

Visualizing Confidence in Cluster-based Ensemble Weather Forecast Analyses

Alexander Kumpf, Bianca Tost, Marlene Baumgart, Michael Riemer, Rüdiger Westermann, and Marc Rautenhaus

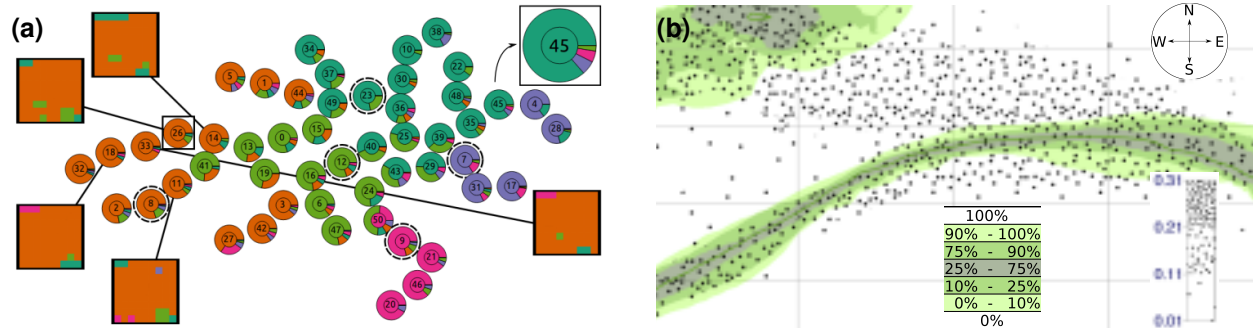


Fig. 1: (a) Analysis of the variation in cluster membership over 81 different clusterings of the case “Tropical Cyclone Karl”, an ensemble of 51 potential vorticity fields. Circular elements represent ensemble members, colors distinguish clusters (member 45 is enlarged: color of inner circles denotes reference cluster, surrounding pie-charts show how often the member was grouped into another cluster). Dashed outlines highlight cluster representative members. Member 26 is picked, for all members with similar cluster membership variation “variability matrix plots” (squared elements encoding cluster membership of all 81 clusterings) pop up. (b) A “contour probability plot” (CPP, different greens show probabilities for contour line occurrence) shows the variability of an iso-contour within a selected cluster. Overlaid stipple pattern shows the spatial variation of the plot with respect to the 81 clusterings.

Abstract— In meteorology, cluster analysis is frequently used to determine representative trends in ensemble weather predictions in a selected spatio-temporal region, e.g., to reduce a set of ensemble members to simplify and improve their analysis. Identified clusters (i.e., groups of similar members), however, can be very sensitive to small changes of the selected region, so that clustering results can be misleading and bias subsequent analyses. In this article, we –a team of visualization scientists and meteorologists– deliver visual analytics solutions to analyze the sensitivity of clustering results with respect to changes of a selected region. We propose an interactive visual interface that enables simultaneous visualization of a) the variation in composition of identified clusters (i.e., their robustness), b) the variability in cluster membership for individual ensemble members, and c) the uncertainty in the spatial locations of identified trends. We demonstrate that our solution shows meteorologists how representative a clustering result is, and with respect to which changes in the selected region it becomes unstable. Furthermore, our solution helps to identify those ensemble members which stably belong to a given cluster and can thus be considered similar. In a real-world application case we show how our approach is used to analyze the clustering behavior of different regions in a forecast of “Tropical Cyclone Karl”, guiding the user towards the cluster robustness information required for subsequent ensemble analysis.

Index Terms—Uncertainty visualization, ensemble visualization, clustering, meteorology

1 INTRODUCTION

In operational weather forecasting and atmospheric research, cluster analysis of ensemble weather prediction data is frequently used as a tool to analyze forecast uncertainty. Applied to scalar fields of the members of an ensemble forecast, cluster analysis groups together members with similar spatio-temporal development not known in advance [70]. One core objective in such analyses is the determination of representative weather scenarios, i.e., the trends, within a given region of the atmosphere. This region can be selected based on its geographic location

(e.g., the operationally computed clustering by the European Centre for Medium-Range Weather Forecasts (ECMWF) uses a European-Atlantic region [13]), or such that it contains a specific weather event of interest, e.g., a trough or a cyclone. Analysis of the computed clusters enables scientists and operational forecasters to distinguish between likely (i.e., forecast by many ensemble members), and unlikely trends. At the same time, by restricting to few representative scenarios, a subsequent analysis can often be sped up significantly.

Before using the cluster information as a basis for further analyses, however, an important question to be resolved is how much confidence can be put into the clustering result. Relying on a determined clustering introduces two major sources of uncertainty: a) The clustering cannot well identify the major trends in the ensemble. A possible cause can be that a region may have been chosen such that the atmospheric features of interest are not entirely covered by this region in all ensemble members, while at the same time features not of interest may be covered that consequently affect the clustering result. Another cause can be that the chosen number of clusters is not sufficient to capture all major trends. This raises the question (Q1) how much “value” is contained in the identified clusters in terms of the similarity of members within a cluster and the distinctness between the different identified clusters. b) The clustering may be sensitive to small changes in the selected region. This

- Alexander Kumpf, Bianca Tost, Rüdiger Westermann, Marc Rautenhaus are with the Computer Graphics & Visualization Group, Technische Universität München, Garching, Germany. E-mail: {alexander.kumpf, bianca.tost, westermann, marc.rautenhaus}@tum.de
- Marlene Baumgart and Michael Riemer are with the Institute of Atmospheric Physics, Johannes Gutenberg Universität Mainz, Mainz, Germany. Email: {mbaumga, mriemer}@uni-mainz.de

© 2017 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

requires to answer the question (Q2) how robust (i.e., representative) the clustering is with respect to changes of the region, i.e., how strongly do clusters vary when the region is changed. Without resolving these questions, any analysis based upon the obtained clustering cannot be assumed reliable. In the context of ensemble forecasts, however, we are not aware of studies that address these questions in a scientifically rigorous way. Often a merely tacit assumption is made that a “useful” clustering result has been computed and that it is safe to interpret subsequent analyses based on this clustering.

If it is found that a clustering is not robust, the domain expert needs to analyze the variation to understand its cause and to judge its “severeness” to find out if the clustering still contains useful information. By considering cluster membership of single members over multiple clustering results obtained from small changes of the region and identifying members with similar membership variations, sub-clusters with similar behavior as well as robust trends can be identified. The “severeness” of the cluster assignment instability can be judged by investigating its effect on per-cluster quantities (e.g., the cluster mean) that are interpreted in a subsequent analysis. Finally, the domain expert can interpret the obtained confidence information by using additional domain knowledge concerning the physical models and processes represented by the ensemble, and investigate the causes of observed cluster variations, inter-cluster relations, and member similarities and variations.

1.1 Contribution

We introduce novel visual analytics solutions for evaluating the confidence of an obtained clustering result; an overview of the proposed workflow is provided in Sect. 3 and Fig. 2. An artificial data set with known characteristics is used to motivate and introduce the visual encodings and interaction techniques we propose with regard to the requirements of the application task (Sect. 4). We apply our solutions to analyze a real-world ensemble forecast that predicted “Tropical Cyclone Karl” in 2016, and demonstrate that the visualization content we provide generates new and important insights (Sect. 5).

To resolve Q1, we provide a combination of abstract cluster-centric views with a linked map (Fig. 2c). The displays enable comparative visualization of the similarity of ensemble members within an identified cluster and the similarity of the members in the unclustered ensemble. In particular, we display where geographically the standard deviation (STDEV) of the clustered ensemble is reduced compared to the STDEV of the unclustered ensemble.

To resolve Q2 and the follow-up robustness analysis, we provide a set of interactive and linked cluster-centric, member-centric, and spatial views (Fig. 2d). A “cluster-centric robustness display” (CRD) visualizes changes of the clusters relative to a reference clustering (e.g., the clustering result for the initially chosen region). This view shows at a glance whether a clustering is stable or whether it changes significantly when the region is changed. Furthermore, the CRD communicates how many members change their cluster membership, and into which clusters these members change. The CRD is linked to a “member-centric robustness display” (MRD). It uses circular pie-chart diagrams for each ensemble member, arranged in a 2D coordinate system spanned by the first two major principal components (PCs) of the ensemble (Fig. 1a). We focus in particular on the requirement to visualize the cluster membership variation for each ensemble member when the clustering region is changed, so that the frequency of membership changes per member can quickly be perceived. The user can pick a member in the MRD to let the interface generate a matrix plot for this member in which the changes in cluster membership for each chosen region are shown. Simultaneously, matrix plots of all members having a similar cluster membership variation are displayed. This provides a very intuitive mechanism to instantly find sub-clusters of ensemble members that behave in a similar way.

Finally, we propose a cluster-centric view combining a variation of contour boxplots [68] and contour variability plots [15] with an overlaid stipple pattern. These “contour probability plots” (CPPs) (Fig. 1b) indicate the point-wise probability that within a selected cluster an isovalue is exceeded; stippling is used as a distinct visual channel to emphasize variations due to changes in the clustered region.

2 RELATED WORK

Clustering analysis is a well established statistical tool to identify groups of samples in a dataset that are similar with respect to a similarity measure. General introductions to data clustering are provided in the books by, e.g., Everitt et al. [12] and Gan and Ma [21]. For the presented work, related topics of particular relevance are the application of cluster-based analysis in meteorology, cluster-based ensemble visualization, and visual methods to judge the robustness of clustering results.

