

Interactive comment on “An interaction network perspective on the relation between patterns of sea surface temperature variability and global mean surface temperature” by A. Tantet and H. A. Dijkstra

A. Tantet and H. A. Dijkstra

a.j.j.tantet@uu.nl

Received and published: 23 October 2013

Version: First Revision

Point by point reply to reviewer #1

23 October 2013

We thank the reviewer for his/her careful reading, the positive assessment and for the
C524

useful comments on the manuscript.

1. p744 l 13f: *Please moderate your statement, as in my opinion you cannot conclude, that the Indian Ocean and North Atlantic determine most of the GMST variability. You only get a hint (for more details see comment p758 l3-10).*

The abstract will be modified, also accounting for the suggestions of Referee # 2 to:

”On interannual-to-multidecadal time scales variability in sea surface temperature appears to be organized in large-scale spatiotemporal patterns. In this paper, we investigate these patterns by studying the community structure of interaction networks constructed from sea surface temperature observations. Much of the community structure can be interpreted using known dominant patterns of variability, such as the El Niño/Southern Oscillation and the Atlantic Multidecadal Oscillation. The community detection method allows to bypass some shortcomings of Empirical Orthogonal Function analysis or composite analysis and can provide additional information with respect to these classical analysis tools. Secondly, the study of the relationship between the communities and indices of global surface temperature shows that, while El Niño/Southern Oscillation is most influent on interannual time scales, the Indian-West Pacific and North Atlantic may also play a key role on decadal time scales. Finally, we show that the comparison of the community structure from simulations and observations can help detect model biases.”

2. p745 l3ff: *“The AMO is the dominant pattern of SST variability in the North Atlantic on decadal-to-multidecadal time scales (Enfield 2001).” This sentence should be placed in line 7 to have the ENSO and AMO sentences together .*

Suggestion will be followed.

3. p748 l22: *"We found that in the worst case the p-value of the 95.*

Will be changed to:

"a p-value of the 5% significance level."

Note that for some reason p-value in our version is written "p value" in the typeset manuscript.

4. p749 l20: *"First-neighbours map provide insight into the connections of all nodes in the network to a selected group of nodes." This sentence is hard to understand. Maybe better: "First-neighbours map provide insight into the connections of one particular node in the network to a selected group of nodes."*

Will be clarified in the revisions as:

"First neighbours maps describe the fraction of nodes belonging to a given group any node in the network is connected to. They are defined as

$$FN_{i \rightarrow G} = \frac{1}{N_G} \sum_{n_j \in G} a_{ij}$$

where G is the selected group of N_G nodes. A node of a first-neighbour map reaches a maximum (minimum) value of 100% (0%) if it is connected to all (none) of the nodes of the selected group of nodes."

5. p751 l8ff: *"The communities are ordered by the total PageRank of their nodes (Brin and Page, 1998), which corresponds to the steady state flow of random*
C526

walkers through the nodes of the community. The larger the PageRank of a community the higher the probability of random walkers to travel through the community." What does this mean exactly? The explanation of PageRank is hard to understand. A more descriptive example would be desirable. Something like: "A PageRank of 50

Will be clarified in the revision as:

"The communities are ordered by the total *PageRank* of their nodes (Brin and Page, 1998). The *PageRank* corresponds to the fraction of random walkers which would flow through the nodes of a community out of a population of random walkers traveling around the network by the links. In a directed network, a large flow of random walkers can arise from the large inward-degree of the nodes they go through but also from the large inward-degree of the nodes pointing to them and so on. However, in our case, the network is undirected so that the flow of random walkers through a node is equal to its degree centrality divided by the edge density. The total *PageRank* of a community is thus given by

$$PR_i = \frac{1}{\rho} \sum_{j \in C_i} d_j = \frac{1}{\rho} \sum_{j \in C_i} \sum_k a_{jk}$$

where C_i is the ensemble of nodes belonging to community i . Consequently, the *PageRank* of a community is a measure of the covariability its nodes."

6. p752 l10ff: *"Such a community cannot be considered as a coherent physical pattern of variability so that, in the context of our study, it would be preferable not to associate nodes of weakly modular regions to any community or to distribute them into several small communities, and this is what the Infomap algorithm does." Why? Please explain! Also p753 l1: "which are not representative of any physical pattern variability." Are the small pattern representative of any*
C527

physical pattern variability? I think as every statistical analysis method you also need in community analysis additional investigations to find out, if a community represents a physical pattern. Please clarify this in the text.

Will be clarified in revision as:

"Such a community cannot be considered as a coherent physical pattern of variability so that it would be preferable not to associate nodes of weakly modular regions to any community. The community detection algorithms we know must distribute every node to a community no matter how modular the network is. Thus, it would be preferable to distribute nodes of weakly modular regions into several small but densely inter-connected communities (even if small means one node, in the case of very weakly modular regions). For our networks, the infomap algorithm was the most able to divide weakly modular regions into small communities. Because these small communities are more inter-connected, they are more likely to represent physical spatial patterns of variability than the broad sparsely inter-connected communities detected by the modularity based algorithms. However, we focus in this study on the dominant patterns. For these reasons, we decided to filter out nodes connected to a fraction of their community smaller than twice the density of the network (i.e. connected to less than $\rho(N_i - 1)$ nodes of the community, where N_i is the number of nodes in the community i) and to remove communities including less than 2% of the total number of nodes in the network.

