Workshop on Visualisation in Environmental Sciences (EnvirVis) (2023)
S. Dutta, K. Feige, K. Rink, and D. Zeckzer (Editors)

Interactive Visual Analysis of Regional Time Series Correlation in Multi-field Climate Ensembles

Marina Evers[1], Michael Böttinger[2], and Lars Linsen[1]

[1]University of Münster, Germany [2]German Climate Computing Center (DKRZ), Germany

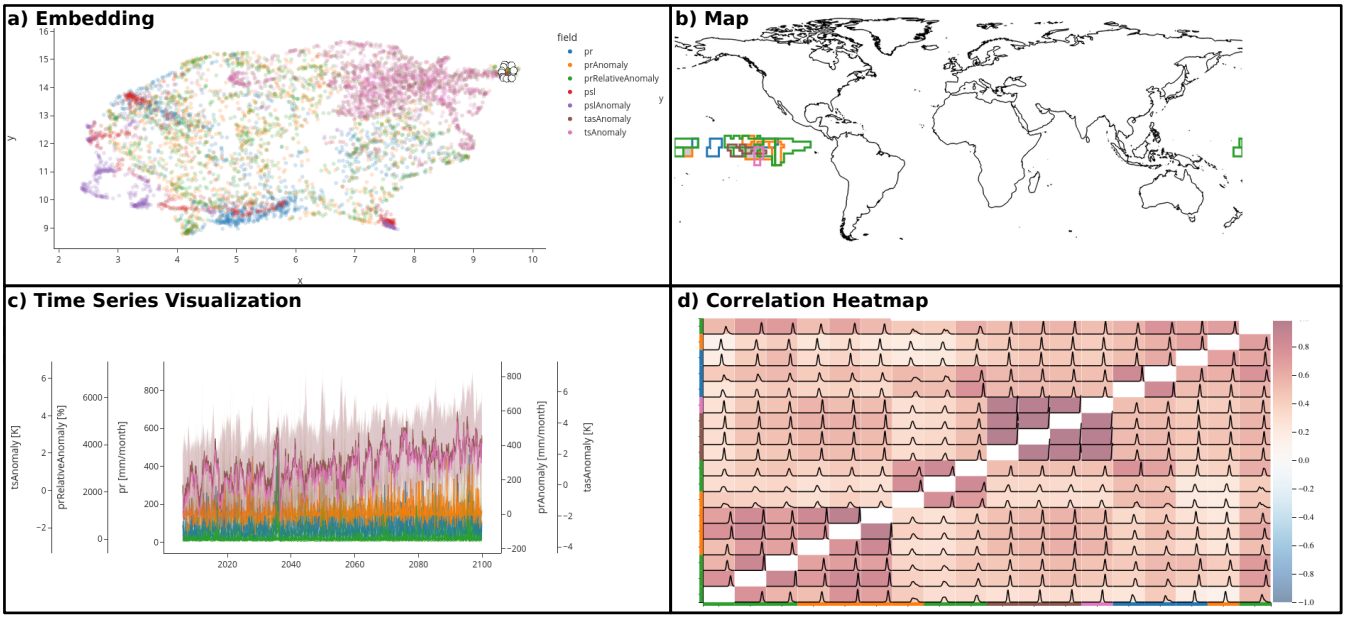

Figure 1: *Visual analysis tool for multi-field climate ensembles: An embedding of spatial segments based on their time series correlations (a) allows for cluster detection and selection. The selected segments are highlighted on a geographical map for spatial context (b). The respective time series can be visualized as a graph over time (c), while the correlations among the segments and their distribution over the ensemble are shown in a heatmap (d).*

Abstract

Spatio-temporal multi-field data resulting from ensemble simulations are commonly used in climate research to investigate possible climatic developments and their certainty. One analysis goal is the investigation of possible correlations among different spatial regions in the different fields to find regions of related behavior. We propose an interactive visual analysis approach that focuses on the analysis of correlations in spatio-temporal ensemble data. Our approach allows for finding correlations between spatial regions in different fields. Detection of clusters of strongly correlated spatial regions is supported by lower-dimensional embeddings. Then, groups can be selected and investigated in detail, e.g., to study the temporal evolution of the selected group, their Fourier spectra or the distribution of the correlations over the different ensemble members. We apply our approach to selected 2D scalar fields of a large ensemble climate simulation and demonstrate the utility of our tool with several use cases.

1. Introduction

To quantify the uncertainty of weather forecasts and climate projections, ensemble simulation techniques have been used for many years. By running the same model several times with slightly different initial conditions, the internal variability of the modeled system can be captured and its response to climate forcing can be

DOI: 10.2312/envirvis.20231108

estimated. Nowadays, *single model initial-condition large ensembles* (SMILEs) with ensemble sizes of the order of 100 or more [MML21] are increasingly used to obtain statistically robust estimates of the internal variability. However, the use of ensemble techniques adds an additional complexity to the multivariate, time-dependent spatial datasets, increasing their size and the effort required to analyze and visualize them.

Analyzing the correlations in climate simulation ensembles allows for studying long-range interaction in single fields as well as among multiple fields. As the number of correlations grows quadratically with the number of data points, most approaches use a reference point to which the correlations are computed. However, the choice of a suitable reference point is a difficult task, which strongly influences the analysis. A global analysis approach, instead, allows for a more comprehensive analysis.

To support the visual correlation analysis of such large and complex multi-field ensemble datasets, we developed an interactive application specifically designed to assist the user in identifying and further analyzing spatial regions with similar temporal behavior in one or more physical parameters. Starting from an embedding of the data points representing regions based on the similarity of their simulated time series, clusters of highly correlated regions can be interactively selected by brushing interactions. The properties of the clusters can, then, be analyzed in more detail in a geographical map, a correlation heatmap that also shows the variability over the ensemble, and a line plot showing the underlying time series or a Fourier transform thereof. Based on a 2D subset of the Grand Ensemble simulations of the Max Planck Institute for Meteorology (MPI-GE) [MMSG*19], we present several use cases that demonstrate the utility of our tool.

