Vision, Modeling, and Visualization (2019)
H.-J. Schulz, M. Teschner, and M. Wimmer (Eds.)

Clustering Ensembles of 3D Jet-Stream Core Lines

Michael Kern[1] and Rüdiger Westermann[1]

[1]Technische Universität München, Chair of Computer Graphics and Visualization, Germany

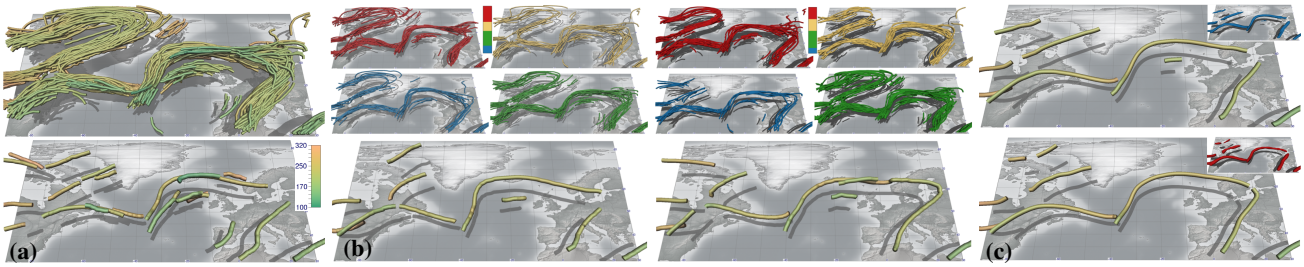

Figure 1: ***(a)*** *Top: Spaghetti plot of jet-stream core lines, extracted from numerical weather prediction data for September 28, 2016. Bottom: Ground truth of the core line on the same day. Line features are colored by pressure level in hPa to infer vertical variation.* ***(b)*** *Top: 4 core line groups obtained from clustering on fields of directional wind speed derivatives (upper left image) and vector-to-closest point volumes (upper right image). Bottom: The best cluster representatives closest to the ground truth.* ***(c)*** *Ridge lines extracted from central tendency volume (inset, upper image) computed from case (a) and ridges from the visitation map (inset, lower image).*

Abstract

The extraction of a jet-stream core line in a wind field results in many disconnected line segments of arbitrary topology. In an ensemble of wind fields, these structures show high variation, coincide only partly, and almost nowhere agree in all ensemble members. In this paper, we shed light on the use of clustering for visualizing an ensemble of jet-stream core lines. Since classical approaches for clustering 3D line sets fail due to the mentioned properties, we analyze different strategies and compare them to each other: We cluster the 3D scalar fields from which jet-stream core lines are extracted. We cluster on a closest-point representation of each set of core lines. These representations are derived from the extracted line geometry and can be used independently of the line orientation and topology. We cluster on the 3D line set using the Hausdorff distance as similarity metric. In the resulting clusters, we visualize core lines from the most representative ensemble member. We further compute ridges in a single 3D visitation map that is build from the ensemble of core lines, and we extract the most central core line from the ensemble closest-point representation. These "averages" are compared to the clustering results, and they are put into relation to ground truth jet-stream core lines at the predicted lead time.

1. Introduction

Nowadays, numerical weather forecasting is routinely performed at weather centers and research institutions. The improvement of weather forecasts depends heavily on documenting and understanding complex three-dimensional structures in the atmosphere. Therefore, these structures need to be defined and extracted from simulated physical fields. Once available, a three-dimensional (3D) visualization of these structures provides a greatly increased capacity to understand the key relationships and dependencies between multiple atmospheric processes. Beyond single forecasts, ensemble weather forecasts are well established in meteorology to understand and assess the uncertainty inherent in numerical weather predictions. Ensemble methods perform multiple simulations using perturbed initial conditions or different forecast models to predict possible future states of the atmosphere. Analysis of the temporal evolution and variability of an ensemble forecast is then used to estimate the likelihood of certain weather events. For many important structures, i.e. line features, in the atmosphere, ensemble analysis becomes notoriously difficult. This is because such structures are often fuzzy and cluttered, consisting of many disconnected segments of arbitrary topology. An example is the jet-stream, regions of high wind speed, typically encountered near to the top of our principal weather systems. The extraction of jet-stream core lines can lead to many separate line features in 3D [KHS*18]. In an ensemble of wind fields, these features show high variation, coincide only partly, and almost nowhere agree in all ensemble members.

To determine and characterize major trends in operational routines, forecasters often actively look for some kind of mean or me-

DOI: 10.2312/vmv.20191321

dian solution for a certain feature in an ensemble, involving clustering features on the 2D chart in their mind if the patterns suggest so to do. We aim to automate this process and, in particular, lift it to 3D. This, however, is challenging for the type of feature we consider. So-called spaghetti plots, i.e., the simultaneous display of all ensemble members (see 1a), quickly produce visual clutter and cannot easily convey major trends, outliers, and statistical properties of the feature distribution. Statistical feature-based approaches, on the other hand, like contour and curve box-plots [WMK13; MWK14] or variability plots [FBW16] cannot be used in general for highly fragmented line sets with arbitrary topology.

In this work, we shed light on the use of clustering to determine the major trends in ensembles of jet-stream core lines with vastly different geometry and topology. We compare the results to the clustering of lines using the mean distance between any pair of core line (one in each cluster) as distance metric. For clustering, we use Agglomerative Hierarchical Clustering (AHC) with Ward's method [Jr63] which groups the initial core lines in bottom-up fashion. To prepare the data for clustering, the used fields are interpreted as high-dimensional data points and projected to lower dimension using either PCA, for derivative fields, or t-SNE, for closest point fields. The findings, in the form of the mean or most representative core line of each cluster, are put into relation to ground truth features obtained from re-analysis data that best represents the atmosphere data at the considered lead time.

