

Author reply to RC1 (wcd-2022-11)

General Comments

The presented manuscript assesses differences in the statistics of daily (“heavy”) precipitation between a 25km RCM simulation and two kilometer-resolution CPM simulations. This analysis has been done before, the results are quite robust across model codes and well described in the literature. The presented analysis is novel nevertheless, but resorts to rather sophisticated methods (PCA, composites) and indices (PSI, FSS, wet-day percentiles), making them rather complex and difficult to interpret. Unfortunately, some of their implementation details are not fully fit for purpose (see major issues below).

We thank the reviewer for the very enlightening and thorough comments of her/his report. In the following we provide our answers to the general major and minor comments as well as implemented changes in the manuscript. We hope that these changes and our answers, improve our paper to better meet the journal standards.

Regarding the reviewer’s comments, we acknowledge that the combination of different statistical and computational techniques (PCA, composites, PSI, FSS, wet-day percentiles), have made our manuscript complex and less readable. We will revise our manuscript to improve the readability and to better describe these techniques, their purposes, and results. In particular, we considered the PCA and composites analyses suitable because they are a good means to extract relevant modes of precipitation events. The advantage with respect to e.g., individual case assessment is that they allow to combine several “similar” precipitation situations into one mode. Regarding the PSI and wet-day percentiles, we were interested in studying the capabilities of the PSI, that has the advantage of allowing a flexible definition of precipitation threshold and persistence. We thought, the PSI could bring added value in comparison to other simpler indices. Finally, we chose the FSS as a score allowing for validation of precipitation structures in modelled simulations, where we have had good results in previous studies, e.g., (Caldas-Alvarez et al., 2021).

Also, I am not convinced that the presented results allow supporting the interpretations and conclusions made at the end of the manuscript. In particular, I criticize some of the inferences made from the detected differences (see major issues below), albeit alternative hypothesis are discussed in the literature. Finally, the detected differences are rather small, while the necessary statistical quantification is not provided (E.g., L490 and L303). In my role as a reviewer, I usually abstain from requiring stat. tests, but when an established and robust hypothesis (no differences in daily statistics) is refuted in a manuscript, requiring a robustness assessment seems warranted.

We admit that some of the conclusions were either not well formulated or not optimally underpinned. We will reformulate these statements and improve the reasoning, including alternative hypothesis. See specific answers below.

I have the impression that the manuscript title does not reflect the content of the manuscript precisely enough, as neither “scale-dependency” nor “extreme precipitation processes” are actually assessed in the study.

Regarding the title, we will modify it, following the reviewer’s advice to better address the content and results of our work. Also other reviewers commented on the title so we will correct it to meet the reviewers’ suggestions.

In particular, I don't think that the chosen indices and events qualify as "extreme precipitation", and "extreme precipitation processes" are not considered at all. Meanwhile, the term "scale-dependency" is usually used differently from the presented use case.

We agree that some of the analysed events cannot be considered extreme precipitation events. We have adapted the terminology to "heavy precipitation" which we believe is more suitable. Other papers in the field have denominated heavy precipitation events those above 100 mm/d as is the case for those analysed in the HyMeX consortium (Khodayar et al., 2021). This precipitation amount fits better the range of intensities covered in our paper. Furthermore, the index used for event identification (PSI), relies on spatial percentiles implying that locations over dry areas need lower precipitation totals to be highlighted as "rare" or extreme. Hence, some events could be detected with precipitation below 100 mm/d if they happened over typically dry areas (more on this in the specific answers below). To summarize, we agree that the precipitation ranges considered here are not extreme. We believe a better term for those cases is "heavy precipitation". If events with totals below 100 mm/d are detected and mentioned we will specifically refer to them as moderate precipitation events.

