ANSWERS TO THE REFEREES
FOR THE ARTICLE ENTITLED "ASSESSING THE IMPACT OF RESOLUTION AND SOIL DATASETS ON FLASH-FLOOD MODELLING." SUBMITTED TO HYDROLOGY AND EARTH SYSTEM SCIENCES

POINT BY POINT ANSWERS TO MASSIMILIANO ZAPPA (REFEREE 1)

Dear authors,

I found your manuscript as a well organized study with clear goals and satisfying answer to the posed questions.
You acknowledge by yourself, that the number of basins and event is limited, which is a common drawback in flash-flood research.

I have only minor issues included in the attachment to this document.

Best regards

Massimiliano Zappa

Please also note the supplement to this comment:
https://www.hydrol-earth-syst-sci-discuss.net/hess-2018-374/hess-2018-374-RC1-supplement.pdf

⇨ The authors would like to thank Massimiliano Zappa for the positive review and his helpful comments. In the following, we reply to each comment and indicate how the suggestions have been taken into account in the new version of the manuscript.

Page 1, line 3
put the number
⇨ Changes in manuscript: The sentence has been changed into:
"The ISBA-TOP coupled system, which is dedicated to Mediterranean flash-flood simulations, is used with two grid-cell sizes (300 m and 1000 m), *two soil texture datasets and two land use databases*, to model 12 past flash-flood events in southeastern France."

Page 1, line 22
This is the core of:
Zappa, M., Jaun, S., Germann, U., Walser, A., & Fundel, F. (2011). Superposition of three sources of uncertainties in operational flood forecasting chains. Atmospheric Research, 100(2-3), 246-262. https://doi.org/10.1016/j.atmosres.2010.12.005
⇨ Changes in manuscript: Reference has been added:
"Zappa et al. (2011) have investigated the propagation and the superposition of these three sources of uncertainty in a hydrometeorological forecasting system for a catchment of the Swiss Alps."

Page 2, line 5
See reference above. Furthermore, you might write some lines on identification of model structures to account for landscape organization.

E.g.
Gharari, S., Hrachowitz, M., Fenicia, F., Gao, H., and Savenije, H. H. G.: Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration, Hydrol. Earth Syst. Sci., 18, 4839-4859, https://doi.org/10.5194/hess-18-4839-2014, 2014.

⇨ Changes in manuscript: Reference has been added and developed in the introduction as follows:

"Several studies have identified the most appropriate model structure in hydrological modelling while taking into account several landscape complexity levels (e.g. Flügel, 1995 ; Savenije, 2010 ; Gharari et al., 2014a, 2014b). Gharari et al. (2014b) have used models of increasing complexity (the first represents the catchment in a lumped way, the second distinguishes wetlands from the remainder, i.e., hillslopes and plateaus and the third gives a complete representation of the wetlands, hillslopes and plateaus). They showed that by allowing for more landscape-related process heterogeneity in a model (third model), the predictive power increases even without traditional calibration."

Flügel, W. A. (1995). Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the River Bröl, Germany. *Hydrological Processes*, *9*(3-4), 423-436.

Savenije, H. H. G. (2010). HESS Opinions" Topography driven conceptual modelling (FLEX-Topo)". *Hydrology and Earth System Sciences*, *14*(12), 2681-2692.

Gharari, S., M. Shafiei, M. Hrachowitz, R. Kumar, F. Fenicia, H. V. Gupta, and H. H. G. Savenije (2014a), A constraint-based search algorithm for parameter identification of environmental models, Hydrol. Earth Syst. Sci. ,18 (12), 4861–4870.

Gharari, S., Hrachowitz, M., Fenicia, F., Gao, H., and Savenije, H. H. G.: Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration, Hydrol. Earth Syst. Sci., 18, 4839-4859, https://doi.org/10.5194/hess-18-4839-2014, 2014b.

Page 2, line 15
We explored recently the influence of mapping dominant runoff processes.
E.g.
Antonetti, M., Buss, R., Scherrer, S., Margreth, M., & Zappa, M. (2016). Mapping dominant runoff processes: an evaluation of different approaches using similarity measures and synthetic runoff simulations. Hydrology and Earth System Sciences, 20(7), 2929-2945. https://doi.org/10.5194/hess-20-2929-2016

⇨ Changes in manuscript: Reference has been added and commented:

"Antonetti et al. (2016) explored recently the uncertainty of hydrological simulations due to different spatial representations of dominant runoff processes. They found that the simulations with the most complex automatic mapping approach are the closer to the reference map, while those without soil information differed considerably."

