Anonymous Referee #3

The authors perform an evaluation study of a convection-permitting model (CPM) at 3 km resolution. The simulation domain covers the south-east of France and part of the Mediterranean Sea. The CPM downscales a 0.11° model, which was run over the EURO-CORDEX domain. A nice aspect of the study is that there are three realisations of the CPM simulations, each based on a different GCM; this aspect could potentially be given mode attention, as it is unusual in the literature. The authors then evaluate the performance of the CPM and 0.11° models and conclude that the CPM produces more realistic precipitation.

I think the study is a reasonable contribution to the literature and could in principal be published. However, at present the study has some limitations which must first be addressed before publication. These are detailed below in the main comments section.

Response: We thank the reviewer for all the constructive comments and suggestions which do help us to improve our manuscript.

Main Comments

1. Novelty and relation to similar literature. In their abstract, the authors state of their climate-length convection-permitting simulations that "... this approach has never been used in a climate simulation for the Mediterranean coastal region" (L4-5). There's a similar statement in the Introduction (L69-70): "... such long simulations, to the best of our knowledge, have never been done for coastal area in the Mediterranean region".

This is not correct. I can think of at least five studies which perform convection-permitting simulations at climate timescales over the north-western Mediterranean, which the authors don't cite. These studies all cover the area of the CPM domain used by the authors, as opposed to the studies of e.g. Armon et al. (2020) and Zittis et al. (2017) cited by the authors, which are for other parts of the Mediterranean and aren't on climate timescales. The studies I have in mind are (there may be more):

[1] Berthou et al.: https://doi.org/10.1007/s00382-018-4114-6

[2] Vergara-Temprado et al.: https://doi.org/10.1029/2020GL089506

[3] Meredith et al.: https://doi.org/10.1088/1748-9326/ab6787

[4] Adinolfi et al.: https://doi.org/10.3390/atmos12010054

[5] Caillaud et al.: https://doi.org/10.1007/s00382-020-05558-y

Ref. [1] has a specific section on heavy precipitation events in SE France. Refs. [4] and [5] also assess intense hourly and daily precipitation events in CPMs over France using similar observation sets to the present authors. Ref. [3] uses the same annual re-initialization technique as the authors and also focuses on the Autumn months in the NW Mediterranean, just as the present authors do.

Around lines 55-67 it would also be good to cite these climate-scale studies, as most of those presently cited are for case studies or selected events.

The authors need to cite and discuss the relevant literature, not just in the Introduction, but also where appropriate in the Results and Discussion. The results of the present authors should be presented in the context of the pre-existing relevant literature. That means, wherever appropriate, compare your results with those in the pre-existing literature. This is particularly important if your results are different, in which case possible explanations would be helpful.

Response: We thank you for suggesting more new studies investigating convection-permitting simulations, especially their domains that cover the French Mediterranean region. They are helpful for us in improving our introduction as well as provide new material to discuss our results. And we will try to look up more concerning studies to improve our discussion.

We will modify our statement in both the abstract and introduction. For the sentence in line 4 and 5 "However, this approach has never been used ...", we will change to "*This approach has been tested and performed at climate scale in several studies in recent decades for different areas*". We will also remove the sentence in line 69 and 70 "However, such long simulations, ... for coastal areas in the Mediterranean region".

2. Comparison of model data and observations at different spatial scales.

A major issue with the evaluation is that model data and observations on different spatial scales are being compared directly. While it's arguable that model data on a 3 km grid could be compared directly with station data, what do the authors hope to learn by comparing data on a 12 km grid (that means grid box averages over an area of 12 x 12 = 144 km2) with station data (point values)? Or even with the 1 km COMEPHORE product?

Response: We have noticed that there are discrepancies in spatial scales among CPSs, EUR-11s and observations datasets. We agree that the spatial averaging effect could smooth out extremes. However, from a model user perspective, one would compare directly what models produce to local observations to serve, e.g. the climate impact at local scale. For instance many users directly use reanalyses with resolutions from 5 to 50 km for local studies, possibly coupled with some statistical techniques. For such applications, the fact that a higher resolution model has a larger spatial variability is to be counted in the added value of the model. By using a higher resolution without a convection parameterization in model configuration, we found here the improvement in reproducing the extreme in smaller spatial and temporal scales. This improvement can in part be explained not only by removing averaging effects, but also other factors coming into play (e.g. moisture convergence, better resolved mountains and flows). Our goal here is to assess the overall change in comparisons to station data.

