NHESS manuscript : NHESS-2023-82

Does a convection-permitting regional climate model bring new perspectives on the projection of Mediterranean floods?

We would like to thank the reviewers for the time and efforts they have dedicated to review our manuscript and for their useful and relevant comments. We believe that the comments truly helped to improve the manuscript. You can find below the responses to the requests and suggestions.

Reviewer 1

This paper assesses the use of convection-permitting regional climate models (CPMs) to improve the outputs of regional climate models (RCM), as climate projections could incorporate simulation of convection system storms in CPMs. CPMs could be applied to improve climate projections in the Mediterranean area, where floods are usually generated by convective storms. Simulations generated by a CPM are compared with the climate projections supplied by a RCM. Climate projections are bias corrected. Two hydrological models are used to transform climate projections into flood time series.

The paper is well written and organised. It can be considered for publication after addressing the comments included below.

We would like to thank the reviewer #1 for his/her positive evaluation of the manuscript. We answered all the comments below.

General comments

Convection-permitting regional climate models are a potential tool to improve the characterisation of convective storms in climate projections, as such events cannot be represented by the current climate models, given their spatial and temporal resolution. However, the analysis is limited to one climate model and one climate change scenario. Therefore, the conclusions could be highly limited to such a climate model and scenario. Why only one climate model is used? In addition, the selection of such a climate model and no other should be discussed. Why only the RCP 8.5 scenario is used?

To our knowledge, this study is the first one that focuses on Mediterranean floods using a convection-permitting model to force two hydrological models. It is indeed limited to one convection-permitting climate model (CPM) and one regional climate model (RCM) given the availability of these types of simulations. However, ALADIN and AROME are well known climate models that are well-documented and used in the literature.

We are aware of the limitation of using only 1 CPM and 1 RCM. Nevertheless, the aim of the study is not to provide robust future scenarios, but rather to develop a methodology that compares two climate models of different generations to assess both the added value in historical simulations and the differences of the projected signals.

Furthermore, convection permitting models are complex to develop and their simulations require large computational costs and storage capabilities. In this present study, for the sake of simplicity, we used only one CPM developed at CNRM. We chose this CPM as it was at the time of the study the first one available over this domain and all data were easily available and documented through different studies (Fumière et al., 2020; Caillaud et al., 2021; Lucas-Picher et al., 2023). Additional simulations should be performed in 2024.

We used the RCP 8.5 scenario because it is the only future scenario available for this simulation that was initially performed for the EUCP project (Hewitt and Lowe, 2018). Finally, by selecting the worst case scenario (i.e. RCP 8.5), we probably get the strongest climate change signal possible.This likely exacerbates the discrepancies between RCM and CPM future projections.

We are however aware that more CPMs and potentially other RCPs should be included in future work. We highlighted further this limitation in the revised version of the manuscript, at the end of section 4.5.

Furthermore, the paper should clarify if the climate model used in the study belongs to either AR5 or AR6 of IPCC.

We cannot say that the climate models in general belong to any of the assessment reports. Assessment reports are based on numerous studies using climate models that are part of projects, coordinated frameworks or experiences.

In this study, the GCM used was one of many models that participated in the Coupled Model Intercomparison Project Phase 5 (CMIP5), the RCM model was used for the CORDEX experiment and the CPM model was used for the EUCP project. The CMIP5 model simulations have been indeed assessed in the AR5. We added a detail in the revised version line 115 : “as part of the EUCP project”. CORDEX experiment was already stated on line 111.

The CREST distributed hydrological model has been calibrated fixing most of its parameters. A discussion should be included about why some parameters are fixed. A sensitivity analysis of model parameters could be useful to improve the understandability of the paper.

For CREST, different tests have been done with some combinations of parameters. Calibrations are long to run (>4 days). We made several tests, calibrating several free parameters spatially, or averaging parameters over the basins.

Most CREST parameters do have a physical meaning. Among the 13 parameters, we fixed 7 of them. Some reasons explaining why we fixed them are:

- **IM** is the percentage of impervious area in %. The Gardon d’Anduze basin has a complex and pronounced topography with a mean slope of around 20 % (Roux et al., 2011). It is primarily rural and it contains little urbanized areas. Even if soils are thin, the area is not composed of natural impervious areas (Roux et al., 2011), and the presence of buildings and bare rocks is rare. Thus, we considered this percentage to be close to 0 (range between 0.00001 and 0.00002)

- **Ke** is the potential evapotranspiration adjustment factor (unitless). Here we have as input a reliable simulation of PE, and this parameter is not dominant for Mediterranean floods. This parameter is adjusted on a monthly basis within the model outputs, but is not generally calibrated.
- **IWU**: (Initial soil saturation in %) and **isu** (Initial water storage in channel grid cells) are initial conditions, we do not need to calibrate them because we included a warm-up period in our hydrological simulations.
- **under** (Subsurface flow speed) and **leaki** (Reduction in interflow storage) are baseflow parameters. Since we focus on floods that occurred rapidly, these parameters are not the most important parameters to study floods, so we fixed them.
- **th**: Drainage area threshold for channel cells (number of cells). This value can easily be estimated by looking at maps from which drainage area streams start to form and convert them to a cell number.

Here are the results of two tests with free (left side) and the above fixed (right side) parameters. The simulation with fixed parameters shows a better ability to reproduce floods over this catchment, and biases of the river discharges are also lower, explaining why we chose to fix these parameters. Furthermore, we followed closely the advice from the CREST model creators, co-authors of this study, who helped us design this experimental setup.