Our main contributions can be summarized as follows:

- We propose to use dimensionality reduction for a global correlation analysis of time series in different spatial regions among multiple fields at once.
- We present an interactive visual analysis approach that allows for an analysis of correlated regions at different levels of detail.
- We present several use cases from the field of climate research to show the utility of our approach.

2. Related Work

Many different approaches for a wide variety of domains have been proposed recently to analyze different aspects of spatio-temporal ensemble simulations [WHLS19]. Weather and climate simulations as one of the major application areas have led to several approaches that are focused in this domain [BLLS16, KRRW18, ZCL*20]. Jänicke et al. [JBMS09] use wavelets for the analysis of climate data on different scales. Kappe et al. [KBL19] analyze the variability of climate ensembles over time, while Vietinghoff et al. focus on the spatial variation of different climate phenomena [VHB*21]. Segmentations for multivariate data [HBLW21] have been proposed but neither deal with temporal data nor correlations.

The analysis of correlations is a common task where mostly correlations between different fields on a fixed position are computed [JPR*04, ZHQL16, WTZG18] or correlations in a single field between different spatial positions [PW12,ACM*19,BGR*19, ALI*19,EHL21]. Other approaches tackle the high dimensionality of the correlation matrix by visualizing the correlations based on a user-defined reference point [CWMW11, SWMW09]. Sauber et al. [STS06] compute correlation fields among different variables and represent them in a graph which is growing exponentially. Additionally, they are not considering correlations among different spatial regions. Nocke et al. [NBD*15] desribe coupled climate networks to study correlations among multi-fields which have been used to study correlations among two fields [ECKD21, DWW23]. Neither more than two fields nor ensembles have been studied. None of the discussed approaches allow for a global correlation analysis between spatial regions of multiple field, which is the objective of our analysis approach.

3. Visual Analysis Workflow

We propose an interactive visual analysis workflow for investigating correlations in a spatio-temporal multi-field climate ensemble simulation data set as shown in Fig. 2. After preprocessing the data, which is explained in detail in Section 4, we obtain correlations among segments, where the segments describe homogeneous spatial regions, i.e., in which the evolution of the field is highly correlated. Then, the correlations among segments are analyzed interactively as described in Section 5.

In order to support finding segments with similar temporal patterns, the analysis process starts with a 2D embedding, where each segment is represented by a point in the embedding and correlations between segments are mapped to distances between the corresponding points. Clusters in the resulting point cloud represent highly correlated segments. The user can select one or more clusters, of which the underlying data are then shown in linked views. A geographic map view shows the selected segments to provide spatial context. Selections in the map view are also possible such that the corresponding segments are highlighted in 2D embedding. For a more detailed analysis, we additionally include a heatmap visualization of the correlation data. Here, we use color to encode the correlation over the whole ensemble and also visualize the spread of the correlations among all ensemble members (see Section 5.4). We also provide a line plot that shows the median as well as the spread of the ensemble over time. For further analysis of cyclic patterns in the time course of the data for the selected areas, our system offers the possibility of spectral frequency analysis, which is a common tool in climate research. We support temporal selections of which a Fourier transform can be computed and visualized in a line plot. Note that all visualizations are closely linked by using a consistent color coding for the different fields. We implemented our approach in a web-based analysis tool using Plotly and D3.

4. Preprocessing

As a preprocessing step, we divide the data into spatial segments of locations that are highly correlated. This step significantly reduces the data size for subsequent analysis steps while keeping the internal variability of the data. Then, we compute a correlation matrix among the different segments over the whole ensemble as well as each ensemble member separately.

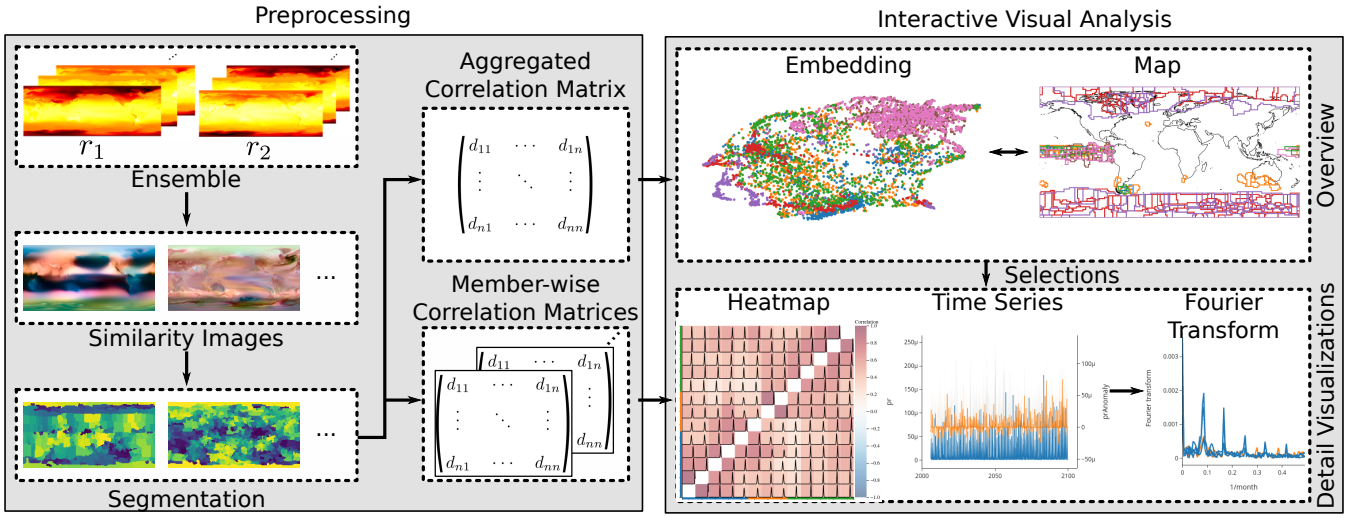

Figure 2: *The analytical workflow starts with a preprocessing step leading to correlations on a level of (homogeneous) segments. These correlations are analyzed in an interactive visual manner at different aggregation levels.*