In the case of the Infomap partition of the H-SST network, 20% of the nodes are removed from their community and 1 community is filtered (Fig. 2b) out of the initial 11 communities (Fig. 2a). When applying the same filtering process to the partitions found by the Multilevel and Leading Eigenvector algorithms more nodes are removed and fewer communities remain. For the multilevel algorithm, 32% of the nodes are removed and the same initial number of 5 communities remain, while for the Leading Eigenvector algorithm, 28% of the nodes are removed and

C528

the same initial number of 7 communities remain.

From these results, we conclude that the ability of the Infomap algorithm to distribute nodes of weakly modular regions into several small but dense communities penalizes its modularity score. On the contrary, the modularity-based algorithms tend to distribute the nodes of these weakly modular regions in a few large but sparsely inter-connected communities which are less likely representative of any physical pattern variability. In the study of a climate network, it is preferable to filter out such weakly modular regions because partitioning them into communities could be misleading."

7. p755 l27f: *"Hence, the community analysis allows to detect more detailed features of SST variability in the climate systems." As every statistical method has pros and cons I am sure that this is not always true. Please do not generalise.*

Will be revised as:

"Hence, the community analysis allowed us to detect more detailed features of SST variability represented in the HadISST dataset."

8. p756 l6f: *"... to filter out the 2 to 7 yr band usually attributed to ENSO." 2 to 7 yr variability is not only ENSO - "... to filter out the 2 to 7 yr band which is strongest influenced by ENSO."*

Suggestion will be followed.

9. p757 l1ff: *"These results indicate that important components of decadal variability are present in the Indian Ocean–West Pacific region and the North Atlantic."*

C529

Please explain why this is the case? Because of the higher PageRank? Then there would be also important components of decadal variability be present in the southern wind-driven gyres community, isn't it?

That is true. The text will be modified in revision as:

"These high PageRanks (or high number of links) are representative of the strong covariability of the nodes in these regions on decadal time-scales indicating that important components of decadal variability are present in the IWP region, the NA and the southern wind-driven gyres communities."

10. *p758 l3-10: Figure 7 is quite important in this study, but additional information are necessary: How large are the regression coefficients? What does the "coefficients of multiple-determination" exactly mean? How do you calculate the time series of community 23 (Just the average over both time series or the average over all nodes of these two communities?)? And most important: From a strong correlation between the time series of the communities and GMST and GLST you cannot state that these two communities explain/determine most of the GMST and GLST variability. You cannot exclude that they are both driven by something else or that GMST/GLST variations drive the SST variations in these regions, as you discuss on p762 l3 – p763 l8. So please moderate your statement here.p758 l15: you should make clear that you do not analyse the ensemble mean.*

To include the regression coefficients and the community names, Fig. 7 and 10 will be replaced by the following Fig. ?? and ?? in the revision.

The text will be clarified in the revision as:

"Fig. 7 represents the linear regressions against the GMST and GLST indices of the time series of the IWP community (#2), the NA community (#3) and of the C530

bivariate time series of the IWP and NA communities (multiple-regression) in a least-square sense. The bivariate fits to the GMST and GLST result in coefficients of multiple-determination (a measure of how much the fitted time series determine the original time series, von Storch and Zwiers, 1999b) of 0.87 and 0.66, respectively. This shows that the patterns of both communities can, together, statistically explain most of the decadal variability of the GMST and the largest part of the decadal variability of the GLST. Using the ENSO time series does not significantly improve the fits. However, this analysis is only statistical and no causality between the time series of the communities and the indices can be stated. Also, the increase in GMST since the 1970s may be explained by the phase synchronization of the time series of the IWP and NA communities, although, once again, the increase of both this index and the time series of the communities could also arise from other factors such as an increased radiative forcing (Fig. 7)."

Also, the following sentence was added to explain how the time series of the communities are computed:

"This spatial averaging is equivalent to projecting the dataset on the community vectors C_j of size N where C_{ij} equals to 1 if node i belongs to community j , 0 instead. Thus, the time series associated with the communities correspond to the expansion coefficients associated with the EOFs."

11. *p759 l13ff: Please explain why you use communities 13. And again please moderate your statement, as you cannot exactly say who the driver is.*

Please see point 10 where statement is moderated. The revision will be clarified as:

"Decadal variability in the GLST within this ESM can be largely explained (statistically) by the ENSO and IWP communities (Fig. 10), but misses the

connection to the NA as found in the observations (Table 4). Contrary to fig. 7, the fits fig. 10 where done without the time series of the NA community, since the regressions to the indices were not significantly improved using this time series."