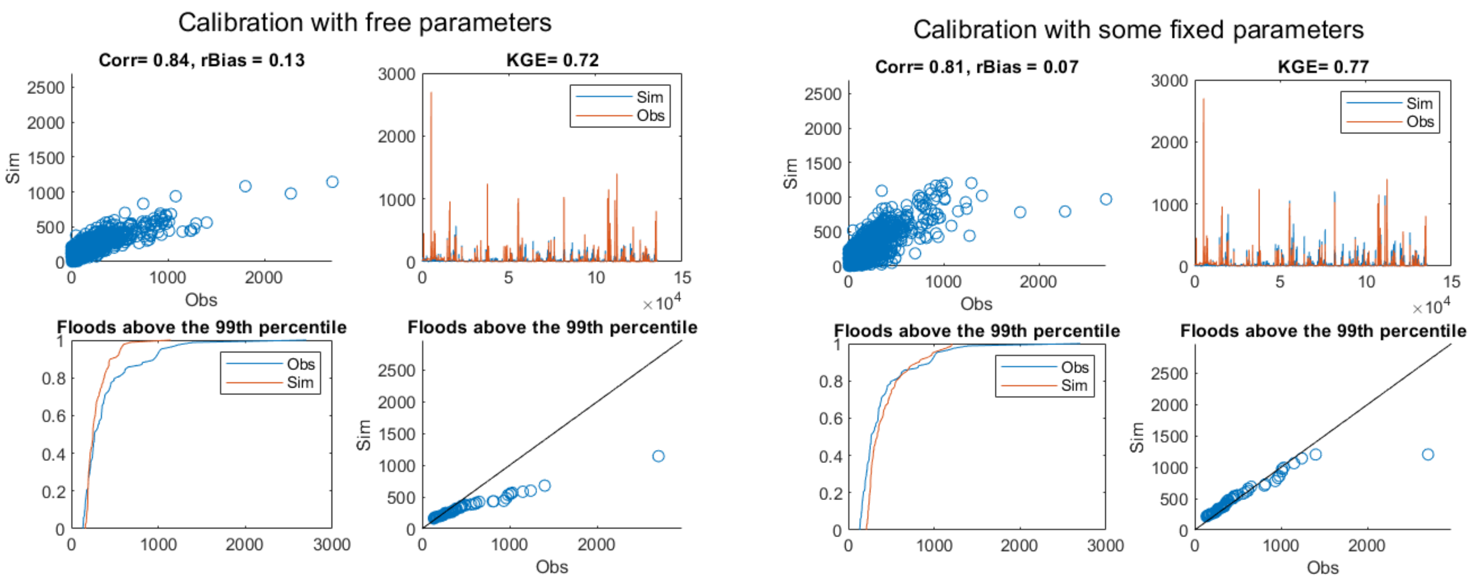

Figure 1: Results of CREST simulations (2002-2018 period) over the Gardon d'Anduze catchment with calibration of some free parameters (left panel) and with fixed parameters (right panel).

This information is too rich to be relevant in this manuscript. We chose to minimize the level of details by staying on the initial sentence : "most parameters have been fixed" (l. 179) as the calibration process is not essential to understand the methodology and interpret the results.

Section 4.1. The results are assessed qualitatively by using figures to compare outputs from different simulations. However, quantitative assessment of the results would be very useful for the reader. For example, comparison between model outputs shown in Figure 3 in terms of cumulative probability distribution could be improved by using similar metric to those used with the hydrological models.

Thank you for this interesting comment. We include below a new table that summarizes a quantitative assessment of the simulated precipitation that was added to the revised manuscript.

Model	**ALADIN**		**AROME**	
	raw	corrected	raw	corrected
Bias on median (%)	-27.2	0.4	-16.5	3.9
Bias on Q95 (%)	-40.4	1.3	-19.6	-3.4
Bias on Q99 (%)	-45.9	5.3	-12.2	0

Table 1: Evaluation of raw and corrected simulated precipitation with respect to the COMEPHORE observed dataset

Table 3 shows that the CREST performance is lower for all the metrics. It could be assumed that a distributed model could represent better rainfall-runoff processes in the catchment that a lumped model. However, if a distributed model has a lower performance than a lumped model, the lumped model should be preferred as it is simpler. Therefore, a comment should be included about why the CREST model is used in this study despite having a lower performance than a simpler lump model. In addition, maybe the lower performance of the CREST model could be increased by improving its calibration.

Distributed models are not necessarily better for representing rainfall-runoff processes, especially in small catchments (Reed et al., 2004; Best et al., 2015). Rainfall-runoff production and routing are indeed more conceptual in lumped models, but the small number of parameters and the efficiency of the calibration algorithms make them powerful tools for reproducing hydrological regimes and extremes. In fact, lumped models as GR models have been used by many studies for hydrological impact studies (Coron et al., 2017; Givati et al., 2019; Arsenault et al., 2020; Séne et al., 2023).

In addition, the study aims to compare the influence of the type of climate model and hydrological model on the simulation of historical floods and future signals. Here, even if CREST performance is a bit lower in reproducing discharge, the simulation of flood distributions is considered acceptable to address the scientific questions. We wanted to keep this model as one of the goals of this work was to assess the impact of the use of different hydrological models on the simulation of future floods. Interestingly, we were able to show that the choice of the hydrological model has a limited impact compared to the choice of the climate model to determine the future climate change signal on floods.

Regarding the improvement of the model calibration, we refer the reviewer to the answer to a previous comment. Many tests were made and we selected the best option in terms of performance.

Specific comments:

74. Do French Mediterranean region face intense floods with important damages and casualties every year? I think that such great floods happen from time to time, not every year.

The French Mediterranean region does face intense floods on average every year. Floods do not occur in every catchment every year, but every year, some basins of this region are affected by floods. The map below released by the French National Weather Agency (Météo France) presents the frequency of daily cumulative rainfall exceeding 200mm by districts (source: http://pluiesextremes.meteo.fr/france-metropole/IMG/sipex_pdf/nbj_sup200_1J_dep.pdf, version of April 2023)

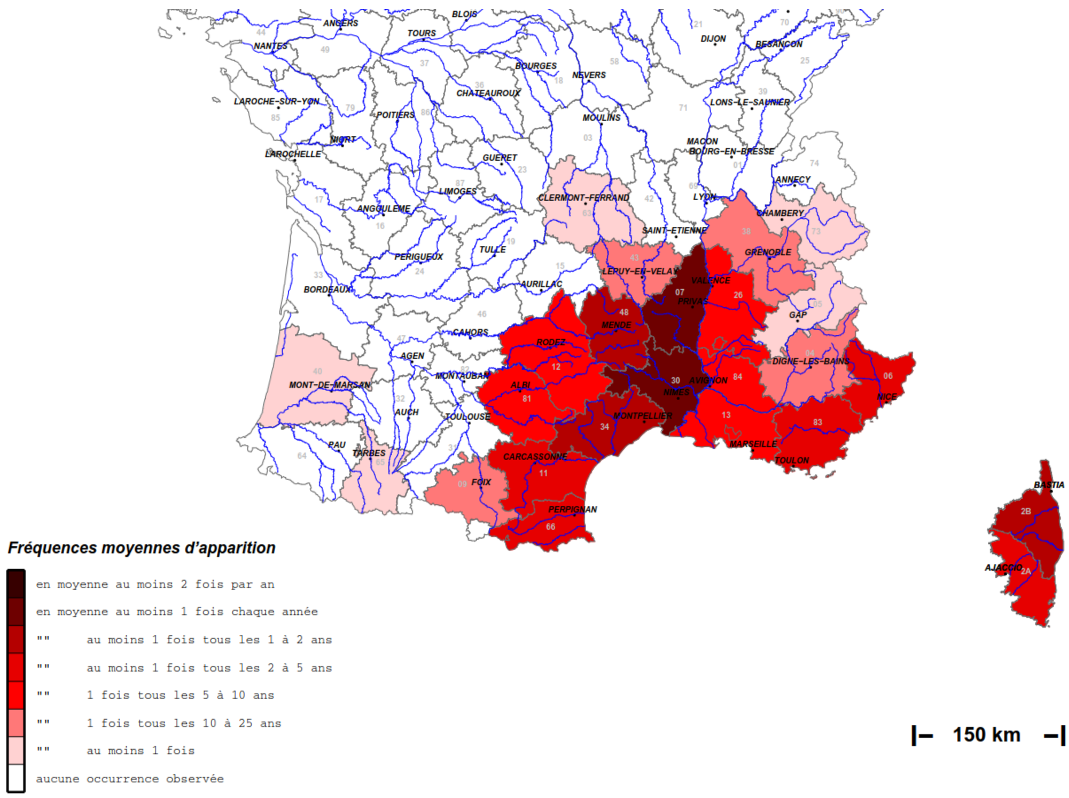

Figure 2 : frequency of daily cumulative rainfall exceeding 200mm over southern France

The color legend is from white (no occurrence) to dark brown (at least twice a year). The Gard and Ardeche districts experience on average one day above 200 mm of daily precipitation per year.