4.1. Segmentation

For obtaining a spatial segmentation of the original data, where each of the segments consists of highly correlated spatial samples, we follow the same approach as Evers et al. [EHL21]. Since the segmentation may be rather different in different fields, we apply this approach for each field separately to reduce the complexity of the data more effectively. In the following, we denote a time series at spatial sample $\mathbf{x}_i$ for the field v and ensemble member r_k as $s_{ivk}(t)$. The correlation between the time series $s_{ivk}(t)$ and $s_{jwk}(t)$ of the same ensemble member at the spatial positions $\mathbf{x}_i$ and $\mathbf{x}_j$ for fields v and w can be computed directly. To obtain a time series that contains all n ensemble members, we use a concatenation. The concatenated time series $f_{iv}(t)$ at spatial sample $\mathbf{x}_i$ and field v can be written as

$$f_{iv}(t) = s_{ivk}\left(t - \sum_{l=1}^{k-1} T_l\right) \text{ for } \sum_{l=1}^{k-1} T_l < t \leq \sum_{l=1}^{k} T_l \,. \qquad (1)$$

The number of time steps of simulation run r_k is denoted as T_k. The pairwise correlations in a single field between two concatenated time series $f_{iv}(t)$ and $f_{jv}(t)$ are computed for each spatial sample pair leading to a correlation matrix for each field. In this work, we use the Pearson correlation coefficient [BCHC09], which is commonly used for time series. Note that the correlation measure can be easily exchanged, for example, to use a non-linear measure.

Based on the correlation matrix, we compute a distance matrix by mapping the correlations from a range $[-1,1]$ to distance in the range of $[0,1]$, where a strong positive correlation is considered as a distance of 0 and a strong negative correlation as a distance of 1. The resulting distance matrix is used as an input to a multi-dimensional scaling where each spatial sample point of the data is mapped to three dimensions. By assigning the three-dimensional values to the spatial positions, we obtain the so-called similarity images where similarity of colors represent the correlation strength.

Similarly to Evers et al., we use a hierarchical watershed segmentation approach to compute the segments from the correlations, but for our work we use only the highest level of detail of the hierarchy. By following this procedure, we obtain a segmentation where each segment contains only highly correlated data points. For each of these segments, we can compute the mean over all spatial samples for each ensemble member separately, which results in n time series for each segment for an ensemble with n ensemble members. By construction, we can assume that the individual segments are highly correlated and, therefore, sufficiently similar to be aggregated into a mean. In the following, we only work on these segment ensemble means which significantly reduces the size of the data. It also ensures that these data depend on the complexity of the underlying data instead of the spatial grid resolution. This can be observed for the case of the MPI-GE data: Pressure shows variability on a large scale, resulting in only 490 segments, while precipitation varies on smaller scales, resulting in 974 segments.

4.2. Correlations

In the following, we consider all fields together and compute correlation matrices for all segments of all fields. For the first correlation matrix, we still use the concatenated time series as defined in Equation 1 over all ensemble members. The values of this correlation matrix can be mapped to distances again as described in the previous section and will be used for the projection. Additionally, these correlation values will be used in the heatmap for a direct visualization. However, to cover the variability among the ensemble members, we also compute pairwise correlations between the time series of the individual ensemble members. This leads to n correlation values for each segment.

5. Visual Design

In the following, we will discuss the design decisions for the different visual encodings included in our interactive visual analysis tool.

5.1. Embedding

As the starting point of the analysis, we use a dimensionality reduction to compute a 2D embedding as shown in Fig. 1a, where each data point represents a spatial segment. For its computation, we chose uniform manifold approximation and projection (UMAP) [MHSG18], because our goal is to identify clusters of highly correlated segments, which are well preserved by UMAP. Additionally, UMAP tends to better preserve the global structure when compared to techniques like t-SNE. The projection is computed based on the correlation matrix that aggregates time series data for the whole ensemble (see Section 4.2). To compute distances from the correlation values, we map the correlations to a range of $[0,1]$ in the same way as we did for the computation of the similarity images. Therefore, we obtain an embedding that contains the structure of the complete multi-field ensemble where the different fields are encoded by color. Since we are computing correlations of time series, it is possible to include multiple fields representing different physical properties without performing normalizations. Depending on the analysis goal, highly anticorrelated time series shall be considered to be similar. Therefore, we provide the users with the option to alternatively compute the distance d_{ij} between segments i and j as $d_{ij} = 1 - |c_{ij}|$ where c_{ij} is the correlation between these segments.

The result of the UMAP algorithm strongly depends on the number of neighbors included for each point. As the optimal choice of this parameter is not obvious,the user can change the parameter and find a suitable setting. The density of the embedding can also be modified. By adjusting the parameters, it is also possible to find different features in the data. Analyzing multiple fields at once can lead to visual clutter in the scatterplot of the projection result. To address this issue, the user can interactively select which fields to display without affecting the fields included in the analysis within the linked visualizations. However, if any fields should be generally excluded, it is also possible to make a global selection of the fields that affects all visualizations. Groups of points can be selected by brushing in the scatterplot. To highlight the selected points, the saturation of the other sample points is decreased.

