We wish to thank the referee for their careful evaluation of the manuscript which would be hopefully more intelligible thanks to their constructive comments. Below are our detailed responses (in blue) to the comments (in black).

Sincerely yours,

———————————————————————————————

This paper proposes a methodology for the analysis of catchment hydrological behaviors during flash floods, based on the introduction and comparison of several hypotheses in a distributed hydrological model. This topic is of broad interest for the hydrological scientific community, and is fully relevant in my opinion for a publication in HESS. However, in its current form the paper suffers from a lack of detail and explanations on several aspects (calibration procedure, explanations related to some figures, ..), causing difficulties for a detailed understanding of the research content. The presentation of sections 4 and 5 should particularly be improved in my opinion (and maybe organized in a slightly different way) to facilitate the overall understanding of the results and related analyses. The paper well illustrates the difficulties in the interpretation of modelling results, due to equifinality issues and lack of internal observations to confirm the nature of the main hydrological processes. Therefore, even if some solutions to cope with these difficulties are proposed here, I think the conclusions relative to the catchments behaviors (section 5.6) should finally be relativized and presented in the discussion section as the most reasonable assumptions, provided the modelling results obtained here.

We are grateful for the constructive comments. Most of them shift toward quite deep modifications of the sections 4 and 5. Following the reviewer's suggestions, the organization of section 5 has been changed, and many efforts were made to describe the methodology (section 4). The answers to the comments are listed below. The modifications made in the paper are quoted in italics. Line references correspond to the unmarked manuscript (attached file). Please note also that the new version of the manuscript proposed here integrates as well the modification related to the comments of the other referee.

———————————————————————————————

Specific comments:
The abstract is very short and could be slightly more detailed. The abstract has been reworded, giving now more details about the results (page

1

1, lines 5 - 10 :). Below, the reworded abstract :

"A method of multiple working hypotheses was applied to a range of catchments in the Mediterranean area to analyse different types of possible flow dynamics in soils during flash flood events. The distributed, process-oriented model, MARINE, was used to test several representations of subsurface flows, including flows at depth in fractured bedrock, and flows through preferential pathways in macropores. Results showed contrasted perfomances of the submitted models, revealing different hydrological behaviours along the catchment set. The benchmark study offered a characterization of the catchments reactivity through the description of the hydrographs formation. The quantification of the different flow processes (surface, intra soil flows) were consistent with the scarse in-situ observations but remains uncertain, as a result of equifinality issue. The spatial description of the simulated flows over the catchments, made available by the model, enabled to spot counterbalancing effects between internal flow processes, including the compensation for the water transit time in the hillslopes and in the drainage network. New insights are finally proposed into strategical monitoring and calibration constraints setting up."

References would be welcome in section 1.2. References to relevant publications are added page 1-2, lines 18 - 5. (note that according to the comments of the second reviewer, this section has been merge with the section 1.1 to reduce the introduction.

Comments about the section 4:

The description of the calibration procedure (section 4.1) and of the metrics for evaluation (section 4.2) are not sufficiently clear in my opinion, and should be improved:

- Please indicate how the "confidence intervals" are obtained for observations ($y_i \pm 2\sigma_i$?) and also for modelling results. This should clarify why the uncertainties ranges mentioned in the text (respectively 20% for observations and 10% for modelling errors) are consistent with eq.(6) and eq.(7). The definition of the confidence interval of the observed flows is now explicitly written page 12, lines 8 - 10 : *"The envelop $\left((\hat{y}_i \pm 2\sigma_{\hat{y}_i}), \ i = 1...n \right)$ consequently defines the 95 % confidence interval of the observed flows."*

- please clarify the reason why the metric used for evaluation (Qmed_INT) is different from the one used for calibration (DEC) ? The Qmed_INT is here used as the meaning of this criteria is easier to interpret and

2

understand, compared to the DEC value that is a standard error. This is clarified, page 13, lines 20-25: *"While the DEC score has provided a standard assessment of the modelling errors enabling a reasonable weighting of the simulations, for a sake of easy understanding ,the percentage of acceptable points of the simulated median time series - Qmed_INT [%] (Douinot et al. 2017) - was chosen to evaluate the ability of the models to reproduce overall flows, rising flood waters and high discharges. A point is defined as acceptable when the median simulated value stands within the modelling acceptability zone, the latter one being determined by σ_{mod} et $\sigma_{\hat{y}}$"*

- the definition of the "acceptability zone" should be provided ($y_i \pm 2\sigma_{y_i} \pm 2\sigma_{mod,i}$?). A definition of the acceptability zone is added page 14, lines 5-21 : *"Finally, the overall overarching envelop $\left((\hat{y}_i \pm 2\sigma_{\hat{y}_i} \pm 2\sigma_{mod,i}),\ i = 1...n \right)$ defines hereafter the acceptability zone, that is to say the interval into which any simulated flow would be considered as acceptable, according to the modelling and measurement uncertainty definitions. "*

- the "a priori" and "a posteriori" modelling errors are not defined. This clearly limits the interpretation of figure 11 (see hereafter). Page 17, lines 5-20 : In order to clarify the variables used in the figure 11 and the related comments, 3 paragraphs were added in the section 4.2 (Metrics and key points in model evaluation and comparison):

The evaluation was then completed through the description of the modelling errors (section 5.2). The objective was to identify those that were inherent in the choice of model structure, regardless of the calibration methodology adopted. In that respect, attention was paid on the a priori and a posteriori confidence interval of the model simulations respectively defined by $\left([y_i^{prior-5th}, y_i^{prior-95th}],\ i = 1...n \right)$ and $\left([y_i^{DEC-5th}, y_i^{DEC-95th}],\ i = 1...n \right)$ where $y_i^{prior-5th}$ and $y_i^{prior-95th}$ are the 5th and the 95th percentile of the 5000 model simulation values at time i, and where $y_i^{DEC-5th}$ and $y_i^{DEC-95th}$ are the 5th and the 95th percentile of the same but weighted series according to the DEC calibration criterion.

Those confidence intervals were standardized according to the DEC modelling error definition (equation 10), respectively defining the a priori and a posteriori confidence intervals of the modelling errors:

$$\epsilon_i^{\alpha-xth} = \begin{cases} 0 & if \quad \mid y_i^{\alpha-xth} \mid \leq 2 \cdot \sigma_{\hat{y}_i} \\ \frac{y_i^{\alpha-xth} \pm 2 \cdot \sigma_{\hat{y}_i}}{2 \cdot \sigma_{mod_i}} & otherwise \quad (- if\ y_i^{\alpha-xth} > 0\ ;\ + if\ y_i^{\alpha-xth} \leq 0) \end{cases}$$

with $\epsilon_i^{\alpha-xth}$ is the xth percentile of the α modelling errors distribution at time i.

The latter definition allows for an informative translation of the prior and posterior confidence intervals (Douinot et al. 2017): a value of $\epsilon_i^{\alpha-xth}$ equal to 0 indicates that the $y_i^{\alpha-xth}$ bound lies within the discharge confidence interval; if $0 < \epsilon_i^{\alpha-xth} \leq 1$, the $y_i^{\alpha-xth}$ bound lies within the acceptability zone; and if $\epsilon_i^{\alpha-xth}$ is larger than 1 then errors of modelling is detected or remained. In addition, the benchmark of both a priori and a posteriori confidence intervals allows for highlighting which were the remaining modelling errors that were induced by the model's assumptions. For those reasons, $\epsilon_i^{\alpha-xth}$ were used as the baseline of the modelling errors analysis.

Comments about the section 5 (results)

According to the comments, the presentation of the results has been reorganized as follow:

- section 5.1: Performance of the models. In these section are exclusively presented the assessment of the models through the metric scores (that are defined in the section 4.2).

  - section 5.1.1: Overall performances of the models. It merges the paragraph that has been written into the previous section 5.1.1 and 5.1.2

  - section 5.1.2: Detailed performances of the models: assessment of the models when simulating the different stages of an hydrograph. It contains the previous section 5.1.3 and 5.1.4

- section 5.1.3: Summary of the assessment: This part has been added, in order to present a global overview of the results after detailed comments in the aforementioned sections.

- section 5.2: Modelling errors inherent in the models'structures: It contain the previous 5.1.5 section.

- section 5.3: Analysis of relevance of the internal hydrological processes simulated: As suggested, the previous sections 5.2 and 5.3 have been merged into one element.

- section 6: Discussion: We propose a novel section in oder to separate the strict description of the results (section 5), and the interpretation done from it (section 6). It finally contains the previous sections 5.1.6 and 5.4.

The presentation of modelling results (section 5.1) could also be enhanced : I think the separated presentation of each metric (overall hydrograph, rising limb, high-discharges, recession) does not help to give a synthetic overview of results. It seems that three main situations can be distinguished here: clear hierarchy (Gardons and Salz), contrasted hierarchy (Ardche), or no clear hierarchy (equi- finality, Hrault). These three situations could be illustrated based on a common analysis of all the metrics. The comments about the performances of the models were mainly reworded. In the novel version, we separated the presentation of global metric (overall hydrograph, section 5.1.1) from those that are focused on a specific stage of the hydrograph (section 5.1.2). The objective of this organization is to highlight the differences between what we learned with a global point of view, and what we learn if we focus on the representation of one part of the hydrograph. In addition, the figures 8 and 10 were modified in order to support the new organization of the section. The figure 8 (now figure 6, page 15) presents only the global performances while the figure 10 (now figure 8, page 17) presents the detailed performances. Finally, for a sake of clarity, a summary is done in a last subsection (section 5.1.3). Those modifications can be find from page 14 line 22, to page 18 line 10.

Explanations in section 5.1.4 (now 5.1.2) are poorly supported by figure 10 in my opinion. Please try to clarify this section and figure. We modified the way to assess the good simulation of the flood recession, using another metric score. This is defined page 14 from lines 25 to line 5 :

"Conversely, Qmed_INT was not relevant for the evaluation of the capacity to reproduce recessions, because the calculation of this score during the recession interval strongly depends on performance at high discharges. Instead, we used the A_{slope} score defined in the equation 9. It calculates the average standard error in simulating the decreasing rate of the discharge during the flood recession interval. Through the consideration of the A_{slope} score here, it is assumed that the recession rate is a relevant feature of the catchment's hydrologic properties Troch et al., 2013; Kirchner, 2009. We therefore choose to make a visual comparison of the simulated and observed recession curves, $Q(t) = f\left(log\left(-\frac{dQ(t)}{dt}\right)\right)$, which are characteristic of a catchment's hydraulic discharge properties.

$$A_{slope} = \frac{\sum_{i=k}^{l} |\frac{dy_i}{dt} - \frac{d\hat{y}_i}{dt}|}{\sum_{i=k}^{l} \frac{d\hat{y}_i}{dt}} \quad (1)$$

where $\frac{d\hat{y}_i}{dt}$ and $\frac{dy_i}{dt}$ are respectively the observed and the simulated recession rate at a time step i which belongs to the flood recession interval $(i = k...l)$.

The assessment focused on the simulation of the recession is then presented in a similar way than those on the simulation of the rising flood waters and

the high flows. Consequently, the three assessments are presented in a same figure (figure 8).

The analysis proposed in section 5.1.5 (now 5.2) is also difficult to follow based on figure 11, which does not well illustrate in my opinion the differences in models behaviors. Figure 11 indeed is difficult to understand: ♯ DDEC is not defined, the definition of prior and posterior errors is again missing. Moreover, is not clear why the width of the acceptability zone does not vary with y_i (not consistent with equation (7)). Please try to clarify this section and figure.
In order to clarify this figure (figure 10 now), definition of the specific variables are written in the section 4.2, page 14, lines 6-21. In addition several comments were added to better link the underlying description of the figure with the possible interpretation of the modelss performances (page 18-19).

Section 5.1.6: (now 6.1) I think this interpretation on catchment behaviors arrives too early here. I think it would be better to put this in the discussion section, and to present these analyses as plausible assumptions, according to the modelling results.
As suggested, the comment of the previous section 5.1.6 are now the basis of subsection 6,1 of the discussion section.

Section 5.2 (now 5.3) may be renamed in a more explicit way, such as: "Analysis of relevance of the internal hydrological processes simulated". It could include both considerations on proportion of surface runoff (current section 5.2), and detailed analysis of velocities and water contents in the case of Hérault (current section 5.3)
As said below, we incorporated the suggestion into the new organization of the result section.

Comments about the conclusion

"For each catchment, the best performing models were those where results reflected the available knowledge and observations on the overall hydrological functioning of the catchments ...". Actually, it seems that very limited information is available on the real hydrological behavior, excepted maybe for the Gardon where detailed measurements were performed. Therefore, I would rather conclude that the modelling results help to draw consistent assumptions on hydrological behaviors, that can in some (rare) cases be confirmed by the existing knowledge and local observations.
As suggested this conclusion was reworded, taking into consideration that we actually have very limited information on those catchment (hence the

interest of the study) page 29, lines 19-21 :

"The modelling results help to draw consistent assumptions on hydrological behaviours, which corroborate when available, the knowledge and observations on the overall hydrological functioning of the catchments, or the experimental estimations of flow processes."

"distinction in hydrological behaviours revealed between the catchment of the Gardons and the Ardèche may explain that taking into account the spatial nature of precipitation in a flash flood forecasting method results in an improvement only on the Gardon and not on the Ardèche ..." I think this conclusion is not really supported by the content of the paper. Moreover, another explanation could just be a difference in the rainfall spatial variability, which seems to be more pronounced in the Gardons catchment for climatic reasons.

This last statement is introduced as an open conclusion about the potential value of the study results facing to the flash flood forecasting issue. Revealing the contrasted hydrological behaviours of the Gardons and the Ardche catchments - the first one clearly more reactive that the second one it might shift towards different considerations when setting up a flash flood forecasting method over those contrasted catchments. We referred to the Douinot et al, 2016 study as it actually corroborates the fact that different considerations should be done, to develop a flood forecasting method. It shows contrasted sensitivities of the catchments to the rainfall spatial variability, which could either be a consequence of the contrasted hydrological behaviours of the catchments revealed here, or we agree - be due to contrasted climatic forcing. We suggested to reword the statement as following (page 30, lines 4-11):

"Lastly, identifying the most pertinent hydrological models for each catchment enables the key elements in the generation of flash floods to be highlighted, which, in turn, could serve to further develop methods for forecasting flash floods. For example, distinctions in hydrological behaviour revealed between the catchments of the Gardon and the Ardèche - the first one appearing more reactive with important runoff and subsurface flows through preferential flowpaths - might shift towards different considerations when setting up a flash flood forecasting method. It corroborates the results of Douinot et al 2016, which highlighted contrasted impacts of taking into account the spatial variability of precipitation in a flash flood forecasting method. These contrasted impacts can indeed be explained by the more pronounced spatial variability of the rainfall over the Gardon catchment, but also by the local more pronounced dynamic of the soil water content in the Gardon catchment revealed in the present study."

Technical corrections

- Section 2.2: The reference Ministre de lEcologie (2015) just corresponds to an URL, which could be added directly in the text. The modification has been done (page 6, table 1 and page 7, line 3).

- "These measurements were calibrated by forecaters at the French Flood Forecasting service by monitoring a network of rain gauges ...". Sentence not clear, please reformulate. The sentence has been reworded, page 17, lines 15-18, as follow : *"The French flood forecasting service (SCHAPI: Service central d'hydrométéorologie et d'appui à la prévision des inondations) used then the CALAMAR patented software (Badoche-Jacquet et al. 1992) to produce rainfall depth data by combining these radar measurements with raingauge data."*

- Figure 5 (now figure 3): are θ_s and θ_i really the current and initial water contents respectively. Shouldnt rather θ_s be the saturation water content? This is true. The error has been corrected (figure 3, page 9).

- Section 3.2, description of the modelling principles: the equations (1) to (4) and description of variables should be placed in the text with reference to figures 6 and 7. The equations and the description of the variables were inserted into the text: page 10, lines 15-20 and page 11, lines 7-15; 21.

- Section 4.2: y and mod rather than y and mod. The modification has been done on page 13, line 24.

- Section 5.1.5: the variation interval of the modelling errors: I dont really understand, please define this. The expression has been reworded, page 19, line 5 as follow: *"the width of the confidence interval of the modelling"*.

- I finally suggest to check the overall quality of English. The text has been proofread by a professional translator.

We wish to thank the referee for his careful evaluation of the manuscript. Please find below, the details responses (in blue) to the comments (in black). Please note also that the new version of the manuscript proposed here integrates as well the modification related to the comments of the other referee. The mentioned line refer to the unmarked manuscript.

Sincerely yours,

Audrey Douinot

The article presents the test of three versions of an event-based model (MARINE) on Mediterranean catchments in France. The authors investigate the impact of the subsurface flow and deep infiltration on model response using three modelling alternatives. They try to relate their results to the a priori knowledge on hydrological processes on the studied catchments. First, I found that the originality of the proposed methodology is not clearly explained compared to existing works. Second, my main concern is that the results and discussion section (section 5) is excessively long and verbose. The authors discuss all the results with great detail, but the reader gets lost in all the information provided (at least I got lost). At the end, it is a bit difficult to extract the main findings. I suggest reducing the size of this section to highlight the most important results. Besides, I found that the discussion on the link between results and the a priori knowledge on processes remains very qualitative. Though the explanations are sensible, there is no clear demonstration that the results are actually the consequence of the perceptual knowledge on processes invoked by the authors. There are so many possible causes to explain modelling results. I found that the reasons found by the authors only remain hypotheses and should be more clearly presented this way.

The general comments are mostly related to the unclear presentation of the results and the insights of the work. As those comments are quite similar to the those from the other referee, deep modifications have been done on the last sections (section 5 and the conclusion):

- the section 5 has been splitted into a section "results" that exactly describes the results when applying the method presented in section 4; and a section "discussion" where those results are interpreted, as the witnesses of the hydrological behavior of the catchment set.

- the conclusion was reworded to clarify which insights are clearly demonstrated from those that are proposals, that still need to be checked with additional field observations.

Specific comments:

1. General: Though the English writing is generally good, some sentences remain unclear. I suggest that the article be checked and corrected by native English.

The text has been proofread by a professional translator, native english.

2. Abstract: The main results are summarized in three lines. I find it difficult to fully understand what was done in the article by reading the abstract only.

The abstract has been reworded, giving now more details about the results (page 1, lines 5 - 10). Below, the reworded abstract :

"A method of multiple working hypotheses was applied to a range of catchments in the Mediterranean area to analyse different types of possible flow dynamics in soils during flash flood events.

1

The distributed, process-oriented model, MARINE, was used to test several representations of sub-surface flows, including flows at depth in fractured bedrock, and flows through preferential pathways in macropores. Results showed contrasted perfomances of the submitted models, revealing different hydrological behaviours along the catchment set. The benchmark study offered a characterization of the catchments reactivity through the description of the hydrographs formation. The quantification of the different flow processes (surface, intra soil flows) were consistent with the scarse in-situ observations but remains uncertain, as a result of equifinality issue. The spatial description of the simulated flows over the catchments, made available by the model, enabled to spot counterbalancing effects between internal flow processes, including the compensation for the water transit time in the hillslopes and in the drainage network. New insights are finally proposed into strategical monitoring and calibration constraints setting up."

3. Sections 1.1 and 1.2: This introduction of the context is interesting but quite classical and does not really bring essential material to understand the work done. I suggest reducing these parts to a few lines only.

The sections 1.1 and 1.2 - that is to say the beginning of the article - has been reduced as follow (page 1 and 2, from lines 15 to 26). Below, the reworded introduction:

Flash floods are "sudden floods with high peak discharges, produced by severe thunderstorms that are generally of limited areal extent". (IAHS-UNESCO-WMO (1974); Garambois (2012); Braud et al. (2014)). They are often linked to localised and major forcings (greater than 100 mm, Gaume et al. (2009)) at the heads of steep-sided, meso-scale catchments (with surface areas of 10-250 km^2).

The large specific discharges, and intensities of precipitation, makes the flash floods being classified as extreme. Nevertheless, those events are not scarce nor unusual since on average, there were no fewer than five flash floods a year on the Mediterranean Arc between 1958 and 1994 (Jacq, 1994), and they tend to be amplifed against a background of climate change (Llasat et al., 2014; Colmet Daage et al., 2016). Flash floods constitute a significant hazard and, therefore, a considerable risk for populations (UNISDR 2009, Llasat et al. (2014)). They are particularly dangerous due to their characteristics: (i) the suddenness of events makes it difficult to warn populations in time, and can lead to panic, thus increasing risk, when a population is unprepared (Ruin et al., 2008); ii) the traditional connected monitoring system are not adapted to the temporal and spatial scales of the flash floods (Borga et al., 2008; Braud et al., 2014); iii) the magnitude of floods implies significant amounts of kinetic energy, which can transform transitory rivers into torrents, resulting in the transport of debris ranging from fine sediments to tree trunks, as well as the scouring of river beds and the erosion of banks (Borga et al., 2014).

A major area of interest for flash floods is, therefore, better risk assessment, to enable them to be forecast and the relevant populations to be pre-warned. Greater knowledge and understanding is required to better identify the determining factors that result in flash floods. In particular, in order to implement a regional forecasting system, the properties of the catchments, and the climatic forcing and linkages between them which lead to flash flood events need to be characterised.

4. Section 1.3 (now 1.2): This section appears to be mostly centered on the French context. A more general perspective could be given to this literature review.

This is true. The bibliography here was actually quite consciously centered to the North-West Mediterranean context, as behind the "flash flood" term, there are different types of hydro-meteorological events, basically depending on the area over the world concerned. As example in Europe, events from the South East of France, Northern Italy and North Eastern Spain (Catalonia) are characterized by more intensive rainfall during larger time extents, and exhibit different climatic conditions than the flash floods occuring in the east part of Europe (along the Carpathian

mountains and the continental Alps region, Tarolli et al. (2012)). As our study is specially focused on catchments of the French Mediterranean area, the bibliography similarly took interest in flash flood events with similar meteorological context. Nevertheless, some references has been added to offer a first broader overview over the current researches on the topic, and to complete the statements found in the litterature on flash flood processing over the North-West Meditterranean area. This focus, on this particular content has also been specified (page 2, lines 21-22).

5. P4, L15-20: Clark et al. (2015a; 2015b) also proposed the SUMMA framework, applicable to distributed models. The authors should more clearly explain what is new and original in the approach they propose compared to these past works.

The proposed study relies on the cited approaches. The originality is to apply the approach on an event-based model, that - as far as we know - has not been done before. The objective here is to bring insights on hydrological understanding in a specific case (flash flood processing in the North-Westhern Mediterranean area) using an alike promising approach described in the past works. I added the reference you suggested, as clearly missing (page 3, line 24-29). In addition, the bold font, highlighting the objectives of the paper has been modified (It would have lead to confusion).

6. Section 2.1: This section could be presented in a more synthetic way, which would help the reader to more easily compare the study catchments. I suggest not repeating in the text information already contained in Table 1.

The description of the catchments has been reduced to the description of the contrasted geology - as the most significant information with the objective of the study, and to the current knowlege on the hydrological processes of the studied area obtained through field experimental studies. You will find the reworded subsection in page 4-5, lines 15-15, or here below:

The main physiographical and hydrological properties of the catchments are presented in Table 1. Figure 2 shows the contrasted geological properties of the studied area : the catchments are marked by a clear upstream / downstream difference. The Ardèche catchment upstream of Ucel sits essentially on a granite bedrock with some sandstone on its edges, while downstream, the geology changes to a predominantly schist and limestone formations. Similarly, the upstream part of the Gardon catchment consists of schistose bedrock while, downstream, the bedrock is impermeable marl-type and granite formation. The Herault catchment is splited into mostly schist and granitic head watersheds (the Valleraugue and la Terrisse sub-catchments) and a predominantly limestone plateau (Saint Laurent le Minier sub-catchment). Finally, the Salz is characterised by sedimentary bedrock comprising sandstone and limestone (Figure 2).

The Ardèche and the Gardon catchments have been subject to intensive monitoring and studies (see lter reference, https://data.lter-europe.net/deims/site/rbv_fr_13), leading to prior knowledge on hydrological understanding. Both the local in-situ experiments (Ribolzi et al., 1997; Braud and Vandervaere, 2015; Braud et al., 2016a,b) and the modelling studies focused on this area (Garambois et al., 2013; Vannier et al., 2013) tend to a hydrological classification according to those contrasted geological properties and, in agreements with the usual hydrogeological signature found in the litterature (Sayama et al., 2011; Pfister et al., 2017). Marls, sandstone and limestones without karst are characterized by limited storage capacities, resulting in higher runoff coefficients, and high sensitivity to the initial soil moisture (Ribolzi et al., 1997; Braud et al., 2016a). In contrast, in granite and schist transects located on hillslope of the Ardèche catchment, infiltration tests and analysis of electrical resisitivity signals show high permeability of the geological substratum in depth (measured up to 2.5 m in depth); and high storage capacities reaching up to 600 mm in 7 out of 10 assessments with artificial forcing, the 3 remaining test suggesting local unaltered bedrock (Braud et al., 2016a,b). The natural resistivity profile suggests a regular soil bedrock interface

when the latter consist in schist, while the granite one presents a more chaotic structure. Finally, the continous comparative study of two experimental sites over surface areas of the order of one km² - one located on the schist upstream part of the Gardon catchment, the other one on it granite downstream part - suggests rapid subsurface flow processing on the schist area, while flow formation appears to be controlled by the extension of the saturated zone related to the river on the granitic site (Ayral et al., 2005; Maréchal et al., 2009, 2013).

7. Fig. 1: For those not knowing France, maybe a small location map within France could be added.

The France map has been added.

8. Figs. 2-3-4: I suggest grouping these three figures.

We followed your suggestion and grouped the figures.

9. Table 1: Be clear that QD2 and QH10 are maximum discharge. The HYDRO code could be introduced in the table. The meaning of Ls, L and Lsi should be made clear in the caption. Say ii the caption that bold values are dominant geology. In column ID, use the same detailed ID as those used in Table 2 for consistency. Not sure Vogue is the right spelling.

