

Response to Reviewers

We would like to thank the editors and reviewer for their insightful and constructive comments. Because of these comments, the revised version of the manuscript is significantly improved compared to the original one. We have addressed all concerns in this document and most of them in the revised manuscript. Below, we provide a detailed response to each of the reviewers' comments. For convenience, we put the **reviewer comments in black font**, and **author responses in blue**. In our responses, the line numbers, when specified, refer to the track-change version of the revised manuscript.

Comment on hess-2021-236

Anonymous Referee #2

Review of "Impact of correcting sub-daily climate model biases for hydrological studies", Faghih al.

In this paper the authors present a study on bias correction of precipitation and temperature projections at the sub-daily scale, noting that bias correction of these variables is a necessary ingredient to reliable projections of the hydrological response of catchments to climate change. Overall the paper is well structured and written and the topic is relevant and interesting to the readership of this journal.

Thanks for the kind and constructive comments.

General comments:

While the research that is developed is as said well developed and interesting, I was perhaps a little disappointed in the discussion of the results, which although they highlight to some extent the findings, their could have been a little more depth in the exploration of the results, and how these relate to the particular hydro-meteorological processes in the set of catchments considered. The MOPEX dataset is a well-established dataset, and has been studied widely, which also means there is a lot of research to benefit from to support a more in-depth discussion. The main explanatory variable that the authors propose is catchment size, but I could imagine there may well be other explanatory variables. The catchments selected are in a relatively concentrated geographic area, but I would also expect that there

may be quite different dominating hydrological processes. I would for example assume that some of the catchments around the Southern great lakes are dominated by spring snowmelt and perhaps convective storms in the summer, while the more coastal catchments east of the Appalachians may well be dominated by frontal precipitation from mid-latitude storm depressions. I myself am not an expert in the climate nor the hydrology of the Eastern USA, but I could imagine that the relative importance of these different hydro-meteorological processes is relevant to the bias correction of precipitation and/or temperature.

We have made several changes to the manuscript following your comments. In particular, several elements have been added to the discussion. Most of the changes suggested above are discussed in more detail in our answer to the detailed comments below.

It is also clear in the diurnal precipitation patterns, that JJA rainfall is largely convective and thus falls in the afternoon/evening, while in the remainder of the seasons it is perhaps more frontal in nature. How the different types of rainfall that are dominant in a catchment and how this changes may also be important to biases and their correction. Resolution of convection is raised in the introduction, but not discussed much further.

Over the region that is covered in this work, you are correct that a significant part of the JJA rainfall is indeed convective. With respect to bias correction, there are two issues at play: the ability of the climate model at accurately representing convective precipitation (and other physical processes), and secondly the ability of the bias correction method to correct for those deficiencies. Modern bias corrections methods are extremely powerful and can map any distribution onto another. For a pure *model output statistic* point of view, it does not matter whether or not the precipitation is convective, orographic or synoptic scale, if it is biased compared to another distribution, it can be corrected. It is therefore quite easy for example to bias correct Tokyo precipitation data to perfectly match London's precipitation. The question is whether or not it should be done. This is more difficult to answer as it relates to the origin of the biases. If the biases come from bad model physics it is harder to justify bias correction than if they come from an insufficient representation of the topography. We chose not to address these issues in the original manuscript since while they are very relevant to climate change impact studies, they are a lot less relevant to the very narrow problem we are addressing in this paper, which is simply whether or not sub-daily bias correction is relevant.

We have added additional comments below (first detailed comment L.52) and have decided to modify the discussion to tackle the issue of scale and model physics in the revised version.

I think the discussion could be much strengthened if the author would go beyond catchment size as the explanatory variable. Perhaps it is the case that it is indeed the most significant influence in explaining how important the bias correction is, as well as the bias correction of either variable, but it is worth exploring, and the interest as well as the scientific contribution of the paper would be benefitted by at least some discussion of the different hydrological characteristics of the selected basins. I am sure that there is sufficient other research on the characteristics of the MOPEX dataset to support such a discussion. I would recommend the authors revise the discussion section to provide more depth and insight into the proposed bias correction methods through the lens of the hydroclimatology of the catchments. I feel some findings are not well addresses (see detailed comments) due to this lack of depth from this perspective.