5.2. Spatial Visualization

Since the UMAP projection by definition does not show spatial relationships, an additional linked map view is needed to analyze regional features in the data. For this reason, we included a world map with continental outlines that provides the geographic context, as shown in Fig. 1b. When the user selects any points in the UMAP projection, the outlines of the corresponding segments are drawn on the world map. Thus, a direct connection between the sample points in the embedding and their spatial positions is possible.

It is also possible to apply a lasso selection directly in the map view. Here, each segment which at least partially overlaps with the selected region, is selected. The selection is also closely linked to the embedding where the points representing the selected segments are shown as selected.

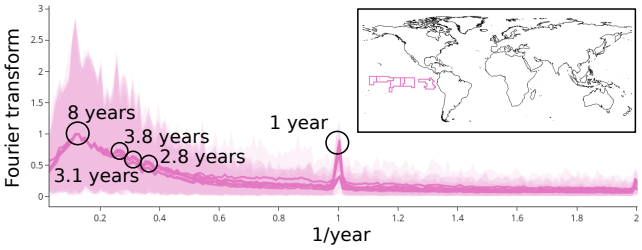

Figure 3: *The Fourier transform of the temperature anomaly exhibits two dominant peaks in the ensemble mean, several smaller peaks, and a strong variation over the ensemble. While the frequency of 1/year indicates seasonal variations, the other frequencies might be attributed to the El Niño phenomenon.*

5.3. Time Series Visualizations

For the segments of interest that have been detected, the time series are also visualized directly. We choose to visualize the time series as a graph over time where we also show the uncertainty introduced by the ensemble as shown in Fig. 1c. Here, we follow the same approach as Evers et al. [EHL21] which is inspired by functional boxplots [SG11]. We visualize the median of the set of time series together with a band spanning the range from the minimum to the maximum. We did not use traditional functional boxplots, as the ensemble data are relatively noisy, leading to many outliers. For linking the time series to the other visualizations, we apply the same color coding for the different fields as in the other visualizations. While this color-coding does not allow the user to distinguish segments of the same field, it allows for a clear separation of the different fields. As our approach aims at investigating correlations between time series of different fields, we show all time series of all fields together. Since the different fields cover very different ranges of values, we use separate y-axes for each field. This allows us to plot the values together without having to perform normalization, which would lose the scaling of the individual values.

Certain climate phenomena such as the El Niño/Southern Oscillation (ENSO) phenomenon occur at quasi-regular intervals, but the frequencies of their occurrence might change over time due to a changing climate. To enable studying frequencies as well as their variations over selected time intervals, we include a Fourier analysis in our approach. The users can brush in the time series visualization to select a time interval for which then a Fourier transform is computed. The Fourier transform is computed for each ensemble member individually. For the visualization, we compute the mean to reduce the noise introduced by the ensemble. Additionally, we remove the linear trend to avoid that the global warming dominates the frequency spectrum. To show the variability within the ensemble, we use a band ranging from the minimum to the maximum value. As we are mainly interested in the frequencies and want to compare different fields directly, we normalize the Fourier transform such that the mean spectrum lies in the range of $[0,1]$. The Fourier transforms of all chosen segments are shown together in a line chart as shown in Fig. 3.

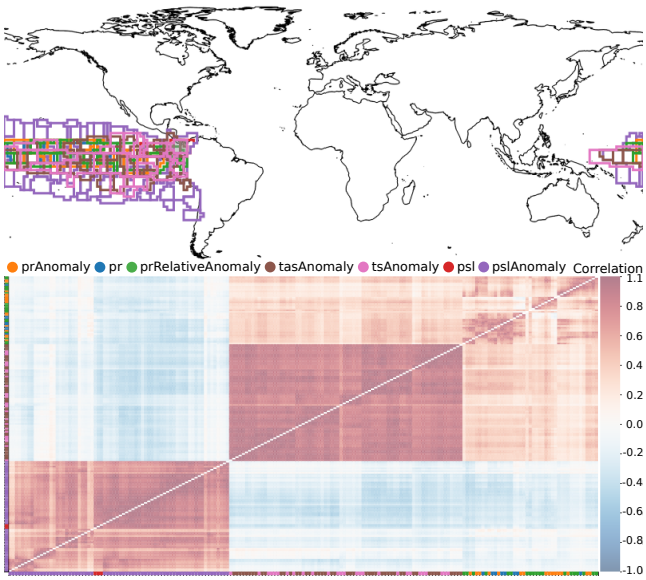

Figure 4: *For several variables, the heatmap (bottom) shows correlations between the temporal developments in the selected segments shown in the geographical map (top). The pressure fields (psl and pslAnomaly) are anticorrelated to the other variables. The correlations among the temperature anomaly fields (tsAnomaly and tasAnomaly) and those among the precipitation fields (pr, prAnomaly, and prRelativeAnomaly) are mostly higher than those between the different groups.*

5.4. Correlation Heatmap

One analysis goal of our approach is finding clusters of highly correlated segments. However, dimensionality reduction techniques might introduce projection artifacts. The avoid misinterpretations and investigate the correlation values in more detail, we visualize them directly by using a heatmap as shown in Fig. 1d. In the heatmap, we apply a color coding to the correlation values that cover the entire ensemble. We use a diverging color map, where red encodes positive correlations and blue encodes negative correlations.

The interpretability of heatmap visualizations strongly depends on the sorting of its rows / columns. We apply a hierarchical clustering using Ward's minimum variance method [WJ63].Thus, we obtain a sorting where similar rows / columns of the matrix are placed closely together. Then, clusters appear as blocks in the heatmap. To allow for a quick interpretation to which field the corresponding row / column of the matrix belongs, we use the colors for the fields instead of labels. This method also scales well when the number of rows / columns increases. To obtain spatial information of the segments, the users can hover over the cells in the heatmap and the corresponding segments are highlighted in the map view by filling them with semi-transparent grey color. Not using the full opacity ensures that the map providing spatial orientation as well as potential segments from other fields remain visible.

