DOI: 10.1111/cgf.14326
Eurographics Conference on Visualization (EuroVis) 2021
R. Borgo, G. E. Marai, and T. von Landesberger
(Guest Editors)

Volume 40 (2021), Number 3

Uncertainty-aware Visualization of Regional Time Series Correlation in Spatio-temporal Ensembles

Marina Evers†, Karim Huesmann†, and Lars Linsen

Westfälische Wilhelms-Universität Münster, Germany

Abstract

Given a time-varying scalar field, the analysis of correlations between different spatial regions, i.e., the linear dependence of time series within these regions, provides insights into the structural properties of the data. In this context, regions are connected components of the spatial domain with high time series correlations. The detection and analysis of such regions is often performed globally, which requires pairwise correlation computations that are quadratic in the number of spatial data samples. Thus, operations based on all pairwise correlations are computationally demanding, especially when dealing with ensembles that model the uncertainty in the spatio-temporal phenomena using multiple simulation runs. We propose a two-step procedure: In a first step, we map the spatial samples to a 3D embedding based on a pairwise correlation matrix computed from the ensemble of time series. The 3D embedding allows for a one-to-one mapping to a 3D color space such that the outcome can be visually investigated by rendering the colors for all samples in the spatial domain. In a second step, we generate a hierarchical image segmentation based on the color images. From then on, we can visually analyze correlations of regions at all levels in the hierarchy within an interactive setting, which includes the uncertainty-aware analysis of the region's time series correlation and respective time lags.

1. Introduction

Numerical simulations of mathematical models describing *spatio-temporal* natural phenomena are key to research in various fields of natural sciences. To capture *uncertainties* in the models, one frequently runs *simulation ensembles* using varying model settings. When analyzing the simulation outcome, a main question is what are the spatial structures that form with respect to the given uncertain time series. More precisely, one would be interested in observing whether there are *regions*, i.e., spatially connected components, that have a similar temporal behavior for the entire ensemble, and whether different regions exhibit a common pattern, i.e., whether the temporal behavior of the different regions is similar or exhibit some causality.

Similarity of time series is commonly quantified using a *correlation* measure. Thus, spatial samples within a homogeneous region would have high correlations. Moreover, when considering spatially separate regions with some connection, they would also exhibit high correlations, where both positive and negative correlations can be of interest. In addition, when observing some correlation between two regions with a *time lag*, this is a basis for hypothesizing about some causality relation between the two regions. Correlations between spatial positions are computed pairwise. Thus, when given n spatial samples, the number of correlations is quadratic in n. When the spatial resolution is high, there are many time steps, and the spatio-temporal data are collected for an ensemble, keeping the computation times for pairwise correlation analysis within a feasible range is rather challenging. Furthermore, developing a visual representation that allows for a *global* analysis of the correlations, while still providing the necessary details for an analysis of all facets on demand (including spatial regions at adjustable granularity, temporal behavior, time lags, and uncertainty), is equally challenging.

We propose a two-step visual analysis of global time series correlations of spatial locations for spatio-temporal ensembles. The first step is dedicated to detecting regions of high correlations and visualizing the global correlations. To do so, we compute pairwise correlations of time series ensembles for all spatial samples. The correlation matrix is used to compute a 3D embedding. The 3D embedding is mapped to a color space such that each spatial sample is assigned a color and color differences are related to the pairwise correlations. The outcome of this first step is a visual representation of global correlations in the form of color-coded spatial visualizations. Analyzing the details based on all pairwise correlations is not feasible within an interactive setting for common data sizes. We, therefore, propose to operate on homogeneous regions in our second step. We extract regions using a hierarchical segmentation algorithm. Then, we can operate on the segments to analyze all facets of the data within an interactive visual analysis tool. Based on the

† The authors contributed equally.

hierarchical segmentation output, our tool allows for the analysis of the time series at a continuous level of spatial detail. The time series analysis includes the visualization of trends, time lags, and uncertainties.

Our main contributions can be summarized as follows:

- We propose a global correlation visualization using color-coded spatial visualizations, where color differences approximate correlations.
- We propose a continuous spatial level-of-detail control based on a hierarchical segmentation of global correlations.
- We present an interactive and uncertainty-aware visual analysis of time series correlations including time lags and locally adaptable spatial granularity.

2. Related Work

Analyzing *spatio-temporal simulation ensembles* is complex due to the multifaceted nature of the data [WHLS19,Cro18,KH13,WP09]. Thus, different approaches focus on different facets of the data, for example, they provide visualizations aiming at an understanding of a temporal evolution [HHB16,OJ14,BLLS17,FML16]. Visual analytics approaches for the analysis of spatio-temporal ensemble data often focus on statistical properties [PWB*09,SZD*10]. There are other approaches though that define *spatial regions* in simulation data. Bordoloi et al. [BKS04] propose techniques to segment probability density function data without taking time dependency into account. Kappe et al. [KBL19] use a topology-based feature detection to detect extrema. Köthur et al. [KSDD14] try and compare different similarity measures and clustering methods in the spatial domain and decide to use Euclidean distances and *k*-means. They aim at a comparison between model output data and reference data. Similarly, Shu et al. [SGL*16] use an inversion of the Euclidean distance together with a *k*-means clustering to partition the spatial domain. Topological methods like Morse-Smale complexes have also been used to create a partitioning of scalar data [TFL*17,RWS11,GKK*11,GBP18]. All these approaches try to define spatial regions, but none of them investigates correlations between the regions.

Correlation is a common similarity measure to define the similarity of time series [PDW*14,FSE12]. Sauber et al. [STS06] analyze the correlations between different fields in multi-field scalar data and create a correlation field based on the result. Other approaches also investigate the correlation between different fields [JPR*04,CWMW11]. These approaches do not consider the correlations between different regions. Pfaffelmoser and Westermann [PW13] showed that local correlations contain a lot of information about the structure of the data. A relatively new approach to the analysis of global correlations in climate data is the use of so called climate networks, where a network is established based on the correlation of spatial data samples [ACM*19,BGR*19]. This network is analyzed using methods from complex network theory. Nocke et al. [NBD*15] summarize the state of the art in visual analytics of climate networks. They highlight that occlusion and visual clutter is a common problem and mainly focus on approaches to improve node-link diagrams drawn on a map. Still, for global correlation analysis, occlusion remains an issue when rendering node-link diagrams over spatial domains. Nocke et al. argue that matrix-based visualizations have the disadvantage of being unable to convey the geographic information. We instead propose to use matrix-based visualizations and link them to spatial visualizations, which convey the region information.

In machine learning, *correlation clustering* is a common approach to cluster data based on pairwise correlations [BBC04]. The methods aim to find an optimal number of clusters, where the correlation inside a cluster is maximized and the correlation between different clusters is minimized. Zhang et al. [ZHQL16] build on *k*-means clustering and apply it on a distance metric for time-varying multi-variate data that combines correlations with distances. Sukharev et al. [SWMW09] propose to treat time lines as high-dimensional points and either apply *k*-means directly or reduce the dimensionality using principal component analysis (PCA) and then use an image segmentation method. The latter variant is closest to our correlation-based segmentation approach, but they operate on the sample points directly instead of correlations.

Kumpf et al. [KRRW18] introduced the notion of *correlation cliques*, which are regions of highly correlated points to a chosen seed point, to analyze the sensitivity of weather forecasts based on correlations. They allow for a subdivision of these regions, but do not target the analysis of global correlation schemes. Thus, their approach strongly depends on the choice of seed points. An interactive approach targeting the analysis of correlations in climate data was proposed by Antonov et al. [ALI*19]. Their approach focuses on discovering so-called teleconnections, which are long-range interactions of weather phenomena. However, they do not apply a hierarchical approach which limits the analysis to one level of detail. Additionally, their visualizations might get cluttered quickly because drawing lines on maps quickly leads to occlusion if many lines are needed. Like Kumpf et al., they also allow for selecting a seed point and then only show local correlations to alleviate the occlusion issue, which does not work for global correlation analysis. Jänicke et al. [JBMS09] presented a method to make wavelet analysis applicable to larger datasets that also contains a clustering step. However, they use wavelet analysis as a base instead of full correlations, and their approach does not apply to ensembles.

