

Anonymous Referee #1:

The authors have come up with a clever way of checking the robustness of scaling estimates. They show the necessity of standardisation of rainfall and given the lack of dew point temperature measurements, using mean monthly dew point temperatures as a covariate is a useful addition to the literature. However, in my mind, some issues still remain.

The authors would like to thank the reviewer again. Our paper has improved as a result of their dedicated effort.

Major comment:

There is one thing bothers me which I hope the authors can test. Figure 9 shows large differences across months. Visser et al (2021) was able to reconcile this through standardisation to a standard normal variable, i.e., not just standardising on the mean, but the variance also. Shouldn't this be tested here?

Thank you for bringing this up. We understand "differences" in this comment to be pointing to our results showing substantially larger scaling rate values in the summer months than the winter months as depicted by the boxplot in Figure 9c? This difference is indeed what we intend to show and is one of the main results of the paper. In the following we explain in detail what we mean and in the end suggest a brief adaptation in the manuscript to improve clarity.

The fact that there are large differences between the scaling rates across months and across stations within individual months is one of the main results of the paper. This highlights the need to estimate scaling rates down to the station/month level. Just as Visser et al. (2021) showed the influence that seasonality can have on the scaling rate at one station (shown in Figure 1 below), we show the seasonality of the scaling rates for each month of the year and for all of the stations in the UCRB. In Fig. 9a and 9b, we show the spatial distribution of the P-T scaling rates which we show for each station and for each month of the year. We used our best performing models for Rx1hr and Rx1day to estimate the scaling rates. Those models corresponded to the exponential regression model using the normalized data over a 3-month temporal window and a 100 km window for Rx1hr and a 50 km window for Rx1day. So, consider that we would like to estimate our scaling rate for Rx1hr at a single station for the month of July. Then, the normalized data for the months June-July-August from the stations within a 100 km radius, from the station of interest, are used to fit the regression model for the case of Rx1hr. This will give us the scaling rate corresponding to Rx1hr, for that one station, and for the month of July. Given these temporal and spatial windows, we only use data that is climatologically similar to begin with, but we also control for the differences by using the normalized data. We then proceed through each station and each month of the year to obtain a unique scaling rate in each case. These are the values in Figs. 9a and 9b, for the scaling rates of Rx1hr and Rx1day, respectively. Then, Fig. 9c is a boxplot showing the distribution of the scaling rates across the spatial domain

for each month of the year. So, the gray box and whiskers for January shows the distribution of the scaling rates corresponding to the stations in the left-most subplot of Fig. 9a (i.e., Rx1hr). Similarly, the white box and whiskers for January shows the distribution of the scaling rates corresponding to the stations in the left-most subplot of Fig. 9b (i.e., Rx1day). The large differences from month to month is simply showing that the data tells us that the scaling rates in the summer months, on average, are substantially greater than in the winter months. The large differences between the Rx1hr and Rx1day scaling rates for the month of June tells us that hourly extreme precipitation rates scale more strongly with dew point temperature than daily extreme precipitation. In Figure 9, we plot the scaling rates which we obtain from the slopes of our exponential regression. This can be thought of as analogous to the slopes of the red and blue dashed lines found in Figure 7 from Visser et al. (2021), which we show below in our Figure 1 in this document. Perhaps there is some confusion about the vertical distance between the red and blue dashed lines in Figures 1a and 1b (corresponding to Fig. 7a and 7b from Visser et al., 2021), and the slopes of those lines. Since we are, in fact, dealing with the slopes of the lines, Visser et al. (2021) actually found the difference between the slopes to increase, and not decrease, after standardization (compare 11.1% - 7.6% in Fig. 7b to 10.5% - 7.5% in Fig 7a).