In terms of hydrological impacts, according to Boissier and Vinet (2009), it has been shown that above 100 mm/d, flood-related deaths are possible. The threshold of 190 mm in 24 hours, which is used by Météo France to describe a "torrential" downpour, causes at least one death on average one time out of three.

The website http://pluiesextremes.meteo.fr/france-metropole/-Evenements-memorables-.html makes an inventory of extreme rainfall events. From that website, Mediterranean districts can be selected and the occurrence of such events and associated impacts become available.

129. Kendon (2010) cite is missing in the References Section.

Done, Kendon et al. (2012) was added to the Reference section.

147. What are CCLM and WRF? These acronyms should be either introduced in the paper or explained to readers.

These are acronyms of regional climate models. CCLM means COSMO-CLM (Climate version of Lokalmodell) developed by the German Weather Service (Baldauf et al., 2011). WRF means Weather Research & Forecasting Model developed by the National Center of Atmospheric Research (NCAR) in the USA (Powers et al., 2017). As these CPM names are details for the bibliography and are not essential for the understanding of the manuscript, the CPM names are removed, and this sentence is finally changed to “The same conclusion was found by (Reszler et al., 2018) using simulations from two CPMs as input for…” (line 69 of the revised manuscript).

224-225. Some previous studies have used CNRM-AROME for assessing climate change, such as Monteiro et al. (2022). Such an application of CRRM-AROME should be investigated reviewing published works. In addition, the statement should be changed.

Thank you for highlighting this omission. We will add this study in the revised version of the paper.

297. Vrugt et al. (2009) cite is missing in the References Section.

We added this reference to the Reference section.

303-306. No information is supplied about how the soil parameters in both CREST and GR5H are estimated from available soil information.

The two hydrological models are calibrated. No soil information based on available local soil information is necessary to calibrate the models.

312. Only a few parameters of the CREST model have been considered in the calibration process. A description about why some parameters are fixed and others not should be included in the paper. In addition, a sensitivity analysis of the model parameters could be included in the paper to identify the parameters that can be fixed and the parameters that should be calibrated. Finally, a discussion should be included in the paper about what hydrological processes are fixed and what processes are considered in the calibration process.

Thank you for this comment. We have already answered a similar comment above (third comment). To complete the previous answer, most studies using CREST have fixed parameters to produce hydrological simulation over catchments, for example Li et al. (2022).

361-362. Simulations are corrected in each cell separately. A sensitivity analysis could be included comparing the results correcting bias in climate projections at the regional scale and cell by cell. In addition, has spatial correlation between cells been considered? Correcting bias cell by cell could change spatial correlation patterns in precipitation.

We agree with the reviewer that using a bias correction method at the pixel scale can locally change the spatial coherence of precipitation patterns. However, we verified the precipitation patterns of the most intense rainfall events over the basin before and after bias correction. The precipitation patterns are minimally impacted by the bias correction. See the figure 4 in the reviewer 2 section.

Table 1. Why nQ equals n plus 1? A comment could be included in the paper.

Thank you for this comment. It is a mistake of the transcription of the algorithm. n_Q is directly the number of timesteps exceeding the threshold. The way it is computed in the algorithm is $n_Q = j_{end} - j_{start} + 1$ with j_{end} and j_{start} respectively the indices of the last and first timestep exceeding threshold, but we prefer limiting this level of detail in the paper.

We will modify this table cell to " n_Q : number of timesteps exceeding the threshold".

483-489. After bias correction, both models underestimate precipitations associated with the highest quantiles. In addition, in Figure 3 for the period 2000-2018 both distributions are similar for quantiles close to 99.99. The main improvement of CPMs are supposed to be a better characterisation of the most extreme events. However, such an improvement cannot be seen in Figure 3. A comment should be included in the paper.

Bias correction performs as expected. All quantiles were corrected. As a result, the distributions of precipitation of AROME and ALADIN are similar after bias correction. We believe that it is not necessary to specify this as it is normal behavior for bias correction. The added value of CPM for simulating extreme events is visible on the raw CDF (top left panel) where the red curve is closer to the black one than the blue one. Note that the x axis of figure 3 is using a Gumbel transformation. The highest quantiles are stretched so the highest quantile here (0.9999) is an interpolation between the second and the last rainfall event.

614-616 and Figure 6. How simulated discharge with observed data represented by the green line in Figure 6 is obtained? The paper should include a description about how these simulations were obtained.

We will clarify this aspect in the revised version of the paper by adding a sentence line 311 : "The green line on the figure represents the hydrological simulations forced by the COMEPHORE observed dataset that has been used for the calibration and could be considered as the reference simulation".

619-623. This discussion should be clarified to improve the understandability of the paper.

We agree with the reviewer that those lines are unclear. We will rephrase them as follows: "An improvement in flood frequency modeling is noticed after bias correction for all simulations. However, the CREST simulations still have a positive bias for flood frequency with higher values than the COMEPHORE driven simulation (2.9 and 3.3 floods per year for AROME and ALADIN respectively)." (lines 329-331).

623-624. CDFs of both climate models are similar after bias correction. In addition, it seems that ALADIN performs somewhat better than AROME, as ALADIN CDF is closer to observations than AROME CDF. A deeper discussion about why CPM improves flood characterisation in this catchment should be required.

We want to thank the reviewer for this remark. As shown by our results, the flood simulations are better with the raw outputs of the CP-RCM than the RCM. The results after bias corrections are impacted by the bias correction method and cannot be used to estimate which model compares the best with observations. This has been recalled in new sentences : " Above all, only non-corrected retrospective climate simulations can be used in the scope of assessing the performance of climate models for reproducing floods. Firstly, Figure 6 shows a large underestimation of POT distributions by CREST…" (lines 314-316).

744-745. There are two brackets after a and b. One bracket should be deleted.

Done.

Section 4.5. Comments about np and POT boxplots in Figure 8 are missing.

A comment about the length of precipitation events (np) will be added as follows :

For np (lines 380-382) "There is no clear signal about the changes in the length of precipitation events (np) and the total precipitation (Ptot) amount of flood producing rainfall events. However, the maximum…"

For POT metric, the complete description and comments about flood peaks are in section 4.4, as figure 7 is a detailed version of POT boxplots of figure 8.

Figure 8. In the boxplot of FI, the y-axis scale should be changed, as boxplots are excessively small to assess the results.

