Interactive comment on "A Climate Data Record (CDR) for the global terrestrial water budget: 1984–2010" by Yu Zhang et al.

Anonymous Referee #1
Received and published: 20 May 2017

In this manuscript the authors describe the development of a global terrestrial water budget time series at 0.5 degree spatial resolution and spanning more than two decades at monthly intervals. This work expands upon previous efforts by some of the co-authors and represents an important next step in bringing together a variety of data sources at the global scale while addressing the problem of water budget closure at the grid scale. The authors describe a rather comprehensive consideration of precipitation, evapotranspiration, runoff, and storage change datasets and how the variations in dataset extent and consistency are addressed to produce merged spatial time-series for each budget component. They then describe the grid-scale water budget constrained data assimilation process and results, notably including a presentation and discussion of attribution of closure errors. Finally, the authors present comparisons of the derived product against independent observations, noting overall adequate agreement but regions of poorer match and potential reasons for this. Overall the manuscript is well-written and describes the process and results well. Figures are consistent, descriptive, and illustrate important points from the text.

Comment 1 (C1): General comments: The authors propose the developed dataset as a publicly-available reference for understanding climate variability and trends. The bulk of this manuscript describes how the derived data product captures mean behavior of the terrestrial hydrology. While this is fundamental and important, less attention is paid to the extent to which the data product captures inter-annual variability. Additional text and perhaps a figure describing the climatologically-relevant variability or cycles in the produced dataset would greatly improve the manuscript and be valuable to users of the product.
This manuscript describes a temporally and spatially disaggregated global terrestrial water budget. It would be helpful to put the results of this work into the broader context of global water budget quantification by including comparisons of the derived average water budget components with previously published global budgets (e.g. Trenberth et al, 2007).

Response 1 (R1): Thanks for the reviewer's comments. The authors have added a paragraph in the "discussion and future work" section to (1) compare with other studies about the global water budgets (2) describes the climate variability from the CDR. Corresponding figures (Figures S6-8) were added into supplement II to support the text. A paper is under preparation that reviews global water budget estimates from ~20 historical studies. Including these studies in the current paper would make it simply too long, so we summarize some results and put some results in a supplement.

"Currently the authors are carrying out another study in comparing the CDR water budget records against around 20 high-impacted studies, at multiple spatial scales (i.e. continental and global). This on-going study is the first attempt to gather and compare global water budget estimates from studies as early as 1974 (i.e. Budyko 1974) to the current study in order to provide a comprehensive overview of global water budget estimates, even though the studies focused on different periods using different data sources and have different global coverage (e.g. some of them exclude Antarctica or Greenland or both). Figure S6 in supplement II gives an example comparison with (Trenberth et al., 2007, T2007 hereafter), which estimated the water budget during 1979-2000 and excluded Antarctica. The total precipiation is quite close to this study ($114{\times}10^3 km^3/yr$) to T2007 ($113{\times}10^3 km^3/yr$). By converting the water budgets into mm/yr based on the global coverage information available in each of those studies, the long-term mean precipitation is around 28 mm/yr (vs. 32 mm/yr in the CDR and 27 mm/yr from T2007), ET is around 78 mm/yr (vs. 78 mm/yr in the CDR and 77 mm/yr from T2007), and runoff is around 47 mm/yr (vs. 46mm/yr in this study and T2007). Figure S7 further provides an example of how the CDR precipiation time series captured the 1998-1999 US drought. The 6-month SPIs exceeds the threshold of exceptional drought (which is defined by the US drought Monitor system; http://droughtmonitor.unl.edu/AboutUs/ClassificationScheme.aspx) around the year 1998. The CDR developed in this study, as a time series of measurements of sufficient length, consistency and continuity, can also be applied to determine climate variability. Figure S8 in the supplement II, as an example, provides the inter-annual variability of the available water (P-ET) over the globe during the CDR period 1984 -2010. "

Specific comments:
C2: Page 1 - Abstract: The method used is described as 'optimal' or 'optimizing'. Provide further explanation in the body regarding what is meant by 'optimal' in the process used.

R2: It still remains a challenge to find a "best" approach to merge multiple data sources into a single data set over the globe due to limited observational data. This study assumes the deviation from the ensemble mean of all data sources, for the same budget variable, as a proxy for the uncertainty/error in individual products and uses this error information to produce the merged water budget variables, which we considered as an 'optimal' approach. This is also described in the beginning of section 2.2.1:
"There is no best estimate or observation of each individual water budget component at the grid scale over the globe due to the limited spatial coverage of in-situ measurements. This is especially true for evapotranspiration observations from the flux tower networks. Thus, the limited availability of gridded ground observations makes it impossible to quantify the error in each water budget component. Therefore, in this study, the deviation from the ensemble mean of all data sources for the same budget variable is used as a proxy of the uncertainty/error in individual products. The merging procedure for each budget component is a weighted averaging, where the optimal merging weight w_i is given by the following equation…"

C3: Page 2, Line 14: The authors describe the product developed as a climate data record (CDR) that is defined as "a time series of measurements of sufficient length, consistency and continuity to determine climate variability and change". It is not necessarily clear from the description given, though, the extent to which the produced global hydrologic budget product meets this standard. Additional explanation should be included as to the nature and validity of variability and change captured in the data product.

