A. Kumpf, M. Rautenhaus, M. Riemer, R. Westermann
Visual Analysis of the Temporal Evolution of Ensemble Forecast Sensitivities
in IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 1, Jan. 2019.
doi: 10.1109/TVCG.2018.2864901
URL: https://ieeexplore.ieee.org/Xplore/home.jsp

Visual Analysis of the Temporal Evolution of Ensemble Forecast Sensitivities

Alexander Kumpf, Marc Rautenhaus, Michael Riemer, and Rüdiger Westermann

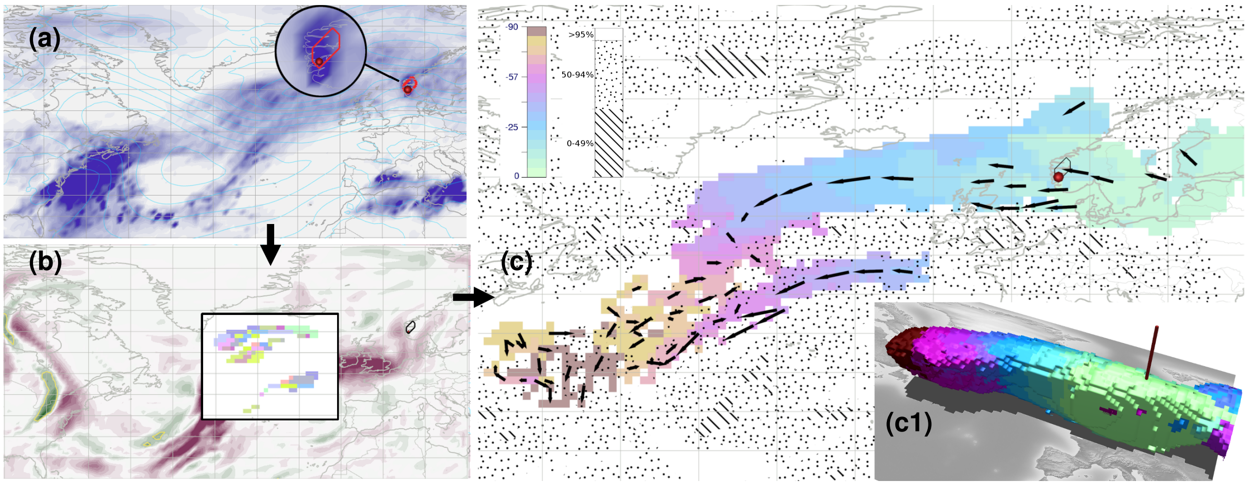

Fig. 1: Analysis of the ensemble sensitivity of forecast variable precipitation (top left) to moisture flux (bottom left) over Norway. **(a)** A high precipitation event is picked and a stochastically coherent clique is computed. **(b)** Sensitivity regions are extracted. Color shows sub-regions with high mutual correlations. **(c)** For a time sequence, sensitivity regions are matched between time steps and tracked over time. A "swipe-path" colors (time in h) all locations covered by a selected sensitivity region over time according to the first time of coverage. Stippling covers statistically insignificant regions. **(c1)** The proposed workflow operates on 2D and 3D data.

Abstract—Ensemble sensitivity analysis (ESA) has been established in the atmospheric sciences as a correlation-based approach to determine the sensitivity of a scalar forecast quantity computed by a numerical weather prediction model to changes in another model variable at a different model state. Its applications include determining the origin of forecast errors and placing targeted observations to improve future forecasts. We—a team of visualization scientists and meteorologists—present a visual analysis framework to improve upon current practice of ESA. We support the user in selecting regions to compute a meaningful target forecast quantity by embedding correlation-based grid-point clustering to obtain statistically coherent regions. The evolution of sensitivity features computed via ESA are then traced through time, by integrating a quantitative measure of feature matching into optical-flow-based feature assignment, and displayed by means of a swipe-path showing the geo-spatial evolution of the sensitivities. Visualization of the internal correlation structure of computed features guides the user towards those features robustly predicting a certain weather event. We demonstrate the use of our method by application to real-world 2D and 3D cases that occurred during the 2016 NAWDEX field campaign, showing the interactive generation of hypothesis chains to explore how atmospheric processes sensitive to each other are interrelated.

Index Terms—Correlation, clustering, tracking, ensemble visualization.

1 Introduction

Ensemble sensitivity analysis (ESA) has been established in meteorology as an ensemble-based approach to estimate the sensitivity of a scalar forecast quantity J (in meteorology referred to as "forecast metric") computed from numerical weather prediction (NWP) model output to changes in the same or another model variable at an earlier state. ESA has been introduced by Ancell and Hakim [1] and Torn and Hakim [42]. It is a correlation-based approach which considers the variations of physical quantities at different domain locations *relative* to each other.

- *Alexander Kumpf, Marc Rautenhaus*, Rüdiger Westermann are with the Computer Graphics & Visualization Group, Technische Universität München, Garching, Germany. *M. R. is now with Universität Hamburg, Regional Computing Center, Hamburg, Germany. E-mail: {alexander.kumpf, marc.rautenhaus, westermann}@tum.de*
- *Michael Riemer is with the Institute of Atmospheric Physics, Johannes Gutenberg Universität Mainz, Mainz, Germany. Email: mriemer@uni-mainz.de*

ESA has been used in a number of applications, including identification of the origins of forecast errors (e.g., [23]), investigation about relations between specific atmospheric processes to other processes of interest in an NWP model (e.g., [43]), and planning of "targeted observations" aiming at improving future forecasts by increasing observation density in critical regions of the atmosphere (e.g., [24]). The use of ESA, however, is associated with a number of challenges.

One major challenge (C1) is how to assess the confidence that can be put into the computed sensitivities. The metric J is commonly obtained by manually selecting a spatial region in which a certain weather event has been identified, and by using a single representative measure for the entire region (e.g., mean or root-mean-squared error of corresponding meteorological parameter) to compute J for every ensemble member. This approach can become a major source of uncertainty, since a single measure does not consider the distribution of values over a region and, thus, might not sufficiently capture the event of interest in all members.

Furthermore, the stochastic stability of the event in the selected region, i.e., the likelihood of variations in the predictions relative to each other, cannot be inferred from a single measure. Thus, the chance of occurrence of the selected event might be very low.

Once a metric J has been defined, ESA computes the correlation between J (the target parameter) and the ensemble data of another model variable (the source parameter) at every point in the domain (cf. [42]). In the resulting correlation field sensitive regions can be identified. Then, another challenge (C2) is to infer the geographical origin of forecast errors from the temporal evolution of theses sensitivities (e.g., [23,47]). Since the manual tracking of sensitive regions (e.g., [47]) becomes tedious especially if many parameters are of interest, addressing C2 entails the development of automatic approaches for tracking sensitive regions in a reliable way. Furthermore, it needs to be analyzed whether determined regions exhibit sufficient stochastic stability (C3) regarding the variation of model parameters across these regions. If this is not the case, it is very likely that the corresponding atmospheric structure will fall apart over time, so that conclusions drawn from ESA have to be carefully evaluated.

Another challenge (C4) put forward by our collaborators from meteorology was the application of ESA to analyze sensitivities in three-dimensional space. In existing meteorological workflows ESA is solely applied to analyze the sensitivity between field variables on two-dimensional atmospheric levels (e.g., [23,43]). To the best of our knowledge, there is no study that applied ESA in full three-dimensional space, even though such an extension would be very beneficial to investigate the interrelations of atmospheric processes as well as the structure of related weather events in the inherently three-dimensional atmosphere.

1.1 Contribution

We introduce a novel visual analysis workflow for addressing C1 to C4; an overview of this workflow is provided in Sect. 3 and Fig. 2. We demonstrate its application by analyzing weather systems observed during the 2016 North Atlantic Waveguide and Downstream Impact Experiment (NAWDEX, [37]), an atmospheric research field campaign in which two of the authors were involved. The NAWDEX research on predictability of weather investigates how different physical processes can play different roles in cyclone evolution; our presented ESA approach aims at facilitating insights into sensitivities of the weather evolution related to Tropical Cyclone "Karl" in September 2016.

To address C1, we provide options to determine sub-regions with low stochastic variation in the region in which the selected weather event occurs. Therefore, the user can pick a location at which the event has a significant occurrence, and via correlation clustering—adapted from Pfaffelmoser and Westermann [27]—the set of locations (including the picked one) with mutual correlations above a threshold is determined. A region comprised of such locations is said to be coherent with respect to the predicted event, and we call such a region a *correlation clique*. By repeating this process, multiple coherent regions can be determined, and the most representative one can be selected and used to restrict the computation of J to its locations.

We address C2 by letting the user, in the source parameter field, interactively select connected components of superlevel sets, i.e., the set of locations at which a threshold in the normalized sensitivity field is exceeded, and track these components over time. Furthermore, and to address C3, we provide an indication of the robustness of the ESA result by visualizing correlation clusters in a selected superlevel set. Computed clusters indicate low or high stochastic variation of model parameter values across the selected region, letting the user infer the structural coherence in the corresponding weather event.