However, we can test here whether standardizing the data provides any benefit over using the normalized data that we advocate for using in our paper. We therefore, went through the process of estimating scaling rates using our normalized data and separately using standardized values (i.e., z-scores of dew point temperature and Rx1day precipitation). We can first illustrate, with an example case, the model fits when using either normalized or standardized data. The example case corresponds to the orange station shown in Figure 2 in this document. Since we will compare using normalized versus standardized values to estimate Rx1day scaling rates, we pool data from stations up to 50 km away and using a 3-month window. The stations within 50 km are shown in magenta in Figure 2 below. Figure 3, below in this document, shows the relationship between the normalized and standardized data for these set of stations for all of the months in the year. There is a correlation coefficient between normalized and standardized dew point temperature of 0.98. Likewise, the correlation coefficient between normalized and standardized Rx1day is also 0.98. For these stations, the normalized values explain 96% of the variance of the standardized values, showing that not much additional information is provided by the standardization. Next, in Figure 4 below in this document, we show the exponential regression fit to the normalized data for the winter months (Dec-Feb) and the summer months (Jun-Aug). The blue and red curves, respectively, show the model fits for the winter and summer. In Figure 5 below in this document, the same data is used except that it has been standardized instead of normalized. Again, the blue and red curves respectively show the model fits for the winter and summer. One of the main challenges when using standardized data is that the results are not as easily interpretable as when using the normalized data. We obtain a model fit which gives us changes in terms of standard deviations from the mean for both our predictor and our predictand. So, for 1 standard deviation above the mean dew point temperature, we might expect something like 0.3

standard deviations above the mean precipitation, for example. This is not intuitively understood, and not easily comparable to something like C-C, which is provided as %/°C. For a given scaling rate using standardized data, we can produce a set of predictions and convert both the predictors and the predicted values to normalized data space to give us a scaling rate to compare with our methodology from the paper. We then proceed with obtaining scaling rate estimates for all of the stations and for all of the months of the year using either normalized or standardized data. Figure 6, below in this document, compares these scaling rates obtained using normalized data versus standardized data. Fig. 6b is the same as Fig. 9b from the paper, while the white boxes and whiskers in Figs. 6c and 9c are the same. These both show the scaling rates when using normalized data. Fig. 6a and the gray boxes and whiskers show the associated scaling rates if we had used standardized data instead for Rx1day. We find very similar results when using either normalized or standardized data. However, the normalized data is better suited to provide us with scaling rates that are easily interpretable. In the case of the UCRB, we do not find any advantage in using standardized data instead of normalized data. While it is beyond the scope of our current study, future work can investigate how well standardized data can be used to leverage pooled data in comparison to normalized data. Nevertheless, we have still added this text at line 250 of the resubmitted paper: "With regards to the fourth model, we also investigated fitting the exponential model to standardized anomalies (i.e., z-scores, or standard deviations from the mean). Using standardized data in the model produced very similar results to what we found using the normalized data (not shown). However, our proposed approach of using normalized data benefits from having less complexity and provides more easily interpretable scaling rates expressed as % per °C."