R3: We're sympathetic to the reviewer's comment, and in many ways for any CDR variable being proposed by GCOS this is an issue. Please see R1 (the response to your general comments.) We have added text regarding this issue in section 5. (Discussion and future work) that we feel addresses the comment.

C4: Page 4, line 9: The authors list accounting for the Earth's oblateness as one of the advances in this study. It is not apparent from the rest of the manuscript what precisely this refers to. Is this in reference to the use of a geographic coordinate system rather than a regular square grid? Please include a short description, where relevant, that specifies what is meant by this.
R4 It means when the spatial mean of water budget variable (e.g. the numbers listed in Figure 12) was calculated, it is not simply considered as the arithmetic mean of all the grid cell values but a weighted averaged value based on the area of each grid cell which considers the Earth's oblateness.

C5: Page 5, lines 19-22: Given the proportion of land mass in Europe, Asia, and North America that exists poleward of 50 degrees N latitude, do you expect that using datasets that do not extend beyond that latitude might also account for part of the variation seen for those continents and river basins?

R5: We tried not to mix the fully global datasets with datasets that only span 50S-50N when calculating the continental or basin averages. The data sets are at 0.5 degree spatial resolution and are averaged onto continents in Figure 2 and basins in Figure 3. For the Asia, Europe and North America that exist poleward beyond 50 degree north, the variation seen from Figure 3 are only calculated from the grids between 50N to 50S. This is clarified in the caption of Figure 3. Similarly for basins like Lena, Mackenzie and Yukon that are either above 50N or are across 50N, only the grids between 50N-50S were used. So we believe that the variation seen for those continents and basins come from the datasets that do not extend beyond 50N latitude.

C6: Page 5, lines 22-25: The spread among seasonal precipitation values for the Danube and Mississippi appears larger than that shown in Pan et al, 2012 as a result of the inclusion of the CSU dataset. Do the same potential explanations apply here, and specifically to the CSU dataset, i.e. a more dense gauge network can lead to more variability in resulting product as a result of variable application of undercatch adjustments and gridding procedures. It would be helpful to include a brief note explaining this along with the discussion of PGF, GPCC, and CHIRPS.

R6: The different, larger spread among seasonal values in the Danube (the Mississippi was not displayed in Pan et al. 2012) shown in this study relative to Pan et al. 2012 should be sourced from different data sources applied during the different time period. And yes, "a denser gauge network can also lead to more variability in the resulting product as a result of variable application of under-catch adjustments and gridding

procedures." And this is discussed in the manuscript in section 2.1.1 using as example basins the Danube and Mississippi.
"It is interesting to note that the average discrepancy between the highest estimates (CSU) and the lowest (CHIRPS) over Europe is around 15mm/month throughout the year (Figure 2). This discrepancy is more prominent at basin scales; for example, the monthly mean difference between CSU and CHIRPS in the densely gauged basins such as Danube and Mississippi is around 20 mm/month (Figure 3). CHIRPS is a blended precipitation product (e.g. precipitation climatology, remote sensing from multiple sources, seasonal forecast form Climate Forecast System Version 2 (CFSv2), and in situ observations) but it is dominated by gauge corrections in regions with higher gauge density such as Europe and North America, and therefore in basins such as the Danube and Mississippi. The differences among the three gauge-merged products PGF, GPCC and CHIRPS might possibly be from the different data sources that they merge rather than from gauge observations, different numbers of gauges used and under-catch corrections."

C7: Page 6, lines 15-18: What additional information or value do the cross-combined SRBCFSR, SRB-PGF and PM/PT datasets bring to the overall analysis and assimilation? In other words, what aspects of ET quantification do these combinations of algorithms provide or cover that are not addressed in the other 6 datasets? A brief justification would help clarify this point.