Here is the corrected figure :

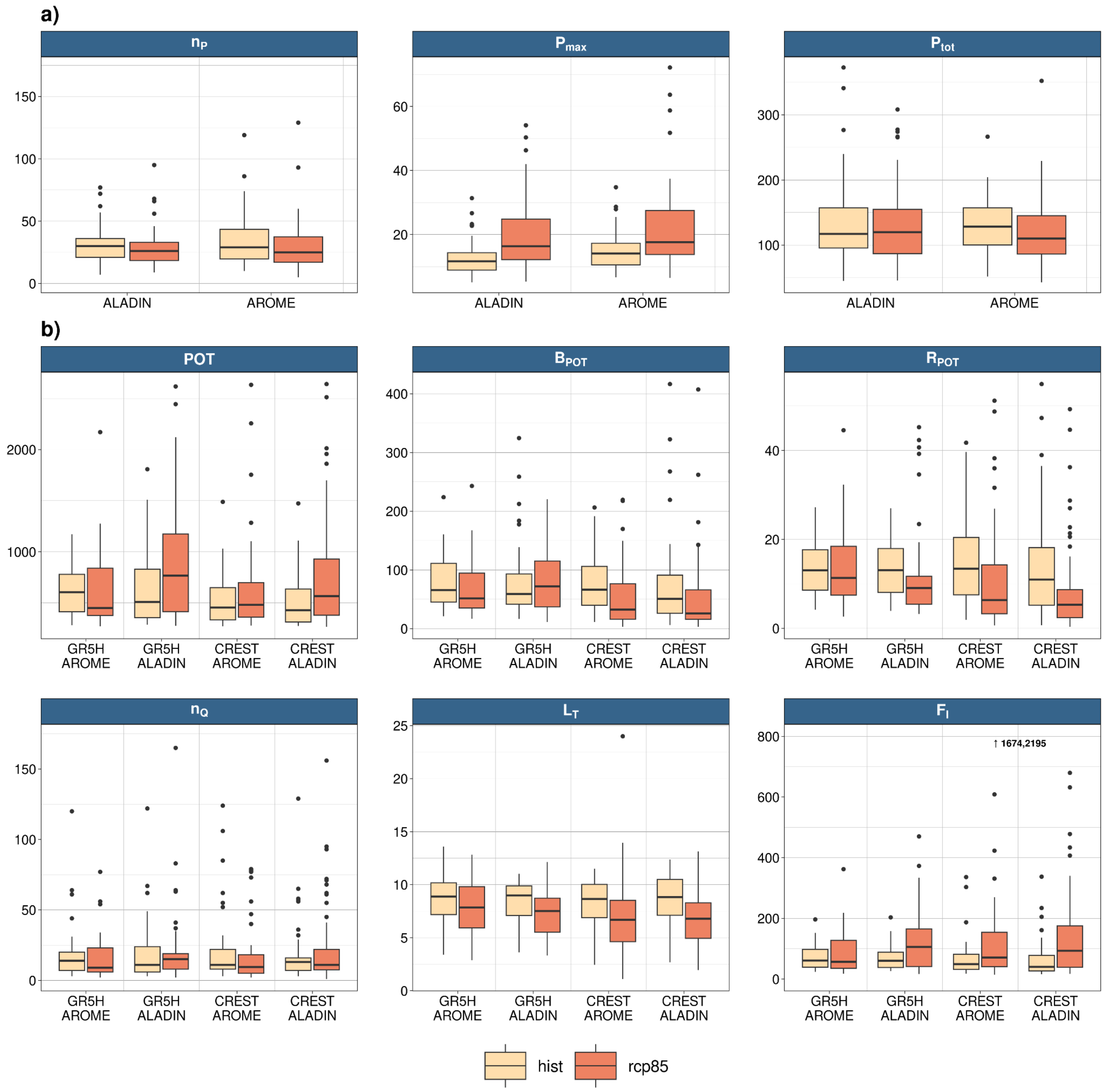

Figure 3 : Updated version of paper figure 8 with hidden outliers

813-815. The added value of CPM can be seen before bias correction. After bias correction, outputs of CPM are similar to those obtained with RCM. This should be clarified in the conclusions.

Indeed, this will be specified in the conclusions, lines 425 : “The added value of the CPM can be clearly seen on rainfall simulation before correction, with a much better representation of extremes with AROME compared to ALADIN, the latter showing a strong underestimation. However, the bias-correction reduces the difference between models since they are both corrected to match the observations. .”

821-824. It is stated that similar results are obtained regardless the hydrological model used. A comment about the limitations of the distributed model to simulate floods in this catchment should be added, given that a simpler lumped model supplies better results.

We added this detail in the conclusion. The second paragraph has been rephrased to:

“Yet, both climate simulations required a bias-correction to reproduce the observed discharge and notably flood events. Similar simulations of flood events have been obtained with the two hydrological models considered, a lumped conceptual model, GR5H, and a spatially-distributed, process-based model, CREST, with slightly better results with the GR5H model. Very similar future projections have been obtained with the two hydrological models, highlighting the robustness of the results given the two different types of model structures.”

Reviewer 2

This paper evaluates the skill of the CNRM-AROME Convection-Permitting regional climate Model (CPM) in projecting floods, using the Gardon at Anduze catchment in southern France as a case study. The CPM demonstrates superior accuracy in reproducing extreme hourly rainfall events compared to traditional models. The study underscores the potential of CPMs in future flood predictions in a warming climate.

The manuscript is interesting and generally well written, although some important aspects must be addressed before being suitable for publication in NHESS:

We thank reviewer #2 for reviewing the manuscript. The answers to all the comments are indicated below.

MAJOR COMMENT

The focus of the paper is not clear to me. I understand there has been a great effort to run two different hydrological model approaches, with two climate projection standards CPM vs high resolution but non-CPM models. The results and conclusions are too much focused on describing the experiments output with barely no interpretation. Indeed the results sections consist of what should be a caption inserted in the text and a description of the figure. What is the research question attempted to be responded by this study? The outlook for future flood scenarios is clearly not an objective as no attention to uncertainties is put.

Thank you for this important comment, we will improve the description of the results.

To clarify, the main objective of the study consist to assess the differences between using a CPM and a RCM to simulate extreme precipitation and projecting mediterranean floods. We agree that some more in-depth discussions are warranted to clarify this point, so we will add the following sentences.

For the description of figures (minor comments below):

- l337-338 : "We now analyze the flood distributions from the hydrological models forced with the AROME and ALADIN climate simulations under the historical and the future RCP 8.5 scenario (Figure 7)."
- l375-376 : "We analyzed the flood and associated rainfall events characteristics (Figure 8) simulated by historical and future bias-corrected climate simulations"".

For the interpretation of climate models flood projections :

- after line 404 : "However, the projected changes over this catchment should be interpreted with caution as it comes from a single pair of RCM and CPM simulations. With these preliminary results, the different impacts of climate change on flood characteristics highlight the fine scale benefits of CPM in simulating underlying hydrological processes. The robustness of the climate change impact on flash floods yet need to be confirmed with a comprehensive study that includes an uncertainty assessment."

MINOR COMMENTS

The title is too ambitious provided the type and robustness of the conclusions reached in the manuscript. The conclusions do not offer “new perspectives”.

Given that the title is a question and not a statement that we try to answer, it seems hard to say that it is too ambitious. However, the following lines will be added to highlight what we call “new perspectives” and guide future work on this topic :

after line 404 : “However, the projected changes over this catchment should be interpreted with caution as it comes from a single pair of RCM and CPM simulations. With these preliminary results, the different impacts of climate change on flood characteristics highlight the fine scale benefits of CPM in simulating underlying hydrological processes. The robustness of the climate change impact on flash floods yet need to be confirmed with a comprehensive study that includes an uncertainty assessment.”

