Vision, Modeling, and Visualization (2023)
T. Grosch and M. Guthe (Eds.)

# Uncertain Stream Lines

J. Zimmermann, M. Motejat, C. Rössl, H. Theisel

Visual Computing Group, University of Magdeburg, Germany

**Abstract**

*We present a new approach for the visual representation of uncertain stream lines in vector field ensembles. While existing approaches rely on a particular seed point for the analysis of uncertain streamlines, our approach considers a whole stream line as seed structure. With this we ensure that uncertain stream lines are independent of the particular choice of seed point, and that uncertain stream lines have the same dimensionality as their certain counterparts in a single vector field. Assuming a Gaussian distribution of stream lines, we provide a visual representation of uncertain stream lines based on a mean map and a covariance map. The extension to uncertain path lines in ensembles of time-dependent vector fields is straightforward and is also introduced in the paper. We analyze properties, discuss discretization and performance issues, and apply the new technique to a number of flows ensembles.*

**CCS Concepts**

• ***Human-centered computing*** → ***Scientific visualization;***

## 1. Introduction

Uncertainty representation is still one of the main challenges in Visualization. While it is commonly accepted that the representation of the underlying uncertainty of a data set is inevitable for a thorough data analysis. It poses significant challenges to the visualization because considering uncertainty results in a significant increase of the data to be represented.

A common approach for uncertainty visualization is to show some objects or features along with their confidence intervals. This has been done for local objects, which describe data at a point like scalars, vectors, or tensors, as well as for global objects, which consist of integrated data like isolines, isosurfaces, or vortices. For each of these, a significant amount of research has been conducted to model and visualize the uncertainty. While established visualization techniques differ in many aspects, they have one thing in common: they do not increase the dimensionality of the data. One example are isosurfaces of a 3D scalar field. Here, the cardinality of the set of all uncertain isosurfaces is identical to the cardinality of "certain" isosurfaces if each "certain" isosurface is equipped with an uncertainty distribution. Adding uncertainty information to isosurfaces does not produce more isosurfaces but gives additional information to existing ones.

In this paper, we introduce a new model and visual representation of *uncertain stream lines* in an ensemble of vector fields. Existing solutions set a point of interest $\mathbf{x}$ in the domain and either consider all stream lines starting from $\mathbf{x}$ or start some probabilistic integration from $\mathbf{x}$. This gives *different* uncertain stream lines for every choice of $\mathbf{x}$, which leads to a higher number of uncertain stream lines than "certain" stream lines. As an example, consider a 2D steady vector field. The set of all stream lines is generated from a *one-dimensional* structure (see, e.g, [RT12]): find a finite set of seed curves such that stream lines starting from all points of these curves cover the whole domain. Contrary, for an ensemble of 2D vector fields, the set of uncertain stream lines – using existing definitions – is *two-dimensional*: every point of interest in the domain results in a different uncertain stream line. We consider this increase of dimensionality when modeling uncertainty a serious drawback of existing concepts. Not only does it increase the amount of data to be analyzed, it is also not stable under perturbation, which limits the information that is related to one stream line. While an uncertain stream line gives uncertainty information also in a distance to the seed point, a small change of the seed may change the global behavior of the stream line drastically.

We present the first approach to representing uncertain stream lines with dimensionality of "certain" and uncertain stream lines being identical. The main idea is to consider not a point of interest $\mathbf{x}$ but instead a whole parametric curve of interest $\mathbf{x}(t)$. Based on this concept, we analyze not one particular curve of interest but all stream lines of a vector field to be uncertain stream lines. As a result, the (arbitrary) seed point of an uncertain stream line does not appear as an artifact in the visualization since it does not bear any particular information. Figure 1 illustrates the difference of existing concepts and ours.

## 2. Related work

In this section, we give a brief review of uncertainty in visualization and in particular for vector fields, stream lines and path lines.



DOI: 10.2312/vmv.20231246

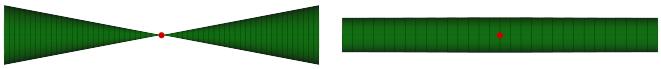

**Figure 1:** *Uncertain stream lines. Left: Existing concepts depend on the choice of a seed point (red), resulting in double-cone-like uncertainty structures. Right: Our model does not depend on a seed, resulting in tube-like uncertainty structures.*

### Uncertainty in Visualization

Finding adequate visual representations of uncertainty in scientific data has created a large amount of research in the last decade. Uncertainty can come from different sources that can be classified into uncertainty observed in sampled data, uncertainty generated by the modeling process, or uncertainty from the visualization process [BHJ*14].

