Interactive comment on ”MSWEP: 3-hourly 0.25°global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data” by H. E. Beck et al.

R. Wartenburger

December 1, 2016

This well-written manuscript describes MSWEP, a multi-source 3-hourly precipitation product. For the first time, the authors successfully attempt to combine satellite, reanalysis and gauge-based precipitation products by performing a 3-step weighted merging procedure. A validation exercise states that MSWEP generally outperforms other data products. Although the manuscript has undergone some substantial improvements since initial submission and I clearly recommend it for final publication in HESS, a few points still need to be addressed.

Major Comments

1. Definition of T_a threshold to distinguish between rain and snow (pp. 4, 10, 15): The choice of the temperature threshold is inconsistent: to derive the snowfall fraction map for the CR-based bias correction, the authors use 1°C, but for the merging they use 3°C. It is unclear to me a) what the reason for this difference is and b) how / why these numbers have been chosen (e.g., common practice?). Did the authors attempt to use other thresholds? It would be good to state this at some point in the manuscript.

2. Bias correction based on Q observations (p. 7): The Zhang et al. 2001 correction is only applied for regions with snowfall and/or complex topography. One might argue that this choice could lead to inconsistencies at the transition zones between regions for which the adjustment has been applied versus those where it hasn't been applied. Have the authors tested a set-up where the corrections are always applied? This would allow undercatch corrections in non-mountainous, snow-free

regions (e.g., for catchments located in the mid latitude westerlies, coastal stations affected by stronger wind speeds). In my opinion MSWEP might even gain some quality when applying the corrections to all gauge observations, even though this won't change the persistent P undercatch in the polar regions that the authors have pointed out on page 22. Some readers might also wonder how this would affect the agreement with Adams et al. (2006) shown in Figure 2c.

3. Generation of weight maps (p. 10): The median weight of the 10 most nearby gauges is used to interpolate station weights to grid weights. This can lead to some temporal inhomogeneities at time steps where one or several stations have gaps. This could be at least partly be resolved by some sort of gap-filling prior to computing the weights (e.g., following Andersson et al. 2012, see below). I realize that this could cost some effort, so it might also be a feature to be implemented in a future version of MSWEP. However, the reader should at least know that this issue exists. Andersson, J. C.M., Zehnder, A. J.B., Wehrli, B. and Yang, H. (2012), Improved SWAT Model Performance With Time-Dynamic Voronoi Tessellation of Climatic Input Data in Southern Africa1. JAWRA Journal of the American Water Resources Association, 48: 480–493. doi:10.1111/j.1752-1688.2011.00627.x

4. Generation of weight maps (p. 10): besides the point above, I am also a bit doubtful on the decision to use the 10 most nearby gauges. This will ultimately yield lower median weights as compared to choosing the 10 gauges with the highest correlation. E.g., the very localized precipitation variability of a mountain station near the centre of the 0.25°grid cell might not be well reflected by the evaluated gridded data product, while all other nearby stations (within some cone of influence) have a higher correlation. One could now argue that the correlation value of the station with the local P signal should thus not have an influence on the weight assigned to its associated 0.25°grid cell. Similar issues might arise in the third merging stage (equation in line 16 on page 10) when setting D_o to a constant value of $25\,km$, which in fact depends much on terrain roughness. Of course one should not overcomplicate things, but this should at least be kept in mind.

5. P undercatch in FLUXNET (p. 15): I wonder if it would not be even more useful to evaluate MSWEP against undercatch-corrected FLUXNET data. However, as I am not an expert in this field, I am not aware if this is feasible without loosing the independence criterion and without introducing any substantial additional biases. Undercatch-corrected FLUXNET data would, however, have the potential to state whether the undercatch adjustments in MSWEP are in fact reasonable or not, and it should underline MSWEP's superiority relative to the other data products (also in terms of the absolute bias). If it is not an option to correct the FLUXNET data, the reader would certainly appreciate if the authors could at least elaborate a bit more on this point.

6. Median NSE scores increase with increasing distance to the closest P gauge (p. 17): The authors state (indirectly) that this is primarily due to the dominance of (semi-)arid catchments (which tend to exhibit higher NSE scores) in densely gauged groups. Have the authors applied a test-wise removal of the (semi-)arid catchments or how do they come to this conclusion?

Minor Comments

- Dataset choice for MSWEP (p. 8): While it is clear to me how the set of satellite and reanalysis P datasets for use in MSWEP was defined, I miss information on how the gauge-only gridded products (CPC Unified and GPCC) were chosen. I assume the reasoning follows the list of issues pointed out on p. 3, but it is nowhere clearly stated that CPC Unified and GPCC are the only gauge-only datasets that meet all criteria.
- Calculation of B (p. 11): How is the long-term mean defined here? Does it correspond to all time steps available in the observational record? I would appreciate as a reader if this was shortly mentioned in the text.
- Central Asian mountain ranges (p. 12): I would not count the mountain range in Iran to central Asia, maybe better rephrase this to something like "mountain ranges in Central Asia and Iran".
- Figure 11 (p. 24): Why is the colour scale inverted here (wrt. e.g. Figure 7)? I think that most readers will, at the first glance, expect that red colours correspond to high NSE scores.