R7: Basically they expand the ensemble of algorithms. As part of the MeaSUREs satellite products, the combined SRB-CFSR, SRB-PGF and PM/PT utilize Surface Radiation Budget as an input but apply different algorithms (i.e. PM and PT) than the other satellite product GLEAM and other reanalysis and modeled products. Satellite observations require retrieval algorithms to estimate the geophysical variable, even though the observations are at fine spatial resolution and have comprehensive coverage. The retrieval algorithms make it possible to estimate the water budget in sparsely gauged regions. For ET, a number of satellite products are needed for the algorithms and were used in this study. This is also described in the text:
"As parts of the MeaSUREs products, the four other satellite products are derived using two algorithms, the Penman-Monteith (PM) and Priestly-Taylor (PT), cross-combined with two forcing inputs that are different from the other six *ET* products, the SRB-CFSR (Surface Radiation Budget – Climate Forecast System Reanalysis) and SRB-PGF. These four products are referred as: SRB-CFSR-PM, SRB-CFSR-PT, SRB-PGF-PM and SRB-PGF-PT (Vinukollu et al., 2011). Satellite remote sensing, carries the mission of observing Earth at fine spatial resolution and comprehensive coverage and makes it possible to estimate water budget in sparsely gauged regions. Therefore, 5 satellite *ET* products are merged into the CDR."

C8: Page 8, Line 11: Does the resampling in space and time introduce additional error or imposed correlation that warrants treatment in the merging and data assimilation process?

R8: Resampling in time is actually an up-scaling (aggregation) process, which sums the high temporal resolution data to the monthly scale. This should not introduce additional error, but should reduce the uncertainty. But for resampling in space, aggregation of those high-resolution (e.g. 0.25 deg for CSU and PGF) onto 0.5 deg might smooth the spatial variability. Nevertheless, the resampling in space and time is a necessary step to organize all the data sources into a uniform spatial and temporal resolution for data assimilation.

C9: Page 9, Section 2.2.2: How was the error calculated for the runoff component? Were all three sources (VIC, CLM, NOAH) used? Please clarify.

R9: We added text in section 2.2.2 to clarify how the runoff error was assumed. "In this study, the error of runoff is simply assumed as 10%, as VIC is the single source of runoff." And we are aware that this number is highly empirical and based on the authors' knowledge and confidence about the model calibration given there is no global grid level (0.5 degree in this study) runoff observations to quantify the error.

R10: Page 10, Section 3.2 and Figure 12: The description of the example water budget constrained assimilation for the Amazon suggests that the precipitation component for the assimilation received the highest non-closure error attribution. If the error covariance for data assimilation is based on the spread of ensemble values for each water budget component (as described in Section 2.2.2), which appears comparatively low (10-20 mm for precipitation compared to >30 mm for ET, based on plots in Figures 3 & 5, respectively), how does this translate to the attributions reported? Perhaps this is obscured by the fact that the plots in Figures 3 and 5 are seasonal averages whereas the water budget closure assimilation is done monthly? Some additional explanation here (Section 3.2) or in the brief Section 2.2.2 would help clarify these sorts of apparent inconsistencies and guide the reader through the process.

R10: Yes, Figures 3 and 5 are seasonal averages but the water budget closure assimilation is done monthly. In order to avoid the intuitive "inconsistence" feeling, one sentence is added at the end of section 2.2.2 to further clarify the water budget closing procedure: "The water budget closure is done monthly based on variational error from month to month."

C11: Page 11 - Line 9-11: Given that human activity can impact long term water storage (multi-decadal groundwater storage decline, filling or removal of dams and reservoirs, etc), it seems that a long-term mean TWSC might not be appropriate in some locations. This assumption needs additional justification. Additionally, how do the authors reconcile the assumption of a long-term zero trend in terrestrial water storage with studies that indicate recent trends in continental water storage (e.g. Reager et al 2016 - 'A decade of sea-level rise slowed by climate-driven hydrology' Science)?

R11: We'd like to thank the reviewer for pointing out this. We struggled with this issue, but the data sources for water management, particularly groundwater extraction and comprehensive reservoir storage changes, are simply unavailable at this time. We agree that at both local and regional scales, some of the places have experienced groundwater depletions such as US high plains and central valley, western Iran, India, etc., starting from different years. But from the global perspective, over the ~three decades covered by the study period 1984-2010, in light of the lack of data the authors assume the long term *TWSC* to be zero and thus apply the de-trending.

R12: Page 13 - Line 6: What is meant by "non-significant correlations" here? What portion of the total was filtered out for comparison?

R12: 33 medium basins (out of 362) and 36 small basins (out of 862) were filtered out by running a test of significance to remove those catchments with non-significant correlations between GRDC runoff observations and CDR runoff records. This was done in order to remove those basins such as Indus and Senegal which might have incorrect observational data. These observational data records that we believe to be incorrect are also discussed in the text:
"Note that the seasonal peaks from Noah and VIC are in agreement for the Indus basin but their peaks precede the peak from the GRDC observations, which strangely happen in November. Comparing to other studies for the Indus River (Bookhagen and Burbank, 2010), show that the discharge peak occurs in the summer time , which is consistent with VIC and Noah. Likewise for Senegal River, records from regional studies (Andersen et al., 2001) and (Stisen et al., 2008) show runoff peaks in August to September instead of April to May from the GRDC record."