In particular, we make the following contributions: First, we compare different clustering approaches and analyze their potential to convey major trends in a weather forecast ensemble to improve the predictability of certain weather situations. Second, we propose two novel representations for clustering meteorological features comprised of line sets with arbitrary topology: We cluster on the derivative fields from which features are extracted, and on a topology-invariant representation of line sets based on a closest point representation. This enables to operate on the extracted jet-stream core lines, regardless of their shape and topology. Third, we compare the results of clustering to those achieved via visitation maps, a density field showing the frequency of occurrence of line features, and via central tendency fields characterizing points most central in between sets of lines. Ridge detection in such volumes is used to extract approximate features which are then used in the comparative analysis. To demonstrate the practical relevance and usefulness of our work, we discuss the strengths and weaknesses of the methods by means of real-world scenarios using numerical weather simulation data from the European Center for Medium Weather Forecasting (ECMWF).

2. Related Work

The analysis of 3D line feature ensembles is closely related to ensemble visualization techniques. Parametric statistical distributions and distribution shape descriptors for scalar-valued ensembles were presented by Love et al. [LPK05]. Different variants of confidence regions were introduced to obtain major geometric trends in ensembles of closed isocontours and streamlines [WMK13; MWK14; FBW16; FKRW16]. Demir et al. [DJW16] proposed a closest-point representation to convey the central tendency of an multi-dimensional shape ensemble.

In a number of works, scalar- and vector-valued ensemble fields were modeled via mixtures of probability density functions to compactly classify complex distributions and their evolution over time [LLBP12; JDKW15; DS15; WLW*17]. Demir et al. [DDW14] propose to linearize 3D data points and visualize the data distributions using bar-charts. Poethkow and Hege [PH13] and Athawale et al. [ASE16] use location-wise estimators of non-parametric distributions from ensemble members to estimate the spread of surface and vector field features. Recently, Hazarika et al. [HBS18; HDSC19] presented a copula-based framework for large multivariate data sets, where they partition the domain and compute statistical quantities over those parts.

Alternatively, clustering has been used to group ensemble members, either fields or curves in such fields, regarding similar data characteristics [BM10; TN14; OLK*14; FBW16; FFST19]. The comparative studies by Zhang et al. [ZHQL16] and Moberts et al. [MVv05] provide overviews of similarity measures using geometric distances between curves. Strehl and Ghosh [SG02] apply different clustering techniques to one single ensemble, and combine the results into a single clustering. Ferstl et el. [FKRW17] cluster different time-steps of the same ensemble in a hierarchical way to convey the change of clusters over time. Clustering of atmospheric data is surveyed by Wilks [Wil11]. An example is the operational clustering of ensemble members at ECMWF [FC11], where forecast scalar fields of geopotential height are clustered in three different time windows for a static data region. Kumpf et al. [KTB*18] use multiple k-Means clusterings on ensemble fields, with slightly varying clustering domain, and analyze the robustness of clusters. Clustering using the Hausdorff metric was applied to group similar streamlines (cf. [RT12]) in flow-fields, and fiber bundles [BBP*05] in diffusion tensor imaging.

Building upon first definitions of jet-streams, e.g., [Rei63], were the "layer of maximum wind" (LMW) was defined to identify the 3D jet-stream axis representing the core line, many different methods have been developed to automatically identify jet-streams. For example, Strong and Davis [SD05] detect core lines on their LMW by computing wind speed maxima via finite differencing in the y-direction only. Molnos et al. [MMP*17] introduced a network-based scheme using shortest paths to detect a jet-stream core as a continuous, globe-circumventing line. Spensberger et al. [SSL17] consider 2D wind fields on a "dynamical tropopause", an isosurface of 2 potential vorticity units.

Aforementioned methods make a priori assumptions about certain jet-stream characteristics, such as being westerly oriented or fully connected large-scale features. Our approach builds upon the detection method by Kern et al. [KHS*18], which avoids such assumptions and detects arbitrarily oriented and possibly disconnected jet-stream core lines.

3. Data

We extract the jet-stream core lines by applying ridge detection to a given wind field, as proposed by Kern et al. [KHS*18]. Numerical weather prediction data, generated from the Numerical Weather Prediction (NWP) model at ECWMF, serve as input for feature detection. Ensemble forecasting is used to compute m predictions,

each comprised of multiple fields such as wind direction and precipitation. For jet-stream core line detection, we use the 2D wind components u, v and wind speed V_s for each vertical layer. From these inputs, we derive new fields from which the features are extracted (see eq. 1). Let $\vec{V} = (u, v)$ be the wind vector at a grid point x, and V_s be the wind speed resolved into the direction $\vec{s}$, the local wind direction at x. Let $\vec{n}$ be the vector normal to $\vec{s}$. The loci (candidates) of jet-stream core lines are defined at each grid point where V_s is maximal along $\vec{n}$ in the horizontal, and along $\vec{z}$ in the vertical:

$$\frac{\partial V_s}{\partial n} = 0, \frac{\partial V_s}{\partial z} = 0 \tag{1}$$

Upon detecting candidate lines via ridge extraction, filtering is applied to consider domain-specific constraints. In particular, domain experts look for jet-stream core lines with at least $V_s \geq 40ms^{-1}$ and large features with a feature length above $500km$. Minima lines of the wind field are removed by testing whether all eigenvectors of the 2D Hessian matrix at each line vertex are negative. Filtered ridge lines are shown in Fig. 1a and Fig. 2.

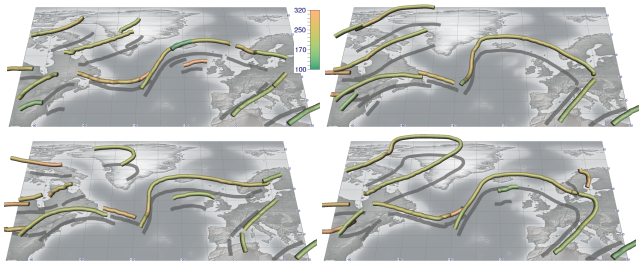

Figure 2: *Extracted jet-stream core lines rendered as 3D tubes for 4 different member of the ensemble in Fig. 1a. Images demonstrate the variation of line topology. Tubes are colored by pressure in hPa to show variation in the vertical.*

4. Clustering of Jet-Stream Core Lines

In this section, different strategies to cluster an ensemble of jet-stream core lines are examined and put into relation regarding the main trends they can convey. We consider clustering on the fields from which the core lines are extracted, and on the extracted line geometry. By using clustering, we overcome in particular the limitations of spaghetti plots (cf. Fig. 1a). In the shown scenario, spaghetti plots suggest long coherent features. By visualizing single members (cf. Fig. 2), however, it can be observed that the ensemble predicts vastly diverging features comprised of many disconnected line segments. We introduce three different strategies to better reveal the major trends in an ensemble of jet-stream core line of arbitrary geometry and topology. Fig. 3 gives an overview of these strategies, their application is shown in Fig. 1b. Furthermore, we compare the extracted line geometries to ridge lines in fields derived from a closest point representation, and ridges in visitation maps derived from the line geometry (cf. Fig. 1c). Both variants can be computed either for each cluster to provide cluster representatives, or for the entire ensemble.