To track a selected feature, i.e., a superlevel set, we use an area-weighted optical flow based approach. To match features in two successive time steps, for all locations covered by a feature in the current time step, we vote using an uncertainty-aware correspondence measure whether these locations again belong to a feature in the next time step.

We account for uncertainties in a feature's boundary by assigning target locations close to a boundary, yet outside a feature, to the closest location inside that feature so that correspondence is established. Additionally, since the features we consider can vary significantly in size over time, the optical flow is used in forward and backward direction in combination with correspondence statistics to obtain a measure that better takes into account the size of matched features.

To effectively convey sensitivities between events at different atmospheric levels as well as the temporal changes these events undergo (C4), we have designed a system comprised of two linked views: An abstract view displaying a graph structure showing split and merge events, and a single track visualization using a so called swipe-path in multi-dimensional space. These visualizations enable the user to shed light on the robustness of tracked events over time as well as their temporal changes in shape, size, and location at different atmospheric levels.

2 Related work

ESA has been introduced by Ancell and Hakim [1] and Torn and Hakim [42]. It has been used in various meteorological publications with a prominent use case being the estimation of the impact of additional observations on forecast accuracy [42]. Although alternative sensitivity analysis techniques are available and used in meteorological applications (an overview is provided e.g., by [24]), due to the widespread use of ESA, and in particular its use by our meteorological partners, we focus exclusively on this method in the current work.

With respect to visualization research, our approach is related to techniques for ensemble visualization, and in particular correlation visualization and region-based tracking in time-varying multi-dimensional scalar fields. A thorough overview of ensemble visualization techniques in meteorology can be found in the recent survey article by Rautenhaus et al. [29].

Finding a suitable visual representation of non-local statistical quantities such as correlation structures is one of the most challenging tasks in visualization. Visualizing the correlation matrix directly is unfeasible for large datasets and cannot show spatial relationships. Global correlation structures were determined by Bansal et al. [3] via correlation clustering, which groups objects based on pair-wise similarities (positive correlation) and dissimilarities (negative correlation) using graph partitioning. Liebmann et al. [21] clustered correlations based on distances on hyperspheres. Paffelmoser and Westermann [27] introduced correlation-based region growing to determine clusters in which the degree of dependency between the data at the cluster centroid and the cluster locations does not fall below a threshold. In contrast, in our approach we search for clusters of locations for which the *mutual* data correlations are above a threshold, to ensure that the stochastic stability of the entire cluster is high.

Paffelmoser and Westermann [28] derived a model to represent local anisotropic correlation structures and used this model to distinguish between correlations along and orthogonal to isosurfaces in 3D scalar fields. By using this approach they analyzed the possible variations of isosurface structures in uncertain scalar fields. Other approaches restricted the analysis to the correlations between the data values at the *same* location in different datasets [16,36], or they analyzed the data variations at the same locations over time, often via variants of time-activity curves [9,40]. Chen et al. [6] used a sampling-based approach to summarize temporal correlations between voxels in multi-variable and time-varying datasets with 3D spatial references. Recently, Zhang et al. [46] developed a temporal multi-variable structure that can express temporal information at a location in multi-dimensional space. This was combined with a dissimilarity-preserving cluster algorithm that characterizes time-varying patterns and spatial locations.

In visualization, a number of techniques have been developed to track regions over time, i.e., to establish the correspondence between regions from one time step to the next. In its simplest form, such methods track the connected components of regions where the data values are entirely below or above a given threshold [35,38], yet also more sophisticated global feature analysis based on scalar field topology and statistics has been used [2,5,7,34,39,45].

The correspondence between regions can either be found using overlap calculations [35,38] or by matching attributes that describe specific properties of each region [31,33,35]. When data with high temporal resolution is given, region overlap methods require only a matching function which measures the degree of similarity of regions in different time steps. In early work this was done using area/volume of the overlap region, yet more advanced matching functions using

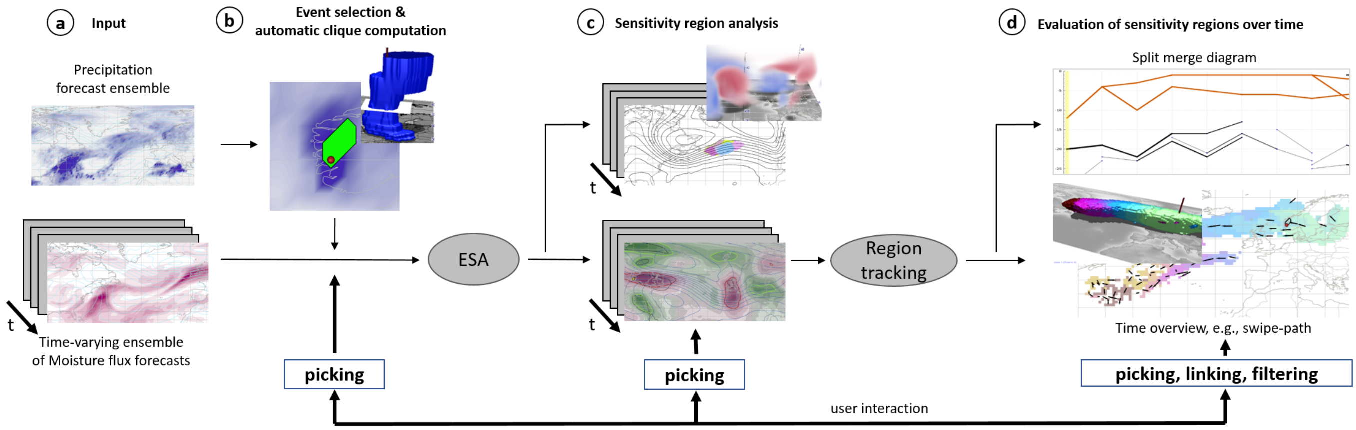

Fig. 2: Method overview: (a) Input are two ensembles of 2D (3D) scalar fields, one of them resolved over time. (b) In the input field of parameter J, a point of interest is picked. A stochastically coherent clique around this point is computed. ESA is then used to compute sensitivity values at every domain point to this clique. (c) Sensitivity fields are displayed, and selected regions can be further analyzed by means of correlation clustering. Selected regions are tracked over time, and (d) different visualizations like split-merge diagrams and swipe-paths are computed and used to analyze sensitivity tracks.

the Earth Mover distance [17] or the Jaccard distance as isosurface similarity metric have been employed recently [12, 34]. However, since these metrics do not consider the size of the overlap regions relative to the features' spatial extent, the matching result can be misleading; for instance, if a large feature splits into multiple very small features.

When data with low temporal resolution is given and features vary strongly in shape and position over time, as it commonly occurs in our meteorological use cases, overlap-based approaches face an additional problem: Before analyzing the overlap, the direction into which the feature has moved needs to be predicted. Muelder and Ma [26] addressed this problem by predicting a feature's next location from previous locations using trajectory extrapolation, and then computing the overlap between the shifted and a target feature. Doraiswamy et al. [7] and Valsangkar et al. [2] used the optical flow [15, 22] between brightness temperature and pressure fields, respectively, to obtain an initial guess of the location of a cloud or cyclone feature in the next time step. Two features are matched if there is at least one hit between these features, i.e., a location in the region of the source feature is connected to a location in the region of the target feature.

In meteorological and climate research tracking is often reduced to single points like extremum points in pressure or precipitation fields [14], and these points are then connected together across time to form tracks. A popular method to find corresponding points in successive time steps is the first-guess method [44], which predicts the location of a selected point in the next time steps by linear continuation and then searches for the closest feature point in the next time step. Gambheer and Bhat [13] tracked clouds by considering the overlap of closed feature contours across time steps. More recently, Fiolleau and Roca [11] considered the time-varying two-dimensional input as a three-dimensional volume, and tracked clouds by tracking seed points within this volume.

3 Method Overview

Given an ensemble of forecast runs, our proposed visual analysis workflow enables meteorologists to interactively analyze the sensitivity of a selected model parameter, e.g., precipitation, to another one, e.g., moisture flux. Fig. 2 shows an overview of the proposed workflow. We consider scalar field data from the ECMWF Ensemble Prediction System (ENS; e.g., [20]). 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 [41] 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 dependent parameter for which the sensitivity shall be analyzed. Following common practice in meteorology, we show a 2D map of ensemble values, e.g., the point-wise mean or maximum values (Fig. 2a). In workflows from past meteorological publications (e.g., [24, 43]), users select a target region enclosing a significant weather event for which ESA is performed. The parameter values in the selected region are then used to compute the single value J per ensemble member, e.g., a mean value or a root-mean-squared error. Then, at every domain location the sensitivity of these values to the ensemble values of the independent parameter at these locations are computed.