The main points above are discussed in more details below. As a result, we have significantly modified the discussion in the revised version of the paper to incorporate these elements.

Detailed comments:

Line 52: The authors comment that many GCM and ESM have coarse resolutions of up to 100km (order 1 degree). I would think that it would be better to refer to the resolution in degrees (as this is common practice). Also, this is a fast-moving field, and the resolution of such models has decreased gradually and continues to do so (see e.g. discussion on the convergence of GCM and ESM by Bierkens, WRR (<https://doi.org/10.1002/2015WR017173>)). Also, the HTESSEL model that underlies the ERA5 data that the authors use could be considered a coupled ESM/Land Surface Model (LSM), and has a higher resolution. It may be good qualify the comment here, but also to include this in the discussion. Will this gradual decrease in resolution in space and time reduce the need for such bias correction as proposed here?

Model resolution has been modified to degrees in the revised version. GCM/ESM spatial resolution is indeed decreasing, although this is not consistent across all modelling centers. Many modeling centers have kept the same resolution in the CMIP5 and CMIP6 experiment, opting instead for additional complexity in the models (e.g. carbon cycle, surface model). Since computing time is a function of spatial resolution to the fourth power (x,y,z,t), there are physical limits as to how rapidly model resolution

can decrease. Regional climate models have comparatively seen larger decreases in spatial resolution, and especially so for convection-permitting models, which need a horizontal spatial resolution below 2-4km to numerically resolve convection. However, the same computational physical limits still apply to regional climate models, and this increase in spatial resolution has been at the expense of a progressively smaller computational domain. Does a decrease in space and time resolution reduce the need for bias correction? For most impact studies, sadly, our answer to that question is 'no'. Model improvements have been shown to reduce the biases. These improvements come from the increased resolution (better representation of local topography and land surface) and from better physics. However climate models remain an imperfect representation of the real climate system, and the sensitivity of impact models to input data (e.g. precipitation, temperature) will still require some level of post-processing to insure realistic outputs from impact models. The Climex ensemble used in this study has a very high resolution (0.11°) and still, it clearly requires bias correction. Using uncorrected climate model data results in unrealistic streamflow simulations (e.g. Figure 5). However, with better and higher-resolution models, there is hope that post-processing methods will only end up correcting minor model deficiencies, and not correcting bad physics over a given area (e.g. incorrect precipitation annual cycle). Bias-correction methods such as the one used in this paper are powerful and can map any distribution (good or bad) onto another. However, just because we can, does not mean we should. In a perfect world, every climate model should be evaluated over a given region prior to post-processing its outputs, as it is hard to justify correcting a model with bad physics over a region. In the real world, this is not an easy task, and impact modelers are generally not well equipped for this task and typically end up using all available climate models instead of a careful selection of the best-performing ones. To sum it up, we don't think that better climate models will eliminate the need for bias correction, but rather that it will reduce instances of bad usage of bias correction.

We have substantially modified the discussion to include some of those elements. We have added additional references (including the one suggested above).

Line 110: The figure presenting the different catchments could be bit more informative. The size of the circle as well as the colour are now used to indicate how the catchments fall in the different categories. One of these dimensions is redundant. As there are some larger catchments, the division across the three scales the paper focuses on is not clear. One could perhaps use a different symbol (square, circle, triangle) for the category and colour for the size would be more informative. It is also clear from the table

that the sample of small catchments is in fact quite small. How well are these spread over the different hydro-climatologies, or are they concentrated in the more mountainous Appalachians (it is not so clear in the figure)?