2.1 Cluster-based analysis in meteorology

A general overview of common clustering techniques and their application was provided by Wilks [70], a number of cluster visualization techniques for clustered climate data (no ensembles) were discussed by Nocke et al. [45]. In weather forecasting, clustering is commonly used in a static, pre-defined way. A representative example is the operational clustering of ensemble members at ECMWF [13], where forecast scalar fields of geopotential height are clustered in three different time windows for a static data region covering the ECMWF member states. Results are visualized in static matrix plots containing small forecast maps of the cluster representatives [13]. In meteorological research, ensemble cluster analysis has been used to improve understanding of various aspects of atmospheric predictability. For example, weather-regime related predictability was studied by using k-means clustering of jet wind profiles and comparing identified forecast regime transitions to observed climatological probability [17]. Clustering of leading PCs was used to study forecast scenarios in relation to tropical cyclones [1, 22, 30] and a heat wave that occurred in 2010 [53].

Clustering results were sometimes displayed by color-coded points in 2D PC plots [22, 53]. Different forecast scenarios were illustrated by cluster mean [1, 22] or by representative cluster members [30, 53]. Harr et al. [22] provided subjective criteria for the optimal number of clusters based on a discussion of the underlying meteorological charts. The sensitivity of the analysis to the number of PCs used in clustering is briefly discussed based on (subjective) changes in the cluster mean. Sensitivities with respect to the choice of the analysis domain are not discussed. In case of bifurcation-type behavior, clusters are often defined ad-hoc, for example as “good” and “bad” forecasts (e.g., [41, 66]). Statistically significant differences in the cluster mean of meteorological variables are sought to explain the forecast bifurcation.

2.2 Cluster-based ensemble visualization

With respect to visualization research, our approach is related to techniques for ensemble visualization – a sub-field of uncertainty visualization, for which a number of surveys exist [3, 40, 48, 52]. Uncertainty in scientific data is often estimated by means of ensembles – a set of representative realizations of a simulated phenomenon, obtained from simulations with different initial conditions and/or physical models. Such data is typically spatiotemporal, multivariate, and multivalued [29, 37], making the analysis and visualization processes difficult. Several methods have been proposed to reduce this complexity, e.g., by visualizing statistical summaries including mean and STDEV with color maps, contours, surface deformation, opacity, and variations of boxplots [11, 25, 37, 49, 51].

Clustering is another approach to reduce the complexity of ensemble data. Bordoloi et al. [4] proposed realization- and distribution-based hierarchical clustering to reduce the amount of information to be visualized. Bruckner and Möller [6] used density-based clustering to identify similar volumetric time sequences in physically-based ensemble simulations. Beham et al. [2] used hierarchical clustering to group similar geometric shapes. Reh et al. [58] clustered similar pores in industrial XCT data into mean objects (Mobjects) and then visualized the per-voxel probability of belonging to an Mobject using transfer functions. Hummel et al. [26] clustered using Minimum Spanning Trees for trend analysis, to compare the material transport in flow ensembles. Other techniques have used clustering to group iso-contours in scalar fields [8, 64] or streamlines in flow fields [47]. Bruckner and Möller [5] proposed the use of signed distance functions to analyze different iso-contours of the same scalar field, Rathi et al. [55] used such functions for shape analysis, and Ferstl et al. [14, 15] demonstrated

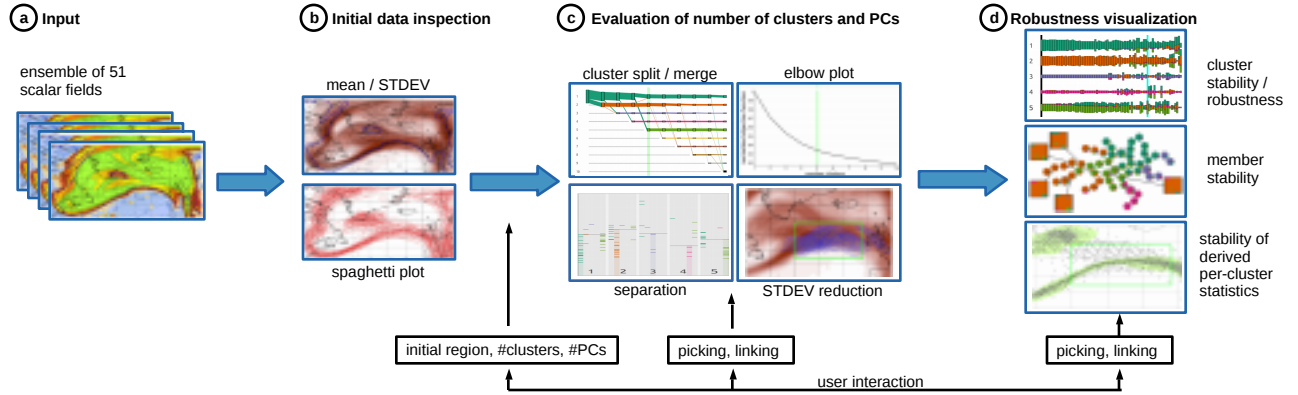


Fig. 2: Method overview. (a) Input is an ensemble of 2D scalar fields, which is (b) first inspected using statistical summaries and spaghetti plots. The user selects the region to be clustered and (c) uses different visualizations to interactively select a suitable number of clusters and of principal components to be used for clustering. Finally, (d) by means of linked views in which specific clusters and ensemble members can be selected, the value in a specific clustering and the robustness of clusters with respect to changes in the input region are analyzed.

the use of clustering to find visual abstractions for conveying the major trends in ensembles of trajectories and iso-contours. Recently, Ferstl et al. [16] proposed the use of clustering to detect similarities in the temporal variation of ensemble members.

2.3 Quality and robustness of clustering results

Common techniques to judge the quality and robustness of clustering results are surveyed in the book by Everitt et al. [12, Ch. 9]. Graphical depictions include banner plots, silhouette plots, and stripes plots (the latter are also used in our workflow). García-Escudero et al. [20] review robust clustering methods.

For the visualization of single clusterings, a number of standard techniques are known, including dendrograms to show hierarchical clusterings, heatmaps and clustergrams [61, 69], 2D principal component plots [44], and graph-based visualization techniques. Most commonly, clusters are visualized by giving each data point a location in a 2D map, so that similar data points are grouped together. For high-dimensional data, dimensional reduction is typically used to find appropriate locations [7, 24, 39]. Another option is to display clusters by means of parallel coordinates [23, 27]; a number of techniques have been proposed to extend parallel coordinates towards the visualization of clusters [19, 38, 50, 72], including the combination with heat maps [46] and radial tree displays [2, 32]. Lex et al. [36] introduce the Caleydo Matchmaker technique to visualize clusterings of subgroups of data using special parallel coordinate plots. This framework can also be used to compare different clustering strategies.

Only few approaches, to the best of our knowledge, combine the output of several clusterings into one; examples include the “cluster ensembles” by Strehl and Ghosh [62]. They applied different methods to obtain different clustering outcomes, and generated a single clustering from all outcomes. The probability accumulation by Wang et al. [67] aggregated multiple clustering outputs into a new one.

3 WORKFLOW OVERVIEW

Our proposed visual analytics solution enables meteorologists to interactively analyze the confidence in clustering results obtained from a selected spatial region. Fig. 2 shows an overview of the proposed workflow. Input (Fig. 2a) is a set of scalar forecast fields representing physical quantities including air temperature and potential vorticity (PV, a quantity computed from vorticity and stratification of an air mass; frequently employed to analyze the dynamics of weather systems). In this work, we consider scalar field data from the ECMWF Ensemble Prediction System (ENS; e.g., [35]). The ensemble comprises 50 perturbed members and an unperturbed control forecast (that is started from the “best” initial conditions). Past forecast data is available from the TIGGE archive [63] on a regular longitude–latitude grid in the horizontal; in the vertical, data is available on levels of constant pressure.

The user first inspects the data by means of non-cluster-based visualizations of the ensemble data (Fig. 2b). We provide maps of ensemble mean and STDEV, and spaghetti plots of relevant features including isocontours. This initial inspection is used to select the data region over which the clustering should be performed. Additionally, the user defines a range of numbers of clusters that are subsequently evaluated.

Following common practice in meteorology (cf. the operational ECMWF clustering [13] and research studies [1, 22]), our meteorological collaborators have explicitly requested the application of principal component analysis (PCA; e.g., [28]) to reduce input data dimensionality and k-means clustering. For the user-selected parameters, different types of diagrams facilitate determination of the optimal number of clusters suitable for separating major trends in the data (Fig. 2c). Cluster split-merge diagrams (CSMD) and “elbow plots” [65] indicate the number of clusters beyond which no significant gain in the cluster-specific objective function is achieved. “Stripes plots” [34] and displays of the reduction of STDEV achieved by a clustering convey how well clusters are separated and how well similar members have been identified (to resolve Q1).

For the determined number of clusters and the corresponding clustering (the “reference clustering”), a further CSMD visualizes changes in the clusters relative to the reference clustering when the number of used principal components (PCs) is changed. By default, as many PCs are used as to explain 80% of the variance of the data (default at ECMWF [13]; yielding on the order of 5 to 15 PCs), yet to judge robustness with respect to the number of PCs, the user is concerned with finding the smallest number of PCs that explain sufficient variance such that the clustering result does not change if more PCs are added.

Finally, with the selected region, number of clusters, and number of PCs, linked cluster- and member-centric views facilitate investigation of Q2, the robustness of the clustering with respect to changes in the selected region (Fig. 2d). If the clustering is found to be not robust, the user can focus on a particular cluster or ensemble member and visually analyze variation in cluster composition and membership to investigate the questions of: (Q3a) Which clusters are robust and which are not robust, and under which region changes are ensemble members changing cluster assignment? (Q3b) Which ensemble members are robust and which are changing ensemble assignment, and are subgroups of ensemble members changing in a similar way? (Q4) What are the effects of these changes on per-cluster quantities that are interpreted in a subsequent analysis?