Page 2, line 27
Two

⇨ Changes in manuscript: Done.

Page 2, line 27
Why these two. I will surely read about it later on.
⇨ This has been discussed later in the manuscript (page 6).

Page 3, line 7
Think you can put the real number.
⇨ Changes in manuscript: Done.

Page 4, line 27
Do you mean "operational" or ex-post experiments in real-time mode?
Is the model implemented somewhere now in real time (Bulgaria, isn't?)
⇨ We mean in real time during the first Special Observation Period of Hymex, from 05 September to 06 November 2012. The model is also implemented now in a real time flood warning system in Arda river basin. Arda River is a cross border river, that springs in Bulgaria and continues into Greece.
Changes in manuscript: The initial paragraph has been modified into:
"As part of the international HyMeX program, the ISBA-TOP coupled system has been used for real-time prediction of discharge for four catchments in the Cévennes-Vivarais region and the French Riviera, during the first Special Observation Period of Hymex, from 05 September to 06 November 2012. Case studies have also been performed with ISBA-TOP for Italian (Nuissier et al., 2016) watersheds. ISBA-TOP is also currently used in real time by the National Institute of Meteorology and Hydrology (NIMH) of Bulgaria for operational flood forecasting for the Arda River Basin (Artinyan et al., 2016)."

Artinyan, E., Vincendon, B., Kroumova, K., Nedkov, N., Tsarev, P., Balabanova, S., & Koshinchanov, G. (2016). Flood forecasting and alert system for Arda River basin. *Journal of Hydrology*, *541*, 457-470.

Page 4, line 29
Reference is incomplete
⇨ Changes in manuscript: Reference has been removed.

Page 5, line 9
With respect to which soil indicator?
⇨ According to Tubiello et al. (2016), "the accuracy of the HWSD was never estimated in the literature. [Tubiello et al.] used in [their] paper an accuracy of 75% for the soil information, based on results from other soil map products (Fisher, 1993) and expert opinion of specialized FAO staff."
No change in the manuscript.

Tubiello, F. N., Biancalani, R., Salvatore, M., Rossi, S., & Conchedda, G. (2016). A worldwide assessment of greenhouse gas emissions from drained organic soils. *Sustainability*, *8*(4), 371.

Fisher, P. F. (1993). Visualizing uncertainty in soil maps by animation. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *30*(2-3), 20-27.

Page 5, line 17
It reads quite awkward. Can you re-formulate the sentence?
⇨ Changes in manuscript: This sentence has been changed into:
“The spatial distribution of the soil texture is not so highly contrasted for LUCAS. In this second dataset, there is less clay over the Vidourle and Hérault catchments.”

Page 6, line 17
Please include here the motivation of selecting these two resolution. Is there any way to have for some analyses presented an evaluation of more resolutions? (Just a question, no mandatory request).
⇨ No other resolutions have been evaluated. These resolutions (300m and 1km) were selected mainly because of:
- the spatial resolution of the meteorological forcing data used in this study is 1km.
- ECOCLIMAP Second Generation is produced at a 300m resolution.

Changes in manuscript: These reasons have been added in the manuscript:
“These both resolutions were selected, because the spatial resolution of meteorological forcing data used in this study is 1km and because the new land ecosystem database ECOCLIMAP Second Generation is produced at a 300m resolution.”

Page 6, line 24
As I understand the 1000 m resolution is used in only one configuration of the experiments, on different events and basins. I am not sure if this is enough to conclude on its difference to 300 m.
Please discuss.
⇨ Until now, ISBA-TOP was always used with a 1000 m resolution (e.g. Artinyan et al., 2016 ; Vincendon et al., 2016 ; Edouard et al., 2018) and thus it constitutes our reference in this study which aim is to investigate the impact of a higher resolution together with various soil descriptors. This study is certainly not exhaustive. Indeed, we could have tested the other configurations with a 1000 m resolution ($R_1T_2C_1$, $R_1T_2C_2$, $R_1T_2C_3$). Moreover it could be also interesting to compare the model performance with these two resolutions and different temporal resolutions (15 or 30 minutes instead of 1 hour for example), even if it is not the scope of the article. The purpose of this article is however not to focus only on the differences generated by the use of these two resolutions but rather to investigate and rank the impacts of the spatial resolution and the terrain descriptors on flash-flood modelling.
No change in manuscript.