The figure was modified accordingly. In the new figure the square, circles, and triangles represent small, medium and large watersheds respectively. As can be seen from the distribution of the squares, the small catchments are distributed across the study area and are not limited to the Appalachians. We would have preferred having a larger sample of small catchments, but it is a limitation of the MOPEX database.

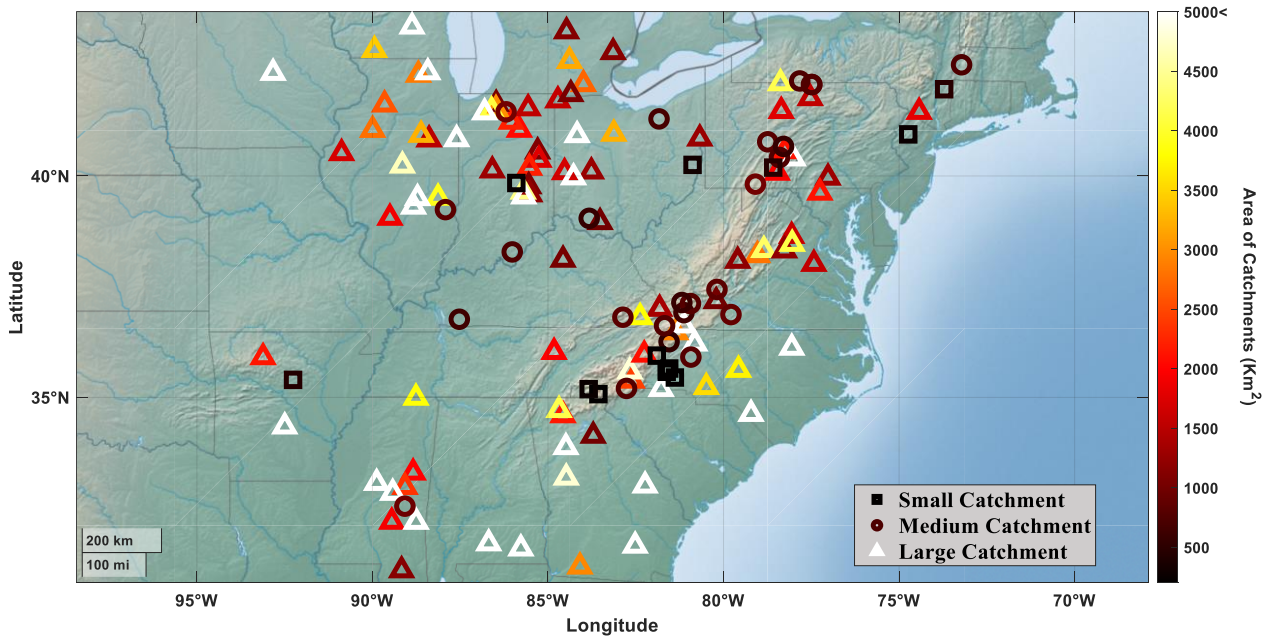


Figure 1 Distribution of watersheds across North Eastern America. Circles and triangles symbols correspond to small, medium and large catchments respectively.

Lines 180-185: The authors present the setup of the calibration the model using SCE-UA and N-S as a performance indicator. I think this is an approach reasonably familiar to the HESS readership, so in some cases the description could be more brief (e.g. one could consider leaving out the equation and simply providing a reference). I am also fine with the fact they do not split the sample, considering that a well

calibrated model is important but it is not the focus of the study. However, that does raise the question on how this was organised for the bias correction approach. Was a split sample used? This is not clearly described (in the previous section). On line 215 mention is made that if the bias correction was applied at the hourly time scale, then this would have fitted the observations exactly (for the mean I would assume). This suggest that a split sample is not used, thus raising the question of how well the corrections applied are a generalisation of the biases. This may warrant some discussion.

A split-sample approach was not used in this paper as our main goal was to investigate whether or not sub-daily bias-correction would improve hydrological simulations on small catchments. Moving to a validation period brings an additional variability which is the stationarity of the biases. Our guess is that sub-daily biases are likely not stationary (because of internal variability), and that the advantages of the sub-daily bias correction method would be somewhat reduced when tested over an independent validation period, as found by Chen et al., (2018) in a comparison study of multivariate vs univariate bias correction methods. This issue has been added to the discussion in the revised version.