It is here were we place our first improvement of the workflow: We let the user determine a representative region in which an event is coherently predicted, i.e., a correlation clique (cf. Sec. 4.1). Therefore, the user selects a seed location, and the system computes instantly a correlation clique to this location (Fig. 2b, Fig. 3). The size of this clique and its strength, i.e., the strength of the mutual dependencies between the data values at the locations covered by the clique, indicate the expected reliability of the ESA results (C1). The user can interactively select different seed locations to find the largest or strongest clique, or abort the process if no such clique is found. For computing the metric J, the parameter values are then condensed only over the selected clique.

The sensitivity values computed by ESA (cf. Sec. 4.1) are plotted in a separate map (Fig. 2c). To highlight regions which can be deemed a significant influence on the target region, the superlevel sets with respect to a sensitivity threshold—selected based on a significance test—are visualized. To address C2, the user selects a region, and the system automatically tracks this region over time by matching against regions in successive time steps (Fig. 2d). Prior to tracking, however, our second improvement comes into play to address C3. By means of correlation clustering it is indicated whether the parameters in the selected superlevel set show low or high stochastic variation (Fig. 2d). If, for instance, two large clusters with inverse correlation to each other exist, the corresponding weather event might be unstable and, thus, rather unsuited for a temporal evaluation.

The tracking process (cf. Sec. 4.3) is guided by the optical flow to predict the percentage of area of one region that overlaps with another region in the next time step. By using the optical flow forward and backward in time, we detect split and merge events, which are encoded in an abstract split-and-merge tree (Fig. 2d) that is linked to the spatial view. This allows for an interactive inspection of detected events, by picking objects in the tree and visualizing them in their spatial surrounding. For a selected region we compute a so-called swipe-path (Fig. 2d), to reveal the spatial and geometric variations a region undergoes over time. A swipe-path shows in one image the temporal evolution of a region, including all split and merge events by coloring and blending all single time states together (Sec. 4.3.2).

To address C4, all components of our proposed visual analysis workflow can be performed optionally using 2D or 3D data (see inlets in Fig 2). To ease navigation, region selection is always performed on 2D maps, and in a synchronized view the 3D region is shown simultaneously. When 3D data is used, direct volume and isosurface rendering via GPU raycasting is used to visualize extracted regions and show their tracks over time.

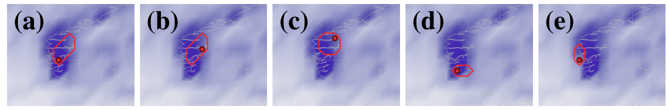

Fig. 3: 24 hours ensemble maximum precipitation over Norway, computed between 27 - 28 September 2016, 18:00 UTC. Comparison of correlation cliques computed from different seed locations.

4 Ensemble sensitivity analysis

We introduce details of our proposed interactive visual analysis workflow for using ESA, and we demonstrate its outcome and derived hypotheses on a real-world dataset. We elaborate specifically on the additional indications and suggestions that our workflow offers compared to the common use of ESA in meteorology.

4.1 Statistically coherent input regions

The ESA approach diagnoses statistically the sensitivity of a selected forecast quantity (the input metric) for a target location to another quantity at other locations and prior times (the initial condition state variable). ESA determines these statistical relations using correlation measures. In particular, the sensitivity $(\partial J/\partial s)$ of the target forecast metric J to a selected state variable s is computed at all points in the model domain. This sensitivity relationship can be expressed as

$$\frac{\partial J}{\partial s} = \frac{\mathrm{cov}(J,s)}{\sigma(s)}, \quad \text{with} \quad \mathrm{cov}(J,s) = \frac{1}{n-1}\sum_{k=1}^{n}(J_k - \bar{J})(s_k - \bar{s}).$$

Here J and s are n-dimensional data vectors, with n being the ensemble size, and the covariance (cov) and standard deviation (σ) are computed over all ensemble members. Thus, the sensitivity measure is the Pearson correlation coefficient multiplied by the standard deviation of the state variable $\sigma(s)$ at a certain location.

As the data values in regions exhibiting very low correlation can be assumed independent of each other, the effect of uncertainty on a feature in such a region is to a large extent arbitrary. Since $\sigma(s)$ can be seen as a measure of the ensemble uncertainty in the state variable at a certain location, points with both high correlation and high uncertainty are emphasized, causing weaker structures to be filtered out. The underlying rational is that changes in terms of standard deviations to grid points with small uncertainty will influence the outcome of a weather prediction system only slightly.

Using the ESA measure, so called hypothesis chains between physical events can be generated to identify physical processes that potentially cause uncertainty in the forecast. The combination of meteorological fields for the input metric and state variable (e.g., precipitation and moisture flux) generates the ensemble that is analyzed to quantify the contribution to uncertainty induced by the state variable with respect to the natural variability estimated from the meteorological models. Critical to the sensitivity analysis is the selection of a suitable input region over which the metric J is considered.

Using the input metric ensemble values at a single location and computing sensitivities of these values to the state variable ensemble values at other locations is highly sensitive to the chosen location. Hence, in practice a region that contains a significant weather event, i.e., a meteorological process of interest like heavy precipitation, is selected, and for each ensemble member a single measure for the entire region is computed and used as input metric J. Also more sophisticated selection procedures exist, for instance, which split manually the selected region into coherent sub-regions with respect to orography or texture, yet all these approaches require profound meteorological knowledge to select a meaningful set of input locations. Furthermore, as indicated by Fig. 4, computed sensitivities are highly sensitive to the selected input region.

We determine automatically a statistically coherent set of locations over which the input metric is computed, by computing *correlation cliques*. A correlation clique is defined as the set of locations (including a user-selected seed location) with pair-wise correlations between the data values at these locations above a threshold. A clique can consist of multiple connected components. The input metric J is then computed ensemble-member-wise over these locations, and the sensitivity is computed to the resulting ensemble of values of J. In this way, only locations that show a statistically similar data distribution and that are likely to deviate into the same direction in data space are considered. Averaging over regions with completely different statistical characteristics is avoided. Fig. 3 shows correlation cliques for 5 different seed locations that were selected in the depicted precipitation field. It can be seen how strongly the cliques can vary, and how few locations of the high precipitation event they can cover.

Figure 4 shows sensitivities computed by ESA for J computed over two manually selected rectangular regions and a correlation clique. Since the orography over Norway is vastly different for every grid-point, it is difficult to select a rectangular region covering a coherent region. The use of an automatically adapted correlation clique can alleviate this problem.

4.1.1 Computation of correlation cliques

To compute a correlation clique, the user first selects a domain location (seed location) at which a significant weather event is predicted, e.g., a location with high precipitation. Given this location, the clique is computed via Algorithm 1.

input : X, $\hat{x}$, α
$x_c = \hat{x}$; $\tilde{X} = X$; $i = 0$;
sort $\tilde{X}$ w.r.t. decreasing $\mathrm{corr}(s(x_j), s(x_c))$, $x_j \in \tilde{X}$;
while $i < |\tilde{X}|$ **do**
    $\tilde{X} = \tilde{X} \setminus \{x_j | \mathrm{corr}(s(x_c), s(x_j)) < \alpha,\ x_j \in \tilde{X}\}$;
    $x_c = x_{i+1}$;
    $i = i + 1$;
end
return $\tilde{X}$;

Algorithm 1: Our algorithm for computing a correlation clique, with X containing all locations (i.e., grid points) in the domain, $\hat{x}$ the seed location, x_i the i^{th} element in $\tilde{X}$, $s(x_i)$ the value vector at location x_i, and α the selected correlation threshold.

To perform the computation efficiently, we first determine all locations with a correlation to the seed location that is greater than or equal to the threshold, sort these locations with respect to decreasing correlation to the seed location, store the sorted sequence, and proceed as follows: As long as there are locations in the sequence, we select the one with highest correlation to the seed location, and remove all locations from the sequence with a correlation to the selected one below the selected threshold. This procedure is successively applied until all locations have been processed, and the remaining locations belong to the computed correlation clique. The correlation threshold has a significant effect on the size of the cliques; the lower (higher) it is, the more (less) locations will be assigned to a clique. From a number of experiments we have found empirically that a threshold around 0.7 usually leads to plausible cliques. In particular, we have observed that slight variations around 0.7 often do not cause any significant changes, and that resulting cliques were considered representative for the considered event by our partners from meteorology.