4 VISUALIZATION TECHNIQUES

We motivate and introduce details of our interactive visual analysis workflow using a synthetic dataset that illustrates data characteristics that our users need to be able to analyze. This dataset with known clustering allows, e.g., to intentionally perform robust and non-robust

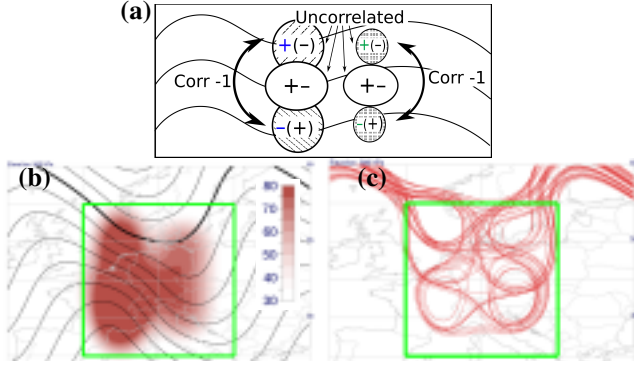


Fig. 3: The synthetic dataset. (a) Each member contains two dipoles (hedges) and in between each one a monopole. The polarities for all 4 poles are randomly chosen for each member. Neither (b) ensemble mean (contour lines) and STDEV of geopotential height in meters (color) nor (c) spaghetti plot of the thick black contour line in (b) indicate the dataset’s features or a suitable number of clusters. The green box is chosen as clustering region.

parameter changes in order to demonstrate correct identification by our methods.

4.1 Synthetic dataset as guiding example

The synthetic 2D dataset was generated with the same data modality as the real-world ECMWF ENS forecasts, comprising 51 members. We use a horizontal grid-spacing of 1° in longitude and latitude. Fig. 3a illustrates the patterns contained in the dataset. The individual ensemble members contain a wave-like pattern (the same in all members) that mimics the wave-patterns typically encountered in atmospheric data. In each member, noise is added to slightly distort the waves.

Clustering information is added by constructing features that uniquely group the members but are not discernible in plots of ensemble mean, STDEV or in spaghetti plots. Four independent features—two dipoles and two monopoles—are added to each member with a randomized polarity for each feature and member. The dipole located in the western part adds a stronger disturbance to the background wave than the eastern dipole. It is hence expected to dominate the clustering result when covered by the clustering domain. The uncorrelated monopoles are centered between the northern and southern parts of each dipole to further disturb the field so that an increased number of clusters can be identified.

In total, 16 different cases are generated, randomly distributed over the 51 members. However, due to the dominance of the western dipole and the high degree of non-correlation between the features, fewer clusters are expected to be detected by clustering. For instance, if both dipoles are covered by the cluster region, it is unlikely that a clustering result can separate both signals.

Figs. 3b, c show ensemble mean and STDEV, as well as a spaghetti plot for the initial data inspection. In the ensemble mean field, the features average out; the ensemble STDEV does not separate distinct features (Fig. 3b). The spaghetti plot of a contour line shows high uncertainty but gives no indication for the existence or number of clusters (Fig. 3c). Based on these plots, a user may choose the cluster region shown in Fig. 3b, covering the region containing high uncertainty.

4.2 Visualizing clustering “value”

Fig. 4 shows the interactive linked views we provide to determine the most suitable number of clusters by judging the value of the computed clusterings for different numbers (Q1). Here it is important that quantitative (about the number of members per cluster) and qualitative (about the spread of the members per cluster) information is provided, so that the domain expert can analyze which ensemble members belong robustly to a certain trend, and how representative the trend indicated by a certain cluster is. This information allows judging which clusters can be relied upon in the following meteorological application (Q4).

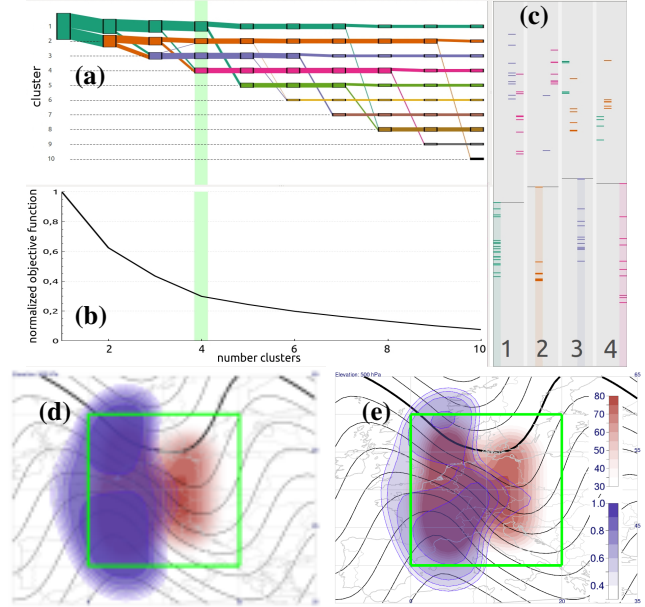


Fig. 4: Visualizations to judge clustering value for the synthetic case. (a) Cluster split-merge diagram (CSMD) for increasing number of clusters. (b) Elbow plot. (c) Stripes plot showing separation of clusters for the clustering selected in (a) (highlighted by green bar). (d) Reduction (blue) in fraction of STDEV (red, in meters) for four clusters is significantly improved compared to (e) three clusters.

An abstract CSMD (Fig. 4a) shows how clusters split as the number of clusters is increased and depicts where members change into different clusters. Each cluster is visualized by a vertical bar with constant width and unique color. A bar’s height represents the number of ensemble members contained in the corresponding cluster. Connecting bands between bars in subsequent clusterings indicate the number of ensemble members that stay in the same cluster or leave a cluster and merge into a different one. The view is augmented by an elbow plot (cf. Thorndike [65]), a simple plot of how the objective function minimized by the clustering algorithm decreases as the number of clusters is increased (Fig. 4b). The elbow plot indicates a suitable number of clusters; the point of the strongest “bend” in the curve can be interpreted as a trade-off between a small number of clusters and a small objective function. Unfortunately, the shape of the curves do not always allow this bend to be unambiguously identified.

For the synthetic use case (Fig. 4), we compute clusterings using $k = 2, \dots, 10$ clusters. Evaluation of the elbow plot indicates 4 clusters. The CSMD shows that for 5 clusters only the first cluster is split and all other clusters stay the same. For even more than 5 clusters we judge the further decrease of the objective function to be insufficient. So, we choose 4 clusters and keep in mind that the first cluster could be split into two if it does not show a clear trend.

To further evaluate a specific clustering, the user can select this clustering, i.e., by moving the green vertical line in the split-merge-diagram to this clustering. Upon movement, a linked stripes plot (Fig. 4c) is updated and displays specific information for this clustering. The stripes plot provides an abstract view on cluster separation; it has been proposed by Leisch [34] and shows the distance of each member to its cluster center and the distance to its second closest cluster center. If those two distances differ only little, the clusters are poorly separated, indicating that a different number of clusters may yield better results. The stripes plot, which notably does not provide any robustness information, indicates good separation for all numbers of clusters in the synthetic case, so in this case, it does not argue against any. For real world cases, clusters are usually less separated (e.g., clusters 1 and 5, Fig. 13). Linked spatial map views (Fig. 4d, e) provide further information on spatial characteristics of a selected clustering,

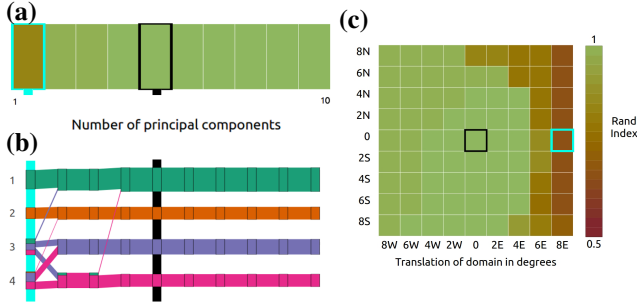


Fig. 5: Visualizations of cluster robustness. (a) Rand index and (b) CSMD visualizing the robustness with respect to changes in the number of principal components (PCs). For the synthetic case, using more than 4 PCs does not lead to further changes. (c) Rand matrix shows high sensitivity (red) to translations of the region to the east.

i.e., about the spread of the members in one cluster. A central quantity in our workflow is the reduction of STDEV per model grid-point [60], providing direct information about the average compactness of a cluster (Fig. 4d, e). STDEV reduction represents the ratio of STDEV within each cluster to the STDEV of the entire ensemble, thereby summarizing how much the identified clustering has grouped similar members into clusters. It can be computed for individual clusters and for the set of all clusters; in our workflow we display the total reduction of all clusters. STDEV reduction is computed as

$$r_k = 1 - \sqrt{\frac{f_k(\mathcal{S})}{f_1(\mathcal{S})}} \quad (1)$$

where f_k is the same as the k-means objective function

$$f_k(\mathcal{S}) = \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2, \quad \{S_1, \dots, S_k\} \in \mathcal{S}, \quad (2)$$

which is the sum of the intra-cluster sums of the squared errors. S_i denotes the set of all members in cluster i , x a member in cluster S_i , μ_i the mean of cluster S_i . \mathcal{S} is the set of all disjoint partitions of members $\{1, \dots, n\}$ into k clusters. The k-means algorithm minimizes Eq. 2 globally, it hence makes sense to use this measure for per-grid-point calculations as well. However, other measures exist which can be used for different clustering algorithms [31].

For the synthetic use case, STDEV reduction confirms four clusters to be appropriate (Fig. 4d); it is significantly improved compared to three clusters (Fig. 4e) but only marginally worse compared to five clusters (not shown).

