. (2007). Streamflow response and water-table dynamics in a sub-mediterranean research catchment (central pyrenees). *Journal of Hydrology*, *347*(3–4), 497–507. https://doi.org/10.1016/j.jhydrol.2007.09.037

Largeron, C., Cloke, H., Verhoef, A., Martinez-de-la Torre, A., and Mueller-Quintino, A. (2018). Impact of the representation of the infiltration on the river flow during intense rainfall events in Jules. Technical report, ECMWF, doi: https://doi.org/10.21957/nkky9s1hs.

Latron, J., & Gallart, F. (2008). Runoff generation processes in a small mediterranean research catchment (Vallcebre, eastern pyrenees). *Journal of Hydrology*, *358*(3–4), 206–220. https://doi.org/10.1016/j.jhydrol.2008.06.014

Le Moine, N. (2008). Le bassin versant de surface vu par le souterrain: une voie d'amélioration des performances et du réalisme des modèles pluie-débit ? [Unpublished PhD thesis], UPMC, Cemagref. https://webgr.inrae.fr/publications/theses

Ledoux, E., Girard, G., Marsily, G., Villeneuve, J. P., & Deschenes, J. (1989). Spatially distributed modeling: Conceptual approach, coupling surface water and groundwater. In *Unsaturated flow in hydrologic modeling* (pp. 435–454). Springer. https://doi.org/10.1007/978-94-009-2352-2_16

Leleu, I., Tonnelier, I., Puechberty, R., Gouin, P., Viquendi, I., Cobos, L., Foray, A., Baillon, M., & Ndima, P.-O. (2014). La refonte du système d'information national pour la gestion et la mise à disposition des données hydrométriques. *La Houille Blanche*, *100*(1), 25–32. https://doi.org/10.1051/lhb/2014004

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.

Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P., & Loumagne, C. (2014). When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events. *Hydrology and Earth System Sciences*, *18*(2), 575–594. https://doi.org/10.5194/hess-18-575-2014

Loritz, R., Hrachowitz, M., Neuper, M., & Zehe, E. (2021). The role and value of distributed precipitation data in hydrological models. *Hydrology and Earth System Sciences*, *25*(1), 147–167. https://doi.org/10.5194/hess-25-147-2021

Manus, C., Anquetin, S., Braud, I., Vandervaere, J.-P., Creutin, J.-D., Viallet, P., & Gaume, E. (2009). A modeling approach to assess the hydrological response of small mediterranean catchments to the variability of soil characteristics in a context of extreme events. *Hydrology and Earth System Sciences*, *13*(2), 79–97. https://doi.org/10.5194/hess-13-79-2009

Mathevet, T., Michel, C., Andréassian, V., & Perrin, C. (2006). A bounded version of the Nash-sutcliffe criterion for better model assessment on large sets of basins. *IAHS Publication*, *307*, 211–219. https://hal.inrae.fr/hal-02588691

McDonnell, J. J., Spence, C., Karran, D. J., van Meerveld, H. J. I., & Harman, C. J. (2021). Fill-and-spill: A process description of runoff generation at the scale of the beholder. *Water Resources Research*, *57*(5), e2020WR027514. https://doi.org/10.1029/2020wr027514

McMillan, H., Booker, D., & Cattoën, C. (2016). Validation of a national hydrological model. *Journal of Hydrology*, *541*, 800–815. https://doi.org/10.1016/j.jhydrol.2016.07.043

McMillan, H., Westerberg, I., & Branger, F. (2017). Five guidelines for selecting hydrological signatures. *Hydrological Processes*, *31*(26), 4757–4761. https://doi.org/10.1002/hyp.11300

Melsen, L. A., Addor, N., Mizukami, N., Newman, A. J., Torfs, P. J. J. F., Clark, M. P., Uijlenhoet, R., & Teuling, A. J. (2018). Mapping (dis)agreement in hydrologic projections. *Hydrology and Earth System Sciences*, *22*(3), 1775–1791. https://doi.org/10.5194/hess-22-1775-2018

Mishra, S. K., Jain, M. K., & Singh, V. P. (2004). Evaluation of the SCS-CN-based model incorporating antecedent moisture. *Water Resources Management*, *18*(6), 567–589. https://doi.org/10.1007/s11269-004-8765-1

Mizukami, N., Rakovec, O., Newman, A. J., Clark, M. P., Wood, A. W., Gupta, H. V., & Kumar, R. (2019). On the choice of calibration metrics for "high-flow" estimation using hydrologic models. *Hydrology and Earth System Sciences*, *23*(6), 2601–2614. https://doi.org/10.5194/hess-23-2601-2019