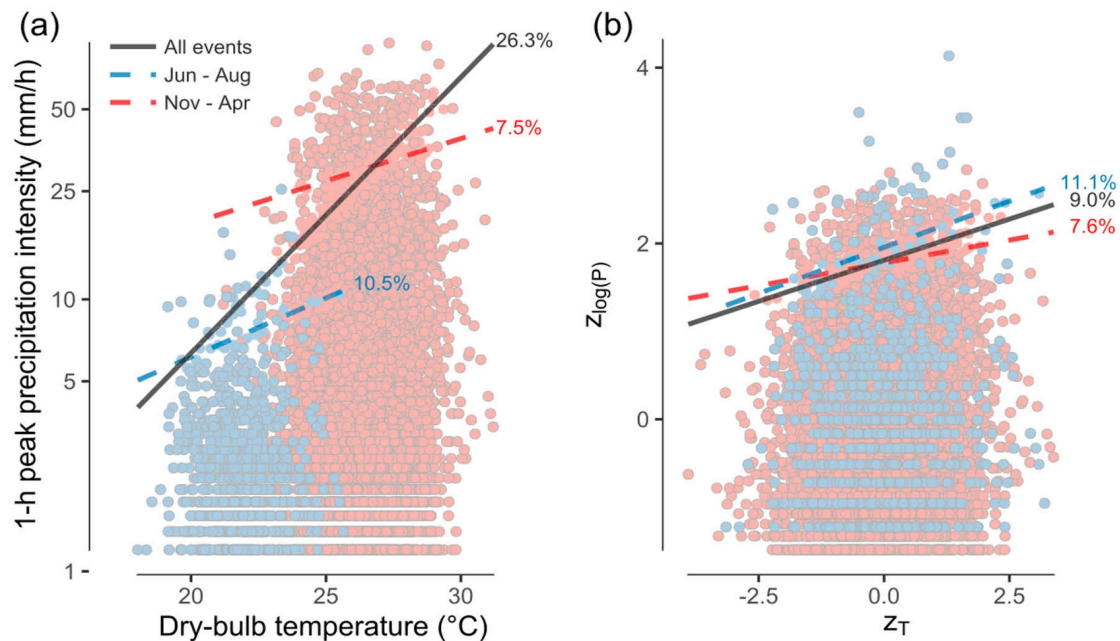


Figure 1: Scaling of 95th-percentile 1-h peak precipitation intensity to dry-bulb temperature for Cairns Airport station (ID: 031011) for (a) nonstandardized and (b) standardized precipitation-temperature pairs. Scaling results are presented for all events and seasonal subsets based on the high-rainfall (November to April) and low-rainfall (June to August) seasons. Figure from Visser et al. (2021).

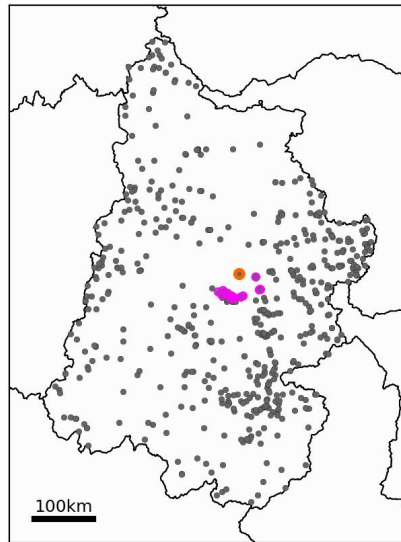


Figure 2: The orange and magenta stations (i.e., these are <50 km from the orange station) are used to estimate the scaling rates for different months of the year for the orange station.

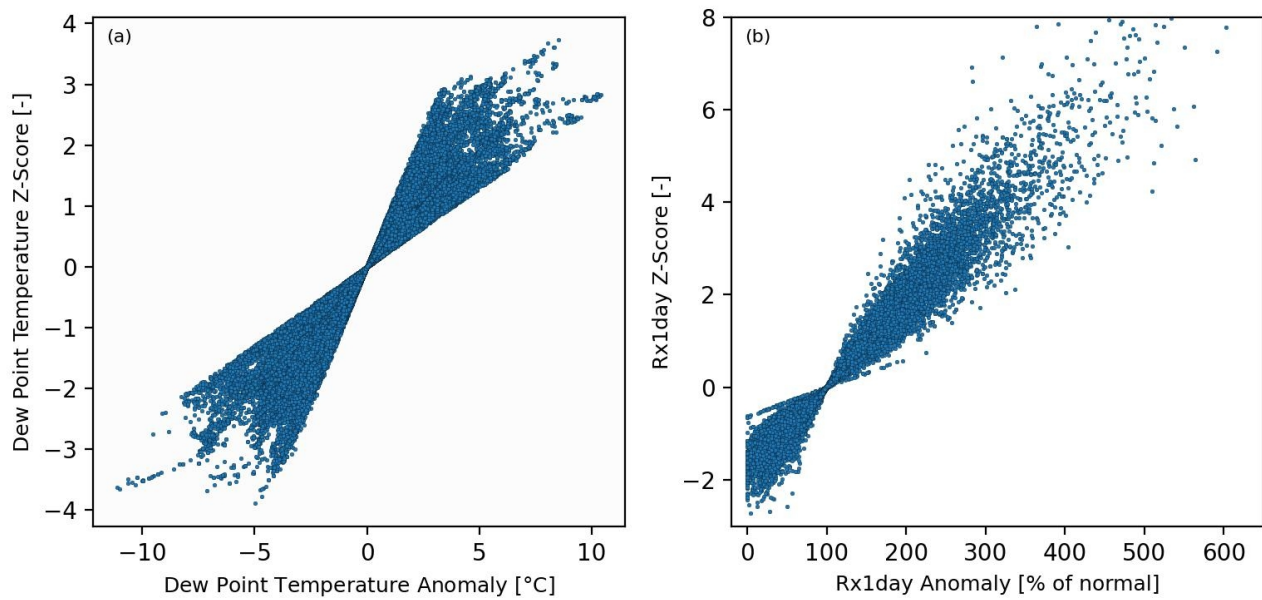


Figure 3: Showing the relationship between the normalized versus the standardized values for the stations indicated by the orange and magenta stations from Figure 2. (a) shows normalized versus standardized dew point temperatures, while (b) shows normalized versus standardized Rx1day values.