On Figure 7, the future flood signals between the two types of climate models are significantly different. This major difference is potentially linked to a better simulation of rainfall event shape (convective peak) with the CPM. The attenuation of the RCM drizzle effect with the CPM yields a reduced bias in the soil moisture. For all these reasons, we think CPM can effectively bring new perspectives in projecting Mediterranean floods. This study is the first to use this modeling chain over a Mediterranean catchment. It would be interesting to extend such a study using a multi-basins and multi-CPM approach to assess further if the findings are robust..

1. **, 56 “all floods are projected to increase… the moderate floods are expected to decrease.” More precision in the language is required. Are these sentences referring to the frequency of floods? the magnitude? the flashiness?**

These sentences refer to the flood magnitude. We will rephrase it in “With the CNRM-ALADIN RCM, the magnitude of all floods is projected to increase. A threshold effect is found for simulations driven by the CNRM-AROME CPM, where the magnitude of the largest floods is expected to intensify while the magnitude of the less severe floods is expected to decrease.” (lines 14-16).

2. **“Thober et al. (2018) showed a decrease of high flows and flood magnitudes for different levels of future global warming.” Where?**

This change was stated for the Mediterranean region. After a closer look at the French Mediterranean region, we will change this statement to “Thober et al. (2018) showed no clear signal for high flows and flood magnitudes in the French Mediterranean.”

122-127 The argument presented in these lines seems vague and unfounded. Can you identify studies that directly attributes the contradictory results to the underrepresentation of sub-daily extremes?

Even if hydrological processes are complex and non linear, in the Mediterranean region, the strongest floods are caused by the heavy precipitation events (HPE). It can be assumed that a poor simulation of these extreme rainfall events yields poor simulations of the most intense Mediterranean floods. Furthermore, most of the RCMs do not simulate precipitation at an hourly timestep but every 3h or 6h, leading to biases in the simulation of flashy and sub-daily floods.

The contradictory results on the French Mediterranean region are stated from lines 43 to 48 of the revised manuscript. To our knowledge, no study has ever made a review on flood projections specifically on this region.

1. **The resolution required to explicitly simulate convective processes is not necessarily determined at 4km. This statement is too absolute.**

Indeed, this statement is too strict. The resolution of a CPM can reach up to 4.5km and it is still unclear if the deep convection parameterization schemes can be removed for resolutions between 5 and 10km (Vergara-Temprado et al., 2020). We modified slightly the text line 57 to make this statement less absolute.

2. **"get rid of" seems too colloquial.**

Changed to "remove the need of".

167/176 and 168. floods→ flood and emission→emissions

Done.

1. **"Evaluate the added value of the CPM on extreme rainfall", you mean "the simulation of"? "the prediction of"**

We meant "the simulation of", This was added in the text.

2. **The COMEPHORE is a high quality dataset but is not an "observation". As the text already states, it is a high resolution analysis gridded field of precipitation. This should be corrected throughout the manuscript.**

We agree with the reviewer and we will change "observation" to COMEPHORE when possible. As sometimes observation refers to both temperature and precipitation in the manuscript, we have added a sentence to prevent confusion : "COMEPHORE rainfall dataset and PET derived from interpolated temperature are the best observation-based estimates in this area and are therefore considered as the reference observational datasets in this manuscript." (lines 136-138).

242 and 243. Can you find alternative expressions to refer to the radars used that avoid the relative reference as “foreign” of the Swiss and Jersey Islands radars. They are not “foreign” to some readers.

“Foreign” is to be placed in the context of “French radar network”. We can change this sentence to “In 2019, COMEPHORE was built using data from 29 radars comprising the French radar network (ARAMIS), in addition to radars from the Swiss network and another one on Jersey Island” (line 128).

243-245. Are there any references that support this statement. “is still considered” should be backed by a referenced source.

This sentence will be changed in “COMEPHORE can be considered as the best national precipitation product for studying hourly rainfall at high spatial and temporal resolutions”. No paper has ever compared all the French precipitation datasets, but Fumière et al. (2020) compared it to another dataset called SAFRAN and Caillaud et al. (2021) used COMEPHORE as a reference dataset for the study of Heavy Precipitation Events.

1. **I think the change of subject towards temperature use for PET must imply a new line/paragraph.**

Changed.

2. **Can you find a more rigorous description of the differences between the two hydrological models than “physical concepts”?**

Our aim here was to provide a general overview of the different concepts between the two models. The differences are better explained in the two following paragraphs describing each hydrological model.

3. **Describing the use of a hydrological model as a system that “transforms” precipitation into discharge sounds naïve.**

This statement will be improved as follows : “this model uses catchment-aggregated hourly precipitation and potential evapotranspiration data to simulate hourly discharge”

296-297. What is the calibration set of flood cases used? Are these extreme floods? How does this affect the application of the model for extreme events in the experiments?

Both hydrological models have been calibrated over the complete 2002-2018 period. This period contains several floods (see figure 5, black curve) and especially the major 2002 flood. However, we want to stress that since these hydrological models are run in a continuous way, agreement with the best practice in hydrological modeling, the entirety of the full time series (i.e. also including non-flood time steps) has been used.

The Nash and Sutcliffe Efficiency (NSE) criterion used for the calibration gives more weight to the simulation of high flows compared to low flows. These observed extreme floods are therefore well considered during the calibration.

344-345. The presence of biases in climate simulations hampers their use in virtually all impact studies, not only in hydrological applications.

Yes, that is true. We will replace “using hydrological models” by “such as for hydrological impact modeling”.

359-362. How do calibrated precipitation maps look like? Is the spatial correlation of the original pcp fields lost?

The spatial correlation is little affected by the bias correction. Figure 4 presents a few rainfall maps of three heavy rainfall events. Raw data is on the left and bias corrected data on the right. From these pictures, we can notice that the precipitation patterns are roughly the same. We added a sentence lines 213-214 in the revised manuscript to specify this aspect.

