

Interactive comment on “HydroGFD3.0: a 25 km global near real-time updated precipitation and temperature data set” by Peter Berg et al.

Anonymous Referee #2

Received and published: 23 November 2020

General comments:

The authors present a new version of the HydroGFD data set, which contains (nearly) global daily precipitation and temperature reanalysis data since 1979 at 0.25° resolution. This data set is based on the ERA5 reanalysis data from the ECMWF, which are corrected on the basis of several observation-based gridded data sets. HydroGFD3.0 is meant to be used e.g. as atmospheric forcing data set for hydrological modelling and for impact studies.

This kind of bias adjusted data sets is certainly very useful, in particular since i) it has a global coverage (except Antarctica), and ii) it will be prolonged in near real time, making it suitable for operational applications.

C1

Besides some ambiguities (see below), the paper is pleasant to read. However, I have two major remarks about the methodology:

1) About the evaluation The authors mainly compare the HydroGFD3.0 data set with the observational data sets they use to calculate the bias adjustment and with the WFDE5 data set, which also relies on these same observational data sets. Thus, I cannot completely agree with the author's conclusion on l. 281. Comparing HydroGFD3.0 to the data sets it relies on (observation-based data sets and ERA5) is very interesting as it evaluates the impact and the usefulness of the bias correction applied to the ERA5 data. However, the consistency and reliability of HydroGFD3.0 could be shown in a much more convincing way, if the authors could add an in-depth evaluation by comparing HydroGFD3.0 with at least one independent data set. Further, as HydroGFD3.0 is a new version of already existing data sets (HydroGFD1.0 and HydroGFD2.0), it might be interesting to add a comparison with these previous versions. This would show the improvements and the added value of v3.0 compared to the previous versions and thus the interest of using this new data set.

2) About the long term changes in the input data sets To my understanding, any changes in the variability within or between data sets used for the bias adjustment, but also long term trends within these data sets will be transposed to HydroGFD3.0. As both the variability and long term trends might be different from one data set to another, the consecutive use of different data sets over time might induce discontinuities in the HydroGFD3.0 data. This aspect, along with a short comparison of the variability and the long term trends between consecutive data sets, should be discussed in the manuscript.

Specific comments:

Section 2 : It would be very useful for the understanding of the paper, if the authors could give more details about the data sets they use, e.g. better differentiate if a data set is a reanalysis or interpolated observations, the specificities of each data set and

C2

especially in which way it is complementary to or different from the other data sets. The authors should also expand the acronyms (CRU, CPC, etc.) the first time they are used.

Section 2 : The authors talk about background climatology (l. 56), historical period (l. 57, which period ?), climatological adjustment (l. 60, 61), etc. which are only explained later in sections 3 and 4. It might be easier to follow if section 2 only described the data sets used (see comment above), and if paragraph l. 56 was moved at a more appropriate place in section 3.

l. 53 : “with a similar model as that used for ERA5”. It would be useful to add some more information here (Is it the same model ? Are the biases supposed to be similar ?).

Section 3 : It would be interesting if the authors could add some explanation on why they calculate the correction on a monthly basis and not e.g. on a 30-day running mean basis around each day, which would smooth the correction curve applied to ERA5.

Section 3 : As the authors use monthly, daily, and hourly data, and also process the data on monthly, daily, and hourly time scales, it would be very useful for understanding to always clearly mention the time scale e.g. on which corrections are applied, anomalies are calculated, especially in this section, but also throughout the whole manuscript. For example, it is not clear to me whether the monthly corrections are applied to hourly or daily P.

l. 75: How / in comparison to which reference were the issues in CHPclim identified ?

l. 80 and 82 : Is the remapping of T and Nwet also conservative as for P ?

l. 83 : “retaining the grid points that are available consistently in all data sets and all months.” Does this mean that HydroGFD3.0 data are not available for some land grid points ?

l. 102 : The dry and wet days should be defined clearly here at latest. Is the threshold

C3

of 1 mm/day mentioned in the caption of Fig. 7 used to separate dry from wet days ?

l. 106 : Does N_clim_wet come from the CRU data set ?