4.3 Visualizing clustering robustness

Figs. 5 and 6 show the visualizations we provide to judge the robustness of the obtained clustering (Q2). In a first step, for the determined number of clusters and the corresponding clustering (the reference clustering), a plot of the Rand index [54] and a CSMD (Fig. 5a, b) are used to visualize changes in the clusters relative to the reference clustering when the number of used PCs is changed. The Rand index is a central summary measure to compare clusterings with differing input parameters; it measures agreement and disagreement of all pairs of members in two partitions. An index of 1 implies identical, an index of 0 maximally distinct clusterings. An index of 0.5 already indicates large changes; we clamp the transfer function at this value. Wu et al. [71] compared external clustering validity measures for k-means and recommended the use of the Rand index among others. For the synthetic case, 4 PCs are sufficient since more do not lead to changes in the cluster assignment anymore.

Next, the user can analyze the robustness of the identified clusters (Q3a, cluster-centric visualization), the robustness of the per-member assignment to specific clusters (Q3b, member-centric visualization), and the robustness of statistical quantities (including spatial fields, Q4)

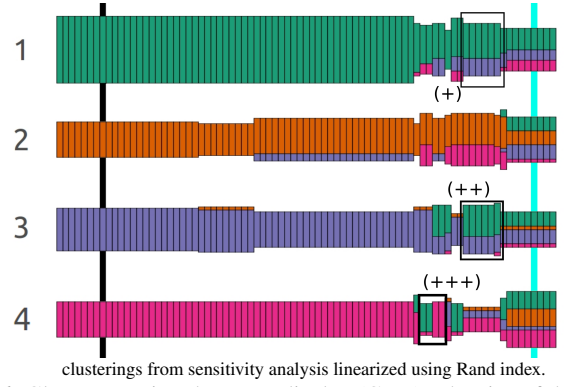


Fig. 6: Cluster-centric robustness display (CRD). The size of the bars represents cluster size, colors represent membership in reference clustering (black background bar). In clusterings (+, ++), the blue and green clusters exchange members; in clusterings (+++), cluster 4 changes almost completely. Turquoise background bar corresponds to turquoise clustering in Fig. 5c.

derived per-cluster to summarize cluster statistics. The latter includes robustness of cluster mean and STDEV, and the robustness of trend plots including CPPs.

For the synthetic case we evaluate the robustness of the clustering result with respect to domain translations up to 8° in each cardinal direction, using a 9×9 grid of 2° translations. Fig. 5c shows an example of the Rand index color-mapped to a 2D matrix representing these horizontal and vertical translations. The matrix highlights those region translations which strongly affect the clustering result (Q3a). The Rand matrix shows that for small perturbations the clustering is very robust; in fact, it hardly changes at all. For larger translations towards the east, however, the clustering changes significantly. This change could be caused by outliers, or it can indicate a new feature not captured in the initially selected region that starts to dominate. In the following, we introduce an additional type of diagram to further analyze the changes in clustering results due to region translations.

4.3.1 Cluster-centric robustness display

To provide to the user an overview of cluster changes for different region translations relative to the reference clustering (Q3a), we provide a cluster-centric robustness display (CRD; Fig. 6). The CRD is linked to the Rand matrix, to let the user select a clustering by picking on the corresponding matrix entry, and highlighting the corresponding information in the CRD. The CRD depicts on the horizontal axis the different clusterings for all possible region translations. Cluster sizes are indicated by the height of the bars, the colors show which proportion of which cluster of the reference clustering is contained in the selected clustering. To reduce overlap, the clusters are centered and the total space used by the bars in vertical direction can be adjusted.

To fit two-dimensional parameter changes into this view, a simple linearization of the clustering, or alternatively, an ordering minimizing the number of cluster membership changes, can be selected. To compute this ordering, all permutations of all clusterings are tested, which works well for up to 10 clusters. For more clusters, the Kuhn-Munkres algorithm [33] can be used, for which efficient implementations exist with a runtime complexity of $\mathcal{O}(n^3)$. When reducing the horizontal size of the bars, connections between them become visible which show where and how many members changed from one cluster to another (e.g., in Fig. 5b). Further, any realization can be selected in the CRD and all connected views automatically update using the selected realization. In Fig. 6, it can be seen that many translations do not affect the clustering at all, followed by some only affecting the orange and blue clusters. For extreme translations to the east, completely new cluster compositions arise, indicating that a new feature is starting to dominate.

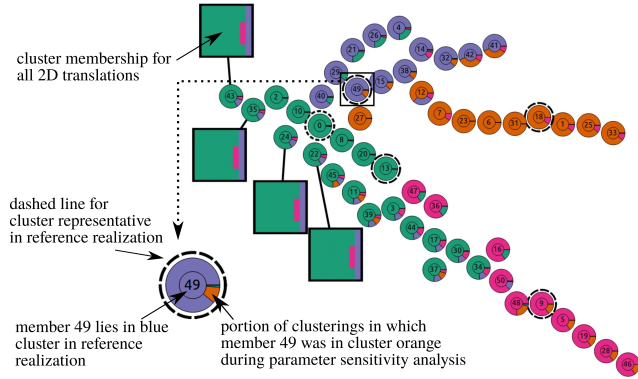


Fig. 7: The member-centric robustness display (MRD) shows a pie-chart for each ensemble member, each encoding robustness information. Stable and unstable members can be identified by means of cluster membership information encoded in the pie-chart. Matrix pop-ups show cluster membership changes with respect to region translations, arranged as in Fig. 5c.

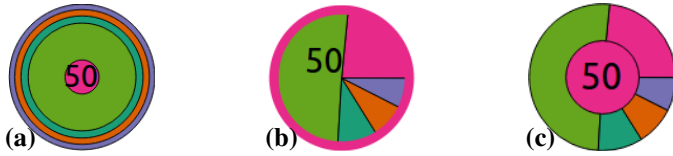


Fig. 8: Possible glyph designs for the MRD. (a) Spherical arrangement of membership information. (b) Membership information represented by a pie-chart, with reference cluster information surrounding the chart. (c) As (b) but reference cluster information in the chart center.

4.3.2 Member-centric robustness display

A further important task is the analysis of the changes in cluster membership of the ensemble members (Q3b). Such an analysis, in particular, has the purpose to identify members that remain together in the same cluster when region translations are performed or the number of clusters or PCs are changed. To visualize this member-centric information, we propose a member-centric robustness display (MRD) using a separate glyph for every member (see Fig. 7 for the synthetic use case).

In the design of the MRD the following criteria were considered: a) Each glyph needs to show simultaneously different types of information, i.e., the cluster of the member in the reference clustering, the member's unique ID, the frequency of changes in cluster membership for the performed region translations, and a detailed view of the relation between cluster membership and region translations. b) A large number of glyphs, i.e., one for each ensemble member, needs to be shown at once to enable a comparative study of members. Due to b), our first design decision was to separate the visual encoding of the latter type of information in a) from the others, to avoid visual overloading and clutter. Our second design decision was with respect to the glyphs shape and structure. In Fig. 8, we show possible circular (to minimize coverage) glyph designs, of which the third one is the design we propose.

In our design, each member is represented by an inner circle and a surrounding pie-chart. The color of the inner circle indicates the reference cluster of the member, and the member ID is shown in the inner circle. The pie-chart shows how often member i was in cluster j , with the frequency sorted decreasingly –but always starting with the reference cluster– and displayed counter-clockwise. If the member is the cluster representative in the reference clustering, a dashed border around the outer circle is added. When looking at the two alternative designs, one can clearly observe the following: In design 8a, where the pie chart is replaced by circular rings depicting the membership variations, the inner circle and rings might become so small and thin that the important information cannot be perceived clearly any more. In

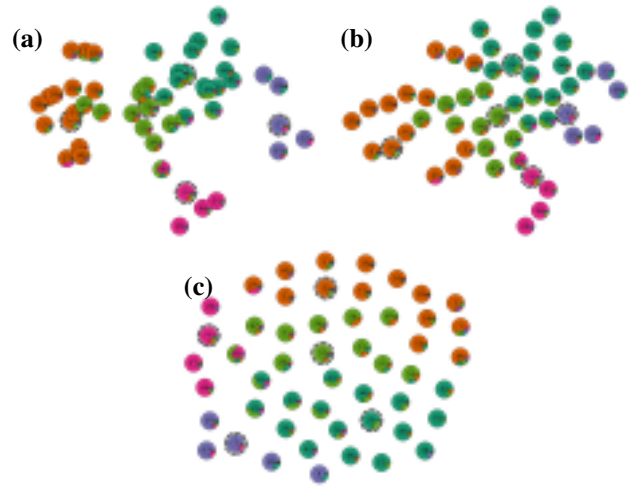


Fig. 9: MRD layouts based on (a) the first two PCs, (b) as (a) but squeezed and without overlap, (c) a force directed graph layout.

design 8b, the member ID obscures some of the pie-slices, and the outer ring encoding the reference cluster can become too thin to be clearly observable (under the constraint that the glyph cannot be arbitrarily large and the pie-chart needs sufficient space).

A click on a pie-chart opens a matrix plot giving a detailed view of the member-specific relation between cluster membership and region translations. Color in the displayed matrix indicates the member's cluster assignment in all clusterings. Double-clicking on a pie-chart opens matrix plots for all members with similar cluster change characteristics, which can immediately reveal similar patterns across multiple members. The pop-ups are placed around the convex hull of the already drawn pie charts to avoid occlusions. In case the closest pop-up position leads to overlap with another pop-up, it is translated to the next closest position without producing overlap.

For the placement of pie-charts on the 2D canvas, we use the members' locations in the 2D coordinate system spanned by the 1st and 2nd PCs (Fig. 9a). Despite more PCs being used for the clustering in most cases, the first two components provide the most separation between members and, thus, yield least overlap when glyphs are drawn. Nevertheless, overlaps can occur, requiring strategies to avoid them.