Although this will be developed further later in this document, we agree that there is a problem with our approach to "extreme precipitation processes". Model variables other than precipitation, that have an influence on the simulation of the analysed events are studied in our manuscript using composites. Our aim is to highlight the differences in the model variables affecting precipitation between the two configurations RCM (25 km) vs. CPM (3 km). We agree that this is not the same as studying "heavy precipitation processes". Hence, we will refer to differences in model variables between our RCM and CPM simulations during the heavy precipitation days. Furthermore, we will reformulate our PCA and composite analyses to work only with heavy precipitation days, as opposite to the previous version where "all days" were considered. Instead of using "all days" we will derive our PCA and composites from precipitation days in the upper percentiles of the distribution.

Also, regarding terminology, we adopted the term "scale-dependency" to discuss model differences due to resolution (and their corresponding configurations for our climate model). This term has been used in this same sense in other papers of the field (e.g., Helsen et al., 2020; Glotfelti et al., 2020; Tölle et al., 2020).

Overall, the simulation configurations are not described in sufficient detail and some important aspects are missing. E.g., for the ALPS-3 simulation, the authors mostly refer to Coppola et al.,(2018), which is a MIP overview paper, and thus does provide the necessary detail to ensure scientific reproducibility. For instance, the info that KLIWA-2.8 and ALPS-3 were conducted using two different major releases of the code was not highlighted. I consider this detail is relevant context when suggesting that ALPS-3 is merely a continuation of KLIWA-2.8.

The reviewer is of course correct. The setup for both simulations differs in many details, although there are as well similarities like resolution, overlapping domain and the basic characteristics of CPM simulations with CCLM. Therefore, we applied it for the specific purpose of identifying heavy precipitation events over an extended period (Fig. 6). The reviewer mentions, the KLIWA ensemble is just used in chapter 4. We will clarify the approach in several ways (see also specific comments): 1) Improved description of the simulations, 2.) Better clarify the specific purpose (see answer to major comment 2). 3. Change figure 5 and 4.) indicate the temporal separation between the "KLIWA period" and the "ALP-3 period"

Major Issues

L153: "In spite of these small inconsistencies, we combine both CPM simulations to attain a sufficiently large investigation period for comparison with the RCM simulation and observational data sets" I don't think that the term "small inconsistencies" is justified here. The KLIWA-2.8 and ALP-3 configurations differ in virtually every aspect, apart from their overlapping computational domains and their use of the COSMO code (in two different major releases)

We agree that the formulation is not adequate and will correct that. We will rewrite the model description section and be more specific, as was requested by the reviewer. It was not meant in the sense, that they can be used as one homogeneous data set. In fact, the paper heavily relies on the ALP-3 data and KLIWA2.8 is only used for a specific purpose in section 4 (see reply to the next question).

Also, what defines "sufficiently large" KLIWA-2.8 is 29 years long. I'd consider that sufficient for all of the presented analysis and the qualitative conclusions of the study. Note that KLIWA2.8 is actually only used in Section 4

We agree that the wording of this sentence is incorrect. We correct it in the newer version of the manuscript.

As the reviewer mentions, KLIWA is only used in section 4 and Figs 5 and 6. We decided to include it to have a longer period for the event detection. We believe this is useful because if we only focus on the e.g., top 500 events, in a shorter period this would correspond to one event per week. On the other hand, combining KLIWA and ALP-3 allows us to cover a 45-year period and the TOP 500 events would correspond to one event per month.

On another note, both simulations are at least comparable in the grid resolution and provide an overlapping area, which seems to use sufficient for the purpose. We will clarify this in the text and indicate the separation between to two datasets in Figure 6. We will also replace Figure 5 and include information separately for ALP-3 and KLIWA2.8, which will enable a good comparison of the two datasets

L170 Does the 80th all-day-percentile really characterize "high grid-point intensity"? I think the 80th percentile represents a few mm/days, which is rather typical for a rainy day in Germany

We agree that the 80th all-day percentile corresponds to a few mm/days. We will correct this in the manuscript. The selection of heavy precipitation events comes in the second step of the process where, after having calculated a PSI value for each day, we retain only events in the upper 90th, 95th percentile of the PSI distribution.