Section 3.4 : This synthesis of the correction steps is very useful. However, it might be a little bit more detailed : - To my understanding, the very first step is the preparation of the climatology. - Step 1 : “Calculate monthly anomalies in observation data” ? - Step 4 : the removal of the weakest excessive wet days should be (more clearly) explained in section 3.3. - Steps 5 and 6: This is not clear to me. Does it only concern P ? If not, I do not understand what the ratio stands for.

l. 182-183 : It might be interesting to add some more discussion, e.g. remember as said before that the relative biases for P are much higher in these dry (or snowy) regions. Moreover, as no correction is applied over Greenland and the biases are quite huge, should one conclude that the climatology is not reliable there ? It might be useful to briefly discuss the interest of not excluding Greenland as it is done for Antarctica.

l. 191 : “For T, we compare to cru only, since cpct is used to build the climatology”: - This is confusing as following e.g. section 4 (e.g. l. 145), fig. 3, and l. 291, cru seems to be used for the climatology, but l. 56 mentions cpct. - In any case, it might be interesting to compare with both data sets to see the impact of the bias adjustment, as it is done for P with gpch.

l. 203-204 : It would be interesting to add some explanation on why the biases are larger in these regions (see also comment for l. 182-183).

l. 210 : Correct “... dry regions ... have more dry days ...”. Are the differences really significant / worse to mention ? The values seem to be very similar.

Paragraph l. 226 : It seems that many regions show an annual cycle. This should be briefly discussed, as it is done for T.

l. 283 : Which was the resolution of the previous versions ?

C4

l. 286 : I agree and this is a major advantage of this data set. However, would such a switch between data sets not introduce biases (see e.g. the general comment on the variability and long term trends) ?

l. 287-288 : As section 8 follows directly, it is redundant to already mention this here.

Technical corrections:

l. 8 : Correct "... as well as the number of wet days ..."

l. 49 : "tiers" are explained in section 4 "Data sets" not in the methodology. It might be useful to already explain here in a few words what is meant by "tiers".

l. 52 : Expand "SMHI".

Tab. 1 : Nwet acronym is not defined.

l. 71 : "climatological period" Is this 1980-2009 ?

Section 3.1 : Use data set names of Table 1 (e.g. CHPclim, GPCCv8).

l. 94 : Correct "1989-2009" to "1980-2009".

l. 128 : Correct "It happens that the land sea masks ..."

l. 132 : Correct "... the output will resort no adjustment ..." ?

l. 145 : Correct "... respectively for P and T."

Fig. 4 and l. 172 : With respect to which data is the bias of e5 shown here ?

Fig. 4 : It might be more consistent to represent Greenland in another way, e.g. like Antarctica, as no bias could be calculated.

l. 157 : Correct "... cpcp and cpct products ..."

l. 164 : Correct "... cpcp and cpct products ..."

l. 172, 206 : Correct "HydroGFD3"

C5

l. 184 : Correct "Arabian peninsula."

l. 190 : Correct "...where both gpcch and cpcp show ..."

l. 195 : Correct "... they are due to differences in elevation ..."

l. 216 : Add reference to Figure 8.

l. 219 : Correct "Orographic effects on T were ..."

l. 234 : Correct "... have similar mean ... and show generally ..."

l. 242 : Should it not be 2016 instead of 2017 ? If not, Fig. 3 should be adapted.

l. 244 : Correct "... a significant offset between the data sets ..."

l. 245 : "... reduce the offset to cru." Should it not be to e5 ?

Fig. 9 and 10 : - Are these monthly anomalies ? - Correct "... and are evaluated ..."

l. 258 : Correct "... using this method is to be able ..." ?

l. 292 : Switch gpccm and cpct to be consistent with the previous sentence.

Interactive comment on Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2020-236>, 2020.

C6