To investigate the variation within the ensemble, we enrich each cell of the heatmap with a visualization of the distribution of the correlation values of the different ensemble members. In this way, the variety within the ensemble can be estimated and it is possible to observe whether the aggregated correlation is representative for the ensemble data. The distribution is shown as a graph which we obtain by applying a kernel density estimate for smoothing. Here, we use the Epanechnikow kernel [Epa69], which minimizes the mean square error. The resulting graph is shown in each matrix cell as an overlay above the color coding.

For a large number of selected elements, the limited scalability of this method becomes obvious. The overlayed distributions hide the color-coded correlation cells below. One option to overcome this problem would be the use of an approach similar to responsive matrix cells [HBS*20], which only shows a detail visualization on demand. However, as our detail visualization is identical in each cell and it is desirable to show or hide all of them at once, we leave it to the users to decide whether to show the distributions or not. By hiding the ensemble spread, sorting the matrix and using colors for the fields as labels, the heatmap scales well also for large correlation matrices, as can be seen in Fig. 4.

6. Use Cases

For our use case, we apply our tool to a 2D subset of MPI-GE, the Max Planck Institute for Meteorology Grand Ensemble Simulations [MMSG*19] (available at https://esgf-data.dkrz.de/projects/mpi-ge/). The dataset consists of a pre-industrial control experiment, a historical experiment (1850-2005), a 150-year 1%-CO_2 increase experiment starting in 1850, and the three different future scenarios RCP2.6, RCP4.5, and RCP8.5 for the time period 2006-2100. For our application examples, we chose precipitation, sea level pressure, near-surface air temperature, and surface temperature of the simulations of scenario RCP8.5. This scenario is the one with the strongest forcing, i.e., a business-as-usual scenario with strongly increasing greenhouse gas concentrations throughout the century.

In weather and climate data, the signal-to-noise ratio of the correlation between the temporal developments of a physical parameter at different locations can be strongly impacted by the large variability caused by the daily and the annual cycle. Depending on the variable and the temporal scale of the desired analysis, it may be necessary to derive the anomaly of a quantity relative to a reference state first. For example, in order to analyze spatial correlations of temperature changes based on monthly data, it is desired to compute temperature anomalies relative to a reference climate, i.e. the multi-year mean monthly temperatures of a reference time interval. We use the years 1986-2005 of the historical experiment as the reference climate for analyzing climate change within the RCP8.5 runs.

6.1. North-Atlantic Oscillation

The North Atlantic Oscillation (NAO) is a fluctuation of the air pressure difference between the Icelandic Low and the Azores High in winter seasons. The pressure difference between the two systems has a strong influence on the strength and direction of westerly winds and the location of storm tracks over the North Atlantic,

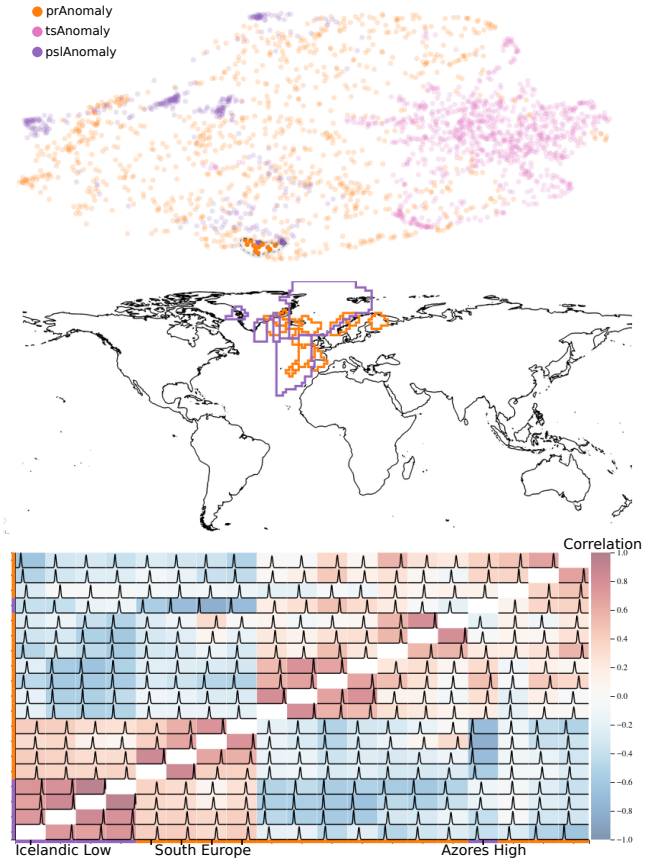

Figure 5: *The set of points selected in the embedding corresponds to regions north-west of Europe. Several correlations and anticorrelations between the regions containing the Azores High and the Icelandic Low in the pressure anomaly can be observed and investigated in the heatmap.*

and thus on winter weather in Europe. Since a strong NAO is characterized by very high air pressure values in the Azores region and particularly low values near Iceland, we look here for a strong anticorrelation between the temporal patterns of the two regions.

To easily select a set of points representing correlations as well as anticorrelations, we consider the absolute value of the correlation for the UMAP computation as described in Section 5.1. Here, we select segments for the pressure and precipitation anomaly that contain the regions for the Icelandic Low as well as the Azores High, as can be seen in Fig. 5. Using the heatmap reveals a dominant anticorrelation between the Icelandic Low and the Azores High with little variation within the ensemble. Additionally, the pressure anomaly in the region of the Icelandic Low is correlated with the precipitation anomaly in southern Europe and anticorrelated with the precipitation anomaly segments in northern Europe. For the segment that covers the Azores High, we observe the contrary even though the correlation with the precipitation segments in the north is often weaker. These regions and the correlations among the different fields are related to the NAO.

