Visual Analysis of Spatio-Temporal Trends in Time-Dependent Ensemble Data Sets on the Example of the North Atlantic Oscillation

Dominik Vietinghoff[*]
Leipzig University

Christian Heine[†]
Leipzig University

Michael Böttinger[‡]
Deutsches Klimarechenzentrum

Nicola Maher[§]
Max Planck Institute for Meteorology

Johann Jungclaus[¶]
Max Planck Institute for Meteorology
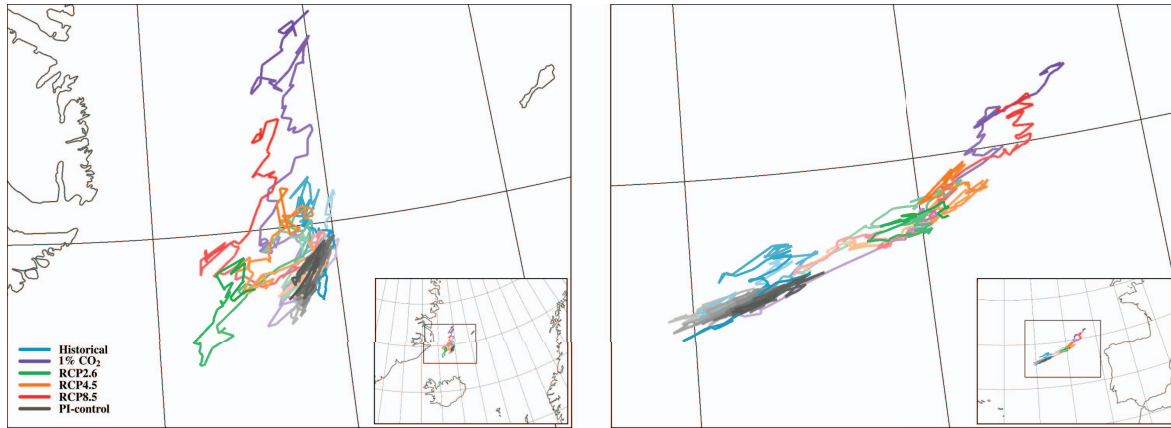
Gerik Scheuermann[‖]
Leipzig University

Figure 1: Systematic movement of the Icelandic Low (left) and the Azores High (right) for different global warming scenarios. Paths were derived using a 30-year sliding window approach and cover the time periods 1850–2004 (historical), 1850–1998 (1% CO_2), 2005–2099 (RCP2.6, RCP4.5, RCP8.5), and 1850–2099 (PI-control). Increasing color saturation denotes advancing time.

ABSTRACT

A driving factor of the winter weather in Western Europe is the North Atlantic Oscillation (NAO), manifested by fluctuations in the difference of sea level pressure between the Icelandic Low and the Azores High. Different methods have been developed that describe the strength of this oscillation, but they rely on certain assumptions, e.g., fixed positions of these two pressure systems. It is possible that climate change affects the mean location of both the Low and the High and thus the validity of these descriptive methods. This study is the first to visually analyze large ensemble climate change simulations (the MPI Grand Ensemble) to robustly assess shifts of the drivers of the NAO phenomenon using the uncertain northern hemispheric surface pressure fields. For this, we use a sliding window approach and compute empirical orthogonal functions (EOFs) for each window and ensemble member, then compare the uncertainty of local extrema in the results as well as their temporal evolution across different CO_2 scenarios. We find systematic northeastward shifts in the location of the pressure systems that correlate with the simulated warming. Applying visualization techniques for this analysis was not straightforward; we reflect and give some lessons learned for the field of visualization.

—————————————————————

[*]e-mail: vietinghoff@informatik.uni-leipzig.de

[†]e-mail: heine@informatik.uni-leipzig.de

[‡]e-mail: boettinger@dkrz.de

[§]e-mail: nicola.maher@colorado.edu

[¶]e-mail: johann.jungclaus@mpimet.mpg.de

[‖]e-mail: scheuermann@informatik.uni-leipzig.de

1 INTRODUCTION

The North Atlantic Oscillation (NAO) is an oscillating meteorological phenomenon that strongly affects the weather in Western Europe in wintertime (cf. Deser *et al.* [2]). It is characterized by the difference in sea level pressure (SLP) of the Azores High and that of the Icelandic Low. In the case of a strong positive NAO, both pressure systems are enhanced. This causes a strengthening of both the clockwise rotational flow of air around the Azores High and the counterclockwise rotation of air around the Icelandic Low. As a result, the flow of mild humid air towards Western Europe is enhanced, causing a milder winter season (typically December, January, February, and March, abbreviated as DJFM). Conversely, if the pressure difference is weak, less mild and humid air from the Atlantic is transported to Europe, resulting in a colder than average winter. This constellation is called negative NAO. For a more comprehensive overview of the NAO, we refer to Hurrell *et al.* [10].

In the past, different approaches have been proposed to express the NAO phenomenon as a dimensionless index. Most approaches fall within one of two classes: those based on differences of pressure levels at two fixed geographic locations or regions [9, 11, 22, 30] and those based on linear regression by finding the first empirical orthogonal function (EOF) of a time series of sea level pressure differences [1, 18, 23]. Figure 2 depicts such a leading EOF computed over the winter months of 2030/31–2059/60 in the area shown. It features a dipole pattern between two anticorrelated areas, one located to the north of Iceland ① and the other one to the northeast of the Azores ②, which is characteristic for the NAO. We will refer to the local extrema of these areas as their *centers of action* (COA).

As global warming is expected to cause changes in the atmospheric circulation, it would be useful to know whether the NAO is also affected, and, in particular, if the centers of these anticorrelated areas move. In that case, using methods that employ fixed geographic regions or locations to derive the NAO index can mislead.
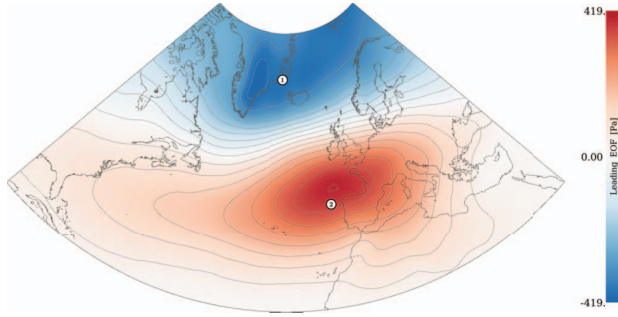
Figure 2: Leading EOF of SLP over the geographic region 90°W–40°E, 20°–80°N and the winter periods of years 2030/31 to 2059/60. Isocontours are drawn at 50 Pa intervals.

Already in 1999, Ulbrich and Christoph [28] found a shift of the NAO in a simulated warming climate. Later, Jung *et al.* [13] and Wang *et al.* [33] confirmed shifts in the NAO in observational data and Hu and Wu [8] in another global warming simulation. However, for their simulation-based studies, both Ulbrich and Christoph [28] and Hu and Wu [8] analyzed the time series of a single simulation run, even though the NAO is a highly variable phenomenon. This raises the question whether the simulated and empirically observed changes are robust shifts or due to random movement.

In this paper, we explore the added value of ensemble simulations to discriminate between internally generated and externally enforced changes of the NAO and to observe whether the movement of the anticorrelated areas is more pronounced in scenarios of stronger climate change. For this, we use one of the largest climate change ensembles that have been produced to date, the Max Planck Institute Grand Ensemble (MPI-GE) [16]. It comprises 100 ensemble members for different scenarios and includes a long preindustrial control simulation with a stable climate. Abstractly, the data can be described as multiple ensembles of 2D time-dependent scalar fields, and we aim to analyze and compare their spatio-temporal trends of oscillating phenomena. Our approach combines several techniques. For each ensemble member separately, we determine the first empirical orthogonal function (EOF) for subsequences of time-steps using a sliding window approach. Describing the NAO via the first EOFs does not rely on fixed locations, but requires a time-dependent signal from which the variation can be estimated. The sliding window approach trades off precision in time with the fact that climate phenomena can only be identified on a multidecadal perspective. Our method then aggregates the information of corresponding time windows by applying local extrema detection for ensemble data according to the method by Mihai and Westermann [17]. This gives confidence regions for the true extrema of the EOF within each time window and allows us to depict the movement of these confidence regions in an animation. To integrate the movement information in one image, we track the weighted barycenter of these regions and depict the movement of these centers as lines. The application of this approach on sliding time windows covering the entire simulation period reveals that the EOF does change, but not as strongly as Ulbrich and Christoph [28] and Hu and Wu [8] found. The movements were found to be more pronounced when climate change is stronger, and, in particular, that no significant movement occurs when climate remains stable.

