Hydrol. Earth Syst. Sci.
hess-2020-46

Dear Reviewer #1,
Thank you for your review and the detailed comments. Below please find our point by point response to your suggestions and questions. The Reviewer's comments are in regular font and our response is in bold.

Response to Referee #1
In their study, the authors adapt a general mathematical method that was published by them earlier (2017) that can be used to determine and correct the biases related to the spatial aggregation of modeled, gridded evapotranspiration fields. The method is exemplarily applied for Switzerland, based on the GLEAM evapotranspiration model. I consider the contribution as innovative and as relevant for the field of hydrometeorological modeling and I recommend its publication after the following points were adequately addressed:

We thank the reviewer for his/her interest in this work.

General comments
Is it always that with higher resolution data models give more realistic estimates of ET? In the introduction you mainly address biases caused by rescaling of ET fields, but how does that rely to observations? Is there evidence in literature for the assumption that higher resolution data usually provides more realistic rates? You use GLEAM to prove your concept. But looking at the comparisons of true and estimated biases in Fig. S2 and S3, it seems that your approach does not work well for resolutions smaller than 0.25 (which is the target resolution of GLEAM). So maybe GLEAM is kind of optimized to this resolution and is not too realistic for higher ones? How would you explain the increased scatter between true and estimated biases for the 1/32 and 1/16 resolutions?

First of all, it is important to remember that we are not comparing GLEAM with real-world measurements and therefore we cannot evaluate the realism of GLEAM at any resolution. We are not assuming that higher-resolution data is more realistic; instead, we use the higher-resolution estimates as a benchmark for synthetic experiments that examine how these ET estimates change with aggregation scale. As we note on lines 217-219, we use 500-m ET estimates (derived from GLEAM) as virtual "truth" and then see how these estimates, averaged over a range of larger scales, compare with GLEAM estimates of ET obtained from averages of temperature, net radiation, and soil moisture over those same larger scales.

We further looked into the point raised by the reviewer regarding the increased scatter between true and estimated biases for the 1/32 and 1/16 resolutions plots of figures S1 and S2. We noticed that due to a coding error, equations 10b, 13b, and 14b were not implemented correctly, meaning that the stress factor function was considered nonlinear in the full range of soil moisture and not only when soil moisture was between 0.1 and 0.6.

The stress factor function is nonlinear between volumetric soil moisture values of 0.1 and 0.6 as it is defined in GLEAM, and is equal to 0 or 1 outside this soil moisture range. Therefore the first and second derivatives of ET function with regard to soil moisture are equal to 0 (eq10b, 13b, and 14b). Unfortunately we noticed that this point was overlooked in our original calculations in the code and the stress factor function was mistakenly considered as a nonlinear function for the entire range of soil moisture. We have now corrected this glitch and

verified that the script is handling the 0.1 and 0.6 soil moisture conditions and the corresponding variability of soil moisture in this range correctly. The supplementary figures corresponding to estimated averaging error versus true averaging error for the two days also exhibit much less scatter than before. In fact, with this correction the R^2 of the scatter plot of the 1/32 degree resolution increases to 0.94 on May 31st 2004 and 0.92 on July 21st 2004 after this correction. We will rerun the script and redraw all the figures in the revised manuscript.

After correcting this glitch, the estimated aggregation biases in Figures S1 and S2 were quite close to the one-to-one line for almost all the points, regardless of the resolution. This indicates that our method for predicting the aggregation bias generally works well. At the highest resolutions (smallest grid cells), however, there are a few cells that lie farther from the 1:1 line. These correspond to individual points in which the absolute values of ET are very small (snow-covered or glacierized landscapes), so even small prediction errors can appear as large percentage errors. But because these large percentage prediction errors are small in absolute terms, they mostly disappear when they are aggregated to larger grid cells. Thus the mean averaging error across Switzerland decreases sharply (almost exponentially) as the resolution increases.

Specific comments:
16: I would say that the drivers for droughts and heatwaves are precipitation, radiation, wind, temperature and soil moisture but not ET. Heatwaves occur because of the advection of warm and dry air. Droughts are caused by lacking precipitation. 42: Can you give a rough number (in percent) of typical deviations?

We will correct this statement to: "Due to its feedbacks to large-scale hydrological processes and its impact on atmospheric dynamics, ET is one of the drivers of droughts and heatwaves".

140: Priestley-Taylor was already cited before in L 101.

This citation is directly relevant to the PET formula and we found it is helpful to keep it where the equation is presented.

167-173: You should cite your 2017 paper here again, is cited in the introduction but when I read the equations below a quick link to where they have been derived would be helping; also you should explain shortly the meaning of the variance and covariance terms here. They are only explained in L 246.

We added the "Rouholahnejad Freund and Kirchner, 2017" paper as a reference and edited the explanation for equation 7 as: where $\overline{\mathrm{ET}}$ is the estimate of the true average of the nonlinear ET function over its variable inputs, $\widehat{\mathrm{ET}}$ is the ET function evaluated at its mean inputs, and the derivatives are understood to be evaluated at the mean values of the variables ($\overline{R_n}, \overline{w_w}, \overline{T}$) and multiplied by the corresponding variances and covariances among the finer-resolution input data."

177-179: Eq. 8 is not a derivative

We corrected the corresponding statement to make this point clearer: "For the specific case of the GLEAM model, the ET function is evaluated at its mean inputs ($\widehat{(\mathrm{ET})}$) and these derivatives are derived analytically from the ET function described by Eq. 6, directly yielding the following expressions:"

179-188: Why was the interception term of Eq. 6 been skipped in the derivative calculations?

In GLEAM, interception loss is explicitly modelled according to Gash's analytical model (Gash, 1979; Valente et al., 1997). Following this approach, the volume of water that evaporates from the canopy is estimated as a linear function of the daily rainfall using parameters that describe the canopy cover, canopy storage, and mean rainfall and evaporation rate during saturated canopy conditions.

Because the interception loss in GLEAM is a linear function of amount of rainfall necessary to saturate the canopy, it has negligible effects on the aggregation bias.

221-230: What algorithm was used for averaging?

-These are pure arithmetic averages (sum of values divided by number of values).

271-280: Are there dates where other variables than soil moisture have an increased impact?

The terms have different positive and negative contributions (increasing or decreasing effects on total bias) on the two days, with some of the variance and covariance terms being negative or positive. For example, the Rn and SM covariance term on May 31st 2004 is slightly negative (-0.53) but this same term is slightly positive on July 21st 2004 (0.88).

On most of the days of the year 2004, the soil moisture variance term is the dominant driver of the aggregation bias. However, there are some days in which other factors such as the T and Rn covariance term is the dominant factor (e.g, on days 285 and 297 of the year 2004, the T and Rn covariance term constitutes 74.5 % and 90.2 % of the aggregation bais).

309, 390, 391: The section references seem to be broken.
Thanks for pointing this out. We will revise the section numbers.

Fig. S2a) / S3a), please put the 6 maps into two rows, the color key numbers are hard to read
We will do that.

References: unify format, many DOIs are missing, some are printed as links, some have no preceding "DOI" (pleas stick to HESS typesetting rules); Use en-dash for page ranges instead of simple dash
We will do that.

522: "Uber" -> "Über"
We will revise "Uber" to "Über"

Minor:
15: feedbacks -> feedback 124: please change to "I is interception loss" or "I are interception losses" 367: two times "These biases can" maybe replace by "and"
We will revise these points.