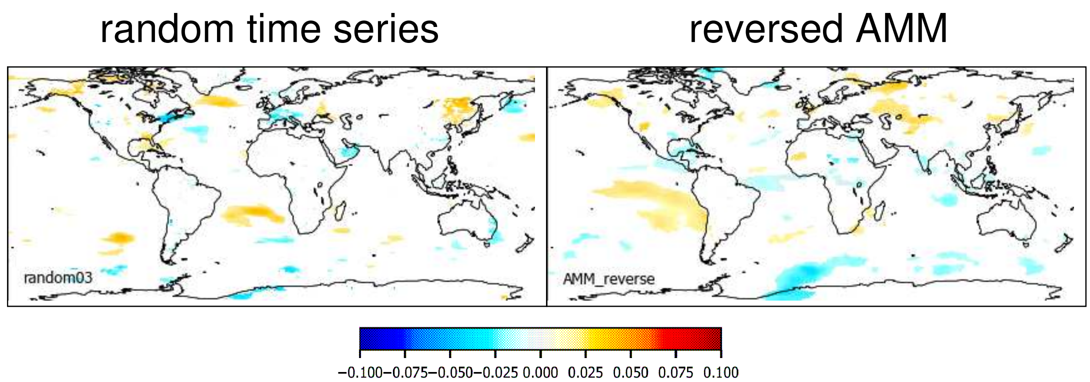
Response to Reviewer 2

We thank the reviewer for the constructive and helpful remarks suggestions. We followed most of these suggestions as described below. Before we give our detailed answers to the individual comments, we first summary the most important changes to the manuscript.
The following major changes were made to the revised version of the manuscript:

A) The motivation and description of the empirical method was improved.

We agree with both reviewers that the description of new method had errors and was complicated. It was also not well motivated.
In the revised version we shifted many technical details of the new method to the appendix (e.g. the description of our normalisation approach as new appendix A1, or the investigation of the effect of time shifts as new appendix A3).
We also added more details about the fit function to section 4, and more details about the calculation of the temporally reversed indices to section 5.1.
For the motivation of our new approach we added the following information to section 5 (see also new Fig. 6):
'... Information about the significance of the fit results can be obtained from the fit function itself. However, in practice, the significance information from the fit has several limitations:
a) The determination of the significance is based on several assumptions about the data sets, e.g. that all data points of the time series have the same uncertainties and follow a normal distribution. However, the errors of the individual data points can be very different, e.g. the effect of clouds on the errors of the satellite TCWV data set can be very different for different seasons and regions. Also, the uncertainties are not only random but contain also systematic contributions. It is difficult (if not impossible) to quantify the uncertainties of the involved time series.
b) The determination of the significance is based on prescribed significance levels. The choice of such a significance level is arbitrary and the obtained significance information depends on this choice.
c) In several tests we fitted artificial time series to the TCWV data set. These tests showed that even for such non-geophysical time series 'significant' fit results can be obtained (see the examples in Fig. 6). On the left side of this figure, fit results for a time series containing only white noise, and on the right side fit results for a temporally reversed teleconnection index are shown (the temporally reversed index is obtained from the original index by mirroring the time axis). The blue and red areas show fit coefficients for both time series, which are classified as significant by the fit.
Based on these findings, we conclude the use of the significance of the detection of an index derived from the fit itself is not straight-forward.
To address these difficulties, we developed and applied an empirical approach to determine threshold values for the delta RMS values to decide whether an index is significantly detected in a global data set. The new

procedure is described in the next section. It has the following two main advantages:
-the threshold values are determined empirically. Thus no assumptions on the properties of the time series or the significance levels have to be made.
- the method provides a clear procedure and in particular a metric which can be applied in a consistent way to different data sets and thus allows a quantitative comparison (see section 6).'



New Fig. 6 Global maps of the fit results for an artificial time series containing only white noise (left) and a temporally reversed teleconnection index (AMM, right). The white areas represent fit results, that are classified as non-significant by the fit routine (for a 5% significance level).

B) The Scope and aims were made more clear.