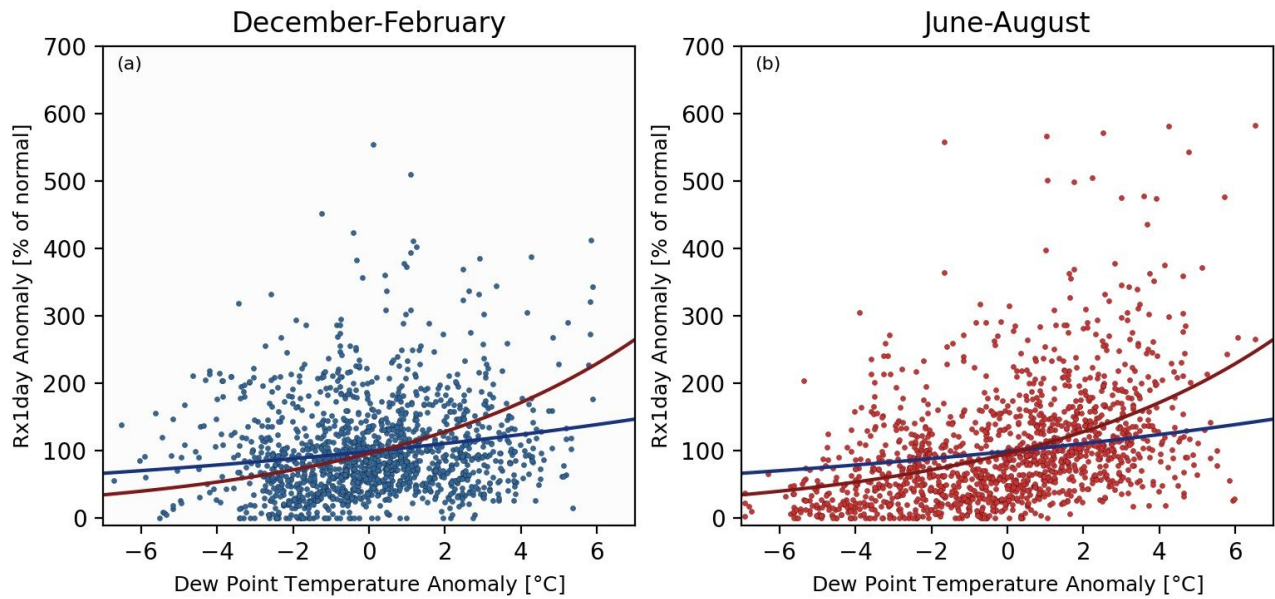


Figure 4: Scaling rates fit to normalized values of dew point temperature and precipitation, as used in our paper. These are the exponential model fits which correspond to the orange station from Fig. 2 in this document and for the months centered about January, plotted in (a), and July, plotted in (b).