A force-directed graph layout as proposed by Fruchterman and Reingold [18] can be used to enforce that members of the same clusters are placed close to each other. Here, the members are initially placed on a circle, then "pulled" together by using their clustering distance measure as a force. In the resulting plot, the members are placed very homogeneously in the 2D image (Fig. 9c). On the downside, the visual separation of the clusters which is present in Fig. 9a gets lost. Due to this, we perform a different placement strategy, which tries to resolve the overlaps that are introduced by the initial approach. First, the distance between the two pairwise closest members is determined, and these two members are then considered being processed. All remaining members are successively moved in the direction of the processed members until their distance to them is approximately the same as the minimum distance. After all glyphs are shrunk together, the domain is rescaled to fit the available space. The result is shown in Fig. 9b. In the MRD in Fig. 7, there are multiple entirely stable members in the blue and green cluster (e.g., member 0), indicating stable sub-clusters. When picking member 35, the pop-ups display 4 members also changing only for extreme eastward translations. Those 4 members are always clustered together forming a stable sub-scenario.

4.3.3 Robustness of per-cluster summary statistics

Finally, we provide information on the robustness of per-cluster summary statistics, including plots of major-trends and cluster mean and STDEV. To visualize major trends, variability plots [15] and contour boxplots [68] are well suited, however, since they do not show point-

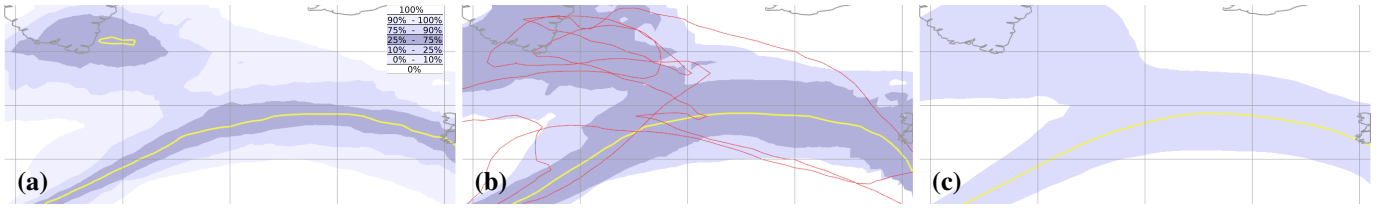


Fig. 10: We propose (a) “contour probability plots” (CPPs), lobes enclosed by iso-contours of constant probability of a set of members exceeding the threshold value for which the lobe is shown. The lobes have similar shape as (b) contour boxplots [68] (50% inner lobe) and (c) variability plots [15] (1.5 STDEV) but are more straightforward to interpret when overlaid with robustness stippling.

wise quantitative information that can be overlaid by clustering robustness information, we propose a variation, which we call “contour probability plots” (CPP). A CPP depicts lobes which indicate the probability that a contour line is locally contained in the lobe. CPPs are generated by grid-point-wise computation of the probability that the considered isocontour threshold is exceeded in the ensemble. Isocontours of this probability field are then used to construct the lobes, e.g., the 25% and 75% isocontours enclose the 50% band. The example in Fig. 10a displays lobes indicating bands for the inner 50%, and outer 15% and 10% on each side. When CPPs for multiple clusters are overlaid, a reduced set of lobes is beneficial to avoid cluttering. A distinct advantage of CPPs is that the sensitivity, calculated as the STDEV of this probability over different clusterings, can be interpreted as the average translation of the contour plot borders. This allows the user to see right away how the contours of a cluster might change when using a different clustering region.

To visualize the robustness of a cluster, i.e., its variation in terms of membership changes, a distinct visual channel that can be combined with CPPs is required. Retchless and Brewer [59] investigated different types of overlays to depict uncertainty in climate variables. Following their study, we use stippling since it can be clearly distinguished from the colored underlay and does not interfere with other communication channels. Our rendering technique follows the texture-based approach presented by Maskey and Newman [42].

Fig. 1b shows the sensitivity of a CPP, Fig. 11 that of STDEV reduction plots via stippling. In Fig. 1b, the stippling shows a large area where the STDEV of the probability is on the order of 0.1–0.2, indicating, e.g., that the lobes could well extend into the white region if the selected region is changed. The displayed cluster characteristics are hence very uncertain. In Fig. 11, a high uncertainty is visible throughout the initially clustered area, indicating that for translated domains the geographical areas in which variance is reduced by clustering also change. For instance, when considering a clustering region translated to the east (Fig. 11b), we see a very different area being clustered “well”. Here, the two eastward dipole features were detected. The contour probability plots clearly confirm this detection of the distinct features by different cluster regions (Fig. 11c and d).

5 RESULTS

All visualization techniques, as well as PCA and clustering methods, have been implemented in the open-source meteorological ensemble visualization tool “Met.3D” [43, 57]. Met.3D provides a suitable infrastructure that already included an ensemble data processing pipeline and visualization functionality for meteorological maps. Stippling functionality, as well as all abstract and linked charts presented in this article, have been added in the context of this work. Clustering is implemented using the library “Cluster 3.0” [9].

To compensate for the denser geometric grid-point spacing towards the poles in the regular longitude–latitude grid, each grid-point needs to be weighted by the size of its grid-cell. We follow Ferranti and Corti [13] and weigh each grid point with the cosine of its latitude prior to applying PCA to the data. Clustering and cluster matching for different realizations require most computing resources; on an Intel® Xeon CPU E5-1650 v2 with 6 cores @ 3.50GHz, each cluster run took from a few seconds up to two minutes depending on the data. Cluster runs are executed in parallel, for the presented case study clustering

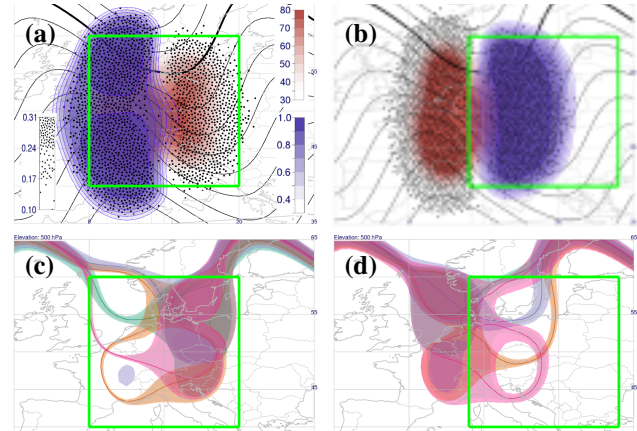


Fig. 11: (a) Reduction (blue) in fraction of ensemble STDEV (red, in meters) for the synthetic case, overlaid with robustness due to region translations (stippling shows STDEV of reduction). The intense stippling indicates that region translations can cause clusterings to reduce ensemble STDEV in different regions. (b) STDEV reduction caused by a region shifted eastward. (c, d) CPPs 85% bands of the respective regions indicate that different features dominated the clusterings.

and cluster matching required on the order of 20 minutes. Once the data are clustered, visualization performance is interactive.

5.1 Tropical Cyclone Karl

To demonstrate the application and value of our method, we discuss a real-world forecast issued during the North Atlantic Waveguide and Downstream Impact Experiment (NAWDEX [10]), a field campaign in which three of the authors have been involved. Tropical Cyclone Karl crossed the North Atlantic in late September 2016 and was associated with heavy precipitation in Norway. A number of days prior to measurement flights it posed significant difficulties in forecasting due to high uncertainty in the forecasts. A major objective of the data analysis activities of the campaign is an investigation of ensemble behavior and determination of the physical processes that caused the uncertainty.

We consider the ECMWF ENS forecast from 00:00 UTC 22 September 2016 (data are at 1° horizontal resolution, on pressure levels in the vertical and at 6-hourly time steps) and focus on the development of Karl in subsequent days. In this ensemble, very different developments for Karl were predicted. Of critical interest for the analysis of the situation is the interplay between upper-level PV (of particular interest is the 2-PV-units (PVU) isosurface used to represent the tropopause, the boundary between the troposphere and stratosphere), the upper-level jet stream (strong winds in the vicinity of the tropopause), lower-level temperature and moisture fields, and Warm Conveyor Belts (airstreams that lift warm and moist air from near the surface to the upper troposphere; c.f. [56]). In the analysis presented here, we –the meteorological domain experts in the author team– are interested in whether we can determine distinct scenarios (i.e., clusters) in the upper-level PV field to investigate the relationship between PV and lower-level developments, e.g., the low-level potential temperature field.