L193: I don't understand in Fig 2 the 11th June event is below the 90th percentile of both respective indices, no? This is exactly opposite to the argument made in the text.

Thank you for pointing the mistake out, we correct this part of the analysis (see next question).

L198: I don't think that the analysis supports this claim, rather to opposite. As indicated in Fig.2 the Spearman's rank correlation between PSI and fldsum is 0.98/0.96. Also it is evident that the solid line mostly tracks the dotted line. That is, at lower amplitude, which does not matter for a ranked index. This means that after applying the second threshold (i.e. L236, L301) almost the exact same events are chosen as would have been when using the fldsum. In other words, Fig.2 actually demonstrates that PSI is an unnecessarily complex choice for the presented use-case.

Figure 2 is a bad choice to show the added value of the PSI. We modify this analysis and graphs in the manuscript. Instead, we focus on the rank correlations between *fldsum* and PSI and include a new discussion of the PSI capabilities.

The main points of our new analysis are summarized in the following:

- PSI performs similarly to *fldsum* when we choose a low percentile threshold and zero days for the $RR_{perc_{ij}}$ and d parameters, respectively. As the reviewer points out "the solid line mostly tracks the dotted line" in such a configuration (for instance with $RR_{perc_{ij}} = RR_{80_{ij}}$ and $d = 0$).
- The PSI performs differently to *fldsum* if $RR_{perc_{ij}}$ and d are set to a higher threshold and $d \neq 0$. In fact different events are detected and the rank correlation with *fldsum* is somewhat lower (0.86) with $RR_{perc_{ij}} = RR_{95_{ij}}$ and $d = 2$.
- Different events are detected, because: a) persistence plays a role (an HPE lasting for 2 days will be preferred to a 1-day HPE of the same intensity); b) "rarer" events are preferred (events occurring over a dry areas or a very heavy events, to surpass the 80th, 95th percentile of precipitation over a grid point).
- A different application of the PSI is suggested where instead of using a percentile to define the threshold ($RR_{perc_{ij}}$), an absolute value could be chosen, e.g., 100 mm/d. In this case only grid points with precipitation larger than this value are included in the calculation.
- PSI has larger flexibility for extremes detection than *fldsum* due to the fine-tuning of the aforementioned parameters.

To show these points we are preparing a new table of ranked events detected with *fldsum* and different combination of PSI settings. We will include spatial distributions to illustrate how the PSI detects different events than the *fldsum* which could be more useful to the needs of a potential user of the PSI.

L240: Why are the EOFs computed using the RCM and not ERA-Interim directly? My point is that the first few EOFs of the 500 hPa geopotential need to be almost exactly the same in RCM and driving data. After all, if those EOFs of the RCM simulation would become systematically different than those in the driving data, the RCM approach would become somewhat questionable.

We agree with the reviewer. This was exactly our assumption to derive EOFs of geopotential height at 500 hPa from the RCM simulation. As the reviewer says, the first EOFs will not be substantially different between RCM and the driving model as no relevant differences are expected between these two simulations for geopotential height at 500 hPa. Hence, we decided to proceed with RCM.

L285: The analysis presented in Figure 5 is rather difficult to interpret, since the boxplot parameters (median, quartiles, ...) depict percentiles of a conditional index (i.e., the wet-day precip., $P > 1$ mm/day). Please consult the following study for a thorough discussion of the problems related to deriving percentiles of conditional indices: Schär et al., (2016)

Following the indications of the reviewer and the conclusions of Schär et al., (2016) we will modify the study of percentile differences between RCM and CPM to consider an all-day index. We agree that using a wet-day index, to obtain conclusions for percentile variations (between RCM and CPM or observations) can be misleading. Instead an all-day index guarantees that the same absolute probabilities are used for comparison.

