Workshop on Visualisation in Environmental Sciences (EnvirVis) (2018)
R. Bujack and K. Rink and D. Zeckzer (Editors)

Interactive Visual Exploration of Teleconnections in Atmospheric Datasets

Anatoliy Antonov[1], Gerrit Lohmann[2], Monica Ionita[2], Mihai Dima[3], and Lars Linsen[1,4]

[1] Jacobs University, Bremen, Germany
[2] Alfred Wegener Institute, Bremerhaven, Germany
[3] University of Bucharest, Romania
[4] Westfälische Wilhelms-Universität Münster, Germany

Abstract

Traditional analyses of geoscientific data and their features require a lot of manual scripting to organize various tools and software libraries. We present a tool developed to cover the typical workflow of the task of analyzing dependencies between regions of the climate system. We propose an interactive visual analysis tool that uses a series of automated analysis steps combined with an interactive visual exploration of patterns in the data. We base our analysis on the statistical approach of strongest negative correlations for discovery of teleconnections. The interactive visual analysis process uses selecting, highlighting, and filtering in four coordinated views. The views represent teleconnectivity information within a teleconnectivity map and a teleconnectivity links list as well as correlation information within a correlation map and a projection view of the correlation space. We apply the tool to different datasets to demonstrate its capabilities for the analysis and comparison of correlation patterns.

1. Introduction

Teleconnections represent relationships between different regions in the climate system such that a certain condition in one region sets an expectation for the condition in the connected regions. For example, increase of pressure at one location may correspond to decrease of pressure at another. Various studies identified a number of prominent patterns. It is of interest how well climate simulations reproduce these patterns, and how much the configuration of the patterns change over time.

In a typical scenario, analysis is performed with scripts written for a particular situation, where each change in parameters requires restarting the script. It makes exploration of the data slow and cumbersome. This script-based approach lies in the foundation of web-based tools for data post-processing [PSD, KNM, Cli]. To the best of our knowledge, there exists no standard software addressing a common workflow for teleconnection research in an interactive and intuitive way, which is also reported in a recent survey on visual analytics of climate networks by Nocke et al. [NBD*15].

We present an interactive visual analysis tool to explore teleconnectivity in climate data. It is based on four coordinated views that encode visually different aspects of teleconnectivity and correlation. Interaction mechanisms allow for a user-centric intuitive analysis process. We provide examples of applying the tool to the data from the 20th Century Reanalysis Project and the respective time period extracted from the COSMOS simulations of the last millennium, see Section 7.1. We demonstrate the capabilities of the tool for the comparison of datasets, the extraction of patterns, and the study of the pattern's components.

We employ one of the main methods of analyzing patterns in the atmosphere, which is based on the use of the correlation coefficient [WG81]. It defines teleconnectivity for a point as the absolute value of the most negative correlation between the time series of this point and others. Plotting teleconnectivity values for all points together in a map provides a way to see the strongest centers of teleconnectivity and understand their relationships with each other. The one-point correlation maps for these centers reveal the patterns.

In addition, we provide an assessment for the quality of data. Typically, the mean of a climate model output is compared to observations. Teleconnectivity and relevant data structures can serve as another, more in-depth validation method. From this point of view, the work demonstrates the bridging of the gap between disciplines, namely, how modern methods of software development can be used to respond to a specific geoscientific question and create a more convenient, powerful, and easy-to-use solution. In terms of perspective, it is similar to the EDEN system [SRS*13]. However, our system has a particular focus on relationships between time series data.

2. Teleconnection Patterns

We provide two examples of teleconnection patterns as a reference to the subject of our studies.

DOI: 10.2312/envirvis.20181131

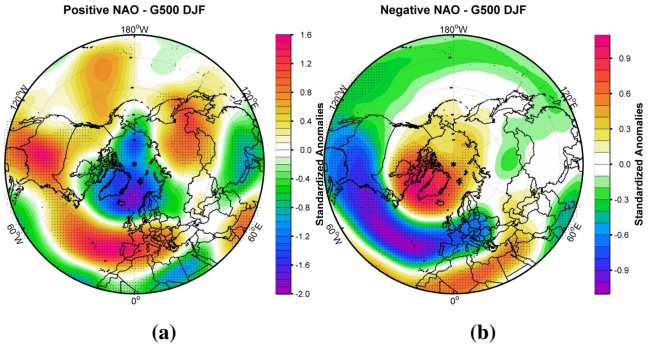

Figure 1: *Composite map between normalized NAO index and normalized geopotential height at 500 mb for positive phase (a) and negative phase (b).*

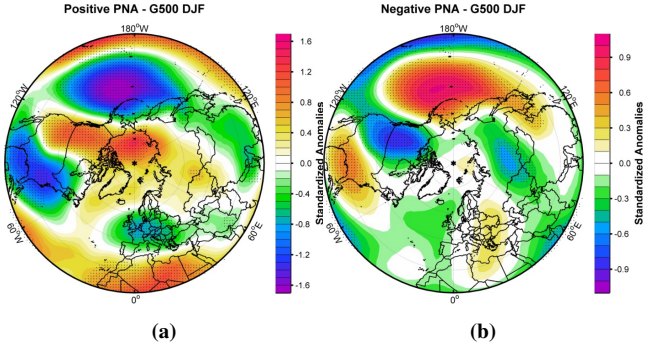

Figure 2: *Composite map between normalized PNA index and normalized geopotential height at 500 mb for positive phase (a) and negative phase (b).*