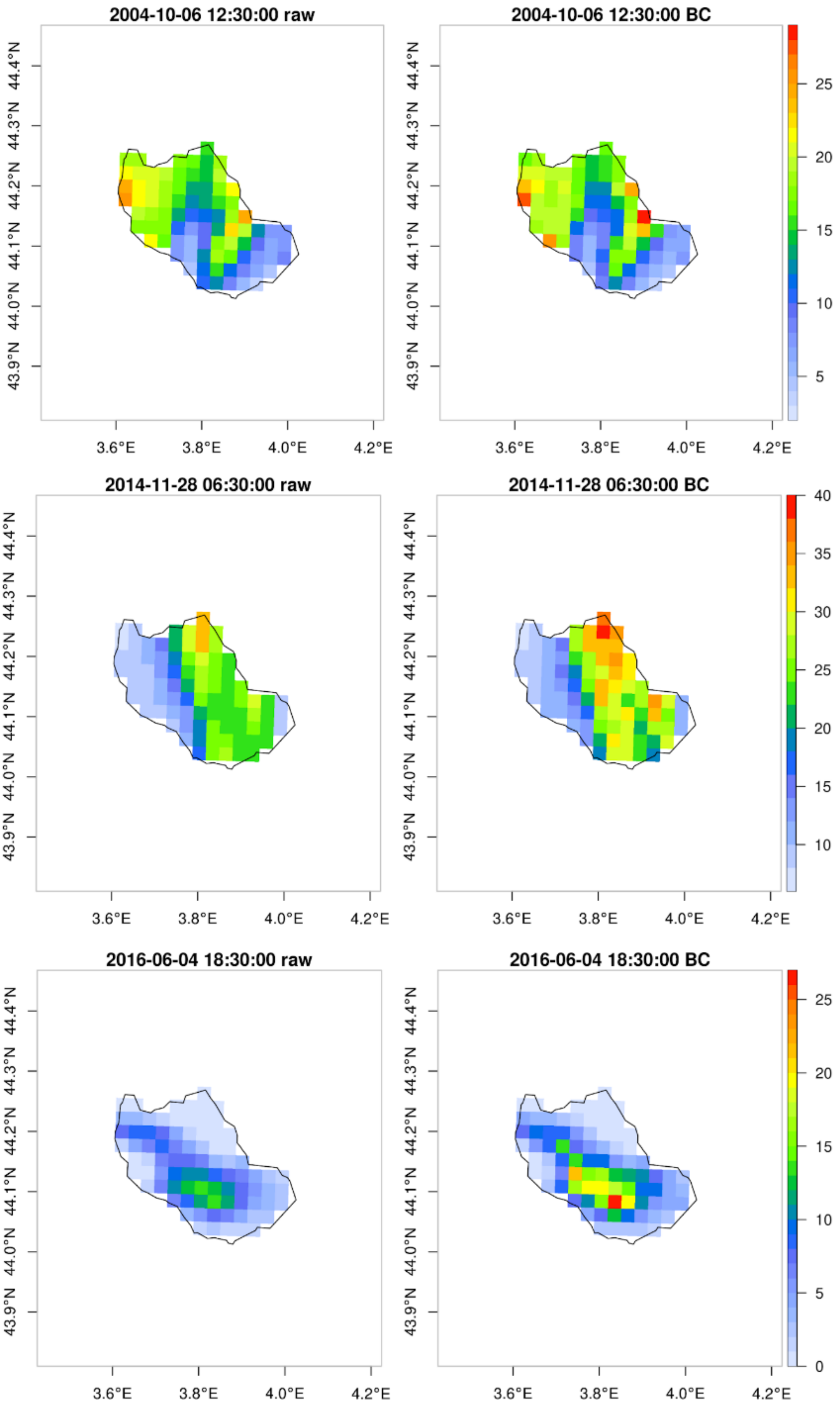

Figure 4 Hourly simulated rainfall before (raw) and after (BC) bias correction for some events

1. **Add "…periods of the same…"**

Sorry but we cannot find where in the text this comment comes from.

405-406. "The rainfall thresholds are related to our knowledge of the river basin dynamics and hydrological expertise." Can you provide any hint or supporting evidence?

We based this assessment on the analysis of flood hydrographs and associated rainfall events. We compared different thresholds to extract the flood events and different parameters to delimitate related rainfall events. Figure 5 is an example of an event. The yellow array represents the barycenter of the rainfall event computed over the 24h (window) before the flood peak. Given the relatively small sample size (one catchment only), it is more efficient to validate manually this threshold rather than using an automatic algorithm.

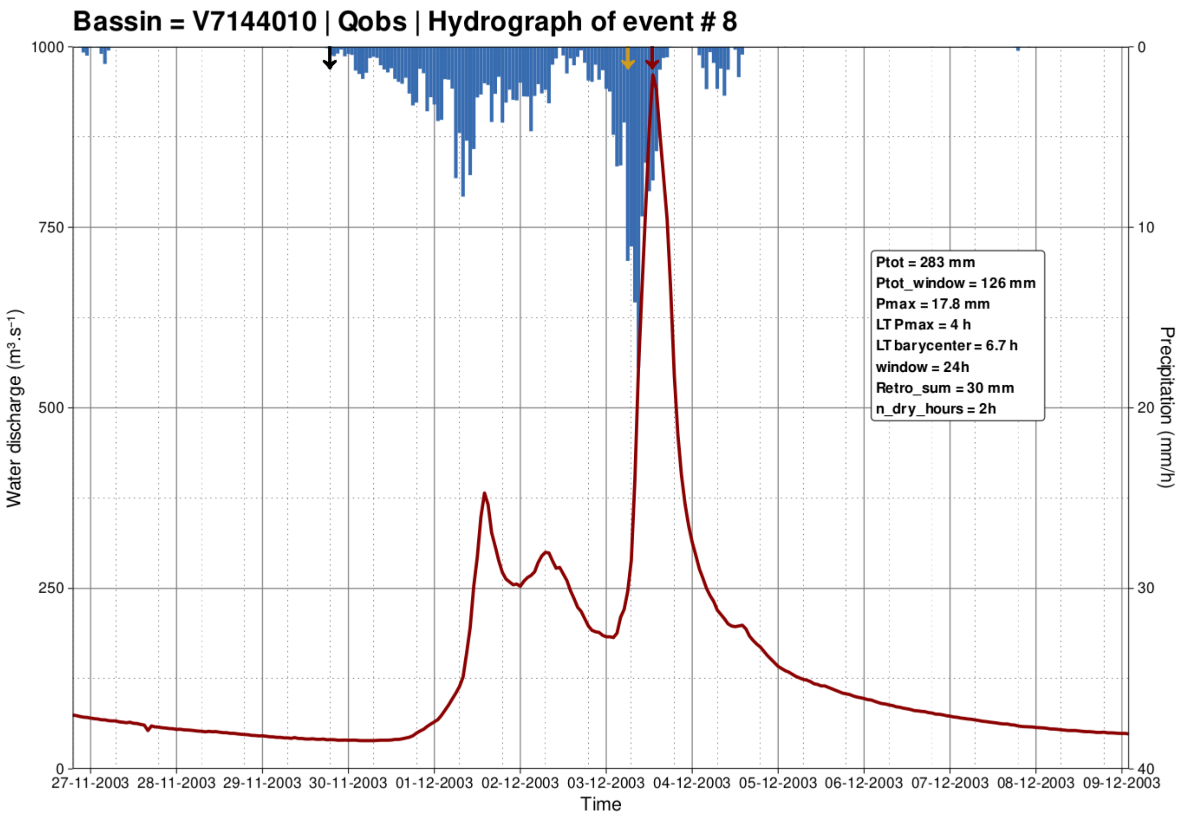

Figure 5 : Example of a hyetograph and hydrograph for an event

459-461. Is this spring-autumn connection possible provided the intense hydric deficit that characterizes Mediterranean climate, that acts as a drying reset to the hydric cycle in most surface basins in the region?

Thank you for this relevant comment. This statement is an hypothesis. It is true that the summer months are really dry over this basin leading to a strong hydric deficit, even after a wet spring. The water stored after wet springs in ALADIN could potentially go through the summer months depending on the representation of hydrological processes and soil storage inertia in the models and depending on the presence of drizzle effect during the early summer. To illustrate that, we performed two simulations of the GR5H model on catchment Y4615020 for the period March to December 2006. The two simulations consist of the same model with different initial states as of 1st March 2006: one with average amount of water in the model storages, one with high levels of water in the initial storages, to illustrate a wet bias during spring. Figure 6 presents the simulated discharge over March to December 2006. Logically, the model with high storage levels for the initial state presents higher simulated flows for the first time steps. However, what is interesting is that even after the dry summer period, where simulated discharges are very close for both models, we see that during the first autumn floods the model with the high storage levels on 1st March still presents higher simulated discharges. This is well visible on Figure 7. Although presented here as a fictive case, this illustrates that summer, even in the Mediterranean area, does not necessarily reset the hydric states in the model. It therefore makes our hypothesis plausible.