The remainder of this paper is structured as follows: After briefly surveying related work on the visualization of ensembles of time-dependent scalar fields in Section 2, we give background on the used methods in Section 3. In Section 4, we describe our workflow for the extraction and tracking of dominant local extrema in EOFs of ensemble data sets. In Section 5, we apply our method to an example in climate research to reveal a temporal shift of the Icelandic Low and the Azores High due to global warming. We discuss our approach and give some lessons learned in Section 6. Finally, we conclude the paper in Section 7.

2 RELATED WORK

Several efforts have been made regarding the analysis and visualization of ensemble and/or time-varying data. Recently, the state of the art in this field has been summarized by Wang *et al.* [31] and, with a special focus on meteorology, by Rautenhaus *et al.* [21]. For the sake of brevity, we will focus on those works that are most closely related to our own, that is, those dealing with the analysis of 2D spatio-temporal ensemble data.

When it comes to the visualization of ensemble data, especially in meteorology and climate research, a widely used approach is the extraction of isocontours for each ensemble member, which are then collectively displayed in so-called *spaghetti plots* [21]. A common problem of those plots, however, is the amount of visual clutter induced by the multiplicity of curves present in one diagram. To overcome this issue, Sanyal *et al.* [24] use a glyph-based approach to aggregate the individual contours to ribbons encoding the variation within the ensemble. As such, this approach does not directly address the analysis of time-dependent data. To this end, Ferstl *et al.* [3] extend the concept of spaghetti plots to ensembles of time series. By employing a clustering method, they are able to track members with common contours through time and find points at which they begin to diverge. A different approach for extending isocontours to temporal data was taken by Fofonov *et al.* [4]. They introduce distance measures on the isocontours and use projection methods to visualize their evolution over time. One drawback of techniques based on isocontours, especially in 2D data sets, is the fact that by visualizing only one isolevel at a time, the user might lose the overview over the entire domain and thus might oversee potentially interesting features in the data set. Furthermore, these methods require the choice of good thresholds for the seeding of isocontours.

Obermaier *et al.* [19] compute clusters of ensemble members for each space-time point and compare them across adjacent time steps to identify temporal trends within the ensemble. Kappe *et al.* [14] also use a clustering approach to preprocess time-dependent ensemble data for their interactive visualization. Therein, clusters are shown as nodes of a Sankey diagram connected by edges for the individual members, allowing the user to study the dynamic recombination of clusters over time. In their work, Shu *et al.* [26] use clustering to extract spatial regions of high similarity among ensemble members. The temporal development of those regions is then presented as a time line and differences between members of a cluster can be analyzed in a *comparison view*. Wang *et al.* [32] cluster ensemble members of weather simulations by the accuracy with which they predicted the actual weather conditions. For user-selected clusters the parameter sets used for the simulation of their individual members are presented in a *Nested Parallel Coordinate Plot*, allowing to investigate which parameter combinations led to accurate predictions. Two more clustering approaches are presented by Hao *et al.* [6]. They either cluster the time steps of the individual members to obtain an ensemble of condensed time series or cluster the member items of each time step separately to obtain a time series of dominant patterns. In both cases, *contiguous item sequential pattern mining* is used to extract similar temporal behavior of members and frequent shape changes in the ensemble.

In their framework, *EnsembleVis*, Potter *et al.* [20] offer a variety of methods for the exploratory analysis of ensemble data sets. However, in order to examine time series, the user is presented with a film strip showing the juxtaposition of all time steps, which might be overwhelming, especially for long time series.

The methods mentioned in this section aim at directly comparing members for differences and similarities or identifying groups of members that show similar behavior in spatial regions and/or time

ranges. They are general in that they do not look for a specific type of pattern, but find subsets of the data that are similar. Some of these works do not consider the semantic difference between time steps of a time series and members of an ensemble. These general systems require the user to interactively explore the data, which in turn requires some degree of expertise in the respective area in order to know what to look for and how to extract it with the system. The method presented in this paper aims at a different goal. We analyze spatio-temporal trends of oscillating phenomena in ensemble data sets. In our case, domain experts already know the phenomenon of interest and which features in the data are relevant for its analysis. However, they require means of extracting those features and ways to visualize the temporal development of the phenomenon that allows to communicate those trends to experts and laypersons on the respective field. In this paper, we present a workflow to extract and track oscillating features in uncertain scalar fields over time and present them in a way suitable for the general public.

3 BACKGROUND

In this section, we provide the notation and mathematical background required for the remainder of the paper.

3.1 Deterministic and Uncertain Scalar Fields

We call a function $f : \mathscr{D} \to \mathbb{R}$ a *(deterministic) scalar field* over the d-manifold $\mathscr{D}$. Due to the finite nature of computers, one typically uses a finite representation of such fields by sampling them on the finite set of vertices of a discrete grid $\mathscr{G}$, $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subseteq \mathscr{D}$.

Uncertain scalar fields, on the other hand, can be regarded as multivariate random variables $X : \Omega \to \mathbb{R}^n$, assigning every event $\omega \in \Omega$ a concrete realization $X(\omega) \in \mathbb{R}^n$. Here, every coordinate $(X(\omega))_i$ describes the scalar value at the respective grid point $\mathbf{x}_i \in \mathscr{G}$. That is, the value at each grid point is not a single scalar, as for the deterministic field, but rather it can take on any real value, each one occurring with a certain probability.

In practice, it is rarely possible to describe the uncertain scalar field explicitly, e.g., by specifying its probability density function, as the phenomena described by the field are too complex. To reduce the complexity, one simulates the actual phenomenon using mathematical models. These models rely on certain input parameters that, to some extent, describe the sample space Ω. As the dependency of the model's output from the chosen input parameters is typically still too complex to be described explicitly, one again falls back to sampling the parameter space by choosing multiple parameter sets with which the model is executed. The set of resulting fields is called an *ensemble* and each field a *member* of the ensemble.

3.2 Empirical Orthogonal Functions

The computation of empirical orthogonal functions (EOFs), also known as proper orthogonal decomposition (POD), is a statistical procedure to separate a signal into orthogonal basis functions, characterizing the main directions of variation in that signal. It is conceptually similar to principal component analysis (PCA) and is also referred to as geographically weighted PCA in geophysics [7, 15].

The input to this method is a set of scalar fields $f^{(i)} : \mathscr{D} \to \mathbb{R}$, $i = 1, \ldots, m$ from which it finds a set of functions $g^{(j)} : \mathscr{D} \to \mathbb{R}$ and for each $f^{(i)}$ a set of scalars $\alpha_j^{(i)}$, $j = 1, \ldots, \ell$ such that for each $k \leq \ell$ the reconstructed functions

$$\hat{f}^{(i;k)} := \sum_{j=1}^{k} \alpha_j^{(i)} g^{(j)}, \quad i = 1, \ldots, m \tag{1}$$

minimize the residual error in the least-square sense:

$$\sum_{i=1}^{m} \langle \hat{f}^{(i;k)} - f^{(i)}, \hat{f}^{(i;k)} - f^{(i)} \rangle \to \min. \tag{2}$$

Here, $\langle f, g \rangle := \int_{\mathscr{D}} f(\mathbf{x}) g(\mathbf{x}) \, d\lambda(\mathbf{x})$ denotes the standard scalar product on $L^2(\mathscr{D})$, the space of square-integrable scalar functions on $\mathscr{D}$.

It turns out that the $g^{(j)}$'s are eigenfunctions to the ℓ largest eigenvalues of a certain symmetric linear operator and are pairwise orthogonal, hence they are called *empirical orthogonal functions*. The α's are called *EOF coefficients*. Often, the signal varies around a non-zero mean function $\bar{f}$. In this case, the variation can be better characterized if the EOFs are computed for the functions given by $f^{(i)} - \bar{f}$ instead of the functions $f^{(i)}$ directly.

EOFs can be interpreted as a pattern caused by some latent factor, the eigenvalue corresponding to each eigenfunction as how much of the overall variation is explained by this factor, and the EOF coefficients as how strong each factor is present in function $f^{(i)}$. Usually, the eigenfunctions $g^{(j)}$ are ordered by the magnitude of their corresponding eigenvalue, i.e., the strength of their influence.