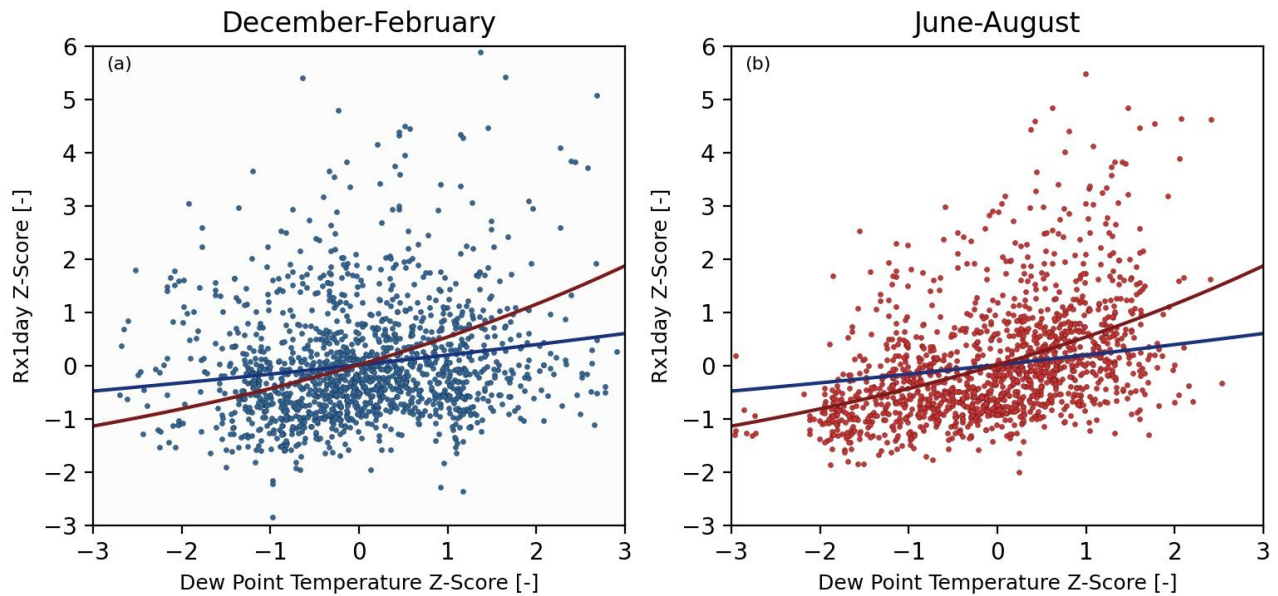


Figure 5: Same as Figure 4 except the scaling rates are fit to standardized values (or Z-scores) of dew point temperature and precipitation.

