Referee #1

In their manuscript "Global terrestrial water storage connectivity revealed using complex climate network analyses", the authors investigate correlation networks which reflect mutual dependencies between the temporal variation of terrestrial water storage at different locations. Estimates of terrestrial water storage (TWS) are based on two data products, the GRACE satellite mission and the NOAH model (GLDAS-NOAH). The network topology indicates different river basins to show different connectivity patterns, pointing to teleconnection structures which seem to be present for some basins (Ganges, Mississippi, Tigris) but not for others (Amazon, Congo, Yangtze). Maps of average node connection length (in Euclidean sense) reveal regions with long connections (such as the Pacific Northwest) and those with short connections (such as the Middle East). Nodes of large area-weighted connectivity (coined "supernode regions" by the authors) seem to reflect combined effects of climate variations and anthropogenic activities. The manuscript provides a network perspective on global terrestrial water storage which is a relevant and timely topic for readers of Nonlinear Processes in Geophysics. Combining insights gained from the network analysis might possibly help to identify TWS predictors and to improve land surface models. The employed methods - network measures as well as network inference by thresholding a matrix of Pearson correlation coefficients - are not new and have been frequently applied in many different scientific disciplines. Applying these methods to GRACE and GLDAS data products establishes the novelty of this work. I have only minor remarks and recommend publication after revision.

Reply: Thank you.

Minor remarks:

1) There is a certain ambiguity when defining network nodes which is natural for such datasets. Nevertheless, how does the spatial resolution of the used data products influence the results of the network analyses? Are the results reported in this study "stable" when network nodes are defined in a different way (e.g. after coarse graining)? How does network structures change when the number of nodes is increased or decreased? I expect the maximum number of reasonably defined nodes of a network to be constrained by the numbers of degrees of freedom captured in the data (which, in case of GRACE, seems to be related to the finite resolution of the measuring instruments).

Reply: These are excellent questions. Before addressing these questions, we first present a background on GRACE data processing done by distribution centers (this information is also available under Section 3 of the paper). The standard GRACE product are sets of spherical harmonic coefficients describing monthly variations of the earth's gravity field. Two filters are typically used to post-process the raw GRACE harmonic coefficients, a de-striping filter that removes correlations between certain spherical harmonic coefficients, and Gaussian averaging filter with a half-width of 300 km that reduces random errors in higher degree spherical harmonic coefficients not removed by ''de-striping'' process; finally, the gridded GRACE product used by this paper is obtained by converting spherical coordinates into geographic coordinates to create mass variations at 1-degree globally [Landerer and Swenson, 2012]. A direct consequence of such a smoothing process is that the pixel-wise correlation and the degree of centrality are increased. To ensure a fair comparison, a standard practice in the GRACE community is to apply the same filtering steps to the model data and, in our case, this is the GLDAS-NOAH data. Therefore, any discrepancy between the networks derived from the two products can be attributed largely to model assumptions and deficiencies. Given this background, our answers to the particular questions are listed below.

How does the spatial resolution of the used data products influence the results of the network analyses?

In general, spatial resolution of a remotely sense product affect the level of details in the calculated measures. The spatial resolution of data products does not affect our comparison results because the two datasets have gone through the same spatial filtering process and considered to have the same "resolution". The spatial resolution is related to the inherent limitation of the GRACE instrumentation, not to any of the network techniques we adopted in this paper.

Are the results reported in this study "stable" when network nodes are defined in a different way (e.g. after coarse graining)?

Yes. The results are stable because the smoothing process essentially reduces the variations/noise and fixes the degree of freedom. We didn't do coarse graining or upscaling in this work; all the pre-processing has been done by NASA/JPL and other distribution centers. Each network node corresponds to a pixel in the dataset. The edge-density-based correlation cutoff threshold is the only parameter used for changing the complexity of networks. Figure 1 of the paper shows that the two products show very similar edge density functions. Because the edge density function is rather smooth, a slight change in cutoff threshold won't affect network. Also, the 300-km filtering radius only introduce local smoothing. The global hotspots are preserved.

How does network structures change when the number of nodes is increased or decreased?

The number of nodes is fixed in this study, which is equal to the number of non-Null, 1-degree pixels given in each dataset.

2) Figure 6 shows GRACE area-weighted connectivity derived from the maximum of the cross correlation functions. Does a corresponding map of the identified lags of the maximum cross correlation reveal any interesting information?

<u>Reply</u>: Accept. We will provide two additional plots in Appendix B of the paper. Figure B1 shows the maximum correlations for six basins chosen in Figure 2 and Figure B2 shows the corresponding phase lags. Recall these plots show the correlation between basin centroid and all other cells in the TWSA dataset. The phase lag plot (normalized by 18) shows that each river basin is in phase with itself and the immediate surrounding regions, but there are significant phase shifts between each river basin and other river basins.