Probably one important misunderstanding was that we gave the impression to the reader that we aim to investigate the influence of teleconnections on the TCWV. This has probably even led to the expectation that we could predict monthly TCWV using teleconnections. This was not our intention. To make this more clear in the revised version of the manuscript, we removed the term 'influence of...', in all parts of the text. And in the introdiction we added the following information: 'Here it should be noted that we do not aim to identify causal relationships or even to predict the TCWV based on teleconnection indices' At the end of the introduction, we added more details and explanations to our research goals. We also restructured the conclusions accordingly and provided respective answers to the research questions formulated in the introduction. The following modified text was added to the conclusions:
'....Based on the obtained results, we could derive the following main conclusions related to the science questions mentioned in the introduction:
a) We developed a new empirical approach to determine whether a teleconnection index is significantly detected in a global data set. This approach avoids problems of existing algorithms for the determination of significance, because no assumptions on the significance level or the measurement uncertainties have to be made. We applied the new method to a global data set of the TCWV derived from satellite observations and found that 40 teleconnection indices could be significantly detected.
b) We applied the same method also to TCWV from the ERA interim data set. Here we used two versions of the model data sets, one including all data, the other only clear sky data. The results for both versions agree in general very

well with those for the satellite data set. This confirms both the quality of the satellite and model data sets. It also indicates that the satellite observations can be seen as representative for all day mean values. For some teleconnections, however, also systematic differences, mainly over northern Africa, were obtained. Since these differences are not found for the majority of the teleconnection indices, we conclude that they are very probably not related to systematic errors of the satellite data set, but rather indicate shortcomings of the model over these regions.

c) We also applied our method to a variety of other data sets, which are usually used in teleconnection studies (surface temperature, surface pressure, geopotential heights and meridional winds at different altitudes). For most of these data sets less teleconnection indices were significantly detected than for the TCWV data sets, while for zonal winds, more teleconnection indices (up to >50) were significantly detected. These results indicate that our global TCWV data set is well suited for teleconnection studies. In our view, this is an important aspect, because our data set is exclusively based on measurements. The strongest teleconnection signals were detected for the data sets of tropospheric geopotential heights and surface pressure. This finding is consistent to the fact that most teleconnection studies are based on these quantities. Another interesting finding is that in none of the global data sets, non-teleconnection indices (like the solar variability, the stratospheric AOD or the hurricane frequency) were significantly detected.

d) We investigated the spatial distribution of the teleconnection patterns. In particular we calculated global maps for the cumulative effect of all teleconnection patterns. For that purpose we first orthogonalised the teleconnection indices to avoid the effect of correlation between the indices. Compared to the original set of indices, much less of the orthogonalised indices (20 compared to 42) were significantly detected in the TCWV data set. Our global map of the cumulative effects of all significantly detected orthogonalised teleconnections showed the strongest teleconnection signals in the global TCWV data set over the Tropics and in polar regions. These spatial patterns point to importance of different driving mechanisms in different regions.'

C) The relationships between different indices and the motivation for the orthogonalisation of the indices was made more clear:

We added new columns in Table 2 (see below). We now show separate columns for indices similar to ENSO, polar atmospheric indices, MJO indices, as well as other oceanic and atmospheric indices.

Indices similar to ENSO (7)	**Other oceanic indices (16)**	**Atmospheric polar indices (8)**	**MJO indices (15)**	**Other atmospheric indices (8)**	**Others indices (7)**

BEST N34 TPI ONI ENSO N4 IND	HAW PDO PMM N1 TNI NTA TNA WHWP IPO CAR AMO DMI AMM STA TSA EA_ersst	SCA AAO EAWR NAO EPNP AO PE WP	MJ1 MJ2 MJN VPM1 VPM2 VPMN RMM1 RMM2 RMMN OOMI1 OOMI2 OOMIN FMO1 FMO2 FMON	PNA SOI NOI EA QBO Q30 Q50 Q70	Solar indices: RI MGII SWO S107 AP HUR (hurricane frequency) SAOD (stratospheric AOD)

In section 3 we added the following explanation:
'Many of these indices (describing the same phenomenon), but also many of the other teleconnection indices are highly correlated. The strength of these correlations is presented in Fig. 3 as a matrix with correlation coefficients between the different indices (after the seasonal cycles were removed). In spite of the correlations amongst the teleconnection indices, we decided as a first step to include them all in our study, because beforehand it is not clear which index might be best suited to represent a teleconnection phenomenon. Using our empirical approach, however, it becomes possible to quantify the significance and strength of the different indices and thus to select the best suited index for a given teleconnection phenomenon. Finally, we apply an orthogonalisation for the most significant indices (see section 7) to minimise the effect of the correlations and to identify the dominant temporal teleconnection patterns in our TCWV data set.'