If presented with discrete fields sampled on a finite grid, we can obtain the required input functions by means of interpolation. To this end, let $\{p_i\}_{i=1}^{n} \subset L^2(\mathscr{D})$ be a set of linearly independent functions. We can then extend functions defined on n discrete grid points by linearly combining the p_i's to functions on $\mathscr{D}$:

$$f(\mathbf{x}) := \sum_{i=1}^{n} f_i p_i(\mathbf{x}), \quad f_i \in \mathbb{R}.$$

The interpolation allows to avoid the function notation in favor of operations on vectors and matrices: Let $f = \sum_{i=1}^{n} f_i p_i, g = \sum_{i=1}^{n} g_i p_i$ denote two functions and $\mathbf{f} = (f_1, \ldots, f_n)^T, \mathbf{g} = (g_1, \ldots, g_n)^T$ the vectors of their interpolation coefficients, then their inner product becomes two matrix products:

$$\langle f, g \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} f_i g_j \langle p_i, p_j \rangle = \mathbf{f}^T \mathbf{W} \mathbf{g},$$

with $w_{ij} = \langle p_i, p_j \rangle$. It is easy to show that if the p_i's are linearly independent, then $\mathbf{W}$ is symmetric and positive definite, which is sufficient so that $\mathbf{W}^{1/2}$ and $\mathbf{W}^{-1/2}$ are well-defined. $\mathbf{W}$ serves for correcting the influence of the choice of interpolation when moving from the original problem in a function space to a finite-dimensional vector space; in particular, it corrects for variations in the density of grid points. We will discuss the concrete choice of an interpolation basis in Section 4.1.

Now, if $f^{(i)}$ and $g^{(j)}$, and therewith $\hat{f}^{(i;k)}$, are discrete functions that are interpolated in this fashion, we find that Equation (2) is solved if $\mathbf{W}^{1/2} \mathbf{g}^{(j)}$ are eigenvectors of the matrix

$$\frac{1}{m-1} \sum_{i=1}^{m} \mathbf{W}^{1/2} (\mathbf{f}^{(i)} - \bar{\mathbf{f}})(\mathbf{f}^{(i)} - \bar{\mathbf{f}})^T \mathbf{W}^{1/2}, \tag{3}$$

with $\bar{\mathbf{f}}$ denoting the pointwise arithmetic mean of the $\mathbf{f}^{(i)}$'s, and $\alpha_j^{(i)} = (\mathbf{g}^{(j)T} \mathbf{W} \mathbf{f}^{(i)})/(\mathbf{g}^{(j)T} \mathbf{W} \mathbf{g}^{(j)})$. This can be seen by iteratively determining the root of the derivative of Equation (2) for increasing values of k. A detailed derivation is provided in the supplementary material. Note that Equation (3) can be interpreted as the weighted sample covariance matrix of the $\mathbf{f}^{(i)}$'s.

Thus, to perform the EOF analysis, we can proceed as follows. First we subtract the mean from each field and premultiply the result by $\mathbf{W}^{1/2}$. We then compute the eigenvectors $\mathbf{W}^{1/2} \mathbf{g}^{(j)}$ of (3) and finally revert the weighting by premultiplying each eigenvector by $\mathbf{W}^{-1/2}$ to obtain the sought fields $\mathbf{g}^{(j)}$.

3.3 Confident Critical Points

Given a smooth function $f : \mathscr{D} \to \mathbb{R}$ on a differentiable d-manifold $\mathscr{D}$, a point $\mathbf{p} \in \mathscr{D}$ is called *critical* if the gradient of f at $\mathbf{p}$ vanishes: $\nabla f(\mathbf{p}) = 0$; all other points are called *regular*. Isolated critical points can be local extrema or saddle points. This type is determined by the signs of the eigenvalues of the Hessian at $\mathbf{p}$, i.e., the matrix of second-order partial derivatives: $(H(\mathbf{p}))_{i,j} = \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} f(\mathbf{p})$. In 2D, the

type of critical point can be uniquely determined from the trace and the determinant of the Hessian: Saddles yield a negative and local extrema a positive determinant. In the latter case, the trace of the Hessian is negative for local maxima and positive for local minima.

Using the mathematical quantities above, one can detect local extrema in single scalar fields, but we wish to integrate the information about local extrema from multiple scalar fields. For this, we use the method by Mihai and Westermann [17], which determines a confidence value for critical points and their type for sets of scalar fields that are the outcome of a probabilistic process such as simulations with randomly perturbed initial conditions. The core idea of this method is to estimate the distributions for the gradient as well as the trace and determinant of the Hessian from the multivariate distribution over the scalar fields. Given these distributions, confidence values are computed that indicate whether the distribution of gradients is likely to contain the zero vector and the likely signs of the trace and determinant of the Hessian. For instance, the components of the gradients at each point are described using a multivariate probability distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{S}$ and the used test statistic is the Mahalanobis distance between the mean gradient and the hypothesized expected gradient $\mathbf{m_0} = \mathbf{0}$:

$$d_M(\mathbf{m_0}) := \sqrt{(\mathbf{m} - \mathbf{m_0})^T \mathbf{S}^{-1}(\mathbf{m} - \mathbf{m_0})}.$$

The test checks whether the gradient's estimated distribution is sufficiently far from the origin to reject the hypothesis of a point being critical. Different thresholds for d_M can be used to reflect the properties of the stochastic process generating the data. For the sign of the trace and determinant of the Hessian, the procedure is similar: First, confidence intervals for the probability distributions of these two quantities are computed, then it is checked whether each interval contains zero or is entirely positive or negative. In the first case, the data does not give a clear indication of the sign of the respective quantity and, therefore, the type of the critical point. Since the input data is usually not smooth, Mihai and Westermann use finite differences to derive the probability distributions for the gradient and Hessian from the distributions of each grid point's neighbors.

A different approach is being followed by Günther *et al.* [5], who find *mandatory critical points* in ensemble data sets. This approach identifies regions where critical points occur in any realization. However, their method requires the probability distribution (or its estimator) at each point to have finite support, while Mihai and Westermann [17]'s method does not restrict the actual distribution. We therefore preferred to use the latter.

4 METHODS

In this section, we describe our workflow for extracting the trajectories of the centers of action (COAs) from time-dependent ensemble data sets in a general setting; in Section 5 we discuss details and parameters that are specific to the data set we consider. We start by giving an overview of our method (see also Figure 3), followed by a more detailed discussion in Sections 4.1–4.4.

Input. A family of discrete, time-dependent ensemble fields $\mathbf{f}^{(e,t)} = (f_i^{(e,t)})_{i=1}^n \in \mathbb{R}^n, e = 1, \ldots, \eta, t = 1, \ldots, \tau$, with η ensemble members, each having τ time steps at n grid points.

Output. The trajectories visualizing the spatial movement of the COAs over time.

Step 1: Compute leading EOFs on sliding windows. For each ensemble member e individually, we group the time steps into smaller ranges by moving a sliding window of length ω along the time dimension, yielding intervals $\mathscr{T}_i := \{i, \ldots, i + \omega - 1\}$, $i = 1, \ldots, \tau - \omega + 1$. For each such window, we then compute the leading EOF (i.e., the EOF corresponding to the largest eigenvalue), giving us $\tau - \omega + 1$ fields for each member.

Step 2: Compute confident critical points. Using the method by Mihai and Westermann [17] (see Section 3.3), we now determine confident critical points, i.e., points that are likely to be critical. As
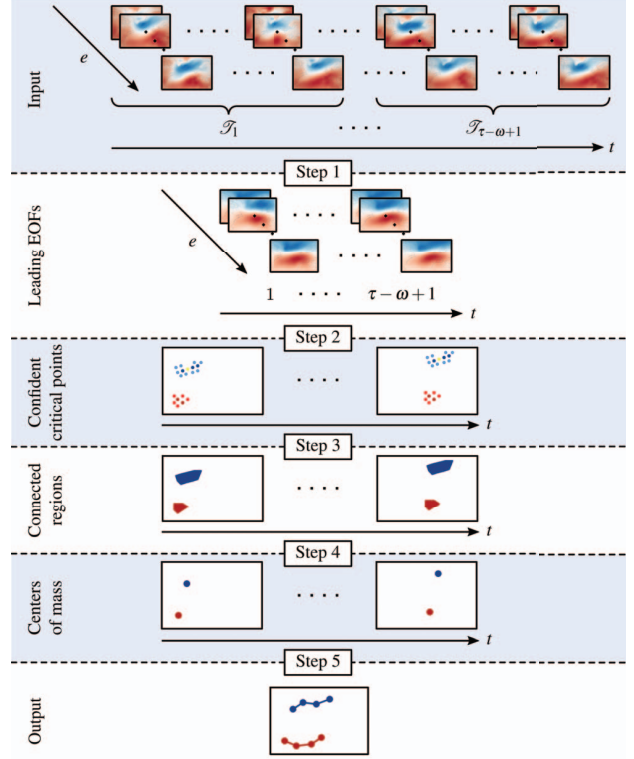

Figure 3: Our workflow for the extraction of the paths on which the centers of action move over time.

the probability of a point being critical gradually decreases with rising Mahalanobis distance, one has to define a cut-off value d_{conf} above which points are considered non-critical.

For the remaining points, falling bellow that threshold, we determine the most probable type. For this purpose, we classify the critical points in seven categories based on the signs of determinant and trace of the Hessian: confident maxima, confident minima, confident saddles, maximum or saddle, minimum or saddle, maximum or minimum, and ambiguous (cf. [17, Figure 3]).

