

Earth Syst. Sci. Data Discuss., author comment AC1  
<https://doi.org/10.5194/essd-2021-115-AC1>, 2021  
© Author(s) 2021. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Reply on CC1

Vassilis Aschonitis et al.

---

Author comment on "Correcting Thornthwaite potential evapotranspiration using a global grid of local coefficients to support temperature-based estimations of reference evapotranspiration and aridity indices" by Vassilis Aschonitis et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-115-AC1>, 2021

---

We respect the opinion of the reviewer, but we disagree with his comment based on the following three aspects:

- **Regarding the monthly or daily time step.** The paper focuses on the correction of a specific methodology: the original Thornthwaite potential evapotranspiration formula; which is a formula that inherently provides only monthly estimations and not daily. Thus, daily reference evapotranspiration comparisons could not be anyway performed based on the corrections of this formula against FAO Penman-Monteith. The paper also deals with aridity indices that use monthly records and not daily. The potential evapotranspiration method of Thornthwaite is still used in these indices and the corrected version of it can significantly improve their performance using only temperature data (as it was shown by the results of this study) reducing the differences that are observed in aridity indices using different methods as indicated by Proutsos et al. (2021).
- **Regarding the machine learning models.** We provided a model that uses one independent variable with correction coefficients for each position in the globe. Machine learning methods are useful for describing phenomena with many independent variables and not just one (i.e., mean temperature monthly or even daily). If we assume that this was not a problem, a machine learning method should be calibrated-validated based on the whole global grid of records providing a unique global model. This would require the machine learning model to describe internally the different combined local effects of the non-participating parameters (wind speed, relative humidity, solar radiation) for each position in the globe, which in our opinion would be quite difficult (if not impossible). In this case, the global machine learning model would make an average description of the different combined local effects of the non-participating parameters (wind speed, relative humidity, solar radiation) for the whole globe that would lead to large errors in the much colder and much warmer regions. We know this, because we tried it during preliminary trials (non-published) for a previous paper (<https://doi.org/10.5194/essd-9-615-2017>) where machine learning techniques were not able to reach the accuracy of Hargreaves-Samani method corrected with local coefficients. That is why we developed the methodology of local weighted coefficients in that paper and used it also in this paper. Other option would be to build one machine learning model for each pixel (or cluster of pixels with similar conditions) covering the whole globe using only mean temperature (again the problem of one independent variable and quite difficult as task even with more independent variables for a global

application). Other option that we tested using machine learning techniques was to describe the local  $c_{th}$  factors using latitude, longitude and altitude, but the results were not as good as the direct use of the local coefficients.

- **Regarding the fact that the weather variables are nowadays easier accessible.**

We agree that every year the number of stations that provide complete datasets of climate variables is increasing. However, the spatial coverage over the globe is still not sufficient (e.g., in developing countries the number of stations is limited). Moreover, even in developed countries, the current number of stations is not sufficient to capture the climatic variation within watersheds. For example, the number of meteorological stations in mountainous areas\* is too limited. Lastly, even in many developed countries complete datasets are also not freely or easily available.

\* We have to stress the fact that the initial databases of Hijmans et al. for temperature and precipitation at 1 km resolution, which were used in this study, were built using ANUSPLIN-SPLINA package for interpolation, using latitude, longitude, and elevation as independent variables. This allowed to describe better temperature and precipitation patterns in mountainous areas.