To better motivate the orthogonalisation, we modified the respective information in section 7 to:
‚To account for correlations between the different indices, we thus applied an orthogonalisation approach. For the orthogonalisation (based on the Gram–Schmidt process), all ‚significant' original indices and significant temporal derivatives (see Figure A11) were considered (in total 57 indices). The order of indices used in the iterative orthogonalisation process was from highest to lowest p99 values. The result of the orthogonalisation approach is a set of modified teleconnection indices, which shows zero correlation amongst each other (for the considered time period). Thus this new set of orthogonalised indices can be used to determine the number of independent significant teleconnection patterns in the global water vapor data sets. We applied our new method to the new set of orthogonalised indices to test which of the modified indices have p99 values above the significance threshold.'

D) The logical flow of the paper and the appearance was improved.
As mentioned above several technical parts were shifted to the appendix. The science questions were better motivated in the introduction, and the corresponding answers were added to the conclusions.
Several Figures were shifted/deleted/modified:
-Fig. 3 was shifted to the appendix
-Fig. 8 was shifted to the appendix
-the upper part of Fig. 9 was shifted to the appendix
-Figs A1 and A2 were deleted as suggested
-the quality of Fig. A4 was improved and the number of the sub-figures was largely reduced (by a factor of 3)
-the quality of Fig. A9 was improved

E) We added a new sub-section (6.1) for the comparison of the spatial patterns of the measured and simulated TCWV.
While for most teleconnection indices very good agreement of the spatial patterns is found between the measuremed and simulated TCWV, for some indices also substantial differences are detected. These differences can point to shortcomings in either the satellite or model data sets (or both) and might be helpful for corresponding improvements.
We added a new Fig. 8 (see below) and the following new text:
'For most of the teleconnection indices, very similar spatial patterns are found in the TCWV data sets obtained from satellite or ECMWF data (see Fig. A9). This confirms both the high quality of the satellite measurements and model simulations. However, for some indices, also substantial differences are found (see Fig. 8). The most obvious differences are found over northern Africa. In principle, they could be caused by errors of both the satellite or model data sets. However, since very good agreement over northern Africa is found for most of the indices, we can very probably exclude systematic measurement biases (like e.g. effects from the high surface albedo over the Sahara). Thus we conclude that the observed differences probably indicate deficiencies in the model simulations, possibly related to the sparseness of observational data over northern Africa used in the model. It is interesting to note that the differences are found for both oceanic and atmospheric indices which have rather different frequencies. These comparison results might help to improve the model performance over norther Africa (and to a lesser degree also over other regions).'

New Fig. 8:

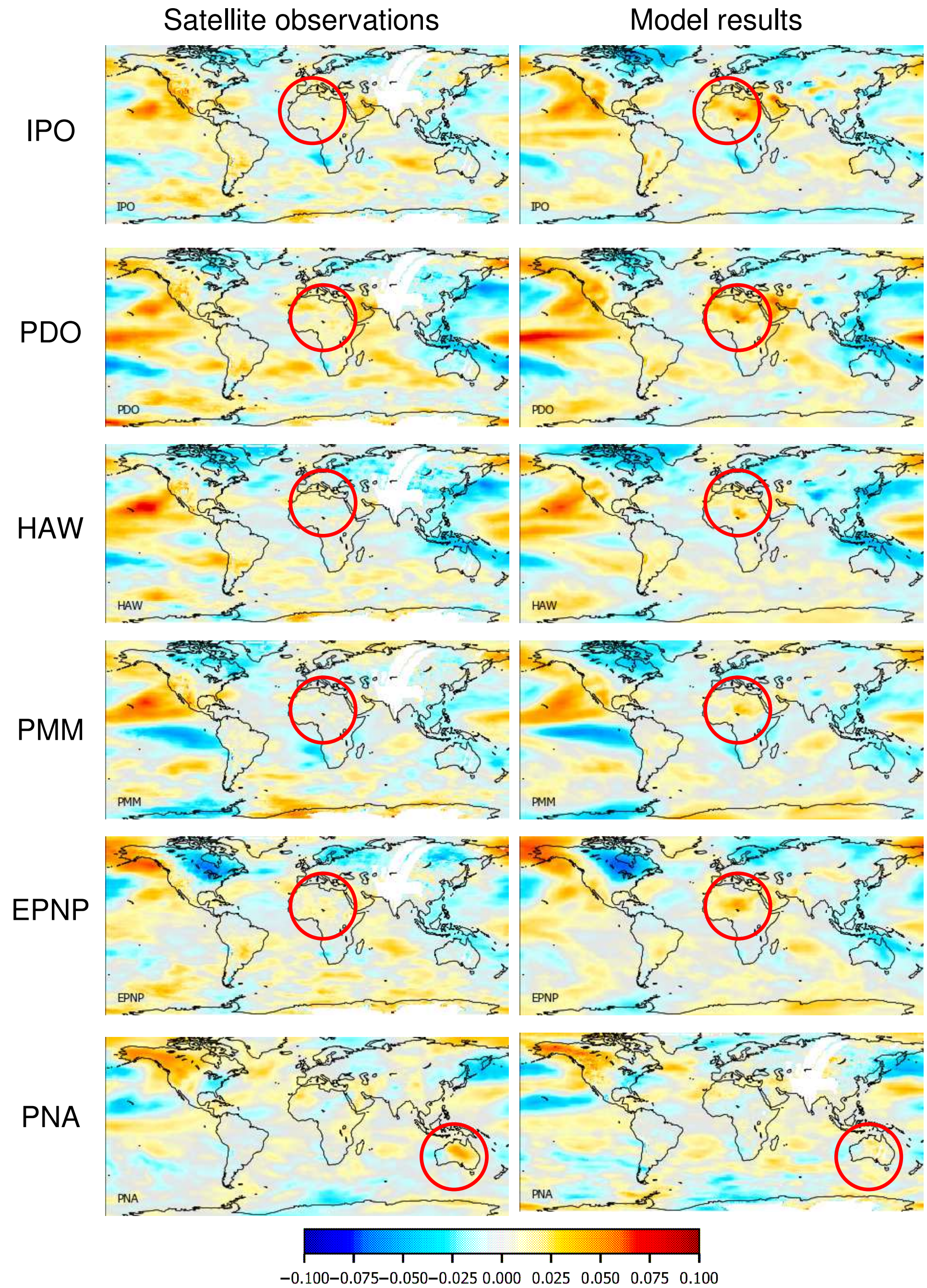


Fig. 8: Fit coefficients for selected teleconnection indices, for which different patterns were found in the TCWV data set from satellite observations (left) and model simulations (right). The red circles indicate regions with substantial differences between the results for both data sets.

More details about the changes are given in the individual replies to the Reviewer comments below:

Recommendation:
General comments: The authors are attempting to show how total column water vapor (TCWV) can be used to reveal the presence of atmospheric teleconnections seen in other datasets. This method is certainly interesting and could be of value, at least in the context of demonstrating the utility of TCWV in revealing existing teleconnections. However, the presentation in this paper was extremely difficult to follow as the authors jumped from one analysis to another with no clear direction as to why. There were many different technical approaches employed within this study, and while these likely have value in the context of what the authors' research goals are, the reasons for using the methods they employ were not well established.

Author reply:
We are thankful for this feedback and agree that our paper was partly difficult to read. We applied major restructuring and added missing information, see points B and D) above.

Reviewer comment:
Further, the authors state early in the study that they are going to compare the results with similar results from pressure, temperature, etc. fields more traditionally utilized in teleconnection studies. I did not see these comparisons.