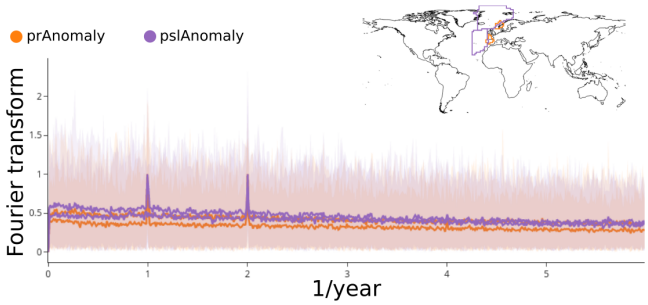

Figure 6: *The Fourier transform of the precipitation and pressure anomalies show two dominant peaks in the ensemble mean and a strong variation over the ensemble. Besides yearly patterns, there is also a clear signal for repetitions twice a year.*

The Fourier transform of four selected segments in these regions (see Fig. 6) reveal only two dominant peaks. The first one indicates the yearly seasonal cycle. As we consider the anomalies, a clear sign for the frequency of one year indicates that the yearly cycle did not stay constant with reference to the historical runs. The second peak at a frequency of 2/year can also be explained by seasonal fluctuations.

6.2. El Niño

The El Niño/Southern Oscillation (ENSO) phenomenon is the strongest mode of internal climate variability on interannual timescales. El Niño is characterized by a pronounced positive anomaly in the sea surface temperature of the eastern and central tropical Pacific in the winter season, occurring on average every 4 years [LK09]. It also has a strong impact on the weather in other regions worldwide.

To place points of correlated and anticorrelated segments close to each other in the scatter plot, we consider both cases as a small distance. When selecting a set of points in the embedding as shown in Fig. 7, we observe long-range correlations and anticorrelations also referred to as teleconnections. The region west of South America corresponds to the region where the El Niño phenomenon occurs. The heatmap in Fig. 7 shows that the temperature fields in this region is positively correlated to the temperature in East Australia and to the North of South America. These findings are in agreement with the IPCC report [CRS*21]. Also, we observe a negative correlation to the relative precipitation anomaly in the South of Africa.

To study the frequency of the ENSO phenomenon, we select a subset of the segments and compute the Fourier spectrum for the full time interval as shown in Fig. 3. Here, we observe a large peak at a frequency of 0.12/year, which corresponds to a period of approximately 8 years. This is not a common period for the El Niño phenomenon. Several smaller peaks, instead, show typical period lengths. Additionally, the peak at a frequency of 1/year indicates a change in the seasonal pattern, as the seasonal cycle was removed based on historical simulation runs. To better understand changes over time, we compute the Fourier spectrum for the time period of 2006 to 2035 and compare it to the years from 2070 to 2100 (see Fig. 8). The most obvious change is the appearance of the 1-year

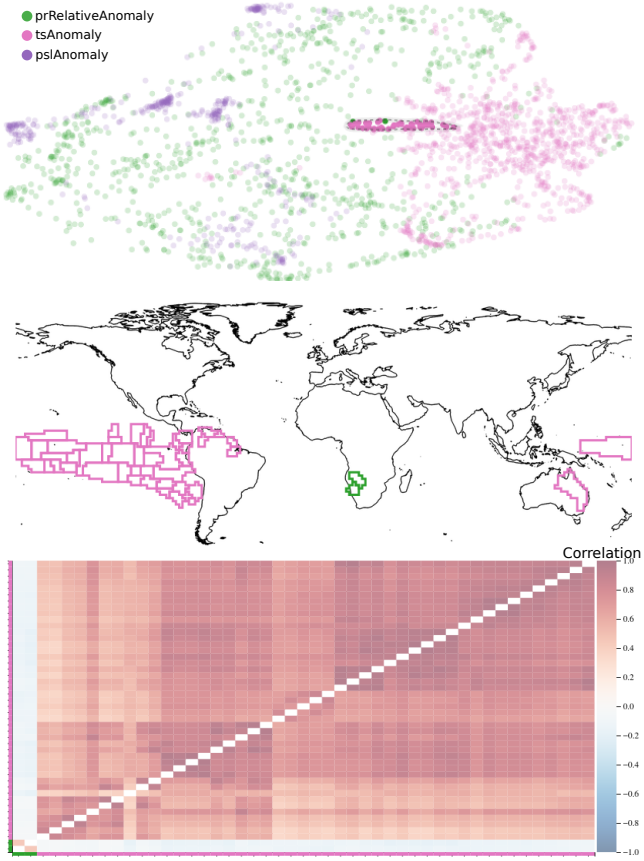

Figure 7: *The selected points correspond to the temperature anomaly in the El Niño region, northern South America, and eastern Australia, and the relative precipitation anomaly in South Africa. The heat map shows a positive correlation (red) between the temperature anomalies and a negative correlation (blue) with the precipitation anomaly.*

periodicity which is barely present in the earlier period. A closer look at the visualization of the time series clarifies that the strength of the seasonal cycle increases over time. Additionally, we observe a small peak indicating a periodicity of 3.2 years for the earlier time interval, while it is located at 2.6 years for the later time interval. This observation might provide a hint towards a shortening of the period length for the El Niño phenomenon which had been observed earlier. However, as these peaks are not very prominent and we observe ensemble means, it might not affect all ensemble members, which requires further investigation.