There is a variety of techniques for uncertainty representation. A standard way to model uncertainty is the assumption of probability distribution functions (PDF) at the grid points. PDFs are visualized in [ESG97, KLDP02, LKP03]. As an alternative, glyphs were proposed for the visualization of uncertainty [Jon03, CR05, GS06]. Uncertainty is also represented by introducing ridges, roughness and oscillations on surfaces or images [LFC02, GR04]. We refer to the recent STAR [BHJ*14] on uncertainty visualization for a detailed review.

### Local uncertainty in vector fields

Local approaches describe uncertainty as a feature that can be evaluated at a point inside the vector field domain without considering the field's "long-term" integral behavior. Sanderson et al. [SJK04] describe patterns of uncertainty using a reaction-diffusion model, while Botchen et al. [BWE05] introduce a texture-based visualization technique, that represents local reliabilities by cross advection and error diffusion. The same authors used additional color schemes to emphasize uncertainty [BWE06]. Another approach by Zuk et al. [ZDG*08] uses bidirectional vector fields to illustrate the impact of uncertainty. Hanser et al. [HKR*18] provide an approach to analyze 1D parameter sensitivity of time-dependent flow simulation ensembles.

### Uncertainty for stream lines and path lines

There exist several approaches to capturing global behavior of stream lines and path lines for ensembles of vector fields in the literature. The perhaps simplest and most direct approach are spaghetti plots which provide straightforward overviews but tend to produce visual clutter [FBW16]. Mirzargar et al. [MWK14] introduce curve box-plots for the visualization of curve-like features based on the concept of statistical data depth. Otto et al. [OGHT10, OGT11] present a topological approach that is based on the integration of vector PDF. A related method by He et al. [HCLS16] for integrating uncertain stream lines is based on a Bayesian model. Hollister and Pang [HP20] present a method to measure uncertainty and to visualize member stream lines from an ensemble of vector fields by incorporating velocity probability density as a feature along each member stream line. Ferstl et al. [FBW16] derive stream line variability plots by computing a probabilistic mixture model for the stream line distribution, from which confidence regions can be derived in which the stream lines are most likely to reside. For a detailed overview of visualization on ensemble data we refer to the recent survey by Wang et al. [WHLS18] give an overview

All approaches mentioned here have something in common: They start with a particular seed point and consider all stream lines starting from there. This gives a "sharp" structure near the starting point that is blurred or spread out during integration (see Figure 1, left). This makes uncertain stream lines depend on an initial, "certain" decision: the placement of the seed point that is not assigned any uncertainty.

## 3. A new approach to uncertain stream lines

In this section and in the remainder of this paper we use the following notation: Given is an ensemble of $n$ (steady) velocity fields $\mathbf{v}_1(\mathbf{x}),...,\mathbf{v}_n(\mathbf{x})$ and their corresponding flow maps $\phi_1(\mathbf{x},\tau),...,\phi_n(\mathbf{x},\tau)$, where $\tau$ denotes the time interval. Furthermore, $\mathcal{N}$ is the normal distribution.

### 3.1. An illustrating example

We start with a small example to illustrate the shortcomings of existing approaches and provide a general impression of the new approach.

We introduce a simple 2D vector field ensemble that consists of the three vector fields

$$\mathbf{v}_1 = \begin{pmatrix} \cos\alpha \\ \sin\alpha \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} \cos\alpha \\ -\sin\alpha \end{pmatrix} \quad \mathbf{v}_3 = \begin{pmatrix} p \\ 0 \end{pmatrix} \tag{1}$$

with $\alpha = \frac{\pi}{4}$ and $p = 1$. Figure 2 (top row) illustrates the ensemble members. Note that all three ensemble members are constant vector fields. How many stream lines are in $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$? This is a one-parametric family: each stream line of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ has a unique intersection, e.g., with the $y$-axis of the underlying coordinate system. The $y$ coordinate of the intersection with the $y$-axis can be considered as a unique parameter for generating a stream line. And seeding stream lines from all points of the $y$-axis covers the whole domain with stream lines.