Performing this step on all leading EOFs for each time interval $\mathscr{T}_i$, thereby accumulating over the ensemble dimension, we obtain those seven sets of points for each time interval.

Step 3: Extract connected components. The previous step will not give us isolated critical points, but rather connected regions where multiple grid points could be critical. Hence, the next step is to extract such regions. Depending on the data set used and, therewith, the output of the previous step, different definitions for connectedness (e.g., 4-point or 8-point, single or multi-hop) might be considered.

Step 4: Compute centers of mass. In order to reduce the connected components to single points representing the COAs, we compute the center of mass (COM) from each connected component by calculating the mean of all contained points weighted by a measure of their respective confidence, e.g., $d_{\mathrm{conf}} - d_M(\mathbf{0})$.

From the obtained points, we select those that are most relevant for the analysis task at hand, e.g., those where the accumulated confidence of the corresponding connected components is the largest. This leaves us with a collection of points for each time interval $\mathscr{T}_i$.

Step 5: Connect COMs to paths. Finally, we connect corresponding COMs from adjacent time steps with line segments to obtain the path they took over time.

4.1 Interpolation basis

In this work, we consider only uniform rectilinear grids, as they allow for a simpler definition of finite difference formulas to approximate first- and second-order derivatives. However, our method can easily be adapted to work with grids on which a differentiation schema for first- and second-order derivatives can be defined.

To account for differences in the sampling density of the grid on which the input fields are given, an appropriate interpolation schema for the reconstruction of continuous functions from those discrete fields is required. In geoscience, for instance, data is typically given on a two-dimensional planar grid describing locations on a 2-sphere (the Earth). However, it is topologically impossible to define such a grid in a way that preserves the actual distances on the globe. All those representations, therefore, contain regions that are more densely sampled by grid points than others. Without accounting for such irregularities, more densely sampled regions enter the EOF analysis over-proportionally, distorting its outcomes.

We can avoid these effects by regarding not only the field's value at each grid point, but also the area that we believe is affected by this value. By doing so, the values in regions containing more grid points (implying that each grid point describes a smaller subregion of the domain $\mathscr{D}$) enter the EOF analysis with a smaller weight than those in less densely sampled regions.

The most straightforward way to take the area of influence into account is to use a piecewise constant interpolation basis. That is, to use $p_i = \mathbb{1}_{U_i}$, where U_i are pairwise disjoint neighborhoods of $\mathbf{x}_i \in \mathscr{G}$ covering the entire domain $\mathscr{D}$ and $\mathbb{1}_U$ is the characteristic function on U, i.e., $\mathbb{1}_U(\mathbf{x}) = 1$ if $\mathbf{x} \in U$ and $\mathbb{1}_U(\mathbf{x}) = 0$ otherwise. Thus, if $\mathbf{f} = (f_1, \ldots, f_n)^T \in \mathbb{R}^n$ are function values at grid points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathscr{G}$, we obtain the piecewise constant function $f := \sum_{i=1}^n f_i p_i$ with function value f_i on each neighborhood U_i of $\mathbf{x}_i$, respectively. In this case, the entries of the weighting matrix $\mathbf{W}$ become

$$w_{ij} = \int_{\mathscr{D}} \mathbb{1}_{U_i}(\mathbf{x}) \mathbb{1}_{U_j}(\mathbf{x}) \, d\lambda(\mathbf{x}) = \delta_{i,j} \operatorname{Vol}(U_i), \qquad (4)$$

where $\delta_{i,j}$ denotes the Kronecker delta. That is, $\mathbf{W}$ is a diagonal matrix containing the surface areas of the regions U_i on its diagonal. Note, that this special form also allows for an easier computation of $\mathbf{W}^{1/2} = \operatorname{diag}(\sqrt{w_{11}}, \ldots, \sqrt{w_{nn}})$.

Depending on the grid, other interpolation schemata can also be considered, e.g., bilinear interpolation on a rectilinear grid. Furthermore, the way the data values have been acquired should be taken into consideration.

4.2 EOFs

The minimization problem given in Section 3.2 does not yield an unique solution, as the cost function in (2) is invariant with respect to the norm and sign of the functions $g^{(j)}$ and the scalar values $\alpha_j^{(i)}$. More precisely, if $g^{(j)}$ and $\alpha_j^{(i)}, j = 1, \ldots, \ell, i = 1, \ldots, m$ minimize (2), then so do $\beta_j g^{(j)}$ and $\alpha_j^{(i)}/\beta_j, j = 1, \ldots, \ell, i = 1, \ldots, m$ for any sequence of non-zero scalars β_j, as those factors cancel out in the computation of $\hat{f}^{(i;k)}$ in Equation (1). This is also reflected in the fact that $\mathbf{W}^{1/2} \mathbf{g}^{(j)}$ turn out to be the eigenvectors of Equation (3), as the eigenvectors of a matrix are also only determined up to a multiplicative non-zero factor.

In practice, this means that after we determine the eigenvectors of (3), we are still left with the choice of how to scale the eigenvectors (i.e., the EOFs) and therewith the EOF coefficients. We can use this freedom to select a scale that allows for an easier interpretation of the results. If the input fields have already been cleansed of their mean, i.e., by using $\mathbf{f}^{(i)} - \bar{\mathbf{f}}$ instead of $\mathbf{f}^{(i)}$, we have $\sum_{i=1}^m \alpha_j^{(i)} = 0$. Then, by normalizing the EOF coefficients such that $\sum_{i=1}^m (\alpha_j^{(i)})^2 = m - 1$, the coefficients resemble a set of independent random variables with mean 0 and variance 1. This allows us to interpret their magnitudes as how extreme the individual input fields deviate from the average along the corresponding EOF. Furthermore, since the matrix in Equation (3) is the weighted sample covariance matrix of the $\mathbf{f}^{(i)}$'s, the unweighted EOFs $\mathbf{g}^{(j)}$ allow for a direct interpretation of the variance contained in the respective mode with the same unit of measurement given by the input fields $\mathbf{f}^{(i)}$.

Finally, we have to choose the signs of the eigenvectors in such a way that the eigenvectors of all members point in a similar direction. Otherwise, we would not be able to extract the concise type of the critical points, as maxima and minima would get mixed-up. We choose the sign of the eigenvectors so that their scalar product with the mean field $\bar{\mathbf{f}}$ adjusted for its own (spatial) mean is positive. This has the advantage of the eigenvectors resembling the mean field, i.e., regions with values above average tend to coincide with regions of high values. It further allows the signs of the α's to be used for interpretation: positive values indicate that prevalent conditions tend to be emphasized and negative values that prevalent conditions are de-emphasized or reversed.

4.3 Time Windows

The length ω of the time windows will affect the degree of temporal smoothing. While small intervals allow covering variations on a smaller timescale, they also contain more noise, therewith making the analysis of long-term trends increasingly difficult. Long time windows, on the other hand, smooth out noise in the data, but also disguise small-scale features.

4.4 Cut-off value d_{conf}

The parameter d_{conf} imposes an upper bound on the Mahalanobis distance between the zero vector and the mean gradient at each grid point. Points of constant Mahalanobis distance d describe an ellipse whose axes are aligned with the principal component axes of the underlying distribution of gradients. The lengths of its semiaxes correspond to d times the variance in that direction. Therefore, the Mahalanobis distance can be seen as a measure for how many standard deviations a point lies away from the mean value.

To obtain an intuition of the meaning of a particular choice of d_{conf}, it is helpful to study the special case where the gradients are multivariate normally distributed. In this case, we can test whether a point is not critical by using Hotelling's T^2 test. Given the sample mean $\mathbf{m}$ and sample covariance matrix $\mathbf{S}$, it checks the null hypothesis "$\mathbf{m} = \mathbf{m_0}$" for some hypothetical mean vector $\mathbf{m_0}$ (here $\mathbf{m_0} = \mathbf{0}$) against the alternative hypothesis "$\mathbf{m} \neq \mathbf{m_0}$" [12]. We chose this test over the hypotheses test based on the χ^2 distribution, indirectly proposed in [17], as the former compensates for small sample sizes.