L327: Why are events chosen that entail bad quality/non-existent observations (since over ocean)? To verify a model, I would intuitively choose case studies that have abundant high-quality observations available. Also, I am not fully convinced about the usefulness of the MSWEP-11km product.

We decided to validate our simulations against MSWEP-11km to also profit from their coverage over the Mediterranean. We did not select events that specifically affect the Mediterranean Sea, but events which were heavy and were representative of the area analysed (greater Alpine area) and the two seasons (summer and winter). We considered the coverage of water surfaces an advantage of MSWEP-11km with respect to HYRAS and EOBs.

We also opted for MSWEP-11km because they contain station data from the Climate Prediction Center (CPC) and the Global Precipitation Climatology Centre (GPCC), in addition to remote sensing, e.g., PERSIANN, TRMM. Furthermore, we chose MSWEP-11km because of its good performance in previous evaluations against station data, globally (Beck et al., 2017, 2019; Xiang et al., 2021) and over specific geographies (Du et al., 2022; Peña-Guerrero et al., 2022). In this sense the station data included in MSWEP from the CPC and the GPCC, have a good coverage over Europe, hence we believe this product to be a good choice for our event validation.

These reasons are emphasized in the new version of the manuscript.

“We use the MSWEP product to profit from its high accuracy, shown in previous studies, globally (Beck et al., 2017, 2019; Xiang et al., 2021) as well as in specific geographies (Du et al., 2022; Peña-Guerrero et al., 2022). MSWEP has the advantage of covering sea surfaces and is adequate for precipitation event evaluation because it includes station data from CPC and GPCC.”

L361: The statement needs to be qualified, also w.r.t. internal variability and accuracy of specific humidity obs with radiosondes. In fact, I was surprised to see only a difference of 0.1 — 0.2 g/kg when comparing profiles of a limited-area climate simulation to more-or-less “instantaneous” soundings. I think, my view is corroborated well when considering Fig. S2, instead of the differences. In contrast to the authors, I think these results are actually rather promising.

We will revise our conclusions from this analysis to better assess the performance of the RCM and CPM simulations.

We acknowledge that we overstated the incapability of COSMO-CLM to represent the temperature and humidity profiles and agree with the reviewer that a deviation of 0.1 to 0.2 g/kg is actually a promising result. We will rephrase our conclusions in this paragraph to better provide this information.

Our point is that humidity deviations of less than 1 g/kg in the lower troposphere can have relevant implications for precipitation representation in our RCM and CPM set-ups. This was actually shown in previous publications in our working group (Caldas-Alvarez and Khodayar, 2020; Caldas-Alvarez et al., 2021) for two HPEs and in other sensitivity studies (Honda and Kawano, 2015 and Lee et al., 2018).

Here, we cannot assert whether the observed differences in the humidity profiles between RCM and CPM have an influence on precipitation differences as only 8 cases are considered, and this model sensitivity is not isolated. However, we include this analysis to provide the reader with useful information about the magnitude of the model biases with respect to observational soundings. We also think this comparison against observations is useful since later in the paper humidity and temperature differences between RCM and CPM will be discussed.

To conclude, we will revise this analysis to provide a fairer description of the models' performance and to provide an estimation of the magnitude of the model biases with respect to observations.

L384ff: I am not completely sure if the chosen procedure is appropriate, but honestly, I do not fully understand why it has been chosen in the first place. First of all, why would EOF1 be associated with “heavy precipitation”? I thought that EOF1 portrays the mode with the largest variance, right? That is (by definition) rather unlikely a percentile at the tail of the precip. Distribution, no? Second, EOFs seems rather complicated provided that the results (Fig. 9) exert a very similar pattern as much simpler indices, e.g., the standard deviation (cdo timstd).