How do established approaches for uncertain stream lines perform on this data set? Figure 2 (bottom row, left) shows a typical application of spaghetti plots: A point of interest $\mathbf{x}_0$ is chosen and all stream lines starting from $\mathbf{x}_0$ are visualized. Figure 2 (bottom row, right *a*) shows the result of an uncertain stream line integration, as done in [OGHT10] or [HCLS16]. These methods also depend on the choice of a seed point $\mathbf{x}_0$, which results in a double-cone shaped structure centered at $\mathbf{x}_0$. A similar statement holds for advanced techniques like curve boxplots [MWK14] and stream line variability plots [FBW16]: They also require picking a point of interest $\mathbf{x}_0$, and consider all stream lines passing through this point, which similarly results in a double-cone structure. This shows again that existing tools for modeling uncertainty of stream lines are *point-oriented*: they require fixing a seed point $\mathbf{x}_0$. As a consequence, the set of

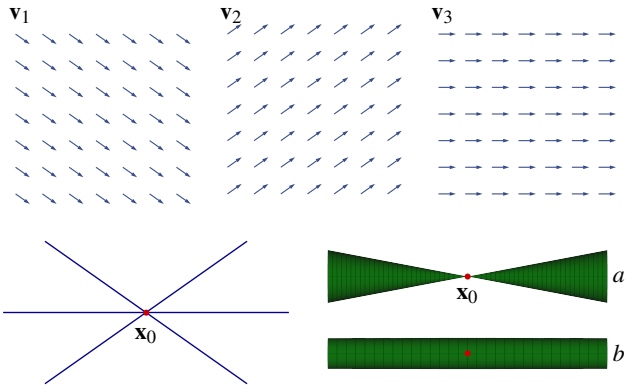


**Figure 2:** *A simple ensemble consisting of 3 constant vector fields. Top row: 3 ensemble members $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$. Bottom row, left: spaghetti plot starting from seed point $\mathbf{x}_0$. Bottom row, top right: uncertain stream line integration, curve boxplots, stream line variability plots starting from $\mathbf{x}_0$ create double-cone structures. Bottom row, bottom right: analyzing a stream line of $\frac{1}{3}(\mathbf{v}_1+\mathbf{v}_2+\mathbf{v}_3)$ to be an uncertain stream line.*

all uncertain stream lines is a *two-parametric family*: every seed $\mathbf{x}_0$ results in a different stream line representation.

Contrary to existing approaches, our method is not point oriented but *curve-oriented*: We consider a parametric curve of interest $\mathbf{x}(t)$ as input and analyze the confidence that $\mathbf{x}(t)$ is a stream line in all ensemble members. In fact, our approach creates a parametric covariance matrix $\mathbf{C}(t)$. The smaller $|\mathbf{C}(t)|$, the more certain is $\mathbf{x}(t)$ a stream line in all ensemble members. On the contrary larger $|\mathbf{C}(t)|$ indicate a lower confidence of $\mathbf{x}(t)$ being an uncertain stream line. In the example in Figure 2 (bottom row, right), a curve $\mathbf{x}(t)$ is analyzed to be an uncertain stream line. As the former method *a* is clearly point based and its extends are fast growing with two cones. The uncertain stream line presented in Figure 2 (bottom row, right) *b* shows a tube with constant extends which represents a constant covariance $\mathbf{C}(t)$ along the analyzed curve. Details on how to compute and visualize $\mathbf{C}(t)$ follow later in this paper.

Our approach does not only consider one particular curve of interest $\mathbf{x}(t)$ but all stream lines of a vector field of interest $\mathbf{w}(\mathbf{x})$ simultaneously. In fact, the curve of interest $\mathbf{x}(t)$ in figure 2 (bottom row, right) is a stream line of the average field $\mathbf{w} = \frac{1}{3}(\mathbf{v}_1+\mathbf{v}_2+\mathbf{v}_3)$. Note that the seed point of the uncertain stream line is of no importance because the covariance matrix has a similar behavior along the whole line and does not deviate near the particular seed.