Moussa, R., & Chahinian, N. (2009). Comparison of different multi-objective calibration criteria using a conceptual rainfall-runoff model of flood events. *Hydrology and Earth System Sciences*, *13*(4), 519–535. https://doi.org/10.5194/hess-13-519-2009

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

Nijzink, R. C., Almeida, S., Pechlivanidis, I. G., Capell, R., Gustafssons, D., Arheimer, B., Parajka, J., Freer, J., Han, D., Wagener, T., Nooijen, R. R. P., Savenije, H. H. G., & Hrachowitz, M. (2018). Constraining conceptual hydrological models with multiple information sources. *Water Resources Research*, *54*(10), 8332–8362. https://doi.org/10.1029/2017wr021895

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *Journal of Hydrology*, *303*(1), 290–306. https://doi.org/10.1016/j.jhydrol.2004.08.026

Pang, S., Wang, X., Melching, C. S., & Feger, K.-H. (2020). Development and testing of a modified SWAT model based on slope condition and precipitation intensity. *Journal of Hydrology*, *588*, 125098. https://doi.org/10.1016/j.jhydrol.2020.125098

Parajka, J., Merz, R., & Blöschl, G. (2007). Uncertainty and multiple objective calibration in regional water balance modelling: Case study in

320 austrian catchments. *Hydrological Processes*, *21*(4), 435–446. https://doi.org/10.1002/hyp.6253

Payrastre, O., Nicolle, P., Bonnifait, L., Brigode, P., Astagneau, P. C., Baise, A., Belleville, A., Bouamara, N., Bourgin, F., Breil, P., Brunet, P., Cerbelaud, A., Courapied, F., Devreux, L., Dreyfus, R., Gaume, E., Nomis, S., Poggio, J., Pons, F., … Sevrez, D. (2022). The 2nd October 2020 Alex storm in South-Eastern France: A contribution of the scientific community to the flood peak discharges estimation. *Hydroscience Journal*. https://doi.org/10.1080/27678490.2022.2082891

Peredo, D., Ramos, M.-H., Andréassian, V., & Oudin, L. (2022). Investigating hydrological model versatility to simulate extreme flood events. *Hydrological Sciences Journal*, *67*, 628–645. https://doi.org/10.1080/02626667.2022.2030864

Perrin, C., Andréassian, V., Serna, C. R., Mathevet, T., & Moine, N. L. (2008). Discrete parameterization of hydrological models: Evaluating the use of parameter sets libraries over 900 catchments. *Water Resources Research*, *44*(8), W08447. https://doi.org/10.1029/2007WR006579

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

Perrin, C., Oudin, L., Andréassian, V., Rojas-Serna, C., Michel, C., & Mathevet, T. (2007). Impact of limited streamflow data on the efficiency and the parameters of rainfall—Runoff models. *Hydrological Sciences Journal*, *52*(1), 131–151. https://doi.org/10.1623/hysj.52.1.131

Pool, S., Vis, M., & Seibert, J. (2021). Regionalization for ungauged catchments—Lessons learned from a comparative large-sample study. *Water Resources Research*, *57*(10), e2021WR030437. https://doi.org/10.1029/2021wr030437

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rakovec, O., Kumar, R., Attinger, S., & Samaniego, L. (2016). Improving the realism of hydrologic model functioning through multivariate parameter estimation. *Water Resources Research*, *52*(10), 7779–7792. https://doi.org/10.1002/2016wr019430

Ruiz-Villanueva, V., Borga, M., Zoccatelli, D., Marchi, L., Gaume, E., & Ehret, U. (2012). Extreme flood response to short-duration convective rainfall in south-West Germany. *Hydrology and Earth System Sciences*, *16*(5), 1543–1559. https://doi.org/10.5194/hess-16-1543-2012

Saadi, M. (2020). Représentation de l'urbanisation dans la modélisation hydrologique à l'échelle du bassin versant [Unpublished PhD thesis], Sorbonne Université, https://tel.archives-ouvertes.fr/tel-03250292.

Saadi, M., Oudin, L., & Ribstein, P. (2019). Random forest ability in regionalizing hourly hydrological model parameters. *Watermark*, *11*(8), 1540. https://doi.org/10.3390/w11081540

Savenije, H. H. G. (2010). HESS opinions: "Topography driven conceptual modelling (FLEX-topo)". *Hydrology and Earth System Sciences*, *14*(12), 2681–2692. https://doi.org/10.5194/hess-14-2681-2010

SCS. (1956). *National engineering handbook*. Sois Conservation Service USDA.