The North Atlantic Oscillation. A pattern which dominates the lower atmosphere variability in the North Atlantic sector is the North Atlantic Oscillation (NAO). It refers to a north-south oscillation in atmospheric mass between the Icelandic low- and the Azores high-pressure centers [WB32]. The NAO is the dominant pattern of near-surface atmospheric variability over the North Atlantic, accounting for one third of the total variance in monthly sea-level pressure in winter. A standard visualization in the form of a schematic representation of the spatial signature and climate impacts of NAO is given in Figure 1. The positive phase (Figure 1a) is associated with higher than normal surface pressure south of 55 °N combined with a broad region with anomalously low pressure throughout the Arctic and subarctic. Consequently, this phase is associated with stronger-than-average winds across the mid-latitudes of the Atlantic onto Europe. During the negative phase, both the Icelandic low and Azores high-pressure systems are weaker than normal (Figure 1b). The negative phase brings higher-than-normal pressure over the polar region and lower-than-normal pressure at about 45 °N. The negative phase allows cold air to plunge into the midwestern United States and Western Europe, and storms bring rain to the Mediterranean.

The Pacific North American Oscillation. The Pacific North American oscillation (PNA) is an alternating pattern between pressures in the central Pacific Ocean and centers of action over western Canada and the southeastern U.S., most pronounced in winter. The PNA is associated with a Rossby wave pattern and refers to the relative amplitudes of the ridge over western North America and the troughs over the central North Pacific and southeastern United States. A positive phase (Figure 2a) of teleconnection occurs when deeper than normal troughs occur over the eastern United States and the region of the Aleutians. In the negative phase (Figure 2b) the troughs are filled and a ridge over the Rockies is lowered. The positive phase is associated with above-average temperatures over western Canada and the extreme western U.S., and below-average temperatures across the south-central and southeastern U.S. The associated precipitation anomalies include above-average totals in the Gulf of Alaska extending into the Pacific Northwest of the United States, and below-average totals over the upper midwestern U.S.

3. Feature Extraction

We first describe how features (foremost correlations between spatial locations or regions) are extracted from the data, which are then visually encoded for the interactive analysis.

3.1. Correlation

The initial step of the approach is the computation of correlation for all pairs of grid points. For the purposes of this paper, we focus on a single two-dimensional data field. For each pair of grid points p, q, defined by their latitude and longitude, the Pearson correlation coefficient is applied to the associated time series as

$$r_{pq} = \frac{\sum_t (p_t - \bar{p})(q_t - \bar{q})}{\sqrt{\sum_t (p_t - \bar{p})^2}\sqrt{\sum_t (q_t - \bar{q})^2}},$$

where p_t, q_t are the values of the respective time series at time t, and $\bar{p}$, $\bar{q}$ are the mean values of the time series.

Due to the fact that random time series can exhibit a certain level of correlation, it becomes necessary to estimate statistical significance of computed values. Moreover, a time series of physical simulation output is not purely random and some inertia is observed in the system, i.e., each value influences the subsequent ones. To account for this dependency, autocorrelation for each time series is computed, which is defined as the correlation of a time series with itself shifted by a timestep. The resulting coefficient r_{i0} is used to estimate an effective number of data series items [vSZ02]

$$n_{ie} = n\frac{1 - r_{i0}}{1 + r_{i0}},$$

where n is the length of the time series. Testing for statistical significance is then performed by executing the Student t-test

$$t(r_{ij}) = r_{ij}\sqrt{\frac{n_{ie} - 2}{1 - r_{ij}^2}}$$

where i and j are indices of two points p_i, p_j in the array of all grid points and r_{ij} denotes the Pearson correlation of the respective time series at the points.

The correlation matrix with entries r_{ij}, where each grid point is represented by a row and a column, form the data space studied in the visual analysis. By definition, the matrix is symmetric ($r_{ij} = r_{ji}$)

and entries on the diagonal are 1 ($r_{ii} = 1$). In the visual encodings (Section 4), the correlation matrix is exploited as follows: (1) The correlation map displays one selected column at a time, cf. Figure 4(b). (2) The teleconnectivity map displays a column derived from the strongest anti-correlations in each row, cf. Figure 4(a). (3) The projection view shows the selected column overlaid on the inherent structure of the matrix, cf. Figure 4(d).

3.2. Projection

Typically in geoscientific research, points are placed on a map where their positions correspond to their latitudes and longitudes in the physical space. Here, we suggest projection as an alternative representation of the correlation space of data, where placements of points are defined not by their geographical coordinates, but instead by mutual correlations of the respective time series. To construct this representation, a distance matrix is derived from the correlation matrix, by defining a distance

$$d^*_{ij} = d^*(p_i, p_j) = \frac{1 - r_{p_i p_j}}{2}$$

between two points p_i, p_j, where d^*_{ij} reaches its minimum if the respective time series at the points are perfectly positively correlated, and maximum if they are perfectly negatively correlated.

For the projection, we use Sammon's mapping [Sam69], which minimizes

$$E = \frac{1}{\sum_{i<j}[d^*_{ij}]} \sum_{i<j}^{N} \frac{[d^*_{ij} - {d^{(2)}}_{ij}]^2}{d^*_{ij}}.$$

with ${d^{(2)}}_{ij}$ denoting the Euclidean distances between the i-th and the j-th grid point. The result of this transformation is presented and explored in the projection view (Section 4).