Hotelling's T^2 test is based on the fact that

$$X := \frac{m(m-p)}{p(m-1)} d_M^2(\mathbf{m_0}) \sim F_{p,m-p},$$

where m is the number of samples, p is the number of dimensions, and F_{ν_1, ν_2} is the F distribution with parameters ν_1 and ν_2. For concrete values of p and m and a significance level α, the critical value $F_{p,m-p,\alpha}$ where $P(X \geq F_{p,m-p,\alpha}) = \alpha$ can be determined. Solving $X \geq F_{p,m-p,\alpha}$ for $d_M(\mathbf{m_0})$ then yields the critical value for the Mahalanobis distance above which the null hypothesis is rejected at the significance level α, that is, the test indicates that the grid point is not critical. Table 1 gives an overview of resulting Mahalanobis distances for $m = 100$, $p = 2$, and often used confidence levels α. For instance, to reject the null hypothesis at the $\alpha = 1\%$ level,

α	5%	1%	0.5%	0.1%	0.01%
$d_M \geq$	0.250	0.312	0.336	0.387	0.452

Table 1: Minimum Mahalanobis distances $d_M(\mathbf{m_0})$ required to be able to claim that a point is not critical at the significance level α based on Hotelling's T^2 test on a two-dimensional sample of size 100.
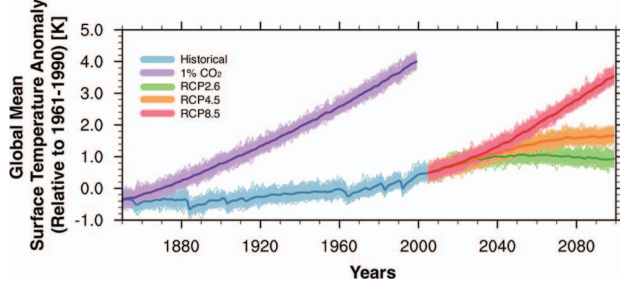
Figure 4: Global mean surface temperature anomalies (relative to 1961–1990) for the historical (blue), 1% CO_2 (purple), RCP2.6 (green), RCP4.5 (orange), and RCP8.5 (red) simulations as depicted in [16, Figure 1 (a), courtesy of Yohei Takano]. The ensemble members (pale lines) are shown together with the ensemble mean (dark line).

$d_M(\mathbf{m_0}) \geq 0.312$ must hold. That is, points where the Mahalanobis distance of the origin to the mean gradient is larger than 0.312 can be considered to be not critical on the 1% significance level. On the other hand, this, unfortunately, does not imply that points having distance smaller than 0.312 need necessarily be critical points. There is just not enough evidence to warrant deeming them not critical. This problem is a shortcoming of the approach described in [17] and approaches based on statistical testing in general.

5 RESULTS

In this section, we discuss how we applied our method to a data set from climate research with the goal to study the temporal variation of the centers of action defining the NAO index.

5.1 Data

For our analysis, we used data from the Max Planck Institute for Meteorology's (MPI-M) Grand Ensemble (MPI-GE) [16]. This data set consists of 100 ensemble members for the simulated historical past as well as for different future scenario simulations. The historical simulation covers the time period from 1850 to 2005, followed seamlessly by the scenario simulations for the years 2006 to 2099. Each year comprises twelve monthly mean fields for several physical quantities, including the sea level pressure, which we used for our analysis of the NAO. Each field covers almost the entire Earth (180°W–180°E, 89.5°N–89.5°S) on a 192×96 grid (resolution of approx. $1.875° \times 1.865°$).

The three future scenarios RCP2.6, RCP4.5, and RCP8.5 describe different developments of the atmospheric CO_2 concentration (resulting from different possible socio-economic developments), which, due to the greenhouse effect, impacts the global energy balance. The scenarios are named according to the respective additional radiative forcing, i.e., the additional energy that influences the climate in comparison to the preindustrial state, reached in the year 2100. These so-called representative concentration pathways (RCPs) [29] correspond to $2.6\,\mathrm{W/m^2}$, $4.5\,\mathrm{W/m^2}$, and $8.5\,\mathrm{W/m^2}$, respectively.

Furthermore, the data set includes an idealized, 150 years long global warming experiment, where the atmospheric CO_2 concentration increases by 1% per year (getting to double the preindustrial value after 70 years), and a 2000 years long preindustrial (PI) control simulation describing the Earth's climate without changes in the CO_2 concentration. Showing the simulated change in the global mean temperature for all experiments (except for PI-control), Figure 4 (taken from [16, Figure 1 (a)]) gives a nice overview of the experiments.

5.2 Preprocessing

As our analysis is focused on the wintertime NAO, we precomputed the mean sea level pressure (SLP) values for the extended wintertime using *Climate Data Operators* (CDO) [25] for all consecutive DJFM periods. Following the approach in [18], we choose the axis-aligned box 90°W–40°E, 20°–80°N as input to our method. We observed that the approximation of derivatives by finite differences sometimes caused small features to disappear for short periods while they moved between adjacent grid points. To avoid missing important features (especially in the less densely sampled southern regions), we therefore supersampled the DJFM mean fields on a $1° \times 1°$ grid.

For control purposes, we compared the systematic movement of the COAs caused by the external forcing in the simulated climate change scenarios with the statistical fluctuations that occur in the absence of those forcings. To this end, we utilized the PI-control run provided with the ensemble. But as it only consists of a single 2000 years long simulation, in contrast to the 100 members given for each scenario run, we could not apply our method directly, as it expects time-dependent ensemble data as input. Hence, we had to split the single run into multiple equally long members. The individual members of the historical ensemble, which later on branch into the different scenario runs, are themselves branched at different time steps from the PI-control run (cf. [16, Table 1]). To obtain an ensemble of PI-control runs that is best comparable to the scenario runs, we extracted 100 time series, each covering 250 consecutive years of the PI-control simulation starting at the same branch time as the respective member of the historical simulation. Each such time series was then regarded as a member of a PI-control ensemble.

5.3 Choice of Parameters

Some parameter choices depend on the data set. Here, we describe the choices made for our analysis of the NAO.

Interpolation basis. Since the data set consists of piecewise constant fields, we also used piecewise constant functions for the interpolation between grid points, as described in Section 4.1.

Furthermore, the data is given on an equirectangular grid. That is, each grid point is described by its longitude φ (angle towards some prime meridian) and latitude θ (angle towards the equator). In this case, we can choose the spherical segment with edge lengths corresponding to the grid resolution centered at the $(k, \ell)^{\mathrm{th}}$ grid point as the region of influence of each grid point, i.e.,

$$U_{(k,\ell)} := [\varphi_k^{(0)}, \varphi_k^{(1)}] \times [\theta_\ell^{(0)}, \theta_\ell^{(1)}].$$

The entries of the diagonal weighting matrix $\mathbf{W}$ therefore are (cf. (4))

$$w_{ii} = \mathrm{Vol}(U_{(k_i,\ell_i)}) = (\varphi_{k_i}^{(1)} - \varphi_{k_i}^{(0)})(\sin(\theta_{\ell_i}^{(1)}) - \sin(\theta_{\ell_i}^{(0)})).$$

Time windows. In climate research, one is interested in long-term changes in Earth's climate. We, therefore, performed our analysis on rather large windows of 30 and 50 years.

Cut-off value d_{conf}. We used a cut-off value of $d_{\mathrm{conf}} = 0.452$ (corresponding to an α-value of 0.01% in the Gaussian distributed case, cf. Table 1), i.e., it is very unlikely that a point is critical if the origin falls outside 0.452 standard deviations of the mean gradient.

The Mahalanobis distance field for a single time window, mapping each grid point to the respective distance of the mean gradient at that position to the zero vector, is depicted in Figure 5. The larger that distance at a certain grid point is (the darker the shade of green), the more unlikely it is for the mean gradient to vanish, and the more convinced we can be that the point is not critical.

Connectedness. To choose an appropriate definition of connectedness, we plotted the critical points in different colors representing their respective type. A representative configuration is shown in Figure 5. There, one can see that a cluster of points corresponding to the Azores High ① consists of points being regarded as confident maxima (dark red) as well as points that might also be saddles (light red). As in the latter case, the probability of the point actually being a saddle is lower than the probability of it being a maximum [17], we decided to extract the connected component of both these classes.
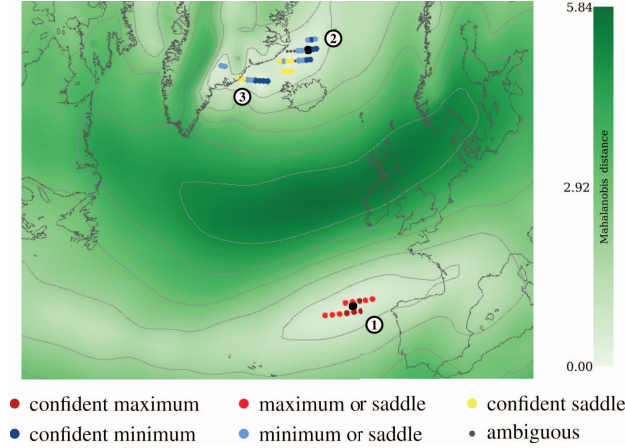
confident maximum • maximum or saddle • confident saddle
confident minimum • minimum or saddle • ambiguous

Figure 5: Critical points for scenario RCP4.5, time window 2030/31–2059/60, and cut-off value $d_{conf} = 0.452$. The color map in the background encodes the Mahalanobis distance $d_M(\mathbf{0})$. Isocontours are drawn in intervals of 1 starting at 0. Potential critical points are colored according to their type as specified by the key at the bottom. Larger black points represent the extracted centers of mass.