The missing caption on the table 1 have been added, and the table has been reorganized (the columns order has been changed to consistency group the properties of the catchments. Finally, the outlet "Vogue" has been changed to the correct French spelling "Vogüé", here in the table and in all the manuscript. Hereafter, the new caption of the table 1:

Physiographic properties and hydrological statistics of the 12 catchments ID: coding name of the catchments used at figure 1 and table 2; area [km²]; mean slope [-]; soil properties: mean soil depth [m] and main soil texture (Tx) : Ls = sandy loam texture, L = loam texture; Lsi = silty loam texture; Geology: percentage of bedrock geology [%] including sandstone (Sa), limestone (Li), granite and gneiss (GG), marls (Ma) and schists (Sc) subcategories - $^{(i)}$ bold values are the dominant geology; mean annual precipitation (P[mm]) ; Hydrometry: discharge time-series availability (Period); mean inter-annual discharge ($Q[m^3.km^{-2}.s^{-1}]$); 2 year return period of maximum daily discharge ($Q_{D2}[m^3.km^{-2}.s^{-1}]$); 10 year return period of maximum hourly discharge ($Q_{H10}[m^3.km^{-2}.s^{-1}]$). Hydrometric statistics are calculated from HydroFrance databank, (de l'Ecologie du développement durable et de l'énergie, 2015) (http://www.hydro.eaufrance.fr/) and the pluviometric ones using rainfall data from the raingauge network of the French flood forecasting services.

10. P9, L2-4: The information on flow data availability could be added in Table 1.

The flow data availability have been added in Table 1.

11. P9, L5-6: Not sure this QD2 threshold is actually the alert threshold everywhere in France. Though there may be link(ed), I am pretty sure the alert threshold is not determined using a statistical approach, but rather by a local analysis.

The flood warning system in France, had actually been restrustured, in the beginning of the 00's, after dramatic consequences of several flash flood events in 1999 and 2002. The main objective was to improve not only the flood forecasts, but also the communication with local authorities. To meet that objective, a subdaily flood warning map of the main rivers in France is broadcoast through a unique website (https://www.vigicrues.gouv.fr/). For a sake of clarity in the communication, a uniform color code is used: yellow for peak discharge ranging from the 2-year to the 10- year flood, orange for peak discharge ranging from the 10-year to the 50-year flood, and red for peak

discharge exceeding the 50-year flood. I didn't find any reference in english, but you can find some description in (Javelle et al., 2014): *In real-time, to describe the potential severity of the ongoing event along the river network, the estimated peak discharges are represented with a colour code based on three flood frequency categories: yellow for peak discharge ranging from the 2-year to the 10- year flood, orange for peak discharge ranging from the 10-year to the 50-year flood, and red for peak discharge exceeding the 50-year flood. These real-time products, delivered every 15 min, are used as input for a web site dedicated to French local authorities.*

12. P9, L5-11: The event selection process ignores all the rainfall events that did not generate high flows, but which would still be interesting to investigate, especially to check that the model is not over-reactive on such events. Was this analyzed in separate work? A few words could be added on this issue.

Those events have been not studied yet as the model has not been built to simulate this kind of responses. For instance there is no percolation and no groundwater recharge. When looking for the assessment of the models as a flood forecast tool, those event could have be to integrated on the study. Here, looking at the flow processes, they were not included, the precipitation being mostly infiltrated. Also, as the discharge threshold is moderate (2 years return period of the maximal daily discharge), the selection already includes intense rainfall events, with only moderate hydrological response, as suggest the different runoff coefficients of the events. Consequently we added this statistic on the table 2 (page 8).

The following sentence has also been added page 7, lines 10-14: *The aim of this selection was to be able to analyse, more broadly, overall catchment behaviour during intense ~~events~~ hydrological activity. Note also that, moderate or intense rainfall events without respective hydrological response might be abducted from the analysis. Nevertheless the first alert threshold used here is small enough to have a selection of flood events with contrasted runoff coefficient (see table 2.)*

13. P9, L14: Which FFS is it?

Here, it specially concerns two regional flood forecasting services (SPC): the SPC Grand Delta and the SPC Med-Ouest. However, to be synthetic, we refer to the national French flood forecasting service SCHAPI that is at the head of the regional ones (page 9, line 15).

14. P9, L14: What RHEA and CALAMR mean? Any reference?

CALAMAR is a patented software developed by a private company RHEA. The sentence has been reworded to be more understandable and to add the reference of the patent (page 9, lines 14-16). The reworded sentence is: *The French flood forecasting service (SCHAPI: Service central dhydrométéorologie et dappui à la prévision des inondations) used the CALAMAR patented software (Badoche-Jacquet et al., 1992) to produce the rainfall depth inputs of the model by combining these radar measurements with raingauge data.*

15. P9, L26: What SIM means?

SIM for the models used into the operational chain : the Safran model, a meteorological analysis system; the ISBA model simulated the Interaction between the Soil, the Biosphere, and the Atmosphere; and the Modcou model, a hydrogeological model. Those details have been succintly added page 8, lines 4-6:

This was done using spatial model outputs from Météo-France's SIM operational chain (Habets et al., 2008), including a meterological model (SAFRAN,Vidal et al. (2010)), a soil - vegetation - atmosphere model (ISBA, Mahfouf et al. (1995)) and a hydrogeological model (MODCOU, Ledoux et al. (1989)).

16. Section 2.2: Maybe I missed something but I did not find information on how the events were split into calibration and validation. Given there are only a few events per catchment, I guess results may be quite sensitive to this selection. This is not commented. Typically, if the authors had reversed the roles of the two events sub-sets (calibration / validation), would results be the same? If yes, this would strengthen the proposed analysis. If not, this may add further uncertainty in the analysis.

Here, each event set was splitted into calibration and validation sets, according to the work of Garambois et al. (2015). As suggested in the latter cited paper, a first individual calibration on each event was done. Events presenting atypical sensivity to the soil depth parameter, has been removed from the calibration. The extreme events were kept for validation. And finally, events were splitted in order to have a wide range of soil moisture initial condition. The following sentences have been added to detail the event set splitting (page 7-8, lines 18-2):

Each rainfall product is firstly assessed through an individual sensitivity analysis of the standard MARINE model (DWF model, see section 3.1). When presenting an atypical sensivity to the soil depth parameter, the rainfall event is discarded of the study, as suggesting questionable measurements. Depending on the availability of the results of rainfall and hydrometric measurements, 7 to 14 intense events were selected for each catchment (Table 2). Each set is finally splitted into a calibration and validation subsets as follow: the extreme events were kept for validation. A minimum number of calibration events is chosen in order to cover the wide range of soil moisture initial condition.

17. Table 2: For Qpeak, is it the mean of peak flows?

Exactly, the caption has been corrected.

18. Section 3.1, title: From the description, it appears that MARINE is a model, not a framework.

The title of the section 3.1 has been modified.

19. P11, L3-10: Indicate units in brackets for parameters.

The parenthesis have been replaced for brackets.

20. P11, L8: Write Module 2 (i.e. subsurface downhill flow)

The specification has been added.

21. P11, L12 (and elsewhere): Check the place of brackets around references.

Ok. Brackets around references have been check in the proofreading.

22. P14, last line: hourly maximum discharge

Ok.

23. P15, L17: This point was not fully clear for me. Please explain a bit more.

A different mean of evaluation is used for the hydrograph recession as the assessement based on the simulated discharge values - Qmed_INT - and restricted to the recession interval is actually representative whether the high discharge values occuring before the recession are well simulated or not. For this reason, to assess the hydrograph recession, a score based on the recession rate is rather used, as it enables to avoid such a dependency of the assessement with the high discharge

values, but also as it is very well known in the litterature as representative of the hydrological behavior of the catchment.

According to your comments and those from the other referee, another assessment is proposed in the reviewed manuscript. Instead of having a special figure for the assessment of the hydrograph recession, the detailed performances over the different stages are grouped in a same figure (figure 8). To meet that objective, a novel score based on the recession rate is proposed and used to evaluate the simulations of the related stage of the hydrographs. You can find these modifications on the section 4.2,page 13-14, lines 25-5:

Conversely, Qmed_INT ~~was not used~~ was not relevant for the evaluation of the capacity to reproduce recessions, because the calculation of this score - based on simulated discharge values - during the recession interval strongly depends on performance at high discharges. Instead, we used the A_{slope} score defined in the equation 9. It calculates the average standard error in simulating the decreasing rate of the discharge during the flood recession interval. Through the consideration of the A_{slope} score here, it is assumed that the recession rate is a relevant feature of the catchment's hydrologic properties (Troch et al., 2013; Kirchner, 2009). ~~We therefore choose to make a visual comparison of the simulated and observed recession curves, $Q(t) = f\left(log\left(-\frac{dQ(t)}{dt}\right)\right)$, which are characteristic of a catchment's hydraulic discharge properties. Lastly, the evaluation was completed by a description of the a priori and a posteriori modelling errors in order to identify those that were inherent in the choice of model structure, regardless of the calibration strategy adopted.~~

$$A_{slope} = \frac{\sum_{i=k}^{l} |\frac{dy_i}{dt} - \frac{d\hat{y}_i}{dt}|}{\sum_{i=k}^{l} \frac{d\hat{y}_i}{dt}} \tag{1}$$

where $\frac{d\hat{y}_i}{dt}$ and $\frac{dy_i}{dt}$ are respectively the observed and the simulated recession rate at a time step i which belongs to the flood recession interval $(i = k...l)$.

24. Section 5: As mentioned above, I think the authors should make an effort to much reduce this section. In several sub-sections, the comments detail so many cases that it is very difficult to get a broad picture.

According to your comments and those from the other referee, the presentation of the results has been reorganized as follow:

- section 5.1: Performance of the models. In these section are exclusively presented the assessment of the models through the metric scores (that are defined in the section 4.2).
  - section 5.1.1: Overall performances of the models. It merges the paragraph that has been written into the previous section 5.1.1 and 5.1.2
  - section 5.1.2: Detailed performances of the models: assessment of the models when simulating the different stages of an hydrograph. It contains the previous section 5.1.3 and 5.1.4
- section 5.1.3: Summary of the assessment: This part has been added, in order to present a global overview of the results after detailed comments in the aforementioned sections.
- section 5.2: Modelling errors inherent in the models'structures: It contain the previous 5.1.5 section.
- section 5.3: Analysis of relevance of the internal hydrological processes simulated: As suggested, the previous sections 5.2 and 5.3 have been merged into one element.

- section 6: Discussion: We propose a novel section in oder to separate the strict description of the results (section 5), and the interpretation done from it (section 6). It finally contains the previous sections 5.1.6 and 5.4.

25. Fig. 8: The distribution of mean results over all the catchments together would be useful to add. Is there any version that appears better on average? Please indicate in figure caption that the x axis refers to the catchments ID in Table 1.

The distribution of mean results over all the catchments together has been added and the caption was modified.

26. Fig. 9: Colors are not very useful (especially if the article is printed black and white). Maybe use different symbols instead.

Suggested modification were done.

27. Section 5.2: Difficult for me to extract the main points from this long discussion...

The section 5.2 (now 5.3.1) has totally been reworded (page 21).

28. Section 5.3: I was not fully convinced by the usefulness of this section.

The results of the section 5.3 (now 5.3.2) follows the section 5.3.1 about the assessment of the proportional volumes of the water up the hydrographs, that arise from the thre main path : on the surface, through the top or the deep layer of the soil. While those assessements are incomplete because of large uncertainties, the section 5.3.2 details their origins, revealing how the different models can involve at some points different internal dynamics and how two parameter sets can lead to simulate similar hydrographs, allowing a wide range of velocities, and then counterbalancing transfer time offsets between internal flow processes. We kept this section as, in our opinion, being relevant for the benchmark of the models, and interesting insights for further studies.

29. Fig. 13: Are the simulation shown obtained in calibration or in validation? It would be useful to have the dates instead of the time steps on the x axis.

As suggested, the x axis has been modified. Among the simulation shown, three are calibration events and one is a validation event. This has been added in the caption. In the same way, this detail has been also added one figure 10.

30. P30, L20 (and elsewhere): I think the term demonstrate is not appropriate. The work done here is not a demonstration. The links established between model results and actual processes remain hypotheses in the work, which may simply be more likely than others.

The term "demonstrate" was replaced by "indicate".

31. List of references: There are several incomplete references. The authors often give two URL; only keep the one for doi. Several titles are in French; please at least add the English translation in brackets, so that the non-French reader can more easily understand the topic of the cited references. I personally find it is not good practice to cite discussion papers if they were not ultimately accepted. The reference Ministere de lEcologie is strange looking.

The list of references has been checked and we propose a translation for the French cited papers.

References

Ayral, P.-A., Sauvagnargues-Lesage, S., and Bressand, F.: Contribution à la spatialisation du modèle opérationnel de prévision des crues éclair ALTHAÏR.(Contribution to the spatial distribution of the operational flash flood forecasting model ALTHAÏR), Etude de géographie physiqueéographie physique, 32, 75–97, 2005.

Badoche-Jacquet, G., Blanchet, F., and Blanchet, B.: Procédé et dispositif pour l'évaluation des précipitations sur une zone de terrain. (Procedure and apparatus for evaluating field precipitation), Patent EP0578544B1, p. Rhea, 1992.

Borga, M., Gaume, E., Creutin, J. D., and Marchi, L.: Surveying flash floods: gauging the ungauged extremes, Hydrological processes, 22, 3883, https://doi.org/10.1002/hyp.7111, 2008.

Borga, M., Stoffel, M., Marchi, L., Marra, F., and Jakob, M.: Hydrogeomorphic response to extreme rainfall in headwater systems: flash floods and debris flows, Journal of Hydrology, 518, 194–205, https://doi.org/10.1016/j.jhydrol.2014.05.022, 2014.

Braud, I. and Vandervaere, J. P.: Analysis of infiltration tests performed in the Cladugne catchment in May-June 2012, contribution to WP3.4 Documentation and mapping of soil hydraulic properties, soil geometry and vegetation cover of small catchments, Tech. Rep. 66, FloodScale (2012-2016) ANR project, URL `http://mistrals.sedoo.fr/?editDatsId=1321.`, 2015.

Braud, I., Ayral, P., Bouvier, C., Branger, F., Delrieu, G., Le Coz, J., Nord, G., Vandervaere, J., Anquetin, S., Adamovic, M., Andrieu, J., Batiot, C., Boudevillain, B., Brunet, P., Carreau, J., Confoland, A., F, D. L. J., Domergue, J., Douvinet, J., Dramais, G., Freydier, R., Gérard, S., Huza, J., Leblois, E., Le Bourgeois, O., Le Boursicaud, R., Marchand, P., Martin, P., Nottale, L., Patris, N., Renard, B., Seidel, J., Taupin, J., Vannier, O., Vincendon, B., and Wijbrans, A.: Multi-scale hydrometeorological observation and modelling for flash-flood understanding, Hydrology and Earth System Sciences, 18, 3733–3761, https://doi.org/10.5194/hess-18-3733-2014, 2014.

Braud, I., Andrieu, H., Ayral, P.-A., Bouvier, C., Branger, F., Carreau, J., Delrieu, G., Douvinet, J., R., F., E., L., Le Coz, J., Martin, H., Nord, G., Patris, N., S., P., Renard, B., Seidel, J.-L., and Vandervaere, J.: FloodScale : Observation et modélisation hydro-météorologique multi-échelles pour la compréhension et la simulation des crues éclairs. Multi-scale hydrometeorological observation and modelling for flash-flood understanding,, final report of the projet ANR-2011 BS56 027 01, 2016a.

Braud, I., Ayral, P., Bouvier, C., Branger, F., Delrieu, G., Dramais, G., Le Coz, J., Leblois, E., Nord, G., and Vandervaere, J.: Advances in flash floods understanding and modelling derived from the FloodScale project in south-east France, in: 3rd European Conference on Flood Risk Management, Innovation, Implementation, Integration (FLOODrisk 2016), vol. 7, p. 04005, Lyon, France, https://doi.org/10.1051/e3sconf/20160704005, fLOODrisk 2016, 3rd European Conference on Flood Risk Management, Innovation, Implementation, Integration, Lyon, FRA, 18-/10/2016 - 20/10/2016, 2016b.

Colmet Daage, A., Ricci, S., Sanchez-Gomez, E., and Valérie, B. E.: Climate change impacts on precipitation extremes, flows and flash floods in Mediterranean mesoscale catchments, in: International Congress on Environmental Modelling and Software, 2016.

de l'Ecologie du développement durable et de l'énergie, M.: HYDRO, http://www.hydro.eaufrance.fr/, URL `http://www.hydro.eaufrance.fr/`, 2015.

Garambois, P., Roux, H., Larnier, K., Labat, D., and Dartus, D.: Characterization of catchment behaviour and rainfall selection for flash flood hydrological model calibration: catchments of the eastern Pyrenees, Hydrological Sciences Journal, 60, 424–447, https://doi.org/10.1080/02626667.2014.909596, 2015.

Garambois, P.-A.: Etude régionale des crues éclair de l'arc méditerranéen franais. Elaboration de méthodologies de transfert des bassins versants non jaugés (Regional study of the flash flood events in the French Mediterranean area: on the methods for transfering information to ungauged catchment), Ph.D. thesis, Institut de Mécanique des Fluides de Toulouse - IMFT (Toulouse, France), 2012.

Garambois, P. A., Roux, H., Larnier, K., Castaings, W., and Dartus, D.: Characterization of process-oriented hydrologic model behavior with temporal sensitivity analysis for flash floods in Mediterranean catchments, Hydrology and Earth System Sciences, 17, 2305–2322, https://doi.org/10.5194/hess-17-2305-2013, 2013.

Gaume, E., Bain, V., Bernardara, P., Newinger, O., Barbuc, M., Bateman, A., Blaskovicova, L., Blschl, G., Borga, M., Dumitrescu, A., Daliakopoulos, I., Garcia, J., Irimescu, A., Kohnova, S., Koutroulis, A., Marchi, L., Matreata, S., Medina, V., Preciso, E., Sempere-Torres, D., Stancalie, G., Szolgay, J., Tsanis, I., Velasco, D., and Viglione, A.: A compilation of data on European flash floods, Journal of Hydrology, 367, 70 – 78, https://doi.org/10.1016/j.jhydrol.2008.12.028, 2009.

Habets, F., Boone, A., Champeaux, J.-L., Etchevers, P., Franchisteguy, L., Leblois, E., Ledoux, E., Le Moigne, P., Martin, E., Morel, S., et al.: The SAFRAN-ISBA-MODCOU hydrometeorological model applied over France, Journal of Geophysical Research: Atmospheres (1984–2012), 113, https://doi.org/10.1029/2007JD008548, 2008.

Jacq, V.: Inventaire des situations précipitations diluviennes sur la région Languedoc-Roussillon, la Provence-Alpes Cotes dazur et la Corse, période 1958-1994 (Inventory of the extreme rainfall events that have occured between 1958-1994 in the Languedoc-Roussillon,the Provence-Alpes Cotes dazur and the Corse regions), Phénomènes remarquables, 1994.

Javelle, P., Demargne, J., Defrance, D., Pansu, J., and Arnaud, P.: Evaluating flash-flood warnings at ungauged locations using post-event surveys: a case study with the AIGA warning system, Hydrological sciences journal, 59, 1390–1402, 2014.

Kirchner, J. W.: Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, Water Resources Research, 45, https://doi.org/10.1029/2008WR006912, 2009.

Ledoux, E., Girard, G., De Marsily, G., Villeneuve, J., and Deschenes, J.: Spatially distributed modeling: conceptual approach, coupling surface water and groundwater, in: Unsaturated Flow in Hydrologic Modeling, pp. 435–454, Springer, 1989.

Llasat, M. C., Marcos, R., Llasat-Botija, M., Gilabert, J., Turco, M., and Quintana-Seguí, P.: Flash flood evolution in North-Western Mediterranean, Atmospheric Research, 149, 230 – 243, https://doi.org/10.1016/j.atmosres.2014.05.024, 2014.

Mahfouf, J., Manzi, A., Noilhan, J., Giordani, H., and Déqué, M.: The land surface scheme ISBA within the Météo-France climate model ARPEGE. Part I. Implementation and preliminary results, Journal of Climate, 8, 2039–2057, https://doi.org/10.1175/1520-0442(1995)008¡2039:TLSSIW¿2.0.CO;2, 1995.

Maréchal, D., Ayral, P.-A., Bailly, J.-S., Puech, C., and Sauvagnargues-Lesage, S.: Sur lorigine morphologique des écoulements par lanalyse dobservations hydrologiques distribuées. Application à deux bassins versants cévenols (Gard, France). Relation between the flow processes and the catchment morphology through the analysis of the distributed hydrological observations : case of two Cevenol catchments (France), Géomorphologie: relief, processus, environnement, 19, 47–62, https://doi.org/10.4000/geomorphologie.10120, 2013.

Maréchal, J.-C., Ladouche, B., and Dörfliger, N.: Analyse hydrogéologique de la contribution de l'eau souterraine à la crue éclair des 6 et 8 septembre 2005 à Nîmes (Hydrogeological analysis of groundwater contribution to the 6-8 September 2005 flash flood in Nîmes), La Houille Blanche, pp. 88–93, 2009.

Pfister, L., Martínez-Carreras, N., Hissler, C., Klaus, J., Carrer, G. E., Stewart, M. K., and McDonnell, J. J.: Bedrock geology controls on catchment storage, mixing, and release: A comparative analysis of 16 nested catchments, Hydrological Processes, 31, 1828–1845, https://doi.org/10.1002/hyp.11134, hYP-16-0708.R1, 2017.

Ribolzi, O., Moussa, R., Gaudu, J.-C., Valls, V., and Voltz, M.: Etude des crues de transition entre priode sche et priode humide, par traage naturel sur un bassin versant mditerranen cultiv, Comptes Rendus de l'Acadmie des Sciences - Series IIA - Earth and Planetary Science, 324, 985 – 992, https://doi.org/https://doi.org/10.1016/S1251-8050(97)83983-3, 1997.

Ruin, I., Creutin, J.-D., Anquetin, S., and Lutoff, C.: Human exposure to flash floods - Relation between flood parameters and human vulnerability during a storm of September 2002 in Southern France, Journal of Hydrology, 361, 199–213, https://doi.org/10.1016/j.jhydrol.2008.07.044, tY - JOUR, 2008.

Sayama, T., McDonnell, J. J., Dhakal, A., and Sullivan, K.: How much water can a watershed store?, Hydrological Processes, 25, 3899–3908, https://doi.org/10.1002/hyp.8288, 2011.

Tarolli, P., Borga, M., Morin, E., and Delrieu, G.: Analysis of flash flood regimes in the North-Western and South-Eastern Mediterranean regions, Natural Hazards and Earth System Sciences, 12, 1255–1265, https://doi.org/10.5194/nhess-12-1255-2012, URL https://www.nat-hazards-earth-syst-sci.net/12/1255/2012/, 2012.

Troch, P. A., Berne, A., Bogaart, P., Harman, C., Hilberts, A. G. J., Lyon, S. W., Paniconi, C., Pauwels, V. R. N., Rupp, D. E., Selker, J. S., Teuling, A. J., Uijlenhoet, R., and Verhoest, N. E. C.: The importance of hydraulic groundwater theory in catchment hydrology: The legacy of Wilfried Brutsaert and Jean-Yves Parlange, Water Resources Research, 49, 5099–5116, https://doi.org/10.1002/wrcr.20407, 2013.

UNISDR: UNISDR Terminology on Disaster Risk Reduction, URL http://www.unisdr.org/eng/terminology/terminology-2009-eng.html (2009), 2009.

Vannier, O., Braud, I., and Anquetin, S.: Regional estimation of catchment-scale soil properties by means of streamflow recession analysis for use in distributed hydrological models, Hydrological Processes, https://doi.org/10.1002/hyp.10101, 2013.

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

Using a multi-hypothesis framework to improve the understanding of flow dynamics during flash floods

Audrey Douinot[1], Hélène Roux[1], Pierre-André Garambois[2], and Denis Dartus[1]

[1]Institut de Mécanique des Fluides de Toulouse, IMFT, Université de Toulouse, CNRS - Toulouse FRANCE
[2]Laboratoire des Sciences de l'ingénieur, de l'informatique et de l'imagerie (ICUBE) - INSA Strasbourg, Strasbourg, France

Correspondence to: Audrey Douinot (audreydouinot@gmail.com)

Abstract. A method of multiple working hypotheses was applied to a range of catchments in the Mediterranean area to analyse different types of possible flow dynamics in soils during flash flood events. The distributed, process-oriented model, MARINE, was used to test several representations of subsurface flows, including flows at depth in fractured bedrock, and flows through preferential pathways in macropores. Results showed contrasted perfomances of the submitted models, revealing different hydrological behaviours along the catchment set. The benchmark study offered a characterization of the catchments reactivity through the description of the hydrographs formation. The quantification of the different flow processes (surface, intra soil flows) were consistent with the scarse in-situ observations but remains uncertain, as a result of equifinality issue. The spatial description of the simulated flows over the catchments, made available by the model, enabled to spot counterbalancing effects between internal flow processes, including the compensation for the water transit time in the hillslopes and in the drainage network. New insights are finally proposed into strategical monitoring and calibration constraints setting up.