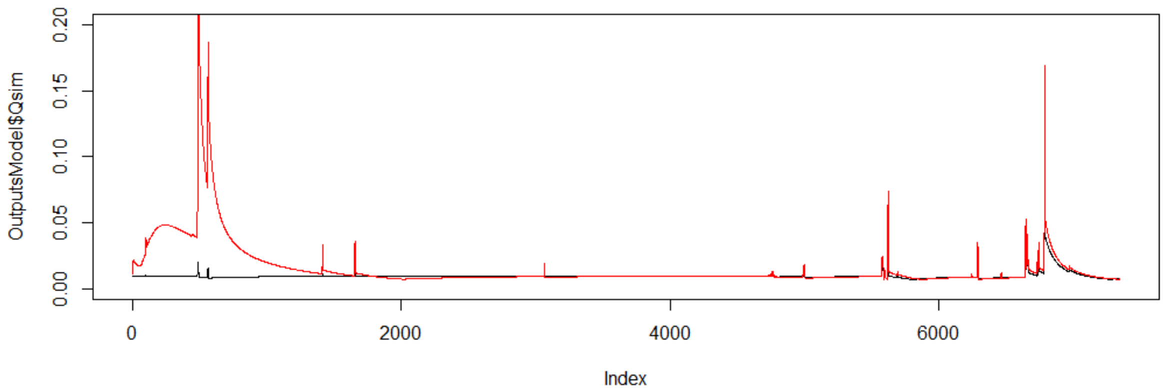

Figure 6: Simulated discharge for the same model with high (red) or normal (black) initial states as of 1st March 2006 (first time step of the model).

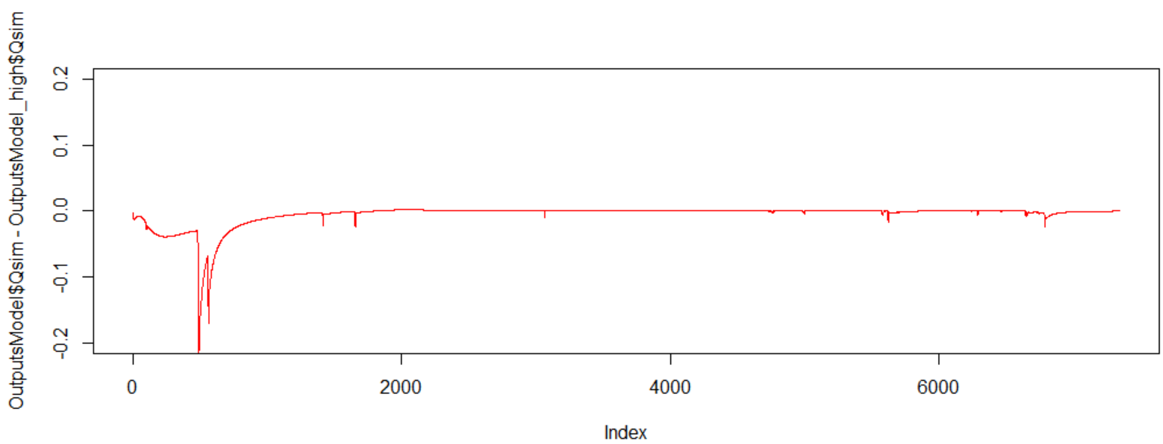

Figure 7 : Difference between the simulated discharge of the normally initialised model and the highly initialised model

Figure 3. Being COMEPHORE an analysis, it very likely underestimates actual precipitation peaks. Despite that, none of the two models reach its extremes. Can you comment on that and the implications for the projected scenarios and the derived hydrological conclusions?

Indeed, the first paper (Laurantin et al., 2012) describing COMEPHORE concluded that COMEPHORE has a tendency of underestimating some extreme values. A sentence has been added lines 129-130. However, the river basin is located in a region with many radars and weather stations and that COMEPHORE has been improved over time (Caillaud et al., 2021). Extreme events are by definition rare and it is possible that for the short simulated periods, the CPM and RCM cannot reproduce the most extreme observed events, even though the observed extreme are likely underestimated. It is also known that despite the added value of CPM in simulating precipitation extremes, such models still underestimate the most extreme events. We will clarify these aspects in the revised manuscript. Concerning hydrological implications, it is not as straightforward. Observed discharges could be biased too, especially during the highest floods, because of gauging difficulties and uncertainties related to the rating curve. We cannot quantify the importance of the COMEPHORE biases for future flood projections, but they are very likely to be minor compared to the biases of the river discharge that are partly compensated through the calibration of the hydrological model parameters.

555-558 After asking the reader to keep in mind an aspect, the authors are expected to make an important statement that requires to keep that in mind. What is it here?

This aspect is recalled in the following lines: “some of the parameters of CREST are fixed and the ones of GR5H are free”.

Figure 6. The bias corrected results show that AROME underestimates Flood Peak Discharges more than ALADIN. Doesn’t that tell opposite messages than the main point of the paper?

On Figure 6 of the manuscript, the reference dataset is not the observation, but the green curve where the different hydrological models have been forced by COMEPHORE. A sentence will be added to explain that l311 : “The green line on the Figure 6 represents the hydrological simulations forced by the COMEPHORE observed dataset that has been used for the calibration could be considered as the reference simulation.”

For GR5H, after correction, both distributions are quite similar even if the shape of the AROME CDF(red one) is closer to to the green one (less concave).

For CREST, the ALADIN curve tends to overestimate a bit between quantiles 0.6 to 0.8 and the AROME curve tends to underestimate the same quantiles.

Furthermore, the main point of the paper is the comparison between AROME and ALADIN under future projections.

1. **Can you provide a better section title?**

We can propose : “Hydrological models evaluation”

2. **“determine how the flood distribution will evolve in the future” This statement is way too pretentious. I don’t think we currently have tools that can do that. At most, current tools generate projection, but not outputs that “will” occur.**

We agree with the reviewer. We will rephrase it to: “This section aims to provide an overview of the flood signal suggested by future climate projections and whether this projected evolution…”

643-645. This is a caption for Figure 7. Avoid using this type of sentence in the argumentation text throughout the results section.

Indeed, it will be changed to “We now analyze the flood distributions coming only from the hydrological models forced with the AROME and ALADIN climate simulations under the historical and the future RCP 8.5 scenario (Figure 7).”

744-745. Idem.

Ok. Changed to : “Then, we analyzed the flood and associated rainfall events characteristics (figure 8) simulated by historical and future bias-corrected climate simulations”.

795 if→ whether

Ok.

797-801 These two sentences seem to state contradictory messages. “Until now” “regional models … cannot” and “In the last 10 years …CPM”. Are regional models that now allow (…) or CPMs?