Artinyan, E., Vincendon, B., Kroumova, K., Nedkov, N., Tsarev, P., Balabanova, S., & Koshinchanov, G. (2016). Flood forecasting and alert system for Arda River basin. *Journal of Hydrology*, *541*, 457-470.

Vincendon, B., Édouard, S., Dewaele, H., Ducrocq, V., Lespinas, F., Delrieu, G., & Anquetin, S. (2016). Modeling flash floods in southern France for road management purposes. *Journal of Hydrology*, *541*, 190-205.

Edouard, S., Vincendon, B., & Ducrocq, V. (2018). Ensemble-based flash-flood modelling: Taking into account hydrodynamic parameters and initial soil moisture uncertainties. *Journal of Hydrology*, *560*, 480-494.

Page 6, line 31
soil temperature

⇨ Changes in manuscript: This has been added.

Page 6, line 33
Please expand how you make this or reference to previous work where this is described and evaluated.
⇨ Changes in manuscript: Explanation has been reported:
"The data were downscaled over the 1-km ISBA domain, using the nearest-grid-point interpolation method as in Edouard et al. (2018)."

Edouard, S., Vincendon, B., & Ducrocq, V. (2018). Ensemble-based flash-flood modelling: Taking into account hydrodynamic parameters and initial soil moisture uncertainties. *Journal of Hydrology, 560*, 480-494.

Page 7, line 8
(Figure 3) Symbols on the lines are too small.
⇨ We agree. Changes in manuscript: Figure 3 has been modified.

Page 7, line 10
In zone A you have 11 events and several outlets with discharge observation. The resulting Figure 3 is quite strongly aggregated to me. Can you design a new figure showing one more dimension a least (e.g. variability between events or variability among stations)?
⇨ The aim of Figure 3 is to provide a synthetic view of the results for the Nash scores which is really not easy to obtain from a visualization for each individual watershed and event (in addition to add several figures to this manuscript). As mentioned in the manuscript, the closer the points are to the bottom-right corner of the figure 3, the better the skill. From Figure 3, we are able to draw information on the general behaviour of the score and not a specific behaviour per watershed. The variability of the skill according to the watersheds can be deduced from Figure 4. Indeed, LNP is a linear combination of the Nash criterion and the error of the peak time and discharge. No change in manuscript.

Page 7, line 23
Can you find in literature other studies supporting this?
⇨ According to Dutta & Nakayama (2009), the decrease of grid resolution leads to a decrease of the average slope which leads to a reduction of flow velocity. Moreover, Vázquez et al. (2002) found that the increase of grid resolution, which increases the number of river branches within the river network, leads to faster overland and channel response of the catchment.

Changes in manuscript: References have been added in the revised manuscript:
"The increase in the grid resolution appears to significantly improve the simulated peak time (see Figure 4c). This might be due to the more detailed description of the river network and of the average slope over the watershed which influence the flow velocity (Dutta & Nakayama, 2009 ; Vázquez et al., 2002). "

Dutta, D., & Nakayama, K. (2009). Effects of spatial grid resolution on river flow and surface inundation simulation by physically based distributed modelling approach. *Hydrological Processes: An International Journal, 23*(4), 534-545.

Vázquez, R. F., Feyen, L., Feyen, J., & Refsgaard, J. C. (2002). Effect of grid size on effective parameters and model performance of the MIKE-SHE code. *Hydrological processes*, *16*(2), 355-372.

Page 7, line 25

The sentence is supported by the data you show as one can evaluate with visual inspection. Nevertheless, I am not sure that "significant" in statistical terms is adequate here.

⇨ We agree. Changes in manuscript: the sentence has been modified.

Page 7, line 28

Enumerate, please

⇨ Changes in manuscript: Done.

Page 8, line 9

Nice analysis!

Page 8, line 19

Can you spend some lines on how interception storage is accounted for? Is this the reason for differences in surface runoff amounts depending on the land cover product used?