Chen, J., et al. (2018). Impacts of correcting the inter-variable correlation of climate model outputs on hydrological modeling. *Journal of Hydrology*, 560, 326-341.

Line 262: Figure 5 is not very easy to read. It is mentioned that the light blue curves overlap – but it is very difficult to identify what is what in the figure. This could well be improved. It is also not fully clear of the average biases, as the lower part of the different lines is obscured. Given the information the authors want to convey, would it perhaps be more appropriate to present in the form of a flow duration curve? This may well not show the full variability, but it does provide information on the average annual spread. I would suggest the authors try in any case to make the figure a little clearer.

The transparency of the blue line has been changed in the revised version. It definitely allows for a better comparison between both curves.

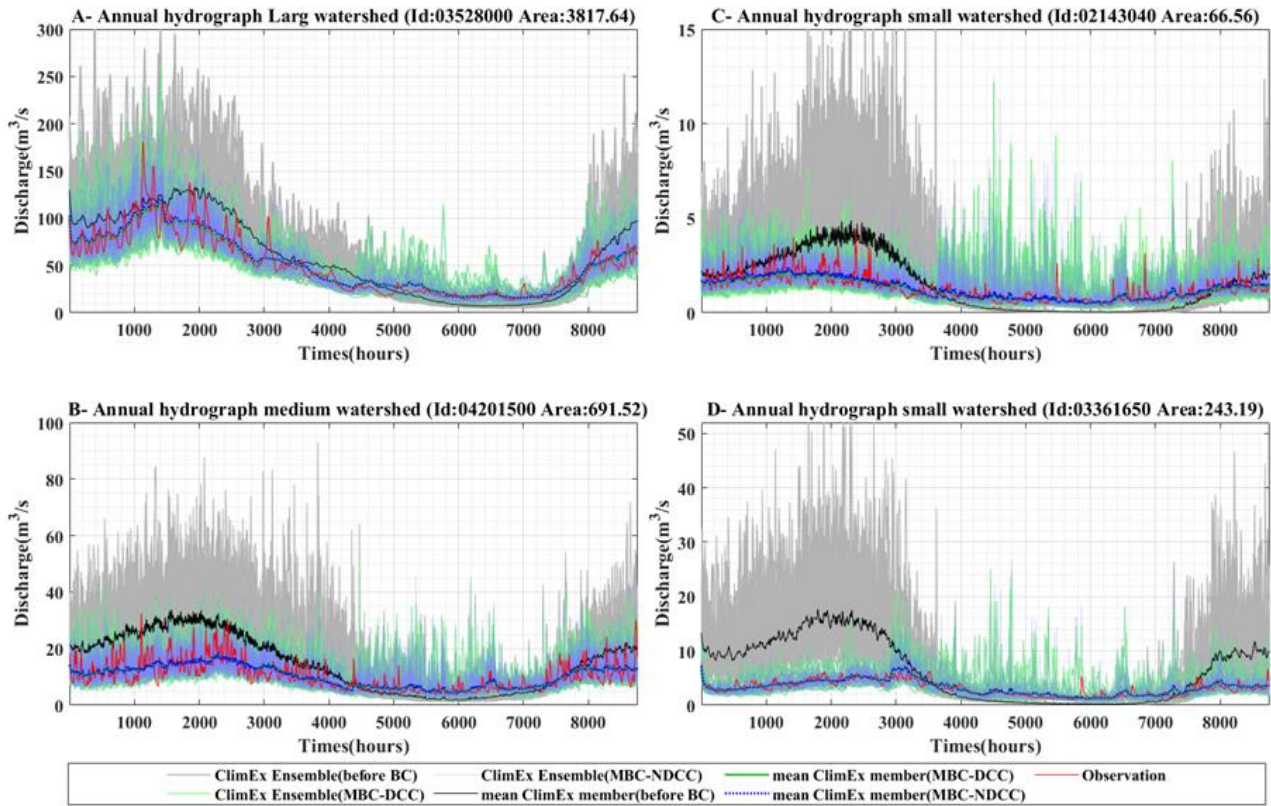


Figure 5 Hydrograph annual cycles for four selected catchments. Catchments A and B are classified as large and medium size respectively. Catchments C and D are classified as small. 0 represents January first at 0h00, and 8760 is December 31st at 24h00.