In the following we introduce our approach formally. First, we consider one parametric curve as input (section 3.2). Afterwards we do a generalization to path lines 3.3).

### 3.2. One curve in a vector field ensemble

Given an ensemble of vector fields and their mean field $\mathbf{w}$, we consider a stream line of the mean field $\mathbf{x}(t)$ with $t \in [t_s, t_e]$ and want

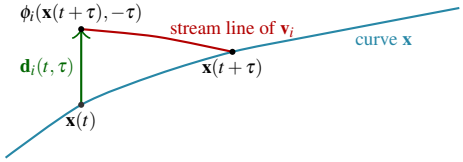


**Figure 3:** *Computation of $\mathbf{d}_i(t,\tau)$.*

to describe the confidence that $\mathbf{x}(t)$ is a stream line of $\mathbf{v}_1,\ldots,\mathbf{v}_n$. The main idea of our approach is to consider *all* stream lines of $\mathbf{v}_1,\ldots,\mathbf{v}_n$ that pass through $\mathbf{x}(t)$. We test stream lines for a Gaussian distribution and construct a parametric covariance matrix $\mathbf{C}(t)$. Our approach is based on the definition of the vector function

$$\mathbf{d}_i(t,\tau) = \phi_i(\mathbf{x}(t+\tau),-\tau) - \mathbf{x}(t) \tag{2}$$

for $t \in [t_s,t_e]$, $\tau \in [t_s-t, t_e-t]$, and $i = 1,\ldots,n$. On a high level, this function is a measure on how strong a stream line of $\mathbf{v}_i$ starting on a curve point $\mathbf{x}(t+\tau)$ moves away from the curve of interest under integration. Figure 3 gives an illustration. We can interpret the field $\mathbf{d}_i(t,\tau)$ as follows: We consider a point $\mathbf{x}(t)$ on the curve of interest and start a stream line integration of $\mathbf{v}_i$ on the curve but at a certain distance of this point: $\mathbf{x}(t+\tau)$. From there we integrate a stream line of $\mathbf{v}_i$ over a time interval $-\tau$: the difference of its end point and the point $\mathbf{x}(t)$ is $\mathbf{d}_i(t,\tau)$. If $\mathbf{x}(t)$ is a stream line of $\mathbf{v}_i$ then $\mathbf{d}_i(t,\tau) = \mathbf{0}$.

In order to analyze the uncertainty at the point $\mathbf{x}(t)$, we do not only consider the stream lines starting from one point of the curve but from all points.

Based on this, we can compute the covariance matrix as

$$\mathbf{C}_i(t) = \int_{t_s-t}^{t_e-t} \mathbf{d}_i(t,\tau)\,\mathbf{d}_i(t,\tau)^{\mathrm{T}}\,\mathrm{d}\tau$$

and

$$\mathbf{C}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{C}_i(t)a(t) \tag{3}$$

where

$$a(t) = \frac{(t_e-t_s)^2}{4(t-t_s)^2 - 4(t-t_s)(t_e-t) + 4(t_e-t)^2} \tag{4}$$

With this, a parametric curve $\mathbf{x}(t)$ is characterized by a covariance matrix function $\mathbf{C}(t)$. As $\mathbf{C}_i(t)$ grows quadratic towards the end for linear vector fields, the weighting function $a(t)$ compensates for this. This is justified by the fact that we want to have control of the (in worst case exponential) moving-off the curve of interest for longer integration times. Figure 4 gives an illustration of all concepts: The left figure shows a blue curve of interest $\mathbf{x}(t)$ on which we consider a particular red point $\mathbf{x}(t_0)$ for illustration. We want to compute the covariance $\mathbf{C}_i(t_0)$ for this particular point. For this, we sample 5 points $\mathbf{x}(t_0-2\Delta t)$, $\mathbf{x}(t_0-\Delta t)$, $\mathbf{x}(t_0)$, $\mathbf{x}(t_0+\Delta t)$, $\mathbf{x}(t_0+2\Delta t)$ on $\mathbf{x}(t)$ for some $\Delta t > 0$ (black points, red point). From these 5 sample points we integrate the vector field $\mathbf{v}_i$ with the integration times $2\Delta t, \Delta t, 0, -\Delta t, -2\Delta t$, respectively, resulting in 5 end points (blue points, red point). The right side of Figure 4 shows these end points without seeds.