A more general approach aiming at the analysis of *global correlation* structures was proposed by Pfaffelmoser and Westermann [PW12]. They focus on uncertain spatial data, not on the correlation of uncertain spatio-temporal data, i.e., they did not consider time series correlations. Moreover, they only include one level of spatial subdivision for a more in-detail analysis, which might not be sufficient in some cases, e.g., for large climate simulation data that also require an in-detail analysis of smaller regions. A hierarchical correlation clustering algorithm was proposed by Liebmann et al. [LWS18]. They find a connection between the clustering dendrogram and a joint tree to visualize the clustering as a topological landscape. However, their approach does not generate spatially connected components as in segmentations.

In conclusion, none of the approaches target global correlation analysis for spatio-temporal ensemble data, but either consider spatio-temporal data or uncertain spatial ensemble data. Also, no existing approach allows for an interactive analysis of global time series correlation for regions at a continuous level of spatial detail.

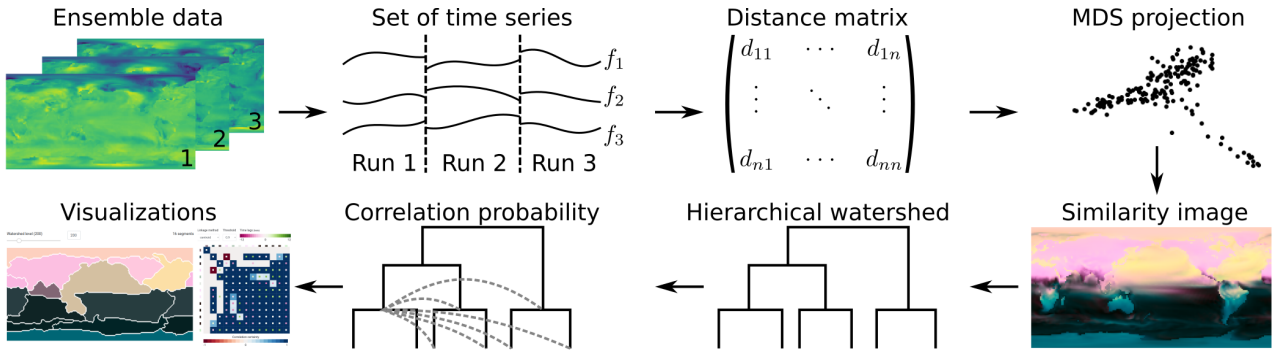

Figure 1: *Overview of processing pipeline from ensemble data via similarity image to interactive hierarchical correlation analysis.*

3. Overview

Given a spatial-temporal ensemble, our approach for the uncertainty-aware visualization of regional time series correlation consists of two main steps that can be further subdivided into a pipeline of sub-steps as depicted in Figure 1.

The first main step is the generation of a color-coded spatial visualization of global correlations (Section 4). The visualization shall provide a global understanding of correlations among spatial samples. We achieve this by assigning to each spatial sample a color such that color distances relate to correlations. For its computation, we first need to define a correlation measure for time series ensembles (Section 4.1). Since we want to assign colors using a 3D color space, the samples' correlations are then mapped to a 3D embedding (Section 4.2). Finally, the samples' positions in the 3D embedding are mapped to colors (Section 4.3), which then are used for the spatial visualization.

The goal of the second main step is to allow for an interactive uncertainty-aware visual analysis of the correlation among spatial regions down to the level of analyzing the relation of time series including time lags (Section 5). It builds upon the first main step by starting with the color-coded spatial visualization and applying a hierarchical segmentation algorithm to it (Section 5.1), which results in a region hierarchy with continuous level of detail. For an in-detail analysis of these regions, we compute multi-level correlations for all regions in the hierarchy including time lags (Section 5.2). The multi-level computation assures that the user can adaptively refine regions during an interactive session. All computation steps up to here can be performed in a pre-processing step. From here on, computations need to fulfill the requirement that they can be executed within an interactive setting. We achieve this by basing all subsequent analysis steps on the region hierarchy only. For the interactive analysis, we use multiple views that link region visualizations (Section 5.3) to correlation visualization (Section 5.4) and uncertainty-aware time series visualizations (Section 5.5) using respective coordinated interactions (Section 5.6).

4. Spatial Visualization of Global Correlation

This section describes the details for the generation of a color-coded spatial visualization of global correlations as introduced in Section 3.

4.1. Time Series Ensemble Correlation

Given n spatial samples and m simulation runs, then we have for each spatial sample $\mathbf{x}_i$ and each simulation run r_k a time series as a function of time t, which we denote by $s_{ik}(t)$. Now, our goal is to define a correlation measure for a time series ensemble to compute the pairwise correlation at two spatial sample locations $\mathbf{x}_i$ and $\mathbf{x}_j$. Thus, we want to compute the correlation between two sets of functions $\{s_{ik}(t) : k = 1,\ldots,m\}$ and $\{s_{jk}(t) : k = 1,\ldots,m\}$. Please note that the number of time steps may differ between simulation runs, but we can assume that two time series $s_{ik}(t)$ and $s_{jk}(t)$ of the same simulation run have the same length.

In general, it is not clear, how a correlation between two sets of functions can be defined. In the case of ensembles, it is obvious that we should compute the correlations for time series that belong to the same simulation run. An immediate choice would then be to compute the correlation for all m time series individually and afterwards compute their mean. However, the mean generally leads to a significant loss of information such as the standard deviation from the mean. Therefore, we decided against using the mean and, instead, propose to use concatenated time series for correlation computation. More precisely, we define the concatenated time series of the m time series given at a spatial sample $\mathbf{x}_i$ piece-wise by

$$f_i(t) = s_{ik}\left(t - \sum_{l=1}^{k-1} T_l\right) \text{ for } \sum_{l=1}^{k-1} T_l < t \leq \sum_{l=1}^{k} T_l \,,$$

where T_k denotes the number of time steps of simulation run r_k with $k = 1,..,m$. The concatenation of the time series is also shown schematically in Figure 1.

Given two concatenated time series $f_i(t)$ and $f_j(t)$, we are now in the position to apply any existing correlation measure. As the Pearson correlation coefficient [BCHC09] is the most commonly applied correlation measure for time series, we decided to use it for our experiments, as well. It describes the linear dependence between the concatenated time series at two samples $\mathbf{x}_i$ and $\mathbf{x}_j$ by

$$C_{ij} = \frac{\sum_p (f_i(t_p) - \mu_i)(f_j(t_p) - \mu_j)}{\sqrt{\sum_p (f_i(t_p) - \mu_i)^2}\sqrt{\sum_p (f_j(t_p) - \mu_j)^2}} \,,$$

where μ_i and μ_j denote the mean of $f_i(t)$ and $f_j(t)$, respectively.