It is not surprising that Figures 4, 5 and 6 show the lowest intensities in the 12 km model, followed by the 3 km model, followed by the point observations. This simply reflects the fact that the extremes are being averaged over ever greater areas as the grid spacing increases, thus the intensities are "smoothed out"; the same applies to the "mean 14/23 stations" in Figs. 2 and 3. Indeed this also applies to the box means in Figs. 2 and 3, because the area mean of high-resolution extremes must by higher than the area-mean of low-resolution extremes. These comparisons don't tell us whether or not the 12 km model is worse than the 3 km model, or vice versa. Suppose your 12 km model was perfect at the 12 km scale: the extreme intensities would still be much lower than those at the 3 km or point scale. Or imagine you aggregated your 1 km COMEPHORE data to the 12 km grid and then compared it against the 1 km data at some point: the 12 km data would have a strong negative bias, even though it's the same dataset. For further discussion of this topic, I suggest the study of Göber et al. (2008, https://doi.org/10.1002/met.78).

Response: We disagree with the reviewer at this point (however we did additional analyses to check what the reviewer suggests, see below). Our goal here is to assess the overall improvement against observed station data (resolution and other processes) with CP set-up (see above) against previous approaches. Our goal is not to disentangle causes of improvement and processes which would require a much longer study with other types of data at climatic time scale. In our simulations, the resolution is not the only factor changed, also the convection scheme is switched off. In addition, a few other parameterizations concerning turbulence in the gray zone were developed for coarser resolutions that can lead to additional systematic biases when applied in CPS configuration (Prein et al., 2020). All these mean that the CPS configuration also introduces different changes and potential sources of uncertainty. Comparing 12-km resolution and the CPS at the same low resolution would only answer the question of how the finer resolution can improve larger-scale phenomena.

In the case of Fig. 4 (temperature scaling), what's important is that the models have similar scaling curves to observations, the intensities don't need to match to validate the models.

Response: We thank you for this suggestion. Indeed, the convection-permitting simulations are able to reproduce similar scaling patterns to observations that should be mentioned in the text.

As pointed out in Göber et al. (2008), the standard/appropriate way to compare observations and model data is by upscaling the observations to the coarsest model grid (EUR-11 in your case). The CPS publications the authors cite all upscale their observations to the coarsest model grid: Kendon et al. (2012), Fosser et al. (2015), Knist et al. (2018), Chan et al. (2013, 2014). Also Refs. [1], [4] and [5] above.

Response: In our opinion, there is no "standard" way of evaluating model performance. This depends on which scientific question is being addressed. In our study, we only focus on how/to what extent the CPSs improve extreme precipitation at a local scale rather than the question why a finer resolution improves the final output. Upscaling simulations at a coarser scale would only partly assess the relative CPSs by answering whether newly-resolved processes improve the coarser scale. We are fully conscious and agree that this is an interesting question, but not the one we address here by comparing simulations with station data.

In the cases of the gridded observations (SAFRAN, COMEPHORE), it is certainly possible to compare models and observations at the same spatial scale (i.e. that of EUR-11) through conservative remapping. In the case of the station data, there's no simple solution. As stated above, comparing the 3 km intensities with stations could be defensible. I don't see much value in comparing the 12 km intensities with stations; but if the authors really want to do this then they need to give a very strong warning to the reader that this has limitations, and these limitations should be communicated in the text.

Another indicator that the results might be being affected by the comparison of different spatial scales is the added value you find for daily precipitation. Studies show that CPMs generally don't add value for daily mean or extreme precipitation, e.g. Refs. [1] and [4] above, Chan et al. (2013), Ban et al. (2014). It's likely that a lot of the added value you find for daily precipitation statistics is simply due to the different spatial scales you're comparing against observations. Having said that, Berthou et al. (2018, Ref. [1]) did find added value at the daily scale for CPMs in the case of autumnal precipitation extremes in the Mediterranean.