Upon computing a correlation clique, it is displayed as an isocontour in the scalar field of input metric J. Locations, i.e., grid points, belonging to the clique are set to one, while all others are set to zero, and the isocontours are extracted from the resulting binary field to the

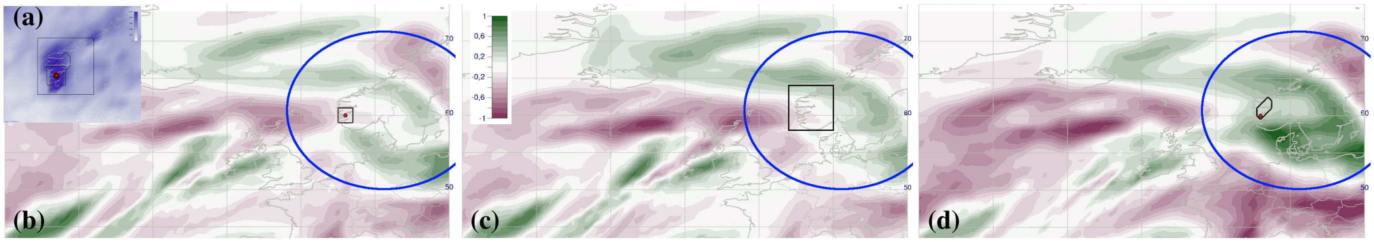

Fig. 4: **(a)** As in Fig. 3, together with the different input regions (framed in black) that were used to compute the ESA metric J. **(b)**-**(d)** Sensitivities are computed with respect to moisture flux at 18:00 UTC 28 September 2016, corresponding to the time right after a high precipitation event over Norway. Hence, high sensitivities are expected mostly east (downstream, blue circle) of the high precipitation region. Rectangular input regions **(b)**, **(c)** show weaker signals than the region that was automatically determined via a correlation clique **(d)**, especially east of the event.

threshold 0.5. The same procedure is applied in 3D, yet the cliques are displayed as isosurfaces in the 3D domain.

4.2 Coherent sensitivity regions

With a selected correlation clique, at every domain point ESA is used to compute the sensitivity between the state variable s at that point and the input metric J. Of special interest are regions of high sensitivity, i.e., superlevel sets to a selected sensitivity threshold, displayed by their enclosing level set. These regions are considered to be coherent with respect to sensitivity, and they are used to indicate where the prediction of J can be improved, i.e., by assimilating additional data from measurements. However, it can be conjectured that measurements placed in a region with high statistical variance will not be able to improve the prediction. Instead, a separate measurement for each statistically coherent region should be preferred.

To further shed light on the statistical variance within a selected region of high sensitivity, one possible solution is to show the correlation matrix for all locations belonging to that region. In general, however, this is unfeasible because the correlation matrix is too large and structures between the values at different locations cannot easily be identified from it. To overcome this limitation, we utilize the algorithm for computing correlation cliques to partition the region into multiple disjoint cliques, with all positive correlations between the data values at locations in one clique.

4.2.1 Correlation clustering

Algorithm 2 describes the procedure for computing a set of correlation cliques that densely cover a selected superlevel set. Our goal is to determine regions with low internal yet high mutual stochastic variation. Multiple cliques with low mutual stochastic variation indicate a rather unstable weather event that may fall apart over time, hinting towards regions where the meteorological models tend to produce different predictions. Especially sub-regions with very low or even negative correlations to the initial clique should be treated separately in subsequent analyses.

The algorithm first computes a correlation clique in the field of state variable s to the seed location with highest sensitivity to J. Next, a second clique is computed, that also has high correlations between the data values at assigned locations, but with low correlation to the first clique. Therefore, we compute the member-wise means of the state variable over the initial clique and select from the remaining locations the one with the minimum correlation to the mean values as seed point for the second clique. This process is then repeated with the second clique as initial clique until all locations in the superlevel set have been assigned.

The resulting partition shows the size and location of cliques with strongly correlated data values, yet with positive and negative mutual correlations. For different scenarios this is demonstrated in Fig. 5. Even though inverse correlations between cliques within one single region with high sensitivity seem surprising at first, such structures can nevertheless occur because the Pearson's correlation coefficient underlying ESA is not transitive.

input : X, $\hat{x}$, α
$\tilde{X} = X$; $i = 0$;
$\text{clique}_0 = \text{correlation\_clique}(\tilde{X}, \hat{x}, \alpha)$;
$\tilde{X} = \tilde{X} \setminus \{x | x \in \text{clique}_0\}$;
while $(\tilde{X} \,!= \emptyset)$ **do**
  $s_c = \text{compute\_metric}(\text{clique}_i)$;
  $x_i = \arg\min_{x \in \tilde{X}} (\text{corr}(s(x), s_c))$;
  $\text{clique}_{i+1} = \text{correlation\_clique}(\tilde{X}, x_i, \alpha)$;
  $\tilde{X} = \tilde{X} \setminus \{x | x \in \text{clique}_{i+1}\}$;
  $i = i + 1$;
end

Algorithm 2: Correlation clustering splits a selected region X into sub-regions with high interior correlation but low mutual correlation.

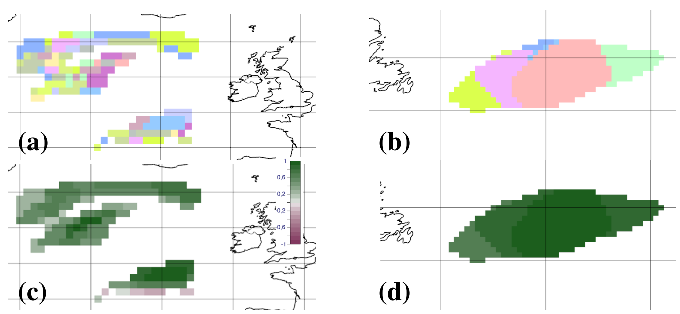

Fig. 5: Partition of superlevel sets using correlation clustering (Alg. 2). Level set in ESA between **(a)** precipitation and moisture flux and **(b)** geopotential height error and geopotential height. Many small clusters (different colors) in **(a)** indicate high stochastic variability. Large clusters **(b)** indicate large statistically coherent regions. **(c,d)** Correlation of first cluster to all others in the region. Even negative correlations (red) appear.

4.2.2 Significance of sensitivity

To test whether an extracted region of high sensitivity is significant, a two-tailed t-test is commonly applied to the correlation values in a selected region ([47]). For the 51 member ensembles used in our work, values higher than 0.276 can be deemed significant with 95% confidence. This, however, can only be seen as a rough indicator, since for every fixed J the t-test assumes normally distributed values for s, which cannot be guaranteed in general. Therefore, we use the test solely to remove regions with low significance. Once a region has significant parts, it is considered as a whole in the upcoming analysis. In our visualizations, we use an additional visual channel to show the significance of the sensitivity values over the domain. According to Retchless and Brewer [32] we use stipple patterns as overlays to depict the significance. In particular, we use a point stipple pattern with constant point density to indicate regions with medium significance, i.e., between 50% and 94%, and a line stipple pattern with constant line spacing to indicate very low significance, i.e., below 50%. The overlays are generated by mapping textures filled with the stipple patterns over the entire domain, and fading out a texture's contribution where the significance values are not in the corresponding interval. The different types of stipple overlays we use are demonstrated in Fig. 1c.

4.3 Tracking sensitive regions

In the majority of meteorological workflows for ensemble analysis today, tracking of regions in which significant weather events are predicted is performed via animation and manual matching of corresponding regions in successive time steps. One of the reasons is that the predictability of regions associated with high-impact weather events, which are often the regions of interest, is low, and predicted events tend to undergo major changes in shape and location over lead time. To support the automatic tracking of regions over time also in this situation, we propose an improved tracking algorithm using the optical flow in forward and backward direction.

4.3.1 Bidirectional OF-based matching

Given a sequence of 2D or 3D time-varying scalar fields, e.g., precipitation, the optical flow (OF) [15, 22] estimates the apparent motion of precipitation patterns in two successive fields, the source and the target field. The main idea behind the OF algorithm is to minimize a global cost function that represents the rate of change of the scalar quantity from the source to the target, under the assumption that the metric does not change during a sufficiently small duration. The result is a displacement field, which indicates for every spatial location and given data value at that location to which location this value should be moved such that the transformed field matches the target field. In our workflow we use the implementation of the Farneback algorithm [10] that is provided by OpenCV [4], with a window size of 12 in either dimension, pyramid scale of 0.4, polynomials of degree 8, 2 iterations, and a smoothing factor of 1.2.

To establish a correspondence between regions in the ESA field at different time steps, which are defined as superlevel sets to the correlation threshold, we apply forward *and* backward OF between these fields (see Fig. 6). In the forward pass, the OF field is used to

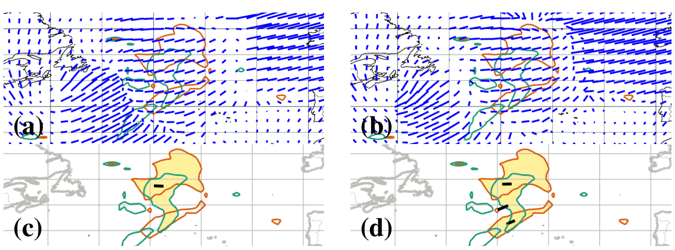

Fig. 6: Forward **(a)** and backward **(b)** OF between sensitivity fields at two consecutive time steps. **(c)** Forward OF tracking with 25% filtering matches a single region (orange outline, yellow interior) to the picked region (green outline, yellow interior). **(d)** Bidirectional OF tracking using the same filtering matches two additional regions, one of them a major split event visible in Fig. 1c.

estimate at which location in the local surrounding of every source location a similar ESA value is found in the next time step. This leads to the predicted motion of the ESA field over the corresponding time period. In principle, this motion field can already be used to establish a correspondence between regions in successive time steps: Correspondence between a source and target region is established if there is at least one location in the source region where the motion vector points at a location in the target region [2]. As manually validated by our domain experts, however, this approach has led to a number of falsely matched regions as well as missed connections in our specific scenario.