1 Introduction

1.1 Flash flood events: an issue for forecasters

Flash floods are ~~defined as~~ "sudden floods with high peak discharges, produced by severe thunderstorms that are generally of limited areal extent". (IAHS-UNESCO-WMO (1974); Garambois (2012); Braud et al. (2014)). They are often linked to localised and major forcings (greater than 100 mm, Gaume et al. (2009)) at the heads of steep-sided, meso-scale catchments (with surface areas of 10-250 km^2). ~~In Europe, particularly intense flash floods are observed predominantly on the north west of the Mediterranean Are, at the level of the mountain foothills. The regions affected are highly specific and marked by the influence of the Mediterranean climate system and mountainous topography. The steep topography and small size of the areas involved explain the rapid responsiveness of the catchments. The orographic effects on atmospheric circulation result in a higher accumulation of precipitation and localised convection cells Flash floods are, thus, the result of particular hydrological (or physiographic) and meteorological conditions.~~

The large specific discharges, and intensities of precipitation, makes the flash floods being ~~that occur being~~ classified as extreme. <u>Nevertheless, those events are not scarce nor unusual since</u> ~~However, this does not necessarily mean their occurrence is exceptional:~~ on average, there were no fewer than five flash floods a year on the Mediterranean Arc between 1958 and 1994 (Jacq, 1994), <u>and they tend to be amplifed against a background of climate change</u> (Llasat et al., 2014; Colmet Daage et al., 2016). ~~The EM-DAT (International Disaster Database, which records natural disasters affecting populations worldwide, also reports 33 thunderstorm episodes in Europe over the last ten years. Moreover, the first observations of global warming on the Mediterranean Arc signal an increase in the frequency and/or severity of events.~~ Flash floods constitute a significant hazard and, therefore, a considerable risk for populations (<u>UNISDR</u> 2009, Llasat et al. (2014)). ~~In general, floods, and the flooding they can cause, represent the world's principal natural hazard . Every year, 280 floods or storms are recorded as being disasters worldwide; whereas, statistically, over the same period, 31 earthquakes and 6 volcanic eruptions will have affected a population somewhere . One of the main explanatory factors is the vulnerability of the areas prone to flooding, which are undergoing increasing urbanisation. Flash floods~~ <u>They</u> are particularly dangerous due to their characteristics: (i) the suddenness of events makes it difficult to warn populations in time, and can lead to panic, thus increasing risk, when a population is unprepared (Ruin et al., 2008); <u>ii) the traditional connected monitoring system are not adapted to the temporal and spatial scales of the flash floods</u> (Borga et al., 2008; Braud et al., 2014); iii) the magnitude of floods implies significant amounts of kinetic energy, which can transform transitory rivers into torrents, resulting in the transport of debris ranging from fine sediments to tree trunks, as well as the scouring of river beds and the erosion of banks <u>(Borga et al., 2014)</u>.

A major area of interest for flash floods is, therefore, better risk assessment, to enable them to be forecasted and the relevant populations to be pre-warned. ~~However, this is not an easy task, because most of the small catchments concerned do not have gauges installed, and they therefore, cannot be connected to an automatic monitoring system . Moreover, weather forecasts remain uncertain, with regard to the intensity of precipitation and, above all, of the location of rain cells. Their use is therefore problematic, especially at the scale of these small catchments~~ Greater knowledge and understanding is required to better identify the determining factors that result in flash floods. In particular, in order to implement a regional forecasting methodology, the properties of the catchments, and the climatic forcing and linkages between them which lead to flash flood events need to be characterised.

1.2 Flash flood events: understanding flow processes

Due to the challenges involved in forecasting flash floods, especially against a background of climate change which is tending to amplify the phenomenon (Llasat et al., 2014; Colmet Daage et al., 2016), there has been considerable research done on the subject over the last ten years. Examples include the HYDRATE project (2006-2010, Gaume and Borga (2013)), which enabled the setting up of a comprehensive European database of flash flood flash events, as well as the development of a reference methodology for the observation of post-flood events; the EXTRAFLO project (2009-2013, Lang et al. (2014)) to estimate extreme precipitation and floods for French catchments; the HYMEX project (2010-2020, Drobinski et al. (2014)) focusing on the meteorological cycle at the Mediterranean scale, and, in particular, on the conditions that allow extreme events to develop; the FLASH project <u>(2012 - 2017, Gourley et al. (2017)) assessing the ability and the improvement of a flash flood forecasting</u>

framework in USA on the basis of real-time hydrological modelling with high resolution forcing; or the FLOODSCALE project (2012-2016, Braud et al. (2014)), based on a multi-scale experimental approach to improve observation of the hydrological processes that lead to flash floods.

In the North-Western Mediterranean context - specially concerned by specific autumnal convective meteorological events -
5 the European cited research demonstrates, in particular, the importance of cumulative rainfall (Arnaud et al., 1999; Sangati et al., 2009),(Camarasa-Belmonte, 2016), previous soil moisture state (Cassardo et al., 2002; Marchandise and Viel, 2009; Hegedüs et al., 2013; Mateo Lázaro et al., 2014; Raynaud et al., 2015) and the storage capacity of the area affected by the precipitation (Viglione et al., 2010; Zoccatelli et al., 2010; Lobligeois, 2014; Garambois et al., 2015a; Douinot et al., 2016). The combined influence of the spatial distribution of precipitation and event-related storage capacities, reported in the study of a num-
10 ber of particular events (Anquetin et al. (2010); Le Lay and Saulnier (2007); Laganier et al. (2014); Garambois et al. (2014) Faccini et al. (2016)), suggests a hydrological reaction, in some areas of the catchments, that arises from localised soil saturation. This statement surmises that there is little direct Hortonian flow, but rather the production of runoff through excess soil saturation, or lateral fluxes in the soil resulting from the activation of preferential pathways.

The geochemical monitoring of eight intense precipitation events, over a 3.9 km^2 catchment area, ~~during the FLOODSCALE~~
15 ~~project~~ (Braud et al., 2014), underlined the dominance of the intra-soil dynamic. First, analysis of the water from ~~revealed a~~ ~~"flushing" phenomenon. In at least~~ the first 40 cm of the soil layer revealed a "flushing" phenomenon, the water present at the start ~~was~~ being replaced by so-called "new" rainwater (Braud et al., 2016a; Bouvier et al., 2017). In addition, even if the peaks of the floods mainly consisted of new water, with a proportion varying between 50% and 80%, it appears that, over the entire period of the events, old water accounts for between 70% and 80% of the total volume of water discharged, which supports the
20 dominance of water pathways in the soil. ~~The proportion of new water at the peak of the flood varied between 50% and 80%~~ ~~depending on the intensity of precipitation and the moisture level at the start of the event. Conversely, over the entire period of~~ ~~the event, it appears that new water accounts for only between 20% and 30% of the total volume of water discharged, which~~ ~~underlines the dominance of intra-soil dynamics.~~

~~Being able to define the storage capacity of the soil column is crucial in explaining the varied responses of the catchments.~~
25 ~~Geological properties, which are crucial physiographic characteristics for determining the total storage capacity of catchments,~~ ~~also appear to be markers of the storage capacities available over the time scales involved in flash floods (which are of the~~ ~~order of a day).~~ Finally the geological properties themselves appear to be markers of the storage capacities available over the time scales involved in flash floods (which are of the order of a day). From simple flow balances of flash flood events (Douinot, 2016), studies of the diverse hydrological responses of several catchments over the same precipitation episode (Payrastre et al.,
30 2012), or the application of regional hydrological models dedicated to flash flood simulation (Garambois et al., 2015b), the literature tends to demonstrate the low storage capacity of non-karst sedimentary and marl-type catchments, and, conversely, the potential for storing large volumes of water in the altered rocks of granitic or schist formations. ~~Flow dynamics during flash~~ ~~floods thus appear to depend on the hydrogeological functioning of the catchments which again emphasises the importance of~~ ~~the saturation dynamics of the "soil + altered substratum" combination.~~

1.3 Applying a multi-hypothesis framework for improving hydrological understanding of the flash flood events

The knowledge gained about the development of the flow processes (for example, the tracing of events carried out during the FLOODSCALE project, Braud et al. (2014)), relates to studies on a number of specific sites where flash floods could be observed while they were taking place. However, being able to generalise the knowledge gained is limited by the specific
5 nature of each study (McDonnell et al., 2007) and by the gap between the spatial scale of forecasts (meso-scale), compared with that of the in-situ observations (<10 km^2) (Sivapalan, 2003). Such hydrological modelling work can be considered as a means of extrapolating knowledge to an extended geographical area, possibly covering catchments with differing physiographic properties.

Moreover, hydrological models viewed as "tentative hypotheses about catchment dynamics" are interesting tools for test-
10 ing hypotheses about hydrological functioning using a systematic methodology. A considerable amount of recently published work has involved comparative studies, using numerical models to develop or validate the hypotheses about the type of hydrological functioning that is most likely to reproduce hydrological responses accurately (Buytaert and Beven, 2011; Clark et al., 2011; Fenicia et al., 2014; Coxon et al., 2014; Ley et al., 2016; Fenicia et al., 2016). ~~For example,~~ Fenicia et al. (2014) ~~show that the performance of different models tested on the Attert Basin in Luxembourg corroborate the various hydrological~~
15 ~~processes known to occur in this catchment; non-linear models are better for modelling the hydrological dynamics of drainage sub-catchment basins on impermeable bedrock layers and those exhibiting threshold behaviour; conversely linear models with parallel storage elements led to better reproduction of the hydrological signature of the catchments with smoother responses.~~

~~The principle of "the method of multiple working hypotheses" is to compare the results from models governed by different assumptions about hydrological processes. Comparisons are even more meaningful if the structure of the models compared~~
20 ~~differs solely in terms of the hypotheses tested, in the form of modules.~~

<u>Using a same model's structure but differing solely in terms of the hypotheses tested, in the form of modules, the comparision is then focused and restricted to the hydrological assumtions tested.</u> Doing this avoids the limitations on interpretation that are often encountered in comparative studies of models (Perrin et al., 2013), where numerical choices can influence results independently of the underlying assumptions. ~~The comparative study makes it possible to conclude either a known hydrological~~
25 ~~functioning, which is distinguished by the better performance of the inherent model, or indeterminacy in the case of an equivalent fit of the models. The equifinality of the models remains instructive because it makes it possible to detect the underlying uncertainties behind the hypothesis of the models, which then helps determine avenues for further research.~~

The multiple working hypotheses framework is usually applied using a flexible conceptual and lumped model framework, such as the FUSE (Clark et al., 2008) or SUPERFLEX (Fenicia et al., 2011). <u>But also,</u> Clark et al. (2015a) and Clark et al. (2015b)
30 have proposed a unified structure to test multiple working hypotheses within a distributed modeling framework. <u>To our knowledge,</u> the case study using the aforementioned frameworks are related to continuous hydrological studies in order to assess hydrological hypotheses through the overall hydrological signature of the catchments. In this work, we extend the method of multiple working hypotheses to the assessment of an ~~distributed, mecanistic and~~ event-based hydrological model framework.

4

The objective is to test a number of proposed hydrological functioning that occur during flash flood events on a set of contrasting catchments in the French Mediterranean area. While the proportion of flows passing through the soil appears to be significant, questions arise about how they form:

– Are they subsurface flows that take place in a restricted area of the root layer, as a result of preferential path activation? Or, are they lateral flows taking place at greater depth comparable to those seen in some aquifer?

– Does the geological bedrock or an altered substratum play a role limited to that of mere storage reservoir, or is it actively involved in flood flows formation?

– Which are the flow processes proportions, according to the events and the catchments? ~~Can the hydrological processes be discerned from the nature of the geological bedrock?~~

~~Other than the observations discussed above, which were made on a specific small site (<10 km², there is little information on the formation of flows in the soil and/or geological layers. While the proportion of flows passing through the soil appears to be significant, questions arise about how they form:~~

The aim of this article is to attempt to answer these questions using a multi-model approach that tests different types of hydrological dynamics. The study was based on MARINE, a physically based, distributed hydrological model (Roux et al., 2011; Garambois et al., 2015a), which was developed specifically to model flash floods in the catchments of the French Mediterranean Arc. Several new representations for the soil column and underground flows were proposed (Douinot, 2016) and included in the MARINE model, in the form of modules that can be used to test different hydrological functions (Section 3). Those different hydrological dynamics were applied to a set of catchments - presented in Section 2 - with physiographic properties representative of the whole of the French Mediterranean Arc. The performance of each model was then examined and subjected to a comparative study (Section 4 and 5). The contributions of the results for improving the hydrological functionnning understanding are lastly discussed in Section 6 before concluding.

~~The structure of the publication is as follows: Section 2 describes the catchments and different datasets used in the study. Section 3 describes the MARINE model and the hypotheses about flow dynamics that were tested. Section 4 describes the evaluation methodology used to characterise the performance of each model. Section 5 presents the key results of the study, in the form of a comparative description of the simulations that resulted from the different modelling choices made. Lastly, the final section sets out conclusions and discusses the works contribution to our understanding of the hydrological functioning of catchments during flash floods and the effectiveness of the methodology adopted.~~

2 Catchments and data used in the study

2.1 Study catchment set

We studied the behaviour of four catchments and eight nested catchments in the French Mediterranean Arc (Figure 1). The catchments (in the order they are numbered in Figure 1) were those of the Ardèche, Gardon, Hérault and Salz rivers; these

were selected for the following reasons: (i) they are representative of the physiographic variability found in areas where flash floods occur; (ii) numerous studies of flash floods have already been carried out on the Gardon and Ardèche (Ruin et al., 2008; Anquetin et al., 2010; Delrieu et al., 2005; Maréchal et al., 2009; Braud et al., 2014), for example. Knowledge of the hydrological functioning of these catchments could guide the interpretation of the modelling results (Fenicia et al., 2014); and

5   (iii) a considerable number of observations of flash flood events are available for these catchments.
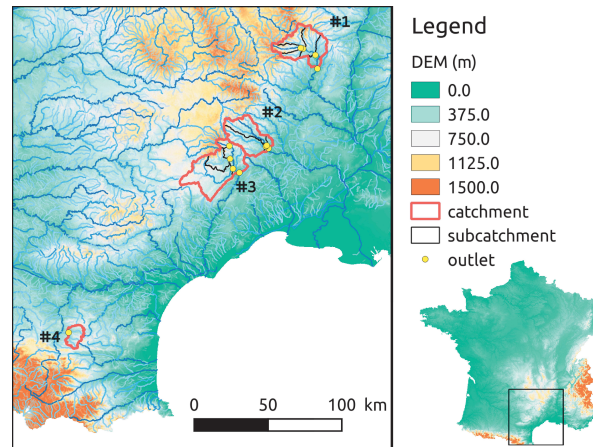

Figure 1. Locations of the catchments studied, with a topographic visualisation at $25\ m$ resolution (Source: IGN, MNT BDALTI)

The main physiographical and hydrological properties of the catchments are presented in Table 1. Figure 2 shows the contrasted geological properties of the studied area : the catchments are marked by a clear upstream / downstream difference. The Ardèche catchment upstream of Ucel sits essentially on a granite bedrock with some sandstone on its edges, while downstream, the geology changes to a predominantly schist and limestone formations. Similarly, the upstream part of the

10   Gardon catchment consists of schistose bedrock while, downstream, the bedrock is impermeable marl-type and granite formation. The Herault catchment is splited into mostly schist and granitic head watersheds (the Valleraugue and la Terrisse sub-catchments) and a predominantly limestone plateau (Saint Laurent le Minier sub-catchment). Finally, the Salz is characterised by sedimentary bedrock comprising sandstone and limestone (Figure 2).

The Ardèche and the Gardon catchments have been subject to intensive monitoring and studies (see lter reference, https://data.lter-

15   europe.net/deims/site/rbv_fr_13), leading to prior knowledge on hydrological understanding. Both the local in-situ experiments (Ribolzi et al., 1997; Braud and Vandervaere, 2015; Braud et al., 2016a, b) and the modelling studies focused on this area (Garambois et al., 2013; Vannier et al., 2013) tend to a hydrological classification according to those contrasted geological properties and, in agreements with the usual hydrogeological signature found in the litterature (Sayama et al., 2011; Pfister et al., 2017a). Marls, sandstone and limestones without karst are characterized by limited storage capacities, resulting in higher runoff

20   coefficients, and high sensitivity to the initial soil moisture (Ribolzi et al., 1997; Braud et al., 2016a). In contrast, in granite and schist transects located on hillslope of the Ardèche catchment, infiltration tests and analysis of electrical resisitivity signals show high permeability of the geological substratum in depth (measured up to 2.5 m in depth); and high storage capacities

reaching up to 600 mm in 7 out of 10 assessments with artificial forcing, the 3 remaining test suggesting local unaltered bedrock (Braud et al., 2016a, b). The natural resistivity profile suggests a regular soil bedrock interface when the latter consist in schist, while the granite one presents a more chaotic structure. Finally, the continous comparative study of two experimental sites over surface areas of the order of one km^2 - one located on the schist upstream part of the Gardon catchment, the other one

5   on it granite downstream part - suggests rapid subsurface flow processing on the schist area, while flow formation appears to be controlled by the extension of the saturated zone related to the river on the granitic site (Ayral et al., 2005; Maréchal et al., 2009, 2013).

The Ardèche catchment at Vogüe has a surface area of 622 km^2. We also studied the behaviour of sub-catchments at Meyras (99 km^2), Pont-la-Beaume (292 km^2) and Ucel (477 km^2). The Ardèche catchment upstream of Ucel sits essentially

10 on a granite bedrock with some sandstone on its edges. Downstream, the geology changes to a predominantly schist and limestone formations (Figure 2). In this area, studies from experimental sites show that flows are mainly due to surface runoff from cultivated soils (Braud and Vandervaere, 2015). The mostly sand-loam soils, covering the entire catchment area, are relatively deep (47 cm) and become shallower as the elevation increases.

The Gardon catchment at Anduze has a surface area of 543 km^2. We also studied the behaviour of the sub-catchments at

15 Corbès (220 km^2) and Mialet-Roucan (240 km^2), which are two separate sub-catchments. The Gardon catchment is marked by clear upstream/downstream differences (Figure ??). The upstream consists of schistose bedrock, and mainly silty soil of shallow depth. Downstream, the bedrock is impermeable marl-type and granite formation, with the latter assumed to be altered. The soil there can be more than a metre deep. Observations of the hydrological functioning of a number of catchments, carried out over surface areas of the order of one km^2, (Ayral et al., 2005; Maréchal et al., 2009; Martin et al., 2005; Maréchal et al.,

20 2013) show that, for the schistose part, flows seem to form rapidly mainly in the subsurface, while on the granitic part of the catchment, flow formation appears to be controlled by the extension of the saturated zone related to the river.

The Hérault catchment at Laroque has a surface area of 912 km^2. The behaviour of the sub-catchments of Saint-Laurent-le-Minier (499 km^2), La Terrisse (155 km^2) and Valleraugue (46 km^2) were also studied. The Hérault catchment has highly contrasting physiographic properties, which are highlighted when it is split into sub-catchments. The sub-catchments at Valleraugue

25 and La Terrisse are on the Cévennes Massif. They sit mainly on schists, but also on granite and gneiss. The catchments are very steep, particularly upstream of Valleraugue, and the soil is mostly silty. Conversely, the sub-catchment upstream of Saint-Laurent-le-Minier is predominantly a limestone plateau, and the slopes are less steep and covered with a silt-loam soil with less capacity for infiltration. The presence of a large karst formation, revealed in particular by a less developed surface hydrographic network (Figure 2), should be noted on this sub-catchment. As a result of the physiographic diversity, there are

30 considerable differences between the mean hydrological responses of the sub-catchments (Table 1).

The Salz catchment at Cassaigne has a surface area of 144 km^2. It is representative of the catchments found in the Corbières (foothills of the Pyrenees), an area frequently affected by flash floods. It is characterised by sedimentary bedrock comprising sandstone and limestone (Figure 2). The slopes of this catchment are less steep than the other catchments studied. Conversely, soils are relatively deep, and the low mean inter-annual discharge is indicative of a low base flow.
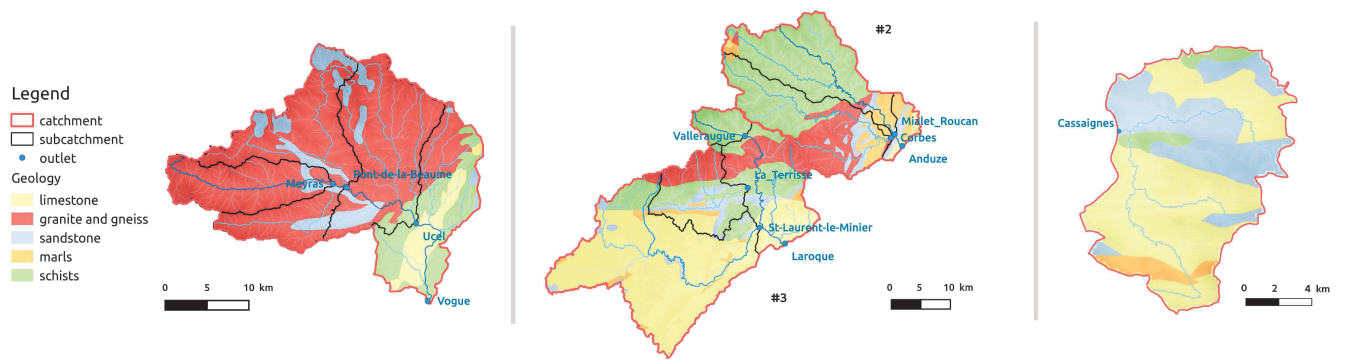
Figure 2. The geology of the Ardèche catchment (left), the Gardon and Hérault catchments (center), and the Salz catchment (sources : BD Million-Géol, BRGM)

Table 1 summarises the main geological, soil and topographical characteristics of the catchments studied.

Table 1. Physiographic properties and hydrological statistics of the 12 catchments *ID*: coding name of the catchments used at figure 1 and table 2; area [km²]; mean slope [-]; soil properties: mean soil depth [m] and main soil texture (Tx) : Ls = sandy loam texture, L = loam texture; Lsi = silty loam texture; Geology: percentage of bedrock geology [%] including sandstone (Sa), limestone (Li), granite and gneiss (GG), marls (Ma) and schists (Sc) subcategories - [i] bold values are the dominant geology; mean annual precipitation ($P[mm]$) ; Hydrometry: discharge time-series availability (Period); mean inter-annual discharge ($Q[m^3.km^{-2}.s^{-1}]$); 2 year return period of maximum daily discharge ($Q_{D2}[m^3.km^{-2}.s^{-1}]$); 10 year return period of maximum hourly discharge ($Q_{H10}[m^3.km^{-2}.s^{-1}]$). Hydrometric statistics are calculated from HydroFrance databank, ~~(de l'Ecologie du développement durable et de l'énergie, 2015)~~ (http://www.hydro.eaufrance.fr/) and the pluviometric ones using rainfall data from the raingauge network of the French flood forecasting services.

ID	River	Outlet	Area $[km^2]$	Slope $[-]$	Depth $[m]$	Tx $[-]$	Sa $[\%]$	Li $[\%]$	GG $[\%]$	Ma $[\%]$	Sc $[\%]$	P $[mm]$	Q	Q_{D2}	Q_{H10}	Period
					Soil properties		Geology[i]						Hydrometry			
													$[m^3.km^{-2}.s^{-1}]$			Period
♯1a	L'Ardèche	Vogüé	622	0.17	0.47	Ls	10.5	5.7	**71.9**	0.0	11.9	1587	0.041	0.62	2.25	00 - 15
♯1b		Ucel	477	0.20	0.45	Ls	13.7	0.0	**84.5**	0.0	1.8	1577	0.046	0.79	2.30	05 - 15
♯1c		Pont de la Beaume	292	0.22	0.39	Ls	14.0	0.0	**86.0**	0.0	0.0	1690	0.056	0.75	2.53	00 - 15
♯1d		Meyras	99	0.24	0.32	Ls	5.4	0.0	**94.6**	0.0	0.0	1720	0.036	0.72	2.92	00 - 15
♯2a	Le Gardon	Anduze	543	0.16	0.25	L	7.2	1.5	18.0	12.1	**61.2**	1370	0.026	0.48	1.82	94 - 15
♯2b		Corbès	220	0.16	0.27	L	9.3	0.0	34.2	9.0	**47.5**	1460	0.022	0.57	2.28	94 - 15
♯2c		Mialet Roucan	240	0.17	0.22	L	2.0	0.6	2.9	9.4	**85.1**	1407	0.023	0.62	2.54	02 - 15
♯3a	L'Hérault	Laroque	912	0.14	0.26	Lsi	6.7	**54.5**	11.7	3.2	24.0	1160	0.019	0.39	1.21	00 - 15
♯3b	La Vis	St Laurent le Minier	499	0.10	0.26	Lsi	4.0	**83.0**	1.0	3.2	8.8	930	0.018	0.42	1.10	00 - 15
♯3c	L'Arre	La Terrisse	155	0.19	0.25	L	19.5	12.3	27.2	6.2	**34.8**	1130	0.027	0.61	2.0	00 - 15
♯3d	L'Hérault	Valleraugue	46	0.27	0.25	L	0.0	0.0	0.0	0.0	**100.0**	1920	0.049	1.13	4.0	08 - 15
♯4	La Salz	Cassaigne	144	0.13	0.37	Lsi	33.5	**56.5**	0.0	5.1	4.9	700	0.008	0.20	1.31	01 - 15

2.2 Forcing inputs and hydrometric data

The hydrometric data were derived from the network of operational measurements (HydroFrance databank, http://www.hydro.eaufrance.fr/). Eight to twenty years of hourly discharge observations were available, according to the dates when the hydrometric stations were installed.

5   Flood events with peak discharges that had exceeded the 2-year return period daily discharge (Q_{D2}, in Table 1, corresponds to the alert threshold for flood forecasting centres in France) were selected as events to be included in the study. Thus, only one criterion for hydrological response was considered. This led to a selection of precipitation events of varying origins (for instance: rainfall induced by mountains, stagnant convective cells; and rainfall occurring in different seasons - mainly in autumn and early spring). Such a selection risked complicating the study because flow processes can vary from one season to another.

10   Nevertheless, it allowed us to test the ability of the model to deal with different (non linear) flow physics regimes. The aim of this selection was to be able to analyse, more broadly, overall catchment behaviour during intense ~~events~~ hydrological activity. Note also that, moderate or intense rainfall events without respective hydrological response might be abducted from the analysis. Nevertheless the first alert threshold used here is small enough to have a selection of flood events with contrasted runoff coefficient (see Table 2).

