Interactive comment on “Using a multi-hypothesis framework to improve the understanding of flow dynamics during flash floods” by Audrey Douinot et al.

Audrey Douinot et al.
helene.roux@imft.fr

Received and published: 6 April 2018

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, downloaded here as a supplement, integrates as well the modification related to the comments of the other referee.

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
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. 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 performances of the submitted models, revealing different hydrological behaviours along the catchment set, and consequently, giving advances in characterising the flash flood processing over the Mediterranean area. Those results are supported by their consistency with the rare available in-situ measurements and the prior knowledge of several catchments. The characterisation is of course carried out within existing equifinality issues. The descriptive potential of the distributed model was then used to spot counterbalancing effects between internal flow processes and to finally propose new insights 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$^2$). In Europe, particularly intense flash floods are observed predominantly on the north west of the Mediterranean Arc, 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 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). 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 forecasted 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 methodology, 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: 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 Mediterrranean area. This focus, on this particular content has also been specified (page 3, lines 1-12).

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-Western Mediterranean area) using an alike promising approach described in the past works. I added the reference you suggested, as clearly missing (page 4, line 24). 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 knowledge on the hydrological processes of the studied area obtained through field experimental studies. You will find the reworded subsection in page 5-6, lines 29-18, 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 split 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 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 literature (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, infiltration tests and analysis of electrical resistivity signals in granite and schist transects located on hillslope 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 continuous 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 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[m³.km⁻².s⁻¹]); 2 year return period of maximum daily discharge (QD2[m³.km⁻².s⁻¹]); 10 year return period of maximum hourly discharge (QH10[m³.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 restructured, 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 broadcast 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 11).

The following sentence has also been added page 10, lines 9-11: 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 10, line 16).


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. The reworded sentence is: The French flood forecasting service (SCHAPI: Service central d'hydrométéorologie et d'appui 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 11, lines 2-6:

This was done using spatial model outputs from Météo-France’s SIM operational chain (Habets et al., 2008), including a meteorological 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 10, lines 20-26):

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.

C12
A different mean of evaluation is used for the hydrograph recession as the assessment based on the simulated discharge values - Qmed_INT - and restricted to the recession interval is actually representative whether the high discharge values occurring 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 assessment with the high discharge values, but also as it is very well known in the literature 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, on the top of the page 17:

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, $Q(t) = f\left(\log(-\frac{dQ(t)}{dt})\right)$, 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.

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

(1)

where $\frac{dy_i}{dt}$ and $\frac{d\hat{y}_i}{dt}$ 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 order 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 been reworded. In addition title were added to the paragraphs in order to clarify the objective of each one.

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 three main path: on the surface, through the top or the deep layer of the soil. While those assessments 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 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 l’Ecologie” is strange looking.

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


Braud, I., Ayral, P., Bouvier, C., Branger, F., Delrieu, G., Dramais, G., Le Coz, J., Leblois,


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 occurred between 1958-1994 in the Languedoc-Roussillon, the Provence-Alpes Cotes d’azur and the Corse regions), Phénomènes remarquables, 1994.


Please also note the supplement to this comment: https://www.hydrol-earth-syst-sci-discuss.net/hess-2017-710/hess-2017-710-AC1-supplement.pdf