Response: We performed additional graphs showing results from the CPSs upscaled to EUR-11 resolution using the conservative remapping (referred as CPS-11, see Fig.1 at the end of this document). This additional analysis shows that the CPS-11s still have good agreement to both gridded and in situ observations though their resolutions/scales are different. The CPS-11s give similar biases of statistics concerning the mean of the Cévennes box against mean of stations within the box for Rx1day, while their maxima of box/stations deviate from the observed values by roughly -24% to 14%. This means that the CPSs do improve the results of Rx1day in our study after considering the upscaling analysis. This also confirmed what was found in Fumière et al., (2019) with AROME model forced by reanalysis data. Therefore, we will mention this point in our manuscript and state that our conclusion remains the same when we perform the upscaling investigation for the CPSs.

Other Comments

1. Ideally this study would have been performed using reanalysis as boundary forcing. Since you are using free-running GCMs, you therefore need to inform the reader early on (i.e. in the methods) that the regional models will inherit biases from the GCMs, and that any biases you find therefore result from a combination of both the GCM and RCM biases. Later on in your results, we see quite different results depending on what the GCM is, so the role of the GCM is clearly not trivial.

Response: We thank you for raising this point. We will discuss this point in the revised manuscript.

2. The CPM simulations cover the Autumn months because this is the time when the most intense events occur in SE France. Maybe not all readers will be aware of this or know why, as many expect the most intense short-duration events to be in the summer. I think a few sentences in the Introduction and/or Methods explaining why the strongest events are in Autumn would be useful. E.g. warmer Mediterranean SSTs, low pressure systems advecting warm moist air at lower levels from the Mediterranean into southern France and then orographic lifting, etc. Maybe the studies of

Labeaupin et al. (2006, https://doi.org/10.1029/2005JD006541) and Toreti et al. (2010, https://doi.org/10.5194/nhess-10-1037-2010) would be of interest to you.

Response: We agree and will add the explanation into the method section of the revised manuscript.

3. Temperature scaling of extreme precipitation (L118-126). What steps have you taken in order to avoid effects from under-sampling? Do you require some minimum value of data points to be in a bin before you compute the percentile? If so, what? Boessenkool et al. (2017, https://doi.org/10.5194/nhess-17-1623-2017) show that the downturn at higher temperatures can simply be a statistical artefact if the bins are not sufficiently populated. In your Figure 4, the deviations away from CC or 2xCC scaling occur at low and high temperatures, exactly the range where there are less events. This could be due to insufficient data points in the bins.

Response: We have not applied any rule to avoid under-sampling effect. Therefore, in this revision, we chose a threshold of at least 300 points to take a bin into consideration. This helps to eliminate a few bins in lowest and highest temperature ranges. However, the hook shape remains at high temperature ranges that suggests that the lack of moisture plays a role (Hardwick Jones et al., 2010).

Also, in Figure 4, do the numbers in the inset table represent the mean scaling rates? If so, how do you compute them? Over the entire range of data? Or is it an average across all stations?

Response: We pooled all stations together and applied the scaling procedure to obtain what was shown in Figure 4. The numbers in legend show mean scaling rates over all bins for each dataset.

4. There are lots of different data sets used: Gridded data, 14 stations, 23 stations, etc. When the biases are presented in the text (Section 3.2), it is sometimes not clear with respect to which data the bias is for. It might help the reader if you state this more explicitly in the text.

Response: All the biases presented in the text came from the comparison of simulations against in situ observations. For daily indices, we compared the simulations with the 14 stations within the Cévennes, while we used 23 stations set for evaluating the 3-hour indices from simulations. We will clarify explicitly in the text.

Minor Comments

L15-16: "because of the limitation in computer resources, deep convection processes have rarely been solved explicitly in long climate simulations". This is again a bit of an exaggeration with respect to the existing literature. There are really quite a lot of CPM studies on climate timescales. For example, there are the studies which you already cite: Ban et al. (2014, 2015), Fosser et al. (2015), Hodnebrog et al. (2019), Kendon et al. (2014), Knist et al. (2018), Vanden Broucke et al. (2019). Then there are the five I've listed under Main Comment 1. There are a lot more if you take a look on Google Scholar, and not just for Europe like those already listed.