15   Precipitation measurements were taken from Météo France's ARAMIS radar network (Tabary, 2007), which provides precipitation measurements, at a resolution of $1\ km \times 1\ km$, every five minutes. The French flood forecasting service (SCHAPI: Service central d'hydrométéorologie et d'appui à la prévision des inondations) used then the CALAMAR patented software (Badoche-Jacquet et al., 1992) to produce rainfall depth data by combining these radar measurements with raingauge data. This processed dataset is here used as inputs of the model. ~~These measurements were calibrated by forecasters at the French~~

20   ~~Flood Forecasting Service by monitoring a network of rain gauges using RHEA's CALAMAR software.~~ Each rainfall product is firstly assessed through an individual sensitivity analysis of the standard MARINE model (DWF model, see section 3.1). When presenting an atypical sensivity to the soil depth parameter, the rainfall event is discarded of the study, as suggesting questionable measurements. Depending on the availability of the results of rainfall and hydrometric measurements, 7 to 14 intense events were selected for each catchment (Table 2). Each set is finally splitted into a calibration and validation subsets

25   as follow: the extreme events were kept for validation. A minimum number of calibration events is chosen in order to cover the wide range of soil moisture initial condition.

Some differences in meteorological forcing and the hydrological responses of catchments can be noted. The Ardèche (♯1) is subject to more significant events in terms of cumulative precipitation, with a notable orographic gradient. In contrast, cumulative precipitations in the Salz catchment (♯4) are the lowest. The highest precipitation intensities have been recorded in

30   the Gardon catchment (♯2). The events selected on this catchment cover a wide range of peak flows despite relatively uniform cumulative precipitation. The Hérault catchments (♯3) at Laroque and Saint Laurent le Minier had more uniform hydrological responses for meteorological forcing similar to that of the Gardon catchment in terms of precipitation, but these were lower in intensity.

10

As the MARINE model is event-based, it must be initialised to take into account the previous moisture state of the catchment, which is linked to the history of the hydrological cycle. This was done using spatial model outputs from Météo-France's SIM operational chain (Habets et al., 2008), including a meteorological analysis system (SAFRAN, Vidal et al. (2010)), a soil - vegetation - atmosphere model (ISBA, Mahfouf et al. (1995)) and a hydrogeological model (MODCOU, Ledoux et al. (1989)).

5   Based on the work of Marchandise and Viel (2009), the spatial daily root-zone humidity outputs (resolution = $8 \, km \times 8 \, km$) simulated by the SIM conceptual model were used for the systematic initialisation of MARINE.

Table 2. Properties of the flash flood events: average on the event set ($\pm$ standard deviation). *ID*: coding name of the concerned catchments (Figure 1 : $\sharp 1$ for the Ardèche; $\sharp 2$ for the Gardon; $\sharp 3$ for the Hérault and $\sharp 4$ for the Salz); N_{evt}: number of observed flash flood events; P [mm] mean precipitation ; $I_{max}[mm.h^{-1}]$: maximal intensity rainfall per event; Q_{peak}: specific flood peak $[m^3.km^{-2}.s^{-1}]$; Hum: initial soil moil moisture according to SIM output (Habets et al., 2008); CR: runoff coeficient [%]

ID	Outlet	N_{evt}	P [mm]	$I_{max}[mm.h^{-1}]$	$Q_{peak}[m^3.km^{-2}.s^{-1}]$	Hum [%]	CR [−]
$\sharp 1a$	Vogüé	10	192 ($\pm$93)	17.3 ($\pm$6.2)	1.33 ($\pm$0.57)	58 ($\pm$6)	0.50 ($\pm$0.16)
$\sharp 1b$	Ucel	10	208 ($\pm$105)	19.1 ($\pm$7.1)	1.41 ($\pm$0.70)	56 ($\pm$5)	0.47 ($\pm$0.17)
$\sharp 1c$	Pont de la Beaume	10	222 ($\pm$122)	20.5 ($\pm$6.2)	1.79 ($\pm$0.82)	56 ($\pm$5)	0.51 ($\pm$0.22)
$\sharp 1c$	Meyras	10	235 ($\pm$141)	25.6 ($\pm$10.6)	2.15 ($\pm$1.15)	56 ($\pm$4)	0.51 ($\pm$0.20)
$\sharp 2a$	Anduze	13	182 ($\pm$69)	26.9 ($\pm$12.6)	2.10($\pm$1.67)	53 ($\pm$7)	0.31 ($\pm$0.13)
$\sharp 2b$	Corbès	14	196 ($\pm$73)	31.4 ($\pm$11.6)	1.90 ($\pm$0.93)	55 ($\pm$7)	0.32 ($\pm$0.15)
$\sharp 2c$	Mialet Roucan	14	177 ($\pm$72)	30.9 ($\pm$13.2)	1.85 ($\pm$0.85)	51 ($\pm$7)	0.33 ($\pm$0.15)
$\sharp 3a$	Laroque	7	188 ($\pm$95)	16.0 ($\pm$8.1)	0.82 ($\pm$0.43)	59 ($\pm$8)	0.45 ($\pm$0.16)
$\sharp 3b$	St Laurent le Minier	7	153 ($\pm$95)	18.4 ($\pm$8.9)	1.14 ($\pm$0.31)	56 ($\pm$9)	0.47 ($\pm$0.16)
$\sharp 3c$	La Terrisse	7	193 ($\pm$103)	22.1 ($\pm$12.1)	1.63 ($\pm$0.87)	52 ($\pm$8)	0.60 ($\pm$0.23)
$\sharp 3d$	Valleraugue	7	156 ($\pm$110)	16.4 ($\pm$8.7)	2.14 ($\pm$1.33)	48 ($\pm$6)	0.62 ($\pm$0.22)
$\sharp 4$	Cassaigne	8	136 ($\pm$47)	17.8 ($\pm$6.2)	1.48 ($\pm$0.64)	57 ($\pm$7)	0.55 ($\pm$0.24)

3   The multi-hypothesis hydrological modelling framework

3.1   The MARINE model ~~framework~~

The MARINE model is a distributed mecanistic hydrological model specially developed for flash flood simulations. It models

10   the main physical processes in flash floods: infiltration, overland flow, lateral flows in soil and channel routing. Conversely, it does not incorporate low-rate flow processes such as evapotranspiration or base flow.

MARINE is structured into three main modules that are run for each catchment grid cell (see Figure 3). The first module allows the separation of surface runoff and infiltration using the Green-Ampt model. The second module represents subsurface downhill flow. It was initially based on the generalised Darcy Law used in the TOPMODEL hydrological model (Roux et al.,

11

2011), but was developed in greater detail as part of this study. Lastly, the third module represents overland and channel flows. Rainfall excess is transferred to the catchment outlet using the Saint-Venant equations simplified with kinematic wave assumptions. The model distinguishes grid cells with a drainage network (where channel flow is calculated on a triangular channel section (Maubourguet et al., 2007)) from grid cells on hillslopes (where overland flow is calculated for the entire surface area of the cell).
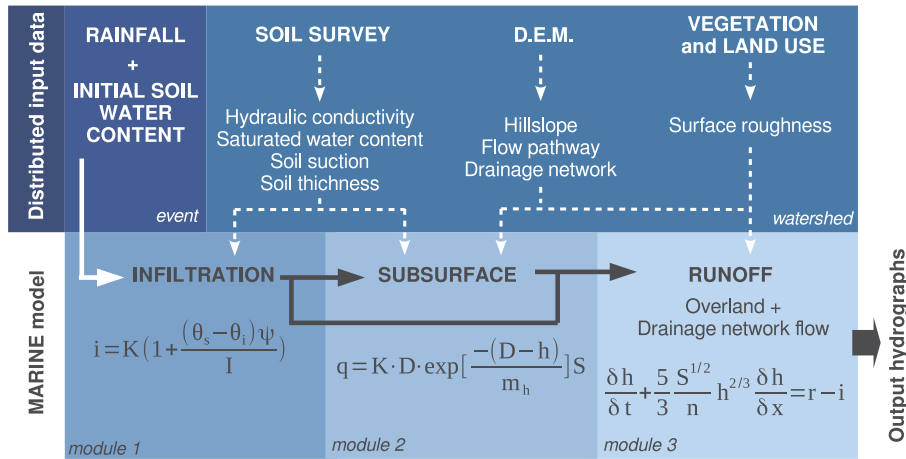

Figure 3. The MARINE model structure, parameters and variables. The Green and Ampt infiltration equation contains the following parameters: infiltration rate i [m.s^{-1}], cumulative infiltration I [mm], saturated hydraulic conductivity k [m.s^{-1}], soil suction at the wetting front Ψ [m], and, saturated ~~current~~ and initial water contents, θ_s and θ_i [m^3.m^{-3}], respectively. Subsurface flow contains the following parameters: soil thickness [m], lateral saturated hydraulic conductivity K [m.s^{-1}], local water depth h [m], transmissivity decay with depth m_h (m), and bed slope S [m.m^{-1}]. The kinematic wave contains the following parameters: surface water depth h [m], time t [s], space variable x [m], rainfall rate r [m.s^{-1}], infiltration rate i [m.s^{-1}], bed slope S [m.m^{-1}], Manning roughness coefficient n [m$^{-1/3}$.s]. The Module 2 described in this figure corresponds to the standard definition applied in the MARINE model. It corresponds, in fact, to the scope of model modifications proposed in this study, which are described in the next section (section 3.2.)

The MARINE model works with distributed input data such as: i) a digital elevation model (DEM) of the catchment to shape the flow pathway and distinguish hillslope cells from drainage network cells, according to a drained area threshold; ii) soil survey data to initialize the hydraulic and storage properties of the soil, which are used as parameters in the infiltration and lateral flow models; iii) vegetation and land-use data to configure the surface roughness parameters used in the overland flow model.

The MARINE model requires parameters to be calibrated in order to be able to reproduce hydrological behaviours accurately. Based on sensitivity analyses of the Garambois et al. (2013) model, five parameters are calibrated: soil depth - C_z, the saturation hydraulic conductivity used in lateral flow modelling - C_{kss}, hydraulic conductivity at saturation, used in infiltration modelling - C_k, and friction coefficients for low and high-water channels - n_r and n_p, respectively, with n_r and n_p uniform throughout the drainage network. C_{kss}, C_k and C_z are the multiplier coefficients for spatialised, saturated hydraulic conductivities and soil

depths. In this study, it was specifically Module 2 (i.e. subsurface downhill flow) that was subject to modifications in order to determine the possible ways that a number of proposals for intra-soil hydrological functioning could be modelled. To do this, modifications were made to the parameters C_z and C_{kss}.

3.2 Modelling lateral flows in the soil: the development of a multi-hypothesis framework

5 ~~The role of altered rocks has been demonstrated in the previous work of~~ Payrastre et al. (2012); Vannier et al. (2013); Garambois et al. (2015b). ~~The integration of this hydrologically active zone into MARINE was done by the calibration of C_z: soil-depth data from the BDsol databases~~ (Robbez-Masson et al., 2002) ~~are artificially increased to take account of the substratum.~~
~~Here, the aim was to integrate hydrological activity at depth, especially given that it seems to differ according to the geological properties of the bedrock~~ (Fenicia et al., 2014; Pfister et al., 2017a). We proposed a number of modifications to
10 Module 2 covering three hypotheses about hydrological functioning:

- Deep Water Flow model (DWF): we assumed deep infiltration and the formation of an aquifer flow in highly altered rocks. In hydrological terms the pedology-geology boundary was transparent. The soil column could be modelled as a single entity of depth D_{tot} (m), which is at least equal to the soil depth D_{BDsol} (m) (see Figure 4). Given the lack of knowledge and available observations, a uniform calibration was applied to the depth of altered rocks - D_{WB} (m) - a
15 level that is rapidly accessible on the scale of a rain event. Groundwater flow was described using the generalised Darcy Law (q_{dw}, Equation 1). The exponential growth of the hydraulic conductivity at saturation, as the water table (h_{dw}) rises, assumed an altered-rock structure where hydraulic conductivity at saturation decreases with depth (the TOPMODEL approach).

$$q_{dw} = K_{dw} \cdot D_{tot} \, exp \left(\frac{h_{dw} - D_{tot}}{m_h} \right) \cdot S \tag{1}$$

20 *with $h_{dw}[m]$, the water depth of the unique water table; $m_h[m]$, the decay factor of the hydraulic conductivity at saturation with soil depth; $S[-]$, the bed slope; $K_{dw} = C_{kdw} \cdot K_{BDsol}[m.s^{-1}]$, the simulated hydraulic conductivity at saturation; and $D_{tot} = D_{BDsol} + D_{WB}$, the soil column depth. Calibrated parameters are in red color.*

- Subsurface Flow model (SSF): We assumed that the formation of subsurface lateral flows was due to the activation of preferential paths, like the in-situ observations of Katsura et al. (2014) and Katsuyama et al. (2005). The altered soil-rock
25 interface acts as a hydrological barrier. The rapid saturation of shallow soils results in the development of rapid flows due to the steep slopes of the catchments and the existence of rapid water flows circulating through the macropores as the soil becomes saturated. The soil column was thus represented by a two-layer model (see Figure 5): an upper layer of depth equal to the soil depth D_{BDsol} (m) and a lower layer of uniform depth D_{WB} (m). The lateral flows in the upper layer were described by the generalised Darcy Law. However, variations in hydraulic conductivity were expressed as a
30 function of the mean water content of the layer (θ_{soil}) and not of the height of water (h_{soil}) that would form a perched water table (Equation 2). Expressing the variability in hydraulic conductivity as a function of the saturation rate indeed appears to be a more appropriate choice for representing the activation of preferential paths in the soil by the increase in

13

the degree to which the soil is filled. The decay factor of the hydraulic conductivity as a function of the saturation rate - m_θ - was set according to the linearized empirical relations, developed by Van Genuchten (1980), between hydraulic conductivity and soil water content for the different classes of soil textures. Flows in the lower soil layer (q_{dw}, Equation 3), in the form of a deep aquifer, were limited by setting the hydraulic conductivity of the substratum as being equivalent to that of the soil divided by 50 (this choice being guided by the orders of magnitude generally observed in the literature (Le Bourgeois et al., 2016; Katsura et al., 2014)). The altered rocks were thus assumed to play, mainly, a storage role. Infiltration occurring between the two layers was initially restricted by the Richards equations which were incorporated using the set hydraulic properties of the substratum (Equation 4). When the upper layer is saturated, filling by a piston effect is allowed. The depth of the soil layer, D_{BDsol}, was set according to the soil data, while the depth of the substratum - D_{WB} - was calibrated in the same way as in the DWF model.

$$q_{ss} = K_{ss} \cdot D_{BDsol}\, exp\left(\frac{\theta_{soil} - 1}{m_\theta}\right) \cdot S \tag{2}$$

$$q_{dw} = K_{dw} \cdot D_{WB}\, exp\left(\frac{h_{WB} - D_{WB}}{m_h}\right) \cdot S \tag{3}$$

$$q_{inf} = -K_{dw}\frac{\delta H(\theta_{soil}, \theta_{WB})}{\delta z} \tag{4}$$

with: h_{soil} and h_{WB} [m], the soil water depth in the upper and lower layer respectively; θ_{soil} and θ_{WB} [−], the soil water content of the upper and lower layer respectively; m_θ [−], the decay factor of the hydraulic conductivity with soil water content θ_{soil}; and $K_{ss} = C_{kss} \cdot K_{BDsol}$ and $K_{dw} = 0.02 \cdot K_{ss}$ [m.s^{-1}], the simulated hydraulic conductivity at saturation of the upper and lower layer in the SSF model respectively.

– The Subsurface and Deep Water Flow model (SSF-DWF): It was assumed that the presence of subsurface flow was due to both local saturation of the top of the soil column, but also the development of a flow at depth, as a result of significant volumes of water introduced by infiltration and a very altered substratum whose apparent hydraulic conductivity was already relatively high. This hypothesis of the process led to a modelling approach analogous to the SSF model (Figure 5), where the hydraulic conductivity at substrate saturation - K_{dw} - was no longer simply imposed, but, instead, calibrated using an additional coefficient, C$_{kdw}$.

$$K_{dw} = C_{kdw} \cdot K_{BDsol} \qquad \text{in SSF-DWF model} \tag{5}$$

The soil water content prior to simulation was, similarly, initialised for each model, in order to ensure, for a fixed depth of altered rock, that the same volume of water was allocated for all models. The SIM humidity indices (Section 2.2) were used to set an overall water content for all groundwater flow models for a given flood, with the two compartments of the SSF and SSF-DWF models then having an equal water content at initialisation.
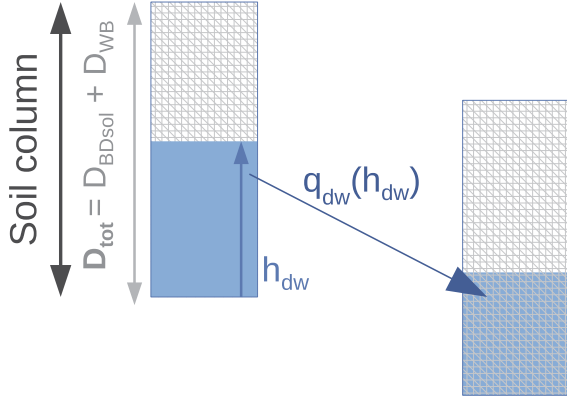
Figure 4. DWF model: flow generation by infiltration at depth and support of a deep aquifer ($q_{dw}(h_{dw})$, equation 1).
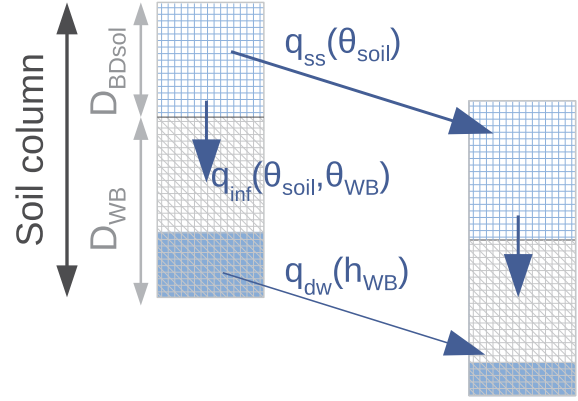

Figure 5. SSF and SSF-DWF models: flow generation by the saturation of the upper part of soil column and activation of preferential paths (q_{ss}), with support flow at depth (q_{dw}), and water exchanges from the upper layer to the lower one according to both soil water content ($q_{inf}(\theta_{soil}, \theta_{WB})$). See equations 2, 3 and 4, for the definition of the flows.

4 Methodology for calibrating and evaluating the models

4.1 Calibration method

The three hydrological models studied - DWF, SSF and SSF-DWF - were calibrated for each catchment by weighting 5,000 randomly drawn samples from the parameter space for each model (the Monte Carlo Method). The weighting was done using the DEC (Discharge Envelope Catching) score (equation 6), discussed by Douinot et al. (2017), in order to integrate the a priori uncertainties of modelling $\left((\sigma_{mod,i}),\ i = 1...n\right)$ (equation 7) and those related to the flow measurements $\left((\sigma_{\hat{y}_i}),\ i = 1...n\right)$ (equation 8). The choice of DEC is justified by the desire to adapt the evaluation criterion to the modelling objectives (for example, by focusing calibration on reproduction of the rise and peaks of floods in order to be able to forecast flash floods) while always being aware of the uncertainties in the reference flow measurements.

Given the lack of information, these uncertainties $\left((\sigma_{\hat{y}_i}),\ i = 1...n\right)$ were set at 20 % of the measured discharge, which is in line with the literature on discharge measurements from operational stations (Le Coz et al., 2014), and increased linearly with the 10-year hourly discharge, beyond which, as a general rule, the observed flow is no longer measured, but derived by extrapolation from a discharge curve, making it less accurate (equation 8). The envelop $\left((\hat{y}_i \pm 2\sigma_{\hat{y}_i}),\ i = 1...n\right)$ consequently defines the 95 % confidence interval of the observed flows.

The modelling uncertainties $\left((\sigma_{mod,i}),\ i = 1...n\right)$ were set at a minimum value - as a function of the basic catchment module, thus ensuring that the evaluation of the hydrographs would not be unduly affected by the reproduction of relatively

low flows which were strongly dependent on initialisation using previous moisture data that were not the subject of this study. In addition, it was assumed that a modelling uncertainty of 10 % around the confidence interval of observed flows was acceptable (equation 7). Finally, the overall overarching envelop $\left((\hat{y}_i \pm 2\sigma_{\hat{y}_i} \pm 2\sigma_{mod,i}),\ i = 1...n\right)$ defines hereafter the acceptability zone, that is to say the interval into which any simulated flow would be considered as acceptable, according to the modelling and measurement uncertainty definitions.

$$DEC = \frac{1}{n}\sum_{i=1}^{n}\epsilon_i^{DEC} \quad = \quad \frac{1}{n}\sum_{i=1}^{n}\frac{d_i}{\sigma_{mod,i}} \tag{6}$$

$$\sigma_{mod,i} \quad = \quad 0.5 * Q + 0.025 * \hat{y}_i \tag{7}$$

$$\sigma_{\hat{y}_i} \quad = \quad 0.05 * \hat{y}_i * \left(1 + \frac{\hat{y}_i}{Q_{H10}}\right) \tag{8}$$

with ϵ_i^{DEC} the DEC modelling error at time i; $\hat{y}_i$ and $\sigma_{\hat{y}_i}$ the observed discharge and the uncertainty of measurement at time i; d_i the discharge distance between the model prediction at time i (y_i) and the confidence interval of observed flows $\left(\hat{y}_i \pm 2\sigma_{\hat{y}_i}\right)$ ~~discharge measurement (that is to say the distance of y_i to $[\hat{y}_i - \sigma_{y_i}, \hat{y}_i - \sigma_{y_i}]$) at time i~~; $\sigma_{mod,i}$ the simulated uncertainty at time i; Q and Q_{H10} respectively the mean inter-annual discharge and the 10-year maximum hourly discharge of the related catchment.

4.2 Metrics and key points in model evaluation and comparison

~~The objective was to evaluate the fit of the models in terms of reproducing the different phases of the hydrographs, and provide a comparative description of the physical processes represented by each model.~~

Results of the models were firstly assessed and benchmarked using performance scores (section 5.1).~~The first step was to evaluate and compare the differences in modelling results from the DWF, SSF and SSF-DWF models.~~ The evaluation focused on the performance of the models in reproducing the hydrographs in overall terms, but also, more specifically, on their ability to reproduce the characteristic stages of floods: rising flood waters, high discharges, and flood recession. These stages were defined as follows:

– Rising flood waters: the period between the moment when the observed flow rate exceeded the mean inter-annual discharge of the catchment and the date of the first flood peak.

– High discharges: this stage includes the points for which the observed flow was greater than 0.25 times the maximum flow during the event.

– Flood recession: this stage begins after a period of t_c (the catchment concentration time according to Bransby's formula (Pilgrim and Cordery, 1992): $t_c = 21.3 \cdot L/(A^{0.1} \cdot S^{0.2})$) after the peak of the flood, and ends when discharge is rising again (or, where appropriate, at the end of the event - the time of peak flooding + 48h).

While the DEC score has provided a standard assessment of the modelling errors enabling a reasonable weighting of the simulations, for a sake of easy understanding ,the percentage of acceptable points of the simulated median time series

16

- Qmed_INT [%] (Douinot et al., 2017) - was chosen to evaluate the ability of the models to reproduce overall flows, rising flood waters and high discharges. A point is defined as acceptable when the median simulated value stands within the modelling acceptability zone, the latter one being determined by σ_{mod} et $\sigma_{\hat{y}}$ ~~Σ_{mod} et Σ_y~~.

Conversely, Qmed_INT ~~was not used~~ was not relevant for the evaluation of the capacity to reproduce recessions, because

5   the calculation of this score - based on simulated discharge values - during the recession interval strongly depends on performance at high discharges. Instead, we used the A_{slope} score defined in the equation 9. It calculates the average standard error in simulating the decreasing rate of the discharge during the flood recession interval. Through the consideration of the A_{slope} score here, it was assumed that the recession rate is a relevant feature of the catchment's hydrologic properties (Troch et al., 2013; Kirchner, 2009). ~~We therefore choose to make a visual comparison of the simulated and observed recession~~

10   ~~curves, $Q(t) = f\left(log\left(-\frac{dQ(t)}{dt}\right)\right)$, which are characteristic of a catchment's hydraulic discharge properties. Lastly, the evaluation was completed by a description of the a priori and a posteriori modelling errors in order to identify those that were inherent in the choice of model structure, regardless of the calibration strategy adopted.~~

$$A_{slope} = \frac{\sum_{i=k}^{l} \left| \frac{dy_i}{dt} - \frac{d\hat{y}_i}{dt} \right|}{\sum_{i=k}^{l} \frac{d\hat{y}_i}{dt}} \tag{9}$$

where $\frac{d\hat{y}_i}{dt}$ and $\frac{dy_i}{dt}$ are respectively the observed and the simulated recession rates at a time step i which belongs to the flood

15   recession interval $(i = k...l)$.

The evaluation was then completed through the description of the modelling errors (section 5.2), in order to identify those that were inherent in the choice of model structure, regardless of the calibration methodology adopted. In that respect, attention was paid on the a priori and a posteriori confidence interval of the model simulations respectively defined by $\left([y_i^{prior-5th}, y_i^{prior-95th}], i = 1...n \right)$ and $\left([y_i^{DEC-5th}, y_i^{DEC-95th}], i = 1...n \right)$ where $y_i^{prior-5th}$ and $y_i^{prior-95th}$ are the

20   5th and the 95th percentile of the 5000 model simulation values at time i, and where $y_i^{DEC-5th}$ and $y_i^{DEC-95th}$ are the 5th and the 95th percentile of the same but weighted series according to the DEC calibration criterion.