3.3. Teleconnectivity

Teleconnectivity at a reference point p_i is defined as the absolute value of the strongest anti-correlation observed at p_i with any other point p_j. Given the correlation matrix, we detect the minimum value of row i and take its absolute value:

$$T_i = \left| \min_j (r_{ij}) \right|.$$

In the following, we refer to the triplet (i, j, T_i), for a cell r_{ij} which exposes such a minimum, as a link from the starting point i to the endpoint j with the teleconnectivity value T_i. The function which establishes the correspondence $i \mapsto T_i$ of each point to its teleconnectivity value defines a teleconnectivity map. Representative links for the dataset are selected from the local maxima of the teleconnectivity map, based on the results of the regions extraction algorithm. To test for statistical significance of teleconnectivity values (Section 3.1), a directional t-test is used, as high values of teleconnectivity are coming from negative values of correlation. Teleconnectivity values and their statistical significance are presented in the teleconnectivity map view, while the representative links are overlaid with the teleconnectivity map and are also enumerated in the teleconnectivity links list (Section 4, cf. Figure 4(c)).

3.4. Region extraction

Region extraction is an intermediate step in choosing representative links. The aim is to extract consistent non-overlapping regions and present teleconnectivity relationships between them. A region is defined by (1) It has high teleconnectivity values: statistically significant and above a threshold (if defined). (2) It contains a local maximum of teleconnectivity. (3) It covers a spatially continuous area. (4) It is consistent with respect to correlations between its points: all points have statistically significant positive correlation with the region's maximum.

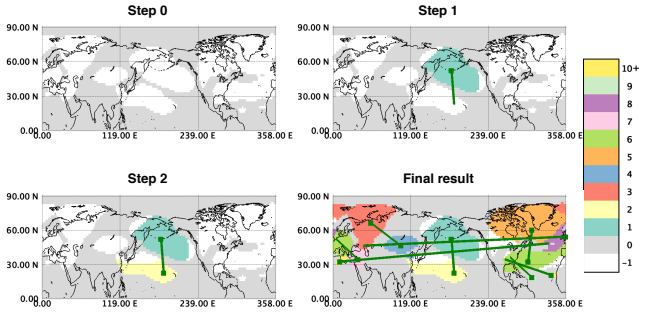

Figure 3: *Region extraction for NCEP dataset with teleconnectivity threshold 0.55. At Step 0, the teleconnectivity threshold is applied, and the excluded area is marked with gray, while regions with values above the threshold stay white. Steps 1-2 show consecutive seed points with respective links and colored regions extracted at first and second iteration. Final result shows the colored extracted regions (color correspond to iteration step) at the state when no local maxima are left in the unmarked areas.*

Our iterative region extraction algorithm starts with excluding the points with teleconnectivity values that are not statistically significant and/or are below the selected teleconnectivity threshold (Condition 1, Step 0 in Figure 3). Then, the algorithm iteratively searches for the largest local maximum of teleconnectivity not yet assigned to any region. This maximum becomes the seed for a new region (Condition 2, point+link in Figure 3). A region-growing approach around the seed point is deployed, including neighboring points which have teleconnectivity above the threshold and are positively and statistically significantly correlated with the seed point. Continuity of the data in the longitude dimension is taken into account (Condition 3). Region growing stops when it reaches its limits (Conditions 1,3,4: all points surrounding the region have teleconnectivity values below the threshold, or are negatively or not statistically significantly correlated with the seed; leading to the colored areas in Figure 3). The process then enters the next iteration step. When no unmarked local maxima are left, extraction of the region stops (Condition 2, final result in Figure 3). The actual links to be shown in the teleconnectivity views are selected afterwards. From the set of local maxima of extracted regions, the strongest links connecting each pair within these regions are chosen in each direction.

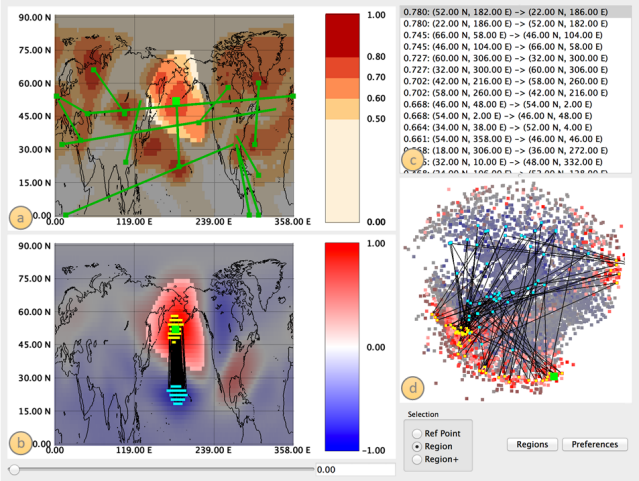

Figure 4: *Interface of the teleconnection analysis tool highlighting the region around the strongest teleconnectivity maximum. Data: NCEP, geopotential height at 500 hPa, northern hemisphere, winter means for 1871–2000. (a) teleconnectivity map, (b) correlation map, (c) teleconnectivity links list, (d) projection view. The reference point is (52 °N, 178 °W) with teleconnectivity $T = 0.78$.*

3.5. Correlation chain

The teleconnectivity map and links, described above, are indicating the points that expose strong relationships with others, and extrapolate these relationships into regions extracted around local maxima of teleconnectivity. The actual patterns of local and global scale are validated through the correlation maps for these points, corresponding to the respective columns in the correlation matrix. A correlation chain includes larger groups of points that have strong correlation relationships, both positive and negative, with a chosen reference point. It provides an additional stability assessment for the teleconnections, improving the understanding on whether correlated points cluster around the ends of a link or tend to be located in other regions of the map. The chain is built iteratively. It starts with the selected reference point. Each subsequent point is the one that has the strongest negative correlation with the current endpoint of the chain and is not in the chain yet. The color alternates between yellow and cyan for even and odd points in the chain, allowing for distinguishing between groups of points that are positively correlated to the reference point and those that are negatively correlated to the reference point, cf. Figure 4(b). For the figures in this paper, the stopping condition is reaching 100 points in the chain.