The main idea of our approach is to find their mean $\mathbf{m}_i(t_0)$ (green

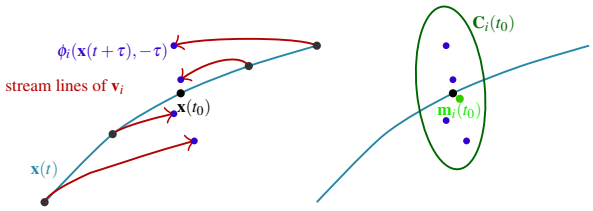


**Figure 4:** *Computation of point on mean curve* $\mathbf{m}_i(t_0)$ *and covariance function* $\mathbf{C}_i(t_0)$ *for a curve of interest* $\mathbf{x}(t)$ *at a particular point* $\mathbf{x}(t_0)$*. Black points are samples on* $\mathbf{x}(t)$*. Red curves are stream lines of* $\mathbf{v}_i$*. Blue points are end points of integration of* $\mathbf{v}_i$ *starting from black points. Green point: mean* $\mathbf{m}_i(t_0)$ *of end points. Green ellipse: their covariance* $\mathbf{C}_i(t_0)$*.*

point) and the covariance $\mathbf{C}(t_0)$ which is depicted as green ellipse here.

**Properties of variance:**

- The curve $\mathbf{x}(t)$ is a perfect stream line of $\mathbf{v}_1,\dots,\mathbf{v}_n$ iff $\mathbf{C}(t)\equiv\mathbf{0}$. This follows directly from $\mathbf{d}_i(t,\tau)\equiv\mathbf{0}$ if $\mathbf{x}(t)$ is a stream line of $\mathbf{v}_1,\dots,\mathbf{v}_n$.
- Invariance to time shift. Let $\tilde{\mathbf{x}}(t)=\mathbf{x}(t+\tilde{t})$ be a time shift reparametrization of of $\mathbf{x}(t)$ in $[t_s-\tilde{t},t_e-\tilde{t}]$. Then we obtain for the covariance function

$$\tilde{\mathbf{C}}(t)=\mathbf{C}(t+\tilde{t})\ . \tag{5}$$

  This is due to $\tilde{\mathbf{d}}_i(t,\tau)=\mathbf{d}_i(t+\tilde{t},\tau)$, which follows directly from the definition of $\tilde{\mathbf{x}}(t)$.

This latter equation (5) describes an important property, which is vital to our approach: All points on $\mathbf{x}(t)$ contribute in the same way to computing $\mathbf{C}(t)$. There is no fixed point of interest and thus no bias or artifacts from such choice. This property ensures that our technique is truly *curve oriented* rather than point oriented.

### 3.3. Generalization to uncertain path lines

So far, we considered uncertain stream lines in ensembles of steady vector fields. The generalization to uncertain path lines in ensembles of *unsteady* vector fields is straightforward. Given the $m$-dimensional time-dependent vector fields $\mathbf{v}_1(\mathbf{x},t),\dots,\mathbf{v}_n(\mathbf{x},t)$ and the time-dependent field of interest $\mathbf{w}(\mathbf{x},t)$, we consider the $(m+1)$-dimensional steady fields [TWHS05] for points $\overline{\mathbf{x}}=(\mathbf{x},t)\in\mathbb{R}^{m+1}$ in space-time:

$$\overline{\mathbf{v}}_i(\overline{\mathbf{x}})=\begin{pmatrix}\mathbf{v}_i(\mathbf{x},t)\\1\end{pmatrix}\quad\text{and}\quad\overline{\mathbf{w}}(\overline{\mathbf{x}})=\begin{pmatrix}\mathbf{w}(\mathbf{x},t)\\1\end{pmatrix} \tag{6}$$

and apply our approach described in the previous section 3.2 to $\overline{\mathbf{v}}_i(\overline{\mathbf{x}})$ and $\overline{\mathbf{w}}(\overline{\mathbf{x}})$. This gives the $(m+1)$-dimensional covariance field $\overline{\mathbf{C}}(\overline{\mathbf{x}})$ with the following properties: The last line and column of $\overline{\mathbf{C}}(\overline{\mathbf{x}})$ is vanishing. This follows from the fact that the last component of $\overline{\mathbf{v}}_i$ and $\overline{\mathbf{w}}$ is constant 1 and applying (2) and (3). Hence, the desired $m$-dimensional covariance field $\mathbf{C}(\mathbf{x},t)$ is encoded as matrix blocks in

$$\overline{\mathbf{C}}(\overline{\mathbf{x}})=\begin{pmatrix}\mathbf{C}(\mathbf{x},t) & \mathbf{0}\\ \mathbf{0} & 0\end{pmatrix}\ . \tag{7}$$