C13: Page 12-13, Section 4.1: The authors refer to the developed dataset alternatively as the CDR and the 'MEaSUREs' dataset within this section. Consider revising for consistency and clarity. The comparison of the developed data product runoff against available gage records (Figures 13 & 14) indicates poorer matches in northern regions and in more arid regions. The authors describe potential reasons for the mismatch in northern basins (lake/wetland influences) and the arid southern Africa data points (poor representation of sporadic rainfall and quick runoff). I'm curious if the poorer match in arid and semiarid regions is potentially attributable to unaccounted-for water management activities (which tend to be more pervasive in water-limited regions) or if there is an underlying hydrologic bias specific to those areas.

R13: Thanks for the comment, and we have changed the text to use consistently CDR.
As for the poorer match in arid and semiarid regions, we don't believe they are due to water management because in the validation data sets the potential water management activities were excluded via those criteria mentioned in the first paragraph of section 4.1.
"Basins under any one or more of the following conditions were excluded: (1) GRDC basins for which the catchment boundaries could not be reliably determined; (2) basins with large dams (reservoir capacity greater than 10% of annual streamflow); (3) basins with urban areas greater than 2% (using the ''artificial areas'' class of the map from GlobCover, version 2.3; (Bontemps et al., 2011)); (4) basins with irrigated areas greater than 2% (using the Global Irrigated Area Map; http://www.iwmigiam.org); and (5) basins with either a gain or loss forest (change in land cover) > 20% of the basin area. For both the medium and small basins, those basins with data records length less than 5 years were also excluded." But for those small basins, "though they were filtered in an attempt to remove basins impacted by factors such as reservoirs, irrigation, urbanization, and so forth, they might be impacted by the scaling issues."

C14: Page 16, Line 1: It seems the runoff and TWSC components of this process could be improved to better represent lake/wetland dynamics which are noted as potential aspects of budget mismatch in certain regions. Do future plans entail addressing these issues?

R14: Yes, the current TWSC term from the LSM does not include lake or wetland dynamics. Lake and wetland modeling was first introduced in VIC model for small lakes (smaller than computing pixel) with no river connection (Bowling et al., 2010) and then re-structured for large lakes with river connections (Gao et al., 2011). So far, a number of major global lakes and wetlands (Melton et al., 2013; Bohn et al., 2013) have been modeled, even though the coverage is not complete globally. The plan for the future is to include all major global lakes and all wetlands such that the storage dynamics can be better captured. This is being done under the NASA Surface Water Ocean Topography mission funding. Again, global data sets are not well organized to include these at this time, but we expect to include then in the future.

Reference:

Bohn, T. J., Podest, E., Schroeder, R., Pinto, N., McDonald, K. C., Glagolev, M., Filippov, I., Maksyutov, S., Heimann, M., Chen, X., and Lettenmaier, D. P.: Modeling the large-scale effects of surface moisture heterogeneity on wetland carbon fluxes in the West Siberian Lowland, Biogeosciences, 10, 6559-6576, https://doi.org/10.5194/bg-10-6559-2013, 2013.

Bowling L C and Lettenmaier D P 2010 Modeling the effects of lakes and wetlands on the water balance of arctic environments J. Hydrometeorol. 11 276–95

Gao, H., T.J. Bohn, E. Podest, K.C. McDonald, and D.P. Lettenmaier, 2011: On the cause of the shrinking of Lake Chad. Environ. Res. Lett. 6 034021, doi: 10.1088/1748-9326/6/3/034021

Melton, J. R., Wania, R., Hodson, E. L., Poulter, B., Ringeval, B., Spahni, R., et al. (2013). Present state of global wetland extent and wetland methane modelling: conclusions from a model intercomparison project (WETCHIMP). Biogeosciences, 10, 753-788. doi:10.5194/bg-10-753-2013.

Minor edits: Page 4, line 20: Check the tense(s)

Page 6, line 15: ‘These four products are referred [to] as. . .’ Page 6, lines 15-18: Check sentence for extra words/order

Page 14, line 4-5: Sentence wording a little unclear - consider revising for clarity

Page 20, Table 1: *CLM and NOAH in grey are analyzed but [NOT] merged into . . .

ERA-Interim & MERRA lines - 1979-present (misspelling)

Page 22 - Table 4: Typo - NoethernDvina -> Northern Dvina

Page 39 - Figure 15: Misspelling on plot axis: ‘infered’ -> ‘inferred’

We’d appreciate the review’s time in pointing out these detailed edits which will definitely lead to an improved manuscript. Those minor edits above have already been modified in the manuscript.