4. Visual Encodings

Exploration of the results of the analysis is performed in an interactive visual setting with four coordinated views: a teleconnectivity map, a teleconnectivity links list, a correlation map, and a projection view. Two of the views present the teleconnectivity analysis results (Figure 4(a,c)), while the other two views present the results of the correlation analysis (Figure 4(b,d)).

Teleconnectivity map and links list. Teleconnectivity map (Figure 4(a)) is a geospatial map-based visualization of teleconnectivity and representative links. The map shows a reference grid and land contours in the geographical area of the dataset, and can present the data in a rectangular or polar setup. Teleconnectivity is visually encoded using color mapping. A user-defined threshold value is used to generate the regions, as described in Section 3.4, as well as for filtering the values in the map. We employ a banded color scheme with decreasing luminance and increasing saturation from the Color Brewer project [Bre0x], with higher values receiving darker and more saturated color shades, see color legend next to the map in Figure 4(a). Two extra colors are used for conveying additional information: gray color is used for the points which are not statistically significant, and the values below the user-defined threshold are colored in white. Additionally, if some group of points is selected for highlighting, then these points are shown in full color, while other points are dimmed down in brightness.

Teleconnectivity links are unilateral relations and visualized with a dot at their starting point and a line connecting the starting point to the endpoint. The starting point is the grid point where the representative teleconnectivity value is observed, and the endpoint is the grid point with which strong anti-correlation is observed. The endpoint is not marked if it reaches its teleconnectivity value with a third point, or if it is in a neighborhood of a stronger maximum. A dot of a slightly bigger size and a brighter color marks the position of the reference point for the correlation views.

The links list (Figure 4(c)) enumerates the links shown on the teleconnectivity map in order of descending teleconnectivity value. For each link, the textual representation of its value, and coordinates of starting and endpoints are shown.

Correlation map and projection view. The correlation map (Figure 4(b)) is another geospatial representation depicting correlations with respect to a reference point within physical space, while the projection view (Figure 4(d)) is a non-geospatial representation of global correlation relationships within the data. The projection visualization provides a novel perspective on the correlation space and opens possibilities for new findings. The color in both views represents the correlations of each point with the current reference point. The used color scheme is a standard diverging (or bipolar) color map that interpolates the color from dark blue for strong negative correlation via white for non-correlated points to dark red for strong positive correlation, see color legend next to the map in Figure 4(b). The current selection is visualized by keeping the selected points bright and dimming the color of the non-selected points. To ease the correspondence, the correlation map uses the same layout as the teleconnectivity map with a rectangular or polar setup, land contours, and the reference grid. It allows for overlaying the color mapping with a geometric representation of the correlation chain. The reference point, which starts the chain, is marked in green, and each consecutive point is marked in cyan for the negatively correlated nodes (an odd number of steps away) and in yellow for the positively correlated ones (an even number of steps away). Nodes are connected with lines.

The projection view also marks the reference point and visualizes the correlation chain in the same way as the correlation map.

Its main contribution, however, is the global representation of pairwise correlations as distances within a 2D plot. This view spatially restructures the information of the correlation map. Instead of being laid out according to the physical coordinates of points, it expresses the relationships between the grid points in the correlation space. Clusters in the projection view represent points that are strongly (positively) correlated with each other (high similarities of their time series) and much less (or negatively) correlated with points outside the cluster. A dense cluster represents a group of points revealing very high mutual correlation and similar correlations of these points with the rest of the dataset. A slightly less dense group of points stretched along the outside of the projection may indicate that while points exhibit strong positive correlations with each other, their correlations with the rest of the dataset are not uniform, or vice versa. Anti-correlated groups of points tend to be placed at the opposite sides of the projection, and this tendency is stronger as the groups get bigger and the strength of their anti-correlation gets higher.

5. Interaction Mechanisms

An important difference of an interactive approach from a traditional script-based approach is that the user is able to interact with the software to examine all aspects of the data at different levels and reconfigure views for multiple analysis tasks and steps after the main resource-consuming operations are completed. Low response times of the system during the interactive analysis enable the user to intuitively explore the data and quickly build knowledge and understanding of the data at multiple scales. In addition, the described approach leverages the concept of coordinated multiple views. It means that several visualizations are shown at the same time, and each action of the user receives fast visual feedback in all views simultaneously, giving a consistent view on multiple aspects of the data and providing assistance and support in making decisions for further actions.

The most basic interaction method is hovering the mouse over points of interest in the views. It triggers small in-place pop-up windows with on-the-spot information, including physical coordinates (latitude and longitude) of the point under the mouse pointer, and its respective value in the view (teleconnectivity for the teleconnectivity map, correlation with the reference point for the correlation map and projection view). Beyond that, the tool accepts three main means of user input: selection of a new reference point, selection of a group of points for highlighting (brushing), and selection of a new teleconnectivity threshold. At the launch of the tool, the default reference point is the top link in the links list (the link with the strongest teleconnectivity), no points are highlighted (no points are dimmed in the views and all points are visualized in full brightness), and the teleconnectivity threshold is zero.

The reference point is used for the color mapping and generating the correlation chain in the correlation map and in the projection view. By selecting a new reference point, the user can validate stability of the teleconnection for the points surrounding the shown link, or simply check the correlation relationships of a point with no link shown in the teleconnectivity map. The reference point is selected by clicking on a point in the point-based views or on a link in the list. The current reference point is reflected in the point-based views as a point of a different color. If it is the starting point of one of the representative links, the respective row is highlighted in the list.