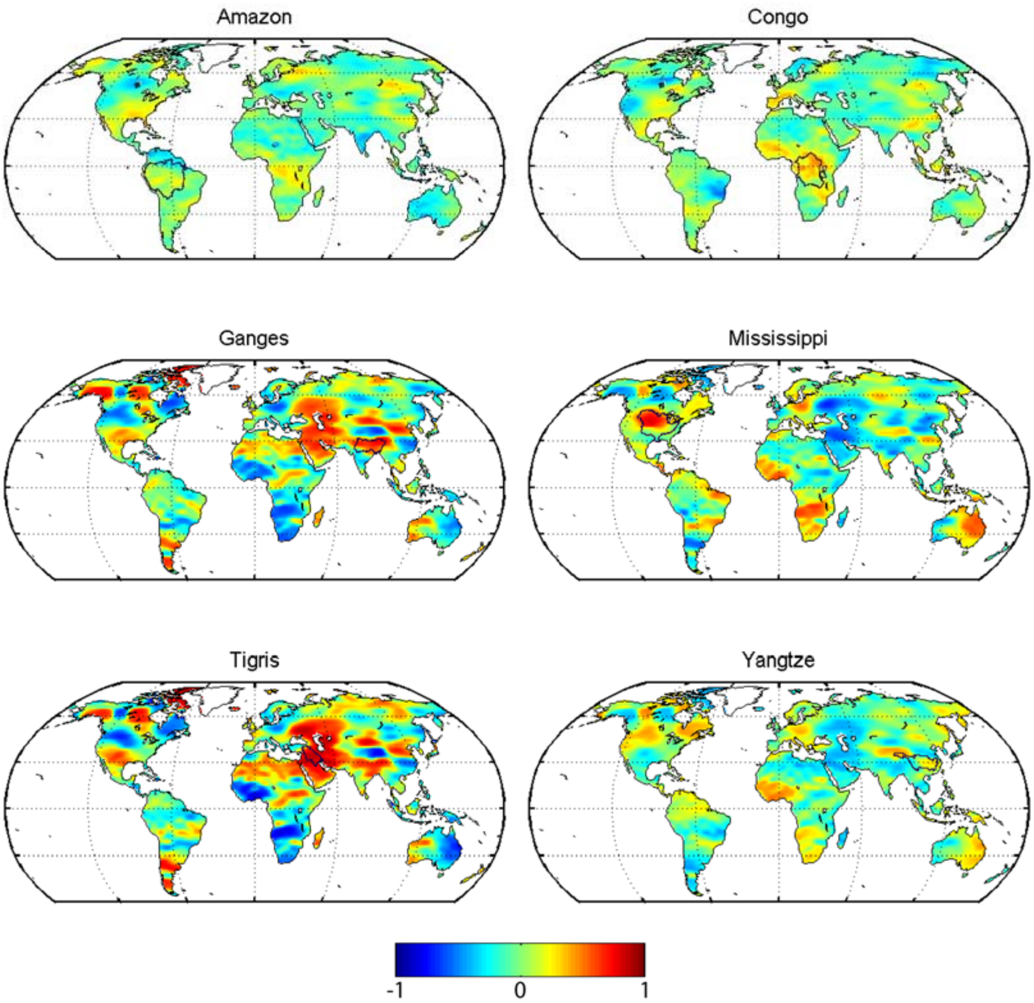

Figure B1. Degree centrality inferred from GRACE TWSA for six river basins, based on the maximum correlation between each basin centroid and all other cells in the grid, and within a window of [-18,18] months.