To weaken this limitation, we have modified the OF-based matching procedure in two ways: Firstly, for every target location that is indicated by the motion vector, we test whether this location is sufficiently close to a target region. This indicates that the location belongs to that region with high probability. In particular, we test for a distance less than the size of one cell of the grid structure at which the ESA values have been computed. The total number of matches from the source to the target region, as well as the percentage of points of the source region that are matched to a certain target region is stored. The number of matches is used in the split-merge diagram described below to indicate the size of matched regions, and the percentage of matched locations is used to filter out region pairs that are only connected weakly.

Secondly, we use the OF, now oriented backward in time, to estimate from where in the source field a physical quantity was moved over the current time step. We could do this via a semi-lagrangian advection step known from fluid simulation, where the motion vectors in the target field are simply reversed, yet the explicit backward step using OF enables us to employ a bidirectional matching according to the distribution of the values in both fields. The result of the backward step is used in the following way: Firstly, we establish a new link in exactly the same way as in the forward step, and we store the number and percentage of matched locations. The computed links are now filtered with respect to a prescribed percentage threshold: If for a given pair of matched regions in both the forward and the backward step this threshold could not be reached, the link is disconnected. Fig. 7 illustrates these two scenarios. Our partners from meteorology liked in particular the possibility to interactively change the percentage threshold and see immediately the resulting matches in the spatial view (see below). By this, it was easy for them to explore all suggested pathways of weather events and select those in best agreement with their domain-specific assumptions.

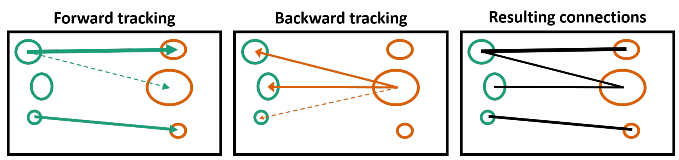

Fig. 7: Schematic illustration of bidirectional OF-based region tracking. Matches (indicated by arrows) identified between regions in time step t_i (green circles) and t_{i+1} (orange circles) are shown for forward and backward OF tracking. Arrow width indicates number of matched locations. Dashed lines indicate weak matches which are filtered out. Final result is the union of remaining matches.

4.3.2 Visual encoding of region tracks

The computed matchings between source and target regions over the entire simulation time are displayed in a split-merge diagram, where every sensitivity region is represented as a node and edges between nodes are established if a match has been established (Fig. 8b). The number and percentage of matched locations for a selected region is used to adjust the width of edges in this graph (Fig. 8b). The diagram orders all extracted regions along the horizontal time axis, yet the regions in one single time step are drawn along the vertical axis in an unordered way. It is clear that this can lead to significant changes in the edge orientations from time step to time step. However, since the split-merge diagram is solely used as an additional support tool for selecting a region that can then be tracked in the spatial view, we did not focus on improving its visualization. Improved layouting strategies for such diagrams are discussed by Widanagamaachchi et al. [45]. In the split-merge diagram time decreases along the horizontal axis. The resulting unusual ordering of time steps was specifically requested by our collaboration partners from meteorology. Since relevant sensitivity structures are defined close to the selected event and then traced backward in time, our partners wanted to see the same temporal evolution in the diagram.

In the diagram the user can now pick a time step and let all extracted regions in that time step being displayed in the linked spatial view. Furthermore, we let the user select a region in the spatial view, and show all regions that were merged with this region or which emerged due to a split of this region. From this set of regions, i.e., the connected

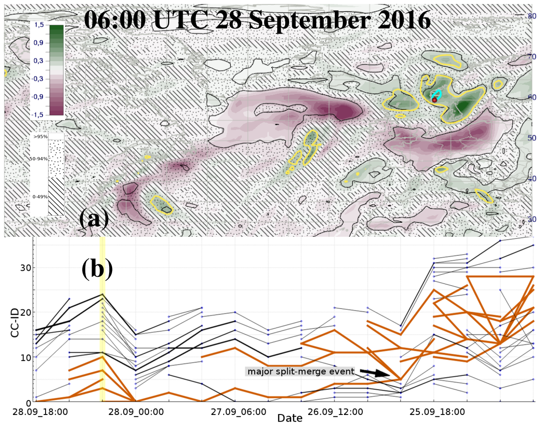

Fig. 8: **(a)** ESA sensitivity between precipitation and moisture flux (color). Stippling covers regions with less than 95% significance. The region with positive ESA close to the user-placed pole (red dot) is selected for tracking, which is performed towards the initial time of the simulation with increasing time-difference to the precipitation event. **(b)** The split-merge graph with 10% filtering shows that the selected region can be traced back (orange path) to the initial time step. There is a merge event which connects two major tracks (black arrow).

component *C* in the diagram containing the selected region, a so-called swipe-path is generated in the spatial view. The swipe-path encodes the spatio-temporal evolution of the selected region, by assigning to every location that was overlaid by a region in *C* a scalar value indicating the first time this location was covered. In this way, the evolution of a region is encoded in a scalar field, which can then be visualized by isocontouring, 2D color plots, or direct volume rendering in 3D (see Figs. 11c and 1c). We do not smooth the resulting spatial structure to keep the grid resolution visible, which was also requested by our partners from meteorology. When mapping scalar values to colors, gradually changing colormaps are beneficial since neighboring time steps can be identified while the whole evolution remains clearly visible.

Even though a swipe-path, conceptually, is a rather simple visual representation of the temporal evolution of selected regions, especially its use for depicting the motion of 3D regions was very well received by our partners from meteorology. For the first time ever, meteorologists could investigate the vertical movement of sensitivity regions, including the geometric changes these regions undergo over time. The domain experts argued explicitly against the use of glyphs for depicting the temporal evolution of a selected region, to be able to directly read both the spatial and temporal changes a region undergoes.

5 Results and Evaluation

All components of the proposed visual workflow have been integrated into the open-source meteorological visualization tool "Met.3D" [25, 30]. The existing data processing pipeline of Met.3D, as well as existing visualization functionality for meteorological maps and direct volume rendering, provided a suitable pre-existing infrastructure. Also, integration into an existing tool eases promulgation of our approach into the meteorological community.

Firstly, we demonstrate the use of the workflow to investigate sensitivities in weather forecasts related to the extratropical transition (ET) of tropical cyclone "Karl" in late September 2016. The case is a focus of current NAWDEX analyses, aiming at identifying atmospheric processes that may have caused deficiencies in the predictability of the subsequent weather evolution (cf. [37]). In the days considered, Karl moved into the middle latitudes, merged with a pre-existing weak extratropical cyclone, rapidly re-intensified during this extratropical transition and thereby impacted the jet stream over the North Atlantic and northern Europe. This period was characterized by large forecast errors. Visualization methods to investigate further aspects of the case were already presented by Kumpf et al. [19] and Kern et al. [18]. We use data from the ECMWF ensemble prediction system (ENS; e.g., [20]); all results are produced from data on a regular latitude–longitude grid with a grid spacing of 0.5°(1° for precipitation), using levels of constant pressure in the vertical.

We consider an extreme precipitation event that occurred along the Norwegian coast at the end of Karl's life cycle on 28-29 September 2016. In some places, more than 116 mm of rain fell in less than 24 hours. A large-scale and a smaller-scale perspective are analyzed, yielding first insights into the atmospheric processes involved.

5.1 Large-scale perspective: Geopotential height error

Fig. 9 shows the forecast error (the difference between forecast and subsequently observed values) of 300 hPa geopotential height (gravity-adjusted height above sea level) of the ensemble control forecast at 00:00 UTC 28 September 2016, valid at six days lead time from the forecast initialized at 00:00 UTC 22 September 2016, as well as the ensemble mean absolute forecast error of the same date. Contour lines of geopotential height show two key features: A large gradient over the North Atlantic indicates the jet stream, a low-gradient region over eastern Europe indicates the remnants of a high-pressure system over Scandinavia that dominated northern European weather in the days before. The decay of this structure (referred to as "block") enabled the jet over the Atlantic to extend over Scandinavia, which steered the cyclone that developed in the wake of Karl into Norway, causing extreme precipitation.

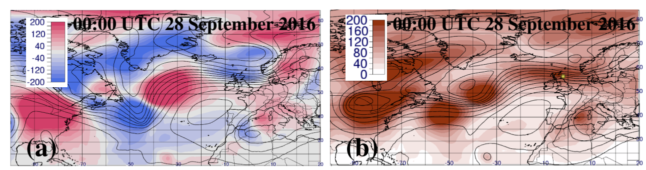

Fig. 9: **(a)** Geopotential height error (m) of control forecast, and **(b)** geopotential height mean absolute error (m), at 300 hPa with a lead time of 144h. Contours show geopotential height of control run **(a)** and analysis **(b)**.