Seibert, J., & Vis, M. J. P. (2012). Teaching hydrological modeling with a user-friendly catchment-runoff-model software package. *Hydrology and Earth System Sciences*, *16*(9), 3315–3325. https://doi.org/10.5194/hess-16-3315-2012

Stein, L., Clark, M. P., Knoben, W. J. M., Pianosi, F., & Woods, R. A. (2021). How do climate and catchment attributes influence flood generating processes? A large-sample study for 671 catchments across the contiguous USA. *Water Resources Research*, *57*(4), e2020WR028300. https://doi.org/10.1029/2020wr028300

Tabary, P., Dupuy, P., L'Henaff, G., Gueguen, C., Moulin, L., Laurantin, O., Merlier, C., & Soubeyroux, J.-M. (2012). *A 10-year (1997–2006) reanalysis of quantitative precipitation estimation over France: Methodology and first results* (Vol. 351, pp. 255–260). IAHS Publication.

Tarasova, L., Basso, S., Wendi, D., Viglione, A., Kumar, R., & Merz, R. (2020). A process-based framework to characterize and classify runoff events: The event typology of Germany. *Water Resources Research*, *56*(5), e2019WR026951. https://doi.org/10.1029/2019wr026951

Thirel, G., Martin, E., Mahfouf, J.-F., Massart, S., Ricci, S., & Habets, F. (2010). A past discharges assimilation system for ensemble streamflow forecasts over France—Part 1: Description and validation of the assimilation system. *Hydrology and Earth System Sciences*, *14*(8), 1623–1637. https://doi.org/10.5194/hess-14-1623-2010

Thirel, G., Martin, E., Mahfouf, J.-F., Massart, S., Ricci, S., Regimbeau, F., & Habets, F. (2010). A past discharge assimilation system for ensemble streamflow forecasts over France—Part 2: Impact on the ensemble streamflow forecasts. *Hydrology and Earth System Sciences*, *14*(8), 1639–1653. https://doi.org/10.5194/hess-14-1639-2010

Tramblay, Y., & Somot, S. (2018). Future evolution of extreme precipitation in the mediterranean. *Climatic Change*, *151*(2), 289–302. https://doi.org/10.1007/s10584-018-2300-5

Tyralis, H., Papacharalampous, G., & Langousis, A. (2019). A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Watermark*, *11*(5), 910. https://doi.org/10.3390/w11050910

Uchida, T., van Meerveld, I. T., & McDonnell, J. J. (2005). The role of lateral pipe flow in hillslope runoff response: An intercomparison of nonlinear hillslope response. *Journal of Hydrology*, *311*(1–4), 117–133. https://doi.org/10.1016/j.jhydrol.2005.01.012

van Esse, W. R., Perrin, C., Booij, M. J., Augustijn, D. C. M., Fenicia, F., Kavetski, D., & Lobligeois, F. (2013). The influence of conceptual model structure on model performance: A comparative study for 237 french catchments. *Hydrology and Earth System Sciences*, *17*(10), 4227–4239. https://doi.org/10.5194/hess-17-4227-2013

van Meerveld, H. J. T., & McDonnell, J. J. (2006). Threshold relations in subsurface stormflow: 1. A 147-storm analysis of the panola hillslope. *Water Resources Research*, *42*(2), W02410. https://doi.org/10.1029/2004WR003778

Verma, R. K., Verma, S., Mishra, S. K., & Pandey, A. (2021). SCS-CN-based improved models for direct surface runoff estimation from large rainfall events. *Water Resources Management*, *35*(7), 2149–2175. https://doi.org/10.1007/s11269-021-02831-5

Verma, S., Singh, P., Mishra, S., Singh, V., Singh, V., & Singh, A. (2020). Activation soil moisture accounting (ASMA) for runoff estimation using soil conservation service curve number (SCS-CN) method. *Journal of Hydrology*, *589*, 125114. https://doi.org/10.1016/j.jhydrol.2020.125114

Vidal, J.-P., Martin, E., Franchistéguy, L., Baillon, M., & Soubeyroux, J.-M. (2010). A 50-year high-resolution atmospheric reanalysis over France with the Safran system. *International Journal of Climatology*, *30*(11), 1627–1644. https://doi.org/10.1002/joc.2003

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1*(6), 80. https://doi.org/10.2307/3001968

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. https://doi.org/10.1016/j.jhydrol.2014.01.017

Yilmaz, K. K., Gupta, H. V., & Wagener, T. (2008). A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resources Research*, *44*(9), W09417. https://doi.org/10.1029/2007wr006716