The views also support highlighting, which helps to focus on the most relevant aspects of the data. Depending on the selection mode, a click in the views does not only change the reference point, but also selects its region or connected component in the graph of regions for highlighting. The user can also make a free-form selection in the projection view. The points inside the selection or belonging to the selected regions keep their bright colors while all other points have dimmed colors.

Change of the teleconnectivity threshold modifies the teleconnectivity map, including points above the threshold and excluding points below it. A higher value of the threshold may cause a region to disappear or split into several subregions according to the strength and the number of the local maxima. A lower threshold may cause a new region to appear or two regions to merge. In case of such changes, different teleconnectivity links will be exposed. By interactively modifying the threshold and noticing the changes in the colored area and links, the user can intuitively compare strengths of the patterns in the dataset, and filter out weaker and less interesting ones.

6. Implementation

The tool is implemented in C++, using Qt for the user interface, OpenGL for the visualizations, the GNU Scientific Library (GSL) for assessment of statistical significance, and the NetCDF library for reading data provided in NetCDF format. Computation times for a non-optimized, single-threaded implementation of the feature extractions (Section 3.4) running on a MacBook Pro with a 2.6 GHz Intel Core i5 processor are shown in Table 1. Note that these precomputations are executed only once. All subsequent computation steps engaging the user during the visual analysis are at interactive rates.

Dataset	Data Size	Correlation	Projection	Loading	Total Time
NCEP	$91 \times 180 \times 129$	6m 33s	4h 58m 11s	2m 19s	5h 7m 3s
CTRL	$48 \times 96 \times 129$	32s	21m 4s	11s	21m 47s
ENS1	$48 \times 96 \times 129$	32s	20m 28 s	11s	21m 11s

Table 1: *Computation times for pre-computed feature extraction, where loading includes generation of teleconnectivity map, region search, finding teleconnectivity links, and building correlation chain for default reference point.*

7. Application Scenarios

7.1. Data

We analyze datasets contain northern hemisphere winter means (DJF) of geopotential height at 500 mb level for a period of 130 years (1871-2000). The reanalysis data "NCEP" are extracted from the 20th Century Reanalysis Project database, version 2 [NCE], for the period 1871-2010. This reanalysis data set is derived through a consistent assimilation and modeling procedure that incorporated most available weather and satellite information [WCWH04, CWS06, CWS*11]. The spatial resolution of the grid is 2° latitude $\times$ 2° longitude. Simulations of the last millennium have

been conducted using the COSMOS [JLT*10]. The Earth system model consists of a general circulation model for the atmosphere ECHAM5 [RBB*03] coupled with the general ocean circulation model MPIOM [MHJ*03] with a full carbon cycle implementation. The carbon cycle model comprises the ocean biogeochemistry module HAMOCC5 [WMRB*06] and the land surface scheme JSBACH [RRK*07]. ECHAM5 is run at T31 resolution (3.75°) with 19 vertical levels, resolving the atmosphere up to 10 hPa and MPIOM applies a conformal mapping grid with a horizontal resolution ranging from 22 km to 350 km. The model is forced by reconstructions of a) Total Solar Irradiance, b) volcanic forcing considering aerosol optical depth at 0.55 μm and effective radius distribution for 10-day time steps and four equi-areal segments, c) land use change, and d) anthropogenic greenhouse gases and aerosols. We use a full-forcing ensemble mean "ENS1" with five members, as well as the control run "CTRL". The whole experiments cover the period 800–2005 AD and the different initial conditions for the ensemble members are derived from a 3000 years control integration forced by constant conditions for 1860. For the presented analysis we use the period 1871–2000 only for a better comparison with the NCEP data set.

7.2. Teleconnectivity analysis results

Figures 5 and 6 provide a comparison of automatically extracted regions and representative links for the three studied datasets showing the teleconnectivity map at two different threshold levels. Figure 5 demonstrates the influence of the threshold on the region extraction. Without thresholding, region growth is only limited by positive correlation with the region seed. This typically leads to the Arctic, the equatorial belt, the Southern Ocean, and the Antarctic being extracted as single regions, possibly under-representing teleconnectivity links between them and other regions.

Figure 6 focuses on the interactive visual analysis of the teleconnectivity. We observe that most of the shown links have points at both ends, i.e., they represent classic teleconnection dipoles, where two points have the most negative correlation on the map with each other. In general, the simulated datasets expose higher amount of teleconnectivity, particularly in the southern hemisphere between Antarctic and the Southern Ocean. In all cases the link between New Zealand and the Ross Sea is shown, which appears as the globally strongest link for the ENS1 dataset with $T = 0.83$. Another link connects south/southeast of the Indian Ocean with the Antarctic continental border around 120°E.

In comparison to the CTRL simulation, ENS1 exposes lower amount of teleconnectivity over equatorial and tropical regions. In these areas, teleconnectivity above 0.6 can only be seen in the central Pacific and the southeastern United States. There is no clearly visible NAO-like teleconnection observed in the ensemble data. The typical southern counterpart of the teleconnectivity center over Greenland is shifted to the east and is located over Western Europe.

There is a noticeable difference in statistical significance between the datasets (the gray areas in Figure 6). At the 0.99 level, all grid points of CTRL exposes statistically significant teleconnectivity, ENS1 has a few not statistically significant points in the tropics, while NCEP shows large areas of not statistically significant points at the equator and in the tropics.