Of interest are forecast errors over northern Europe associated with the decay of the block and the jet impinging on Norway. The error field exhibits very different structure in different members of the ensemble (not shown), yet still in the ensemble mean of the absolute errors, errors associated with this jet are very large, with a maximum over Scotland (Fig. 9b). It is a natural choice to first consider this error maximum; we hence aim at investigating the sensitivity of errors in this region to geopotential at earlier forecast times.

The user selects a seed location at 300 hPa over Scotland, at 00:00 UTC 28 September 2016, and the system computes an extended correlation clique with low stochastic variation (see Fig. 10) It captures the error region over Scotland and extends vertically throughout the entire atmosphere but does not include the southern Scandinavian region. The ESA signal (sensitivities) at the selected time (Fig. 10b) show a distinct dipole correlation pattern indicating that members with smaller forecast errors have lower (higher) geopotential height north (south) of the large gradient. Furthermore, the dipole pattern extends to the east of the large gradient. This pattern indicates that the actual jet was stronger and extended farther to the east than in the ensemble mean. The sensitivities, however, quickly vanish as we trace them backwards in time (illustrated in Fig. 10c for the positive ESA signal north of the large gradient). Evidently, the error maximum within the largest gradient is thus not associated with geopotential features at previous times earlier than 12-24h and no meaningful insight is gained into which earlier processes may have caused the error. As the region of large errors is located in a region of a strong gradient, however, the considered error maximum may be dominated by uncertainties in the north-south location of the jet.

Our interest here, however, is rather on the zonal extent of the jet and thus to investigate the situation further, we consider the lower-gradient region at the eastern end of the jet. A correlation clique in

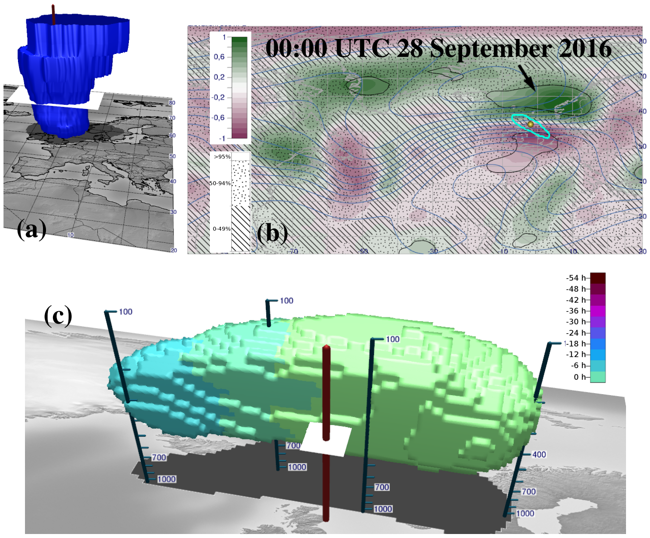

Fig. 10: (**a**) Correlation clique of geopotential height error, seeded at 300hPa over Scotland (pole and white cut). (**b**) ESA at 300hPa, using the clique from (a). (**c**) 3D swipe-path of sensitive region north of Great Britain (black arrow in (b)). The sensitivity signal vanishes after 18 h.

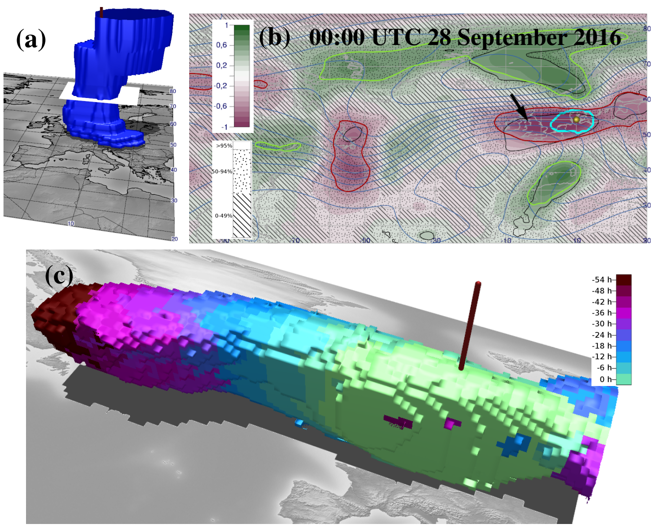

Fig. 11: **(a)** As Fig. 10a, but seeded over Denmark (pole). **(b)** ESA between geopotential height error and geopotential height at 300hPa using root mean square error of clique (turquoise) from **(a)** as input metric. Blue contours show mean pressure of geopotential height, stippling covers statistically insignificant regions to black contours marking 95% confidence. The red region indicated by a black arrow was later tracked. **(c)** Swipe-path of sensitivity feature over Denmark to correlation clique (a) seeded at dark red pole. Color changes in 6 hourly steps. Darker color indicates earlier valid times. Blue colors in the east show that the initial feature must have split before the second time step.

this region, centered at 300 hPa over Denmark (Fig. 11a), also yields a very coherent spatial structure, both in the horizontal and in the vertical, similar in this respect to the one associated with the error maximum discussed above. The ESA signal now exhibits a “tripole” structure with an elongated region of strong negative correlations overlapping the clique and positive correlations to the north and south. Hence, members with smaller forecast errors exhibit a stronger and more northerly jet and a stronger block to its east than the ensemble mean, and in addition more pronounced wave breaking over the western Mediterranean.

Fig. 11c shows the 3D swipe-path obtained from tracking the negative (red) ESA feature. The sensitivity pattern can be traced up- (westwards) and downstream (eastward) in time, respectively, with notably faster speed “upstream” than “downstream”. This error pattern in the large-scale flow therefore has both an upstream component, as well as a more local “block” component. Importantly, the sensitivity pattern gets gradually more confined to the upper troposphere when tracing back in time, which emphasizes the importance of the jet structure over the North Atlantic two days before the development over Scandinavia. Focusing on this upstream propagation of the ESA feature, we find that the statistically-significant signal is lost at 18:00 UTC 25 September 2016 in the ridge over the western North Atlantic. This is the time at which the extratropical transition of Karl started to impact the ridge development ([37]). Inspecting the ESA map at this time (at 300 hPa, Fig. 12a), we note a statistically-significant positive signal near the base of the trough upstream of the ridge. Going further back in time reveals that this signal highlights persistently the region between the tropical cyclone Karl and the upstream trough in the preceding 24 h (illustrated at 12:00 UTC and 00:00 UTC 25 September 2016, Fig. 12b,c). This signal is consistent with the high sensitivity of the outcome of the extratropical transition to the occurrence of the tropical cyclone and the upstream trough [8].

Our approach facilitates the intuitive building of hypothesis chains. As one example, it is now of interest if the error identified in the ridge is indeed associated with tropical cyclone Karl during its extratropical transition. To this end, we select a clique of geopotential errors from the center of the negative ESA signal in the ridge at 00:00 UTC 26 September 2016 and investigate its correlation with low-level features (using 925hPa geopotential). At this time, the ridge error is associated with a statistically significant dipole in 925hPa geopotential elongated in the direction of Karl’s track (not shown), indicating that ridge errors correlate with the position of Karl. This dipole pattern can be traced back for 96h (Fig. 13), using the positive part of the signal, until the start of the forecast. The signal, however, is statistically significant only in the first and last 24h of the considered time period. Still, Fig. 13b indicates that errors in the track of Karl lead to errors in the ridge formation over the North Atlantic, which in turn leads to errors in the large-scale flow over Scandinavia two days later. Application of our method thus greatly facilitates building a hypothesis chain that ultimately relates the extreme precipitation event in southern Norway to the evolution of tropical cyclone Karl. It is a promising task for future work to further elaborate on this hypothesis.

5.2 Smaller-scale perspective: Moisture flux

The second perspective focuses more directly on the extreme precipitation event, aiming at identifying the sensitivity of the forecast precipitation amount to uncertainties in the forecast moisture flux (the product of humidity and wind) at earlier times. When selecting a square target region enclosing a significant weather event, as in common meteorological workflows, a rather weak ESA signal is computed (see Fig. 4b,c). By using our workflow, the user has picked different seed locations (see Fig. 3), until a coherent correlation clique (Fig. 1a) yields clear, significant ESA signals (Fig. 4d). The interpretation of the signal is rather straightforward: The positive signal over southern Norway and upstream (to the west) reveals that stronger precipitation is associated with stronger antecedent moisture flux. Interestingly, our case does not exhibit a dipole pattern close to the region of extreme precipitation, which indicates sensitivity to the location of antecedent moisture flux and has often been found in other studies. Here, instead, the sensitivity is to the magnitude of the moisture flux over a relatively broad area.