We totally agree with the reviewer. We are correcting this in the revised version of the manuscript. The error is in one of our departing assumptions. In the first version of the manuscript we used “PCA as described in Sect. 2.3.2 with daily precipitation in the period 2000-2015 from RCM and CPM.” As the reviewer points out, the leading mode of precipitation is the most common pattern, not heavy precipitation.

To avoid this issue, we correct our approach to obtain the PCs from heavy precipitation days exclusively. We filter, by means of the PSI, days with heavy precipitation in the period (above the upper 90th percentile) and obtain the PCs are obtained from that sample. This is on-going work that will be presented in the new version of the manuscript.

L391: Why for the day prior? I thought the authors suggest the detected heavy precipitation events to be primarily associated with synoptic systems.

We are interested, foremost, in differences in the ground and local conditions (soil temperature and fluxes, cape, surface winds) between RCM (25 km) and CPM (3 km) that could explain the final differences in precipitation amount (in the composites). Selecting the day prior, we try to observe those differences when precipitation has not yet started, or when precipitation is not yet heavy. This is especially relevant for moisture transport that builds-up on the hours prior to precipitation initiation. Our goal is to find out whether changes due to model resolution exist in the pre-conditioning of heavy precipitation are related to precipitation differences.

As the reviewer points out differences in these variables will be most impacting for cases with a weak synoptic forcing. This is more typical of summer events as in winter events.

We have tested both approaches and could not find substantial differences, hence we would continue showing composites of the day prior to profit from the advantage of understanding model differences during the moisture build up.

L398/399: If the first three leading EOFs only explain 39% of the variance. Is that analysis really an appropriate tool? Maybe I do not understand precisely what the authors are trying to achieve.

We expect to explain a larger amount of precipitation variance with the new implementation of the method, i.e., extracting the EOFs from heavy precipitation days exclusively (see previous questions).

We believe the EOFs approach is suitable for our purposes since it helps us encompassing several heavy precipitation days into the single EOFs. By doing so, we can study model differences on a lower number of precipitation spatial patterns.

L415ff: How come? Why not a parameter choice, or a time-step sensitivity in the cloud microphysics parameterization, or the nesting strategy ...

We agree that our presenting of conclusions at this part has been too categorical. We cannot, and should not, conclude that precipitation differences between RCM and CPM being are due to

differences in vertical wind speeds and more triggering of convective cells only. The dominating factors for the differences are difficult to disentangle provided the large number of differences in the modelling configurations, e.g., parameter choice, schemes, etc. However, we believe the differences in vertical wind speeds are important to explain these differences, based on previous publications in the field (Langhans et al., 2012; Barthlott and Hoose, 2015)

We have adapted the wording in the same paragraph to include this remark.

“The findings based on the main modes of precipitation variance, for which EOF-1 is shown as an example, can be summarized as follows: (a) CPM displays larger precipitation than RCM over the mountains for all assessed EOFs and winter and summer seasons. Resolution differences in dynamic processes, e.g., increased vertical wind speeds, larger triggering of convective cells (Langhans et al., 2012; Barthlott and Hoose, 2015) could play an important role in these differences, invigorating the precipitating systems for this mode especially since EOF-1 has a marked orographic pattern.”

L462: No. The statement is conditioned on precipitation being formed the same way in both simulations, given the different surface fields as input. However, the point of the study is that one simulation has the convection scheme active, while the other one does not.

We agree that our conclusions were too categorical (see previous question). While resolution differences in the analysed model variables exist on the day prior to precipitation and are non-negligible, (e.g., surface specific humidity shows differences up to 1.2 mm) we agree that we cannot conclude they are responsible for changes in precipitation. As the reviewer points out, several other changes in the model configurations of RCM and CPM could be responsible for the precipitation differences.

We will then focus in assessing the differences in the model variables without implying causality for precipitation differences. We will likewise provide plausible explanations for the model differences without overstating our findings.

We believe nevertheless that it is relevant for the climate modelling community and potential readers of our paper knowing about how the RCM and CPM show differences in our particular simulations with CCLM.