4.1. Clustering of Physical Fields

The first approach operates on the original fields of wind directions and speed, as explained in Sec. 3 for m ensemble members. A naive

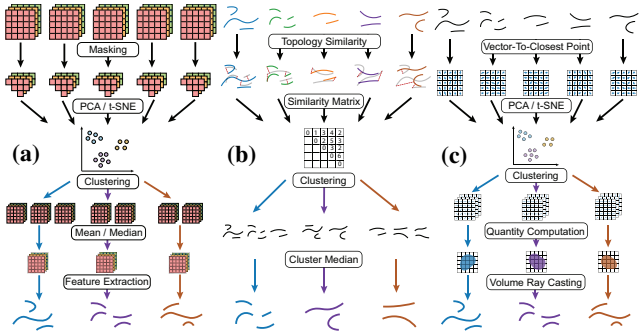

Figure 3: *Clustering jet-stream core lines based on* ***(a)*** *scalar volumes derived from initial fields,* ***(b)*** *geometric distance metrics for lines, and* ***(c)*** *closest point representations derived from line sets.*

approach is to first average all fields over the entire ensemble and perform feature detection in the averaged fields. Feature lines extracted from mean fields of two different scenarios are shown in Fig. 4a and b. For both days, averaging leads to long connected features, while small features vanish due to high variation in the data. In comparison to the ground truth, averaging is not able to convey important trends, such as the long jet-stream core line south of Greenland (Fig. 4b), for regions of high variation. We also discovered that clustering directly on the original scalar fields leads to unsatisfying results due to high variation in the data across vertical levels and ensembles. More specifically, inputs like wind direction and speed over a certain domain do not necessarily correlate locally or globally with the occurrence of line features.

To overcome this problem, we propose the following clustering strategy outlined in Fig. 3a. From the input field, we first derive quantities that describe the characteristics of jet-stream core lines and from which these lines are extracted. In particular, we use directional derivatives of wind (cf. 1) and wind speed as initial scalar fields for clustering. The latter is important as forecasters are only interested in features of high wind speed. Based on these scalar fields, we mask all grid points where the derivatives and wind speed do not approximately match the criteria listed in Sec. 3 for all ensemble members, and eventually remove these points from the initial fields. Next, we concatenate the initial fields and perform dimensionality reduction using Principal Component Analysis (PCA). Here, we only take into account the principal components representing at least 90% of data variance, similar to [FKRW16]. In PCA space, we use clustering to obtain groups of ensemble members with similar set of scalar fields. The fields of each cluster are then either averaged, or features are extracted from the most representative fields per cluster (selected similar to eq. 3).

We propose to choose cluster median over computing the mean for each cluster as the mean leads to less satisfying results with important smaller features being vanished. Fig. 4c,d show outcomes of clustering using the mean across all clusters. These images suggest fully connected line features whereas the ground truth in Fig. 4a,b (inset) reveals trends of interrupted features, for instance, over Florida. Contrary to that, using the scalar field median (cf. Fig. 1b and Fig. 7b) predict interruptions and convey more minor trends, thus, being more similar to the ground truth.

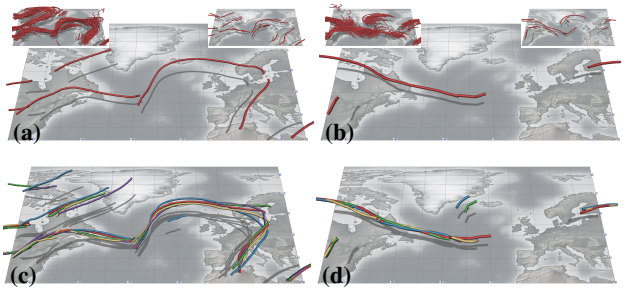

Figure 4: ***(a)*** *Jet-stream core lines extracted from mean input fields of the entire weather ensemble for (a) September 28, 2016, and* ***(b)*** *October 02, 2016, spaghetti plot and the ground truth are shown in the insets.* ***(c)*** *Mean representations for case (a) after averaging clusters obtained from directional derivative fields.* ***(d)*** *Results from clustering similar to (c) for use case (b).*

4.2. Geometry-Based Clustering

Next, we consider the scenario where only the ensemble of extracted jet-stream core lines is given as a set of lines per ensemble member. This requires the computation of feature-to-feature similarities regarding the shape of their line geometries. We introduce two approaches to perform clustering on 3D line feature ensembles.

4.2.1. Geometry Comparison

The first approach is illustrated in Fig. 3b. Given that each ensemble member contains several lines varying in shape and topology, we first measure member-to-member similarity by comparing pairs of lines vertex-wise. We use both the mean closest point distance (MCPD) and the Hausdorff metric (HD) for comparing two sets of lines, similar to [MVv05; BBP*05]. MCPD is the average of the closest point-wise distances while HD computes the maximum of all minimal closest distances between two non-empty subsets. In our experiments, we found that the results using both metrics were quite similar. Yet, since HD produces slightly better results in terms of trend prediction, we use this metric for topology comparison. In the following, we assume that $d_H(A,B)$ represents the Hausdorff distance between two point sets A and B. For each line in one ensemble member, we determine the line with minimum d_H to a line in the other member. For final comparison of two ensemble members, we take the average d_H between two ensemble members for all line-pair-wise minimal HDs. Let E be the ensemble set with ensemble members $i, j \in E$ consisting of line point sets Li, L_j. The member-wise similarity $d_s(L_i, L_j)$ is then defined as:

$$\begin{aligned} d_{He}(L,L_j) &= min_{L' \in L_j}\{d_H(L,L')\} \\ d_s(L_i,L_j) &= mean_{L \in L_i}\{max\{d_{He}(L,L_j), d_{He}(L_j,L)\}\} \end{aligned} \tag{2}$$

Based on d_s, we set-up a distance matrix containing the similarity values for each ensemble member pair used for line clustering.