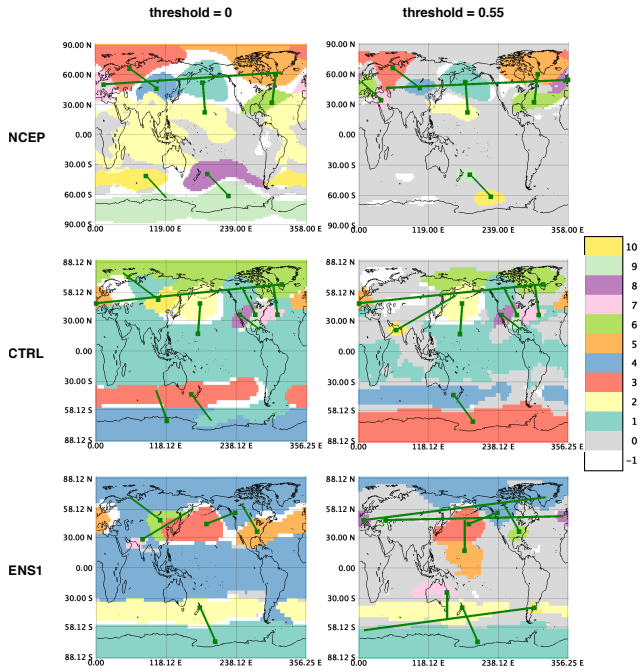

Figure 5: *Comparison of the region extraction results up to the tenth region for different datasets at two threshold levels (NCEP: the reanalysis data; CTRL: the control run data; ENS1: the full-forcing ensemble mean data). Colors correspond to number of iterations executed when the grid point is assigned a value, where zero means no statistically significant teleconnectivity (at 0.99 level) or below the user-defined threshold.*

7.3. Pattern analysis results

For a detailed look at the patterns, we restrict the analysis to the northern hemisphere. Iterating through the list of representative links for the NCEP dataset, we can see that the correlation chain forms a few strong dipoles (between the northern and the central Pacific, between Scandinavia and Asia; between the northern and the western Pacific) and a few patterns with three centers (Greenland, the northern Atlantic and Europe; the eastern Pacific, Canada, and the Gulf of Mexico). There are also mixtures of these patterns, or less defined patterns with multiple weaker centers. The strongest teleconnections (Figure 7) are the dipole in the northern and the central Pacific ($T = 0.78$), the dipole between Scandinavia and Asia ($T = 0.745$), the three-center pattern with Greenland, northern Atlantic, and Europe (the strongest link with $T = 0.727$), and the three-center pattern with the eastern Pacific, Canada, and the Gulf of Mexico (the strongest link with $T = 0.702$). Three-center patterns match the well-known North Atlantic Oscillation and the Pacific/North American pattern very well. Comparing the correlation maps for $T = 0.78$ and $T = 0.702$ we can see that they represent the same pattern. However, a shift from the center at (52°N, 182°E) to the one at (42°N, 216°E) results in a completely different correlation chain, even if both centers belong to the same extracted region.

In the CTRL simulation data (Figure 8, top row), we observe the PNA pattern with a center in the northern Pacific and two centers

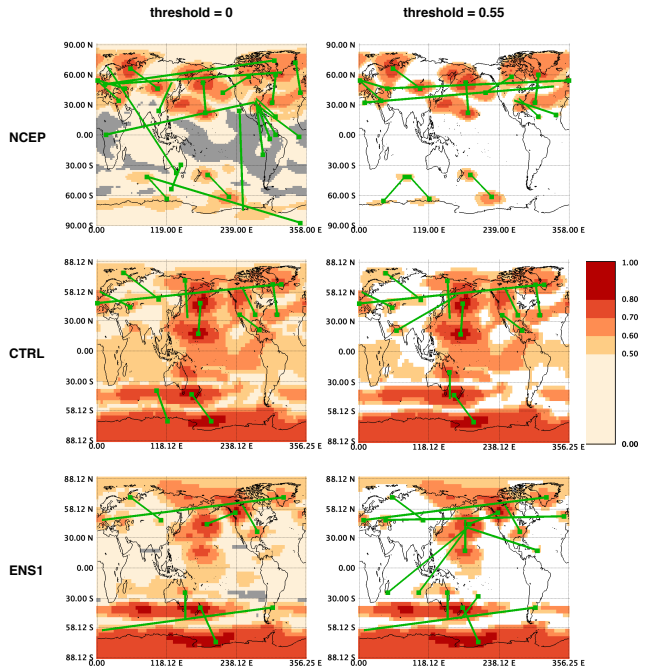

Figure 6: *Comparison of the teleconnectivity map views after completing the region extraction of Figure 5. Gray color indicates no statistically significant teleconnectivity (at 0.99 level).*

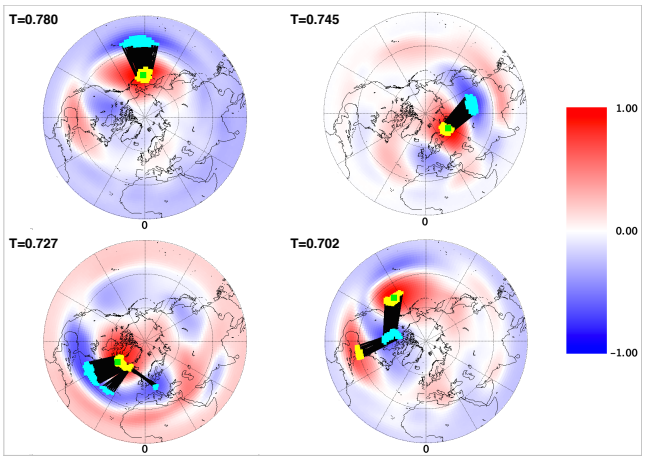

Figure 7: *Strongest teleconnections for the NCEP dataset. (Top-left) $T = 0.78$ at (52°N, 178°W), (top-right) $T = 0.745$ at (66°N, 58°E), (bottom-left) $T = 0.727$ at (60°N, 54°W), (bottom-right) $T = 0.702$ at (42°N, 144°W).*

of the opposite sign in the midwestern Pacific and over Canada (the strongest link with $T = 0.858$). Additionally, we see the NAO pattern with a center in the Davis Strait near Greenland and two centers of the opposite sign in the western Atlantic and over Western Europe (the strongest link has $T = 0.714$). A stable dipole exists between the center of the PNA in the midwestern Pacific and an area over the northeastern border of Eurasia, which serves as an attractor for all shown links with $T < 0.7$. In ENS1 data, for the PNA the strongest link has $T = 0.816$, and the teleconnectivity center suspected for NAO does not produce a clearly separated pattern in the correlation chain (Figure 8, bottom row).

