Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2020-236-AC2, 2020

Open Access Earth System Science Data Discussions

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

Peter Berg et al.

peter.berg@smhi.se

Received and published: 22 December 2020

Thank you for the thorough review of our manuscript, and for the very helpful comments on improvements.

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 0.25âŰę 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.

Printer-friendly version

Discussion paper

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.

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.

We agree that a complete evaluation should use independent data. Truly independent data are rare at a global scale since most large data sets are already included in the gridded observations. However, our aim is to provide a comprehensive overview of the data to present its qualities and also to point to potential issues. To this purpose, we compare to other data sets at a global scale, with specific analyses for sub-regions. The analysis is based on a range of data sets that have been completely or to a large extent independently generated.

For precipitation, the HydroGFD3 climatology is based on the satellite-gauge merged CHPclim climatology, with some minor additions of GPCCv8 for Scandi-

navia and part of Siberia. Therefore CPC-Unified, GPCCv8, wfd-gpcc and wfd-cru can all be considered independent data sets and are valid for the evaluation in Fig. 5. For the daily distribution and timeseries in Figs. 7 and 9, HydroGFD3 is based on climatology with GPCCv8 anomalies, and is therefore independent of CPC-Unified, partly independent of WFDE5-CRU (besides ERA5), and to some extent independent of WFDE5-GPCC regarding the climatology.

For temperature, the HydroGFD3 climatology is based on CPC-Temp, and CRU is therefore independent data. The daily HydroGFD3 version for temperature is based on climatology and CRU, and is indeed strongly dependent of CPC, and partly independent of WFDE5-CRU (besides ERA5). Please note that we have corrected ourselves after your comment below on the inconsistent use of CPC and CRU for temperature.

The daily timeseries are plotted as a visual guide to spot issues with trends of the different data sets, with specific focus on the last years when the different tiers might affect the temporal consistency. It is not a full-fledged evaluation, but is there to give an impression of the quality of the product. Our experience from earlier studies is that in-depth evaluation can only be performed at the local scale, e.g. Fallah et al., (2020; in the reference list), and we will encourage users to pursue such evaluations.

About comparisons to earlier HydroGFD versions. The main purpose of HydroGFD3 is the general improvements of the methodology as well as the increased redundancy options for the realtime production. The earlier versions were discontinued with the end of ERA-Interim. We will consider adding some evaluation on differences between the versions to the supporting material.

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.

This is indeed what we are showing in the timeseries plots. Anomalies of the data sources do contain the trends and potential discontinuities. For this version, an implicit choice was made to let trends be determined by each observational source. However, future versions could work around the issue of discontinuities by adopting the ERA5 trends and to use a shorter time window to construct anomalies of a data set. We will add discussions along these lines in a revised version.

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 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.

Thank you, we will improve the description of the different datasets with this in mind.

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.

Thank you, we will adjust this to increase the readability.

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 ?).

The intention of this sentence is to justify the choice of reanalysis system ERA5 instead of other systems, such as NCEP or JRA. The similarities or differences with the operational products of ECMWF are not relevant to the topic of the paper and we do not want to confuse the reader. The sentence will be re-formulated to clarify this.

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.

That would have been neat, but is not possible because some of the observational data are only available at calendar month resolution.

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.

We will clarify the procedure of applying the scaling to each single timestep, and also that we do not currently produce hourly data. The reason is that we have not carefully evaluated the ERA5 hourly data, which is likely to have strong biases due to the coarse resolution and parameterized convection.

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

By visual inspection. The data set contains obvious unphysical lines for some

months and regions.

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

Thanks, we will describe this better. These two sources are mapped from 0.5 to 0.25 degree using a bilinear method for this reason.

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 ?

Correct, not all land grid points are available. The grids are described in the figures and in the data set itself and most notable is that some islands are missing.

l. 102 : The dry and wet days should be defined clearly here at latest. Is the threshold of 1 mm/day mentioned in the caption of Fig. 7 used to separate dry from wet days ?

The method of calculating wet days does not require a definition, nor does the adjustment. Instead, the definition is imposed implicitly from the wet days in the CRU data set. This is a complicated measure since it is a mixture of instrumentation resolution and interpolation methods.

l. 106 : Does $N_c lim_w et come from the CRU data set?$

Yes, we will clarify this.

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.

Thank you for noticing the issues in the explanation. 1 is correctly interpreted. 5 and 6 is a ratio for precipitation and a difference for temperature. We will revise

and clarify, also regarding the wet day correction in 3.3.

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.

We will consider adding more discussion here. Regarding Greenland, this is a difficult topic. The observations are generally poor, and this goes for other parts of the Arctic coastline as well. Although we do not change the climatology, we still include observed anomalies for Greenland since they might provide some useful information. Antarctica is excluded because the observational data sets do not provide any information there at all.

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 gpcch.

Thank you very much for spotting these inconsistencies! Both cpc and cru versions of the temperature time series have been produced, however the cru-version was chosen for the main data set. We will investigate if there was only a switch in the naming in the plots, or if also the wrong data were used. The plots, evaluation and manuscript will be updated accordingly.

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).

We will consider explanations for the differences and to extend the discussion.

l. 210 : Correct ". . . dry regions . . . have more dry days . . .". Are the differences

really significant / worse to mention ? The values seem to be very similar.

We mention these differences because the extreme values will be scaled to a larger extent than moderate and weak intensities, and the scaling becomes stronger when more days are removed from ERA5. A small difference in the wet days can have a large impact on the tail of the distribution.

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

We will add such discussion.

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

It was 0.5 degrees, we will describe these earlier in the revised manuscript.

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)?

Indeed, the switch does sometimes introduce discontinuities, as shown in the timeseries plots. The use of a common climatology reduces discontinuities that arise from consistent offsets, however, it cannot affect differences in variance, or changes in the data set itself, such as changes in the underlying station network. More elaborate methods are necessary to address such issues.

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

Indeed, it is.

Technical corrections:

Thank you for the detailed reading, we will address the all the following points in our revision.

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”

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.

Printer-friendly version

Discussion paper