Please find below an alternative (often discussed) hypothesis that would yield similar differences as those presented:

The soil in simulations typically dries out during the summer months. It is very probable that in autumn soil-water content in the RCM and CPM slightly differ, since soil conductivity, soil diffusivity, evaporation and infiltration rate were not calibrated to yield the exact same soil state. The differences in soil-water content will yield differences in the partitioning of the surface latent and sensible heat fluxes (as observed), and ultimately slight differences in surface specific humidity. The signal appears in the composites of all EOFs because it is already there in the mean.

We agree that it was not precisely enough formulated and will provide a more detailed discussion, including the effects of the small-scale heterogeneities and the differences between parametrized and not parametrized precipitation. See also answers to the point above. We will consider this explanation in our new version of the manuscript to explain soil moisture and surface heat fluxes differences.

L465: Where does conclusion (c) come from? Maybe a paragraph went missing?

Conclusion c) comes from the study of the remainder EOFs. These are provided in the supplementary material due to the extension of the paper.

However, since the EOFs will be processed and obtain again, using only heavy precipitation days, this paragraph will be analysed again in depth.

L479: No. (i) scale-dependency is something different. (ii) The considered indices and case studies are not “extreme”. (iii) The manuscript is not assessing how “thermodynamical processes” influence precip. You are looking at composites. (iv) Also consider that “thermodynamical processes” is a rather uncommon term.

We will work on the wording of this paragraph to convey the concerns pointed out here. Regarding the term “extreme” precipitation, this will be changed for “heavy” following the new approach for EOF derivation explained before (see previous questions). Likewise, we will not talk about thermodynamic processes as the reviewer suggest, but rather the specific model variables will be mentioned. Finally, scale-dependency refers to modelling studies of simulations with different resolutions. This term has been used in previous literature of the like, e.g., field (e.g., Helsen et al., 2020; Glotfelt et al., 2020; Tölle et al., 2020).

L481: I disagree. The analysis also allows other conclusions than portrayed (see above).

We have corrected this in the new version of the manuscript (see above).

L487: It might be worthwhile to note that only daily statistics are considered. In the European summer, a large share of heavy precipitation relates to diurnal convection (see Ban, Kendon). These events only show up when considering hourly precipitation statistics.

We agree with the reviewer and specify that daily precipitation statistics are used, to avoid confusion with sub-daily temporal scales. The paragraph now reads:

“Using daily precipitation data we find that summer events are associated to either frontal convection on the western sector of elongated upper-level troughs and evolved cut-off lows (EOFs 1, 3 and 4), or due to winter-like synoptic patterns of stationary fronts over central Europe or strong zonal flows (EOFs 2 and 5). Five PCs are sufficient to explain the major part of the natural variability of summer cases”

L495/L498: I do not agree with the term “explained by” (see above).

We agree with the reviewer that resolution differences in the dynamic processes describe do not “explain” alone the precipitation differences. As stated, before we will change the text to better convey this information. The changes in the number of convective cells triggered or the vertical wind speeds are two possible precursors that we consider relevant but we agree that we cannot states these “explain” the changes in precipitation over the mountain.

Minor Issues

Introduction: The introduction discusses heavy precipitation and the CPM approach. However, context on the employed approach is not provided. E.g., why are “weather types and PCA” insightful and useful tools for the proposed questions?

We will include this information in the introduction of the manuscript. We believe our approach using weather types and PCA is useful because it helps us extract the predominating spatial distributions of atmospheric modelling variables, e.g., 500 hPa geopotential and precipitation. By doing so we can comprehend several thousand days of decadal data into a few statistically differentiated groups where we can observe the modelling differences of our two configurations RCM and CPM.

L80ff: Indeed there is some connection between moisture, moisture flux, instability and precipitation. After all, these aspects have been under investigation for many decades. I think the authors need to be more specific when making their argument. Also, what “moisture excess” do you refer to? The term suddenly appears out of thin air (maybe a part of a sentence got lost while editing?).