We would like to point out that time lags are not considered here,

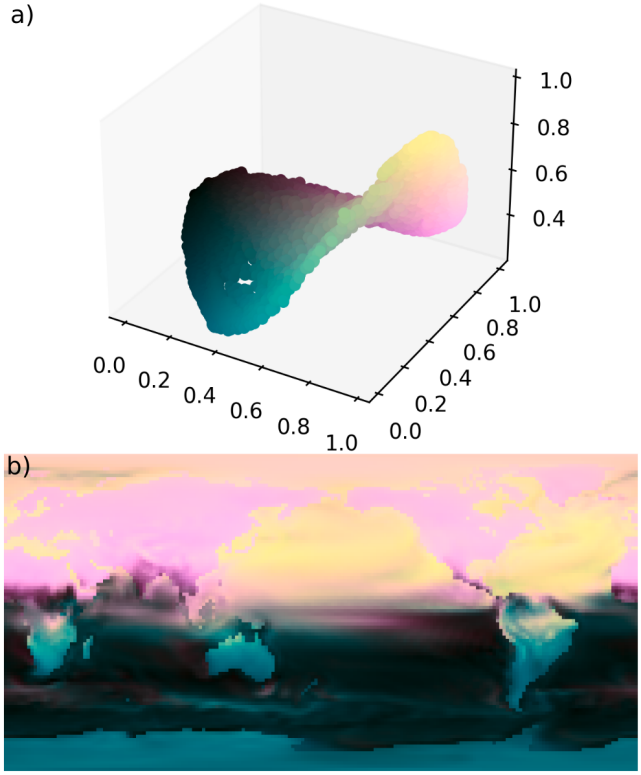

Figure 2: *a) 3D embedding of correlation-based distances between spatial samples with assigned colors. b) Respective similarity image of temperature for global climate simulation. Color differences relate to correlations.*

as two samples that exhibit a high correlation only when applying a time shift shall not be assigned to the same region during the later segmentation step. Additionally, performing the calculation including time lags only on the level of segments, saves a significant amount of computation time.

4.2. 3D Embedding

Having computed the pairwise correlations C_{ij} for all spatial sample pairs, we want to assign colors to the samples such that the color differences relate to the correlations. Since color spaces are three-dimensional, we would like to map the correlations to a 3D space. The values C_{ij} form a correlation matrix. We first normalize the values of the matrix such that its range $[-1,1]$ is scaled to the range $[0,1]$. We transform the normalized correlation matrix to a distance matrix by subtracting it from the one matrix. Thus, the entries of the distance matrix become $d_{ij} = \frac{1-C_{ij}}{2}$. Thus, time series with a strong positive correlation have small distance values and those with a strong negative correlation have large distance values. Similar to our comment above about time lags, we would like to point out that points with a strong negative correlation should not be put into the same region during segmentation.

Given the distance matrix, we would like to find a 3D embedding that preserves relative distances as much as possible. Thus, we are interested in minimizing the stress function

$$\sigma = \sqrt{\sum_{i \neq j} (d_{ij} - \|\mathbf{p}_i - \mathbf{p}_j\|_2)^2} \,,$$

where $\mathbf{p}_i$ denotes the position related to spatial sample $\mathbf{x}_i$ in the 3D embedding. Multi-dimensional scaling (MDS) [Wic03] is a dimensionality reduction method for generating embeddings from distance matrices by minimizing the stress function σ. Consequently, we decided to use MDS as a distance-preserving projection technique to compute points $\mathbf{p}_i$ in the 3D embedding. In general, the distances cannot be fully preserved, if the intrinsic dimensionality is higher than three. However, we will use the resulting visualization as an overview and starting point for in-detail analysis such that small inaccuracies are acceptable here. Moreover, we can perform an investigation of the eigenvalues to estimate the loss, if desired.

4.3. Color Mapping

Given the 3D embeddings $\mathbf{p}_i$ of all spatial samples $\mathbf{x}_i$, we can assign to each point a color in a 3D color space. This can be achieved by simply using the positions of the points in the 3D embedding as coordinates in the 3D color space. Since our goal is to relate color distances to correlations, we should use a color model, where Euclidean distances in color space relate to human perception. We decided to use the CIE $L^*a^*b^*$ color space, as it is designed to be perceptually uniform, i.e., Euclidean distances of points in the color space correspond to perceptual differences of colors.

To map the points $\mathbf{p}_i$ to a color in the CIE $L^*a^*b^*$ color space, we first rotate the points such that the most distant points lie on the diagonal of a cube and then re-scale the points to a range such that we make the best use of the color space. The re-scaling is applied by first scaling the positions $\mathbf{p}_i$ to the unit cube $[0,1]^3$ with the same scaling factor for all axes. Then, we scale the three dimensions by a factor 100 and translate them such that all points lie in the range $[0,100] \times [-50,50] \times [-50,50]$ of the CIE $L^*a^*b^*$ color space. This mapping preserves the relative distances, even though it does not use the complete color range available. However, we use a wide range of colors. An example of a color-coded embedding is shown in Figure 2a. Now, the image of the spatial domain can be shown in the respective colors. The closer the colors appear to an observer, the higher is the correlation between the underlying time series. In the following, we are going to refer to the resulting image as similarity image. An example of a similarity image can be seen in Figure 2b. Since the CIE $L^*a^*b^*$ color space uses colors that may not be perfectly re-producible in the RGB color space, the colors may be a bit off when displaying them on an RGB color screen. However, as this is only the first step of our analysis, where loss had already been introduced by the 3D embeddings, we consider the slightly off colors as being acceptable for our visualizations here. The subsequent segmentation step will benefit from using a larger set of color representations.

5. Interactive Hierarchical Correlation Analysis

This section describes the details for an interactive visual analysis of multi-level correlations within a region hierarchy as introduced in Section 3.

5.1. Hierarchical Segmentation

Given the similarity image as described in the previous section, we want to segment it into roughly homogeneous regions to allow for further analyses based on segment means. A myriad of image segmentation algorithms exists that could be applied. We want to generate a hierarchical segmentation with continuous level of detail, i.e., it should generate a hierarchy with continuously changing granularity. Given the homogeneity of segments at finest granularity, we can assume their smoothness at the highest level of detail. Watershed algorithms are capable to produce such hierarchies. We choose a *graph-based watershed* algorithm [CN11,NCP13], as this algorithm can handle color images, provides large flexibility in the choice of the geometry, and can also cover images defined on a spherical surface like Earth climate data. Moreover, the algorithm itself is very fast with a quasi-linear runtime complexity in the number of pixels. An efficient implementation is available [PCC*19].

We first define an undirected graph $G = (V,E)$ with vertices V and edges E. Each vertex in V corresponds to one pixel in the image. Adjacent vertices are connected by edges. This is shown schematically for a one-dimensional image in Figure 3b. The graph structure allows us to include non-flat geometries like the Earth, where we insert additional edges between the vertices representing the pixels on the left side of the image and those on the right.

Each edge $e = \{v_1, v_2\} \in E$ is weighted by the *gradient* between the two vertices v_1 and v_2 it connects. In the field of image segmentation, a variety of methods for gradient image calculations is available. We propose two methods. The first option is to use the Euclidean distance between color values $f(v)$ of the single pixels associated with vertex $v \in V$. This leads to rather sharp edges and may cause oversegmentation for smooth color transitions, but on the other hand, assigns the pixels accurately for clear edges in the image. The second option is to use a Sobel filter [SF68], which is among the simplest operators for edge detection in images. Here, we first calculate a gradient image, where each pixel contains the gradient magnitude. Each edge is weighted with the mean of the adjacent gradient image pixels. This second method leads to smoother edges than the first one, but is better suited for data with smooth transitions. We compute gradients with respect to the CIE $L^*a^*b^*$ colors for intuitive interpretation, but, in principle, the gradients could be computed in higher-dimensional spaces. The mapping from the 3D embedding to the CIE $L^*a^*b^*$ color space is a linear transformation that does not affect relative distances.

The *algorithm* to create a watershed hierarchy consists of two main steps. First, a binary partition tree is created from the edge-weighted graph. Second, for each edge in the tree, it is marked whether it is a watershed edge or not. An edge is a watershed edge if it merges segments that correspond to different minima. An example is shown in Figure 3c. If the edges are already sorted, this algorithm for the binary partition tree has an asymptotic time complexity of $\mathcal{O}(|E| \cdot \alpha(|V|))$ where $\alpha(\cdot)$ is the inverse of the single-valued Ackermann function. On the resulting binary partition tree, the watershed edges are labeled in a post-processing step that is linear in $|E|$. By removing all edges without labels, we obtain the watershed hierarchy. In this hierarchy, we remove consecutive tree nodes with equal altitudes. Note that the tree representing the hierarchy is not necessarily binary. In the following, we treat the smallest watershed basins as the lowest level of the hierarchy. The whole algorithm is schematically shown in Figure 3.