Zhang, X., Zwiers, F. W., Li, G., Wan, H., & Cannon, A. J. (2017). Complexity in estimating past and future extreme short-duration rainfall. *Nature Geoscience*, *10*(4), 255–259. https://doi.org/10.1038/ngeo2911

How to cite this article: Astagneau, P. C., Bourgin, F., Andréassian, V., & Perrin, C. (2022). Catchment response to intense rainfall: Evaluating modelling hypotheses. *Hydrological Processes*, *36*(8), e14676. https://doi.org/10.1002/hyp.14676

APPENDIX A

A.1 | A Cluster characteristics

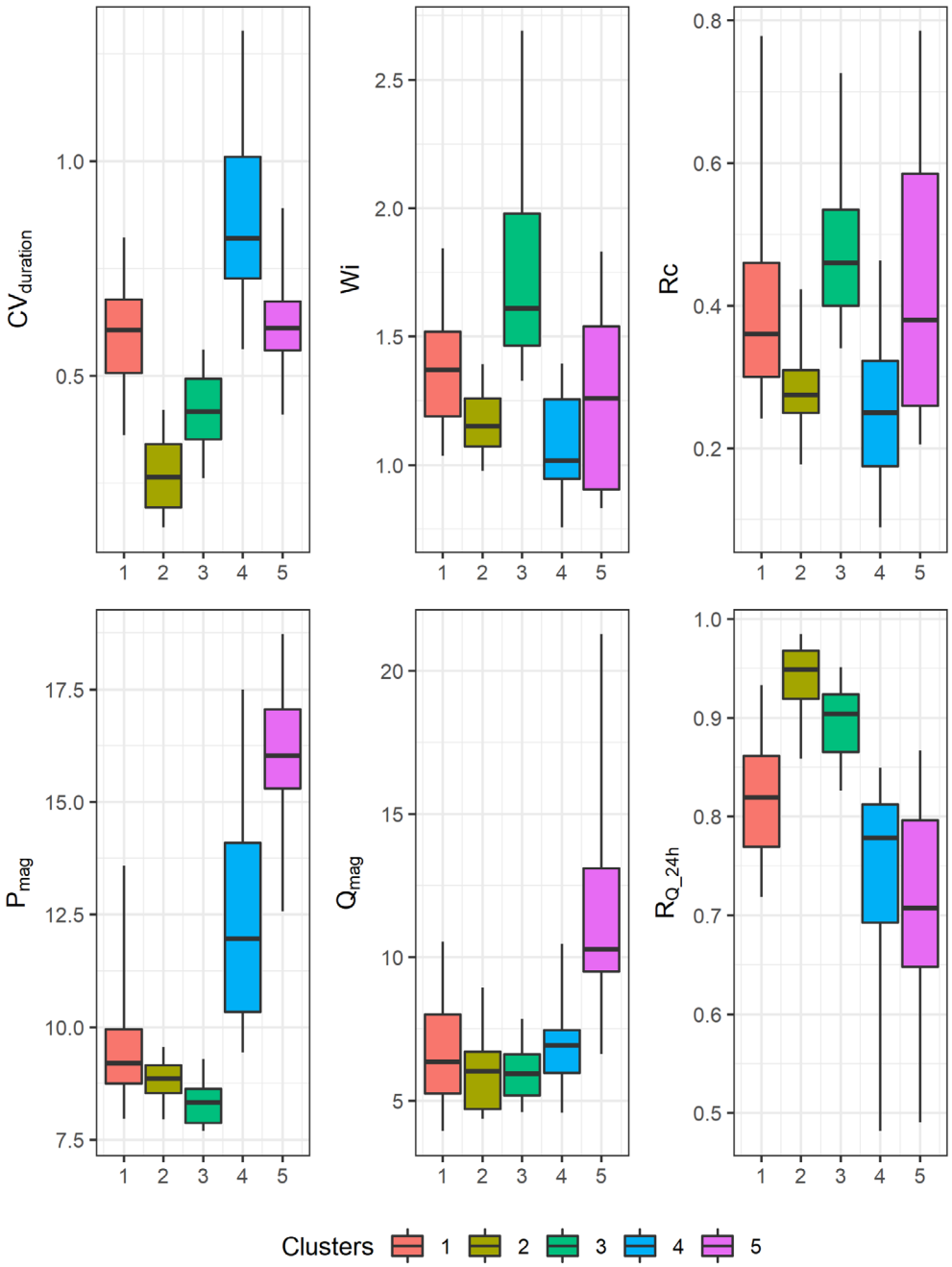

FIGURE A1 Distribution of six characteristics of the five hydroclimatic groups of catchments. The distributions are presented between the 5th and 95th percentiles