Response: We will change two sentences from line 15 to 18: "However, because of limitation in computer resources, … with prognostic variables have been designed to represent this process at local scale (Kendon et al., 2012)" into "*However, deep convection processes have been parameterized in simulations at climate scale for a long period of time. The parameterization methods that are based on statistical properties of convection processes within a grid box and their interactions with prognostic variables have been designed to represent this process at local scale (Kendon et al., 2012).*"

L28-44: Please remember to also cite literature relevant to your study region.

Response: We will improve that paragraph with literature relevant to the Mediterranean area.

L41: "added value" is always singular, i.e. not "added values". Also in other parts of the manuscript. L82: Could you please also give the resolution of the CPM in degrees?

Response: We have replaced "*added values*" by "**added value**" in line 41. The resolution of CPM is 0.0275°. We have added it to the text in line 82.

L103: Could you please state what the model top is? With only 32 levels, the spacing between layers could be quite high. You should avoid having a vertical spacing which is greater than your horizontal spacing, which may be a risk here for your CPM simulations. It's too late to change this now, but it's useful to keep in mind for the future.

Response: We thank you for your suggestion. The top model level is 50 hpa, roughly 20 km. We bear in mind that for such a high horizontal resolution, e.g., 3 km in our study, we should raise the number of vertical levels into at least 60.

L105-109: Is the shallow convection parametrized in the CPM? If so, what scheme?

Response: We do not use an independent shallow convection scheme in our convection-permitting simulations. We used a similar set of physical schemes, except that deep convection was switched off, to EURO-CORDEX simulations, in which shallow convection was tied to deep convection parameterization. This enables us to assess the added value of explicitly resolving deep convection. However, we keep in mind that the horizontal resolution is insufficient to resolve shallow convection explicitly, and that this issue deserves separate investigations.

L124: Maybe you mean "same" instead of "similar"? "Similar" doesn't mean "identical", but "same" does.

Response: We agree and will replace "similar" by "same" in line 124.

L136: Unit of g is m s-2.

Response: We thank you for pointing out this error.

L148-163: The authors could consider making these lines into a separate Section 2.3 for the data sets? If they don't want to, that's also OK.

Response: It is reasonable to dedicate a separated section describing the data. We will make changed to it.

L178-180: Do these biases refer to the bias over the whole box against SAFRAN? If so, the numbers don't agree with my calculations based on the insets in the panels of Fig. 2. please check.

Response: Those numbers refer to the bias of simulations against all stations within the Cévennes box from in situ observations, not SAFRAN.

L203-205: Are these 23 stations for the time period in 3 (h) or 3(j)?

Response: We compare simulations with 23 stations from Figure 3-j. Since the results from 2 different periods of COMEPHORE are quite similar, we will remove the one with a shorter period to avoid confusion.

L220: Instead of "we model the Clausius-Clapeyron relation ...", it would be more correct to say "we investigate if the temperature-precipitation scaling follows the Clausius-Clapeyron relation in observations and models", or similar.

Response: We will change the sentence in line 220-221 to "*In this section, we model the relation between extreme precipitation and daily mean surface temperature, which is theoretically reflected by the Clausius-Clapeyron equation, by a simple non-parametric scaling method described in section 2.2.*"

L230: The EUR-11 model can't be expected to have similar intensities as the point-scale observations, simply because you're comparing at different scales here (see main comment 2). What's important is whether the EUR-11 and CPM have the same scaling rate. Same goes for L243.

Response: The updated scaling analysis shows that the EURO-CORDEX simulations can reproduce scaling rate in case of extreme daily precipitation, while convection-permitting simulations (CPS) tend to slightly overestimate. For extreme sub-daily precipitation, the EURO-CORDEX simulations fail to provide the overall scaling rate, while CPSs show their advantages in capturing convective events.

L235-240: Maybe your super-CC scaling results from the combination of strong moisture convergence in autumn precipitation extremes in SE France (due to onshore moisture advection) and deep convection. These ingredients aren't present simultaneously at other times of the year.