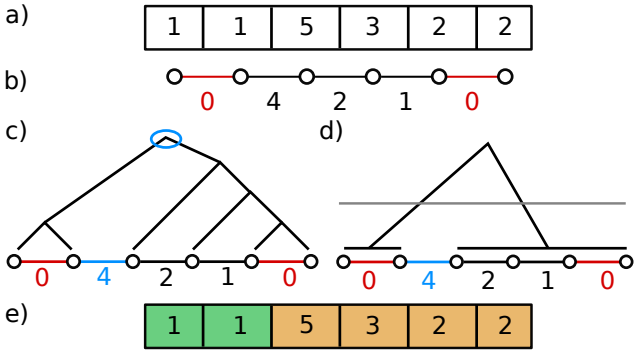

Figure 3: *Example for watershed hierarchy generation: a) 1D image. b) Edge-weighted graph with Euclidean distances as gradients and minima shown in red. c) Binary partition tree with watershed edge (blue). d) Watershed hierarchy with pruning line (grey). e) Watershed segmentation with two segments for pruning line in d).*

We store the obtained hierarchical segmentation in a *tree data structure*, where each node represents one segment. The segments in the lowest level of the hierarchy, i.e., the leaves of our tree, represent the smallest unit on which we will work in the subsequent steps. Thus, the number of nodes in our tree structure does not scale with the resolution of the data, but *scales with the information contained in the data*. Increasing the spatial resolution of the given data without adding further information, does not increase the size of our tree structure.

We store additional *aggregated data* in the nodes of our tree. For leaves, we store the spatial sample points associated with them. For each node, we store means over the segment represented by them. More precisely, we store the mean time series for each ensemble member, to allow for their immediate access during an in-detail analysis at interactive rates. We can assume that the segments at the highest level of detail are by construction sufficiently smooth to work on means. In addition, we store for each segment the minimal and maximal pair-wise correlation of the associated sample points. This allows us to detect projection errors and points the user towards segments with strong internal variations. To also save computation times in the pre-processing stage, we make use of the linearity of the mean: We only calculate the means of the leaves directly from the input data, while for inner nodes and the root we compute the mean of their children's means weighted by the area of the children's segment. The altitude for each node as computed by the watershed algorithm is also stored, which allows for a refinement of single segments based on the watershed level.

5.2. Multi-level Correlations

We calculate pair-wise correlations between the segments for all ensemble members. We can restrict the computations to pairs of segments that do not belong to one path within the tree. Additionally, we make use of the symmetry of the correlation matrix.

To find global correlations, one is interested in regions with a

synchronized behavior, but also in regions where there is a time lag τ between the events hinting towards some causality. Thus, we calculate the *time-shifted cross-correlation* $\rho_{ij,k}(\tau)$ between two time-series $\bar{s}_{ik}(t)$ and $\bar{s}_{jk}(t)$ of ensemble member r_k with time lag τ as

$$\rho_{ij,k}(\tau) = \frac{E[(\bar{s}_{ik}(t) - \bar{\mu}_i)(\bar{s}_{jk}(t+\tau) - \bar{\mu}_j)]}{\bar{\sigma}_i \bar{\sigma}_j},$$

where $E[\cdot]$ is the expected value and $\bar{\sigma}_i$ and $\bar{\sigma}_j$ are the standard deviations of $\bar{s}_{ik}(t)$ and $\bar{s}_{jk}(t)$ respectively. The bars above the variables indicate that these are mean values of the segments. We calculate these correlations for all $\tau \in \{-\tau_{\max}, \ldots, \tau_{\max}\}$ where $\tau_{\max}$ is an application- and data-specific value to be defined by the user.

We compute a total $(2\tau_{\max}+1)m$ correlation values for m ensemble members. We reduce this number by calculating the strongest correlation. Here, strongest correlations could be both positive or negative, thus, we use the maximum absolute value, but store the sign to capture if negative or positive correlations dominate.

As we mainly want to identify whether strong correlations exist, we threshold the correlations values leading to ternary values of strong negative (-1), strong positive (1), or weak (0) correlations. Subsequently, we compute the average of the binary values over all ensemble members and normalize it by the number of ensemble members m. As we assume that the ensemble members are relatively similar, it is very unlikely that there exist ensemble members with strong positive correlations as well as strong negative correlations. If this occurs nevertheless, we alert the user about this issue for further investigation. Otherwise, we obtain an average value, that is either positive representing the percentage of runs with positive correlations above the threshold or negative representing the percentage of runs with negative correlations below the (negative) threshold. This percentage encodes the likelihood of the occurrence of a strong correlation, which relates to the uncertainty within the ensemble.

As it may be of interest to investigate the data for multiple thresholds, we also support this option. These are pre-computations to the interactive visual analysis session, during which the user then can select from the pre-defined thresholds interactively. Calculating lots of different thresholds comes at the cost of large memory requirements, but we expect few thresholds to suffice in all applications we envision.

5.3. Region Visualization

For an in-depth analysis of correlations between regions, it is important that the user can relate the correlations to the regions within the spatial domain. Hence, we support a spatial visualization of the regions. The data sets used within the scope of this paper have 2D spatial domains, which allows us to visualize the regions using a 2D map, see Figure 4. If data sets with 3D spatial domains were used, the 2D maps would need to be replaced by 3D maps and displayed using volume visualizations such as slice-based viewers or even direct volume renderer.

In the 2D map, we show the regions in the form of segments for a user-defined watershed level. The watershed level can be changed interactively to get an overview of the different levels in the segmentation hierarchy. The segments are colored with the mean color

Figure 4: *Region visualization using white borders: Four regions are interactively selected, others are greyed out. Labeled region (center) is shown with black dashed border. Region that is hovered over (right) is shown with pink border.*

of the colors assigned to the sample points in the similarity image (cf. Section 4). This gives the user an impression of the similarity of the segments. A comparison to the similarity image, that is always shown juxtaposed with the segmentation, allows the user to evaluate the quality of the segmentation. The mean color of the segment represents the choice to use the segment's mean of the time series for further correlation investigations. This chosen interpretable color-coding can cause neighboring segments to have similar colors. To assure that segments are clearly distinguishable, we render a white boundary around each segment. The user can interactively label segments for further analysis, where the labels are used in all visualizations. As users typically label segments that are of special interest, we highlight them by replacing the white boundary with a black dashed frame.

Using our continuous level-of-detail hierarchy, segments can be *refined adaptively*. Thus, the dataset can be analyzed following the visual information-seeking mantra "Overview first, zoom and filter, then details-on-demand" [Shn96]. This concept minimizes visual overload in the beginning of the analysis process and supports a correlation analysis on a more global level. Adaptive refinement is performed by interactively selecting a subset of regions and changing the level of detail for the selection only. Thus, segments at different levels of the hierarchy can directly be compared to each other. The respective correlations are all pre-computed and readily available for interactive investigations. Using coordinated interactions with other views (cf. Section 5.6) segments can be filtered. Filtered segments are highlighted by rendering all not selected segments in faint grey, which puts the focus on the filtered segments while still allowing the user to see the boundaries between those segments for context.