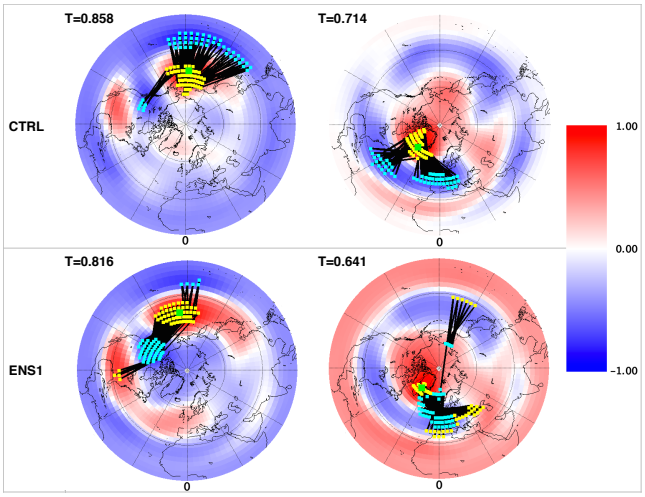

Figure 8: *Teleconnections patterns in the model data. (Left) closest to PNA, (right) closest to NAO. $T = 0.858$ at (46.88°N, 176.25°E), $T = 0.714$ at (65.62°N, 45°W), $T = 0.816$ at (43.12°N, 172.5°W), $T = 0.641$ at (69.38°N, 41.25°W).*

7.4. Pattern components

With the help of the projection view, we can explore the patterns in more depth. The top row in Figure 9 shows highlighting of the automatically extracted region, corresponding to the northern and the northwestern Pacific, in the projected view and in the teleconnectivity map, for the northern hemisphere part of ENS1 dataset. Applying manual selection around the positive end of the correlation chain in the projected view (the middle row in Figure 9), we can see that these points correspond to the northern Pacific, northeastern United States and northwestern Europe. Repeating the same interaction for the negative end of the correlation chain (the bottom row in Figure 9), the central Pacific, Canada, and an area in the central Atlantic receive the highlighting. The areas that are highlighted in the map with these selections are considered to have a similar temporal behavior to the respective parts of the correlation chain (the ones positively correlated with the reference point, and the ones negatively correlated with the reference point). Consequently, the areas of northwestern Europe and the central Atlantic can be experiencing influence of the PNA pattern.

8. Conclusion

We presented a tool for identification and analysis of teleconnections. The strongest negative temporal correlations, taken at each point, define a teleconnectivity map that suggests potential pattern centers. Analysis of neighboring points allows for the extraction of regions, relationships between which are taken as representative of the data structure. A correlation map for a pattern center shows de-

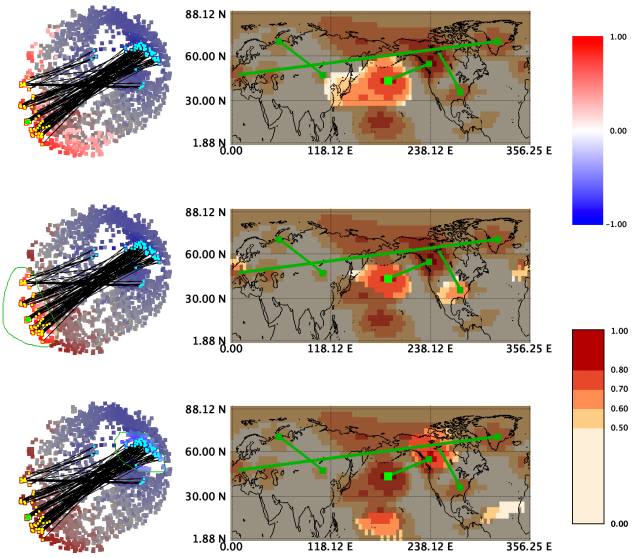

Figure 9: *Exploring the PNA pattern components in the ENS1 data with the help of highlighting in the projection view and the teleconnectivity map. (Top row) region around the start point of the strongest link is highlighted in the views, (middle row) a free-form selection around the positive nodes of the correlation chain in the projection view, (bottom row) a free-form selection around the negative nodes of the correlation chain in the projection view. The reference point is (43 °N, 172.5 °W), with teleconnectivity T=0.816.*

grees of its influence, and the sequence of strongly negatively correlated points starting at the center serves as a stability assessment for the pattern. An additional view provides insights into global correlation relationships within the dataset. By comparing three sample datasets with the help of our tool, it is shown that the model data exposes higher amount of teleconnectivity when compared with reanalysis data (e.g. NCEP). It is also apparent that large variability around the equator in the control run is not present in the ensemble mean data, which may be an artifact of the model simulations or due to the fact that we use an ensemble mean over five members. The patterns extracted from the observation-based NCEP dataset are much better aligned, when compared to model data, with well-known climate modes of variability (e.g. NAO and PNA).