After clustering, we choose the jet-stream core lines from the ensembles which best represent their clusters. Here, we consider the median of the cluster ensemble members to be the the most representative. An ensemble member serves as the cluster median if its similarity d_s is smallest to all other group members. Let $C \subset E$ be a cluster, then the representative line feature is defined as follows:

$$argmin_{i \in C}\{d_s(L_i,L_j) | j \in C \wedge i \neq j\} \tag{3}$$

Note that with the median we obtain actual line feature predictions and display those as possible outcomes for further analysis. An example of clustering results using HD as similarity metric is shown in Fig. 8a, the best cluster representatives are shown in Fig. 7c.

4.2.2. Implicit Line-Feature Representation

Similarity metrics operating on point-to-point distances are very sensitive to the topology and orientation of given line features, strongly influenced by outliers, especially when comparing several disconnected line features per ensemble member. To overcome this limitation, we make use of a method that first transforms a jet-stream core line of a single ensemble member into a domain-filling field in which this feature is given implicitly. Then, clustering operates on the ensemble of derived fields. Since jet-stream core lines are not closed and inside/outside classification is not possible, signed distance fields cannot be utilized for our purpose. Instead, as proposed by Demir et al. [DJW16], we compute at each grid cell the vectors to the closest points of each jet-stream core line in the ensemble, and cluster the resulting closest-point field (see Fig. 3c).

To transform jet-stream core lines to a closest-point representation, the ensemble domain is first discretized using a Cartesian grid of fixed resolution. Vcp volumes are then generated by using a bounding volume hierarchy for all line segments in a bottom-up fashion. At each grid point, we store m vcps pointing to the closest feature in each member (compare Fig. 6b). Vector magnitudes represent the distance to the closest feature. For clustering, we consider members to be similar at a grid point if all vectors point into the same direction and have a similar magnitude. To improve cluster results, vcp volumes are first preprocessed by filtering out grid points which do not point to any close-by feature, with vcp magnitude exceeding a user-defined minimum length λ with $||vcp_i(\vec{x})|| \geq \lambda$ for all members i. λ is set to half the grid distance. Dimensionality reduction is conducted via t-SNE [MH08] to project the vector volume dimension to 2D for improved cluster separation. Clustering is then performed on the resulting data points in 2D. We use t-SNE with a perplexity value of 5 for $m = 51$ ensemble members to maintain rather local trends with at least 5 close data points. Higher perplexity values force t-SNE to consider more global structures in the data, which we found to produce clusters of higher geometric variability.

To analyze different cluster characteristics, each cluster can be visualized separately, or new quantities can be derived from each cluster using the set of vcp volumes assigned to it. For visualization, we use GPU volume ray-casting to render the isosurface for locations at least half of the grid distance away from any line feature. The minimal distance d_{min} of a vcp set to a line feature is defined as

$$d_{min} = min\{||vcp_i(\vec{x})||, i \in C\} \tag{4}$$

Similar to Demir et al. [DW15], we employ a custom interpolation scheme combining tri-linear interpolation with a plane-based approach for smooth isosurface rendering.

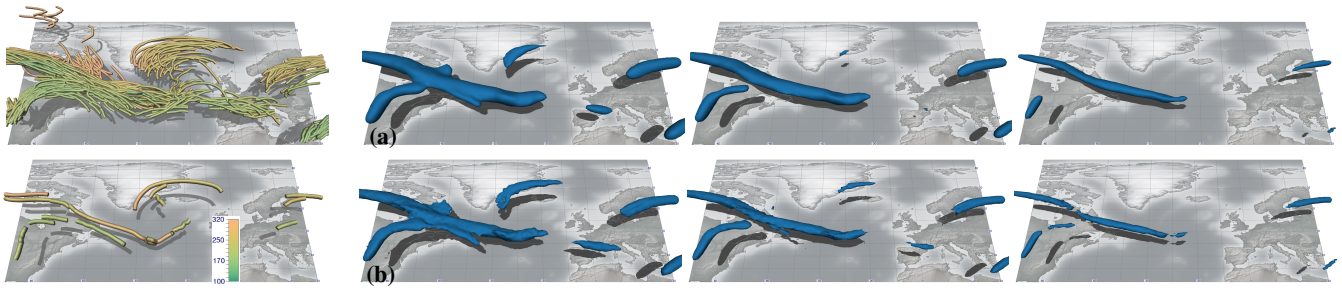

Figure 5: *Left most images: Jet-stream core line ensemble for lead time October 02, 2016 (top) and ground truth core line features (bottom). Top row **(a)**: Visitation map of line features for the entire ensemble with different thresholds for isosurface extraction ($\tau = 0.4/0.5/0.6$). Bottom row **(b)**: Central tendency of line features with different thresholds for isosurface ray-casting ($\sigma = 0.4/0.5/0.6$).*

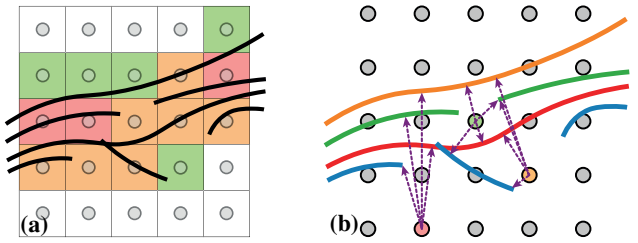

Figure 6: ***(a)** Schematic of a visitation map from a line set (black bold lines). Color represents amount of density related to the number of line crossings (red = 3 crossings, yellow = 2, green = 1). **(b)** Schematic of central tendency computed from closest point vectors (purple dashed arrows). Grid point color indicates amount of centrality (green = highest, orange = medium, red = lowest).*

5. Ridge Line Extraction in Proxy Fields

In addition to clustering, we convey trends in line sets by applying ridge line extraction in fields that are derived from given line geometries, e.g., in vcp fields. Extracted ridge lines represent artificial "mean" representations of the given features, conveying major and minor trends in the feature ensemble. Note that this is contrary to feature detection from averaged scalar fields (of the entire ensemble or single clusters), where minor trends in regions of high variance can vanish (cf. Fig. 4d with best representatives in Fig. 7 and further examined in Sec. 6). In the following, we discuss two different approaches to obtain such fields.