Author reply:
It seems that here was a misunderstanding. The comparison to other data sets was one of the main aims of our study. The comparison results are shown and dicussed in section 6. We added more explanations for the findings of the comparisons to this section:
'For the TCWV data sets, surface temperature and pressure, as well as most of the zonal winds, the largest p99 values are found for indices similar to ENSO. For the TCWV data sets and surface temperature, this can be expected, because the ENSO phenomenon is driven by the surface temperature (over the tropical Pacific). Accordingly, also the TCWV data sets will be strongly affected, because the TCWV depends strongly on the temperature in the lowest atmospheric layers. The strong influence of the ENSO phenomenon (BEST index) on the zonal winds at most levels can probably be explained by the fact that large scale phenomena like ENSO can have a strong influence on the quasi-persistent zonal flow patterns in the tropics and sub-tropics. For the geopotential heights and meridional winds, the largest p99 values are found for the polar atmospheric indices (mostly AAO, but also SCA). For the geopotential heights this might be expected because the polar atmospheric indices are defined based on anomalies of the geopotential heights. Why also for the zonal winds, the largest p99 values are found for the polar atmospheric indices is, however, is not clear to us.'

We added also a comparison of the maximum p99 values to section 6 (we also added a new column to table 3). The respective text in section 6 is:
'Our new method for the determination of the significance level also allows a direct comparison of the strengths at which the different indices are detected in the different data sets. In Table 3 also the maximum p99 values of the delta

RMS normalised by the corresponding significance threshold values are shown. The highest normalised p99 values are found for the geopotential heights (except the 50hPa level) and the surface pressure. This finding is consistent with the fact that these quantities are used in most teleconnection studies and many indices are even defined using these quantities. The lowest normalised p99 values are found for zonal winds, for which also the smallest numbers of significant indices are obtained. Intermediate values are found for the water vapor data sets.'

Reviewer comment:
In general, the authors focused too heavily on the significance of the relationship between their empirical estimates of the TCWV using the teleconnection index and the TCWV itself. It read more like a study attempting to predict monthly TCWV using teleconnections, not a study linking TCWV to teleconnections. Either the study should be reframed in that context or the authors need to do a better job of linking their results back to the teleconnections they are trying to predict.
Which teleconnections were predicted well? Which were predicted poorly? Why? Such discussion was absent from this study and seems directly relevant to the research objectives outlined therein.

Author reply:
We are sorry that we gave the wrong impression here. As mentioned in point B) above, our aim was not to investigate the influence of teleconnection on the TCMV or to predict monthly TCWV using teleconnections. As described in point B) above, this was made more clear in many parts of the manuscript. In the introduction we modified the respective sentence (Page 2, line 66-67) to:
‚In this study we investigate to which extent the temporal patterns of various teleconnections can be identified in the global distribution of the total column water vapor (TCWV).'

Specific comments:

Reviewer comment:
Most of the work done in PCA-based teleconnection studies in pressure/geopotential height is confined to midlatitude and Arctic regions in the Northern Hemisphere owing to the barotropic conditions in the tropical latitudes. This should be better specified by the authors.

Author reply:
This information was added to the introduction.

Reviewer comment:
If multiple indices characterizing the same phenomena exist (e.g. MJO, ENSO), why include them all? How do you reconcile the differences in how those indices are characterizing their teleconnection and relate those differences back to your results? (Lines 135-137).

Author reply:
The same aspect was also mentioned by the other reviewer, and we tried to make our motivation and strategy more clear in the revised manuscript:
Overall our procedure should be seen as a two step approach: in the first step all available indices are used, because it is beforehand unclear, which of them are most significantly detected in the TCWV data set. But by applying our method to all indices, we can answer the question which indices are most signifcantly detected.
In a further step we then apply the orthogonalisation to obtain a new set of indices without any correlation amongst them.
To make our aims and the procedure more clear, we added the following information to the section 3:
'Many of these indices (describing the same phenomenon), but also many of the other teleconnection indices are highly correlated. The strength of these correlations is presented in Fig. 3 as a matrix with correlation coefficients between the different indices (after the seasonal cycles were removed). In spite of the correlations amongst the teleconnection indices, we decided as a first step to include them all in our study, because beforehand it is not clear which index might be best suited to represent a teleconnection phenomenon. Using our empirical approach, however, it becomes possible to quantify the significance and strength of the different indices and thus to select the best suited index for a given teleconnection phenomenon. Finally, we apply an orthogonalisation for the most significant indices (see section 7) to minimise the effect of the correlations and to identify the dominant temporal teleconnection patterns in our TCWV data set.'