## 4. Computation and visual representation

Given the ensemble members $\mathbf{v}_1,\dots,\mathbf{v}_n$ and the field of interest $\mathbf{w}$, For computing $\mathbf{C}(\mathbf{x})$ for a particular grid point, we require numerical integration of (2)–(3). A simple quadrature is sufficient, with $k$ equidistant time samples $[t_s,t_e]$ and apply the midpoint rule. For vector field integration and computation of the flow map, we use a fourth-order Runge-Kutta scheme with adaptive time step.

Computing $\mathbf{C}(\mathbf{x})$ involves multiple parameters, which need to be discussed: the integration time $\tau$ steers the length of the stream lines of $\mathbf{w}$ that are analyzed. The longer $\tau$, the more global is our approach. The number of samples $k$ steering the resolution of the resulting uncertain path lines. Depending on the length of the seeding curve higher values create smoother tubes as this relates to the number of ellipses used. For the examples in the paper, we use $k=31$ samples while 15 were already sufficient for the GEFS and Red Sea dataset. For all examples we used $\tau=2$ unless stated otherwise.

### 4.1. Visual representation

We need to show a stream line of interest $\phi(\mathbf{x},\tau)$ of $\mathbf{w}$ within the time interval $[t_s,t_e]$ with its covariance function $\mathbf{C}(\phi(\mathbf{x},\tau))$. For the visualization of the covariance $\mathbf{C}$, we first encode individual samples $\mathbf{C}(\phi(\mathbf{x},\tau))$ at $\tau$ as ellipse/ellipsoid centered at $\mathbf{x}(\phi(\mathbf{x},\tau))$. These ellipsoids can represented implicitly by

$$\mathbf{x}^T\mathbf{C}^{-2}\mathbf{x}-1=0 \tag{8}$$

or explicitly by

$$\mathbf{r}\to\mathbf{C}\,\mathbf{r} \tag{9}$$

with parameter points $||\mathbf{r}||=1$ on the unit circle/sphere. For varying $\tau$, this results in a sweeping ellipsoid that generates a tube-like structure, which can be interpreted as a generalized canal surface. While the general and precise computation of canal surfaces is non-trivial [PP97], we found that a fairly simple discretization scheme suffices for the purpose of visualization: We take $k$ equispaced time samples in $[-t_s,t_e]$ and compute covariance matrices $\mathbf{C}_1,\dots,\mathbf{C}_k$ and approximate the canal surface by pairwise convex hulls

$$\bigcup_{i=1}^{k-1}\mathrm{conv}(\mathbf{C}_i,\mathbf{C}_{i+1})\ , \tag{10}$$

where conv denotes the convex hull of the two ellipsoids. An alternative visual representation consists in a spaghetti plot of all stream lines starting from $\phi(\mathbf{x},\tau)$ in all ensemble members. While this visualization may give a good overview of a stream line and its uncertainty, it has two potential drawbacks: Firstly, it suffers from visual clutter, as $n\cdot k$ lines have to be drawn close by each other. Secondly, the sampling density on the curve of interest – steered by $k$ – influences the visual representation in different ways. While increasing $k$ increases the numerical accuracy of $\mathbf{C}$, it has a contrary effect on the visualization. This is due to the increase of visual clutter in the spaghetti plots arising from the increased the number of curves displayed.

With these ingredients, several exploration scenarios are possible. One is to interactively move a single stream line of interest. Another one is considering multiple such stream lines simultaneously in the desired area.

### 4.2. Parameters

The presented algorithm has some parameters. We use $k$ as the number of equidistant samples along the domain $[t_s, t_e]$ to discretize the integral in equation 3. To create a seeding structure in the form of a stream line, a position in space and time $\mathbf{x}$ is needed. For the visual representation, the ellipsoids have to be sampled at discrete points. A sufficient amount of points in the examples stated here were 15 or 31 samples.