The Icelandic Low, on the other hand, is made up of two collections of minima (blue) connected by confident saddle points (yellow). The transition from confident minima (dark blue) to confident saddles is made through points that could be either of both (light blue). Since the group of westward minima ③ decomposes and vanishes over time, we chose to track the more stable and, furthermore, more confident eastward group of minima ②. Again, we took into account the confident minima as well as the points that could also be saddles, i.e., both dark and light blue points, for the extraction of the connected component.

In both cases, the Azores High and the Icelandic Low, we used an one-hop, 8-point neighborhood. That is, two points belong to the same connected component if their respective row and column index in the grid differs by no more than one.

Center of mass. As in our use case, the points are given as longitude-latitude pairs, we computed their COM by averaging their 3D-representation on the sphere and then projected the resulting point back to its longitude-latitude location.

Since we are only interested in the strongest maximum and minimum, we select those two COMs where the accumulated confidence of the corresponding connected components is the largest.

5.4 Visual Presentation

We display the superposition of the extracted paths for the different scenarios to allow for a direct comparison (cf. Figure 7). The paths were colored according to the color schema used in [16] (cf. Figure 4): Historical → blue, RCP2.6 → green, RCP4.5 → orange, RCP8.5 → red, and 1% CO_2 → purple. The paths of the PI-control run (not displayed in [16]) were colored in gray. To improve the perception of the temporal development, we encoded the time dimension by gradually lightening the colors for earlier time steps. Occlusion is reduced by drawing the individual segments of all paths sorted by time, such that line segments from previous time steps will not overlap the ones from newer time steps.

To indicate to the viewer the uncertainty of the underlying data and to some extent that of the displayed paths, we further augmented the individual points of each path with the isocontour at d_{conf} in the field of Mahalanobis distances $d_M(\mathbf{0})$, therewith outlining the region that was then aggregated into that point. When interpreting those regions, one has to bear in mind that one cannot argue for the points being contained to most likely be critical, but rather can only say
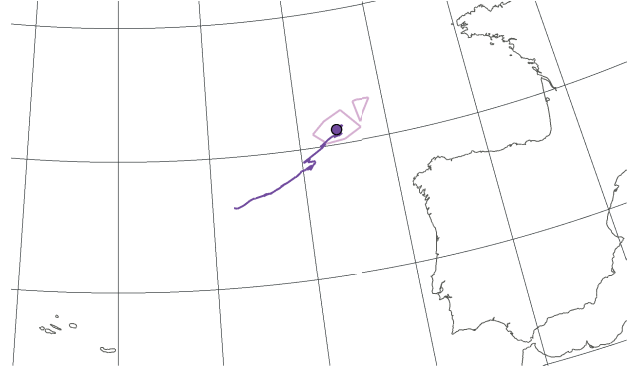

Figure 6: Trajectory of the southern center of action for the 1% CO_2 simulation (dark purple). The location derived for the time window 1949/50–1998/99 is marked by a dot in the same color. The corresponding isocontour $d_{conf} = 0.452$ of the Mahalanobis distance field is shown in a lighter shade of purple.

that points outside the region are with high certainty non-critical. Since displaying the region for each point of every path would result in a large amount of clutter, we generated animations of the paths growing over time where only the uncertainty regions for the current time step are drawn. Again, we choose the same color schema of lighter colors for the uncertainty visualization as [16] (cf. Figure 4). As an example, Figure 6 shows the snapshot for the time window 1949/50–1998/99 and a single path (1% CO_2) up to that time step. The last point of the path (current time step) is marked by a dot and the described isocontour for that time window is displayed in a lighter color shade. We furthermore included the temperature curves depicted in Figure 4 into our animations with the time window corresponding to the current frame highlighted. This allows for a direct assessment of the correlation between rising surface temperatures and the shift of NAO's centers of action. The animations for both 30- and 50-year windows are provided as supplementary material.

As the region under investigation reaches far north, the choice of the map projection becomes important. An equirectangular projection, for instance, would, depending on the latitude, distort the map in x-direction (longitudes), which in turn makes the interpretation of the lengths of the displayed paths a difficult and error-prone task. This issue can be circumvented by the use of an equal-area projection. We chose the Albers equal-area conic projection [27], as it allows us to specify two standard parallels of true scale, that is, two latitudes along which distances are preserved. Since we are interested mainly in the two regions around Iceland (approx. 65°N) and the Azores (approx. 40°N), we selected those two latitudes as the standard parallels for the projection.

To give a geographical context, continent outlines, as well as graticules at 5° intervals, are shown in the map background.

5.5 Analysis and Interpretation

The paths extracted for the different scenarios using 30- and 50-year sliding windows are displayed in Figure 1 and Figure 7, respectively. All in all, we make out a robust, CO_2-forcing-dependent shift in the dominant position of the Icelandic Low and the Azores High, i.e., the centers of action of the NAO. As one might expect, the displacements are more pronounced for scenarios with higher radiative forcings (e.g., RCP8.5 and 1% CO_2) than for those with more moderate or none such forcings (e.g., RCP2.6 and PI-control). The paths of the Azores High are longer and less noisy than those of the Icelandic Low for both window lengths. One possible explanation for this might be that the Azores High is more affected by the poleward shift of the storm tracks, which was discovered to be a robust feature in global warming simulations (e.g., [34]). However, as expected,
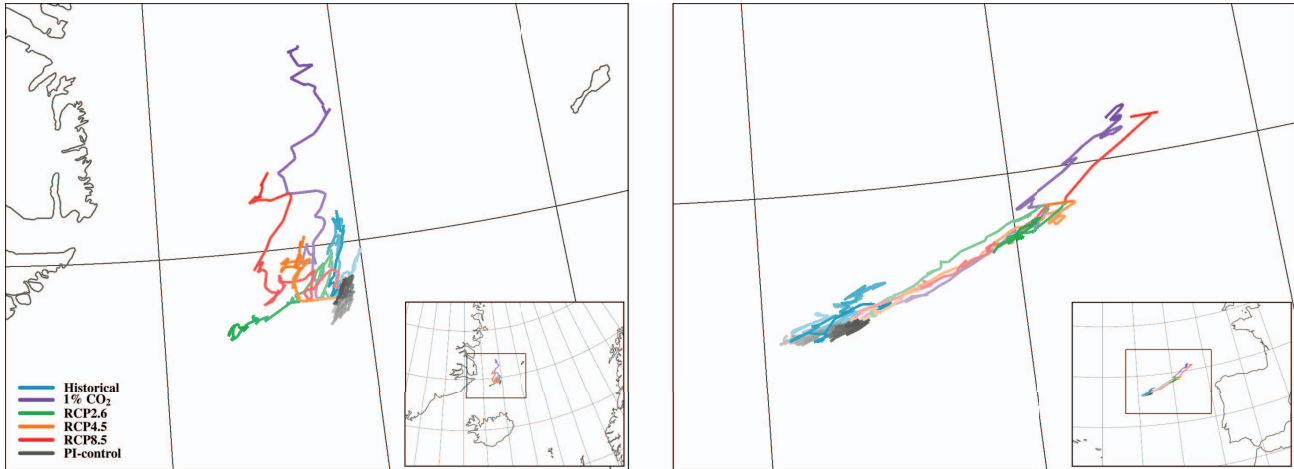
Figure 7: Same as Figure 1, but using 50-year time windows for the path extraction.

the paths derived using 30-year windows show more variability, as fewer time steps are accumulated into the single points of the paths and the overlap of consecutive windows is considerably smaller. As the results produced with 50-year windows show a clearer trend, we focus on them in the following detailed analysis of the Azores High and Icelandic Low:

Azores High. In the right part of Figure 7, we can see that all RCP scenario runs as well as the 1% CO_2 experiment exhibit a systematic northeastward movement within a narrow corridor and merely differ in the amount of displacement along that direction. We measured distances of approx. 240 km, 320 km, 500 km, and 550 km (measuring the great-circle distance between the first and the last point of the path) for the RCP2.6, RCP4.5, RCP8.5, and 1% CO_2 simulation, respectively. Note how the positions derived for the RCP8.5 and the 1% CO_2 scenario are shifted by almost the same amount, which corresponds to the temperature changes reached at the end of the two simulations, as shown in Figure 4.