Line 271-274: I think it is important here to be explicit on the question if the climatic characteristics and the dominant processes that shape the regime (pluvial – nival) of the two catchments compared are indeed similar, thus supporting the claim that the diurnal variability being more apparent in the smaller catchment is indeed attributable to catchment size.

We made sure that both catchments were geographically close and displayed similar climate characteristics. We are therefore confident that the main differences come from the catchment size. This is a good point however, and we have added precisions in the revised manuscript and have modified Figure 6 to show the localization of both catchments (see new Figure 6 below).

Here is a detailed description of the catchment main characteristics. A brief account of the similarities has been given in the modified version.

Large Catchment: Station 03528000 CLINCH RIVER ABOVE TAZEWELL, TN

Drainage area is 1474 sq.mi.

Located in hydrologic unit 06010205 at Latitude: 36:25:30N Longitude: 083:23:54W

Station is in the state of TN (county code 025) and serviced by the TN office.

Channel slope is 6.59 feet/mile

Stream length is 199.00 miles

Mean basin elevation is 2130.00 feet

Storage: 0.02% Lakes: N.A.% Forests: 53.30% Glaciers: N.A.%

Soil infiltration index is 3.89 inches

Annual precipitation is 45.00 inches

Precipitation intensity is 3.00 inches/24 hour (expected on the average once each two years)

January minimum temperature is 28.00 degrees Fahrenheit

Daily and longer averages for 69 water years are acceptable (all years except partial years).

Small Catchment: Station 02143040 JACOB FORK AT RAMSEY, N. C

Drainage area is 25.7 sq.mi.

Located in hydrologic unit 03050102 at Latitude: 35:35:26N Longitude: 081:34:02W

Station is in the state of NC (county code 023) and serviced by the NC office.

Channel slope is 119.70 feet/mile

Stream length is 9.80 miles

Mean basin elevation is 1760.00 feet

Percent of contributing drainage area covered by

Storage: 0.00% Lakes: 0.00% Forests: 90.00% Glaciers: N.A.%

Soil infiltration index is N.A. inches

Annual precipitation is 49.00 inches

Precipitation intensity is 4.00 inches/24 hour (expected on the average once each two years)