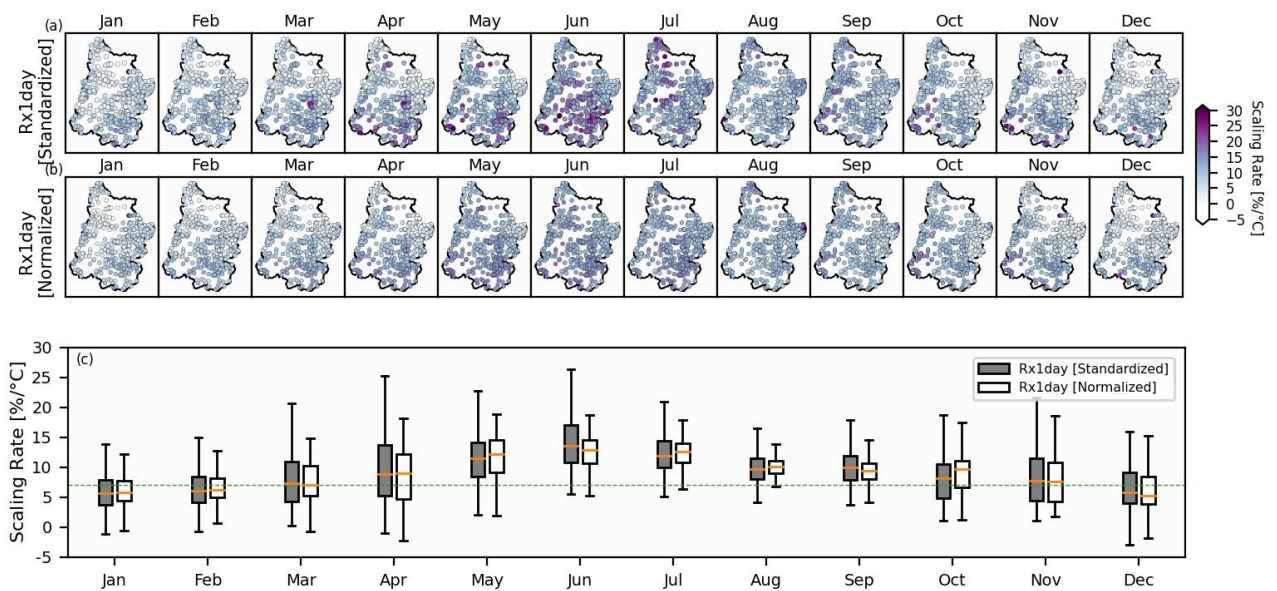


Figure 6: Similar to our Figure 9 from the paper, except we compare estimates of scaling rates using normalized versus standardized values. The top row of (a) plots the scaling rates of Rx1day at each station for each month of the year using standardized values. (b) same as (a) except using normalized values. (c) uses a boxplot to show the distribution of the scaling rates as a function of Rx1day (using standardized values) or Rx1day (using normalized values) along with the month of the year. In (c), each box, median, and whiskers are constructed using the scaling rates from all of the stations in the UCRB corresponding to that month from either (a) or (b). The gray boxplot for July, as an example, is constructed using the scaling rates from the stations in the "Jul" panel in subplot (a).

Shouldn't there be some discussion of the limitations of this study? Negative scaling isn't expected and is generally considered an issue that can be resolved by sampling the temperature just before the event.

Thank you. We had already included some discussion about the limitations of the study, which can be found at line 413 in the resubmitted paper:

"We should note that our methodology has assumed that the scaling rates are stationary over our period of record, which appears to be a good assumption if we compare the results of leave-one-year-out and two-fold cross validation cases for Rx1day. Our focus, here in this paper, has been to illustrate the value of normalizing data in an effort to more accurately estimate P-T scaling rates. However, future research can focus on investigating whether or not scaling rates can be considered stationary. Another issue worth mentioning is the potential impact of the measurement precision of the precipitation data (Ali et al., 2022). In this study, we used data at a measurement precision of 0.1 mm. However, we did additionally try rounding our precipitation data to the nearest 1.0 mm prior to normalization, and this was not found to noticeably impact our results."

We should be clear that the method which we propose in the paper is capable of handling negative scaling. If b from Eq. 6 is less than 1, then we have negative scaling. For example, consider $a=100$, $b=0.95$, and we have a positive dew point temperature anomaly of 3°C . This would give us, $100*0.95^3$, which is 85.7% of normal Rx1hr or Rx1day. See in Figure 9, that in the winter months, some of the stations are found to have scaling rates below zero (seen by some of the whiskers in the box plot from Fig. 9c). Nevertheless, we have added some additional text that explains that the method cannot be reliably used if there is a transition from positive to negative scaling, or vice-versa. See line 420 of the resubmitted paper: "Lastly, while our proposed method is capable of individually modeling either positive and negative scaling rates, it is not capable of handling cases where the normalized data experiences a transition between positive to negative scaling, or vice-versa."

Minor comments:

In the abstract, you say "sub-daily to daily extreme precipitation intensities are expected to increase in a warming climate consistent with the CC relationship" and that "global averages align closely with CC". I don't really agree with this statement including both sub-daily and daily rainfall in it. Referring to this figure from the last IPCC reporting one can only say this confidently for daily rainfall extremes: <http://ipcc.ch/report/ar6/wg1/figures/chapter-8/figure-8-4/> RCMs and observations are (well) above CC.

We have rewritten this, and it is now written as: "Many studies have benchmarked observed extreme precipitation-temperature (P-T) scaling rates against this theoretical value, finding that globally averaged scaling rates of daily extreme precipitation is largely consistent with C-C, while hourly extremes have been observed to scale at rates greater than C-C."

In the data section you use the word "additionally" at least three times by my count. You can omit "additionally". In other sections you write "this paper" or "in this paper". That can be removed also as I believe it is obvious that you are referring to the work pertained within.

Thank you for pointing this out. We have changed the wording in a number of these instances.

Line 60: "to larger sample" -> "to a sample"

Thank you. We have changed this.

Line 61: The novelty of your work would be lost to someone who isn't familiar with Molnar et al (2015). Why not explicitly say what they found: That when pooled the P-T scaling rate was essentially double that when the individual samples were considered? That would highlight very clearly how important your work is!

Thank you for making this point. We have added this text at line 60 of the resubmitted paper: "They found that the scaling rate of the raw, pooled data was nearly double that of the samples which were split by the occurrence of lightning or no-lightning."

Section 3.1/Figure 2: I think the thick line is produced by binning, but it was never explicitly that the line was the result of the binning (the text just says binning was used) and nothing is mentioned in the caption of the figure.