Those confidence intervals were standardized according to the DEC modelling error definition (equation 6), respectively defining the a priori and a posteriori confidence intervals of the modelling errors:

$$\epsilon_i^{\alpha-xth} = \begin{cases} 0 & \text{if} \quad | y_i^{\alpha-xth} | \leq 2 \cdot \sigma_{\hat{y}_i} \\ \frac{y_i^{\alpha-xth} \pm 2 \cdot \sigma_{\hat{y}_i}}{2 \cdot \sigma_{mod_i}} & \text{otherwise} \quad (- \text{ if } y_i^{\alpha-xth} > 0 \,;\, + \text{ if } y_i^{\alpha-xth} \leq 0) \end{cases} \tag{10}$$

25   *with $\epsilon_i^{\alpha-xth}$ is the xth percentile of the α modelling errors distribution at time i.*

The latter definition allows for an informative translation of the prior and posterior confidence intervals (Douinot et al. (2017)): a value of $\epsilon_i^{\alpha-xth}$ equal to 0 indicates that the $y_i^{\alpha-xth}$ bound lies within the discharge confidence interval; if $0 < \epsilon_i^{\alpha-xth} \leq 1$, the $y_i^{\alpha-xth}$ bound lies within the acceptability zone; and if $\epsilon_i^{\alpha-xth}$ is larger than 1 then errors of modelling is detected or remained. In addition, the benchmark of both a priori and a posteriori confidence intervals allows for highlighting which were

30   the remaining modelling errors that were induced by the model's assumptions, and those that were induced by the calibration. For those reasons, $\epsilon_i^{\alpha-xth}$ were used as the baseline of the modelling errors analysis.

17

5 Results

5.1 Performance of the models

5.1.1 Overall performances of the models

Assessment of the performances by catchment : Figure 6 shows the average Qmed_INT scores obtained after calibration of the DWF, SSF and SSF-DWF models for each catchment studied. It also shows the mean and standard deviations obtained from the series of calibration (top) and validation (bottom) events, calculated <u>over all the parts</u> ~~from or parts~~ of the catchments. ~~This section analyses the differences in performance, depending on the model used and the catchment studied.~~ The DWF model assuming deep infiltration and the formation of an aquifer flow in altered bedrocks showed better performance in the Ardèche catchment (♯1), while in the Gardon (♯2) and the Salz (♯4) catchments, the SSF and SSF-DWF models, assuming the formation of subsurface flows due to the activation of preferential flowpaths by local saturation (SSF), and development of flow at depth (SSF-DWF), produced the most accurate results. On the Hérault catchment (♯3), the modelling results obtained with each model, in terms of Qmed_INT, were less obvious, although the SSF-DWF model seemed to stand out to some extent. The differences in model performance were more pronounced for the validation events. The better-performing models tended to be more consistent, with equivalent Qmed_INT scores on calibration and validation events (for example, the DWF model on the Ardèche (♯1) or the SSF and SSF-DWF models on the Gardon (♯2). There was also a deterioration in performance in several models that had already been judged less effective (for example, the SSF and SSF-DWF models on the Ardèche (♯1), or the SSF model on the two catchments of the Hérault, ♯3c and ♯3d).

SSF model versus SSF-DWF model: As a reminder, the difference between the SSF and SSF-DWF models is that the latter has an extra calibration parameter - C_{kdw} - to be able to initialise a significant lateral flow in the subsoil horizons of the soil column (see Equation 3). The lateral hydraulic conductivity in the deep layer is configured using the hydraulic conductivity from BD-sol: $K_{dw} = C_{kdw} \cdot K_{BDsol}$, with C_{kdw} set to $0.02 \cdot C_{kss}$ in the SSF model and calibrated in the SSF-DWF model. The small differences between the SSF and SSF-DWF models showed that this flexibility does not produce any significant improvement, with the exceptions of the Ardèche catchment at Meyras and the Hérault catchment at Valleraugue. These two areas have a number of common features that could explain the similar modelling results: they are at the heads of high elevation catchments with steep slopes (Table 1), and are subject to considerable annual meteorological forcing. Therefore calibration
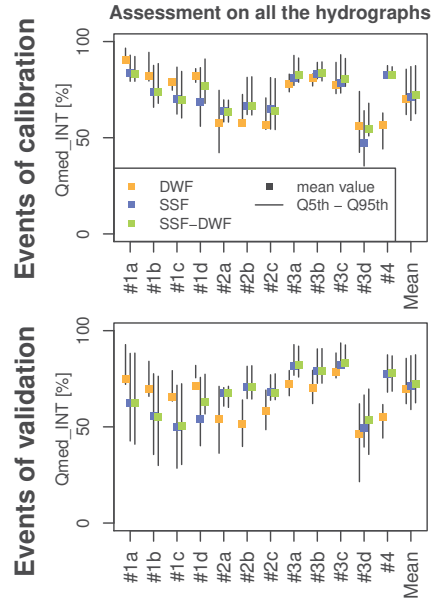
Figure 6. Qmed_INT scores: mean Qmed_INT scores obtained for the calibration (top) and validation (bottom) events, by model and catchment. The Qmed_INT scores were calculated for the whole hydrograph. ~~(left), modelling of the rising flood waters (centre), and modelling of high discharges (right).~~ <u>The x axis refers to the ID number of each catchment (Figure 1). Finally, *Mean* attribute refers to the average results over all the catchment obtained with each model.</u>
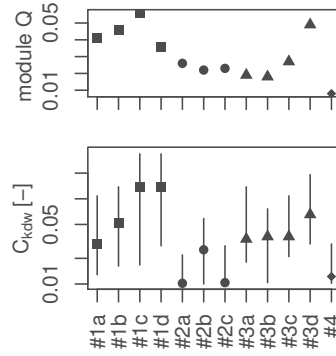

Figure 7. Top: Mean inter-annual discharge ($m^3.km^{-2}.s^{-1}$) for the catchments. Bottom: a posteriori distribution of the calibration of the subsoil horizon hydraulic conductivity in the SSF-DWF model (the C_{kdw} parameter, Equation 3)

of the saturation hydraulic conductivity parameter of the subsoil horizon tended to result in a significant flow at depth for these two catchments ($C_{kdw} \in [0.028, 0.33]$ for $\sharp 1d$ and $C_{kdw} \in [0.03, 0.2]$ for $\sharp 3d$, Figure 7, with this ratio set to 0.02 in the SSF model). In general, the calibration of the C_{kdw} parameter of the SSF-DWF model (Figure 7) seems to be correlated with the more or less sustained, annual hydrological activity of the catchments: the confidence interval of the C_{kdw} coefficient is

restricted to low values for the catchments with low mean inter-annual discharges (Figure 7, ♯2a, ♯2b, ♯2c, ♯3a, ♯3b, ♯4) and inversely for the catchments with high mean inter-annual discharges (♯1, ♯3c and ♯3d).

5.1.2 Detailed performances: assessment of the models to simulate the different stages of an hydrograph

~~Representation of rising flood waters and high-volume discharges~~

5   Figure 8 shows the detailed assessments according to the specific stages of the hydrographs. The objective is to highlight whether the overall performances (Figure 6) reflect uniform results along the hydrographs, or if they actually hide contrasted likelihood of the simulations over the different hydrographs's stages.

    Uniform results are observed on the Gardon catchment at Corbes and Anduze (♯2a and ♯2b) and on the Salz catchment (♯4): the SSF and SSF-DWF models demonstrated clearly superior performances for all stage-specific assessment on those

10 catchments. For the Gardon catchment at Mialet (♯2c), the detailed assessment (Figure 8) shows that the overall superiority of the SSF and SSF-DWF models is mainly due to a better simulation of the rising limb. Nevertheless, for any score, the SSF and SSF-DWF models present either similar of the best modelling results compared to the DWF model.

    On the Ardèche catchments (♯1a, ♯1b, ♯1c, ♯1d), the overall performances reflect the simulation of the high discharges and of the flood recessions. There, the DWF model gives the best results to simulate those hydrograph's stages. Conversely, it deals

15 slightly less well with the simulation of the rising flood waters. As it would be shown in the section 5.2, all the model tend to underestimate initial flows prior to the event and during the onset of a flood. The DWF model, in particular, exhibits this modelling weakness (see, for example, the onset of floods in the hydrographs for the 18/10/2006 and 01/11/2014 events in Ucel (♯1b), Figure 10), which explains the poorer performance. It can be noticed that the SSF-DWF model clearly better simulated the rising flood waters of the Ardèche head watershed (♯1d), explaining the overall good performance as well of this model on

20 this catchment (Figure 6).

    On the Hérault, the detailed evaluation enabled us to distinguish the performance of the different models. On the one hand, on the 2 larger catchments (♯1a and ♯1b), the DWF model get slightly better performances for rising flood waters simulations, while the SSF model gave more clearly better simulations of the flood recessions. On the other hand the SSF-DWF model generated the best simulations of the rising flood waters and of the high flows on the upstream catchments of La Terrisse (♯3c) and

25 Valleraugue (♯3d), while the DWF model simulated better flood recession. These constrated results explained why there is not a specific model that stands out on this catchment. In addition, it suggests a marked influence of the physiographic properties on the development of flow processes because they are correlated with the differences in the geological and topographical properties of the Hérault (♯3; see Figure 2 and Table 1). The hydrological behaviours simulated for the Valleraugue and La Terrisse sub-catchments, which are predominantly granitic and schistose, and where slopes are very steep, can be distinguished

30 from those of Laroque and Saint-Laurent-le-Minier, which are mainly sedimentary and in the form of large plateaus.

    ~~Considering detailed results for all periods covered by hydrographs, the SSF and SSF-DWF models on the Gardon and the Salz catchments produced the most uniform results, since both the simulations of rising flood waters and high-volume discharges demonstrated the superior performance of these models. The results for the Ardèche were not as clear (Fig. 6, (♯1)), because we observed that the DWF model produces the best simulation of high-volume flows. Conversely, the DWF model~~

~~deals slightly less well, overall, with rising flood waters. All the models tend to underestimate initial flows prior to the event and during the onset of a flood. The DWF model, in particular, exhibits this modelling weakness (see, for example, the onset of floods in the hydrographs for the 18/10/2006 and 01/11/2014 events in Ucel (#1b), Figure 10), which explains the poorer performance. On the Hérault, a detailed evaluation enabled us to compare the performance of the different models. On the~~

5 ~~one hand, the DWF model shows a more mixed performance for rising flood waters, reflecting a wider Qmed_INT confidence interval, which indicates greater uncertainty in forecasting the timing of rising flood waters. In addition, this model performed the best on the Hérault catchments at Laroque (#3a) and Saint Laurent le Minier (#3b); while the SSF-DWF model generated the best results for the upstream catchments of La Terrisse (#3c) and Valleraugue (#3d). These results suggest a marked influence of the physiographic properties on the development of flow processes because they are correlated with the differences in the~~

10 ~~geological and topographical properties of the Hérault (#3; see Figure 2 and Table 1). The hydrological behaviours simulated for the Valleraugue and La Terrisse sub-catchments, which are predominantly granitic and schistose, and where slopes are very steep, can be distinguished from those of Laroque and Saint-Laurent-le-Minier, which are mainly sedimentary and in the form of large plateaus.~~
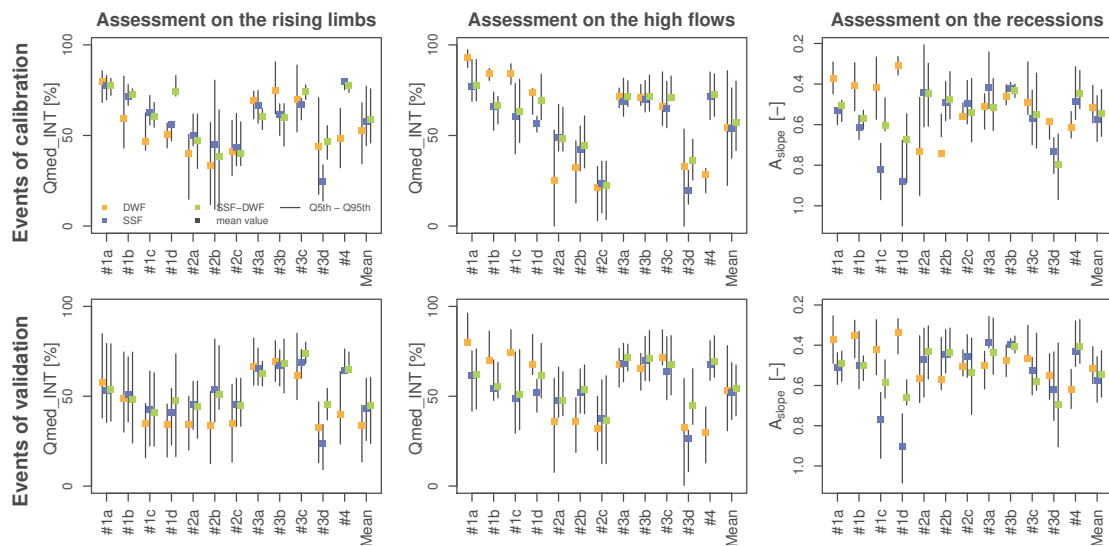

Figure 8. Assessment of the models by catchment over the different stages of the hydrographs. Left : Qmed_INT scores calculated over the rising flood waters stage; center: Qmed_INT scores calculated over the high discharges stage; right: A_{slope} scores. High Qmed_INT scores and conversely low A_{slope} values indicate good performances of the model. ~~Comparison of the modelled and observed characteristics of flood recession. Black: flood recession points for observed flows; orange: flood recession points for flows modelled using the DWF model; blue: flood recession points for flows modelled with the SSF model; green: flood recession points for flows modelled with the SSF-DWF model. Note that the y axis of the A_{slope} values are reversed in order to have all the good performances displayed on the top of the windows.~~

~~Representation of flood recessions~~

~~Visual inspection of the hydrographs showed that some models produced a better fit on certain catchments. An example of this was the DWF model on the Ardèche catchments (for example, the simulation of hydrographs at Ucel; ♯1d, Figure 10), which provided a much better fit for flood recession. The DWF model more accurately simulated the slow flood recession in the Ardèche catchment.~~

5   ~~The flood recession characteristics, $Q(t) = f(log(\frac{dQ(t)}{dt}))$, reflected the catchment's release properties. Figure 10 compares the simulated and observed flood-recession curves for each catchment. The catchments can be divided into three groups. For the Ardèche catchments (♯1a, 1b, 1c and 1d), the DWF model is considerably more accurate in reproducing flood recession, especially at the moment when the waters begin to recede. For the Salz-Cassaignes (♯4) and Gardon-Anduze and Corbes (♯2a, ♯2b) catchments, the SSF-DWF and SSF models performed better in reproducing recession curves. Conversely, for the other~~

10   ~~catchments, there are no distinctions to be drawn on how realistic the models' results are for this criterion, as can be seen for the Hérault catchment at Valleraugue (♯3d). For this third group, either there is no clear hydrological signature of the observed recessions (♯3a, ♯3c), or the characteristic recessions predicted by the models cannot be distinguished (♯2c, ♯3b).~~
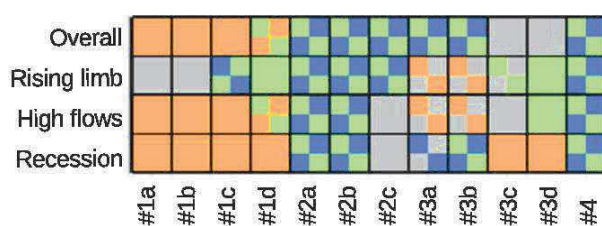
5.1.3   Summary of the assessment

Figure 9. Summary of the models's benchmark. A (2) color(s) is (are) attributed for each score and each catchment when one (or two) models give(s) clearly superior performance: the score of a model is defined as clearly superior when the lower bound of it confidence interval is higher than the median values obtained with the other models. The superiority of a model might be half attributed whether the criteria is only respected for the calibration processes. Color attribution: orange for the DWF model; blue for the SSF model; green for the DWF-SSF model; and grey when the superiority of one's model is undetermined.

The figure 9 sums up the highlighted models according to the assessed hydrograph's stage. It shows when one's model has a

15   clearly higher performance according to the following definition: a model is assessed as clearly superior when the lower bound of the confidence interval of his score is higher than the median values of the scores obtained with the other models. It reveals that the catchments set might be divided in 4 groups:

     – a first group of catchments where the SSF and DWF-SSF models uniformly give either similar or better performances than the DWF models. This is the case for the Gardon (♯2) and the Salz (♯4) catchments;

20   – a second group of catchments where the DWF model gives the best results according to all the scores besides the rising flood waters assessment. This is the case for the downstream Ardèche catchments ($\sharp 1a$, $\sharp 1b$, $\sharp 1c$);

- a third group where the models's results are not really discernible. For those catchment, the DWF model appears to slightly simulate better the rising flood and the high discharge, while the recession is better represented by SSF model. This is the case for the downstream Hérault catchments (♯3a, ♯3b);

- a last group where the SSF-DWF model slighty generated better the rising flood and the high discharge, while the recession is better represented by DWF model. In this group are the head watersheds of the Hérault (♯3c, ♯3d) and of the Ardèche (♯1d) catchments.

5.2 Modelling errors inherent in the models' structures

For the sake of conciseness, only the simulation of the hydrographs of one catchment is presented. Figure 10 shows the simulation results of the three models over the Ardèche catchment at Ucel (♯1b). It shows the simulated hydrographs, and their confidence intervals, compared with observed flows, as well as the inherent errors in the simulations. This highlights the modelling errors due to the choice of model structure (DWF, SSF or SSF-DWF models). When - at a time i - the a priori confidence interval (grey color) does not cross the acceptability region (green color), it means that no parameter set gives an acceptable simulation, and consequently modelling errors due to the structure - assumptions - of the model is detected. When the posterior confidence interval (salmon color) is outside the acceptability zone, modelling error is remaining. Finally whether the prior (posterior) interval is large or small, the model's structure allows for reaching a more or less large range of simulated values (the model prediction is more or less uncertain).

Representing the soil column with either one compartment (the DWF model) or two compartments (SSF or SSF-DWF models) leads to distinct a priori confidence interval of modelling errors (grey). The first structure (the DWF model) constrains the simulated flows at the beginning of the event, before the onset of precipitation, because the width of the confidence interval of the modelling errors is low at that point. More specifically, it tends to underestimate the initialisation discharges because the variation interval of the errors over this period is predominantly negative. This may explain this model's relative difficulty in reproducing the onset of floods, since the calibration of the parameters did not allow the acceptability zone on this part of the hydrograph to be reached. A resulting interpretation applicable to the catchment sets is that good results in modelling the rising flood waters with the DWF model means that the observed rising flow is relatively slow and could be reached in spite of the restrictive modelling structure (as example ♯3a, ♯3b).

Likewise, it can be noted that the one-compartment structure (in the DWF model) allows flexibility in the modelling of high discharges and flood recessions, because the confidence interval of the modelling errors ~~the variation intervals in the modelling errors~~ is quite large over these periods in the hydrograph. ~~However, it also led to the underestimation of high discharges and flood recessions.~~ In fact, the modelling error interval (in gray color) has a negative bias with respect to the acceptability zone. The calibration finally allows the simulations to be selected, at the intersection of the acceptability zones and a priori confidence in modelling errors. This generally corresponds to the calibration of a low-depth altered rock D_{WB}, in order to make the model more sensitive to soil saturation and more responsive, via the generation of early runoff. From that resulted

23

low D_{WB}, the simulated water storage capacity is limited, which might explains the inequacy of the DWF model for catchment with small runoff coefficients (♯2, table 2)

Conversely, the two-compartment structure (the SSF and SSF-DWF models) offers flexibility in modelling the beginning of events, flood warnings and high discharges, but the ability to model flood recessions is more constrained. SSF and SSF-DWF models simulate fast flood recession in comparison to the DWF model, suggesting that good results in modelling the flood recession with the SSF model might be interpreted as fast return to normal or low discharge are observed on the related catchments (as example, ♯2, ♯4). ~~As well, the relative position of the modelling-error confidence interval, with respect to the acceptability zone, shows that the structure leads to an unbiased estimate of the onset of a flood, a slight overestimation of high discharges, and an underestimation of flood recessions.~~

In the SSF and SSF-DWF models, the addition of a flux calibration parameter in the subsoil horizons, not surprisingly, laed to wider variations in the a priori modelling errors. A surprising finding, however, is that the calibration of the lateral conductivity of the deep layer, C_{kdw}, seems to affect only the simulation at the beginning of the hydrographs (the events of 01/11/2011 and 13/11/2014, figure 10), and has very little effect on flood recessions. ~~This last point was also visible in the analysis of flood recessions, where we observed a high degree of similarity in the flood recessions simulated by the SSF and SSF-DWF models, whatever catchment was being studied. The calibration of C_{kdw}, in fact, only influences the support volume from the subsoil horizons, but not the rate of decline (and, hence, not the recession). This is determined by the exponential term in equations 1, 2 and 3). It appears that the limited variations in the speeds of flood recession for the SSF and SSF-DWF models can be explained by the value of m_θ in Equation 2, which determines the rate of decay for most of the runoff as a flood recedes.~~ The high similiraties of the prior modelling intervals of the SSF and SSF-DWF models explain the similar performances of those models. In the same way, when there is improvement of the performnance through the SSF-DWF, it concerns the early rising of the flood, as the detailed perfomances has already shown it, the SSF-DWF enabling high and early start of the flood events.
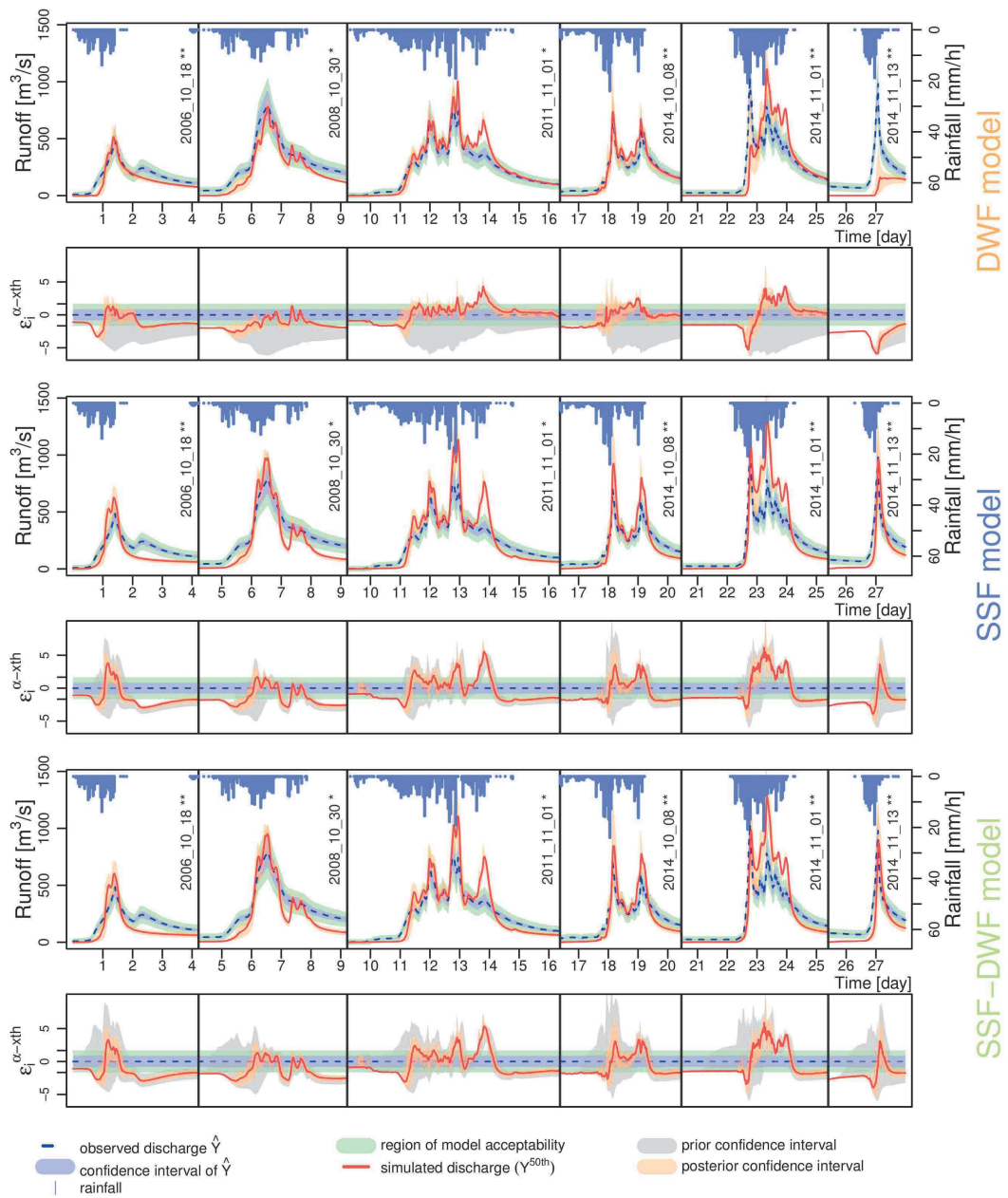
Figure 10. Calibration of the three models for the Ardèche catchment at Ucel, ♯1b). The results of the simulation of five flood hydrographs, and the inherent modelling errors (equation 10) for each model (top: DWF; centre: SSF; bottom: SSF-DWF). The median simulation and the posterior confidence interval are shown, respectively, in red and salmon. The confidence intervals of the measured flows and the acceptability zone are shown, respectively, in green and blue. The a priori confidence interval for each model (i.e. with no calibration) are shown in grey. (∗): event of calibration; (∗∗): event of validation.

5.3   Analysis of relevance of the internal hydrological processes simulated

5.3.1   Characterisation of the hydrological processes simulated

Each time a model is run it generates its own paths for water flow as it attempts to reproduce the hydrograph. The proportional volumes of the water making up the hydrographs, that arise from the three main simulated paths - on the surface, through the
5   top or through the deep layer of the soil - were calculated. Figure 11 shows the results for the runoff contribution, i.e. the water which has not passed through the soil at any point. The contributions of these surface flows on the whole of the hydrograph (Figure 11, left) and those that support high discharges (Figure 11, right) are distinguished. (Note that the other contributions are not detailed, being correlated to the runoff assessment, and therefore leading to a similar analysis).