January minimum temperature is 32.00 degrees Fahrenheit

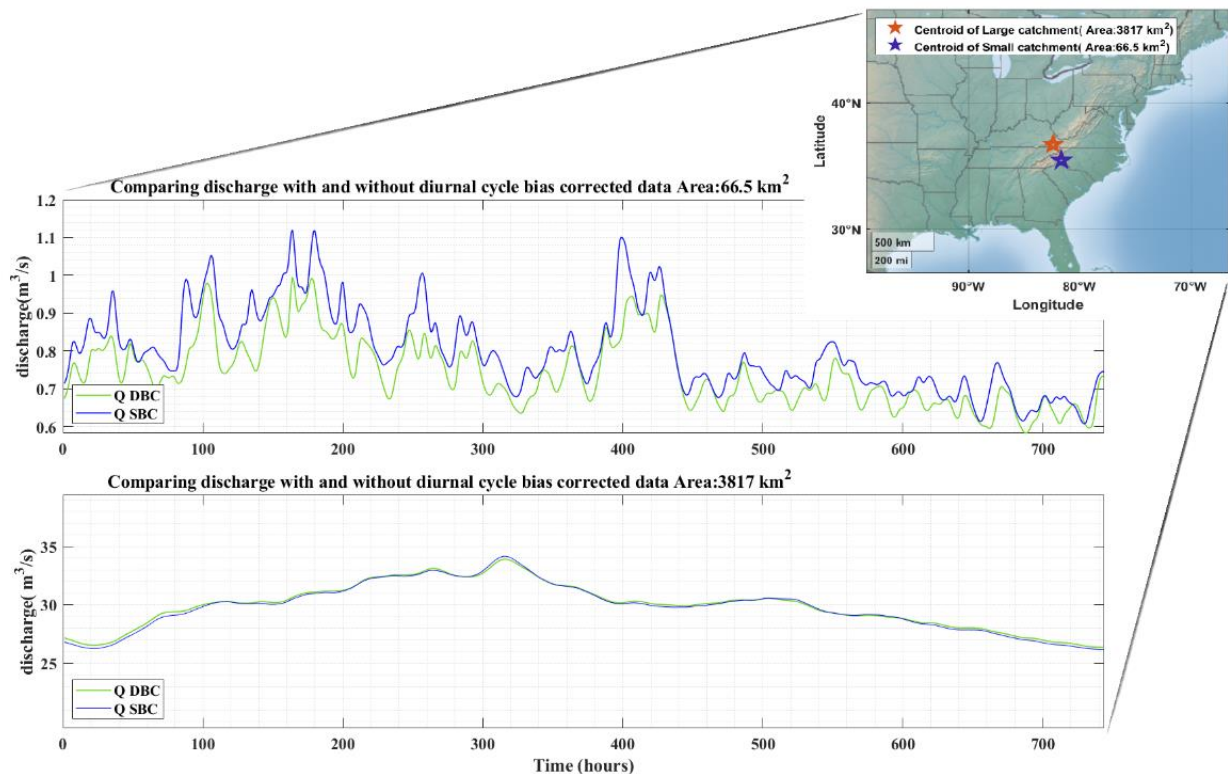


Figure 6 Hydrographs of two sampled catchments for the month of July (744 hours=31 days * 24 hours) (small and large size surface area)

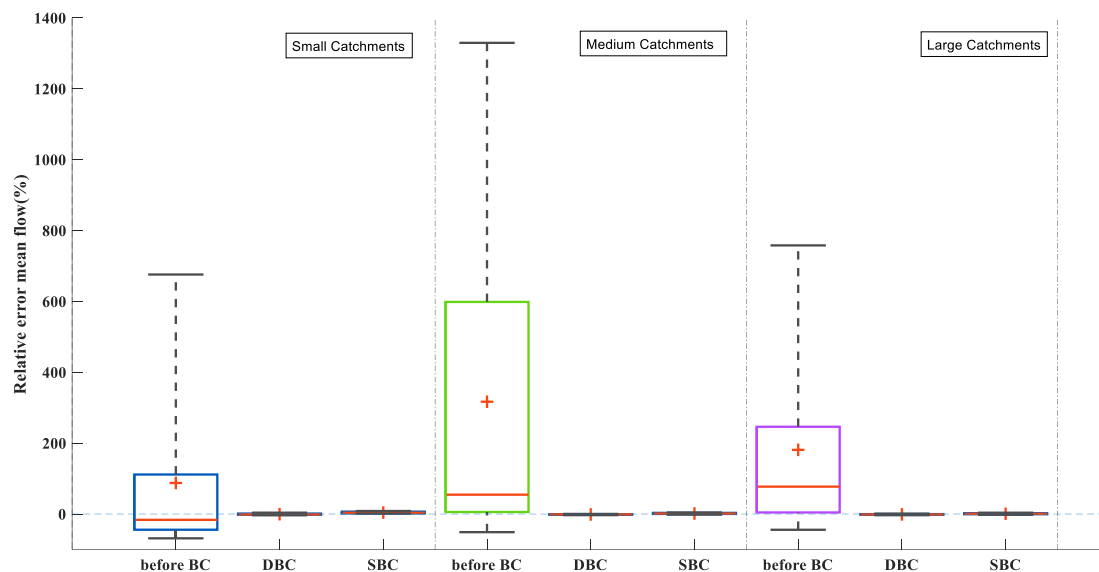
Line 288: Please check the description of the Figure 7. The red crosses appear to be circles. Also, there is a pink/purple cross that is not explained which I assume is the mean of the distribution.