5.1. Visitation Map

Line geometry can be used to analyze the frequency of line feature occurrences within the ensemble domain. This is accomplished by rasterizing line segments into a discrete grid and counting the number of line crossings per voxel cell. Fig. 6a illustrates a visitation map with line feature ensembles of 4 different ensemble members. In particular, we implement visitation maps proposed by Bürger et al. [BFMW12] and used in [FBW16]. Visitation maps make use of a discrete voxel grid and count for each cell the number of line segments crossing the cell. We use rasterization of line vertices to build the visitation maps with each vertex assumed to be a particle – in this context called splat-kernel – of a user-defined diameter. Splat-kernels with support of 4 voxels were chosen for our data and the density values were normalized in a second pass to alleviate thresholding. Next, we use GPU volume ray-casting to obtain isosurfaces with a user-defined density. This reveals regions of high feature density and provides first insights about possible tendencies in the data. Fig. 5b shows a visitation map for 3 different thresholds where major trends of high frequent occlusions are revealed but interruptions are hardly detected.

We also extract ridge lines from visitation maps using classical ridge detection (cf. [PS08]). The ridge lines provide approximate line features which serve as possible predictions for the current ensemble (see Fig. 1c bottom and Fig. 7f,h). Ridge line extraction is restricted by a user-defined minimal density for visitation maps. This threshold has to be set carefully as high values lead to high information loss (see Fig. 5a for threshold comparison). In our work, we found thresholds between 0.4 and 0.6 most suitable for our data. We can compute visitation maps for either the entire ensemble or for each cluster individually. Note that ridge lines are approximations producing new features not contained in the ensemble. In Sec. 6, we compare ridge lines to feature lines obtained from clustering techniques to discuss their potential.

5.2. Central Tendency

We use vcp volumes to also determine the central tendency of the data. Central tendency (centrality) represents locations which are most central in the data set, meaning that the grid points lie in between all closest line features and inherit small magnitudes. An example of three different central tendencies is sketched in Fig. 6b for 4 ensemble members. For centrality, we additionally compute the mean vcp $\vec{\mu}$ at each cell:

$$\vec{\mu} = \overline{vcp}(\vec{p}) = \frac{1}{N}\sum_{i=1}^{N} vcp_i(\vec{p}) \tag{5}$$

Note that we refrain from visualizing the isosurface of $\vec{\mu}$ as this introduces artifacts in rendering, in example for grid points in between two line feature groups. The final central tendency σ of a grid point is computed as follows:

$$\begin{aligned} d_{max} &= max\{||vcp_i(\vec{x})||, i \in C\} \\ \sigma &= max\{1 - d_{max}, \sqrt{f \cdot (1 - ||\vec{\mu}||)}\}, \end{aligned} \tag{6}$$

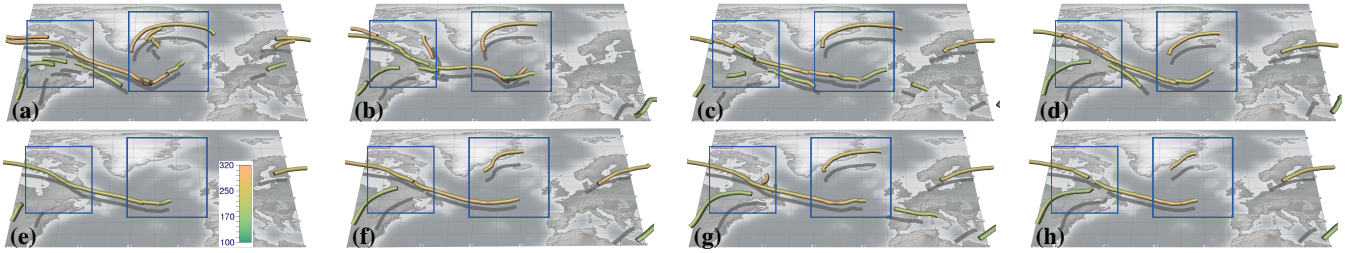

Figure 7: *Best cluster representatives for scenario on October 2, 2016 (lead time) simulated for September 28 (init time) from:* ***(a)*** *ground truth,* ***(b)*** *derivative field clustering,* ***(c)*** *geometry-based clustering,* ***(d)*** *vcp clustering,* ***(e)*** *ensemble mean,* ***(f)*** *ensemble visitation map,* ***(g)*** *ensemble centrality,* ***(h)*** *cluster centrality. Blue rectangles mark regions of varying feature prediction across all ridges.*

where f is the fraction of vcp vectors with a vector magnitude $||vcp_i(\vec{x})|| < \tau$, with τ being half the grid distance. The factor f is important as we want to filter out cases where many vcp of high magnitude cancel out each other, resulting in artifacts with falsely classified points of high centrality. The final rendering of central tendency for the entire ensemble is shown in Fig. 5b with 3 different threshold for σ. From the central tendency volume, we can also extract ridge lines, both for the ensemble and for each cluster. Compared to the visitation maps, extracting ridges from the central tendency differs from ridges in visitation maps in terms of trend inference. In example, features over Florida and Greenland (cf. Fig. 5b) are better revealed than with visitation maps (cf. Fig. 5a) when compared to ground truth. However, central tendency produced a jet-stream core line over Spain which has not been predicted by the ground truth, leading to false interpretations.