Response: In fact, the deep convection in the autumn in SE France is favoured by the moist and unstable low-level jet from the Mediterranean, and triggered by steep mountain range (the Cévennes). This mechanism is crucial for the development of the Mesoscale Convective System (MCS) leading to extreme precipitation over this area (Ducrocq et al., 2008; Khodayar et al., 2016; Lee et al., 2018; Nuissier et al., 2008). However, the manifestation of the super-CC scaling has been observed in many other areas rather than the western Mediterranean where mechanisms leading to heavy convective precipitation events are different. The underlying theory to clarify this deviation from the Clausius-Clapeyron relation is still controversial. One can explain by the property of convective processes itself which is enhanced by the latent heat released during condensation as we discussed in line 234-235 of the preprint. However, the super-CC can also be explained by statistical effects at the transition-temperature range that convective and large-scale precipitation are combined (Berg and Haerter, 2013; Haerter and Berg, 2009; Molnar et al., 2015), or it can be the combination of both where MCS is embedded within a persistent large-scale frontal system and latent heat release favouring the moist updraft is involved in the MCS (Hatsuzuka et al., 2021). The appreciation of this mechanism is beyond the scope of our study and deserves further thorough investigation.

-L249 (Section 3.4): My understanding is that the analysis in this section is based on wet-events, i.e. days without precipitation are excluded. If this is the case, it would be useful for the reader to know what fraction of days contain wet events and if this differs much between the different simulations.

Response: We only consider wet-events (hourly/daily amount >= 0.1 mm) in our analysis. These wet-events account for 30% in observations dataset, from 40% to 50% in convection-permitting simulations (CPS) and from 50% to 60% in the EURO-CORDEX simulations. This also shows that the CPSs reduce the drizzle problem as stated in Kendon et al., (2012).

-L252: Change “either ... or” to “both ... and”.

Response: We agree.

-Figure 3: There’s no panel (i) after (h), so I think you need to change (j) to (i).
Response: Thank you for noticing this error.

-Figure 3: What does the yellow colour over Italy represent? If this is simply an area of no data, then it would be good to mask it in white like in Figure 2 (g).

Response: Yes, that is an area with no data. We will mask it.

Please pay attention that the Fig.1 is put at the end of this document.

Reference

Berg, P. and Haerter, J. O.: Unexpected increase in precipitation intensity with temperature — A result of mixing of precipitation types?, Atmos. Res., 119, 56–61, doi:https://doi.org/10.1016/j.atmosres.2011.05.012, 2013.

Ducrocq, V., Nuissier, O., Ricard, D., Lebeaupin, C. and Thouvenin, T.: A numerical study of three catastrophic precipitating events over southern France. II: Mesoscale triggering and stationarity factors, Q. J. R. Meteorol. Soc., 134(630), 131–145, doi:10.1002/qj.199, 2008.

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., 1–15, 2019.

Haerter, J. O. and Berg, P.: Unexpected rise in extreme precipitation caused by a shift in rain type?, Nat. Geosci., 2(6), 372–373, doi:10.1038/ngeo523, 2009.

Hardwick Jones, R., Westra, S. and Sharma, A.: Observed relationships between extreme sub-daily precipitation, surface temperature, and relative humidity, Geophys. Res. Lett., 37(22), doi:10.1029/2010GL045081, 2010.

Hatsuzuka, D., Sato, T. and Higuchi, Y.: Sharp rises in large-scale, long-duration precipitation extremes with higher temperatures over Japan, npj Clim. Atmos. Sci., 4(1), 29, doi:10.1038/s41612-021-00184-9, 2021.

Kendon, E. J., Roberts, N. M., Senior, C. A. and Roberts, M. J.: Realism of rainfall in a very high-resolution regional climate model, J. Clim., 25(17), 5791–5806, 2012.

Khodayar, S., Fosser, G., Berthou, S., Davolio, S., Drobinski, P., Ducrocq, V., Ferretti, R., Nuret, M., Pichelli, E., Richard, E. and Bock, O.: A seamless weather–climate multi-model intercomparison on the representation of a high impact weather event in the western Mediterranean: HyMeX IOP12, Q. J. R. Meteorol. Soc., 142(S1), 433–452, doi:https://doi.org/10.1002/qj.2700, 2016.