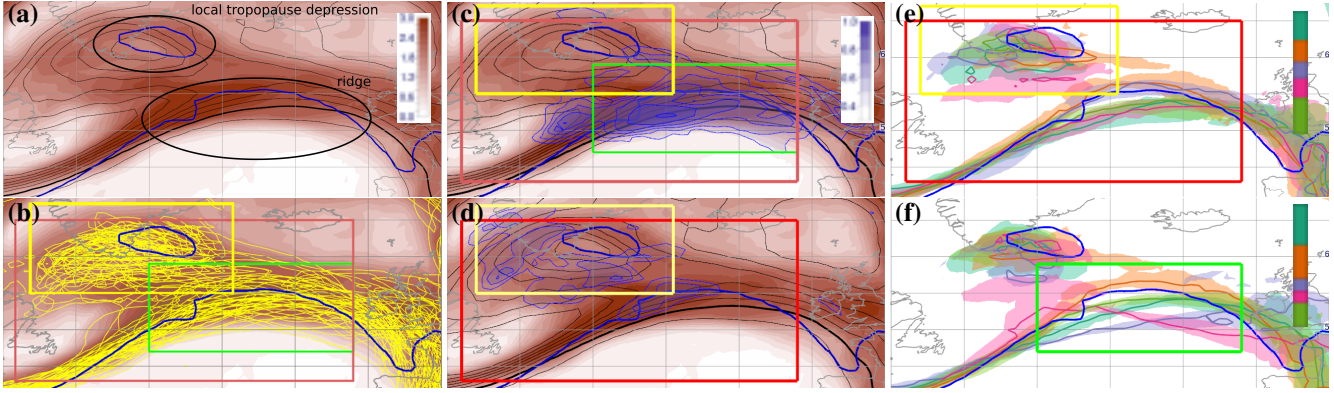


Fig. 12: Case study. A clustering region is chosen based on (a) ensemble mean (black contours) and standard deviation (color in PVU) of potential vorticity at 250 hPa, and (b) a spaghetti plot of the 2-PVU isocontour. Shown is the ECMWF ENS forecast from 00:00 UTC 22 September 2016, valid at 00:00 UTC 26 September 2016. Thick blue contour shows ground-truth observation; red, green and yellow boxes represent clustering regions. (c) STDEV reduction (blue) when clustering the red region. There is almost no STDEV reduction in the yellow region. (d) Clustering the yellow region leads to only little STDEV reduction. (e) CPP (85% bands) of the 2-PVU isocontour showing clusters of the yellow clustering region. No particularly different scenarios have been identified. (f) When clustering the green region, distinct scenarios for the ridge are identified.

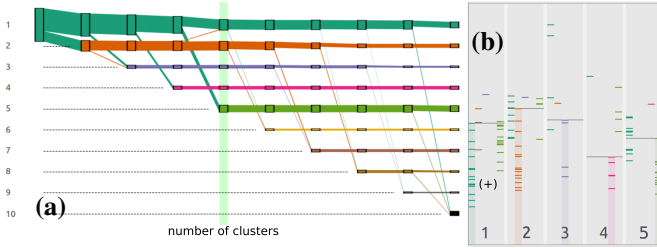


Fig. 13: (a) CSMD and (b) stripes plot (for five clusters) showing weak separation of clusters 1 and 5 (+). The CSMD shows the creation of a robust cluster 5 when increasing the number of clusters from four to five.

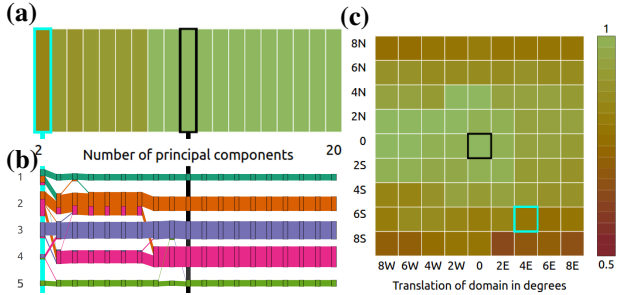


Fig. 14: Case study. (a) Rand index and (b) CSMD for changes in the number of PCs; more than 11 PCs do not change the result. (c) Rand matrix shows high sensitivity to translations of the domain in particular at the northern and southern edges.

For initial data inspection (Fig. 2b), we consider PV at an elevation of 250 hPa at 00:00 UTC 26 September 2016 (96 hours lead time). Fig. 12a shows the ensemble mean PV field, the ground-truth 2-PVU isocontour that was observed later, and the ensemble STDEV computed from all members. The corresponding spaghetti plot of the 2-PVU line is shown in Fig. 12b. Over the North Atlantic, the STDEV field shows high uncertainty in the vicinity and to the north of the tropopause. We are in particular interested in the formation of the ridge (the northward undulation of the 2-PVU line) to the south of Iceland and Greenland and the local tropopause depression over Southern Greenland (which is particularly hard to discern in Fig. 12b).

Based on Fig. 12a and b, we choose the large red clustering region shown in Fig. 12b, encompassing both features. To determine the num-

ber of clusters, a rather smooth curve of the elbow plot does not provide particularly strong visual guidance (not shown), we hence investigate the difference between the clusterings containing four, five, and six clusters in the corresponding CSMD and stripes plots (Fig. 13). At the transition from four to five clusters, a rather robust and substantially sized cluster 5 (light green) is formed. With six clusters, only a very small new cluster is formed (Fig. 13a). While the stripes plot shows that clusters 1 and 5 have only limited separation, STDEV reduction is further improved with five clusters compared to four clusters in the region of meteorological interest, i.e., along the tropopause in which most variability occurred (only shown for the five-cluster realization in Fig. 12c). The optimal number of PCs is found to be 11 (Fig. 14).

Interestingly, while the STDEV is reduced along the ridge, there is virtually no reduction in the vicinity of the local tropopause depression (Fig. 12c). We conclude that the ridge feature dominates the clustering but re-configure our method to use two smaller regions centered on the ridge (green region) and on the local depression (yellow region) for confirmation and to determine if we can obtain a useful clustering of the depression region. Both regions cluster best with five clusters. The green region yields slightly better separated clusters for the ridge, the STDEV reduction largely resembles that of the red region (not shown). Clustering the yellow region, however, still only leads to marginal STDEV reduction even though it is now centered on the depression (Fig. 12d). The stripes plot confirms that identified clusters are not well separated (not shown). Similarly, the CPP displaying the scenarios of the 2-PVU line shows neither well separated trends in the depression region nor in the ridge region (Fig. 12e). We conclude that data in this region cannot be well clustered.

We focus on the clustering obtained from the green region. Fig. 12f shows the corresponding CPP, showing the 2-PVU trends identified by this clustering. Cluster 2 (orange) most closely matches the observation; however, none of the cluster means represents the observation very well. The scenarios differ with respect to amplitude and orientation of the ridge. While the strongest ridge is represented by cluster 2 (orange), cluster 3 (blue) represents the cluster with the weakest ridge. In terms of ridge orientation the largest difference is found between cluster 4 (pink) and 3 (blue). Clusters 5 (light green) and 1 (dark green) are very similar to each other. This is confirmed by the CSMD and stripes plot, they are not strongly separated and cluster 5 contains a large number of members that in the four-cluster realization belong to cluster 1. We perform a robustness analysis for region translations from 2° to 8° in 2° steps in all four cardinal directions. The resulting Rand matrix and CRD are shown in Figs. 15 and 14c. While the clustering result is rather robust for east-west translations, a high sensitivity can be observed for translations in the north-south (N/S) direction. In particular for translations by more than 4° N/S the results change strongly. A closer inspection reveals,

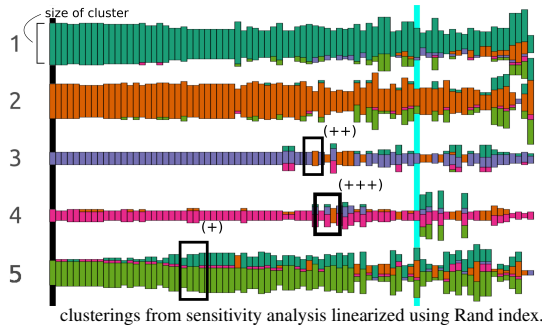


Fig. 15: Case study CRD for region translations. The green clusters (1 and 5) exchange many members (+) and in extreme cases, clusters 3 and 4 change completely (++, +++).

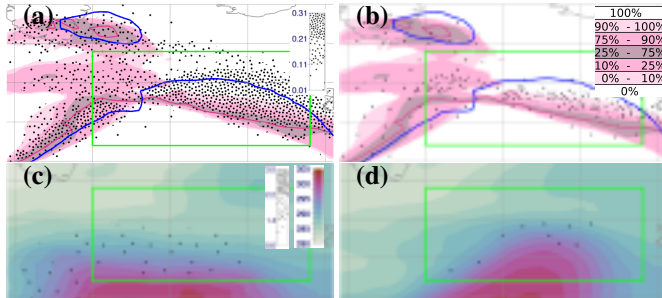


Fig. 16: (a) The CPP of the 2-PVU isocontour at 250 hPa for the pink cluster 4 is very sensitive to changes in the selected region (stippling shows STDEV of probability). (b) It becomes much more robust if the north-south region translations by 6° and 8° are omitted (cf. Fig. 14c). (c) Cluster means for cluster 4 and (d) cluster 1 (dark green) of potential temperature at 850 hPa (stippling shows STDEV of potential temperature with extreme north-south translations omitted) show distinctly different scenarios.

however, that the dominant feature (the variability associated with the strong PV gradient along the tropopause) of the cluster region is more or less removed by these translations. The high sensitivity related to these translations is thereby plausible from a physical viewpoint.

The CRD (Fig. 15) shows that members often change between clusters 1 and 5 when translating the cluster region – this is plausible due to their weak separation. The MRD (Fig. 1a) reveals the members that are switching cluster assignment. Less intuitive are exchanges between the blue and pink clusters indicated in both CRD and MRD, as the CPP shows these two clusters to represent quite different trends (Fig. 12f). Selection of the corresponding realizations in the CRD reveals that these exchanges occur for the extreme N/S translations; in these translated regions the two clusters seem to be more similar.

For further analysis, we are interested in cluster summary statistics as well as individual members that robustly remain within a given cluster. For the sake of brevity, here we only consider upper-level PV and lower-level potential temperature. The MRD (Fig. 1a) shows that for the orange cluster, the members in the upper left region of the plot appear to be robust. Inspection of further robustness details (we have selected member 26; our system also shows further members with the same robustness characteristics) reveals that these members indeed are robust with respect to E/W translations; changes occur only for N/S shifts. As a contrary example, member 50 (pink cluster) is particularly unstable (Fig. 8) and is therefore to be used with caution.