6. Results and Use Cases

We discuss the potential of all proposed techniques by means of real-world weather forecasting cases from the 2016 North Atlantic Waveguide and Downstream Impact Experiment (NAWDEX, [SCW*18]), an atmospheric research field campaign. Forecast data are obtained from the Ensemble Prediction System (ENS) at the ECMWF with 51 ensemble members. The data is a regular latitude-longitude grid with grid spacing of 1.0° and 70 terrain-aligned model levels in the vertical. To assess the potential of our methods in terms of trend prediction, we compare the results to jet-stream core lines extracted from re-analysis data obtained from ECMWF using the ERA5 re-analysis model [HB19]. ERA5 data is similar to the NAWDEX data with 0.5° grid spacing and 276 vertical levels.

To enable a proper quality comparison, we selected cases where trends in the ground truth are predicted by a subset in the ensemble forecast. Furthermore, we will examine weather forecasts 72 and 96 hours from initial simulation time because of the following reasons: a) these hours are commonly used by forecaster to create short- and medium-range forecasts, and b) the high variation of features prediction hampers forecasters from inferring clear trends from standard spaghetti plots. We assess the quality of each approach by comparing the median MCPD of best cluster representatives (predictions closest to ground truth) and ridge lines to all line features in ERA5. To compare the average closest distance of lines contained in feature predictions from our approaches to the ground truth lines in ERA5, and to be insensitive to features not present in the reference, we found that MCPD is most suitable here for quality comparison.

Table 1: *Median of mean point-pair distances of best cluster representatives and ridge lines to ground truth for all techniques: mean over entire ensemble (MEAN), scalar field clustering with mean (SFCμ) and median (SFCM), geometry-based clustering (GC), vcp clustering (VC), cluster central tendency (VCC), ensemble central tendency (C), and cluster visitation map (V).*

Use cases	MEAN	SFCμ	SFCM	GC	VC	VCC	C	V
Sep. 28	4.58	3.58	2.61	2.25	2.62	2.66	2.79	3.08
Oct. 02	4.73	2.70	3.13	2.47	2.60	2.73	4.31	3.01

6.1. September 28, 2016

We regard a scenario of highly varying predictions on September 28 from initial time September 25, 2016, partly predicting either a closed curved or interrupted jet-stream core line over Greenland or Florida. Taking the mean over the entire ensemble (MEAN) does not convey information about interrupted features and minor feature trends (cf. Fig. 4a). Clustering the derived scalar fields and obtaining the mean over the clusters for feature detection led to improved results (cf. Fig. 4c), indicated by smaller values for MCPD (cf. Table 1 MEAN and SFCμ); however, those means did not convey any interruptions. We also applied all three clustering techniques to the data and obtained the best representative feature lines for comparison. In terms of cluster separation, vcp ensemble clustering managed to separate the line ensemble best compared to geometry-based scalar field clustering (shown Fig. 1b). All clustering techniques produced representatives which are similar to predictions in the reference, predicted multiple interrupted lines over Florida, and, thus, outperformed MEAN wrt. MCPD.

Next, we assess the quality of ridge lines identified in the visitation map of a cluster obtained from a geometry-based clustering (cf. Fig. 1c bottom), and ridge lines from the central tendency field (see Fig. 1c top). Both methods are able to provide satisfying results, however, the ridge line of the visitation map fails to convey some minor features (in example west of Great Britain), leading to slightly higher MCPD values. Yet, ridge lines are closer to the reference than lines from MEAN, and those from centrality can compete with all best cluster representatives. In Summary, we could improve MEAN wrt. MCPD with all cluster representatives and ridge lines, best results were achieved with cluster representatives (cf. Table 1).

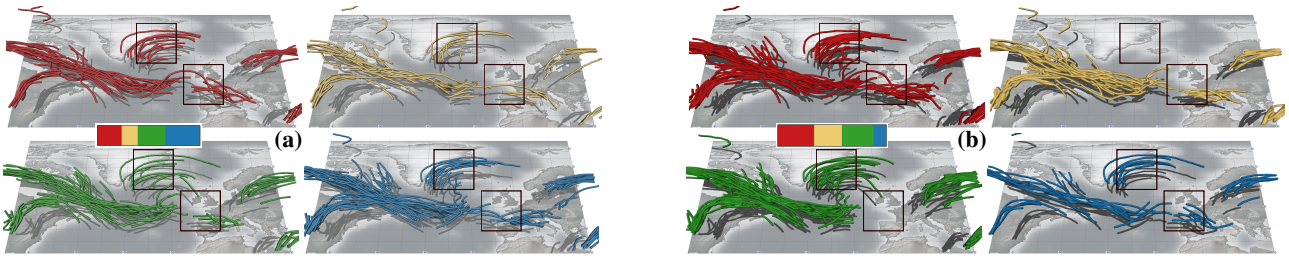

Figure 8: *Clustering results with 4 clusters (each cluster colored)* ***(a)*** *obtained from geometry-based clustering with Hausdorff distance, and* ***(b)*** *from clustering on vector-to-closest point volumes. Inset shows member distribution for all clusters. Black rectangles highlight major differences in trend prediction between both clustering techniques.*

6.2. October 02, 2016

We picked another scenario for October 2 from September 28, 2016 with highly varying predictions, in example feature occurrence south of Greenland, two merging or two distinct jet cores from Florida towards the North Atlantic, and multiple interruptions in between (cf. blue rectangles in Fig. 7). Similar to Sec. 6.1, taking the mean over the entire ensemble led to massive information loss and missed features compared to the reference (see Fig. 4b and d). Best representatives from all clustering techniques are shown in Fig. 7b,c and d. Computing the mean or median from scalar-field-based clustering produced results close to the ground truth, both visually and wrt. MCPD (see. column 2 and 3 in Table 1). All clustering techniques, again, were able to provide good results coinciding with the ground truth. In example, they correctly predicted two distinct features over Florida with multiple interruptions (cf. Fig. 7b,c, and d). In terms of cluster separation, we found that vcp clustering is superior to all other clustering techniques, in example, it excelled line set comparison with HD as shown in Fig. 8. Opposed to clustering, ridge lines in central tendency and visitation map volumes (cf. Fig. 7f,g, and h) could not properly reveal such interruptions in the data. In addition, centrality over the entire ensemble produced a feature over Spain not present in the reference (Fig. 7g), leading to the highest MCPD value compared to all other techniques except the MEAN. All approaches, except centrality, were able to highly outperform MEAN wrt. MCPD where best cluster representatives were closest to the reference.