In section 7 the explanation was extended to:
'To account for correlations between the different indices, we thus applied an orthogonalisation approach. For the orthogonalisation (based on the Gram–Schmidt process), all ‚significant' original indices and significant temporal derivatives (see Figure A11) were considered (in total 57 indices). The order of indices used in the iterative orthogonalisation process was from highest to lowest p99 values. The result of the orthogonalisation approach is a set of modified teleconnection indices, which shows zero correlation amongst each other (for the considered time period). Thus this new set of orthogonalised indices can be used to determine the number of independent significant teleconnection patterns in the global water vapor data sets.'

and in section 8:
‚The cumulative delta RMS map for the orthogonalised indices represents the overall contribution of teleconnections to the variability of the global TCWV distribution.'

Reviewer comment:
In the fit functions, how were the quantities *c* and *b* determined? Were they based on a fit with the satellite data, the ERA, etc.? Nothing is provided in the text in this regard.

Author reply:
We checked the explanation of the quantities used in the fit function and added some more explanation. The definition of the involved quantities should now be more clear.

Reviewer comment:
The authors discuss the use of “reversed datasets” in section 5.1. However, they provide no discussion of what was reversed. Was it just the teleconnection time series? Was it the TCWV time series? Were they reversed in time? Did you just reverse the index numbers directly, as is done frequently in pattern recognition and database type work? I don’t see why, if the reverse was temporal, why the correlations didn’t simply change sign but remain the same magnitude. The authors need to provide a lot more explanation on this aspect of their study as they do not really describe it in much detail. Why did you do this?

Author reply:
Obviously our explanation of the details was not sufficient here. We added more explanations here, see also point A) above.
And in section 5.1 the following clarification was added:
’...Information about the significance of the fit results can be obtained from the fit function itself. However, in practice, the significance information from the fit has several limitations:
a) The determination of the significance is based on several assumptions about the data sets, e.g. that all data points of the time series have the same uncertainties and follow a normal distribution. However, the errors of the individual data points can be very different, e.g. the effect of clouds on the errors of the satellite TCWV data set can be very different for different seasons and regions. Also, the uncertainties are not only random but contain also systematic contributions. It is difficult (if not impossible) to quantify the uncertainties of the involved time series.
b) The determination of the significance is based on prescribed significance levels. The choice of such a significance level is arbitrary and the obtained significance information depends on this choice.
c) In several tests we fitted artificial time series to the TCWV data set. These tests showed that even for such non-geophysical time series ‘significant’ fit results can be obtained (see the examples in Fig. 6). On the left side of this figure, fit results for a time series containing only white noise, and on the right side fit results for a temporally reversed teleconnection index are shown (the temporally reversed index is obtained from the original index by mirroring the time axis). The blue and red areas show fit coefficients for both time series, which are classified as significant by the fit.
Based on these findings, we conclude the use of the significance of the detection of an index derived from the fit itself is not straight-forward.
To address these difficulties, we developed and applied an empirical approach to determine threshold values for the delta RMS values to decide whether an index is significantly detected in a global data set. The new procedure is described in the next section. It has the following two main advantages:

-the threshold values are determined empirically. Thus no assumptions on the properties of the time series or the significance levels have to be made.
- the method provides a clear procedure and in particular a metric which can be applied in a consistent way to different data sets and thus allows a quantitative comparison (see section 6).'

We also added information how the delta RMS values compare to the r^2 values at the end of section 5.1 (see also new Fig. A8):
'It should be noted that instead of the delta RMS values, also the correlation coefficients between the considered data set and the fit function (eq. 1) might have been used since the spatial patterns of both quantities are very similar (see Fig. A8).'