By looking at the animations (provided as supplementary material), we can further analyze the velocity of the movement: The path of RCP4.5 starts slowing down around the time window 2004/05–2053/54, which corresponds nicely to the time when the temperature curve also begins to increase at a slower rate. For RCP2.6, we can even observe a backward trend along the path it came from, just as the temperature also begins to drop slightly. RCP8.5, being the only scenario where the temperature keeps rising at least linearly, shows a high velocity towards the northeast almost during the entire simulation. There, slower speed can only be found at the windows 2007/08–2056/57 to 2029/30–2078/79.

In the historical simulation, hardly any systematic movement can be made out, which is most likely due to the fact that the radiative forcing is comparatively low in this period. This can also be seen in the rather small temperature increase depicted in the blue curve of Figure 4.

For control purposes, we also applied our method on a pseudo-ensemble created from the 2000 year long PI-control run (cf. Section 5.2). This simulation does not contain any changes in the forcing, and the resulting gray path on the right of Figure 7, therefore, indicates the internal variability of the position of the Azores High. The detected positions stay within a narrow corridor of about 140 km length in the east-northeast direction. One clearly sees that the fluctuation is most dominant in the direction of the other paths, suggesting that the observed shift is in parts a consequence of statistical fluctuations. Nevertheless, the maximal

extent of the fluctuations in the PI-control run is rather small compared to the expansion of the paths from the other scenarios.

Furthermore, in the last frames of the animation, we can see that at least the confidence regions for the RCP4.5 and RCP8.5 scenarios are disjoint of the respective regions for the PI-control simulation. This implies that at the end of the simulation, the probability of the COAs in those scenario runs being located in proximity to the PI-control run is very small. Since the PI-control simulation reflects the development of the Earth's climate in a non-warming environment, we conclude that global warming causes a shift of the southern COA.

Icelandic Low. The paths extracted for the Icelandic Low, depicted on the left of Figure 7, are noisier than those of the Azores High. Most paths show a slight westward shift, and, depending on the warming reached, an additional shift in the north-south direction. While the detected position for the RCP2.6 scenario moves south-westwards over time, those of the two remaining RCP scenarios drift northward, although the shift found for the RCP4.5 scenario is significantly smaller than that found for the RCP8.5 scenario. In contrast to the movement found for the Azores High, the 1% CO_2 scenario features a more pronounced northward shift than that found for the RCP8.5 scenario. This could be caused by differences in the spatial warming patterns of the two experiments.

For the Icelandic Low, we measured distances of approx. 100 km, 80 km, 150 km, and 270 km (measuring the great-circle distance between the first and the last point of the path) for RCP2.6, RCP4.5, RCP8.5, and 1% CO_2, respectively.

The historical simulation mainly exhibits a significant (compared to the PI-control run) north-south fluctuation, but without any trend in either direction. Again, this lack of a systematic movement, especially in comparison with the scenarios RCP8.5 and 1% CO_2, can be explained by the lower forcing in the historical ensemble.

Although the movement for the Icelandic Low is much smaller than that of the Azores High, comparison with the PI-control run (fluctuation of about 50 km) shows that a major portion of the movement is a systematic movement rather than statistical fluctuations. The final frames of the animation furthermore show that at least the confidence regions of the RCP8.5 scenario are disjoint from those of the PI-control run. That is, the COAs at the end of the RCP8.5 scenario are very unlikely to appear close to the COAs in the PI-control simulation. Thus, we conclude that the position of the northern COA is also affected by global warming.

In summary, we discovered similar movements as [28] and [8]. However, especially the paths extracted for the northern center of action did not show as much movement as discovered in those previous studies. We hypothesize that this is mainly caused by the fact that there only a single simulation run is used, whereas, in this work, we utilized the bundled information of 100 ensemble members yielding a statistically more robust result. In [28, Figure 6], the COAs extracted for the time windows covering the period 2020–2099 are not labeled with the respective window, thus they do not allow to reconstruct a temporal order, while the labeled points in [8, Figure 7 (a)] show large fluctuations in the direction of the movement, suggesting that the larger displacement might be the consequence of statistical variability in the used simulation model. Similarly, the movement of the Azores High—which was not detected in [28] at all—is also more pronounced in [8]. But there again, the points show large fluctuation in the direction of the movement (cf. [8, Fiugre 7 (b)]).

6 DISCUSSION

Due to the natural variability of the climate system, climate data can contain quasi-periodic signals on multiple temporal and spatial scales. A major task in climate research is the analysis of the dominant variability patterns. We found that the straightforward application of existing visualization techniques used for feature extraction are not well suited for such an analysis, as their results are dominated by the particularly noisy character of the data. For example, applying Mihai and Westermann's method [17] directly to the ensemble data resulted in very large regions for the potential positions of the Azores High's and the Icelandic Low's centers of action (cf. supplementary material), from which no concise information on the temporal variability or shifts in the position due to climate change can be seen. Similarly, if the spaghetti plot technique is applied directly to the data (cf. supplementary material), the individual contours are spread rather widely, making the identification of a significant center and its consistent motion over time a difficult task.

To detect, visualize, and analyze relevant variations, the data, therefore, may need to be filtered first. A proven means for this task is the computation of empirical orthogonal functions. They provide a decomposition of the signal into the main directions of variation, therewith cleansing the data of less-relevant fluctuations. If we, e.g., apply spaghetti plots to EOF fields of the data (cf. supplementary material), we obtain a more concise picture where isocontours are bundled more densely. Still, it is hardly possible to extract information about the temporal evolution of the COAs. Therefore, a more sophisticated algorithm for the extraction and tracking of topological features in uncertain time-dependent data sets is required. To the best of our knowledge, there is no preexisting method specifically designed for this purpose.

Therefore, we developed a workflow that combined the filtering abilities of EOFs with the method described by Mihai and Westermann [17] for the extraction of critical points from uncertain fields and a tracking mechanism to track the identified points over time. We identified some shortcomings of Mihai and Westermann's method in the process. First, the implicitly performed hypothesis test is designed to assess whether a point is regular but does not allow any conclusions about the significance of points discovered to be not regular. This makes the interpretation of the results challenging and renders the use of the proposed qualitative indicators (cf. [17, Equation (13)])—evaluating the test statistic precisely in the non-regular case—questionable from a statistical point of view. Second, it performs tests independently for each grid point, although information from neighboring grid points and neighboring time steps could be used to give a clearer picture. Nevertheless, this method allowed us to extract regions where we may expect critical points to occur, which we were then able to track over time. However, our method of tracking is preliminary and currently only works if the extracted

regions (cf. Section 4, Step 3) remain disjoint and do not split or merge with other regions over time. Especially, the evolution of the Icelandic Low prompted the domain experts to ask whether the two pressure systems only move or whether they could also split due to climate change and how likely this event would be. Currently, there does not seem to be a technique for answering such a question for time-dependent uncertain data.

Despite our focus on the NAO phenomenon, domain experts deemed our method generally useful to investigate changes in oscillating patterns characterizing the Earth's climate and indicated their interest to use it to investigate further climatological phenomena such as El Niño-Southern Oscillation (ENSO), the Indian Ocean Dipole (IOC), the Southern Annular Mode (SAM), and the East Atlantic (EA) Pattern, the latter one being described by the second strongest mode of the SLP over the region analyzed in this paper. The investigation of those phenomena, as well as their temporal evolution, is of great relevance, as they play an important role within the climate system, and shifts in their frequency, strength and position can lead to severe changes in prevailing weather conditions. The shift of the NAO pattern found in this paper, for instance, could cause the flow of maritime air towards Europe to make landfall in more northern territories and could therewith lead to severe changes in European winter weather. Preliminary tests suggested that our method can also be applied to usefully analyze these phenomena.

Overall, we learned the following lessons during the work on this paper and our collaboration with the domain experts:

- Time dependent ensemble data sets often feature a lot of noise, making it hard to visually spot temporal trends from the data directly. Some form of filtering, e.g, EOFs, to separate the relevant signals from that noise is therefore indispensable.
- We found that existing methods for critical point extraction from ensemble data have their shortcomings and see potential for further development in this area.
- A challenge was to translate the resulting pictures into statements that are meaningful for the domain experts. This involved both the interpretation of EOFs metaphorically as the effects of latent causal factors as well as the interpretation of the confident regions ("The outside can be characterized as being free of critical points for the given level of confidence.").
- Parameter choice in applying techniques should serve easing interpretation of results. The EOF method, for instance, admits multiple solutions of equal quality and we used the opportunity to pick solutions to ease interpretation by domain experts and laypersons.
- To create confidence in the results, visualization systems should combine few techniques, preferably techniques whose intermediate results can be shown if the domain expert requests it and which do not interact in such a way that a visual feature cannot be retraced to the technique that gave rise to it.
- Accuracy in interpretation plays a higher role for experts, who are more interested in the evolution of the confidence regions of local extrema, than laypeople, for whom the qualitative message that the tracks transport is sufficient.