Our argument is that it is worthy investigating what the implications of the moisture wet biases in RCM and CPM are for precipitation representation. Our paper touches, to some extent, upon this topic as surface humidity biases, and moisture flux differences in both configurations (RCM and CPM) are analysed.

“Moisture excess” means moisture wet bias, but we agree that this term can be misleading and we change it for moisture wet bias.

L100ff: Discussion of precipitation under-catch and spatial representativeness, in particular for mountainous regions might be worthwhile.

In our manuscript we do not describe the implications for precipitation under-catch so we believe it would be better not to introduce this topic in the manuscript. Regarding spatial representativeness over mountain terrain we will briefly introduce the implications and differences between RCM and CPM.

L188: I am skeptic that this dataset qualifies as an “independent” source for validation. I didn’t know it before, but the description reads like it is mostly based on other models.

We are not sure, what the reviewer means here. The paragraph described the effect of different parameters in the PSI calculation on the results. The examples are taken from the HYRAS gridded observation and are not meant as a model evaluation.

L136: “nominal resolution” Do you mean grid spacing?

Correct. We meant the grid spacing here and will change the phrasing

L143: It might be worthwhile to note that KLIWA-2.8 is embedded into three nests with 0.44 °, 0.0625 °, and 0.025° grid spacing respectively, as outlined in Hackenbruch et al., (2016).

We will describe the nesting strategy better in the restructured chapter 2.2

L146: Could you confirm that ALP-3 was directly nested into ERA-Interim? I am skeptic because (i) the use of intermediate nests wasn’t mentioned for the KLIWA-2.8 simulation either, and (ii) the CCLM-5-0-9-KIT contribution outlined in Coppola et al. (2018) specifically mentions an intermediate nest.

The description in Coppola et al. (2018) is correct. The ALP-3 km simulation was nested within the EUR-22 simulation used here. We will clarify that better in Chapter 2.2, which will be rewritten anyway

L155: Don’t you disregard the lateral relaxation zone + additional margin for spinup. Btw: What upper boundary condition is used?

We excluded the lateral relaxation zones and the additional margin for the spin-up of 23 grid boxes or about 70 km on each boundary for ALP-3 and 11 grid boxes for the EUR-22 domain (~275 km). CCLM uses Raleigh-damping at the upper boundary above 11.000 m which is the default setup for a European domain. We will include the information in Chapter 2.2

L173: For d=2, PSI considers the sum of 3 days, right?

Yes.

L300ff: I am not familiar with the analysis presented in Figure 6. I guess its main purpose is a visualization of the spearman rank correlation, and the mentioned “clustering”. Could you elaborate on the argument made using that figure? For now, the text mainly refers to the correlation numbers (0.94, 0.41, 0.48).

Yes, a dot diagram is used to represent a series of events in a time span. The figure aids visualizing which events are represented in the model or found in the observations.

L304: Please define “hit rate”.

We include a brief definition.

L321: Define “percentage of affected area”

Ok.

L331: Define “ECOs”

Is defined in line 260.

“Elongated Cut-Off (ECO)”

L348: Why not simply write specific humidity and temperature in the figure labels? Same for the remaining figures.

We will incorporate this suggestion in the labels.

L351: How is “spread” calculated?

It is the standard deviation of the differences.

Figure 5: “The kernel density at each precipitation intensity is shown by the shaded areas.” This shaded area is rather confusing, since the corresponding scale is missing. Also, shouldn't the area be the same size across all presented datasets (i.e., sum up to 1), or is it normalized to one dataset?

It is normalized to each data set.

What does “maximum grid point precipitation” describe? Is that the highest value ever encountered a any grid point in the analysis domain? If yes, does this mean that the data is pooled before computing the boxplot?

It is the maximum grid point precipitation of the analysis domain, represented by each simulation or observations) for the full period 1971 to 2015.