5.4. Regional Correlation Visualization

Having defined hierarchical regions, we want to investigate their correlations including time lags and uncertainty. The task is closely related to climate networks known from climate analysis. They are usually visualized by node-link diagrams on top of a map to include the spatial information. However, such visualizations suffer from occlusion, which may lead to severe visual clutter. Visually encoding additional information such as time lags or uncertainty seems impossible. Therefore, we decided to encode the correlations in a separate linked view. Our visualization shows a heatmap of the matrix containing the correlation probabilities which are calculated as described in Section 5.2. The heatmap scales much better to large amount of segments, allows for customization during inter-

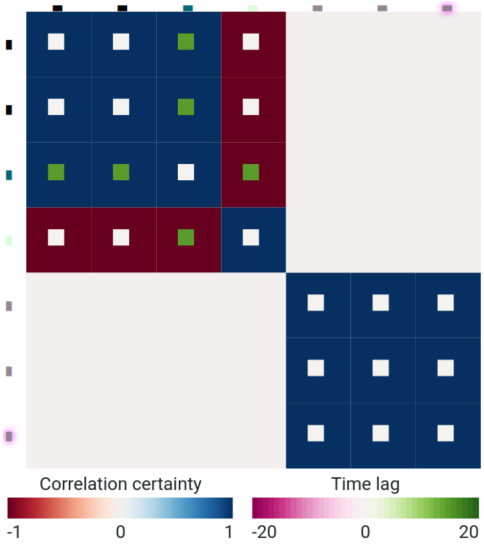

Figure 5: *Correlation heatmap for correlation threshold* 0.8*: Correlation probability is shown using a diverging red-white-blue color map. The colors on the axes correspond to the colors of the segments in the region visualization. The pink highlighting of the lowest row/rightmost column indicates that the corresponding segment in the region visualization is hovered over. The small squares inside of each heatmap field show the time lag for its strongest correlation using a diverging magenta-white-green color map.*

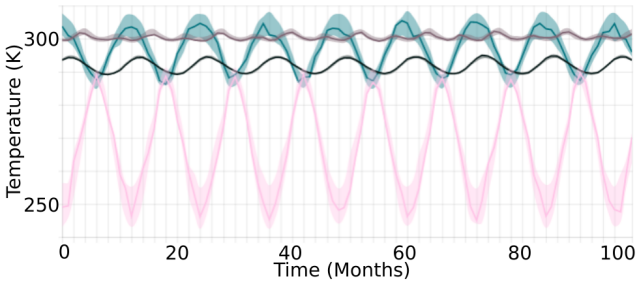

Figure 6: *Uncertainty-aware visualization of time series ensemble displays median curves of the segment means with uncertainty bands depicting the ensemble's variance. Curves are color-coded according to the color of their respective segments.*

active analysis, and can be enhanced with additional information. Relations to spatial information about the regions can be easily established via coordinated interactions (Section 5.6) to the region visualization (Section 5.3) as well as by using a consistent coloring scheme for segments in all views (in addition to the option to assign labels). Using separate views also comes with the advantage of intuitive interactions such as brushing in both views.

The *heatmap* uses color to encode the percentage of ensemble members that surpass the selected correlation threshold, either positively or negatively. Since we want to differentiate between positive and negative correlations a diverging color map should be used. We also support displaying only positive or negative correlations, respectively. The heatmap view can be *customized* in multiple ways. First, since we want to analyze strong correlations, one may reduce the size of the matrix by suppressing rows and columns with no strong correlations. This decision is left to the user though. The size of the heatmap can further be reduced via interactive filtering operations within the heatmap view or in linked views. When regions are selected, we render the selection in the heatmap with full saturation and decrease saturation of all other regions.

To identify groups of highly correlated segments, we use *matrix reordering* methods as they are known from network analysis. The thresholded correlation matrix can also be seen as an adjacency matrix where the percentage of strong correlations among the segments corresponds to the strength of the connection. Our main goal of the matrix reordering is the identification of clusters which could then be investigated in more detail. In a survey about matrix reordering methods, Behrisch et al. [BBHR*16] conclude with a guideline on which clustering algorithm should be used in which application case. They state that for cluster identification, a hierarchical clustering is recommended. We follow this recommendation. As the choice of the linkage scheme used in hierarchical clustering might strongly depend on the underlying dataset, we leave this choice to the user, who can interactively test different linkage algorithms during the exploration of the matrix view. We include (i) single linkage which uses minimal distances between two points, (ii) complete linkage which uses maximal distances, (iii) average linkage with the unweighted pair-group method for arithmetic averages (UPGMA), (iv) linking based on the distance of the centroids of the clusters, (v) median linkage which is similar to the previous method but uses the average of the child centroids instead of calculating the centroid from the original points, and (vi) Ward's minimum variance method. For the hierarchical clustering, we use the implementation provided by SciPy [VGO*20].

Our heatmap encodes probabilities of strong correlations and the type of correlation (positive or negative). In addition, it is of interest to investigate, for what *time lag* τ this correlation was observed. We enhance our heatmap visualization by rendering a color-coded square inside of the field of the matrix that represents the corresponding correlation probability, see Figure 5. Since τ can have positive as well as negative values, a second diverging color map should be used that is distinguishable from the one used above and automatically fitted to $[-\tau_{max}, \tau_{max}]$. It is also possible to select a certain time lag and only show the regions that have a strong positive or negative correlation (above the threshold) with the selected time lag.

5.5. Uncertainty-aware Time Series Visualization

Having detected strong correlations between selected regions, one would then still like to see the actual time series. Thus, we provide a direct visualizations of the time series data as graphs of a function over time. We visualize the time series means over the chosen segments. To convey the uncertainty captured by the ensemble, we visualize their statistical data using an approach that was inspired by functional boxplots [SG11], see Figure 6. As for functional boxplots, we determine the median by ordering the time series by band

Figure 7: *Correlation analysis for synthetic dataset. a) Regional characteristics of time series (ground truth) for ensemble generation. Underlying functions are shown here, too. b) Similarity image for global correlations. c) Segmentation of similarity image matches ground-truth regions.*

depth choosing the one with the largest band depth [SP10]. This corresponds to the most central time series. We surround the median by bands that capture the variation within the ensemble. However, here we deviate from the functional boxplots approach, since we experienced the ensemble data to be rather noisy, which led to a large number of outliers in functional boxplots. Thus, we decided to instead display bands that spread to the full range that is covered by the ensemble members, which is also referred to as the 100% band.

To link this view to the other views, we again use the colors that were assigned to the regions also for the respective time series. Even though this might lead to rather similar colors for highly correlated segments, we found this not to be problematic and noticed that the different colors for negatively correlated or uncorrelated time series facilitate the analysis.

5.6. Coordinated Interactions

The region visualization (Section 5.3), the correlation heatmap (Section 5.4), and the time series view (Section 5.5) display different facets of the spatio-temporal ensemble data. Thus, the views complement each other. For a comprehensive analysis, these facets have to be considered jointly, which we support by *coordinated interactions* between the three views, which are best documented in the accompanying video.

Hovering over the region map directly *highlights* the respective region in the heatmap, and vice versa. The highlighting of a segment using a pink frame is shown in Figure 4. In order to *select* individual regions, we can directly select them in the region map or brush in the heatmap. Not selected regions are de-emphasized by attenuating the colors. Time series of selected regions can be investigated in the time series view. We also support selecting a specific time lag or a specific correlation probability by clicking the respective colors in the color legends at the bottom of Figure 5. Having made a selection, the data can be *filtered* accordingly. To easily go back to the previous level of detail, a new tab is created for the analysis of the filtered data, where the same interaction mechanisms are available. In this new tab, the non-selected segments are greyed out in the region map and are hidden in the heatmap, which frees space in this view for the in-depth investigation of the selection. Such an in-depth investigation may, for example, involve an adaptive refinement of the regions by adjusting the watershed level. The filtering mechanisms can be executed consecutively at different levels of detail, thus, supporting a top-down adaptive hierarchical investigation. At every point of the investigation, one can easily switch back to the previous level of detail by just going back to the respective tab, thus providing *provenance* information.

6. Results

In the following, we will demonstrate the effectiveness of our methods by applying them to a synthetic dataset with known ground truth (Section 6.1) as well as to a real-world dataset from climate simulations (Section 6.2) which we discuss with a domain expert.