Thanks for the catch. The text was modified accordingly.

Line 297: Figure 7 shows the relative biases of the small, medium and larger catchments after bias correction. The unbiased model results are not shown. Perhaps this was an obvious choice, but it may be valuable to include these, at least in Fig 7, to understand the relative improvement for each of these scales. Also, in Figures 7,8 & 9 it is clear that the diurnal bias correction always results in a more negative bias than the standard bias correction. This is commented on in the text, but the authors do not provide possible explanation. Is the bias corrected precipitation consistently lower; the temperature and consequent evaporation consistently higher? I would assume this is the case, as if it simply a diurnal

redistribution of moisture, then this may not be so consistently seen in the larger catchments. Please comment.

We have not included the boxplots of uncorrected simulations because the biases are up to two order of magnitude larger than for the corrected ones, as can be seen in the Figure below. While these conclusively demonstrate the need for bias correction, it is simply not possible to outline the difference between both bias correction methods if the uncorrected boxplots are shown on the same Figure. We have instead added a sentence in the revised version explaining why the uncorrected model results are not shown.



The reasons as to why the diurnal bias correction results in less flows (we would not say more negatively biases, since Q99 is positively biased for example) is not entirely clear. Since there is less water when the diurnal cycle is corrected, it implies that evapotranspiration is increased. We have tested this on a few catchments (as shown below for one small catchment), and in all cases, this was verified. What we believe is happening is that the climate model diurnal cycle for temperature is flatter than for the observations (see Figure 3 for example). By correcting it, we get higher afternoon temperatures, leading to an increase in potential evapotranspiration (and evapotranspiration), since we are using a temperature

ETP formulation. We have mentioned this in the revised manuscript although we have not added any additional Figures.

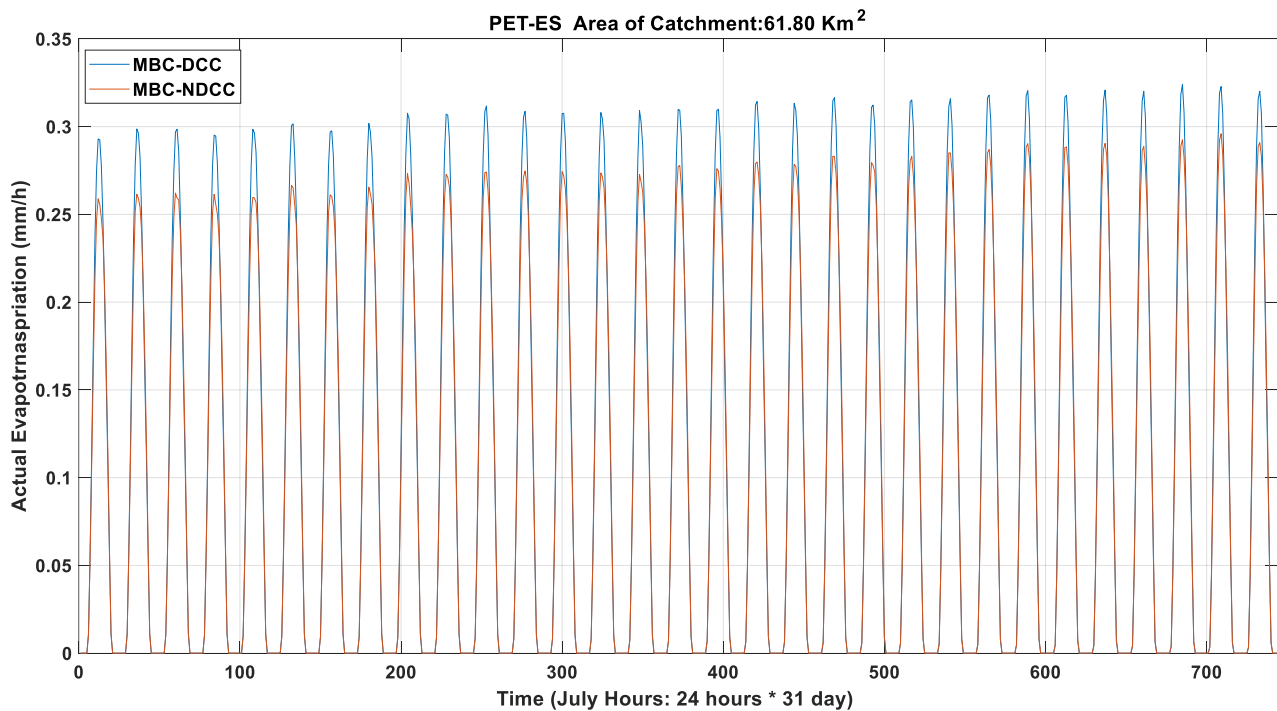
Small Catchment:

Average PET of July-MBC-DCC= 0.1077

Maximum PET of July-MBC-DCC=0.3347

Average PET of July-MBC-NDCC= 0.1018

Maximum PET of July-MBC-NDCC=0.3122



Line 342: I find the discussion on the errors in the tipping bucket somewhat suggestive. I agree that at very small rain rates the time of the tip may not well represent the time of precipitation, but I would hardly think this is a major contribution across all catchments and rainfall intensities in the datasets used here. Is there any research that substantiates this statement?

We agree that this is unlikely to be a major contributor. Still this is a type of error that will average itself out over a longer time period (e.g. daily) and is therefore more relevant to the shorter time steps. We

have modified this section to do a slightly more detailed analysis of potential problems based on the recent work of Segovia-Cardozo et al., (2021), and in the process, we have deemphasized the importance of the time offset mentioned in the original version.

Segovia-Cardozo, D. A., Rodríguez-Sinobas, L., Díez-Herrero, A., Zubezu, S., & Canales-Ide, F. (2021). Understanding the Mechanical Biases of Tipping-Bucket Rain Gauges: A Semi-Analytical Calibration Approach. *Water*, 13(16), 2285.

'Measuring issues related to the use of tipping bucket rain gauges have been reviewed by Segovia-Cardozo et al. (2021). Those issues are an underestimation of total amounts, and especially so for high intensity rainfall and light drizzle, losses from evaporation and non-linear response to rainfall intensity. In addition, at the sub-daily scale, the above may result in small shifts in the actual recording of small precipitation.'

Line 371: A discussion is provided on the diurnal cycle due to evapotranspiration in the smaller catchments, with this being better represented after bias correction. I agree that this is may well be detectable in flow records for small catchments. However, when looking at figure 3 it would appear that there is a diurnal cycle in the temperature data. Indeed, the temperature in JJA has such a cycle but a clear positive bias, which one would expect would increase evapotranspiration further. Yet in Figure 10 there is no diurnal cycle at all in the unbiased simulations. This does seem somewhat incongruent. It may be good to explore this a little further. Overall Figure 10 can be improved in clarity also as it is not very clear.

Your comments led us to understand that we were not clear enough in our description of Figure 10. There are diurnal cycles for both temperature and evapotranspiration on all catchments, independent of size. This suggest that there is also a surface runoff diurnal cycle which results from the above factors on all catchments, irrespective of size. However, this surface runoff cycle can only be seen on catchment which have a reaction time smaller than approximately 12 hours so that the catchment outlet can register differences between day and night processes. This is what we were trying to show with Figure 10. We have added to the Figure description in the revised version of the manuscript and made changes to the Figure as demanded by Reviewer 1.