7. Conclusion

In this work, we have introduced and compared different approaches to cluster 3D line sets of arbitrary orientation and topology for trend inference. In particular, we clustered (a) on derived scalar fields, (b) on line geometry using the Hausdorff metric, and (c) on vector-to-closest point representations invariant to line topology and orientation. We have shown that all clustering approaches have the potential to infer major and minor trends from weather ensemble forecasts close to trends contained in re-analysis data considered to be the ground truth. In addition, we have provided methods to analyze feature occurrence frequency with visitation maps, and we derived central tendency fields from closest-point volumes to classify points most central between line sets. From these fields, we have extracted ridge lines as an artificial mean representation and have demonstrated that they can compete with clustering when applied to cluster results only, yet produced slightly inferior results. In terms of clustering, we can recommend two approaches: If the initial physical fields are given, we propose to operate on derived scalar fields characterizing points close to potential features. In case line geometry is provided only, clustering based on vector-to-closest-point representations is the most effective technique to obtain clear distinct clusters and trends in the data.

Note that the examination of cluster quality for each proposed clustering technique is beyond the scope of this work and requires further investigations in the future. This includes automated selection of best clusters and best cluster representatives by comparing multiple clustering results and incorporating information about additional physical quantities. However, our proposed techniques are not limited to jet-stream core lines and can be applied to arbitrary scalar fields, feature detection methods, and line sets. Note that parameters for t-SNE and ridge detection have to be adapted depending on the number of data elements (ensemble members) and their dimensionality. We have integrated our techniques in the visualization tool "Met.3D" (see [FKRW17; KHS*18; KTB*18]), enabling domain experts to further judge the quality and likelihood of best representatives by plotting additional atmospheric processes along with jet-stream core lines. We are confident that this analysis can help domain experts in estimating the best predictions and trends from vastly diverging line features in the future.

8. Acknowledgements

This research has been done within the subproject A7 of the Transregional Collaborative Research Center SFB/TRR 165 Waves to Weather funded by the German Research Foundation (DFG).

References

[ASE16] ATHAWALE, T., SAKHAEE, E., and ENTEZARI, A. "Isosurface Visualization of Data with Nonparametric Models for Uncertainty". *IEEE Transactions on Visualization and Computer Graphics* 22.1 (Jan. 2016), 777–786 2.

[BBP*05] BLAAS, J., BOTHA, C. P., PETERS, B., VOS, F. M., and POST, F. H. "Fast and reproducible fiber bundle selection in DTI visualization". *VIS 05. IEEE Visualization, 2005.* Oct. 2005, 59–64 2, 4.

[BFMW12] BÜRGER, K., FRAEDRICH, R., MERHOF, D., and WESTERMANN, R. "Instant Visitation Maps for Interactive Visualization of Uncertain Particle Trajectories". *Proceedings Visualization and Data Analysis 2012*. Vol. SPIE 8294. 2012, 82940P-82940P-12 5.

[BM10] BRUCKNER, S. and MÖLLER, T. "Isosurface Similarity Maps". *Computer Graphics Forum* 29.3 (2010), 773–782 2.

[DDW14] DEMIR, I., DICK, C., and WESTERMANN, R. "Multi-Charts for Comparative 3D Ensemble Visualization". *IEEE Transactions on Visualization and Computer Graphics* 20.12 (Dec. 2014), 2694–2703 2.

[DJW16] DEMIR, I., JAREMA, M., and WESTERMANN, R. "Visualizing the Central Tendency of Ensembles of Shapes". *SIGGRAPH Asia 2016 Symposium on Visualization*. SA '16. ACM, 2016 2, 4.

[DS15] DUTTA, S. and SHEN, H.-W. "Distribution driven extraction and tracking of features for time-varying data analysis". *IEEE transactions on visualization and computer graphics* 22.1 (2015), 837–846 2.

[DW15] DEMIR, I. and WESTERMANN, R. "Vector-to-Closest-Point Octree for Surface Ray-Casting". *Vision, Modeling & Visualization*. Ed. by BOMMES, D., RITSCHEL, T., and SCHULTZ, T. The Eurographics Association, 2015 4.

[FBW16] FERSTL, F., BÜRGER, K., and WESTERMANN, R. "Streamline Variability Plots for Characterizing the Uncertainty in Vector Field Ensembles". *IEEE Transactions on Visualization and Computer Graphics* 22.1 (Jan. 2016), 767–776 2, 5.

[FC11] FERRANTI, L. and CORTI, S. "New clustering products". *ECMWF Newsletter* 127 (2011), 6–11 2.

[FFST19] FAVELIER, G., FARAJ, N., SUMMA, B., and TIERNY, J. "Persistence Atlas for Critical Point Variability in Ensembles". *IEEE Transactions on Visualization and Computer Graphics* 25.1 (Jan. 2019), 1152–1162 2.

[FKRW16] FERSTL, F., KANZLER, M., RAUTENHAUS, M., and WESTERMANN, R. "Visual Analysis of Spatial Variability and Global Correlations in Ensembles of Iso-Contours". *Computer Graphics Forum (Proc. EuroVis)* 35.3 (2016), 221–230 2, 3.

[FKRW17] FERSTL, F., KANZLER, M., RAUTENHAUS, M., and WESTERMANN, R. "Time-Hierarchical Clustering and Visualization of Weather Forecast Ensembles". *IEEE Transactions on Visualization and Computer Graphics* 23.1 (Jan. 2017), 831–840 2, 7.

[HB19] HENNERMANN, K. and BERRISFORD, P. *ERA5 data documentation*. Accessed 11 June 2019. 2019 6.

[HBS18] HAZARIKA, S., BISWAS, A., and SHEN, H. "Uncertainty Visualization Using Copula-Based Analysis in Mixed Distribution Models". *IEEE Transactions on Visualization and Computer Graphics* 24.1 (Jan. 2018), 934–943 2.