6.1. Synthetic Dataset

We *verify* our approach using a synthetic dataset. For each spatial sample, we create a time series using different functions as depicted in Figure 7a. Each function in is sampled at 300 time steps with a resolution of 0.1. We generate 10 ensemble members by adding noise (with uniform distribution between $[0, 0.1]$) to the respective functions. The resulting *similarity image* can be seen in Figure 7b. Here we observe 5 different colors, where the color belonging to region R6 is similar but not identical to that of R4, R5, and R8, which was expected as they have a correlation > 0.8. Colors of regions R1 and R7 are identical though, as their time series are perfectly correlated. Region R2, which by definition is anti-correlated to R1, R7, and R9 exhibits a quite different color. When applying our proposed segmentation technique to this similarity image with watershed level 20, we obtain a *segmentation* as shown in Figure 7c. The detected segments are identical to the ground truth in Figure 7 a). To verify that our segmentation works correctly for circular datasets as in global climate science, we also assume a circular domain here. In our segmentation, regions R1 and R9 as well as R3 and R8 form indeed connected regions. Regions R3, R4, and R8 share the same function, but due to the spatial distance between R4 to R3/R8 they indeed are represented by two separated segments.

The corresponding *correlation heatmap* is shown in Figure 5. The lower-right block shows that the gray segments (R3/R8, R4, and R6) are strongly correlated (≥ 0.8) to each other, but not to any other time series, which is the expected result. When increasing the

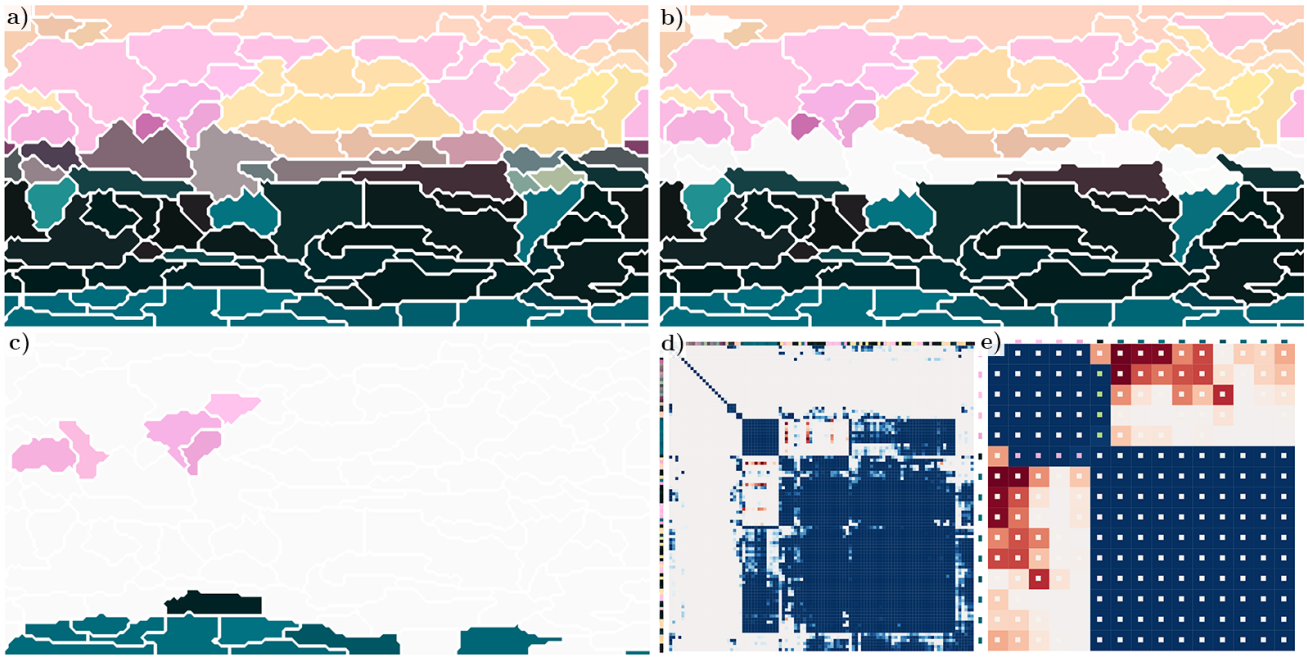

Figure 8: *Correlation analysis for MPI-GE temperature field: a) Regions visualization for watershed level 20. b) Regions visualization filtered for regions with 6-months time-lag correlations. c) Regions visualization filtered for regions with anti-correlations. d) Correlation heatmap for selected regions of a) with threshold* 0.9. *e) Correlation heatmap for selected regions of b) with threshold* 0.9.

threshold, we can observe that the correlation between the regions with a linear (R3/R8 and R4) and a quadratic function (R6) vanishes. As expected, the orange-colored region (R2) shows an anti-correlation to the black segments (R1 and R7). The cyan-colored segment (R5) is anti-correlated to R2 and correlated to R1, R9, and R7 with a time lag of 15, which is equivalent to shifting the curve by approximately $\pi/2$. We can conclude that all observations are in line with our expectations and that our proposed approach is working for this kind of dataset.

6.2. Global Climate Simulation

We also applied our approach to the Max Planck Institute Grand Ensemble [MMSG*19] dataset (MPI-GE) and discussed the results with a scientist who has been working with climate data for decades. With 100 members for each scenario, MPIM-GE is to date one of the largest climate simulation ensembles worldwide. We first used the mean monthly surface air temperature fields of the simulation of scenario RCP8.5, which covers the years 2006-2099 at a spatial resolution of 192x96.

First, we looked into the raw temperature fields. The similarity image in Figure 2b exhibits a prominent difference in the perceived colors between the northern and southern hemisphere. The northern regions have a bright color, the southern regions a dark color, which highlights the strong difference of the underlying time series. The boundaries between continents and oceans are clearly visible, e.g., the contours of the northern continents can be identified by the pink, the southern continents by the blue colors. For our analysis we start with watershed level 20, dividing the similarity image into 111 segments, see Figure 8a. The contours of the continents are still recognizable. In the *correlation heatmap* in Figure 8d, some segments with anti-correlations (red) stand out. We *interacted* with the *time lag* color map to choose a time lag of 6 months to select all segments that show the corresponding time-shifted correlation to another segment, see Figure 8b. We obtain the northern and southern hemispheres without most equatorial regions. This time lag of 6 months between North and South is very plausible due to the natural seasonality of the Earth's global temperatures. We also filtered regions by *negative correlations* to obtain the segmentation and correlation heatmap in Figures 8 c) and e), respectively. Each of these segments shows a negative correlation ≤ -0.9 to at least one other segment, but we observe that the uncertainty of the correlation varies. Some regions in North Africa and East Asia are strongly (≤ -0.9) anti-correlated with Antarctica. The northernmost Antarctic segment shows a strong correlation with the shown pink segments, but only when applying a time lag of 5 months (pink/green squares in the heatmap's corresponding rows/columns). Thus, the correlation with time lag of 5 months is larger than the given anti-correlation without time lag.