## 5. Results

We evaluate our method for three flow ensembles: an analytic flow of two sinks with a saddle, a numeric ensemble simulation of current in the Red Sea and a worldwide weather forecast. As the analytic flow ensemble consists of steady flows only, we are computing uncertain streamlines. In contrast the numeric ensemble flows are unsteady which enables us to compute uncertain path lines. The Red Sea data set was generated by Toye et al. [TZG*17] and was used as a benchmark in the IEEE SciVis contest in 2020. It spans 30 days and consists of 50 high resolution ensemble members with a total size of of 1.5 TB.

### 5.1. Two Sinks with Saddle

Before we demonstrate this technique on unsteady flows like the ocean simulation of the Red Sea or wind prediction by the Global Ensemble Forecast System, we would like to introduce a steady ensemble data set. This data set consists of two sinks with a saddle in the middle with a slight offset in each ensemble member:

$$\mathbf{v}_1(\mathbf{x}) = \begin{pmatrix} x(1-x)(1+x) \\ -y \end{pmatrix} \tag{11}$$

$$\mathbf{v}_2(\mathbf{x}) = \begin{pmatrix} (x+\frac{1}{10})(1-x)(1+x) \\ -y \end{pmatrix}$$

$$\mathbf{v}_3(\mathbf{x}) = \begin{pmatrix} (x-\frac{1}{10})(1-x)(1+x) \\ -y \end{pmatrix}$$

Figure 5 shows a set of uncertain stream lines at the top. The arrows show the mean of the vector fields. The dark green ellipsoids are very thin due to the identical $y$ component of all ensemble members. The uncertain stream lines are all seeded at the same height indicating uncertainty and flow behaviour at different regions. The center one is enhancing the saddle structure and its uncertainty due to the shifted critical point. The uncertain stream lines to the left and right become thinner as they approach the sink, as all ensemble flows have the same sink positions. This also show that uncertain stream lines are well defined near critical points. The bottom of Figure 5 shows an example with the former method with cones for comparison. It can be observed, that all cones diverge with increased integration time but the very right one. This creates overlapping and makes it therefore more challenging to understand the flow. The cones are also only a representation for the exact seeding point without any information for its close proximity. A minor advantage is less overlap of the uncertain stream line while its major advantage is its retained dimensionality. With lower dimensionality our method reduces the search space by one and enables the explorer to use less uncertain stream lines to cover the same area of interest. All uncertain path lines in this section are computed with a 90% confidence interval.

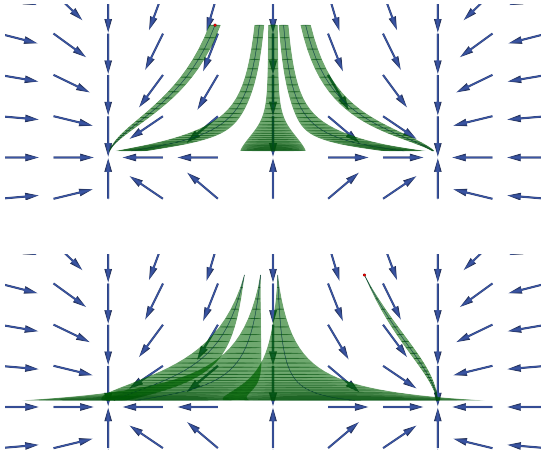

**Figure 5:** *Uncertain stream lines with $k = 31$ for vector fields $\mathbf{u}_1(\mathbf{x})$, $\mathbf{u}_2(\mathbf{x})$ and $\mathbf{u}_3(\mathbf{x})$ with different methods. The quiver plot shows the normalized flow direction of $\mathbf{u}_1(\mathbf{x})$. Top: uncertain stream lines with multiple seeding structures covering the domain. Bottom: old method of uncertain stream lines. The dark green ellipsoids clarify the underlying geometry which is used to compute the tube-like structure.*