12. p759 l14f: ". . . and IWP communities, but misses the connection to the AMO as found in the observations (Table 4 and Fig. 10)." => ". . . and IWP communities (Fig.10), but misses the connection to the AMO as found in the observations (Table 4)."

Suggestion will be followed.

13. p759 l18: "These model biases could be efficiently identified thanks to the network approach." Please be more neutral. e.g. "using the network approach"

Will be changed in revision to:

"These model biases could be efficiently identified using the network approach."

14. p760 l17: "Finding communities in a network is a much more efficient way to reveal non-overlapping spatial patterns of variability of the global climate system than an EOF analysis." I think you cannot generalise this statement, as every statistical method has pros and cons.

Will be changed in revision to:

"Detecting communities in the HadISST and MPI-ESM networks was more efficient in revealing non-overlapping spatial patterns of SST variability than EOF analysis."

C532

15. p760 l26ff: "... with the second community (2) could also be associated with the first community which was also suggested by Guan and Nigam (2009)". => "... with the second community (AMO) could also be associated with the first community (ENSO) which was also suggested by Guan and Nigam (2009)". Makes it easier to follow.

Suggestion will be followed.

16. p761 l12ff: "In short, thanks to the community detection algorithm, ..." Again, please be more neutral.

The revision will be changed to:

"In short, using community detection algorithms, neither prior information nor a choice of a pattern of variability to be studied is necessary, making pioneering studies easier and maybe also less biased."

17. *Technical corrections.*

All technical corrections will be revised.

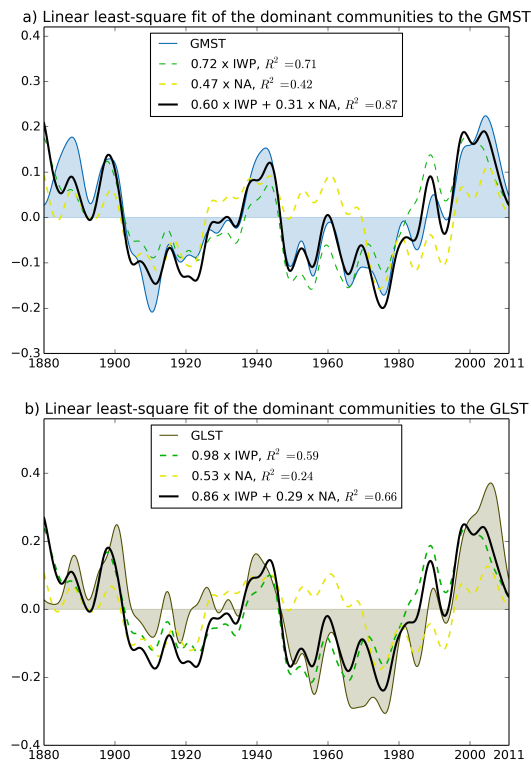


Fig. 1. Regression of the mean time series of the IWP and NA communities of the H-SST-LP8y network to the (a) GMST and (b) GLST.

C534

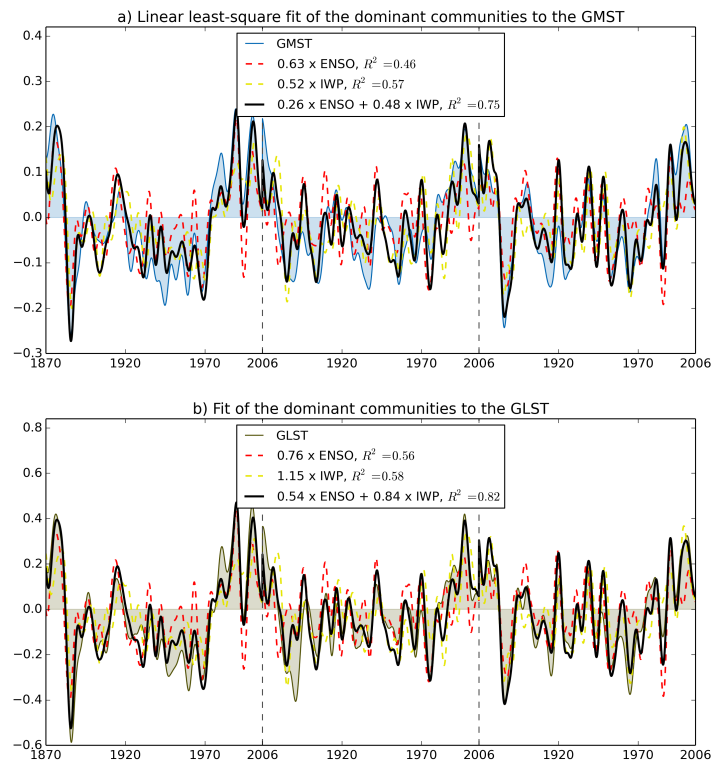


Fig. 2. Regression of the mean time series of the ENSO and IWP communities of the ESM-LP8y network to the (a) GMST and (b) GLST.

C535