The pointwise variance of monthly temperature data is mostly dominated by the seasonal cycle. In order to enable analysis of climate variability and changes, we derived anomalies with respect to the climatological mean monthly values of the period 2006-2015. In order to study the climate variability and its related changes due to increasing greenhouse gas concentrations, we focus on the derived temperature anomaly data in our second analysis. Observing

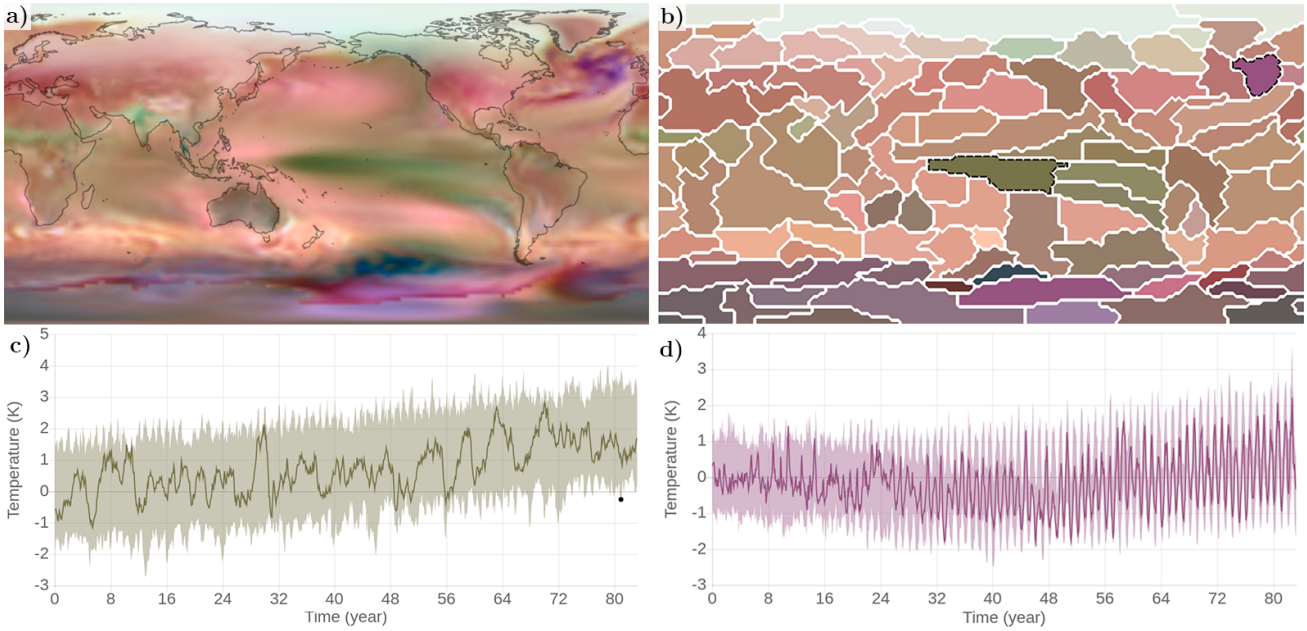

Figure 9: *Analysis of the temperature anomalies: a) Similarity image. We overlaid coastlines for better orientation. b) Region visualization for watershed level 20. Two interesting segments are selected (dashed border) for their time series visualization showing the El Niño (c) and a warming hole phenomenon (d).*

the similarity image as depicted in Figure 9a, we observe that many regions display a similar color, while some exhibit distinct different colors. First, we select the green region in the center of the image (Pacific Ocean) using the segmentation in Figure 9b, and analyze its time series in Figure 9c. While the plot reveals a large variation in the ensemble, the mean curve shows a repeating increase in the temperature anomaly every 3-5 years, which can be explained by the El Niño phenomenon. Another region that stands out and has caught the domain scientist's interest is a dark purple area in the top-right corner of the similarity image (North Atlantic Ocean). This region is known as a "warming hole", i.e., a region that does not encounter much warming or is even slightly cooled in a globally warming climate. This phenomenon is likely linked to ocean circulation changes [KMJ*20]. We investigate this behavior in more detail by selecting the corresponding segment, see Figure 9b and 9d. After around 30 years, we observe a small increase of temperature accompanied by a prominent annual oscillation with decreasing ensemble uncertainty. The domain expert pointed out that our tool is helpful to find characteristic climate phenomena. While he immediately recognized some well-known patterns, he acknowledged that our approach might be helpful to observe new phenomena. For future work, he suggested to include a more in-detail analysis of climate research attributes to better investigate phenomena like the increased annual fluctuations in the northern Atlantic.

7. Discussion and Conclusion

We presented a new approach for global correlation analysis in spatio-temporal ensembles. In a first step, we color-code the correlations in a spatial visualization. In a second step, we hierarchically segment the spatial domain and continue with an in-detail analysis on regions with adaptive refinement. We proposed an interactive visual analysis with coordinate views on different facets of the spatio-temporal ensembles. We presented examples to demonstrate the functionality and effectiveness of our approach. The only *parameters* we introduced are the maximum time lag and the correlation threshold. Those depend on the application and should be left to the domain expert user. Our approach does not depend on the choice of methodological parameters such as weights to combine correlation with spatial information (cf. [ZHQL16]), which are non-intuitive for domain experts and may affect the stability of the approach. Concerning *scalability* of our approach, the bottleneck is the creation of the 3D embedding. The size of the distance matrix grows quadratically with the number of grid points of the simulation data. However, if this becomes an issue, there exist solutions that allow the application of MDS for large data sets [Pla05]. The segmentation scales quasi-linearly in the number of grid points. After segmentation, the approach scales quadratically in the number of regions in the hierarchy, which generally is much lower than the number of spatial samples. Hence, analyses can be performed within an interactive setting. While the presented examples had 2D spatial domains, all steps in principle also work for *3D spatial domains*. The main challenge would be to find a suitable visualization strategy for the segmentation that provides a global overview without suffering from occlusion. We provide our source code at `https://github.com/marinaevers/regional-correlations`.

Acknowledgements We would like to thank Michael Böttinger for his valuable ideas and feedback. 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. doi:10.1038/s41598-019-45423-5. 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. doi:10.1007/s12665-019-8295-z. 2

[BBC04] BANSAL N., BLUM A., CHAWLA S.: Correlation clustering. *Machine Learning 56*, 1-3 (2004), 89–113. doi:10.1023/b:mach.0000033116.57574.95. 2

[BBHR*16] BEHRISCH M., BACH B., HENRY RICHE N., SCHRECK T., FEKETE J.-D.: Matrix reordering methods for table and network visualization. In *Computer Graphics Forum* (2016), vol. 35, Wiley Online Library, pp. 693–716. doi:10.1111/cgf.12935. 7

[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. doi:10.1038/s41586-018-0872-x. 2

[BKS04] BORDOLOI U. D., KAO D. L., SHEN H.-W.: Visualization techniques for spatial probability density function data. *Data Science Journal 3* (2004), 153–162. doi:10.2481/dsj.3.153. 2

[BLLS17] BISWAS A., LIN G., LIU X., SHEN H.: Visualization of time-varying weather ensembles across multiple resolutions. *IEEE Transactions on Visualization and Computer Graphics 23*, 1 (2017), 841–850. doi:10.1109/TVCG.2016.2598869. 2

[CN11] COUSTY J., NAJMAN L.: Incremental algorithm for hierarchical minimum spanning forests and saliency of watershed cuts. In *International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing* (2011), Springer, pp. 272–283. 5

[Cro18] CROSSNO P.: Challenges in visual analysis of ensembles. *IEEE Computer Graphics and Applications 38*, 2 (2018), 122–131. doi:10.1109/mcg.2018.021951640. 2

[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. doi:10.1109/pacificvis.2011.5742369. 2

[FML16] FOFONOV A., MOLCHANOV V., LINSEN L.: Visual analysis of multi-run spatio-temporal simulations using isocontour similarity for projected views. *IEEE Transactions on Visualization and Computer Graphics 22*, 8 (2016), 2037–2050. doi:10.1109/TVCG.2015.2498554. 2

[FSE12] FREY S., SADLO F., ERTL T.: Visualization of temporal similarity in field data. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2023–2032. doi:10.1109/tvcg.2012.284. 2

[GBP18] GYULASSY A., BREMER P.-T., PASCUCCI V.: Shared-memory parallel computation of morse-smale complexes with improved accuracy. *IEEE Transactions on Visualization and Computer Graphics 25*, 1 (2018), 1183–1192. doi:10.1109/tvcg.2018.2864848. 2

[GKK*11] GYULASSY A., KOTAVA N., KIM M., HANSEN C. D., HAGEN H., PASCUCCI V.: Direct feature visualization using morse-smale complexes. *IEEE Transactions on Visualization and Computer Graphics 18*, 9 (2011), 1549–1562. doi:10.1109/tvcg.2011.272. 2