New Fig. A8:

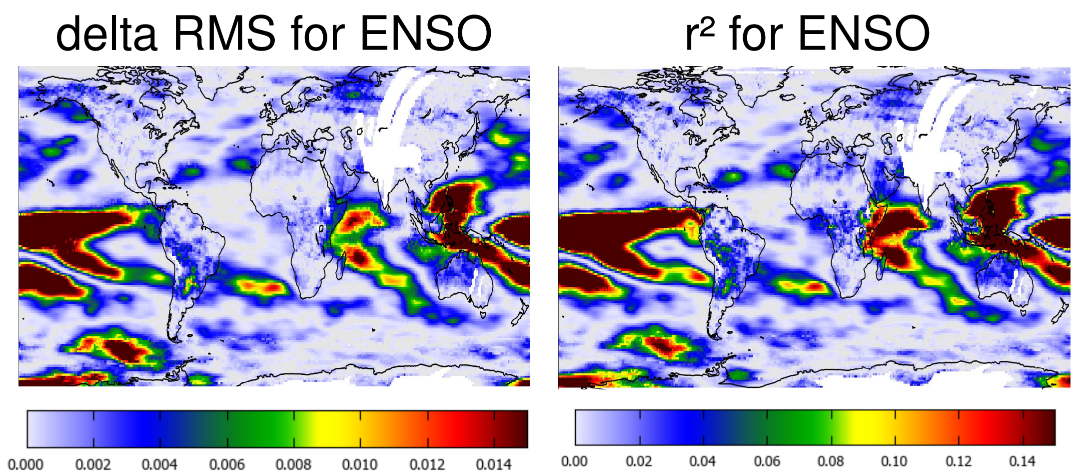


Fig. A8: Delta RMS (left) and r^2 values (right) for the fit of the ENSO index to the TCWV derived from satellite observations.

Reviewer comment:
In section 8 the authors state they "orthogonalized" their indices. What method was used to do this? Why did they do this?

Author reply:
To better explain why we applied the orthogonalisation, we modifed and extended the information given in section 7. We also added the information of the orthogonalisation technique:
'To account for correlations between the different indices, we thus applied an orthogonalisation approach. For the orthogonalisation (based on the Gram–Schmidt process), all ‚significant' original indices and significant temporal derivatives (see Figure A11) were considered (in total 57 indices). The order of indices used in the iterative orthogonalisation process was from highest to lowest p99 values. The result of the orthogonalisation approach is a set of modified teleconnection indices, which shows zero correlation amongst each other (for the considered time period). Thus this new set of orthogonalised

indices can be used to determine the number of independent significant teleconnection patterns in the global water vapor data sets. We applied our new method to the new set of orthogonalised indices to test which of the modified indices have p99 values above the significance threshold.'

Technical corrections:

Reviewer comment:
The e.g. on line 51 can be removed.

Author reply:
Deleted

Reviewer comment:
What is a "time series like index"? (Line 78)

Author reply:
We replaced ‚like' by ‚such as' to make the meaning more clear.

Reviewer comment:
In Figures A1 and A2, are the times over which these averages were computed the same 1995-2015 time period? The ERA have a longer period of record so it would be good to specify this.

Author reply:
Both figures were deleted as suggested by the other reviewer.

Reviewer comment:
It is not clear why Figure A3 is included in the text. There are too many time series and their individual value in the study is not clear.

Author reply:
This figure was included for two reasons:
a) to add information about the sources of the different indices
b) to show the temporal patterns for the considered time period. This information is interesting for two reasons. First, the ‚frequency' of an index can be directly recognised. Second, similarities in the temporal patterns can be easily seen.
For these reasons we decided to keep this figure in the manuscript.

Reviewer comment:
Figure A4 is almost impossible to read. There should be a compelling reason why this figure is included in the text as it includes well over 200 maps. The authors should choose which of those figures best illustrate their point and include those instead of including them all.

Author reply:

We agree that there are too many sub figures. And we want to apologise for the rather bad quality. In the revised manuscript we reduced the number of sub figures by a factor of 3 and improved the quality of the figure. We would like to keep this figure, because the global maps reveal many details of the spatial patterns found for the individual indices. It might be interesting for future studies to compare these patterns to similar results of their own analyses.