Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2018-150-RC2, 2019

Open Access Earth System Science Data Discussions

Interactive comment on "WHU-SGCC: A novel approach for blending daily satellite (CHIRP) and precipitation observations over Jinsha River Basin" *by* Gaoyun Shen et al.

Anonymous Referee #2

Received and published: 21 February 2019

General Comments The manuscript describes a very interesting method to correct random and systematic errors from satellite infrared precipitation estimates based on gauge data and grid points from interpolated precipitation fields. The manuscript is well structured and present links to access all the data that is used in the work. However, I found that the analysis period (summer of 2016, JJA) is too short to make significant conclusions about this precipitation dataset. Also, the manuscript could have explored another ways to analyze and evaluate the dataset other than the very conventional use of statistical metrics, that are definitely needed, but the authors could have explored beyond that. Maybe using a case study or exploring the usefulness of the data for a particular hydrological application would be helpful. The dataset published with the

Printer-friendly version

Discussion paper

manuscript is of good quality and is stored in a standard and easy to read format. I suggest that the authors make another English revision, especially regarding the use of articles and prepositions.

Specific Comments Line 18: It was evaluated not only by categorical indices. Line 28: The number of gauge stations is actually very limited, especially in regions with complex terrain and in the case of gauges that measure solid precipitation. Accuracy of this gauges is also not very good in the case of solid precipitation. It would be a good idea to state that the paper focus is on liquid precipitation (rainfall), and use this term throughout the text. Line 35: I found this line confusing in the way it is phrased. I think this sentence could be phrased this way: "Satellite estimates are susceptible to systematic biases that can influence hydrological modelling." In the introduction, I think that a better description of what is available to estimate precipitation from satellites is missing. For example, GOES-R and GPM are missing in the description. Line 62: That are other sources of uncertainty in the monitoring of rainfall in complex terrain (e.g., orographic enhancement). Line 94: I was not able to understand the average precipitation over the Yangtze River Basin. You could be more specific about what statistic you are presenting here. Usually what is presented is the spatially averaged annual accumulation of precipitation as an indication of precipitation climatology for the region. Line 98: I was not able to comprehend the units for the area of the basin. It should be presented as km2. Line 102: Topography would not exert a temporal variation in climate, since this is not a very dynamic feature of the Earth's surface. It could exert a temporal variation in weather though. Lines 102 and 103: Please try be consistent with the statistics you are using here. Line 134: Maybe explain better what is the CHPClim product. Line 146: Which CHIRPS data? The data from its stations? Or the blended product? Please make this clear here. Line 162: What do you mean about physically similar? Is this means that these pixels are related to others based on its physical attributes (lat, long, elevation, slope, aspect, and curvature)? Or this means that is similar is terms of rainfall distribution? If is in terms of rainfall, I think a better world would be statistically similar rather than physically similar, since

this is based on a cluster analysis. Line 170: Please explain why you chose 30% of the stations/grid points for validation. Since the stations and grid points are of limited number, it would not be better to do a Bootstrap validation instead? Then you are able to use all the stations/grid points in your correction algorithm. Lines 224 and 225: Since your method relies heavily on the cluster algorithm, it would not be better to use some sort of statistical metric to define the number of clusters? Line 234: Pixels should also be similar regarding precipitation characteristics, is that right? Line 244: I think you should use the word relationship or correlation instead of confidence in this line. Line 259: Why the number of decision trees was set to 500? Line 264: Do C1 and C2 pixels included in the R pixel category? Lines 267 and 280: Are you also considering only SCC values with p-value lower than 0.05 here? Line 287: Why did you choose the 10 mm value? Could you please explain the meaning of this constant better? Line 298: Please use the actual percentage here. Line 326: You should be clearer on what the numbers of pixels are here. Is this the number of pixels inside the basin multiplied by the number of days? It would be a good idea to describe the exact number of pixels for each class along with its actual percentage in the text. Line 341: It would be better to use the same x axis scale in both plots. It seems that the gridded data observation has similar biases as CHIRP and thus their CDFs are more similar, providing less improvement in the adjusted dataset. Line 355: It would be a good idea to explore and discuss more the statistics presented in Figure 7. Line 373: NSE values have increased, but still not very good (i.e., still negative). Line 376: Is not intuitive that the evaluation metrics are better for CHIRP than CHIRPS, since CHIRPS adds stations data to their dataset. Could you please clarify this? I am seeing here that magnitude evaluation metrics have not changed considerably, probably because the improvement is seem in the low magnitude events. SCC has a considerable increase, but still cannot explain much of the variability of rainfall in the region. POD values are good. Line 383: Could this means that because the daily precipitation in the lowland region of the basin is higher, the RMSE values are also higher? Line 388: The fact that your method performs well in complex terrain is a very positive point in your manuscript,