With respect to cluster summary statistics, Fig. 16 shows CPPs of the 2-PVU line at 250 hPa for cluster 4 and cluster means of potential temperature at 850 hPa for clusters 4 and 1. The variability of the respective fields due to clustering robustness is encoded as stippling. Confirming the above findings of CRD and MRD, the cluster summary statistics become much more robust when the physically unreasonable, extreme translations in the N/S directions are not considered (Fig. 16a

vs. b). This increased robustness strongly increases our confidence in the subsequent physical interpretation. Clusters 1 and 4 differ with respect to the longitudinal position of the ridge maximum (cf. Fig. 12f), and also, the corresponding potential temperature fields (Fig. 16c vs. d) show distinct structure differences similar to the structure differences of the upper-level ridges. This similarity yields evidence for a baroclinic nature of the evolution, i.e., a strong coupling between the evolutions at upper and lower levels.

We conclude by noting that our preliminary analysis provides the key insight that there are robust and physically-meaningful forecast scenarios associated with the evolution of Karl in the region that was of interest during the NAWDEX campaign. Future analysis will consider in more detail the processes that lead to the divergence of the forecasts into these scenarios and the ramifications to the heavy precipitation event in Norway a few days later. Compared to methodology previously applied in the meteorological community, the method proposed here greatly facilitates the analysis and provides at the same time unprecedented quantification and visual representation of cluster robustness.

6 SUMMARY AND CONCLUSIONS

We have proposed a novel visual analysis workflow to visualize the value and, in particular, the robustness of cluster-based analyses of ensemble weather forecasts. This information is required for analysis and interpretation of the obtained clusters, however, it has been largely neglected in the context of ensemble cluster analysis in the past.

Our approach enables the user to identify a suitable number of clusters by visualizing how clusters split, are separated, and reduce an ensemble's STDEV. To investigate the robustness of a clustering, its sensitivity to changes in the selected clustering region and the number of PCs used for data reduction can be visualized. Cluster-centric and member-centric views show the stability of clusters and how members change cluster assignment when region or PCs are changed; linked visualizations of per-cluster summary statistics are augmented by overlays that indicate the sensitivity of the displayed quantities.

We have developed our methodology in a team of visualization and atmospheric scientists, integrated it into the open-source meteorological ensemble visualization software Met.3D for straightforward application by users, and have demonstrated its benefit with a real-world case study taken from ongoing work aimed at improving the accuracy of numerical weather prediction. Compared to methodology previously applied in the meteorological community, the method proposed here greatly facilitates the analysis and provides at the same time unprecedented quantification and visual representation of cluster robustness. In the near future, the method will actively be used in data analysis activities related to the NAWDEX campaign. Except for the region translations and meteorological views, we believe that the proposed workflow and the CRD and MRD can be used in other domains than meteorology as well. The extension to other deterministic clustering algorithms such as hierarchical clustering would only need small adaptations such as the use of dendrograms. Investigation of its extensibility to further clustering parameters, to time-dependent data, and to further clustering methods including fuzzy clustering would in our opinion be beneficial to the atmospheric community; it is left as an open issue for future work. A further limitation is the current brute-force computation of the clustering realization to obtain robustness information; here, approaches that can estimate this information based on fewer clustering runs would be beneficial.

ACKNOWLEDGMENTS

The research leading to these results has been done within the subprojects B5 "Data-driven ensemble visualization" and A1 "Upscale impact of diabatic processes from convective to near-hemispheric scale" of the Transregional Collaborative Research Center SFB/TRR 165 "Waves to Weather" funded by the German Research Foundation (DFG). The work was partly funded by the European Union under the ERC Advanced Grant 291372 SaferVis: Uncertainty Visualization for Reliable Data Discovery. Access to ECMWF prediction data has been kindly provided in the context of the ECMWF special project "Support Tool for HALO Missions".

REFERENCES

- [1] D. Anwender, P. A. Harr, and S. C. Jones. Predictability associated with the downstream impacts of the extratropical transition of tropical cyclones: Case studies. *Monthly Weather Review*, 136(9):3226–3247, Sept. 2008.
- [2] M. Beham, W. Herzner, M. E. Gröller, and J. Kehler. Cupid: Cluster-based exploration of geometry generators with parallel coordinates and radial trees. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1693–1702, 2014. doi: 10.1109/TVCG.2014.2346626
- [3] G.-P. Bonneau, H.-C. Hege, C. Johnson, M. Oliveira, K. Potter, P. Rheingans, and T. Schultz. Overview and State-of-the-Art of uncertainty visualization. In C. D. Hansen, M. Chen, C. R. Johnson, A. E. Kaufman, and H. Hagen, eds., *Scientific Visualization, Mathematics and Visualization*, pp. 3–27. Springer London, 2014.
- [4] U. D. Bordoloi, D. L. Kao, and H.-W. Shen. Visualization techniques for spatial probability density function data. *Data Science Journal*, 3:153–162, 2004. doi: 10.2481/dsj.3.153
- [5] S. Bruckner and T. Möller. Isosurface similarity maps. In *Computer Graphics Forum*, vol. 29, pp. 773–782. Wiley Online Library, 2010. doi: 10.1111/j.1467-8659.2009.01689.x
- [6] S. Bruckner and T. Moller. Result-driven exploration of simulation parameter spaces for visual effects design. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1468–1476, 2010. doi: 10.1109/tvcg.2010.190
- [7] K. Bunte, S. Haase, M. Biehl, and T. Villmann. Stochastic neighbor embedding (SNE) for dimension reduction and visualization using arbitrary divergences. *Neurocomputing*, 90:23–45, 2012. doi: 10.1016/j.neucom.2012.02.034
- [8] H. Carr, D. Brian, and D. Brian. On histograms and isosurface statistics. *IEEE Transactions on Visualization and Computer Graphics*, 12(5), 2006. doi: 10.1109/tvcg.2006.168
- [9] M. J. de Hoon, S. Imoto, J. Nolan, and S. Miyano. Open source clustering software. *Bioinformatics*, 20(9):1453–1454, 2004. doi: 10.1093/bioinformatics/bth078
- [10] Deutsches Zentrum für Luft- und Raumfahrt. NAWDEX - North Atlantic Waveguide and Downstream Impact Experiment. <http://www.pa.op.dlr.de/nawdex>, 2017. Accessed 30 July 2017.
- [11] S. Djurcilov, K. Kim, P. Lermusiaux, and A. Pang. Visualizing scalar volumetric data with uncertainty. *Computers & Graphics*, 26(2):239–248, Apr. 2002. doi: 10.1016/S0097-8493(02)00055-9
- [12] B. S. Everitt, S. Landau, M. Leese, and D. Stahl. *Cluster Analysis*. Wiley, 5 ed., Feb. 2011.
- [13] L. Ferranti and S. Corti. New clustering products. *ECMWF Newsletter*, 127:6–11, 2011.
- [14] F. Ferstl, K. Bürger, and R. Westermann. Streamline variability plots for characterizing the uncertainty in vector field ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):767–776, Jan 2016. doi: 10.1109/TVCG.2015.2467204
- [15] F. Ferstl, M. Kanzler, M. Rautenhaus, and R. Westermann. Visual analysis of spatial variability and global correlations in ensembles of Iso-Contours. *Computer Graphics Forum*, 35(3), 2016. doi: 10.1111/cgf.12898
- [16] F. Ferstl, M. Kanzler, M. Rautenhaus, and R. Westermann. Time-hierarchical clustering and visualization of weather forecast ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):831–840, 2017. doi: 10.1109/TVCG.2016.2598868
- [17] T. H. A. Frame, M. H. P. Ambaum, S. L. Gray, and J. Methven. Ensemble prediction of transitions of the north atlantic eddy-driven jet. *Quarterly Journal of the Royal Meteorological Society*, 137(658):1288–1297, 2011. doi: 10.1002/qj.829
- [18] T. M. Fruchterman and E. M. Reingold. Graph drawing by force-directed placement. *Software: Practice and experience*, 21(11):1129–1164, 1991. doi: 10.1002/spe.4380211102
- [19] Y.-H. Fua, M. O. Ward, and E. A. Rundensteiner. Hierarchical parallel coordinates for exploration of large datasets. In *Visualization '99. Proceedings*, pp. 43–508, 1999. doi: 10.1109/VISUAL.1999.809866
- [20] L. García-Escudero, A. Gordaliza, C. Matrán, and A. Mayo-Iscar. A review of robust clustering methods. *Advances in Data Analysis and Classification*, 4(2-3):89–109, June 2010. doi: 10.1007/s11634-010-0064-5
- [21] J. W. Guojun Gan, Chaoqun Ma. *Data Clustering: Theory, Algorithms, and Applications*. ASA-SIAM Series on Statistics and Applied Probability. SIAM, Society for Industrial and Applied Mathematics, 2007.
- [22] P. A. Harr, D. Anwender, and S. C. Jones. Predictability associated with the downstream impacts of the extratropical transition of tropical cyclones: Methodology and a case study of typhoon nabi (2005). *Monthly Weather Review*, 136(9):3205–3225, Sept. 2008.
- [23] H. Hauser, F. Ledermann, and H. Doleisch. Angular brushing of extended parallel coordinates. In *Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on*, pp. 127–130. IEEE, 2002. doi: 10.1109/infvis.2002.1173157
- [24] G. Hinton and S. Roweis. Stochastic neighbor embedding. *Advances in neural information processing systems*, 15:833–840, 2003.
- [25] T. Höllt, A. Magdy, P. Zhan, G. Chen, G. Gopalakrishnan, I. Hoteit, C. D. Hansen, and M. Hadwiger. Ovis: A framework for visual analysis of ocean forecast ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 20(8):1114–1126, 2014. doi: 10.1109/tvcg.2014.2307892
- [26] M. Hummel, H. Obermaier, C. Garth, and K. I. Joy. Comparative visual analysis of lagrangian transport in CFD ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2743–2752, 2013. doi: 10.1109/tvcg.2013.141
- [27] A. Inselberg. The plane with parallel coordinates. *The visual computer*, 1(2):69–91, 1985.
- [28] I. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer, 2nd ed. ed., 2010.
- [29] J. Kehler and H. Hauser. Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 19(3):495–513, Mar. 2013. doi: 10.1109/tvcg.2012.110
- [30] J. H. Keller, S. C. Jones, J. L. Evans, and P. A. Harr. Characteristics of the TIGGE multimodel ensemble prediction system in representing forecast variability associated with extratropical transition. *Geophysical Research Letters*, 38(12):L12802+, June 2011.
- [31] F. Kovács, C. Legány, and A. Babos. Cluster validity measurement techniques. In *Proceedings of the 6th International Symposium of Hungarian Researchers on Computational Intelligence, Budapest*, pp. 18–19. Citeseer, 2005.
- [32] M. I. Krzywinski, J. E. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, and M. A. Marra. Circos: An information aesthetic for comparative genomics. *Genome Research*, 19(9):1639–1645, 2009. doi: 10.1101/gr.092759.109
- [33] H. W. Kuhn. The Hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955. doi: 10.1002/nav.3800020109
- [34] F. Leisch. Neighborhood graphs, stripes and shadow plots for cluster visualization. *Statistics and Computing*, 20(4):457–469, 2010. doi: 10.1007/s11222-009-9137-8
- [35] M. Leutbecher and T. N. Palmer. Ensemble forecasting. *Journal of Computational Physics*, 227(7):3515–3539, 2008. doi: 10.1016/j.jcp.2007.02.014
- [36] A. Lex, M. Streit, C. Partl, K. Kashofer, and D. Schmalstieg. Comparative analysis of multidimensional, quantitative data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1027–1035, Nov 2010. doi: 10.1109/TVCG.2010.138
- [37] A. L. Love, A. Pang, and D. L. Kao. Visualizing spatial multivalued data. *IEEE Computer Graphics and Applications*, 25(3):69–79, May 2005. doi: 10.1109/mcg.2005.71
- [38] Y. Luo, D. Weiskopf, H. Zhang, and A. E. Kirkpatrick. Cluster visualization in parallel coordinates using curve bundles. *IEEE Transaction on Visualization and Computer Graphics*, 18, 2008.
- [39] L. v. d. Maaten and G. Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.
- [40] A. M. MacEachren, A. Robinson, S. Hopper, S. Gardner, R. Murray, M. Gahegan, and E. Hetzler. Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science*, 32(3):139–160, July 2005. doi: 10.1559/1523040054738936
- [41] L. Magnusson, J.-R. Bidlot, S. T. K. Lang, A. Thorpe, N. Wedi, and M. Yamaguchi. Evaluation of medium-range forecasts for hurricane sandy. *Monthly Weather Review*, 142(5):1962–1981, 2014. doi: 10.1175/MWR-D-13-00228.1
- [42] M. Maskey and T. S. Newman. Directional texture for visualization - new technique and application study. In *19th International Conference on Information Visualisation (iV)*, 2015, pp. 1–8. IEEE, 2015.
- [43] Met.3D contributors. Met.3D. <http://met3d.readthedocs.org>, 2017. Accessed 30 July 2017.
- [44] T. Metsalu and J. Vilo. Clustvis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. *Nucleic*