[HHB16] HAO L., HEALEY C. G., BASS S. A.: Effective visualization of temporal ensembles. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (2016), 787–796. doi:10.1109/tvcg.2015.2468093. 2

[JBMS09] JANICKE H., BOTTINGER 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 II R. M., BURETTE A., WEINBERG R.: Imagesurfer: A tool for visualizing correlations between two volume scalar fields. In *Proceedings of the Conference on Visualization '04* (USA, 2004), VIS '04, IEEE Computer Society, p. 529–536. URL: https://doi.org/10.1109/VISUAL.2004.46, doi:10.1109/VISUAL.2004.46. 2

[KBL19] KAPPE C. P., BÖTTINGER M., LEITTE H.: Topology-based Feature Detection in Climate Data. In *Workshop on Visualisation in Environmental Sciences (EnvirVis)* (2019), Bujack R., Feige K., Rink K., Zeckzer D., (Eds.), The Eurographics Association. doi:10.2312/envirvis.20191099. 2

[KH13] KEHRER J., HAUSER H.: Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE Transactions on Visualization and Computer Graphics 19*, 3 (2013), 495–513. doi:10.1109/tvcg.2012.110. 2

[KMJ*20] KEIL P., MAURITSEN T., JUNGCLAUS J., HEDEMANN C., OLONSCHECK D., GHOSH R.: Multiple drivers of the north atlantic warming hole. *Nature Climate Change 10*, 7 (2020), 667–671. 10

[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. doi:10.1109/tvcg.2018.2864901. 2

[KSDD14] KÖTHUR P., SIPS M., DOBSLAW H., DRANSCH D.: Visual analytics for comparison of ocean model output with reference data: Detecting and analyzing geophysical processes using clustering ensembles. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 1893–1902. doi:10.1109/tvcg.2014.2346751. 2

[LWS18] LIEBMANN T., WEBER G. H., SCHEUERMANN G.: Hierarchical correlation clustering in multiple 2d scalar fields. In *Computer Graphics Forum* (2018), vol. 37, Wiley Online Library, pp. 1–12. doi:10.1111/cgf.13396. 2

[MMSG*19] MAHER N., MILINSKI S., SUAREZ-GUTIERREZ L., BOTZET M., DOBRYNIN M., KORNBLUEH L., KRÖGER J., TAKANO Y., GHOSH R., HEDEMANN C., ET AL.: 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:10.1029/2019ms001639. 9

[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. doi:10.5194/npg-22-545-2015. 2

[NCP13] NAJMAN L., COUSTY J., PERRET B.: Playing with kruskal: Algorithms for morphological trees in edge-weighted graphs. In *International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing* (2013), Springer, pp. 135–146. doi:10.1007/978-3-642-38294-9_12. 5

[OJ14] OBERMAIER H., JOY K. I.: Future challenges for ensemble visualization. *IEEE Computer Graphics and Applications 34*, 3 (2014), 8–11. doi:10.1109/mcg.2014.52. 2

[PCC*19] PERRET B., CHIERCHIA G., COUSTY J., AES S. F. G., KENMOCHI Y., NAJMAN L.: Higra: Hierarchical graph analysis. *SoftwareX 10* (2019), 1–6. doi:10.1016/j.softx.2019.100335. 5

[PDW*14] POCO J., DASGUPTA A., WEI Y., HARGROVE W., SCHWALM C., COOK R., BERTINI E., SILVA C.: Similarityexplorer: A visual inter-comparison tool for multifaceted climate data. In *Computer Graphics Forum* (2014), vol. 33, Wiley Online Library, pp. 341–350. doi:10.1111/cgf.12390. 2

[Pla05] PLATT J.: Fastmap, metricmap, and landmark mds are all nystrom algorithms. In *AISTATS* (2005). 10

[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. doi:10.1111/j.1467-8659.2012.03095.x. 2

[PW13] PFAFFELMOSER T., WESTERMANN R.: Correlation visualization for structural uncertainty analysis. *International Journal for Uncertainty Quantification 3* (2013), 171–186. doi:10.1615/int.j.uncertaintyquantification.2012003934. 2

[PWB*09] POTTER K., WILSON A., BREMER P.-T., WILLIAMS D., DOUTRIAUX C., PASCUCCI V., JOHNSON C. R.: Ensemble-vis: A framework for the statistical visualization of ensemble data. In *2009 IEEE International Conference on Data Mining Workshops* (2009), IEEE, pp. 233–240. doi:10.1109/icdmw.2009.55. 2

[RWS11] ROBINS V., WOOD P. J., SHEPPARD A. P.: Theory and algorithms for constructing discrete morse complexes from grayscale digital images. *IEEE Transactions on Pattern Analysis and Machine Intelligence 33*, 8 (2011), 1646–1658. doi:10.1109/tpami.2011.95. 2

[SF68] SOBEL I., FELDMAN G.: A 3x3 isotropic gradient operator for image processing. *a talk at the Stanford Artificial Project in* (1968), 271–272. 5

[SG11] SUN Y., GENTON M. G.: Functional boxplots. *Journal of Computational and Graphical Statistics 20*, 2 (2011), 316–334. doi:https://doi.org/10.1198/jcgs.2011.09224. 7

[SGL*16] SHU Q., GUO H., LIANG J., CHE L., LIU J., YUAN X.: Ensemblegraph: Interactive visual analysis of spatiotemporal behaviors in ensemble simulation data. In *2016 IEEE Pacific Visualization Symposium (PacificVis)* (2016), IEEE, pp. 56–63. doi:10.1109/pacificvis.2016.7465251. 2

[Shn96] SHNEIDERMAN B.: The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages* (1996), p. 336. 6

[SP10] SEABOLD S., PERKTOLD J.: Statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference* (2010). doi:10.25080/majora-92bf1922-011. 8

[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. doi:10.1109/tvcg.2006.165. 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. doi:10.1109/pacificvis.2009.4906852. 2

[SZD*10] SANYAL J., ZHANG S., DYER J., MERCER A., AMBURN P., MOORHEAD R.: Noodles: A tool for visualization of numerical weather model ensemble uncertainty. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (2010), 1421–1430. doi:10.1109/tvcg.2010.181. 2

[TFL*17] TIERNY J., FAVELIER G., LEVINE J. A., GUEUNET C., MICHAUX M.: The Topology ToolKit. *IEEE Transactions on Visualization and Computer Graphics 24*, 1 (2017), 832–842. doi:10.1109/tvcg.2017.2743938. 2

[VGO*20] VIRTANEN P., GOMMERS R., OLIPHANT T. E., HABERLAND M., REDDY T., COURNAPEAU D., BUROVSKI E., PETERSON P., WECKESSER W., BRIGHT J., VAN DER WALT S. J., BRETT M., WILSON J., MILLMAN K. J., MAYOROV N., NELSON A. R. J., JONES E., KERN R., LARSON E., CAREY C. J., POLAT İ., FENG Y., MOORE E. W., VANDERPLAS J., LAXALDE D., PERKTOLD J., CIMRMAN R., HENRIKSEN I., QUINTERO E. A., HARRIS C. R., ARCHIBALD A. M., RIBEIRO A. H., PEDREGOSA F., VAN MULBREGT P., SCIPY 1.0 CONTRIBUTORS: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods 17* (2020), 261–272. doi:10.1038/s41592-019-0686-2. 7

[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

[Wic03] WICKELMAIER F.: An introduction to mds. *Sound Quality Research Unit, Aalborg University, Denmark 46*, 5 (2003), 1–26. 4

[WP09] WILSON A. T., POTTER K. C.: Toward visual analysis of ensemble data sets. In *Proceedings of the 2009 Workshop on Ultrascale Visualization - UltraVis '09* (2009), ACM, ACM Press, pp. 48–53. doi:10.1145/1838544.1838551. 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. doi:10.1109/access.2016.2601339. 2, 10