but you will need a longer study period to confirm this finding. Line 402: The boxplots do not show that the higher reduction is seen in the Bias metric. Line 417: I am still confused about which rainfall value is presented is this figure. It does not seem realistic that the average daily rainfall would be $\sim$ 200 mm. Please be more specific about which rainfall statistic you are using in this figure. Line 430: The evaluation metrics for the threshold values are quite similar among the three precipitation products. Just because the values for WHU-SGCC become slightly worse for precipitation higher than 20 mm/day does not mean they are significantly different from the other products. Since the values are very similar, I would suggest to test their differences statistically before making the assumption that WHU-SGCC works better for low magnitude events. This might also be caused by the limitation of the short study period. There is a tendency to use the word scale for a temporal dimension, better to use period instead. I think it would be interesting to add a map with the accumulated precipitation for the study period. The analysis period of 92 days is too short to make conclusive assumptions about the dataset usefulness in the region, which are made several times throughout the text. I understand why the authors want to focus on the summer months to avoid the higher biases introduced by solid precipitation, but I did not understand why they only performed the evaluation for one summer season. Is there a particular reason for that? I still think that is hard to make the conclusions you made based on 92 days, if you add more seasons to the analysis this can become a very interesting manuscript and dataset. The dataset seems to be of good quality. A few comments about it are the following. For a spatial extent of this magnitude I think it would be better to use a geographic coordinate system, rather than a Mercator projection. There is also some artifacts (0 precipitation values) that appear at the same location at multiple days. I was wondering if this is a limitation from the negative values of Rule 4. Is there any way to correct this? This dataset could be very useful if its period is extended to multiple years.

Technical Corrections Line 2: Change “over Jinsha” for “over the Jinsha”. Line 31: Change “distributed” for “spatial distribution”. Line 37: “without adjustment” is men-

tioned twice. Line 68: In table 1, it should be written “PERSIANN-CDR” instead of “PRESSIANN-CDR”. Line 84: Change “in summer 2016” for “in the summer of 2016”. Line 93: Change “proximately covers an area” for “covers an area of approximately”. Line 95: Change “sub-regions” for “sub-basins”. Line 136: Change “precipitation observations” for “surface based precipitation observations”. Line 157: Change “other pixels” for “the remaining pixels”. Line 159: The acronym “SIRC” meaning was not mentioned before. Line 169: The first sentence could be placed before the item 1. Line 171: Is the same phrase that is shown in line 163. Line 172: The flowchart: CHIRP resolution should be 0.05 x 0.05. In the first box of rule 3, change “and gauged” for “with gauged”. In the last box of rule 3, change “can derive” for “can be derived”. In rule 4, change “ration” for “ratio” (this happens twice here). Line 189: Change “as” for “in”. Line 246: Change “p” for “p-value”. Line 283: The word “method” is repeated twice. Line 300: Change “for summer (JJA) 2016” for “for the summer (JJA) of 2016”. Line 315: Change “as” for “in”. Line 326: Change “to be adjusted” for “adjusted”. Line 340: Change “study” for “studies”. Line 400: Change “with especially greatly decreases compared to CHIRPS” for “with greater decreases when compared to CHIRPS”. Line 451: Change “in summer 2016” for “in the summer of 2016”. Line 456: Change “over region has” for “over a region that has”. Line 465: Change “of the precipitation region” for “of precipitation events in the region”. Line 466: Change “short” for “short duration”. Line 468: Change “complicated mountainous” for “complex terrain”. Line 480: Change “topographic and long time series climatic factors” for “topographic factors and longer time series”.

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-150, 2019.