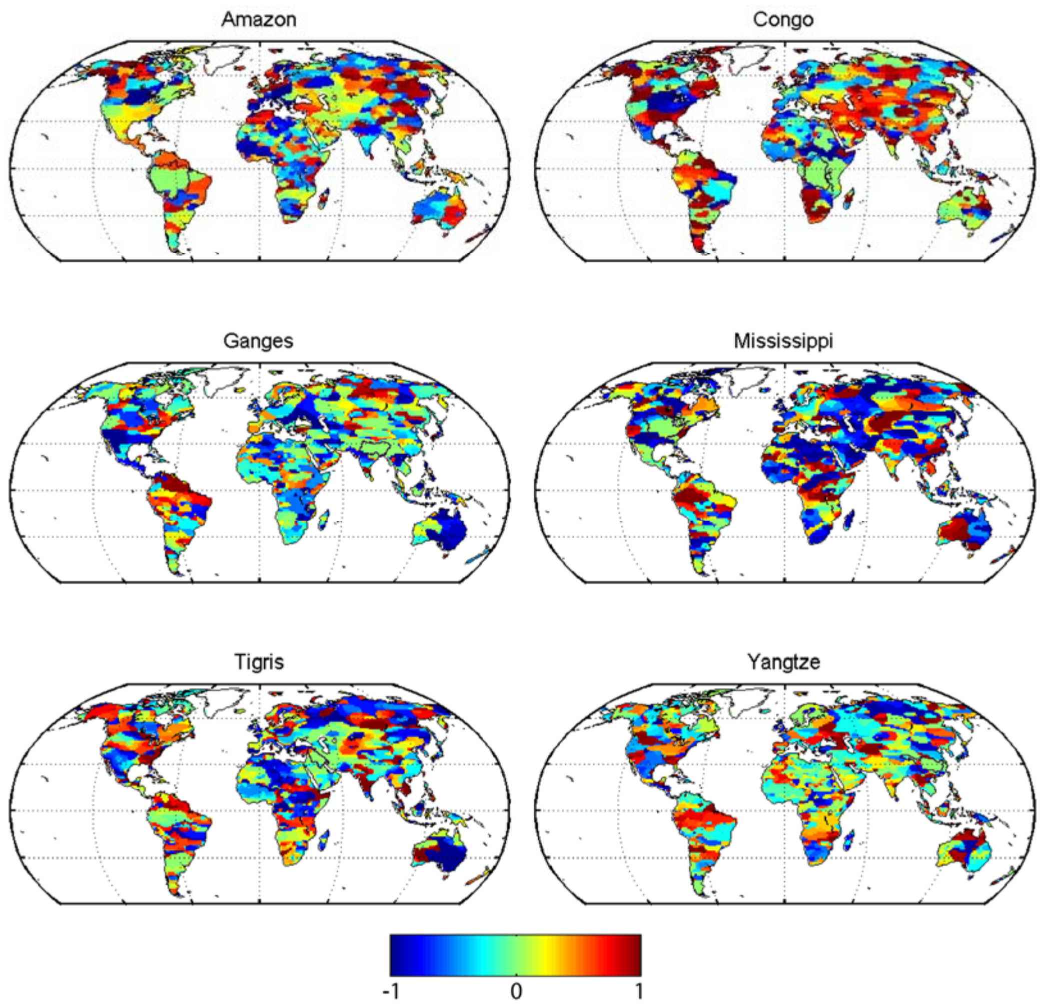

Figure B2. Phase lag of maximum correlations obtained for the six river basins shown in Figure B1 (normalized by 18).

3) p. 788, l. 16: "... \bar{L} provides a measure of network integration." I believe that readers not aware of the spatial aspect of this network measure will profit from a short interpretation of \bar{L}. For the grid-like arrangements of nodes considered in this work, low values of \bar{L} would indicate a grid of nodes which are only locally connected. In a topological sense, this would be a network which is not well integrated since it comes along with a large average shortest path length (which may be somewhat counterintuitive for some network scientists not investigating spatial networks).

Reply: In the context of our work, TWS is connected to the global precip patterns, although TWS is not a flux and does not move as atmospheric fluxes do. Therefore, it is not surprising that most connections are local. Nonetheless, as we mention in the conclusion of the paper, in some regions the connection length distribution is bimodal because of long-range teleconnections (in precipitation). We commented in the conclusion that "*…In terms of connection lengths, the Middle East region is dominated by local connections, whereas regions such as Pacific Northwest, North Central, Colorado River, and North East regions of the US, south Africa, and eastern Australia all have strong bimodal connections*"

4) p. 796, l. 5: "... which are extended for use with gridded datasets." The authors need to explain how exactly they extended the "classic degree of centrality and connection length measures" (l. 4). Is here anything novel that is not yet reported in studies on spatial networks?

Reply: Gridded datasets are special class of spatial networks. The word "classic" is used to distinguish networks that do not require area-weighting. To avoid future confusion, we will rephrase the sentence to "*The constructed networks are further analyzed using the degree of centrality and connection length measures.*"

5) p. 796, l. 19: "... and type of TWS connectivity ...". What do authors mean with "type" here?

Reply: The connection length distribution is bimodal. Loosely speaking, there are two types of connections, the local connection and long-range connection. We will add the following in parentheses after the word type, "*…(i.e., local connection vs. teleconnection)…*"

6) The authors use the notion "complex climate network theory" (CCN theory) at several places in their manuscript. Given that the employed network methods are pretty standard, it is unclear to me what exactly establishes a new theory. I recommend to refrain from using the notion "CCN theory". Instead, authors could use phrases like "applications of complex network theory to climate science" as they already did in their previous publications.

Reply: Accept. We will replace all appearances of "CCN theory" with either "CCN studies" or "CCN applications" in the text.

7) p. 787, l. 10: "... between edge (i,j) ...". Perhaps the authors wanted to write: "... between time series i and j ...".

Reply: Accept. We will replace "…*between edge (i,j)*…" to "…*between time series available at nodes i and j* …"