We have now added this to our Figure 2 caption, "The blue line is produced through the binning method, whereby the average of the top 1% of precipitation rates are computed for each dew point temperature bin."

Section 3.2: I don't think what you have is quantile regression - it is an exponential regression using monthly rainfall maxima as the response, if so, just call it that?

Good point. We have renamed our description of our regression model, and we now call it an exponential regression model throughout the paper.

Section 3.4: Not sure why the evaluation section was in the methods and not the results?

We have now moved Section 3.4 to the Results.

Figure 7: Given that the scaling for Rx1hour is much more robust than the Rx1day across spatial and temporal pooling, does this suggest something of the different mechanisms that cause the rainfall and how the temperature that you are sampling represents these? I feel this would be a good point for discussion.

Actually, our results show the opposite. We find the skill scores of our exponential model which uses normalized data to be higher for the predictions of Rx1day. See the highest skill of the green lines in Figs. 7a and 7e from the paper, along with the colors in Figs. 7d and 7h. When comparing the exponential model which uses non-normalized data, the best skill obtained for Rx1day is still better than that for Rx1hr. In the case of the UCRB, the large drop off in skill as depicted by the red dashed-dotted line in Fig. 7e can primarily be attributed to the fact that climatological differences play a stronger role in the case of Rx1day. We have added the text at line 361 of the resubmitted paper: "Please refer again to Figs. 5e-5h, to observe the impact that controlling for climatological differences in space and time can have on model fitting."

Figure A2 isn't explicitly referenced in the text.

Thank you for pointing this out to us. We have now included text referencing Figure A2.

Anonymous Referee #2:

I commend the authors for this deep revision. I think the study strongly benefitted from it and can now recommend the study for publication. I have some residual comments that I think can be addressed with a minor revision. Unless I misunderstood something, I would insist on points 4 and 5 because I think they would further clarify the methodological approach.

Again, I am happy to see such a positive outcome from a major review round.

The authors would like to thank the reviewer again. Our paper has improved as a result of their dedicated effort.

Kind regards

1. Line 44: I suggest to rephrase to “dew point temperature contains direct information on the available moisture in the atmosphere”

Good suggestion. We have implemented this change.

2. Line 56: I suggest to remove this paragraph break

We have done this. Thank you.

3. Line 173 and figure 3 in general: I am not sure I get why the example is done using 4 hour lag. Is it just an example lag that could also have been 6 hours (for example)?

Yes, we had provided an example lag to show that even when the events are separated by multiple hours they cannot be treated as statistically independent. See more in the next point.

4. The outcome other of the analyses in Fig 3 seem to be that we cannot assume independence for the precipitation values nor for the conditional probability of being in the 1% or 0.1% largest events using 4 hours lag. I was expecting to see an estimate of what time lag would be necessary to safely assume independence but I seem not to find that. Why are these analyses done then? Just to confirm that binning would lead to biased estimates?

As stated above, we had provided a lagged example with a lag of greater than 1 hour to show that even then the data cannot be treated as statistically independent. We now include at the end of this paragraph an estimate of when the data can more safely be treated as independent. See at line 191 in the resubmitted paper: "When it comes to conditional probabilities, we find for data

in the UCRB, that the data on average approaches statistical independence for events separated by approximately 12 hours for the top 0.1% and approximately 24 hours for the top 1%."

5. The model in eq. 6 is a least squares linear regression between monthly maxima and T_d . The quantile part comes from the frequency of hourly and daily monthly maxima. I believe it would be more appropriate to call this "least squares regression of monthly maxima" because quantile regression would lead to confusion (it is a different thing). For example quantile regression requires independence and until this point I was very concerned about that because you showed dependence in Fig.3 and stated you used quantile regression.

We would like to be clear, that our exponential model is making use of R_{x1hr} and R_{x1day} values and therefore does not suffer from the same problems of statistical independence as we depicted in Figure 3. We wanted to point out the problem of statistical independence, as shown in Fig. 3, because this can impact any method which does not account for this. We then offer an approach to address it by using monthly maxima. To avoid confusion, we now describe our model throughout the paper as an exponential regression model.

6. Lines 237-241: it seems these lines explain how to predict extreme precipitation based on your method. I wonder whether they should better fit in section 3.3

Thank you for this suggestion. We have moved those lines to the next section.