⇨ Done. Changes in manuscript: A sentence has been added in paragraph 2.3.2:

"In ISBA-TOP, the land cover product influences both interception storage (through the leaf area index, vegetation height and roughness length) and infiltration capacity (through the root depth), with resulting impacts on the simulated surface runoff amounts."

Page 9, line 18

Have you tried to re-aggregate the 300 m grid on the 1 km grid and make a difference plot?

⇨ Cumulated runoff of each experiment was re-aggregated on the 1 km grid and difference plots between the first experiment ($R_1T_1C_1$) and every others are displayed in the figure below (Figure R1), to assess the contribution of each experiment with respect to the first one.

The plots illustrate that in the main coastal area, from E18 to the east of O15, where most of the impacts are located, the runoff is mostly less intense with the first experiment (the pixels are more greenish). In the southeast of O15 (zone with many observed impacts) $R_2T_2C_2$ and $R_2T_2C_3$ produce less runoff than $R_1T_1C_1$. However $R_2T_2C_3$ produces runoff over a larger area. Near to E18 (other zone with many observed impacts), the differences between simulated amounts of runoff are more difficult to assess, because the location corresponds to the maximum of simulated runoff for $R_1T_1C_1$.

No change in manuscript: The new figure has not been added in the manuscript because it is partly redundant with Figure 9.

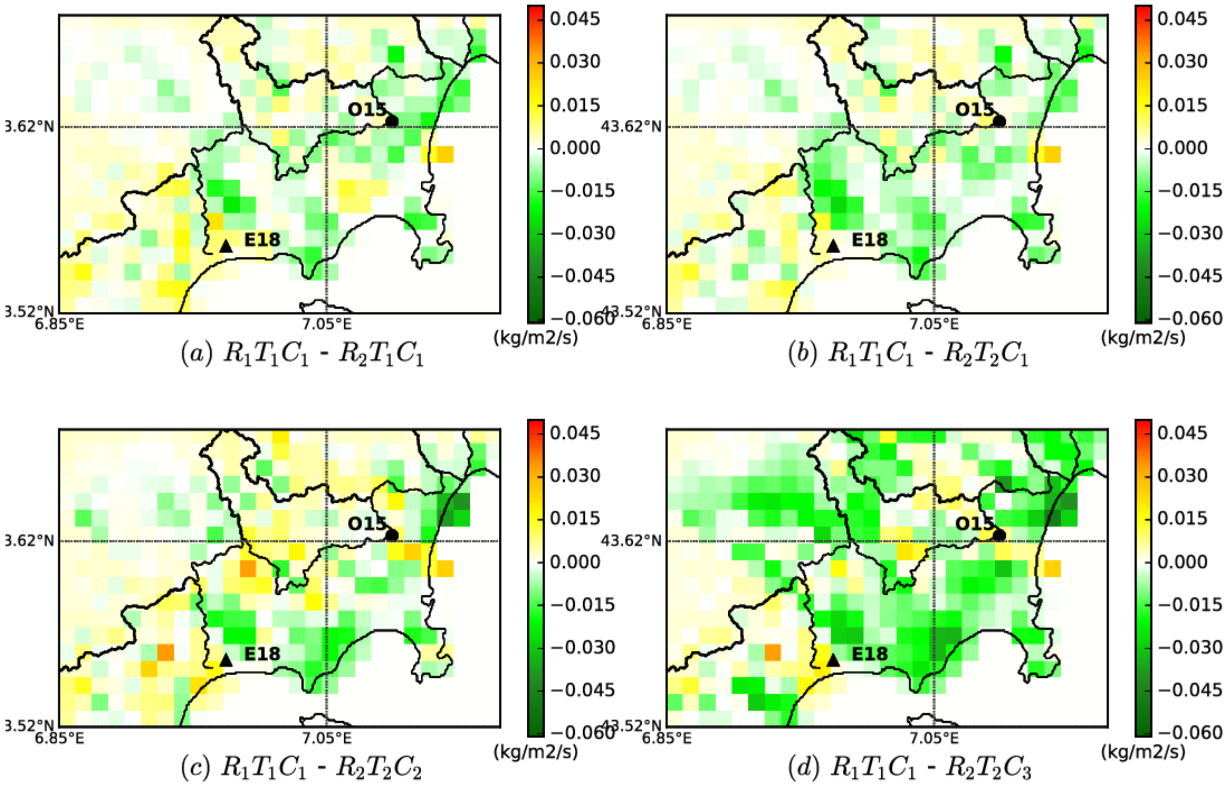

Figure R1. Difference plots of cumulated runoff between the first experiment and the others for the October 2015 event. Note that the range of the colour scale is the same for the various panels; in red/yellow, the differences are always positive and, in green, they are always negative.