Lee, K. O., Flamant, C., Duffourg, F., Ducrocq, V. and Chaboureau, J. P.: Impact of upstream moisture structure on a back-building convective precipitation system in south-eastern France during HyMeX IOP13, Atmos. Chem. Phys., 18(23), 16845–16862, doi:10.5194/acp-18-16845-2018, 2018.

Molnar, P., Fatichi, S., Gaál, L., Szolgay, J. and Burlando, P.: Storm type effects on super Clausius–Clapeyron scaling of intense rainstorm properties with air temperature, Hydrol. Earth Syst. Sci., 19(4), 1753–1766, doi:10.5194/hess-19-1753-2015, 2015.

Nuissier, O., Ducrocq, V., Ricard, D., Lebeaupin, C. and Anquetin, S.: A numerical study of three catastrophic precipitating events over southern France. I: Numerical framework and synoptic ingredients, Q. J. R. Meteorol. Soc., 134(630), 111–130, 2008.

Prein, A. F., Rasmussen, R., Castro, C. L., Dai, A. and Minder, J.: Special issue: Advances in convection-permitting climate modeling, Clim. Dyn., 55(1), 1–2, doi:10.1007/s00382-020-05240-3, 2020.

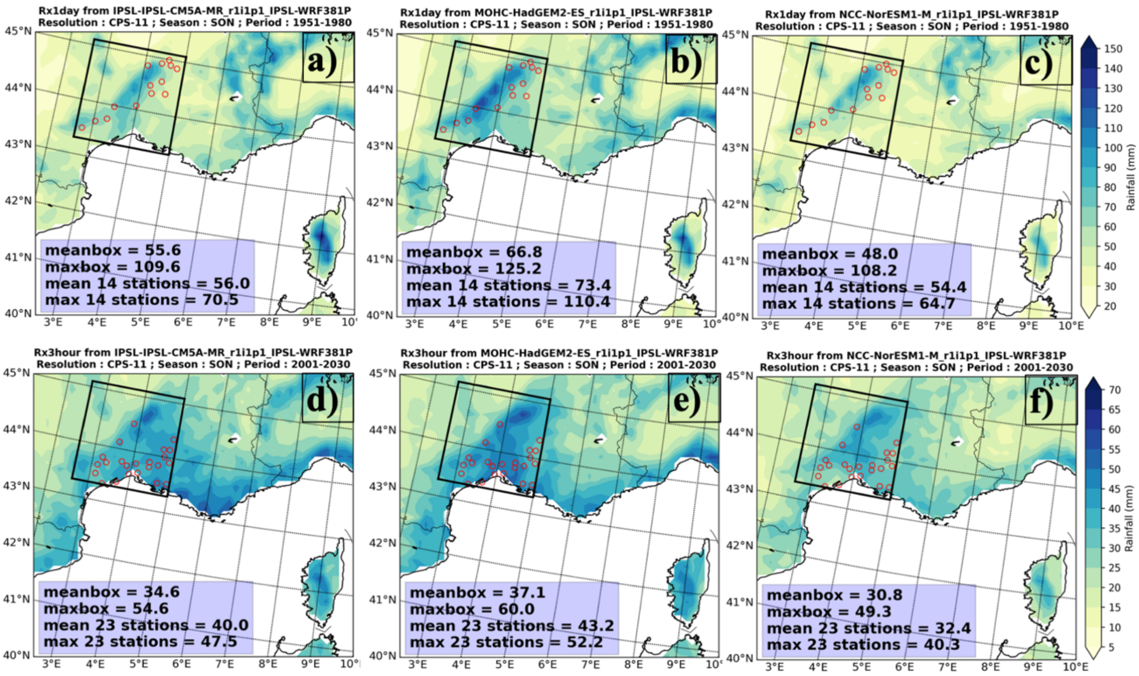

Fig.1: Rx1day (upper panel) and Rx3hour (lower panel) from CPSs upscaled to EUR-11 resolution. From left to right showing downscaling experiments from IPSL-CM5A-MR, HadGEM2-ES and NorESM1-M. Note that the color bar for Rx1day shows a different scale from one for Rx3hour.