Acknowledgements. This work was funded by Helmholtz Association as a part of Earth System Science Research School, as well as the REKLIM and PACES programmes.

References

[Bre0x] BREWER C. A.: Colorbrewer, 200x. URL: http://colorbrewer2.org. 4

[Cli] CLIMATE REANALYZER: Climate Change Institute, University of Maine, USA. URL: http://cci-reanalyzer.org/ [cited 2016/05/13]. 1

[CWS06] COMPO G. P., WHITAKER J. S., SARDESHMUKH P. D.: Feasibility of a 100-year reanalysis using only surface pressure data. *Bulletin of the American Meteorological Society 87*, 2 (2006), 175–190. 5

[CWS*11] COMPO G. P., WHITAKER J. S., SARDESHMUKH P. D., MATSUI N., ALLAN R. J., YIN X., GLEASON B. E., VOSE R. S., RUTLEDGE G., BESSEMOULIN P., BRÖNNIMANN S., BRUNET M., CROUTHAMEL R. I., GRANT A. N., GROISMAN P. Y., JONES P. D., KRUK M. C., KRUGER A. C., MARSHALL G. J., MAUGERI M., MOK H. Y., NORDLI Ø., ROSS T. F., TRIGO R. M., WANG X. L., WOODRUFF S. D., WORLEY S. J.: The twentieth century reanalysis project. *Quarterly Journal of the Royal Meteorological Society 137*, 654 (2011), 1–28. 5

[JLT*10] JUNGCLAUS J. H., LORENZ S. J., TIMMRECK C., REICK C. H., BROVKIN V., SIX K., SEGSCHNEIDER J., GIORGETTA M. A., CROWLEY T. J., PONGRATZ J., KRIVOVA N. A., VIEIRA L. E., SOLANKI S. K., KLOCKE D., BOTZET M., ESCH M., GAYLER V., HAAK H., RADDATZ T. J., ROECKNER E., SCHNUR R., WIDMANN H., CLAUSSEN M., STEVENS B., MAROTZKE J.: Climate and carbon-cycle variability over the last millennium. *Climate of the Past 6*, 5 (2010), 723–737. 6

[KNM] KNMI CLIMATE EXPLORER: Royal Netherlands Meteorological Institute (KNMI). URL: https://climexp.knmi.nl/ [cited 2016/05/13]. 1

[MHJ*03] MARSLAND S. J., HAAK H., JUNGCLAUS J. H., LATIF M., RÖSKE F.: The Max-Planck-Institute global ocean/sea ice model with orthogonal curvilinear coordinates. *Ocean Modelling 5*, 2 (2003), 91 – 127. 6

[NBD*15] NOCKE T., BUSCHMANN S., DONGES J. F., MARWAN N., SCHULZ H.-J., TOMINSKI C.: Review: visual analytics of climate networks. *Nonlinear Processes in Geophysics 22*, 5 (2015), 545–570. URL: https://www.nonlin-processes-geophys.net/22/545/2015/, doi:10.5194/npg-22-545-2015. 1

[NCE] NCEP: The 20th century reanalysis project database, version 2, ESRL NOAA. URL: http://www.esrl.noaa.gov/psd/data/20thC_Rean [cited 2015/02/18]. 5

[PSD] PSD WEB PRODUCTS AND TOOLS: ESRL NOAA. URL: http://www.esrl.noaa.gov/psd/products [cited 2016/05/13]. 1

[RBB*03] ROECKNER E., BÄUML G., BONAVENTURA L., BROKOPF R., ESCH M., GIORGETTA M., HAGEMANN S., KIRCHNER I., KORNBLUEH L., MANZINI E., RHODIN A., SCHLESE U., SCHULZWEIDA U., TOMPKINS A.: The atmospheric general circulation model ECHAM5, Part I: model description. *Max Planck Institute for Meteorology Report 349* (2003), 1–127. 6

[RRK*07] RADDATZ T. J., REICK C. H., KNORR W., KATTGE J., ROECKNER E., SCHNUR R., SCHNITZLER K. G., WETZEL P., JUNGCLAUS J.: Will the tropical land biosphere dominate the climate–carbon cycle feedback during the twenty-first century. *Climate Dynamics 29*, 6 (2007), 565–574. 6

[Sam69] SAMMON JR. J. W.: A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers C-18*, 5 (1969), 401–409. 3

[SRS*13] STEED C. A., RICCIUTO D. M., SHIPMAN G., SMITH B., THORNTON P. E., WANG D., SHI X., WILLIAMS D. N.: Big data visual analytics for exploratory earth system simulation analysis. *Computers & Geosciences 61* (2013), 71 – 82. 1

[vSZ02] VON STORCH H., ZWIERS F. W.: *Statistical Analysis in Climate Research*. Cambridge University Press, 2002. 2

[WB32] WALKER G., BLISS E.: World weather v. *Memoirs of the Royal Meteorological Society 4*, 36 (1932), 53–84. 2

[WCWH04] WHITAKER J. S., COMPO G. P., WEI X., HAMILL T. M.: Reanalysis without radiosondes using ensemble data assimilation. *Monthly Weather Review 132*, 5 (2004), 1190–1200. 5

[WG81] WALLACE J. M., GUTZLER D. S.: Teleconnections in the geopotential height field during the northern hemisphere winter. *Monthly Weather Review 109* (April 1981), 784–812. 1

[WMRB*06] WETZEL P., MAIER-REIMER E., BOTZET M., JUNGCLAUS J., KEENLYSIDE N., LATIF M.: Effects of ocean biology on the penetrative radiation in a coupled climate model. *Journal of Climate 19*, 16 (2015/09/02 2006), 3973–3987. 6