Page 9, line 21
Well, show me numbers and then you can use it. I will try to avoiding using "significant" or "undeniable"and other synonyms when describing findings from visual inspection. If you would try here to make a quantitative estimate of the agreement between "impacts" and your maps in Figure 9, then the whole paper would be more attractive.

⇨ We agree.

To make a quantitative estimate of the agreement between "impacts" and the runoff maps in Figure 9, a sliding window approach has been used: A circular region, centered on each 300 m grid point, slides across all the domain of Figure 9. For each sliding window, the average runoff of the grid points contained in the circular region is reported as well as the impacts of each category (victims, damage, high water marks) are counted. We expect the better experiment to give intense runoff in the neighbourhood of impacts and less runoff if there is no impact. Several values of radius (0.75, 1, 1.5 or 2 km) have been used. A radius of 1 km seems the most appropriate as it allows to compare all the results at the coarser resolution of $R_1T_2C_1$ (i.e. 1 km) without having too much sea points in the circular regions with impacts along the coast.
Figure R2 shows the average runoff in function of the damage number. Clearly, the runoff is larger when impacts are recorded in the 1km neighbourhood. The average runoff increases with the impact number up to 10 damages in the 1km neighbourhood. $R_2T_2C_3$ (in yellow) produces on average more runoff than the other experiments. The largest differences between the experiments are for the 16-25 damages per circular area. In this range, the 1 km resolution simulation ($R_1T_1C_1$ in black) provides the lowest average value. The 16-25 damages per circular region category is mainly

recorded next to O15 where $R_1T_1C_1$ produces less runoff (pink color pixels in Figure 9a of the paper).

