

Comment on **essd-2022-67**

Anonymous Referee #2

Referee comment on "Downscaled hyper-resolution (400m) gridded datasets of daily precipitation and temperature (2008–2019) for the East-Taylor subbasin (western United States)" by Utkarsh Mital et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-67-RC2>, 2022

The manuscript by Mital et al. attempts to downscale PRISM data from 800 to 400m spatial resolution in the East-Taylor subbasin. The topic is of great interest to ecohydrologists. The presentation is very clear. However, there are many questions that need to be addressed by the Authors:

[1] PRISM data is available for the United States (<https://climatedataguide.ucar.edu/climate-data/prism-high-resolution-spatial-climate-data-united-states-maxmin-temp-dewpoint>). If I understood it correctly, the only fine-scale information used in the downscaling approach is related to topography. Then the obvious question is why not extend it to the larger area because 10m DEM is available from NED. I acknowledge that the importance of the selected study area is clearly mentioned in Section 2.1. But similar reasonings (as in Section 2.1) can be presented for some other areas as well.

[2] If the topography data is available at 10m and the method heavily depend on the topography, then why Authors chose 400 m as grid size for the downscaling?

[3] I assume that by precipitation map Authors mean both the snow and rainfall, is it correct? Looking at Figure 2, it is unclear how the SWE maps relate to hyper-resolution precipitation and temperature.

[4] Authors have referred to their previous work in Section 2.3.2 while mentioning about the gap filling in the weather data. This is a very crucial step. The justification is not very clear about using a data-driven imputation approach.

[5] Figure 4 shows the nearest neighbour map. It is not clear whether correlations were calculated based on the values at the grids or at the centroid of the polygons. Authors have rightly mentioned that the nearest neighbor for a grid point is not necessarily the closest station, especially in the area with different topography. Then what it means to have correlated stations, it is unclear.

[6] In the Section 4.2, Authors mention that 'if the downscaled datasets are aggregated to the original resolution, we should get back the original dataset'. This means Authors are trying for disaggregation, not spatial downscaling. Both are not the same. Authors may need to apply IDW type of approach to estimate the value at the centroid of the larger grid using the values at the smaller grids within the larger grid.

[7] In the beginning of section 5, the wet days are defined as precipitation > 1mm. There has to be some justification for this. Why not chose 0.01 or 0.5 mm?

[8] For the downscaled precipitation, only one figure (Figure 5) is presented to visually match the spatial patterns. Authors should look into spatial statistics such as Variogram, skill scores etc. to objectively match the downscaled value to the reference.

[9] Section 5.2 is very confusing. It is unclear why the mean residual error is calculated temporally when Authors are talking about spatial downscaling. Please consider expanding lines 280 to 285 as these are very crucial information to justify the downscaled estimates.

[10] In Section 6, the example use case is of ecohydrological modeling, more specifically in terms of modeling of SWE. Since there is not much mentioned about the methodology in the manuscript, it looks surprising why Authors couldn't use any hydrological models to check the streamflow in the mountainous area, why specifically SWE modeling? The discussion so far was about the downscaled Precipitation and Temperature, which itself has uncertainty. Now Authors have estimated four new meteorological variables and set up two RF models (Figure 11). I assume that all these are done at every grid, then how to take into account the spatial structure of precipitation and temperature.

[11] The title of the manuscript doesn't mention that the Authors have checked the applicability of the downscaled dataset only for the snow season. In Section 7, it is clearly mentioned that for dry periods and summer months, the Authors are not sure about the quality of the dataset. Then I suggest changing the title of the manuscript.