[HDSC19] HAZARIKA, S., DUTTA, S., SHEN, H., and CHEN, J. "CoDDA: A Flexible Copula-based Distribution Driven Analysis Framework for Large-Scale Multivariate Data". *IEEE Transactions on Visualization and Computer Graphics* 25.1 (Jan. 2019), 1214–1224 2.

[JDKW15] JAREMA, M., DEMIR, I., KEHRER, J., and WESTERMANN, R. "Comparative visual analysis of vector field ensembles". *Visual Analytics Science and Technology (VAST), 2015 IEEE Conference on*. Oct. 2015, 81–88 2.

[Jr63] JR., J. H. W. "Hierarchical Grouping to Optimize an Objective Function". *Journal of the American Statistical Association* 58.301 (1963), 236–244 2.

[KHS*18] KERN, M., HEWSON, T., SADLO, F., WESTERMANN, R., and RAUTENHAUS, M. "Robust Detection and Visualization of Jet-Stream Core Lines in Atmospheric Flow". *IEEE Transactions on Visualization and Computer Graphics* 24.1 (Jan. 2018), 893–902 1, 2, 7.

[KTB*18] KUMPF, A., TOST, B., BAUMGART, M., RIEMER, M., WESTERMANN, R., and RAUTENHAUS, M. "Visualizing Confidence in Cluster-based Ensemble Weather Forecast Analyses". *IEEE Transactions on Visualization and Computer Graphics* 24.1 (Jan. 2018), 109–119 2, 7.

[LLBP12] LIU, S., LEVINE, J. A., BREMER, P.-T., and PASCUCCI, V. "Gaussian mixture model based volume visualization". *IEEE Symposium on Large Data Analysis and Visualization (LDAV)*. IEEE. 2012, 73–77 2.

[LPK05] LOVE, A. L., PANG, A., and KAO, D. L. "Visualizing spatial multivalue data". *IEEE Computer Graphics and Applications* 25.3 (2005), 69–79 2.

[MH08] MAATEN, L. V. D. and HINTON, G. "Visualizing data using t-SNE". *Journal of Machine Learning Research* 9.Nov (2008), 2579–2605 4.

[MMP*17] MOLNOS, S., MAMDOUH, T., PETRI, S., NOCKE, T., WEINKAUF, T., and COUMOU, D. "A network-based detection scheme for the jet stream core". *Earth System Dynamics* 8.1 (2017), 75–89 2.

[MVv05] MOBERTS, B., VILANOVA, A., and VAN WIJK, J. J. "Evaluation of fiber clustering methods for diffusion tensor imaging". *VIS 05. IEEE Visualization, 2005.* Oct. 2005, 65–72 2, 4.

[MWK14] MIRZARGAR, M., WHITAKER, R. T., and KIRBY, R. M. "Curve boxplot: Generalization of boxplot for ensembles of curves". *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014), 2654–2663 2.

[OLK*14] OELTZE, S., LEHMANN, D. J., KUHN, A., JANIGA, G., THEISEL, H., and PREIM, B. "Blood Flow Clustering and Applications inVirtual Stenting of Intracranial Aneurysms". *IEEE Transactions on Visualization and Computer Graphics* 20.5 (May 2014), 686–701 2.

[PH13] PÖTHKOW, K. and HEGE, H.-C. "Nonparametric Models for Uncertainty Visualization". *Computer Graphics Forum* 32.3 (2013), 131–140 2.

[PS08] PEIKERT, R. and SADLO, F. "Height Ridge Computation and Filtering for Visualization". *2008 IEEE Pacific Visualization Symposium*. Mar. 2008, 119–126 5.

[Rei63] REITER, E. R. *Jet-stream meteorology*. 1st. University of Chicago Press, 1963 2.

[RT12] ROSSL, C. and THEISEL, H. "Streamline Embedding for 3D Vector Field Exploration". *IEEE Transactions on Visualization and Computer Graphics* 18.3 (Mar. 2012), 407–420 2.

[SCW*18] SCHÄFLER, A., CRAIG, G., WERNLI, H., et al. "The North Atlantic Waveguide and Downstream Impact Experiment". *Bulletin of the American Meteorological Society* 99.8 (2018), 1607–1637 6.

[SD05] STRONG, C. and DAVIS, R. E. "The surface of maximum wind as an alternative to the isobaric surface for wind climatology". *Geophys. Res. Lett.* 32.4 (2005), L04813+ 2.

[SG02] STREHL, A. and GHOSH, J. "Cluster ensembles—a knowledge reuse framework for combining multiple partitions". *Journal of machine learning research* 3.Dec (2002), 583–617 2.

[SSL17] SPENSBERGER, C., SPENGER, T., and LI, C. "Upper Tropospheric Jet Axis Detection and Application to the Boreal Winter 2013/14". *Mon. Wea. Rev.* (2017) 2.

[TN14] THOMAS, D. and NATARAJAN, V. "Multiscale Symmetry Detection in Scalar Fields by Clustering Contours". *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014), 2427–2436 2.

[Wil11] WILKS, D. S. *Statistical Methods in the Atmospheric Sciences*. 3rd. Academic Press, June 2011 2.

[WLW*17] WANG, K.-C., LU, K., WEI, T.-H., SHAREEF, N., and SHEN, H.-W. "Statistical visualization and analysis of large data using a value-based spatial distribution". *2017 IEEE Pacific Visualization Symposium (PacificVis)*. IEEE. 2017, 161–170 2.

[WMK13] WHITAKER, R. T., MIRZARGAR, M., and KIRBY, R. M. "Contour Boxplots: A Method for Characterizing Uncertainty in Feature Sets from Simulation Ensembles". *IEEE Transactions on Visualization and Computer Graphics* 19.12 (Dec. 2013), 2713–2722 2.

[ZHQL16] ZHANG, H., HOU, Y., QU, D., and LIU, Q. "Correlation Visualization of Time-Varying Patterns for Multi-Variable Data". *IEEE Access* 4 (2016), 4669–4677 2.