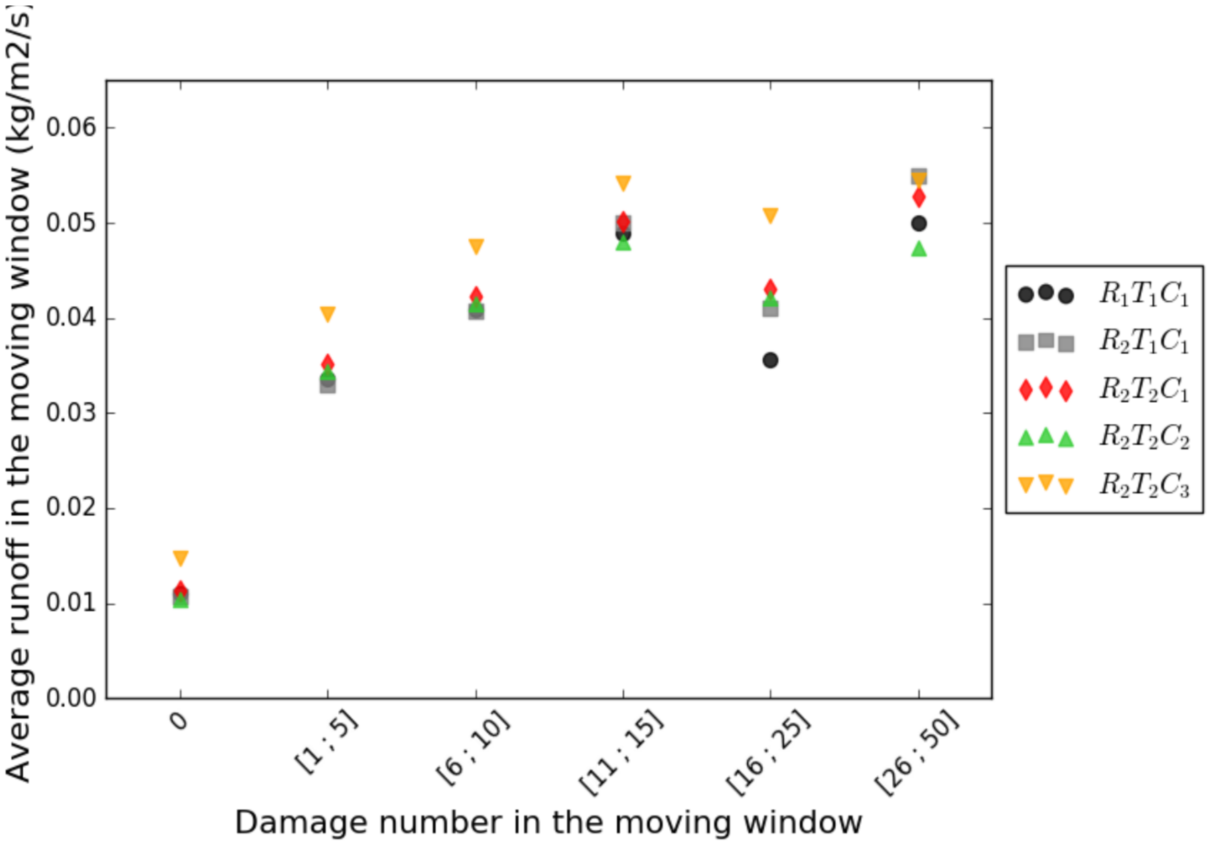

Figure R2: Average runoff in function of the number of damages encountered in the 1-km circular neighbourhood over all the domain of Figure 9.

Changes in manuscript: The figure R2 has been added and the last paragraph of 3.1.3 has been re-written as follows:

" The cumulated runoff for each experiment is displayed in Figure 9. The spatial patterns of the surface runoff simulated by the different experiments are consistent with the surface accumulated rainfall (Figure 6 compared to Figure 9). Differences between the experiments appear primarily east and north of O15, as well as at the limit between the two coastal zones, west of E18. The areas of simulated high runoff match the observed impacts zones (Figure 10), which are located near the coast and close to O15. The matching of the impacts and high runoff zones is assessed using a neighbouring approach, for which a circular region, centered on each 300 m grid point, slides across all the domain of Figure 9 counting the impacts of each category (victims, damage, high water marks) inside the circular region and the average runoff over the circular region. The radius of the circular area is set to 1 km, allowing to compare all the results at the coarser resolution of $R_1T_1C_1$ (*i.e.* 1 km) without having too much sea grid-points in the circular regions with impacts along the coast. Figure "R2" shows the average runoff in function of the damage number. Clearly, the runoff is larger when impacts are recorded in the 1km neighbourhood, in agreement with the visual comparison between Figure 9 and Figure 10. The average runoff increases with the impact number up to 10 damages in the 1km neighbourhood. $R_2T_2C_3$ (in yellow) produces on average more runoff than the other experiments. Figure 9 shows that significant runoff is produced over a larger area for R2 T2 C3. In particular, over the urbanized areas south of the upward catchments, R2 T2 C3 produces more runoff than the other experiments. The largest differences between the experiments are for the 16-25 damages per circular area (Figure “R2”). In this range, the 1 km resolution simulation ($R_1T_1C_1$ in black) provides the lowest average value. The 16-25 damages per circular region category is mainly recorded next to O15 where $R_1T_1C_1$ produces less runoff (pink color pixels in Figure 9a).”

Page 11, line 8
See previous comments

⇨ Changes in manuscript: This has been modified.

Page 11, line 15
Which is a common drawback in flash-flood research.

ANSWERS TO THE REFEREES
FOR THE ARTICLE ENTITLED "ASSESSING THE IMPACT OF RESOLUTION AND SOIL DATASETS ON FLASH-FLOOD MODELLING." SUBMITTED TO HYDROLOGY AND EARTH SYSTEM SCIENCES

POINT BY POINT ANSWERS TO RENATA ROMANOWICZ (REFEREE 2)

⇨ The authors would like to thank Renata Romanowicz for her constructive comments. In the following, we will answer each comment and indicate how the suggestions have been taken into account in the new version of the manuscript.

The authors present a study of the performance of the flash-flood modelling tool, the ISBA-TOPP coupled system under varying grid resolutions and terrain descriptors in order to assess their influence. Two resolution grids were used, 300 m and 1000 m, and it was found that the higher resolution gave better results in reproducing the flood peak. It is not surprising and could be stated without any experiments.

⇨ Change in manuscript: the following comment has been added in the introduction:
"Even if in general higher resolution leads to more accurate simulations, there can be a critical level beyond which the model response is not necessarily improved (Egüen et al., 2012 ; Hengl, 2006)."