The runoff contribution simulated by the DWF model discredits even more that model for representing the hydrological
10   behaviour of the Gardon (♯2) and Salz (♯4) catchments. Really high proportion of runoff contribution over the entire hydrograph were simulated, ranging from 40 to 98 %. In contrast, the few experimental measurements made on the Gardon (Bouvier et al., 2017; Braud et al., 2016a) provide evidence on the proportions of new water - which might be seen as an upper bound for runoff contribution volume - ranging from 20 to 40 % of the volumes in the hydrograph. The SSF and SSF-DWF model conversely gave more reasonable runoff contribution, although remaining high, ranging from 19 % and 62 %.

15   The DWF model suggests a larger contribution from runoff to the generation of high discharges, whatever the catchment modelled. We observed an increase from 15% to 30 % of the proportions of surface flow between the DWF model and the SSF and SSF-DWF models. But when considering the most suitable models, revealed in section 5.1, i.e. the DWF model for the Ardèche catchment and the SSF and SSF-DWF models for the Gardons catchment, the flow contribution assessment seems to be consistant with catchemnt set's diversity. The runoff contributions to the high flows of the hydrographs were slightly
20   lower on the three downstream Ardèche catchments (♯1a, ♯1b, ♯1c, with runoff contributions included between 17 and 57 %) compared to the runoff contributions on the Gardon catchment (♯2a, ♯2b, ♯2c) and on the upstream part of the Ardèche (♯1d, with runoff contributions between 20 and 78 %). It correlates the properties of the catchments and the rainfall forcing, the first catchment subset (♯1a, ♯1b, ♯1c) having deeper soil cover, with more permeable soil texture (see table 1), and being forced by rainfall with lower maximal intensities (see table 2) than the second one ♯2a, ♯2b, ♯2c. Without validating the estimation
25   done, it clearly suggests that the assessment of the flow contributions through the most suitable model's simulations for each catchment are consistant with each other.

~~The SSF and SSF-DWF models suggest very similar proportions of subsurface flows, including those at the catchment heads. Calibration of hydraulic properties at depth influences - as intended - only the proportions of subsurface and deep aquifer flows (which are not shown here). The DWF model suggests a larger contribution from runoff to the generation of high discharges,~~
30   ~~whatever the catchment modelled. In fact, we observed a 15 to 30 % increase in the proportions of surface flow between the DWF model and the SSF and SSF-DWF models.~~

~~The performance of the DWF model was noticeably different on the Gardon (♯2) and Salz (♯4) catchments, where simulated runoff was much more pronounced over the entire hydrograph. Here, the DWF model showed runoff proportions ranging from 40 to 98 %. However, the few experimental measurements made on the Gardon~~ (Bouvier et al., 2017; Braud et al., 2016a)

~~provide firm evidence on the proportions of new water - i.e. water resulting from meteorological forcing during the event - which range from 20 to 40 % of the volumes in the hydrograph. This clearly points to a lower runoff rate. Even though these experimental results only represent activity in the granitic part of the catchment, they appear to call into question the hydrological functioning suggested by the DWF model. Conversely, the observations lend support to the results obtained by the~~

5 ~~SSF and SSF-DWF models, where runoff rates were between 19 and 62 %. On the Salz there are no experimental observations available, and, therefore, observed results cannot be corroborated, by the orders of magnitude of the simulated surface flows. Nevertheless, in view of the extremely large runoff proportions suggested by the DWF model, the SSF and SSF-DWF model structures, with more reasonable assessments, appear to be more pertinent for characterising the types of processes occurring.~~

~~Taking the most suitable models for the catchments studied, an estimate of the degree of contribution of surface flows to the~~

10 ~~hydrographs can be made: (i) between 4 and 31 % for the main Ardèche catchments (♯1a, ♯1b, ♯1c), according to the DWF model, and between 0 and 40 %, or 10 and 29 %, on the Ardèche catchment at Meyras (♯1d), according to the DWF and SSF-DWF models, respectively; (ii) between 17 and 53 % (62 %) on the Gardon (♯2a, ♯2b, ♯2c) catchments; (iii) between 11 and 31 % in the Salz catchment (♯4); iv) between 5 and 58 % on the Hérault at Laroque and La Terrisse (♯3a and ♯3c), 15 and 63 % on ♯3b, and 5 and 34 % on ♯3d according to the DWF model, or between 10 and 43 % on ♯3a and ♯3c, 11 and 58 % on~~

15 ~~♯3b, and 4 and 20 % on ♯3d according to the SSF-DWF model.~~

On the downstream catchments of the Hérault (♯3a, ♯3b), the variation intervals of the surface flows estimated by the three models overlap. It may explain why the three models can achieve good reproductions of the hydrological signal : the calibration step makes possible, from that integrated point of view, to obtain an analogous distribution of the flows processes.

<u>Notwithstanding the uncertainty related to the model's choice when any model has been identified most suitable through</u>

20 <u>the performances,</u> ~~The above~~ the largest uncertainties are related to the parameterization of the models, a consequence of the equifinality of the solutions when calibrating a hydrological model against the solely criterion of the reproduction of the hydrological signal. While, in terms of plausibility, several sets of parameters may be equivalent, even for the same model, these sets of parameters are likely to lead to different hydrological functioning. ~~This is especially the case for the DWF model, for which the relative proportions of processes simulated depend on the choice of D_{WB}.~~

25 **5.3.2 Detailed study of four plausible simulations on the Hérault watershed at Saint Laurent-le-Minier**

~~Analysis of the distribution of the flows between those passing through the soil and those flowing on the surface lends support to the SSF and SSF-DWF models being realistic for the Gardon (♯2) and Salz (♯4) catchments. However, drawing distinctions in the other catchments between the models through such an integrated description of processes is limited by the equifinality of the solutions.~~ In order to detail the different <u>impacts behind the</u> hypotheses on which the models are based, <u>but also to explain</u>

30 <u>the resulting uncertainty when assessing the flow processes distribution</u> ~~of the various likely parameters on the hydrological changes that take place in the catchments~~, other variables, such as (spatialised and integrated) changes in moisture levels in the catchments or the flow velocities generated by modelling choices, have to be considered. Next, are described the detailed results of four simulations, ~~also~~ <u>equally</u> considered to be plausible according to the DEC criterion, obtained from the DWF and SSF models ~~as well as four sets of parameters~~ (2 simulations by model, see Table 3). ~~We considered~~ The Hérault catchment
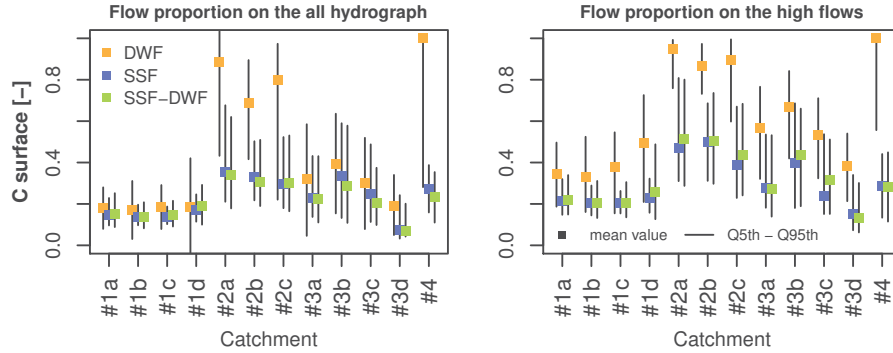
Figure 11. Proportion of surface runoff in the flows at the outlet. Left: The proportion over the whole hydrograph; right: the proportion at high discharges (Observed flow greater than 0.25 times the maximum flow during the event).

at Saint-Laurent-le-Minier (♯3b) ~~has been considered~~ because ~~of the equivalence of the models to represent that catchment~~ ~~the criteria previously used had not shown any one model to be more representative. The objective was to highlight how the~~ ~~models differed in terms of flow development, and what compensations occurred between processes to allow the equifinality~~ ~~of solutions.~~

Table 3. Realistic models and parameter sets for the Hérault catchment at Saint-Laurent-le-Minier (♯3b). C_{soil}: the contribution to the hydrograph of flows passing through the soil ; C_{kdw}/C_{kss}^* : the value of the parameter C_{kdw} for model DWF (Equation 1) or the value of the parameter C_{kss}^* for the model SSF (Equation 2).

ID	NSE	$D_{WB}[m]$	$C_k[-]$	$C_{kdw}/C_{kss}^*[-]$	$n_r[-]$	$n_p[-]$	$C_{soil}[\%]$
DWF1	0.82	0.15	17.3	8711	19.6	19.11	61
DWF2	0.84	0.11	2.34	4416	19.16	7.63	39
SSF1	0.89	0.40	15.81	45284	15.96	5.86	68
SSF2	0.89	0.34	2.08	22543	14.06	6.42	53

5   The figure 12 compares the changes over time in the state of soil saturation and the different simulated flow velocities of four "model + parameter set" configurations (Table 3). Figure 13 compares the spatial distributions of these variables, at a given moment, as an example.

In terms of hydrographs, quite logically given the similar likelihood scores, the simulations differ very little. Overall, the DWF1 configuration anticipate flood peaks; the DWF model (in the DWF1 and DWF2 configurations) generated greater flows

10   at the end of rain episodes; and these same configurations result in a slight underestimation of peaks for floods of average intensity (18/10/2009 and 05/03/2013) and, conversely, an overestimation of the peaks for exceptional floods (12/03/2011 and 01/11/2011), compared with the SSF model (in the SSF1 and SSF2 configurations). The notable difference in the generation of hydrographs is the contribution of the different simulated flowpaths. The proportions of water passing through the soil column

(via sub or surface-soil horizons) are highly variable: with an average of 39 % for the DWF2 model, 53 % for the SSF2 model, 61 % for the DWF1 model and 68 % for the SSF1 model. This is due both to the structural choices (DWF and SSF) which involve a saturation dynamic and the incorporation of different types of flow, and the choice of the parameters which involves flow velocities of differing orders of magnitude.

5   The choice of a model's structure (DWF and SSF) implied differences in soil moisture spatial distribution and dynamic, which in turn impacts the timing of the flow processes. With the DWF structure, the soil moisture distribution is sensitive to the spatial soil depth distribution, as a result of the decrease in the simulated intra soil flows as a function of water table height (cf. Section 3.2, Equation 1). Consequently, the DWF model produced a greater contrast in saturation levels between different areas of the catchment (Figure 13, a, d). With the SSF model, the overall catchment saturation levels is more related to the
10   topography: saturated cells were observed close to the drainage network, and, conversely, lower water content in the upper reaches of the catchments. In fact, for the SSF model, rainfall forcing is mainly involved in saturation of the upper soil layer (the dashed lines in Figure 12-b), which reacts very rapidly to precipitation.

As a result of the contrasted soil moisture dynamic, the flow velocities simulated in the soil showed consecutive differences. At the start of flooding, the SSF structure resulted in an early increase in flow velocities due to a higher and more homogeneous
15   saturation level of the upper soil layer (Figure 12-c). Conversely, with the DWF model that simulated a more heterogeneous spatial saturation of the catchment, the simulated velocities increase was delayed, and the maximum values reached was two to four times lower.

On the surface, only the dynamic in the drainage network is impacted by the choice of the structure, the runoff velocities average over the hillslopes showing the overall same shape whatever the model choose (figure 12- d). The runoff velocities
20   average in the drainage network actually reflected the earlier inlet of the subsurface flow processes through the fast saturation of the upper compartment with the SSF model, as it increased earlier at each beginning of the simulated events (figure 12- e). The DWF model yields a more contrasting variation in the runoff velocities in the drainage network, mirroring variations in soil saturation levels.

The choice of parameters mainly implied different range of value for the velocities simulated, in the soil, on the surface of
25   the hillslope and in the drainage network. For both models, the different parameter sets reflected different time transition and water proportions going through the different pathways. As exemple, the calibration of the C_{kss} and C_{kdw} parameters in the four configurations controlled the order of magnitude in the subsurface velocities (table 3, and figure 13, b, e, h, k). As well, the calibrated C_k (infiltration capacity control) and D_{WB} (depth of the subsoil horizon) parameters controlled the infiltration, leading to more or less high number of cell with saturation excess or infiltration capacity reached (figure 13, c, f, i, l), and
30   consequently to more or less high proportion of runoff over the hillslope (figure 12, d).

Several order of magnitude were actually allowed while respecting the calibration objective because transit time of the different water pathways compensate each other. Taking the example of the four configurations, the selection of plausible parameter sets appears to show a correlation between the parameters C_k, and n_r and n_p. This is actually a general results of the models calibration. There are high values of the Pearson correlation coefficient, especially for the Gardon catchment at Anduze
35   (♯2a): $\rho_{DWF} = 0.46$ and $\rho_{SSF} = 0.18$. This shows the necessity of slowing down flows in the drainage network when a larger

proportion of runoff from the catchments is simulated (i.e C_k is low) and vice versa. In the same way, we noticed positive correlation between C_k ans C_{kss}, C_{kdw} parameters. Thus, as a result of the model calibration, a degree of compensation occurs in the simulated transfer times between the various water paths, from the hillslopes to the drainage network, and from the drainage network towards the outlet.

5   ~~Figure 12 b) shows the different saturation dynamics involved in the DWF and SSF structures. The DWF structure entails continuous drainage of the catchment, including at initialisation. This results in a noticeable continuous decrease in the water content of the soil, including at the beginning of an event, which slows down saturation during the onset of flooding. In contrast, the overall soil water content of the soil profile (solid line) from the catchment simulated by the SSF model are distinguished by a gradual decrease towards a state of equilibrium. Taking Figure 13 (the left-hand column), we can also observe differences~~

10   ~~in spatial dynamics. The DWF model produces a greater contrast in saturation levels between different areas of the catchment. This results from the decrease in the simulated flows as a function of water height (cf. Section 3.2, Equation 1), which makes the draining of each grid cell sensitive to spatial distribution of the soil depth. With the SSF model, the overall catchment saturation levels appear to be more related to the topography: we observe saturation of the cells close to the drainage network, and, conversely, lower water content in the upper reaches of the catchments. In fact, for the SSF model, rainfall forcing is mainly~~

15   ~~involved in saturation of the upper soil layer (the dashed lines in Figure 12 b), which reacts very rapidly to precipitation.~~

~~The flow velocities simulated in the soil (Figure 12 c) are linked to the saturation dynamics. At the start of flooding, the SSF structure results in an early increase in flow velocities due to a higher saturation level of the upper soil layer. Moreover, the flow model chosen and fixing of the parameter m_θ) to simulate the activation of preferential paths in the SSF model (Equation 2 allows a much greater variation of simulated velocities over the short period during which the watershed is~~

20   ~~saturated. Conversely, for the DWF model, the variation interval of simulated velocities is two to four times lower, and the reaction to changes in soil moisture appears to be more linear. The choice of parameters – in particular C_{kss}, here – influences the order of magnitude of the simulated velocities but not the evolution over time, which depends on the structure of the model (the flow modelling equation and the representation of one or two compartments).~~

~~The spatial distributions of the flow velocities in the soil (Figure 13, centre) shows similarities with the areas affected by the~~

25   ~~flows. For the four configurations, the development of flows in the soil only partially reflects the state of saturation, but it is correlated with the physiographic properties of the soil (topography and thickness) and the spatial distribution of meteorological forcing. The different orders of magnitude in the simulated velocities reflect the calibrations of the C_{kss}/C_{kdw} parameters in the four configurations.~~

~~The simulated runoff velocities in the catchment area (Figure 12 d) differs only by their order of magnitude, their evolution being similar. The order of magnitude of the runoff velocities is mainly due to the number of grid cells in the catchment with~~

30   ~~excess infiltration. It reflects the influence of the infiltration control parameter, C_k, and the depth of the subsoil horizon, D_{WB}. At the end of the event, we note the presence of average, non-zero runoff rates on the catchments with the DWF model, a consequence of grid cells that are still saturated.~~

~~The spatial distributions of the flow velocities on the catchments (Figure 13, right) show the two types of functioning~~

35   ~~suggested by the four proposed configurations. Either the runoff is generated by exceeding the storage capacity of the soil; this~~

~~is the case for configurations DWF1 and SSF1, where the grid cells with non-zero runoff velocities correspond to the grid cells where the saturation state of the soil column has been reached, or, runoff is generated by exceeding the infiltration capacity of the soil; this is the case for configurations DWF2 and SSF2 for which the coefficient C_k, set at a low value (cf. Table 3), limits infiltration.~~

5   ~~The changes in runoff velocities in the drainage network (Figure 12-c) reflect the soil saturation dynamics (Figure 12-b). For the SSF model, an early increase in velocities in the drainage network is observed; this is due to the fast saturation of the upper compartment of the soil column, producing consequently interflows through activation of preferential flow paths at the beginning of the event. The DWF model yields a more contrasting variation in the runoff velocities in the drainage network, mirroring variations in soil saturation levels. Finally, it can again be noted that only the structure of the model influences the~~

10   ~~evolution over time of flows in the drainage network, whereas the choice of parameters - particularly, here, n_r and n_p - affects the order of magnitude of the simulated velocities.~~

    ~~Taking the example of the four configurations, the selection of plausible parameter sets appears to show a correlation between the parameters C_k, and n_r and n_p. This is actually a general results of the models calibration. There are high values of the Pearson correlation coefficient, especially for the Gardon catchment at Anduze (♯2a): $\rho_{DWF} = 0.46$ and $\rho_{SSF} = 0.18$. This~~

15   ~~shows the necessity of slowing down flows in the drainage network when a larger proportion of runoff from the catchments is simulated (i.e C_k is low). In all cases, where the values of C_k are low, the transit flows through the ground are also slower (i.e. the values of C_{kss}, C_{kdw} were low). Thus, as a result of the model calibration, a degree of compensation occurs in the simulated transfer times between the various water paths, from the hillslopes to the drainage network, and from the drainage network towards the outlet.~~

20 ## 6   Discussion

6.1   On the hydrological functioning of the catchments studied

~~On the basis of the calibration and performance of the DWF, SSF and SSF-DWF models, the catchments can be divided into several groups:~~

    The benchmark of the models's performance on the catchment set leads to reveal 4 subsets, suggesting 4 distinct hydrological

25   behaviours. According to the modelling assumptions (Section 5.1), the resulting errors in simulating the different stages of the hydrographs (Section 5.2), and according to the catchment properties (Section 2.1), the hydrological behaviour of the catchment can be interpreted subset by subset as follow:

    – The SSF and SSF-DWF models shows better overall performance (with no particular pattern) in the first subset : the Gardon (♯2) and Salz (♯4) catchments. This suggests, on the one hand, rapid catchment reactivity with fast rising flood

30   waters as well as fast flood recession, and on the other, formation of the flows in the soil through local saturation tied to the climate forcing. Although the models exhibit similar performances, the contrasting physiographic characteristics of these catchments suggest that there are different explanations for this better fit of the SSF-DWF model. On the Gardon,
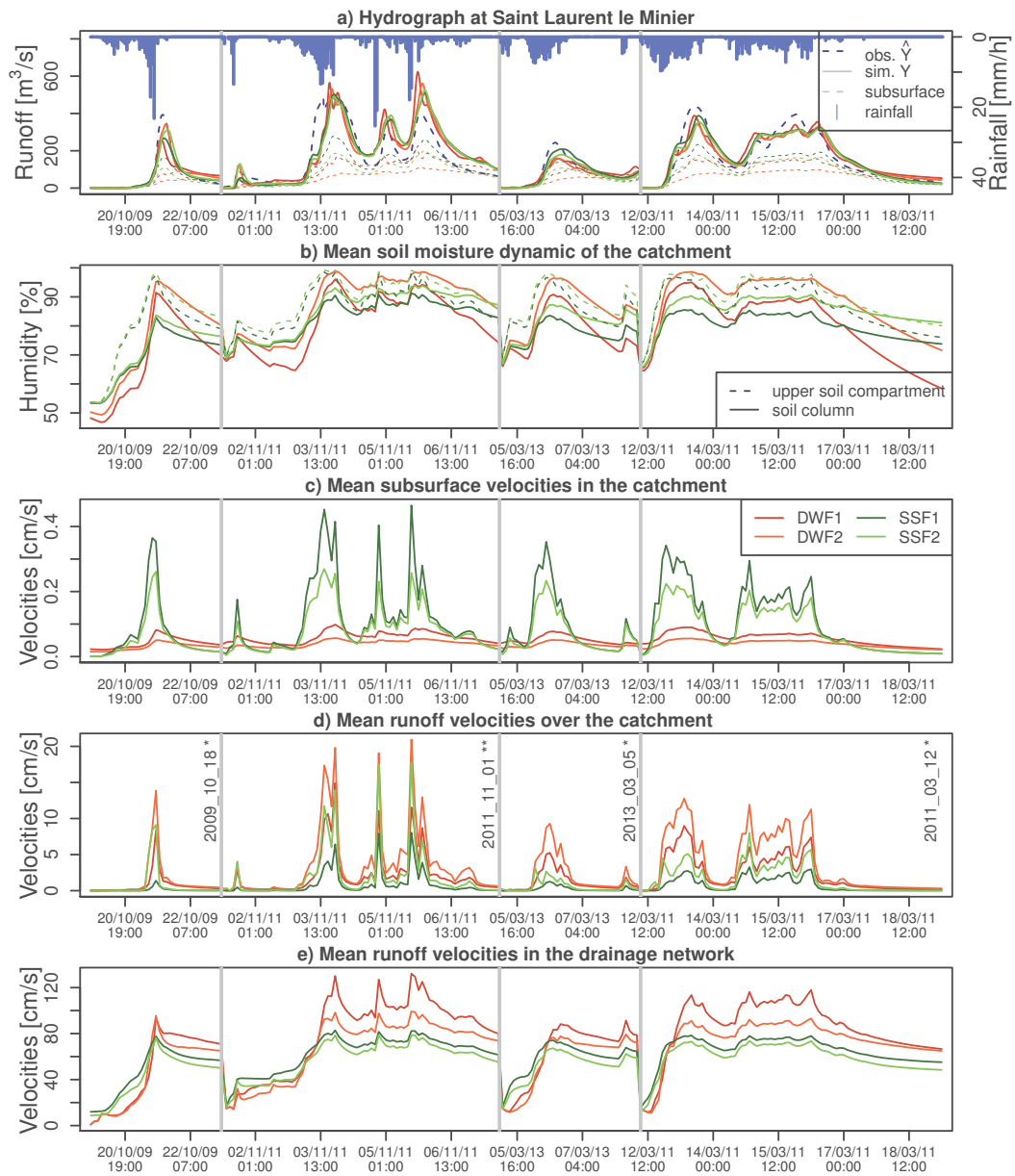
Figure 12. Comparison of the results of four equally plausible simulations on the Hérault at Saint Laurent le Minier (Table 3). a) Flood hydrographs (solid lines) and outlet flows transiting via the soil (dashed lines). b) Evolution in the overall moisture content of the soil column. c) Evolution in simulated mean velocities in the subsoil horizon (DWF model) and in the upper part of the soil column (SSF model). d) Average runoff velocities on the hillslopes. e) Average runoff velocities in the drainage network. (∗): event of calibration; (∗∗): event of validation.
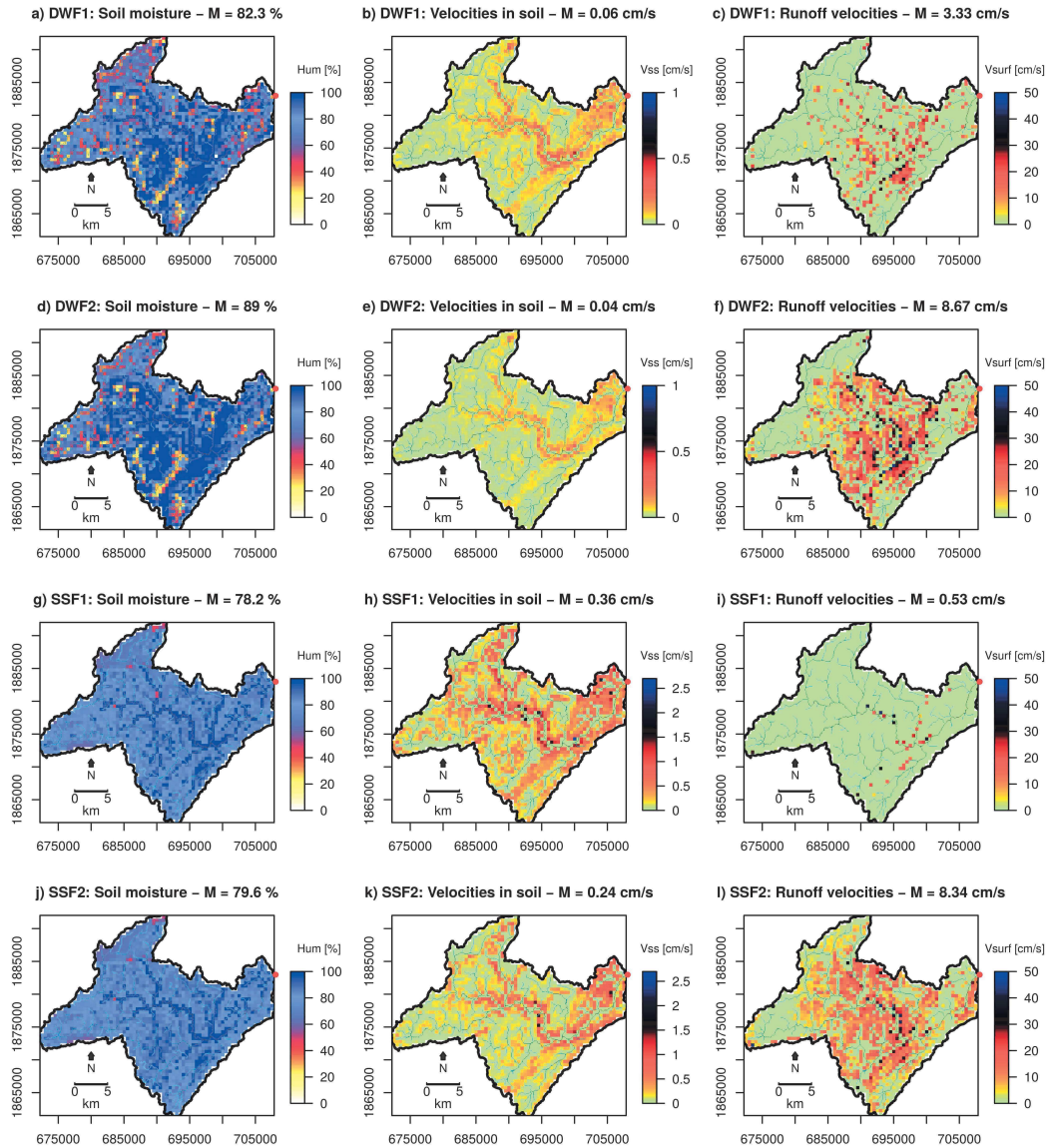
Figure 13. Spatialised outputs for a given moment during the event of 18/10/2009 (during the development of the flood, where $Q = 74\ m^3.s^{-1}$): a-d-g-j) soil moisture conditions simulated, respectively, by the configurations DWF1, DWF2, SSF1, SSF2; b-e-h-k) discharges in the soil simulated, respectively, by the configurations DWF1, DWF2, SSF1, SSF2 (N.B: different colour scheme); c-f-i-l) surface flow velocities simulated, respectively, by the configurations DWF1, DWF2, SSF1, SSF2.

the very high intensities of the observed events (Table 2) and/or the low soil depth (Table 1) may explain the limitations on vertical infiltration due to the properties of the soil and/or geological bedrock. As a result, the rapid formation of a saturated zone at the top of the soil column, favours runoff and subsurface flux by activating preferential paths in

the soil. This interpretation is in agreements with the field studies achieved on a shist upstream sub-catchment of the Gardon, the shist substratum being the predominantly geology of the Gardon catchment (see section 2.1, Ayral et al. (2005); Maréchal et al. (2009, 2013)). On the Salz (♯4), the soil is deeper and the precipitation intensities lower. On the other hand, the geological bedrock composed of marls, sandstone and limestone is assumed to have low permeability and the soil is less conductive due to its predominantly silt-loam texture. As a result, despite the lower forcing intensities, the surface soil can reach saturation, which might explain why the SSF model offers the best fit.

