

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

Dear Graham, Thank you very much for the detailed reading and the useful comments to our manuscript. Below, we answer in bold font to all issues raised.

Overview:

Berg et al. describe the processing procedures needed to generate the HydroGFD3 data set from the ERA5 reanalysis. The data set covers hourly precipitation rate, near surface air temperature and daily mean, maximum and minimum near-surface air temperature at 0.25o x 0.25o spatial resolution. The freely available data cover February 1979 to December 2019, but later, near real-time data (up to 5 days ago) requires a subscription.

C1

A minor correction is in place here. The data set does not include the hourly precipitation rates, only aggregations to the daily scale. It is indeed possible to construct data at this frequency with the methodology, as you also do in Cucchi et al. (2020), we have chosen not to make this data available until we have analysed this more thoroughly. The hourly time scale is often poorly represented in coarse model data, and data for verification is even more difficult to get hold of.

The processing methodology builds on the WATCH Forcing Data (WFD) procedures (Weedon et al., 2011, JHM, Weedon et al., 2014, WRR) that were used in HydroGFD1 (Berg et al., 2018, HESS). The paper compares HydroGFD3 with the precipitation and temperature components of the WFDE5 data (Cucchi et al., 2020 ESSD, though these data are at 0.5o x 0.5o resolution) and with the observational datasets used in the processing to establish the climatologies needed for bias correction.

The descriptions of the processing procedures are clear - though I may well be biased since I designed the WFD methodology on which they are largely based. Comparisons of HydroGFD3 with WFDE5 and the contributing data sets are favourable aside from a problem with the probability of rainfall in the MED and SAH regions (Fig. 7). In these areas the authors have noted/acknowledged (lines 262-264) that they have far higher probabilities at almost all rainfall rates than the comparison data sets probably due to the reduced numbers of wet days used.

Aside from some minor issues and text corrections, there is one key weakness with the manuscript which I will raise below. Otherwise the manuscript is well presented and close to acceptable for publication.

Key weakness:

The key weakness concerns the evaluation. As explained HydroGFD3 derives from HydroGFD1 which was based on the WFD methodology. Consequently, it seems likely that HydroGFD3 will inevitably compare favourably with the WFDE5 data set since the latter was also produced using the WFD methodology (Cucchi et al., 2020,

C2

ESSD). Similarly, the various data sets listed in Table 1 have to an extent been incorporated into HydroGFD3 so again comparison with the new data set is liable to be favourable. My recommendation therefore is that the authors re-evaluate HydroGFD3 using truly independent data set (for example, using global data produced using a non-WFD methodology and without using an ERA reanalysis and/or using site observations e.g. FLUXNET2015 data).

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 Scandinavia 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 another reviewer spotted an inconsistent use of CRU and CPC-Temp data in the manuscript, which will be amended for the revision.

C3

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.

Nomenclature issue:

Line 63: "WATCH forcing data WFDE5" is NOT correct and is misleading. WFDE5 means the "WATCH Forcing Data methodology applied to ERA5 reanalysis" (Cucchi et al., 2020, ESSD). WFDE5 has nothing to do with the WATCH programme (which ended in 2011). The authors should avoid using "wfd" as a shorthand for WFDE5. Correct all uses of "wfd" to WFDE5 or wfde5 throughout the manuscript (i.e. Table 1, 181, 185, 186, 191, 199-200, 213, 219, 226-227, 231, 233, 256, Figs 5, 7, S5, S6, S7, S8).

Thank you for the clarification. We will adopt your suggestion and rename is in the revision.

Update:

Cucchi et al. (2020 ESSD) has been revised and is now fully published – update the reference list.

Thank you, we will of course update this reference.

Issue of interpretation:

My issue of interpretation concerns lines 186-188: "However, the two wfd [WFDE5] datasets also tend to be drier in very dry areas, which is likely to due to the direct use of the number of rain days from the CRU data set. An incompatibility between P and no observed wet days can act to remove P completely for some months, and therefore making a drier data set." This is the wrong explanation for why WFDE5 data appear

C4

drier than HydroGFD3 in Fig. 5. Firstly, the WFDE5 monthly precipitation totals are adjusted after the correction of the number of wet days. So, as long as there are wet days, the wet day correction does not lead to too little monthly precipitation.

Secondly, suppose for a grid box the number of ERA5 wet days were to be set to zero from using CRU wet days even if there is some precipitation according to CRU or GPCC. This situation would mean that all the days with ERA5 precipitation would have been reset to zero in the WFDE5 processing. However, the adjustment of a hourly ERA5 rate involves multiplication by the CRU or GPCC monthly total divided by the (supposedly zero) monthly ERA5 precipitation total. This would generate a "NaN". Such a result would have caused the WFDE5 processing to stop since there is a check for plausible precipitation rates for every grid box and every time step. Hence, there are never occasions in the WFDE5 processing when the CRU number of wet days is zero but the ERA5 and CRU totals exceed zero. As a result, there needs to be a different reason given for the observation that the WFDE5 data are drier than HydroGFD3 in the dry regions in Fig. 5.

Thank you for providing more insight on the WFDE5 processing on this issue. We agree completely in what you are stating. However, not knowing exactly how the situation of division by zero was programmed, we speculated about the reason. We can now conclude that the reason for the differences actually lies in a combination of the background climatology mixed with GPCC and that zero-division is avoided in HydroGFD3 by setting the scaling to zero when this occurs. The manuscript will be updated accordingly.

Minor text changes (L = line number):

We appreciate the detailed checks, and will adjust all in a revised version.

L26 Add Cucchi et al 2020

L74 Used here "conservatively" sounds like it means being cautious rather than en-

C5

surging conservation of total precipitation during interpolation/remapping. Reword the sentence.

L77 boarders > borders

L94 1989-2009 shouldn't this be 1980-2009 as used everywhere else?

L127 version > versions

L131 successfully > successful in

L131 will be filling > is used to fill

L132 data is > data are OR data is > data value is

L139 PFDs > PDFs

L185 "the data" is ambiguous change to "wfde5 data"

Fig 7 cpc > cpcp

Fig 8 cpc > cpct

Fig 9 cpc > cpcp

Fig 10 cpc > cpct

L252 failover > failsafe [wrong use of failover]

L258 as to > is to

L264 enlarging the tail of the distribution. > enlarging the tail of the distribution (e.g. in the MED and SAH region in Fig. 7).

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

C6