Egüen, M., Aguilar, C., Herrero, J., Millares, A., & Polo, M. J. (2012). On the influence of cell size in physically-based distributed hydrological modelling to assess extreme values in water resource planning. *Natural Hazards and Earth System Sciences*, *12*(5), 1573-1582.

Hengl, T. (2006). Finding the right pixel size. *Computers & geosciences*, *32*(9), 1283-1298.

However, it would be interesting to know how fine the grid should be to still give acceptable results and feasible computation costs. In fact, there might be an interesting relationship between the grid size and model performance, obviously depending on the scale of the catchment. I am aware that this could be a separate paper, but I expect the authors to state clearly that the influence of the grid size is not studied here, apart from the comparison of two different grid options.

⇨ As we said in a previous comment, the study of the influence of the grid size is certainly not exhaustive. Indeed, the relationship between the grid size and the model performance, depending on the scale of the catchment, is not the scope of the article. The purpose of this article is not to focus only on the differences generated by the use of different resolutions but rather to investigate and rank the impacts of two different spatial resolutions and different terrain descriptors on flash-flood modelling.
Changes in manuscript: The first sentence in the abstract has been modified into:

"The present study assesses the impacts of *two* grid resolutions and the descriptors of soil texture and land cover on flash-flood modelling at local and basin scales."

When it comes to the terrain descriptors, the authors do not present their comparison in a very clear way. Two different soil texture maps and two different land use maps are applied. Apart from the fact that we know those maps have different sources and give slightly different percentages of clay and sand, or land surface cover, no other analysis of the differences in map descriptors is given. As a result, a discussion of the possible reasons for the

experimental results is impossible. It would be interesting to know where the differences between the results come from. At the moment, we learn only that for the peaks the texture seems to have a larger impact than the land use and that there is no noticeable difference for peak times between the two. It shows that the comparison between different maps is very crude. There are studies showing that land use and in particular, preferential pathways, can have a large impact on the catchment residence times and the time flood wave travels [Bloschl, 2001, 2007]. The authors are advised to add a discussion on those issues. At the moment, I am not sure what is the paper's outcome.

Bloschl, G., 2001. Scaling in hydrology. Invited commentary. Hydrol. Process. 15, 709–711.
Bloschl, G., 2007, At what scales do climate variability and land cover change impact on flooding and low flows?, Hydrological Processes, 21, 1241-1247.

⇨ Changes in manuscript:

1. A sentence has been added in the introduction:
   "Land use and in particular, preferential pathways, can have a large impact on the catchment residence times and the time flood wave travels [Bloschl, 2001, 2007]"

2. After these lines, some sentences on identification of model structures to account for landscape organization have also been added [see the 3rd answer to M. Zappa].

3. A sentence has been added in 2.3.2 on how land use products modify flow pathways and storage in ISBA-TOP [see the 21st answer to M. Zappa].

4. Some sentences have been added in 2.3.1 on how soil textures modify runoff with soil hydrodynamic parameters in ISBA-TOP:
   "Soil texture has an impact on simulated runoff through soil hydrodynamic parameters, which are determined by CH78 pedotransfer functions (Clapp & Hornberger, 1978) in ISBA-TOP. Edouard et al. (2018) investigated the impact of these parameters on runoff simulations. "

5. Some sentences have been added or modified in 3.1 in order to link the results to the map descriptors:
   "The differences in the soil texture databases, which impact the water storage capacity and the ease of water to move through saturated soil, resulted in ..." in 3.1.1
   "This excess runoff is consistent with lower infiltration and drainage capacity associated with clay-rich soils." in 3.1.2

6. Some sentences have been added in the conclusions:
   "Land cover and soil texture influence locally the processes in the catchments. Their spatial variability has an impact on the preferential flow paths, the flow velocities and the water storage. The complexity of the interactions between processes at the catchment scale does not allow us to clearly conclude on how land cover and soil texture, induce differences in simulated flows."

Edouard, S., Vincendon, B., & Ducrocq, V. (2018). Ensemble-based flash-flood modelling: Taking into account hydrodynamic parameters and initial soil moisture uncertainties. *Journal of Hydrology*, *560*, 480-494.

Clapp, R. B., & Hornberger, G. M. (1978). Empirical equations for some soil hydraulic properties. *Water resources research*, *14*(4), 601-604.