– The considerable hydrological responses, in terms of volumes, on the Ardèche second subset, appear to be linked to hydrological activity at depth, including that taking place during intense floods, as suggested by the better fit of the DWF model. Here, in particular, the model gives a better representation of the relatively slow and uniform hydrological recessions from one event to the next, reflecting an aquifer-type flow whose discharge properties are governed by the properties of the catchment bedrock only. Again, this interpretation is enforced by the field studies achieved this time in a granite experimental sub-catchment localised in the downstream part of the Gardon (Section 2.1, Ayral et al. (2005); Maréchal et al. (2009, 2013)), the Ardèche catchment being granitic. The somewhat delayed flood timing that the structure of the one-compartment model imposes seems to indicate that there are more rapid flows at the beginning of an event, which this model structure is not able to represent. An initial explanation for this may lie in the design of the model: the drainage network being structured into 1 km^2 drained areas. The comparison with the observed hydrographic network for the catchment showed an under-representation of the upstream drainage network, which may have resulted in a delay in the modelling of the signal, despite the model offering a good overall fit. A second possible explanation is the default calibration, which uses a uniform depth of active subsoil horizons, D_{WB}, during a flood. This might mask the appearance of local saturation zones, and the subsequent runoff due to shallow soil and discontinuities in the permeable base layer (for example, in the downstream sedimentary layers, where infiltration tests have shown the appearance of runoff, see Section 2.1). In contrast, the SSF and SSF-DWF models do not display this weakness because the varying nature of soil depths (D_{BDsol}, which determines the depth of the upper compartment) allows the rapid development of flows via preferential paths in the soil blocks, thus enabling the simulation of such local dynamics.

– The third subset consists in the downstream part of the Hérault (♯3a, ♯3b). The models's performances constrast with the Hérault catchment heads (♯3c, ♯3d), suggesting a hydrological behaviours related to the contrasted geological properties. An interpretation of hydrological functioning is nevertheless not possible, given the good overall results offered by other models and that no distinctions can be drawn according to other criteria, such as performance in terms of the simulation of flood recession, for example. ~~in spite of the fairly similar results generated by the models, we observed differences in model performance at the catchment heads (♯3c, ♯3d), where the SSF-DWF model performed significantly better, and at the downstream catchments (♯3a, ♯3b) where the DWF model performed better – for these catchments only – in simulating rising flood waters.~~

– The last subset consists in the catchment heads (♯1d, ♯3c, and ♯3d). We observe superior performances from the DWF and SSF-DWF models, with a particular improvement in the forecasting of rising flood waters when using the SSF-DWF

34

model. This suggests the presence of several types of flow in the soil with strong support from flows at depth, which corroborates the high mean inter-annual discharges associated with these catchments, and additionally the presence of rapidly formed flows, providing a good simulation of the rising flood waters. It should be noted that, here again, modelling the drainage network for an area greater than that observed on these steep-sloped catchments can also affect the results.

6.2 Overcoming the remaining uncertainty

The substantiated results in Section 5 explained the remaining equifinality issue and highlighted the points of differentiation between models. The benchmark of modelling configurations, scanning the different simulated processes (section 5.3.2) revealed the actual internal differences - discrepancy in the soil saturation dynamic and in the resulting set up of the intra soil flow processes - and the clearing mechanisms through the counterbalancing effects of the velocities and the flow proportions simulated. This points out direct guidelines for improving the calibration of the models. A multi objective calibration strategy function should be prospected according to the prevailing observation capacities:

– The distinguished spatial saturation patterns generated by the DWF and SSF structures enlighted the interest of a spatial distribution assessment. The current availability of high-resolution telemetry measurements with high spatial coverage (for example, Sentinel-1-based satellite Earth Observation data (Enenkel et al., 2016; Cenci et al., 2017)) offers this opportunity. The temporal resolution (up to six days) is not adapted to flash-flood time scales and prevents their use for real-time evaluation of hydrological simulations. However, observing some saturation patterns for a number of events during, or shortly after, an episode would provide an interesting research avenue, in terms of distinguishing the hydrological reactions of the catchments in a spatialised manner.

– The remaining uncertainty when assessing the different flow proportions points out the need of a specific control to one of those flow processes assessed. Either flow partitioning measurement or the experimental assessment of the transit time along the different pathways, at catchment scale, would be judicious observations to integrate. Imaginatives indirect methods to detect the degree of surface flows in a flood, such as diatom tracing (Pfister et al., 2017b), suspended particles or water turbidity measurements provide new avenues for partitioning the hydrographs. As well, a specific calibration of the drainage network, focused on the flood propagation, through intermediate discharge time series evaluation, is promising, as it would provide stronger constraint on surface flow velocities, and consequently limit the counterbalancing effects between the transit times of the different flow processes .

The description of the a priori modelling errors (section 5.2), representing the spectrum of possibilities for a given rainfall forcing, revealed the hydrograph's simulated points mostly controlled by the model's structure, namely the early rising limb for the DWF model and the flood recession for the SSF and SSF-DWF models. It actually points the way towards an optimal consideration of these parts of the hydrograph, when calibrating the models. In the same way, the relevant information of these parts of the hydrograph must guide toward high temporal resolution of the river level measurements, the rising and receding flood stages being short periods, and toward efforts for reducing the uncertainty of the rating curve at low and moderate flow,

rather than getting extreme discharge measurments, including as example hysterisis of the discharge curves (Le Coz et al., 2014).

~~In this comparison of the simulated processes, the equivalence of the configurations, presented in terms of integrated modelling of the flow at the outlet, the equivalence is refuted by the differences generated in:~~

- ~~the proportions of water passing through the ground or over the surface, linked to the infiltration control mechanism, governed here by the chosen configuration.~~

- ~~the saturation dynamics of the soil, which are linked to the flows developed in the soil, governed here by the structure of the chosen model.~~

- ~~the orders of magnitude of the simulated flow velocities, which are related to both the choice of the model structure and the parametrization.~~

~~A detailed description of the configurations, together with an estimation of the structural errors in the modelling, allows better visualisation of what the different hypotheses of hydrological functioning involve, and points to new options for assessing models, as well as the potential contributions from new knowledge/observations:~~

- ~~The DWF and SSF structures generate vertical dynamics and distinct spatial saturation patterns. The current availability of high-resolution telemetry measurements with high spatial coverage (for example, Sentinel-1-based satellite Earth Observation data~~ (Enenkel et al., 2016; Cenci et al., 2017)) ~~offers the opportunity to conduct a qualitative assessment of soil moisture patterns. The temporal resolution (up to six days) is not adapted to flash-flood time scales and prevents their use for real-time evaluation of hydrological simulations. However, observing some saturation patterns for a number of events during, or shortly after, an episode would provide an interesting research avenue, in terms of distinguishing the hydrological reactions of the catchments in a spatialised manner, which could help confirm the accuracy of the models tested.~~

- ~~The different flow proportions related to the structure of the model selected (use of the DWF model tends to result in more runoff on slopes) and its calibration emerge as new objectives for constraints, because they imply distinct hydrological behaviours. Tracing flows via isotopic measurements is not suited to the meso-scale catchments studied, nor to the spatial representation of the MARINE model, which assumes an instantaneous and complete mixing of the water volume and does not calculate residence times~~ (McDonnell and Beven, 2014). ~~Conversely, the use of an indicator of the presence of runoff, such as diatom tracing~~ (Pfister et al., 2017b), ~~suspended particles or the turbidity of water, offers an indirect means of detecting the degree of surface flows in a flood, and could make it possible to better constrain the partitioning of the hydrographs.~~

- ~~The different proportions in the simulated flows are allowed by the simulation of transfer times, of varying length, on the different water paths: runoff, flows through the soil, and via the drainage network. These arise, in particular, as a result of the calibration of flow velocities having different orders of magnitude. It would be difficult to envisage a~~

constraint on the orders of magnitude of the simulated velocities because the scale of modelling (where, as a reminder, ~~dx 100 m~~) encompasses macrostructures (for example, preferential paths) that cannot be quantified without detailed analysis. Conversely, separate optimisation of the drainage network, and the parameters that control flow on the hillslopes, would limit the possible compensations between the transfer-times modelled. In particular, intermediate hydrometric stations could be used to calibrate the transfer function of the hydrological signal from the drainage network to the outlet.

- Finally, the evaluation based on the ability of the simulations to reproduce the characteristic stages of floods demonstrates the greater impact of the choice of model structure on the rise and recession of floods. They, therefore, point the way towards an optimal consideration of these parts of the hydrograph. The choice of an evaluation score based on a comparison of time series proved pertinent as a result of its sensitivity at the onset of a flood. The uncertainty in flow measurements was systematically taken into account for all catchments. In order to refine the information on the measured flows (and, more precisely, the average range of flows for rising and receding floods), it would make sense to carry out specific calculations for each hydrometric station and its associated discharge curves (using the Baratin Method (Le Coz et al., 2014), for example).

7 Conclusions and Perspectives

7.1 Summary of the study's objectives and methodology

The objective of the study was to improve our understanding of flash flooding on the French Mediterranean Arc. In particular, attention was paid to the dynamics of soil saturation in catchments during these events, and their possible relationship with the physiographic diversity encountered. The method used consisted in considering hydrological models as a diagnostic tool to test hypotheses about the functioning of catchments.

Based on the structure of the MARINE model - a hydrological model with a physical and distributed basis - three types of dynamic of soil saturation were postulated and tested. In the first case (the DWF model), we assumed an aquifer dynamic, with infiltration at depth, and the generation of strong base support, according to the volume of infiltrated water; in the second case (the SSF model), it was the activation of preferential paths at the soil/altered rock interface that generated the majority of the flows passing through the soil, with the lower part of the soil column serving only as a storage reservoir; and in the third case (the SSF-DWF model), there was flow generation via both the activation of preferential pathways, initially by saturation of the top of the soil column, and a significant increase in the base flux via the subsequent infiltration of water present at deeper levels.

The same calibration strategy was used for the three models on a set of 12 catchments which are representative of the diverse characteristics of the Mediterranean Arc. Whether a model offers a good fit was evaluated on the basis of: scores representing overall, or partial model performance in terms of simulating the hydrographs; the proportions of the processes simulated; and the timing and form of flood recession.

7.2   Conclusions on our understanding of the processes involved

From the application and validation of the three hydrological models, the 12 catchments of the study could be classified into four categories: i) the Gardon and Salz catchments, for which the SSF model is better suited to reproducing the hydrological signal. For these catchments, this highlights the importance of local and surface soil dynamics in the generation of flows, especially at the beginning of a flood; (ii) the Ardèche catchments, for which the DWF model most accurately reproduce the observed flows. This ~~demonstrates~~ indic͟a͟t͟e͟s more regular and integrated hydrological functioning at the catchment level, with the flows generated being directly related to the moisture history and rainfall volumes; (iii) the Hérault catchments at Valleraugue and La Terrisse, and the Ardèche catchment at Meyras, which have steep-sloped catchment heads, where the SSF-DWF model stands out, suggesting both sustained and significant hydrological activity at depth during flash floods, and surface activity in the establishment of early flows at the beginning of events; (iv) the Hérault catchments at Laroque and Saint-Laurent-le-Minier, for which no model shows any significant difference.

T͟h͟e͟ ͟m͟o͟d͟e͟l͟l͟i͟n͟g͟ ͟r͟e͟s͟u͟l͟t͟s͟ ͟h͟e͟l͟p͟ ͟t͟o͟ ͟d͟r͟a͟w͟ ͟c͟o͟n͟s͟i͟s͟t͟e͟n͟t͟ ͟a͟s͟s͟u͟m͟p͟t͟i͟o͟n͟s͟ ͟o͟n͟ ͟h͟y͟d͟r͟o͟l͟o͟g͟i͟c͟a͟l͟ ͟b͟e͟h͟a͟v͟i͟o͟u͟r͟s͟,͟ ͟w͟h͟i͟c͟h͟ ͟c͟o͟r͟r͟o͟b͟o͟r͟a͟t͟e͟ ͟w͟h͟e͟n͟ ͟a͟v͟a͟i͟l͟a͟b͟l͟e͟,͟ t͟h͟e͟ ͟k͟n͟o͟w͟l͟e͟d͟g͟e͟ ͟a͟n͟d͟ ͟o͟b͟s͟e͟r͟v͟a͟t͟i͟o͟n͟s͟ ͟o͟n͟ ͟t͟h͟e͟ ͟o͟v͟e͟r͟a͟l͟l͟ ͟h͟y͟d͟r͟o͟l͟o͟g͟i͟c͟a͟l͟ ͟f͟u͟n͟c͟t͟i͟o͟n͟i͟n͟g͟ ͟o͟f͟ ͟t͟h͟e͟ ͟c͟a͟t͟c͟h͟m͟e͟n͟t͟s͟,͟ ͟o͟r͟ ͟t͟h͟e͟ ͟e͟x͟p͟e͟r͟i͟m͟e͟n͟t͟a͟l͟ ͟e͟s͟t͟i͟m͟a͟t͟i͟o͟n͟s͟ ͟o͟f͟ f͟l͟o͟w͟ ͟p͟r͟o͟c͟e͟s͟s͟e͟s͟.͟ ~~For each catchment, the best-performing models were those where results reflected the available knowledge and observations on the overall hydrological functioning of the catchments, and where estimates of the different flow processes corresponded to experimental observations.~~ The results suggest that the behaviour of catchments under extreme forcing is a continuation of the hydrological functioning normally encountered. Several earlier studies have pointed to a potential correspondence between hydrological functioning and the nature of the geological bedrock. This is in evidence on the Hérault, where the evaluation of the three models highlighted different hydrological behaviours which are linked to differences in the geological nature of the catchments. A͟l͟s͟o͟,͟ ͟t͟h͟e͟ ͟G͟a͟r͟d͟o͟n͟ ͟a͟n͟d͟ ͟A͟r͟d͟è͟c͟h͟e͟ ͟c͟a͟t͟c͟h͟m͟e͟n͟t͟s͟,͟ ͟w͟h͟i͟c͟h͟ ͟h͟a͟v͟e͟ ͟r͟e͟s͟p͟e͟c͟t͟i͟v͟e͟l͟y͟ ͟m͟a͟i͟n͟l͟y͟ ͟s͟c͟h͟i͟s͟t͟o͟s͟e͟ ͟a͟n͟d͟ g͟r͟a͟n͟i͟t͟i͟c͟ ͟g͟e͟o͟l͟o͟g͟y͟,͟ ͟e͟x͟h͟i͟b͟i͟t͟e͟d͟ ͟d͟i͟f͟f͟e͟r͟e͟n͟t͟ ͟b͟e͟h͟a͟v͟i͟o͟u͟r͟s͟,͟ ͟i͟n͟ ͟c͟o͟r͟r͟e͟s͟p͟o͟n͟d͟a͟n͟c͟e͟ ͟w͟i͟t͟h͟ ͟t͟h͟e͟ ͟f͟i͟e͟l͟d͟ ͟e͟x͟p͟e͟r͟i͟m͟e͟n͟t͟a͟l͟ ͟s͟t͟u͟d͟i͟e͟s͟ ͟o͟f͟ ͟t͟h͟e͟ ͟r͟e͟g͟i͟o͟n͟.͟ On the other hand, the similar hydrological behaviours of the Gardon and the sedimentary Salz catchment are quite surprising owing to their contrasted geological and other physiographical properties. These results, however, did not contradict the earlier studies, which suggest a relationship between storage capacity in the substratum and the nature of the geological bedrock, while the similarity highlighted here concern the formation of flows in the soil.

A͟n͟o͟t͟h͟e͟r͟ ͟o͟b͟j͟e͟c͟t͟i͟v͟e͟ ͟o͟f͟ ͟t͟h͟e͟ ͟s͟t͟u͟d͟y͟,͟ ͟w͟a͟s͟ ͟t͟h͟e͟ ͟a͟s͟s͟e͟s͟s͟m͟e͟n͟t͟ ͟o͟f͟ ͟t͟h͟e͟ ͟f͟l͟o͟w͟ ͟p͟r͟o͟c͟e͟s͟s͟e͟s͟.͟ ͟O͟w͟i͟n͟g͟ ͟t͟o͟ ͟t͟h͟e͟ ͟e͟q͟u͟i͟f͟i͟n͟a͟l͟i͟t͟y͟ ͟i͟s͟s͟u͟e͟s͟,͟ ͟t͟h͟e͟ ͟a͟s͟s͟e͟s͟s͟m͟e͟n͟t͟ r͟e͟m͟a͟i͟n͟s͟ ͟u͟n͟c͟e͟r͟t͟a͟i͟n͟.͟ ͟N͟e͟v͟e͟r͟t͟h͟e͟l͟e͟s͟s͟,͟ ͟t͟h͟e͟ ͟a͟n͟a͟l͟y͟s͟i͟s͟ ͟o͟f͟ ͟t͟h͟e͟ ͟i͟n͟t͟e͟r͟n͟a͟l͟ ͟p͟r͟o͟c͟e͟s͟s͟e͟s͟ ͟h͟i͟g͟h͟l͟i͟g͟h͟t͟e͟d͟ ͟t͟h͟e͟ ͟c͟o͟m͟p͟e͟n͟s͟a͟t͟i͟o͟n͟ ͟b͟e͟t͟w͟e͟e͟n͟ ͟t͟h͟e͟ ͟f͟l͟o͟w͟ ͟p͟r͟o͟c͟e͟s͟s͟e͟s͟ s͟i͟m͟u͟l͟a͟t͟e͟d͟ ͟-͟ ͟i͟n͟ ͟t͟h͟e͟ ͟d͟r͟a͟i͟n͟a͟g͟e͟ ͟n͟e͟t͟w͟o͟r͟k͟ ͟a͟n͟d͟ ͟i͟n͟ ͟t͟h͟e͟ ͟h͟i͟l͟l͟s͟l͟o͟p͟e͟ ͟-͟ ͟t͟h͟a͟t͟ ͟i͟s͟ ͟m͟a͟d͟e͟ ͟p͟o͟s͟s͟i͟b͟l͟e͟ ͟t͟h͟r͟o͟u͟g͟h͟ ͟a͟ ͟w͟i͟d͟e͟ ͟r͟a͟n͟g͟e͟ ͟o͟f͟ ͟f͟l͟o͟w͟ ͟v͟e͟l͟o͟c͟i͟t͟i͟e͟s͟ ͟s͟i͟m͟u͟l͟a͟t͟e͟d͟,͟ a͟s͟ ͟b͟e͟i͟n͟g͟ ͟t͟h͟e͟ ͟m͟a͟i͟n͟ ͟r͟e͟a͟s͟o͟n͟ ͟o͟f͟ ͟t͟h͟e͟ ͟e͟q͟u͟i͟f͟i͟n͟a͟l͟i͟t͟y͟ ͟i͟s͟s͟u͟e͟.͟ ͟T͟h͟e͟ ͟d͟e͟t͟a͟i͟l͟e͟d͟ ͟d͟e͟s͟c͟r͟i͟p͟t͟i͟o͟n͟ ͟e͟n͟a͟b͟l͟e͟s͟ ͟f͟i͟n͟a͟l͟l͟y͟ ͟t͟o͟ ͟p͟r͟o͟p͟o͟s͟e͟ ͟n͟e͟w͟ ͟s͟t͟r͟a͟t͟e͟g͟i͟e͟s͟ ͟f͟o͟r͟ ͟a͟ ͟b͟e͟t͟t͟e͟r͟ c͟o͟n͟s͟t͟r͟a͟i͟n͟t͟ ͟o͟f͟ ͟t͟h͟e͟ ͟m͟o͟d͟e͟l͟s͟.͟

~~Lastly, identifying the most pertinent hydrological models for each catchment enables the key elements in the generation of flash floods to be highlighted, which, in turn, could serve to further develop methods for forecasting flash floods. For example, distinctions in hydrological behaviour revealed between the catchments of the Gardon and the Ardèche may explain that taking into account the spatial nature of precipitation in a flash flood forecasting method results in an improvement only~~

~~on the Gardon and not on the Ardèche. Indeed, in the present study, the Gardon catchment appears to be more sensitive to the local dynamic of the soil water content than the Ardèche one, corroborating the sensitivity to spatial distribution of the rainfall revealed in.~~ Lastly, identifying the most pertinent hydrological models for each catchment enables the key elements in the generation of flash floods to be highlighted, which, in turn, could serve to further develop methods for forecasting flash floods.

5 For example, distinctions in hydrological behaviour revealed between the catchments of the Gardon and the Ardèche - the first one appearing more reactive with important runoff and subsurface flows through preferential flowpaths - might shift towards different considerations when setting up a flash flood forecasting method over those contrasted area. It corroborates the results of Douinot et al. (2016) which highlighted contrasted impacts of taking into account the spatial variability of precipitation in a flash flood forecasting method. These contrasted impacts can indeed be explained by the more pronounced spatial variability

10 of the rainfall over the Gardon catchment, but also by the local dynamic of the soil water content of the Gardon catchment revealed in the present study.

7.3 Conclusions about the method used

The use of the hydrological model as a diagnostic tool allowed the classification of the catchments studied. It also contributes to the overall knowledge of these catchments in order to improve understanding of hydrological functioning during flash floods.

15 The study also demonstrates: i) the complementarity of field observations in the interpretation of results, ii) the limitations in the evaluation and drawing of distinctions between models when constrained solely on the basis of the reproduction of an integrated response; and iii) the contribution that an analysis of equally performing parameter sets and possible model functioning can make to guide the choice of new and better constraints, and the strategic observations that need to be made in order to differentiate between equally plausible models. Lastly, distinguishing between models based on the evolution of

20 internal variables - flow velocities and soil saturation states - makes it possible to highlight the value added by the descriptive potential of a distributed model with a physical basis, such as MARINE.

References

Anquetin, S., Braud, I., Vannier, O., Viallet, P., Boudevillain, B., Creutin, J.-D., and Manus, C.: Sensitivity of the hydrological response to the variability of rainfall fields and soils for the Gard 2002 flash-flood event, Journal of Hydrology, 394, 134 – 147, https://doi.org/10.1016/j.jhydrol.2010.07.002, flash Floods: Observations and Analysis of Hydrometeorological Controls, 2010.

5 Arnaud, P., Lavabre, J., and Masson, J.: Amélioration des performances d'un modèle stochastique de génération de hyétogrammes horaires: application au pourtour méditerranéen français. (Improving the performances of a hourly time step stochastic model: case of the French Mediterranean area), Revue des sciences de l'eau/Journal of Water Science, 12, 251–271, https://doi.org/10.7202/705351ar, http://id.erudit.org/iderudit/705351ar, 1999.

Ayral, P.-A., Sauvagnargues-Lesage, S., and Bressand, F.: Contribution à la spatialisation du modèle opérationnel de prévision des crues

10 éclair ALTHAÏR.(Contribution to the spatial distribution of the operational flash flood forecasting model ALTHAÏR), Etude de géographie physiqueéographie physique, 32, 75–97, 2005.

Badoche-Jacquet, G., Blanchet, F., and Blanchet, B.: Procédé et dispositif pour l'évaluation des précipitations sur une zone de terrain. (Procedure and apparatus for evaluating field precipitation), Patent EP0578544B1, p. Rhea, 1992.

Borga, M., Gaume, E., Creutin, J. D., and Marchi, L.: Surveying flash floods: gauging the ungauged extremes, Hydrological processes, 22,

15 3883, https://doi.org/10.1002/hyp.7111, 2008.