The automated tracking clearly traces the signal back to the beginning of Karl’s extratropical transition approximately 4 days earlier (Fig. 1c). The split-merge graph (Fig. 8b) and the mean tracking vectors in Fig. 1c indicate that the signal undergoes several merge and splitting events during this time. In particular before 00:00 UTC 26 September 2016, the sensitivity regions are of relatively small scale before they merge into a spatially more coherent pattern (Figs. 8b, 1c). By inspecting the individual ESA maps we visually verified the automated tracking. The ESA signal found in the moisture flux thus corroborates

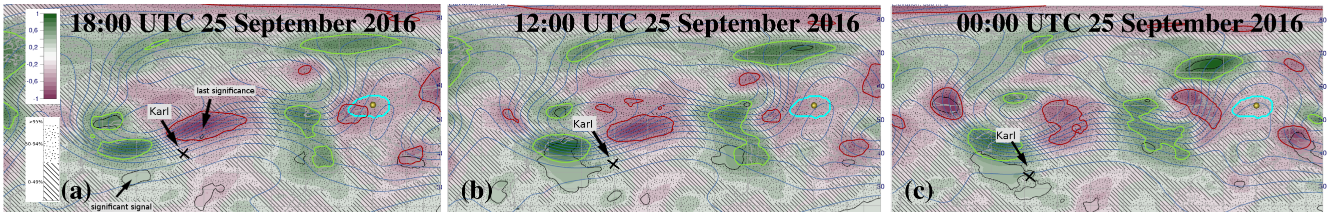

Fig. 12: ESA to clique over Denmark (cf. Fig. 11). **(a)** The last significant part of the tracked red region vanishes. A new significant signal is detected close to Karl. **(b,c)** It appears in earlier time steps as well, indicating a link between the error over Denmark and the early stage of Karl.

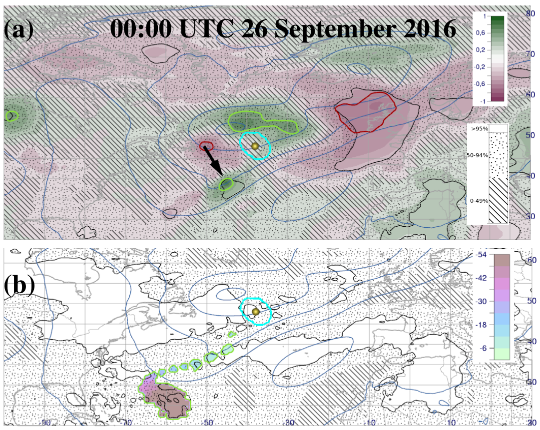

Fig. 13: **(a)** Sensitivities between geopotential height error at 300hPa (clique marked in turquoise) and geopotential height at 925 hPa. **(b)** Swipe path for the positive (green) sensitivity feature marked by the arrow in (a).

the hypothesis chain developed using the geopotential correlations, i.e., that the extreme precipitation event in Norway is sensitively linked to the evolution of tropical cyclone Karl.

Finally, it is important to note that the statistically coherent structures in the moisture flux field are relatively small compared to the ones in the geopotential height field (Fig. 5), i.e., the moisture flux region exhibits more nontrivial smaller-scale structures. Hence, more atmospheric observations are required to adequately sample the sensitive moisture flux region. If targeted observations were collected to reduce the uncertainty in the forecast of the extreme precipitation event, the analysis suggest that it might be more efficient to sample the sensitive geopotential height region using less observations. Depending on the forecast horizon, the observations should be collected in the region between Karl and the upstream trough, to improve the forecast of Karl's track, or directly within the ridge after the completion of the ET.

5.3 User discussion

We discuss our first experiences with the method from the point of view of the meteorological domain experts in the author team. To our knowledge, this is the first interactive end-to-end workflow for ESA (providing guidance from selecting the region for J to tracking of sensitivities, including guidance on robustness of results and possibly further locations of interest, i.e., supporting hypothesis chain building). Compared to our current script-based workflow, ESA is therefore greatly facilitated during all necessary steps of the analysis. Definition of a suitable region for J using correlation cliques greatly facilitates selection of a meaningful (in a statistical sense) region. In our script-based workflow, region selection was guided by meteorological intuition. Now, the selection of the initial point is guided by intuition, the region is proposed by the automated method. The user can check if the proposed region contains the event of interest.

Tracking of sensitivity features and subsequent depiction of a swipe-path provides a quick, succinct overview of the evolution of the sensitivities, thus enabling fast judgement if further investigation of the sensitivities should be conducted. If yes, attention of the investigation is directed to the "end" (in a backward tracking sense) of the swipe-path, which can be the starting point of any further sensitivity analysis (in the context of building hypothesis chains). In particular, the 3D depiction of the swipe-path provides unprecedented insight into the evolution of the 3D structure of the sensitivities. The split-merge diagram provides guidance for the robustness of the signal. If many split and merge events are present, the user needs to evaluate the sensitivities carefully to judge their physical relevance; few, distinct events, on the other hand, may indicate physically meaningful evolutions. For example, the lack of split events between 28 and 26 September in the orange curve in Fig. 8b is guidance to interpret the track as physically meaningful backward propagation of the sensitivity signal from the initially selected weather event of interest to the Western Atlantic near Karl.

6 Conclusion

We have proposed a novel visual analysis workflow to facilitate an interactive analysis of sensitivities of a forecast metric J on another forecast field. Our workflow enables the user to interactively identify regions of intercorrelated grid points from which J is computed, and to automatically track features of high sensitivity through time. A "swipe-path" visualization showing the track of a sensitivity feature in time has been proposed that allows the user to immediately see geographical regions from which sensitivities originated. Swipe-paths are generated from user-selected features and can be displayed both in 2D and 3D. In particular, the novel interactive sensitivity tracking in 3D opens the door to analyses considering all spatial dimensions, as not possible with existing 2D meteorological workflows.

The workflow has been integrated into the open-source software Met.3D. Its benefit has been demonstrated with a real-world case study taken from ongoing analyses of the NAWDEX atmospheric field campaign. Compared to script-based tools commonly applied in the meteorological community, the workflow proposed here greatly simplifies ensemble sensitivity analysis by providing an interactive end-to-end workflow that encapsulates all steps required for sensitivity analysis in a single framework. At the same time, important information about the reliability of the results is provided, and a fully 3D analysis is facilitated. In the near future, the method will actively be used in further data analysis activities related to the NAWDEX campaign. Future work might include detailed statistics of sensitivity structures, a manual correction tool for the tracking and further enhancement of the swipe-path using glyphs to compensate for occlusion.

Acknowledgments

The research leading to these results has been done within the subprojects B5 and A4 of the Transregional Collaborative Research Center SFB/TRR 165 Waves to Weather funded by the German Research Foundation (DFG). Access to ECMWF prediction data has been kindly provided in the context of the ECMWF special project Support Tool for HALO Missions. We thank all the reviewers for their constructive criticism and valuable comments.

REFERENCES

[1] B. Ancell and G. J. Hakim. Comparing adjoint- and Ensemble-Sensitivity analysis with applications to observation targeting. *Monthly Weather Review*, 135(12):4117–4134, Dec. 2007. doi: 10.1175/2007mwr1904.1

[2] A. Anil Valsangkar, J. Merwin Monteiro, V. Narayanan, I. Hotz, and V. Natarajan. An exploratory framework for cyclone identification and tracking. *IEEE Transactions on Visualization and Computer Graphics*, PP:1–1, 02 2018. doi: 10.1109/tvcg.2018.2810068

[3] N. Bansal, A. Blum, and S. Chawla. Correlation clustering. *Machine Learning*, 56:89–113, 2004. doi: 10.1023/B:MACH.0000033116.57574.95

[4] G. Bradski and A. Kaehler. Opencv. *Dr. Dobb's journal of software tools*, 3, 2000.

[5] P. T. Bremer, G. Weber, J. Tierny, V. Pascucci, M. Day, and J. Bell. Interactive exploration and analysis of large-scale simulations using topology-based data segmentation. *IEEE Transactions on Visualization and Computer Graphics*, 17(9):1307–1324, Sept 2011. doi: 10.1109/TVCG.2010.253

[6] C.-K. Chen, C. Wang, K.-L. Ma, and A. T. Wittenberg. Static correlation visualization for large time-varying volume data. In *Visualization Symposium (PacificVis), 2011 IEEE Pacific*, pp. 27–34. IEEE, 2011.