7. Discussion and Conclusion

We presented a new approach for correlation analysis in multi-field climate ensembles. Starting from an embedding, we identify clusters of highly correlated spatial regions in different fields. Sets of selected segments can be analyzed in different detail visualizations. The usefulness of our approach was shown by analyzing the MPI-GE dataset. We identified well-known climate anomalies, studied

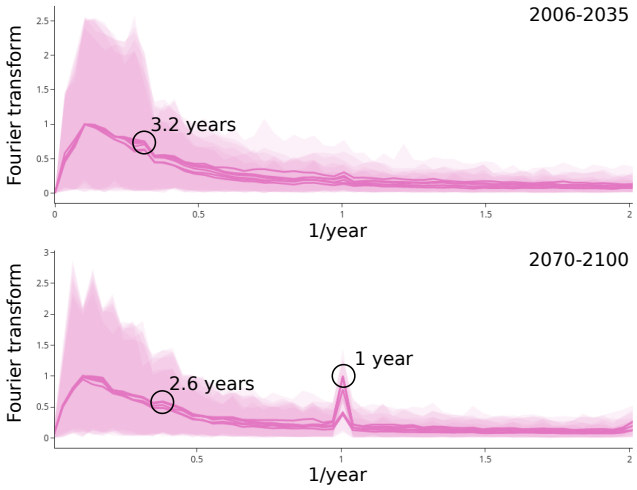

Figure 8: *The Fourier transform of the temperature field changes for different time spans indicate a shortening of the interval between El Niño periods from* 3.2 *years to* 2.6 *years and a change of the seasonal oscillations.*

the correlations and frequencies and also identified some features that could not yet be fully explained, like the dominating 8-year period in the El Niño region. Studying the distribution of the Fourier transform among the ensemble members in more detail provides a future research direction that might lead to additional insights.

Our analysis does not require the choice of a reference point to compute correlations and works on several fields at once. After the preprocessing, we work only on the level of segments. Therefore, the scalability regarding the spatial resolution of the data depends only on the variability of the dataset. With an increasing number of fields, the visualizations get more cluttered. Additionally, a very high number of segments slows down the application as the UMAP projection needs to be applied to more points. However, as the UMAP does not need to be recomputed frequently and due to the selection of fields that should be included in the visualizations, our approach still works on subsets of these kinds of data, where the fields that are included can be changed interactively.

In this work, we have analyzed 2D spatial data from a climate simulation. Nevertheless, all preprocessing steps as well as all visualizations can be directly applied to 3D datasets, except for the map view, which would have to be replaced by a suitable 3D visualization. It is also easy to apply our analysis approach to multi-field simulation ensembles from other fields where correlations between different spatial regions are of interest. We provide our source code at https://github.com/marinaevers/regional-correlations.

Acknowledgements. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) grant 260446826 (LI 1530/21-2).

References

[ACM*19] AGARWAL A., CAESAR L., MARWAN N., MAHESWARAN R., MERZ B., KURTHS J.: Network-based identification and characterization of teleconnections on different scales. *Scientific Reports 9*, 1 (2019), 1–12. 2

[ALI*19] ANTONOV A., LOHMANN G., IONITA M., DIMA M., LINSEN L.: An interactive visual analysis tool for investigating teleconnections in climate simulations. *Environmental Earth Sciences 78*, 10 (2019), 294. 2

[BCHC09] BENESTY J., CHEN J., HUANG Y., COHEN I.: Pearson correlation coefficient. In *Noise reduction in speech processing*. Springer, 2009, pp. 1–4. doi:10.1007/978-3-642-00296-0_5. 3

[BGR*19] BOERS N., GOSWAMI B., RHEINWALT A., BOOKHAGEN B., HOSKINS B., KURTHS J.: Complex networks reveal global pattern of extreme-rainfall teleconnections. *Nature 566*, 7744 (2019), 373–377. 2

[BLLS16] BISWAS A., LIN G., LIU X., SHEN H.-W.: Visualization of time-varying weather ensembles across multiple resolutions. *IEEE Transactions on Visualization and Computer Graphics 23*, 1 (2016), 841–850. 2

[CRS*21] CHEN D., ROJAS M., SAMSET B., COBB K., DIONGUE NIANG A., EDWARDS P., EMORI S., FARIA S., HAWKINS E., HOPE P., HUYBRECHTS P., MEINSHAUSEN M., MUSTAFA S., PLATTNER G.-K., TRÉGUIER A.-M.: *Framing, Context, and Methods*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021, p. 147–286. doi:10.1017/9781009157896.003. 6

[CWMW11] CHEN C.-K., WANG C., MA K.-L., WITTENBERG A. T.: Static correlation visualization for large time-varying volume data. In *2011 IEEE Pacific Visualization Symposium* (2011), IEEE, pp. 27–34. 2

[DWW23] DALELANE C., WINDERLICH K., WALTER A.: Evaluation of global teleconnections in CMIP6 climate projections using complex networks. *Earth System Dynamics 14*, 1 (2023), 17–37. 2

[ECKD21] EKHTIARI N., CIEMER C., KIRSCH C., DONNER R. V.: Coupled network analysis revealing global monthly scale co-variability patterns between sea-surface temperatures and precipitation in dependence on the enso state. *The European Physical Journal Special Topics 230*, 14 (2021), 3019–3032. 2

[EHL21] EVERS M., HUESMANN K., LINSEN L.: Uncertainty-aware visualization of regional time series correlation in spatio-temporal ensembles. In *Computer Graphics Forum* (2021), vol. 40, Wiley Online Library, pp. 519–530. 2, 3, 4

[Epa69] EPANECHNIKOV V. A.: Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications 14*, 1 (1969), 153–158. 5

[HBLW21] HAZARIKA S., BISWAS A., LAWRENCE E., WOLFRAM P. J.: Probabilistic Principal Component Analysis Guided Spatial Partitioning of Multivariate Ocean Biogeochemistry Data. In *Workshop on Visualisation in Environmental Sciences (EnvirVis)* (2021), Dutta S., Feige K., Rink K., Zeckzer D., (Eds.), The Eurographics Association. doi:10.2312/envirvis.20211078. 2