Acids Research, 43(W1):W566, 2015. doi: 10.1093/nar/gkv468

- [45] T. Nocke, H. Schumann, and U. Böhm. Methods for the visualization of clustered climate data. *Computational Statistics*, 19(1):75–94, Feb. 2004.
- [46] M. Novotny and H. Hauser. Outlier-preserving focus+context visualization in parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):893–900, 2006. doi: 10.1109/TVCG.2006.170
- [47] S. Oeltze, D. J. Lehmann, A. Kuhn, G. Janiga, H. Theisel, and B. Preim. Blood flow clustering and applications in virtual stenting of intracranial aneurysms. *IEEE Transactions on Visualization and Computer Graphics*, 20(5):686–701, 2014. doi: 10.1109/tvcg.2013.2297914
- [48] A. T. Pang, C. M. Wittenbrink, and S. K. Lodha. Approaches to uncertainty visualization. *The Visual Computer*, 13(8):370–390, Nov. 1997.
- [49] T. Pfaffelmoser, M. Mihai, and R. Westermann. Visualizing the variability of gradients in uncertain 2d scalar fields. *IEEE Transactions on Visualization and Computer Graphics*, 2013.
- [50] A. Pilhofer, A. Gribov, and A. Unwin. Comparing clusterings using Bertin’s idea. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2506–2515, 2012. doi: 10.1109/TVCG.2012.207
- [51] K. Potter, J. Kniss, R. Riesenfeld, and C. R. Johnson. Visualizing summary statistics and uncertainty. *Computer Graphics Forum*, 29(3):823–832, 2010. doi: 10.1111/j.1467-8659.2009.01677.x
- [52] K. Potter, P. Rosen, and C. Johnson. From quantification to visualization: A taxonomy of uncertainty visualization approaches. In A. Dienstfrey and R. Boisvert, eds., *Uncertainty Quantification in Scientific Computing*, vol. 377 of *IFIP Advances in Information and Communication Technology*, pp. 226–249. Springer Berlin Heidelberg, 2012.
- [53] L.-A. Quandt, J. H. Keller, O. Martius, and S. C. Jones. Forecast variability of the blocking system over Russia in summer 2010 and its impact on surface conditions. *Weather and Forecasting*, 32(1):61–82, 2017. doi: 10.1175/WAF-D-16-0065.1
- [54] W. M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336):846–850, 1971. doi: 10.2307/2284239
- [55] Y. Rath, S. Dambreville, and A. Tannenbaum. Statistical shape analysis using kernel pca. In *Electronic Imaging 2006*, pp. 60641B–60641B. International Society for Optics and Photonics, 2006. doi: 10.1117/12.641417
- [56] M. Rautenhaus, C. M. Grams, A. Schäfler, and R. Westermann. Three-dimensional visualization of ensemble weather forecasts – Part 2: Forecasting warm conveyor belt situations for aircraft-based field campaigns. *Geoscientific Model Development*, 8(7):2355–2377, 2015.
- [57] M. Rautenhaus, M. Kern, A. Schäfler, and R. Westermann. Three-dimensional visualization of ensemble weather forecasts – Part 1: The visualization tool Met.3D (version 1.0). *Geoscientific Model Development*, 8(7):2329–2353, 2015.
- [58] A. Reh, C. Gusembauer, J. Kastner, M. E. Gröller, and C. Heinzl. Mobjects—a novel method for the visualization and interactive exploration of defects in industrial xct data. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2906–2915, 2013. doi: 10.1109/tvcg.2013.177
- [59] D. P. Retchless and C. A. Brewer. Guidance for representing uncertainty on global temperature change maps. *International Journal of Climatology*, 36(3):1143–1159, 2016. doi: 10.1002/joc.4408
- [60] M. R. Rezaee, B. P. Lelieveldt, and J. H. Reiber. A new cluster validity index for the fuzzy c-mean. *Pattern recognition letters*, 19(3):237–246, 1998. doi: 10.1016/S0167-8655(97)00168-2
- [61] M. Schonlau. The clustergram: A graph for visualizing hierarchical and non-hierarchical cluster analyses. *The Stata Journal*, pp. 316–327, 2002.
- [62] A. Strehl and J. Ghosh. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of machine learning research*, 3(Dec):583–617, 2002.
- [63] R. Swinbank, M. Kyouda, P. Buchanan, L. Froude, T. M. Hamill, T. D. Hewson, J. H. Keller, M. Matsueda, J. Methven, F. Pappenberger, M. Scheuerer, H. A. Titley, L. Wilson, and M. Yamaguchi. The TIGGE project and its achievements. *Bulletin of the American Meteorological Society*, 97(1):49–67, Mar. 2015. doi: 10.1175/bams-d-13-00191.1
- [64] D. M. Thomas and V. Natarajan. Multiscale symmetry detection in scalar fields by clustering contours. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2427–2436, 2014. doi: 10.1109/tvcg.2014.2346332
- [65] R. L. Thorndike. Who belongs in the family? *Psychometrika*, 18(4):267–276, 1953.
- [66] R. D. Torn, J. S. Whitaker, P. Pegion, T. M. Hamill, and G. J. Hakim. Diagnosis of the source of gfs medium-range track errors in hurricane sandy (2012). *Monthly Weather Review*, 143(1):132–152, 2015. doi: 10.1175/MWR-D-14-00086.1
- [67] X. Wang, C. Yang, and J. Zhou. Clustering aggregation by probability accumulation. *Pattern Recognition*, 42(5):668–675, 2009. doi: 10.1016/j.patcog.2008.09.013
- [68] R. T. Whitaker, M. Mirzargar, and R. M. Kirby. Contour boxplots: A method for characterizing uncertainty in feature sets from simulation ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2713–2722, Dec 2013. doi: 10.1109/TVCG.2013.143
- [69] L. Wilkinson and M. Friendly. The history of the cluster heat map. *The American Statistician*, 63(2):179–184, 2009. doi: 10.1198/tas.2009.0033
- [70] D. S. Wilks. *Statistical Methods in the Atmospheric Sciences*. Academic Press, 3rd ed., June 2011.
- [71] J. Wu, H. Xiong, and J. Chen. Adapting the right measures for k-means clustering. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 877–886. ACM, 2009. doi: 10.1145/1557019.1557115
- [72] H. Zhou, X. Yuan, H. Qu, W. Cui, and B. Chen. Visual clustering in parallel coordinates. *Computer Graphics Forum*, 27(3):1047–1054, 2008. doi: 10.1111/j.1467-8659.2008.01241.x