[7] H. Doraiswamy, V. Natarajan, and R. S. Nanjundiah. An exploration framework to identify and track movement of cloud systems. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2896–2905, Dec 2013. doi: 10.1109/TVCG.2013.131

[8] C. Evans, K. M. Wood, S. D. Aberson, H. M. Archambault, S. M. Milrad, L. F. Bosart, K. L. Corbosiero, C. A. Davis, J. R. Dias Pinto, J. Doyle, et al. The extratropical transition of tropical cyclones. Part I: Cyclone evolution and direct impacts. *Monthly Weather Review*, 145(11):4317–4344, 2017. doi: 10.1175/mwr-d-17-0027.1

[9] Z. Fang, T. Möller, G. Hamarneh, and A. Celler. Visualization and Exploration of Time-varying Medical Image Data Sets. In *Proceedings of Graphics Interface 2007*, pp. 281–288, 2007. doi: 10.1145/1268517.1268563

[10] G. Farnebäck. Two-frame motion estimation based on polynomial expansion. In *Scandinavian conference on Image analysis*, pp. 363–370. Springer, 2003. doi: 10.1007/3-540-45103-x_50

[11] T. Fiolleau and R. Roca. An algorithm for the detection and tracking of tropical mesoscale convective systems using infrared images from geostationary satellite. *IEEE Transactions on Geoscience and Remote Sensing*, 51(7):4302–4315, July 2013. doi: 10.1109/TGRS.2012.2227762

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

[13] A. V. Gambheer and G. S. Bhat. Life cycle characteristics of deep cloud systems over the Indian region using INSAT-1B pixel data. *Monthly Weather Review*, 128(12):4071–4083, 2000. doi: 10.1175/1520-0493(2000)129<4071:LCCODC>2.0.CO;2

[14] J. Hanley and R. Caballero. Objective identification and tracking of multicentre cyclones in the ERA-Interim reanalysis dataset. *Quarterly Journal of the Royal Meteorological Society*, 138(664):612–625, 2012. doi: 10.1002/qj.948

[15] B. K. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1):185 – 203, 1981. doi: 10.1016/0004-3702(81)90024-2

[16] D. Jen, P. Parente, J. Robbins, C. Weigle, R. Taylor II, A. Burette, and R. Weinberg. Imagesurfer: A tool for visualizing correlations between two volume scalar fields. In *Proceedings of IEEE Visualization*, pp. 529–536, 2004. doi: 10.1109/VISUAL.2004.46

[17] G. Ji and H.-W. Shen. Feature tracking using earth mover's distance and global optimization. In *Pacific Graphics*, vol. 2, 2006.

[18] M. Kern, T. Hewson, F. Sadlo, R. Westermann, and M. Rautenhaus. Robust detection and visualization of jet-stream core lines in atmospheric flow. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):893–902, 2018. doi: 10.1109/tvcg.2017.2743989

[19] A. Kumpf, B. Tost, M. Baumgart, M. Riemer, R. Westermann, and M. Rautenhaus. Visualizing confidence in cluster-based ensemble weather forecast analyses. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):109–119, 2018. doi: 10.1109/tvcg.2017.2745178

[20] 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

[21] T. Liebmann, G. H. Weber, and G. Scheuermann. Hierarchical Correlation Clustering in Multiple 2D Scalar Fields. *Computer Graphics Forum*, 2018. doi: 10.1111/cgf.13396

[22] B. D. Lucas, T. Kanade, et al. An iterative image registration technique with an application to stereo vision. In *IJCAI*, vol. 81, pp. 674–679, 1981.

[23] L. Magnusson. Diagnostic methods for understanding the origin of forecast errors. *Quarterly Journal of the Royal Meteorological Society*, 143(706):2129–2142, July 2017. doi: 10.1002/qj.3072

[24] S. J. Majumdar. A review of targeted observations. *Bulletin of the American Meteorological Society*, 97(12):2287–2303, Apr. 2016. doi: 10.1175/bams-d-14-00259.1

[25] Met.3D contributors. Met.3D. `http://met3d.wavestoweather.de`, 2018. Accessed 31 March 2018.

[26] C. Muelder and K. L. Ma. Interactive feature extraction and tracking by utilizing region coherency. In *2009 IEEE Pacific Visualization Symposium*, pp. 17–24, April 2009. doi: 10.1109/PACIFICVIS.2009.4906833

[27] T. Pfaffelmoser and R. Westermann. Visualization of Global Correlation Structures in Uncertain 2D Scalar Fields. *Computer Graphics Forum*, 31:1025–1034, 2012. doi: 10.1111/j.1467-8659.2012.03095.x

[28] T. Pfaffelmoser and R. Westermann. Correlation visualization for structural uncertainty analysis. *International Journal for Uncertainty Quantification*, 3:171–186, 2013. doi: 10.1615/int.j.uncertaintyquantification.201200393

[29] M. Rautenhaus, M. Böttinger, S. Siemen, R. Hoffman, R. M. Kirby, M. Mirzargar, N. R´ober, and R. Westermann. Visualization in meteorology - a survey of techniques and tools for data analysis tasks. *IEEE Transactions on Visualization and Computer Graphics*, PP(99):1–1, 2017. doi: 10.1109/TVCG.2017.2779501

[30] 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). *Geosci. Model Dev.*, 8(7):2329–2353, 2015.

[31] F. Reinders, F. H. Post, and H. J. Spoelder. Visualization of time-dependent data with feature tracking and event detection. *The Visual Computer*, 17(1):55–71, Feb 2001. doi: 10.1007/PL00013399

[32] 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

[33] H. Saikia, H.-P. Seidel, and T. Weinkauf. Fast similarity search in scalar fields using merging histograms. In H. Carr, C. Garth, and T. Weinkauf, eds., *Topological Methods in Data Analysis and Visualization*, pp. 121–134. Springer International Publishing, Cham, 2015. doi: 10.1007/978-3-319-44684-4_7

[34] H. Saikia and T. Weinkauf. Global feature tracking and similarity estimation in time-dependent scalar fields. *Computer Graphics Forum*, 36(3):1–11, 2017. doi: 10.1111/cgf.13163

[35] R. Samtaney, D. Silver, N. Zabusky, and J. Cao. Visualizing features and tracking their evolution. *Computer*, 27(7):20–27, July 1994. doi: 10.1109/2.299407

[36] N. Sauber, H. Theisel, and H. Seidel. Multifield-graphs: An approach to visualizing correlations in multifield scalar data. *IEEE Transactions on Visualization and Computer Graphics*, 12:917–924, 2006. doi: 10.1109/tvcg.2006.165

[37] A. Schäfler, G. Craig, H. Wernli, P. Arbogast, J. D. Doyle, R. McTaggart-Cowan, J. Methven, G. Rivière, F. Ament, M. Boettcher, M. Bramberger, Q. Cazenave, R. Cotton, S. Crewell, J. Delanoë, A. Dörnbrack, A. Ehrlich, F. Ewald, A. Fix, C. M. Grams, S. L. Gray, H. Grob, S. Gro, M. Hagen, B. Harvey, L. Hirsch, M. Jacob, T. Kölling, H. Konow, C. Lemmerz, O. Lux, L. Magnusson, B. Mayer, M. Mech, R. Moore, J. Pelon, J. Quinting, S. Rahm, M. Rapp, M. Rautenhaus, O. Reitebuch, C. A. Reynolds, H. Sodemann, T. Spengler, G. Vaughan, M. Wendisch, M. Wirth, B. Witschas, K. Wolf, and T. Zinner. The north atlantic waveguide and downstream impact experiment. *Bulletin of the American Meteorological Society*, 2018. doi: 10.1175/BAMS-D-17-0003.1

[38] D. Silver and X. Wang. Tracking and visualizing turbulent 3d features. *IEEE Transactions on Visualization and Computer Graphics*, 3(2):129–141, Apr 1997. doi: 10.1109/2945.597796

[39] B. S. Sohn and C. Bajaj. Time-varying contour topology. *IEEE Transactions on Visualization and Computer Graphics*, 12(1):14–25, Jan 2006. doi: 10.1109/TVCG.2006.16

[40] J. Sukharev, C. Wang, K.-L. Ma, and A. Wittenberg. Correlation study of time-varying multivariate climate data sets. In *Proceedings of IEEE Pacific Graphics 2009*, pp. 161–168, 2009. doi: 10.1109/pacificvis.2009.4906852

[41] 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

[42] R. D. Torn and G. J. Hakim. Ensemble-based sensitivity analysis. *Monthly Weather Review*, 136(2):663–677, 2008. doi: 10.1175/2007mwr2132.1

[43] 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

[44] H. Wernli and C. Schwierz. Surface cyclones in the ERA-40 dataset (1958-2001). Part I: Novel identification method and global climatology. *Journal of the Atmospheric Sciences*, 63(10):2486–2507, 2006. doi: 10.1175/JAS3766.1

[45] W. Widanagamaachchi, C. Christensen, V. Pascucci, and P. T. Bremer. Interactive exploration of large-scale time-varying data using dynamic tracking graphs. In *IEEE Symposium on Large Data Analysis and Visualization (LDAV)*, pp. 9–17, Oct 2012. doi: 10.1109/LDAV.2012.6378962

[46] H. Zhang, Y. Hou, D. Qu, and Q. Liu. Correlation visualization of time-varying patterns for multi-variable data. *IEEE Access*, 4:4669–4677, 2016. doi: 10.1109/ACCESS.2016.2601339

[47] M. Zheng, E. K. M. Chang, and B. A. Colle. Ensemble sensitivity tools for assessing extratropical cyclone intensity and track predictability. *Weather and Forecasting*, 28(5):1133–1156, May 2013. doi: 10.1175/waf-d-12-00132.1