7 CONCLUSION AND FUTURE WORK

In this paper, we presented a method for the analysis of spatio-temporal trends of oscillating phenomena in time-dependent ensemble data sets. We used a combination of a statistical method called empirical orthogonal functions (EOF), a method described in [17], and a sliding window approach to aggregate the vast amount of data that comes with the analysis of spatio-temporal ensemble data sets to paths showing the evolution of dominant features over time.

We used that workflow to analyze the effect of global warming on the wintertime NAO in a large ensemble of climate simulations. A similar analysis on a single simulation run has already been conducted in, e.g., [28] and [8], where systematic northeast shifts of the

NAO's centers of action were observed. We were able to corroborate their results within the ensemble data set, albeit the displacements were less pronounced in our case. This could be a consequence of previous work using only single simulation runs, resulting in higher risk of a chance result.

We worked in close collaboration with climate experts, who gave us consistently positive feedback. However, our method is not restricted to applications from climate research, but may also benefit other fields where time-dependent ensemble simulations are conducted, e.g., particle physics and fluid mechanics.

There are still a few open points that we wish to address in future work. For one, a more sophisticated algorithm for tracking the centers of action would allow us to track more complex features through time. For instance, one could apply a time-hierarchical clustering similar to the one performed in [3], to identify corresponding connected components in neighboring time steps and track their evolution over time. Furthermore, we intend to research possibilities for an enhanced visual representation of the uncertainty of the paths' locations. The current approach using hypothesis testing makes that difficult, as it does not offer us a quality measure in cases where the null hypothesis cannot be rejected (i.e., at points that cannot be declared to be not critical).

REFERENCES

[1] A. G. Barnston and R. E. Livezey. Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Monthly Weather Review*, 115(6):1083–1126, 1987.

[2] C. Deser, J. W. Hurrell, and A. S. Phillips. The role of the north atlantic oscillation in european climate projections. *Climate Dynamics*, 49(9):3141–3157, 2017.

[3] F. Ferstl, M. Kanzler, M. Rautenhaus, and R. Westermann. Time-hierarchical clustering and visualization of weather forecast ensembles. *IEEE TVCG*, 23(1):831–840, 2017.

[4] A. Fofonov, V. Molchanov, and L. Linsen. Visual analysis of multi-run spatio-temporal simulations using isocontour similarity for projected views. *IEEE TVCG*, 22(8):2037–2050, 2016.

[5] D. Günther, J. Salmon, and J. Tierny. Mandatory critical points of 2d uncertain scalar fields. *Computer Graphics Forum*, 33(3):31–40, 2014.

[6] L. Hao, C. G. Healey, and S. A. Bass. Effective visualization of temporal ensembles. *IEEE TVCG*, 22(1):787–796, 2016.

[7] P. Harris, C. Brunsdon, and M. Charlton. Geographically weighted principal components analysis. *Int. J. Geogr. Inf. Sci.*, 25(10):1717–1736, 2011.

[8] Z.-Z. Hu and Z. Wu. The intensification and shift of the annual north atlantic oscillation in a global warming scenario simulation. *Tellus A*, 56(2):112–124, 2004.

[9] J. W. Hurrell. Decadal trends in the north atlantic oscillation: regional temperatures and precipitation. *Science*, 269(5224):676–679, 1995.

[10] J. W. Hurrell, Y. Kushnir, G. Ottersen, and M. Visbeck. *An Overview of the North Atlantic Oscillation*, pages 1–35. American Geophysical Union (AGU), 2013.

[11] L. Jianping and J. X. Wang. A new north atlantic oscillation index and its variability. *Advances in Atmospheric Sciences*, 20(5):661–676, 2003.

[12] R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistical Analysis*. Applied Multivariate Statistical Analysis. Pearson Prentice Hall, Upper Saddle River, NJ, 2007.

[13] T. Jung, M. Hilmer, E. Ruprecht, S. Kleppek, S. K. Gulev, and O. Zolina. Characteristics of the Recent Eastward Shift of Interannual NAO Variability. *Journal of Climate*, 16(20):3371–3382, 2003.

[14] C. P. Kappe, M. Böttinger, and H. Leitte. Exploring variability within ensembles of decadal climate predictions. *IEEE TVCG*, 25(3):1499–1512, 2019.

[15] E. N. Lorenz. Empirical orthogonal functions and statistical weather prediction. Scientific Report 1, Statistical Forecasting Project, MIT, Department of Meteorology, Cambridge, 1956.

[16] N. Maher, S. Milinski, L. Suarez-Gutierrez, M. Botzet, M. Dobrynin, L. Kornblueh, J. Kröger, Y. Takano, R. Ghosh, C. Hedemann, C. Li, H. Li, E. Manzini, D. Notz, D. Putrasahan, L. Boysen, M. Claussen, T. Ilyina, D. Olonscheck, T. Raddatz, B. Stevens, and J. Marotzke. The max planck institute grand ensemble: Enabling the exploration of climate system variability. *JAMES*, 11(7):2050–2069, 2019.

[17] M. Mihai and R. Westermann. Visualizing the stability of critical points in uncertain scalar fields. *Computers & Graphics*, 41:13–25, 2014.

[18] National Center for Atmospheric Research Staff (Eds). The climate data guide: Hurrell north atlantic oscillation (NAO) index (pc-based). https://climatedataguide.ucar.edu/climate-data/hurrell-north-atlantic-oscillation-nao-index-pc-based. Accessed: 2020-03-03.

[19] H. Obermaier, K. Bensema, and K. I. Joy. Visual trends analysis in time-varying ensembles. *IEEE TVCG*, 22(10):2331–2342, 2016.

[20] K. Potter, A. Wilson, P. Bremer, D. Williams, C. Doutriaux, V. Pascucci, and C. R. Johnson. Ensemble-vis: A framework for the statistical visualization of ensemble data. In *IEEE ICDMW*, pages 233–240, 2009.

[21] M. Rautenhaus, M. Böttinger, S. Siemen, R. Hoffman, R. M. Kirby, M. Mirzargar, N. Röber, and R. Westermann. Visualization in meteorology—a survey of techniques and tools for data analysis tasks. *IEEE TVCG*, 24(12):3268–3296, 2018.

[22] J. C. Rogers. The association between the north atlantic oscillation and the southern oscillation in the northern hemisphere. *Monthly Weather Review*, 112(10):1999–2015, 1984.

[23] J. C. Rogers. Patterns of low-frequency monthly sea level pressure variability (1899-1986) and associated wave cyclone frequencies. *Journal of Climate*, 3(12):1364–1379, 1990.

[24] J. Sanyal, S. Zhang, J. Dyer, A. Mercer, P. Amburn, and R. Moorhead. Noodles: A tool for visualization of numerical weather model ensemble uncertainty. *IEEE TVCG*, 16(6):1421–1430, 2010.

[25] U. Schulzweida. CDO user guide, 2019.

[26] Q. Shu, H. Guo, J. Liang, L. Che, J. Liu, and X. Yuan. Ensemblegraph: Interactive visual analysis of spatiotemporal behaviors in ensemble simulation data. In *IEEE PacificVis*, pages 56–63. IEEE, 2016.

[27] J. P. Snyder. *Map projections – A working manual*, volume 1395. US Government Printing Office, 1987.

[28] U. Ulbrich and M. Christoph. A shift of the NAO and increasing storm track activity over Europe due to anthropogenic greenhouse gas forcing. *Climate Dynamics*, 15, 1999.

[29] D. P. van Vuuren, J. Edmonds, M. Kainuma, K. Riahi, A. Thomson, K. Hibbard, G. C. Hurtt, T. Kram, V. Krey, J.-F. Lamarque, T. Masui, M. Meinshausen, N. Nakicenovic, S. J. Smith, and S. K. Rose. The representative concentration pathways: An overview. *Climatic Change*, 109(1):5, 2011.

[30] G. Walker and E. Bliss. Memoirs of the royal meteorological society. *World weather V*, 4:53–84, 1932.

[31] J. Wang, S. Hazarika, C. Li, and H. Shen. Visualization and visual analysis of ensemble data: A survey. *IEEE TVCG*, 25(9):2853–2872, 2019.

[32] J. Wang, X. Liu, H. Shen, and G. Lin. Multi-resolution climate ensemble parameter analysis with nested parallel coordinates plots. *IEEE TVCG*, 23(1):81–90, 2017.

[33] Y.-H. Wang, G. Magnusdottir, H. Stern, X. Tian, and Y. Yu. Decadal variability of the NAO: Introducing an augmented NAO index. *Geophys. Res. Lett.*, 39(21), 2012.

[34] T. Woollings, J. M. Gregory, J. G. Pinto, M. Reyers, and D. J. Brayshaw. Response of the north atlantic storm track toclimate change shaped by ocean-atmosphere coupling. *Nature Geoscience*, 5(5):313–317, 2012.