[HBS*20] HORAK T., BERGER P., SCHUMANN H., DACHSELT R., TOMINSKI C.: Responsive matrix cells: A focus+ context approach for exploring and editing multivariate graphs. *IEEE Transactions on Visualization and Computer Graphics 27*, 2 (2020), 1644–1654. 5

[JBMS09] JÄNICKE H., BÖTTINGER M., MIKOLAJEWICZ U., SCHEUERMANN G.: Visual exploration of climate variability changes using wavelet analysis. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 1375–1382. doi:10.1109/TVCG.2009.197. 2

[JPR*04] JEN D., PARENTE P., ROBBINS J., WEIGLE C., TAYLOR R. M., BURETTE A., WEINBERG R.: Imagesurfer: A tool for visualizing correlations between two volume scalar fields. In *IEEE Visualization 2004* (2004), IEEE, pp. 529–536. 2

[KBL19] KAPPE C., BÖTTINGER M., LEITTE H.: Analysis of decadal climate predictions with user-guided hierarchical ensemble clustering. *Computer Graphics Forum 38*, 3 (2019), 505–515. doi:https://doi.org/10.1111/cgf.13706. 2

[KRRW18] KUMPF A., RAUTENHAUS M., RIEMER M., WESTERMANN R.: Visual analysis of the temporal evolution of ensemble forecast sensitivities. *IEEE Transactions on Visualization and Computer Graphics 25*, 1 (2018), 98–108. 2

[LK09] LATIF M., KEENLYSIDE N. S.: El niño/southern oscillation response to global warming. *Proceedings of the National Academy of Sciences 106*, 49 (2009), 20578–20583. doi:10.1073/pnas.0710860105. 6

[MHSG18] MCINNES L., HEALY J., SAUL N., GROSSBERGER L.: Umap: Uniform manifold approximation and projection. *Journal of Open Source Software 3*, 29 (2018). 4

[MML21] MAHER N., MILINSKI S., LUDWIG R.: Large ensemble climate model simulations: introduction, overview, and future prospects for utilising multiple types of large ensemble. *Earth System Dynamics 12*, 2 (2021), 401–418. doi:10.5194/esd-12-401-2021. 2

[MMSG*19] MAHER N., MILINSKI S., SUAREZ-GUTIERREZ L., BOTZET M., DOBRYNIN M., KORNBLUEH L., KRÖGER J., TAKANO Y., GHOSH R., HEDEMANN C., LI C., LI H., MANZINI E., NOTZ D., PUTRASAHAN D., BOYSEN L., CLAUSSEN M., ILYINA T., OLONSCHECK D., RADDATZ T., STEVENS B., MAROTZKE J.: The Max Planck Institute Grand Ensemble: Enabling the exploration of climate system variability. *Journal of Advances in Modeling Earth Systems 11*, 7 (2019), 2050–2069. doi:https://doi.org/10.1029/2019MS001639. 2, 5

[NBD*15] NOCKE T., BUSCHMANN S., DONGES J. F., MARWAN N., SCHULZ H.-J., TOMINSKI C.: Visual analytics of climate networks. *Nonlinear Processes in Geophysics 22*, 5 (2015), 545. 2

[PW12] PFAFFELMOSER T., WESTERMANN R.: Visualization of global correlation structures in uncertain 2d scalar fields. In *Computer Graphics Forum* (2012), vol. 31, Wiley Online Library, pp. 1025–1034. 2

[SG11] SUN Y., GENTON M. G.: Functional boxplots. *Journal of Computational and Graphical Statistics 20*, 2 (2011), 316–334. 4

[STS06] SAUBER N., THEISEL H., SEIDEL H.-P.: Multifield-graphs: An approach to visualizing correlations in multifield scalar data. *IEEE Transactions on Visualization and Computer Graphics 12*, 5 (2006), 917–924. 2

[SWMW09] SUKHAREV J., WANG C., MA K.-L., WITTENBERG A. T.: Correlation study of time-varying multivariate climate data sets. In *2009 IEEE Pacific Visualization Symposium* (2009), IEEE, pp. 161–168. 2

[VHB*21] VIETINGHOFF D., HEINE C., BÖTTINGER M., MAHER N., JUNGCLAUS J., SCHEUERMANN G.: Visual analysis of spatio-temporal trends in time-dependent ensemble data sets on the example of the north atlantic oscillation. In *2021 IEEE 14th Pacific Visualization Symposium (PacificVis)* (2021), IEEE, pp. 71–80. 2

[WHLS19] WANG J., HAZARIKA S., LI C., SHEN H.-W.: Visualization and visual analysis of ensemble data: A survey. *IEEE Transactions on Visualization and Computer Graphics 25*, 9 (2019), 2853–2872. doi:10.1109/tvcg.2018.2853721. 2

[WJ63] WARD JR J. H.: Hierarchical grouping to optimize an objective function. *Journal of the American statistical association 58*, 301 (1963), 236–244. 5

[WTZG18] WANG L., TANG X., ZHANG J., GUAN D.: Correlation analysis for exploring multivariate data sets. *IEEE Access 6* (2018), 44235–44243. 2

[ZCL*20] ZHANG M., CHEN L., LI Q., YUAN X., YONG J.: Uncertainty-oriented ensemble data visualization and exploration using variable spatial spreading. *IEEE Transactions on Visualization and Computer Graphics 27*, 2 (2020), 1808–1818. 2

[ZHQL16] ZHANG H., HOU Y., QU D., LIU Q.: Correlation visualization of time-varying patterns for multi-variable data. *IEEE Access 4* (2016), 4669–4677. 2