CPM are a type of Regional Climate Models, we specified it in the sentence line 413 to clarify this point.

- **Despite the last paragraph of the conclusions section mentions the lack of attention to uncertainties, previous parts of the text attributes predictive capacity to the set of experiments done. For instance, the 3rd paragraph of the conclusions present results of the experiments with 2080-2099 data as “future” predictions.**
- **The use of the label “future” to describe results is excessive, provided the lack of robustness of the single projection used. I recommend using “projection” and avoid presenting the scenarios calculated as an interesting outcome of the work, provided that not uncertainty analysis is done.**

Thank you for your comment. We removed the label “future” to make these statements less predictive.

Comment to the reviewers and editor

Since the submission of the paper, a problem has been identified in the AROME future simulations. As mentioned by Caillaud et al. (2023) : “For CNRM-AROME41t1, a bug was recently found in the GHG concentrations : they evolve, but do not completely follow the RCP8.5 scenario.”

The impact on temperature change has been tested and is marginal since the RCM emissions and temperature are correct and the lateral boundaries forcing takes over from internal CPM forcing.

This issue will be revealed in a coming note.

Bibliography

Arsenault, R., Brissette, F., Chen, J., Guo, Q., and Dallaire, G.: NAC2H: The North American Climate Change and Hydroclimatology Data Set, Water Resour. Res., 56, e2020WR027097, https://doi.org/10.1029/2020WR027097, 2020.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., and Reinhardt, T.: Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities, Mon. Weather Rev., 139, 3887–3905, https://doi.org/10.1175/MWR-D-10-05013.1, 2011.

Best, M. J., Abramowitz, G., Johnson, H. R., Pitman, A. J., Balsamo, G., Boone, A., Cuntz, M., Decharme, B., Dirmeyer, P. A., Dong, J., Ek, M., Guo, Z., Haverd, V., Hurk, B. J. J. van den, Nearing, G. S., Pak, B., Peters-Lidard, C., Santanello, J. A., Stevens, L., and Vuichard, N.: The Plumbing of Land Surface Models: Benchmarking Model Performance, J. Hydrometeorol., 16, 1425–1442, https://doi.org/10.1175/JHM-D-14-0158.1, 2015.

Boissier, L. and Vinet, F.: Paramètres hydroclimatiques et mortalité due aux crues torrentielles. Etude dans le sud de la France, in: Actes du 22ème colloque de l’Association Internationale de Climatologie, 1er au, 79–84p, 2009.

Caillaud, C., Somot, S., Alias, A., Bernard-Bouissières, I., Fumière, Q., Laurantin, O., Seity, Y., and Ducrocq, V.: Modelling Mediterranean heavy precipitation events at climate scale: an object-oriented evaluation of the CNRM-AROME convection-permitting regional climate model, Clim. Dyn., 56, 1717–1752, https://doi.org/10.1007/s00382-020-05558-y, 2021.

Caillaud, C., Somot, S., Douville, H., Alias, A., Bastin, S., Brienen, S., Demory, M.-E., Dobler, A., Feldmann, H., Frisius, T., Goergen, K., Kendon, E., Keuler, K. G., Lenderlink, G., Mercogliano, P., Pichelli, E., Soares, P. M. M., Tölle, M., and Vries, H. de: Mediterranean Heavy Precipitation Events in a warmer climate : robust versus uncertain changes with a large convection-permitting model ensemble, https://doi.org/10.22541/essoar.168987136.64498273/v1, 2023.

Coron, L., Thirel, G., Delaigue, O., Perrin, C., and Andréassian, V.: The suite of lumped GR hydrological models in an R package, Environ. Model. Softw., 94, 166–171, https://doi.org/10.1016/j.envsoft.2017.05.002, 2017.

Fumière, Q., Déqué, M., Nuissier, O., Somot, S., Alias, A., Caillaud, C., Laurantin, O., and Seity, Y.: Extreme rainfall in Mediterranean France during the fall: added value of the CNRM-AROME Convection-Permitting Regional Climate Model, Clim. Dyn., 55, 77–91, https://doi.org/10.1007/s00382-019-04898-8, 2020.

Givati, A., Thirel, G., Rosenfeld, D., and Paz, D.: Climate change impacts on streamflow at the upper Jordan River based on an ensemble of regional climate models, J. Hydrol. Reg. Stud., 21, 92–109, https://doi.org/10.1016/j.ejrh.2018.12.004, 2019.

Hewitt, C. D. and Lowe, J. A.: Toward a European Climate Prediction System, Bull. Am. Meteorol. Soc., 99, 1997–2001, https://doi.org/10.1175/BAMS-D-18-0022.1, 2018.

Laurantin, O., Tabary, P., Dupuy, P., L'Henaff, G., Gueguen, C., and Moulin, L.: A 10-year (1997—2006) reanalysis of Quantitative Precipitation Estimation over France: methodology and first results, in: IAHS-AISH publication, 255–260, 2012.

Li, Z., Gao, S., Chen, M., Gourley, J. J., Liu, C., Prein, A. F., and Hong, Y.: The conterminous United States are projected to become more prone to flash floods in a high-end emissions scenario, Commun. Earth Environ., 3, 86, https://doi.org/10.1038/s43247-022-00409-6, 2022.

Lucas-Picher, P., Brisson, E., Caillaud, C., Alias, A., Nabat, P., Lemonsu, A., Poncet, N., Cortés Hernandez, V. E., Michau, Y., Doury, A., Monteiro, D., and Somot, S.: Evaluation of the convection-permitting regional climate model CNRM-AROME41t1 over Northwestern Europe, Clim. Dyn., https://doi.org/10.1007/s00382-022-06637-y, 2023.

Powers, J. G., Klemp, J. B., Skamarock, W. C., Davis, C. A., Dudhia, J., Gill, D. O., Coen, J. L., Gochis, D. J., Ahmadov, R., Peckham, S. E., Grell, G. A., Michalakes, J., Trahan, S., Benjamin, S. G., Alexander, C. R., Dimego, G. J., Wang, W., Schwartz, C. S., Romine, G. S., Liu, Z., Snyder, C., Chen, F., Barlage, M. J., Yu, W., and Duda, M. G.: The Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions, Bull. Am. Meteorol. Soc., 98, 1717–1737, https://doi.org/10.1175/BAMS-D-15-00308.1, 2017.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J., and DMIP Participants, and: Overall distributed model intercomparison project results, J. Hydrol., 298, 27–60, https://doi.org/10.1016/j.jhydrol.2004.03.031, 2004.

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, Nat. Hazards Earth Syst. Sci., 11, 2567–2582, https://doi.org/10.5194/nhess-11-2567-2011, 2011.

Séne, S. M. K., Faye, C., and Pande, C. B.: Assessment of current and future trends in water resources in the Gambia River Basin in a context of climate change, In Review, https://doi.org/10.21203/rs.3.rs-3143338/v1, 2023.

Vergara-Temprado, J., Ban, N., Panosetti, D., Schlemmer, L., and Schär, C.: Climate models permit convection at much coarser resolutions than previously considered, J. Clim., 33, 1915–1933, https://doi.org/10.1175/JCLI-D-19-0286.1, 2020.