To derive the box plots, we use all daily grid point precipitation data (observations and simulations).

Table 3: Define “coverage”.

Coverage is the percentage of affected area above the 80th percentile. We include this information in the text.

Fig. 10: Why does the term “heavy precipitation” suddenly appear again? How is it related to EOF-1? Also, at what time of the day is CAPE computed?

We will change the term to heavy precipitation throughout the whole text in the revised version of the manuscript (see above). Moreover, with the new approach for EOF and PC derivation employing

heavy precipitation days, EOF-1 will be the first mode of heavy precipitation variance. CAPE is computed on the day prior to heavy precipitation.

L50: I do not understand what an “energetic low-level” should be.

It has been corrected to “warm and moist” low level.

L54: Why is Khodayar et al., (2021) cited here? I am not convinced it fits the context very well.

We take this reference out.

References

- Barthlott, C. and Hoose, C.: Spatial and temporal variability of clouds and precipitation over Germany: multiscale simulations 705 across the “gray zone”, Atmospheric Chemistry and Physics, 15, 12 361–12 384, <https://doi.org/10.5194/acp-15-12361-2015>, 2015.
- Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., van Dijk, A. I. J. M., Huffman, G. J., Adler, R. F., and Wood, E. F.: Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS, Hydrology and Earth System Sciences, 23, 207–224, 2019.
- Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., and de Roo, A.: MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data, Hydrology and Earth System Sciences, 21, 589–615, 2017.
- Du, Y., Wang, D., Zhu, J., Lin, Z., and Zhong, Y.: Intercomparison of multiple high-resolution precipitation products over China: Climatology and extremes, Atmospheric Research, 278, 106342, 2022.
- Glotfelty, T., Alapaty, K., He, J., Hawbecker, P., Song, X., and Zhang, G.: Studying Scale Dependency of Aerosol–Cloud Interactions Using Multiscale Cloud Formulations, Journal of the Atmospheric Sciences, 77, 3847–3868, 2020.
- Helsen, S., van Lipzig, N. P. M., Demuzere, M., Broucke, S. V., Caluwaerts, S., Cruz, L. D., Troch, R. D., Hamdi, R., Termonia, P., Schaeybroeck, B. V., and Wouters, H.: Consistent scale-dependency of future increases in hourly extreme precipitation in two convection-permitting climate models, Climate Dynamics, 54, 1267–1280, 2019.
- Khodayar, S., Davolio, S., Di Girolamo, P., Lebeaupin Brossier, C., Flaounas, E., Fourrie, N., Lee, K.-O., Ricard, D., Vie, B., Bouttier, F., Caldas-Alvarez, A., and Ducrocq, V.: Overview towards improved understanding of the mechanisms leading to heavy precipitation in the western Mediterranean: lessons learned from HyMeX, Atmos. Chem. Phys., 21, 17051–17078, <https://doi.org/10.5194/acp-21-17051-2021>, 2021.
- Langhans, W., Schmidli, J., and Schär, C.: Mesoscale Impacts of Explicit Numerical Diffusion in a Convection-Permitting Model, Monthly Weather Review, 140, 226–244, <https://doi.org/10.1175/2011mwr3650.1>, 2012.
- Peña-Guerrero, M. D., Umirbekov, A., Tarasova, L., and Müller, D.: Comparing the performance of high-resolution global precipitation products across topographic and climatic gradients of Central Asia, International Journal of Climatology, 42, 5554–5569, 2022.

Tölle, M. H., Schefczyk, L., and Gutjahr, O.: Scale dependency of regional climate modeling of current and future climate extremes in Germany, *Theoretical and Applied Climatology*, 134, 829–848, 2017.

Xiang, Y., Chen, J., Li, L., Peng, T., and Yin, Z.: Evaluation of Eight Global Precipitation Datasets in Hydrological Modeling, *Remote Sensing*, 13, 2831, 2021.