WFDE5: bias adjusted ERA5 reanalysis data for impact studies

Summary and General Comments

This paper presents the WFDE5 dataset, an atmospheric forcing dataset which will further be used to drive and evaluate impact models. It is constructed by applying the WFD methodology to ERA5, the last generation of ECMWF reanalyses. WFDE5 is constructed with the same methodology as the WFDEI dataset, which has been constructed from the previous generation of ECMWF reanalyses, ERA-Interim.

WFDE5 is based on a monthly bias-correction of ERA5 by CRU TS4.03. It contains all the variables required by impact models at the conventional format (cf. ALMA conventions). In order to consider the high uncertainty of precipitation, WFDE5 is available with two different bias-adjustment of precipitation, one has been bias-adjusted by CRU and the other by GPCC. A derived daily dataset, W5E5, has also been created for upcoming ISIMIP phase 3, combining WFDE5 over the land to ERA5 over the ocean. This paper details the improvement and adjustment of the WFD method in order to fit with the ERA5 dataset.

WFDE5 benefits from the ERA5 improvements and in particular from its higher spatial and temporal resolution. Even if both WFDE5 and WFDEI have the same spatial resolution of 0.5°, the WFDE5 dataset integrates more spatial variability than WFDEI being constructed by aggregation instead of interpolation. It also integrates more temporal variability as WFDE5 is available at an hourly temporal resolution instead of 3-hourly for WFDEI The evaluation of the WFDE5 dataset is done in comparison with WFDEI and with ERA5. This is done by comparing (1) their performances over 13 FLUXNET2015 sites well distributed over the world and (2) their performance at forcing an hydrological model (WaterGAP) in order to have a first estimate of the capacity of WFDE5 to drive an hydrological model.

The manuscript is well written and well organized. This dataset is a good contribution to the land modeling community as it permits to use a bias-adjusted version of the last release of the ECMWF reanalyses, ERA5. It is promising as it will benefits from further improvement of ERA5 like, for example, from the future extensions of the period covered by ERA5. The code is publicly available so it will help the community to use the WFD method adjusted for this new version and to generate new forcings at higher resolution (when all the ground-based observation of the variables will be available at these resolutions).

I recommend the publication of this paper after some minor revisions.

Specific comments

Line 127 :
You explain how you process the grid points of CRU TS4.03 and GPCCv2018 that are not considered as land points in ERA5 and declare that "*In this way, the final WFDE5 dataset contains values only for all grid-points which are classified as land or lake by both ERA5 and CRU*".
Have you been confronted to the opposite, land points of ERA5 that are not considered as land points in CRU or GPCC ? If that was the case, how did you proceed to bias-adjust them ?

Validation with a global hydrological model :
The simulations with the hydrological model WaterGAP are used to assess the capacity of WFDE5 to force an hydrological model compared to ERA5 and to WFDEI. The Figure 5 shows the annual cycle of the outflow of the 12 large river basins over the period 1981-2010.
For the basins with a FLUXNET2015 stations, if we can make the hypothesis that the bias over the FLUXNET2015 stations is representative catchment area, crossing the results from the WaterGAP simulation with the previous analysis of the FLUXNET2015 stations may allow to understand which variables are responsible of the differences between CRU/GPCC and between WFDE5/WFDEI.
I think that your analysis is already quite complete but have you considered crossing the results from the WaterGAP simulations with the previous analysis of the FLUXNET2015 stations ?

Technical corrections

Line 39 :
I suggest to add the reference of ERA-40 here. The reference is present but later in the text at l. 64.

Line 118 :
I suppose that the "*validity date-time*" represents the start time of the time step, I suggest that you define "*validity date-time*" so the text would be clearer.

Line 245 :
I suggest to change: "since the assessment of the water balance components are highly dependent on it" to "since the water balance components are highly dependent on it"

Line 255-259 :
I think you forgot to close the parenthesis opened before "*the latter*" in this phrase :
"*The model was driven by ERA5, WFDE5 and WFDEI (the latter ...*"

Line 267-269 :
It is not clear that the variable you are comparing between the CRU and the GPCC version is the river discharge. I suggest to precise : "*difference of discharge : 1825 km 3 yr −1 ...*"

Line 311 :
Please change "*aggegated*" to "*aggregated*"