Borga, M., Stoffel, M., Marchi, L., Marra, F., and Jakob, M.: Hydrogeomorphic response to extreme rainfall in headwater systems: flash floods and debris flows, Journal of Hydrology, 518, 194–205, https://doi.org/10.1016/j.jhydrol.2014.05.022, 2014.

Bouvier, C., Patris, N., Freydier, R., Guilhe-Batiot, C., Seidel, J.-L., Taupin, J.-D., Brunet, P., and Remes, A.: The Floodscale Experiment at the Hillslope Scale: Physico-Chemical Characterization of Runoff Contributing to Flash Flood Generation in the Small Catchment of

20 Valescure, France, in: 10th HyMeX workshop 4-7 July 2017, Barcelona, Spain, 2017.

Braud, I. and Vandervaere, J. P.: Analysis of infiltration tests performed in the Claduègne catchment in May-June 2012, contribution to WP3.4 "Documentation and mapping of soil hydraulic properties, soil geometry and vegetation cover of small catchments", Tech. Rep. 66, FloodScale (2012-2016) ANR project, http://mistrals.sedoo.fr/?editDatsId=1321., 2015.

Braud, I., Ayral, P., Bouvier, C., Branger, F., Delrieu, G., Le Coz, J., Nord, G., Vandervaere, J., Anquetin, S., Adamovic, M., Andrieu, J.,

25 Batiot, C., Boudevillain, B., Brunet, P., Carreau, J., Confoland, A., F, D. L. J., Domergue, J., Douvinet, J., Dramais, G., Freydier, R., Gérard, S., Huza, J., Leblois, E., Le Bourgeois, O., Le Boursicaud, R., Marchand, P., Martin, P., Nottale, L., Patris, N., Renard, B., Seidel, J., Taupin, J., Vannier, O., Vincendon, B., and Wijbrans, A.: Multi-scale hydrometeorological observation and modelling for flash-flood understanding, Hydrology and Earth System Sciences, 18, 3733–3761, https://doi.org/10.5194/hess-18-3733-2014, 2014.

Braud, I., Andrieu, H., Ayral, P.-A., Bouvier, C., Branger, F., Carreau, J., Delrieu, G., Douvinet, J., R., F., E., L., Le Coz, J., Martin, H.,

30 Nord, G., Patris, N., S., P., Renard, B., Seidel, J.-L., and Vandervaere, J.: FloodScale : Observation et modélisation hydro-météorologique multi-échelles pour la compréhension et la simulation des crues éclairs. Multi-scale hydrometeorological observation and modelling for flash-flood understanding,, final report of the projet ANR-2011 BS56 027 01, 2016a.

Braud, I., Ayral, P., Bouvier, C., Branger, F., Delrieu, G., Dramais, G., Le Coz, J., Leblois, E., Nord, G., and Vandervaere, J.: Advances in flash floods understanding and modelling derived from the FloodScale project in south-east France, in: 3rd European Con-

35 ference on Flood Risk Management, Innovation, Implementation, Integration (FLOODrisk 2016), vol. 7, p. 04005, Lyon, France, https://doi.org/10.1051/e3sconf/20160704005, https://hal.archives-ouvertes.fr/hal-01425992, fLOODrisk 2016, 3rd European Conference on Flood Risk Management, Innovation, Implementation, Integration, Lyon, FRA, 18-/10/2016 - 20/10/2016, 2016b.

Buytaert, W. and Beven, K.: Models as multiple working hypotheses: hydrological simulation of tropical alpine wetlands, Hydrological Processes, 25, 1784–1799, https://doi.org/10.1002/hyp.7936, 2011.

Camarasa-Belmonte, A. M.: Flash floods in Mediterranean ephemeral streams in Valencia Region (Spain), Journal of Hydrology, 541, 99 – 115, https://doi.org/10.1016/j.jhydrol.2016.03.019, flash floods, hydro-geomorphic response and risk management, 2016.

Cassardo, C., Balsamo, G. P., Cacciamani, C., Cesari, D., Paccagnella, T., and Pelosini, R.: Impact of soil surface moisture initialization on rainfall in a limited area model: a case study of the 1995 South Ticino flash flood, Hydrological Processes, 16, 1301–1317, https://doi.org/10.1002/hyp.1063, 2002.

Cenci, L., Pulvirenti, L., Boni, G., Chini, M., Matgen, P., Gabellani, S., Squicciarino, G., and Pierdicca, N.: An evaluation of the potential of Sentinel 1 for improving flash flood predictions via soil moisture–data assimilation, Advances in Geosciences, 44, 89–100, https://doi.org/10.5194/adgeo-44-89-2017, 2017.

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

Clark, M. P., Kavetski, D., and Fenicia, F.: Pursuing the method of multiple working hypotheses for hydrological modeling, Water Resources Research, 47, https://doi.org/10.1029/2010WR009827, http://dx.doi.org/10.1029/2010WR009827, w09301, 2011.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Brekke, L. D., et al.: A unified approach for process-based hydrologic modeling: 1. Modeling concept, Water Resources Research, 51, 2498–2514, https://doi.org/10.1002/2015WR017198, 2015a.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Gochis, D. J., et al.: A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies, Water Resources Research, 51, 2515–2542, https://doi.org/10.1002/2015WR017200, 2015b.

Colmet Daage, A., Ricci, S., Sanchez-Gomez, E., and Valérie, B. E.: Climate change impacts on precipitation extremes, flows and flash floods in Mediterranean mesoscale catchments, in: International Congress on Environmental Modelling and Software, 2016.

Coxon, G., Freer, J., Wagener, T., Odoni, N., and Clark, M.: Diagnostic evaluation of multiple hypotheses of hydrological behaviour in a limits-of-acceptability framework for 24 UK catchments, Hydrological processes, 28, 6135–6150, https://doi.org/doi.org/10.1002/hyp.10096, 2014.

de l'Ecologie du développement durable et de l'énergie, M.: HYDRO, http://www.hydro.eaufrance.fr/, http://www.hydro.eaufrance.fr/, 2015.

Delrieu, G., Nicol, J., Yates, E., Kirstetter, P.-E., Creutin, J.-D., Anquetin, S., Obled, C., Saulnier, G.-M., Ducrocq, V., Gaume, E., Payrastre, O., Andrieu, H., Ayral, P.-A., Bouvier, C., Neppel, L., Livet, M., Lang, M., du Châtelet, J. P., Walpersdorf, A., and Wobrock, W.: The Catastrophic Flash-Flood Event of 8 - 9 September 2002 in the Gard Region, France: A First Case Study for the Cévennes Vivarais Mediterranean Hydrometeorological Observatory, J. Hydrometeor, 6, 34–52, https://doi.org/10.1175/JHM-400.1, 2005.

Douinot, A.: Analyse des processus d'écoulement lors de crues à cinétique rapide sur l'arc méditerranéen, Ph.D. thesis, Université Paul Sabatier, Toulouse, 2016.

Douinot, A., Roux, H., Garambois, P.-A., Larnier, K., Labat, D., and Dartus, D.: Accounting for rainfall systematic spatial variability in flash flood forecasting, Journal of Hydrology, 541, 359–370, https://doi.org/10.1016/j.jhydrol.2015.08.024, 2016.

Douinot, A., Roux, H., and Dartus, D.: Modelling errors calculation adapted to rainfall–Runoff model user expectations and discharge data uncertainties, Environmental Modelling & Software, 90, 157–166, https://doi.org/10.1016/j.envsoft.2017.01.007, 2017.

Drobinski, P., Ducrocq, V., Alpert, P., Anagnostou, E., Béranger, K., Borga, M., Braud, I., Chanzy, A., Davolio, S., Delrieu, G., et al.: HyMeX: a 10-year multidisciplinary program on the mediterranean water cycle, Bulletin of the American Meteorological Society, 95, 1063–1082, https://doi.org/10.1175/BAMS-D-12-00242.1, 2014.

Enenkel, M., Reimer, C., Dorigo, W., Wagner, W., Pfeil, I., Parinussa, R., and De Jeu, R.: Combining satellite observations to develop a global soil moisture product for near-real-time applications, Hydrology and Earth System Sciences, 20, 4191, https://doi.org/10.5194/hess-20-4191-2016, 2016.

Faccini, F., Paliaga, G., Piana, P., Sacchini, A., and Watkins, C.: The Bisagno stream catchment (Genoa, Italy) and its major floods: geomorphic and land use variations in the last three centuries, Geomorphology, 273, 14 – 27, https://doi.org/10.1016/j.geomorph.2016.07.037, 2016.

Fenicia, F., Kavetski, D., and Savenije, H. H.: Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, Water Resources Research, 47, https://doi.org/10.1029/2010WR010174, 2011.

Fenicia, F., Kavetski, D., Savenije, H. H., Clark, M. P., Schoups, G., Pfister, L., and Freer, J.: Catchment properties, function, and conceptual model representation: is there a correspondence?, Hydrological Processes, 28, 2451–2467, https://doi.org/10.1002/hyp.9726, 2014.

Fenicia, F., Kavetski, D., Savenije, H. H. G., and Pfister, L.: From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions, Water Resources Research, 52, 954–989, https://doi.org/10.1002/2015WR017398, 2016.

Garambois, P., Larnier, K., Roux, H., Labat, D., and Dartus, D.: 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, 2014.

Garambois, P., Roux, H., Larnier, K., Labat, D., and Dartus, D.: Characterization of catchment behaviour and rainfall selection for flash flood hydrological model calibration: catchments of the eastern Pyrenees, Hydrological Sciences Journal, 60, 424–447, https://doi.org/10.1080/02626667.2014.909596, 2015a.

Garambois, P.-A.: Etude régionale des crues éclair de l'arc méditerranéen français. Elaboration de méthodologies de transfert à des bassins versants non jaugés (Regional study of the flash flood events in the French Mediterranean area: on the methods for transfering information to ungauged catchment), Ph.D. thesis, Institut de Mécanique des Fluides de Toulouse - IMFT (Toulouse, France), 2012.

Garambois, P. A., Roux, H., Larnier, K., Castaings, W., and Dartus, D.: Characterization of process-oriented hydrologic model behavior with temporal sensitivity analysis for flash floods in Mediterranean catchments, Hydrology and Earth System Sciences, 17, 2305–2322, https://doi.org/10.5194/hess-17-2305-2013, 2013.

Garambois, P.-A., Roux, H., Larnier, K., Labat, D., and Dartus, D.: Parameter regionalization for a process-oriented distributed model dedicated to flash floods, Journal of Hydrology, 525, 383–399, 2015b.

Gaume, E. and Borga, M.: Etat des connaissances récentes acquises sur les crues éclair en Europe: bilan du projet de recherches européen HYDRATE (2006-2010) (State of the art on the knowledge recently acquired on flash floods in Europe: results of the European research project HYDRATE (2006-2010)), La Houille Blanche, pp. 24–30, https://doi.org/10.1051/lhb/2013012, 2013.

Gaume, E., Bain, V., Bernardara, P., Newinger, O., Barbuc, M., Bateman, A., Blaskovicova, L., Blöschl, G., Borga, M., Dumitrescu, A., Daliakopoulos, I., Garcia, J., Irimescu, A., Kohnova, S., Koutroulis, A., Marchi, L., Matreata, S., Medina, V., Preciso, E., Sempere-Torres, D., Stancalie, G., Szolgay, J., Tsanis, I., Velasco, D., and Viglione, A.: A compilation of data on European flash floods, Journal of Hydrology, 367, 70 – 78, https://doi.org/10.1016/j.jhydrol.2008.12.028, 2009.

Gourley, J. J., Flamig, Z. L., Vergara, H., Kirstetter, P.-E., III, R. A. C., Argyle, E., Arthur, A., Martinaitis, S., Terti, G., Erlingis, J. M., Hong, Y., and Howard, K. W.: The FLASH Project: Improving the Tools for Flash Flood Monitoring and Prediction across the United States, Bulletin of the American Meteorological Society, 98, 361–372, https://doi.org/10.1175/BAMS-D-15-00247.1, 2017.

Habets, F., Boone, A., Champeaux, J.-L., Etchevers, P., Franchisteguy, L., Leblois, E., Ledoux, E., Le Moigne, P., Martin, E., Morel, S., et al.: The SAFRAN-ISBA-MODCOU hydrometeorological model applied over France, Journal of Geophysical Research: Atmospheres (1984–2012), 113, https://doi.org/10.1029/2007JD008548, 2008.

Hegedüs, P., Czigány, S., Balatonyi, L., and Pirkhoffer, E.: Analysis of soil boundary conditions of flash floods in a small basin in SW Hungary, Open Geosciences, 5, 97–111, https://doi.org/10.2478/s13533-012-0119-6, 2013.

Jacq, V.: Inventaire des situations à précipitations diluviennes sur la région Languedoc-Roussillon, la Provence-Alpes Cotes d'azur et la Corse, période 1958-1994 (Inventory of the extreme rainfall events that have occured between 1958-1994 in the Languedoc-Roussillon,the Provence-Alpes Cotes d'azur and the Corse regions), Phénomènes remarquables, 1994.

Katsura, S., Kosugi, K., Yamakawa, Y., and Mizuyama, T.: Field evidence of groundwater ridging in a slope of a granite watershed without the capillary fringe effect, Journal of Hydrology, 511, 703–718, https://doi.org/10.1016/j.jhydrol.2014.02.021, 2014.

Katsuyama, M., Ohte, N., and Kabeya, N.: Effects of bedrock permeability on hillslope and riparian groundwater dynamics in a weathered granite catchment, Water Resources Research, 41, https://doi.org/10.1029/2004WR003275, 2005.

Kirchner, J. W.: Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, Water Resources Research, 45, https://doi.org/10.1029/2008WR006912, 2009.

Laganier, O., Ayral, P., Salze, D., and Sauvagnargues, S.: A coupling of hydrologic and hydraulic models appropriate for the fast floods of the Gardon River basin (France), Natural Hazards and Earth System Sciences, 14, 2899–2920, https://doi.org/10.5194/nhess-14-2899-2014, 2014.

Lang, M., Arnaud, P., Carreau, J., Deaux, N., Dezileau, L., Garavaglia, F., Latapie, A., Neppel, L., Paquet, E., Renard, B., Soubeyroux, J., Terrier, B., Veysseire, J., Aubert, Y., Auffray, A., Borchi, F., Bernardara, P., Carré, J., Chambon, D., Cipriani, T., Delgado, J., Doumenc, H., Fantin, R., Jourdain, S., Kochanek, K., Paquier, A., Sauquet, E., and Tramblay, Y.: ExtraFlo project (ANR 2009-2013) on extreme rainfall and flood assessment: main outcomes, La Houille Blanche - Revue internationale de l'eau, 2, 5–13, https://doi.org/10.1051/lhb/2014010, 2014.

Le Bourgeois, O., Bouvier, C., Brunet, P., and Ayral, P.-A.: Inverse modeling of soil water content to estimate the hydraulic properties of a shallow soil and the associated weathered bedrock, Journal of Hydrology, https://doi.org/10.1016/j.jhydrol.2016.01.067, 2016.

Le Coz, J., Renard, B., Bonnifait, L., Branger, F., and Boursicaud, R. L.: Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: A Bayesian approach, Journal of Hydrology, 509, 573 – 587, https://doi.org/10.1016/j.jhydrol.2013.11.016, 2014.

Le Lay, M. and Saulnier, G.-M.: Exploring the signature of climate and landscape spatial variabilities in flash flood events: Case of the 8–9 September 2002 Cévennes-Vivarais catastrophic event, Geophysical research letters, 34, https://doi.org/10.1029/2007GL029746, 2007.

Ledoux, E., Girard, G., De Marsily, G., Villeneuve, J., and Deschenes, J.: Spatially distributed modeling: conceptual approach, coupling surface water and groundwater, in: Unsaturated Flow in Hydrologic Modeling, pp. 435–454, Springer, 1989.

Ley, R., Hellebrand, H., Casper, M. C., and Fenicia, F.: Is Catchment Classification Possible by Means of Multiple Model Structures? A Case Study Based on 99 Catchments in Germany, Hydrology, 3, 22, https://doi.org/10.3390/hydrology3020022, 2016.

Llasat, M. C., Marcos, R., Llasat-Botija, M., Gilabert, J., Turco, M., and Quintana-Seguí, P.: Flash flood evolution in North-Western Mediterranean, Atmospheric Research, 149, 230 – 243, https://doi.org/10.1016/j.atmosres.2014.05.024, 2014.

Lobligeois, F.: Mieux connaître la distribution spatiale des pluies améliore-t-il la modélisation des crues? Diagnostic sur 181 bassins versants français (Is the spatial distribution of the rainfall observation improving flood modelling? Diagnostic over 181 French catchments), Ph.D. thesis, AgroParisTech, Paris, https://pastel.archives-ouvertes.fr/tel-01134990, 2014.

Mahfouf, J., Manzi, A., Noilhan, J., Giordani, H., and Déqué, M.: The land surface scheme ISBA within the Météo-France climate model ARPEGE. Part I. Implementation and preliminary results, Journal of Climate, 8, 2039–2057, https://doi.org/10.1175/1520-0442(1995)008<2039:TLSSIW>2.0.CO;2, 1995.

Marchandise, A. and Viel, C.: Utilisation des indices d'humidité de la chaîne Safran-Isba-Modcou de Météo-France pour la vigilance et la prévision opérationnelle des crues (Using the Mété-France Safran-Isba-Modcou data for operational flood warning and forecasting), La Houille Blanche, pp. 35–41, https://doi.org/10.1051/lhb/2009075, 2009.

Maréchal, D., Ayral, P.-A., Bailly, J.-S., Puech, C., and Sauvagnargues-Lesage, S.: Sur l'origine morphologique des écoulements par l'analyse d'observations hydrologiques distribuées. Application à deux bassins versants cévenols (Gard, France). Relation between the flow processes and the catchment morphology through the analysis of the distributed hydrological observations : case of two Cevenol catchments (France), Géomorphologie: relief, processus, environnement, 19, 47–62, https://doi.org/10.4000/geomorphologie.10120, 2013.

Maréchal, J.-C., Ladouche, B., and Dörfliger, N.: Analyse hydrogéologique de la contribution de l'eau souterraine à la crue éclair des 6 et 8 septembre 2005 à Nîmes (Hydrogeological analysis of groundwater contribution to the 6-8 September 2005 flash flood in Nîmes), La Houille Blanche, pp. 88–93, 2009.

Martin, C., Didon-Lescot, J.-F., Jolivet, J., Allignol, F., Castex, J.-M., Fiandino, M., Ray, D., and Usselmann, P.: Observations complémentaires sur les fonctionnements hydrologiques et hydrosédimentaires de la Vallée Obscure (commune de Peyrolles), (Observation of the flow processes in a sedimentary catchment (Obscure valley, Peyrolles, France)), Tech. rep., Technical report for the Syndicat Mixte pour l'Aménagement et la Gestion Equilibrée (SMAGE) des Gardons, 2005.

Mateo Lázaro, J., Sánchez Navarro, J. Á., García Gil, A., and Edo Romero, V.: Sensitivity analysis of main variables present in flash flood processes. Application in two Spanish catchments: Arás and Aguilón, Environmental Earth Sciences, 71, 2925–2939, https://doi.org/10.1007/s12665-013-2668-5, 2014.

Maubourguet, M.-M., Chorda, J., Dartus, D., and George, J.: Prévision des crues éclair sur le Gardon d'Anduze (Flash flood forecasting in the Gardon catchment at Anduze), in: 1st Mediterranean-HyMeX Workshop - Hydrological cycle in Mediterranean Experiment, 9-11 January 2007, Météo-France, Toulouse, France, 2007.

McDonnell, J., Sivapalan, M., Vaché, K., Dunn, S., Grant, G., Haggerty, R., Hinz, C., Hooper, R., Kirchner, J., Roderick, M., et al.: Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology, Water Resources Research, 43, https://doi.org/10.1029/2006WR005467, 2007.

McDonnell, J. J. and Beven, K.: Debates—The future of hydrological sciences: A (common) path forward? A call to action aimed at understanding velocities, celerities and residence time distributions of the headwater hydrograph, Water Resources Research, 50, 5342–5350, https://doi.org/10.1002/2013WR015141, 2014.

MNT BDALTI: MNT BDALTI, Institut Institut national de l'information géographique et forestière (IGN-F), résolution 250 m, 2016.

Payrastre, O., NAULIN, J. P., Nguyen, C. C., and Gaume, E.: Analyse hydrologique des crues de juin 2010 dans le Var (Hydrological analysis of the flash flood of the Var region, June 2010), Research report, IFSTTAR - Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux, https://hal.archives-ouvertes.fr/hal-01272025, 2012.

Perrin, C., Van Esse, W., Booij, M., Augustijn, D., Lobligeois, F., and Fenicia, F.: The influence of conceptual model structure on model performance: a comparative study for 237 French catchments, Hydrology and Earth System Sciences Discussions, 10 (4), 2013, 2013.

Pfister, L., Martínez-Carreras, N., Hissler, C., Klaus, J., Carrer, G. E., Stewart, M. K., and McDonnell, J. J.: Bedrock geology controls on catchment storage, mixing, and release: A comparative analysis of 16 nested catchments, Hydrological Processes, 31, 1828–1845, https://doi.org/10.1002/hyp.11134, hYP-16-0708.R1, 2017a.

Pfister, L., Wetzel, C. E., Klaus, J., Martínez-Carreras, N., Antonelli, M., Teuling, A. J., and McDonnell, J. J.: Terrestrial diatoms as tracers in catchment hydrology: a review, Wiley Interdisciplinary Reviews: Water, 4, https://doi.org/doi.org/10.1002/wat2.1241, 2017b.

Raynaud, D., Thielen, J., Salamon, P., Burek, P., Anquetin, S., and Alfieri, L.: A dynamic runoff co-efficient to improve flash flood early warning in Europe: evaluation on the 2013 central European floods in Germany, Meteorological Applications, 22, 410–418, https://doi.org/10.1002/met.1469, 2015.

Ribolzi, O., Moussa, R., Gaudu, J.-C., Vallès, V., and Voltz, M.: Etude des crues de transition entre période sèche et période humide, par traçage naturel sur un bassin versant méditerranéen cultivé, Comptes Rendus de l'Académie des Sciences - Series IIA - Earth and Planetary Science, 324, 985 – 992, https://doi.org/https://doi.org/10.1016/S1251-8050(97)83983-3, http://www.sciencedirect.com/science/article/pii/S1251805097839833, 1997.

Robbez-Masson, J., Mazzoni, C., Vallejo, S., Ernesty, D., and Falipou, P.: A regional multimedia soil and landscape database on the web in order to help use and teaching of soil data, in: 17th World congress of Soil Science, Bangkok, Thailand, pp. 14–21, 2002.

Roux, H., Labat, D., Garambois, P.-A., Maubourguet, M.-M., Chorda, J., and Dartus, D.: A physically-based parsimonious hydrological model for flash floods in Mediterranean catchments, Natural Hazards and Earth System Science, 11, 2567–2582, https://doi.org/10.5194/nhess-11-2567-2011, 2011.

Ruin, I., Creutin, J.-D., Anquetin, S., and Lutoff, C.: Human exposure to flash floods - Relation between flood parameters and human vulnerability during a storm of September 2002 in Southern France, Journal of Hydrology, 361, 199–213, https://doi.org/10.1016/j.jhydrol.2008.07.044, tY - JOUR, 2008.

Sangati, M., Borga, M., Rabuffetti, D., and Bechini, R.: Influence of rainfall and soil properties spatial aggregation on extreme flash flood response modelling: an evaluation based on the Sesia river basin, North Western Italy, Advances in Water Resources, 32, 1090–1106, 2009.

Sayama, T., McDonnell, J. J., Dhakal, A., and Sullivan, K.: How much water can a watershed store?, Hydrological Processes, 25, 3899–3908, https://doi.org/10.1002/hyp.8288, 2011.

Sivapalan, M.: Process complexity at hillslope scale, process simplicity at the watershed scale: is there a connection?, Hydrological Processes, 17, 1037–1041, https://doi.org/10.1002/hyp.5109, 2003.

Tabary, P.: The new French operational radar rainfall product. Part I: Methodology., Weather & Forecasting, 22, https://doi.org/doi: 10.1175/WAF1004.1, 2007.

Troch, P. A., Berne, A., Bogaart, P., Harman, C., Hilberts, A. G. J., Lyon, S. W., Paniconi, C., Pauwels, V. R. N., Rupp, D. E., Selker, J. S., Teuling, A. J., Uijlenhoet, R., and Verhoest, N. E. C.: The importance of hydraulic groundwater theory in catchment hydrology: The legacy of Wilfried Brutsaert and Jean-Yves Parlange, Water Resources Research, 49, 5099–5116, https://doi.org/10.1002/wrcr.20407, 2013.

UNISDR: UNISDR Terminology on Disaster Risk Reduction, http://www.unisdr.org/eng/terminology/terminology-2009-eng.html(2009), 2009.

Van Genuchten, M. T.: A closed-form equation for predicting the hydraulic conductivity of unsaturated soils, Soil science society of America journal, 44, 892–898, 1980.

Vannier, O., Braud, I., and Anquetin, S.: Regional estimation of catchment-scale soil properties by means of streamflow recession analysis for use in distributed hydrological models, Hydrological Processes, https://doi.org/10.1002/hyp.10101, 2013.

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

Viglione, A., Chirico, G. B., Woods, R., and Blöschl, G.: Generalised synthesis of space–time variability in flood response: An analytical framework, Journal of Hydrology, 394, 198–212, https://doi.org/10.1016/j.jhydrol.2010.05.047, 2010.

Zoccatelli, D., Borga, M., Zanon, F., Antonescu, B., and Stancalie, G.: Which rainfall spatial information for flash flood response modelling? A numerical investigation based on data from the Carpathian range, Romania, Journal of Hydrology, 394, 148 – 161, https://doi.org/10.1016/j.jhydrol.2010.07.019, 2010.

5