### 5.2. GEFS Data Set

The second data set stems from the Global Ensemble Forecast System (GEFS) by the National Oceanic and Atmospheric Administration. It consists of 21 ensemble members, which span 15 days and cover the whole world. We take a closer look at a subset, covering the Indian Ocean and compare the former method 6 with ours 7. The integration time of all path lines in the top images of both figures are four days. Path lines are color coded by their ensemble member and rendered half transparent in order to make all of them visible. The bottom image of figure 6 shows cone-like uncertain path lines of the former method. They only give information about the flow behavior originating from a single starting point. The top of figure 7 shows the spaghetti plot for all path lines seeded by the red curves. We use the mean curves of the former method as seeding curves for our method. There are now $21 \cdot k$ path lines for every seeding curve resulting in an even more cluttered visualization. We used $k = 15$ here. The image below shows the corresponding uncertain path lines of our method, visualizing tubes that represent the expansion of path lines around the mean curve. The density of the transparent path lines may indicate the outlines of the tubes below. It can be observed that most of the tubes, especially above the Indian Ocean stay rather thin, indicating similar flow behavior among the ensemble members. Nevertheless there are some cone like structures above east of Africa with our method but in the inverse direction in comparison to the former method. This indicates converging instead of diverging behavior. Cone-like structures in our method are therefore a result of changing similarity in flow behavior rather than being an artifact of increasing integration time. Due to the increased number of underlying path lines, outliers have less impact, resulting in overall smaller tubes and less overlap of tubes. Therefore more information, higher reliability and less overlap can be archived by our method.

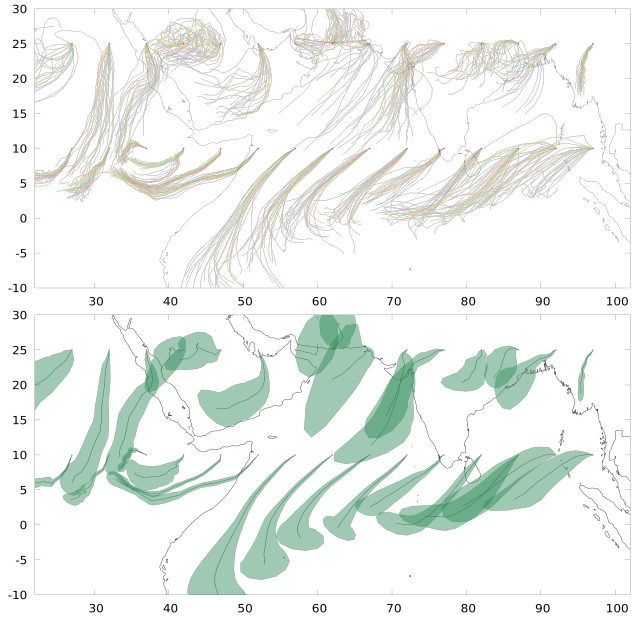

**Figure 6:** *Top: Spaghetti plots of wind flow above the Indian Ocean based on the red seeding points. Bottom: Uncertain path lines visualized based on the path lines from above resulting in cone-like structures.*

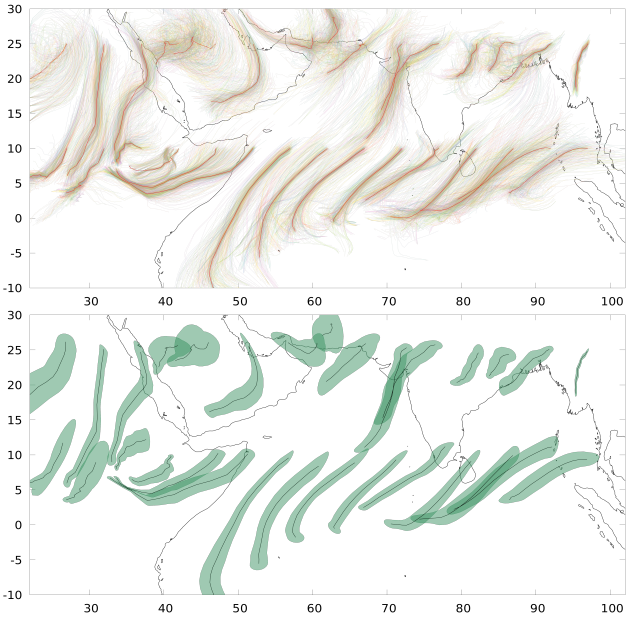

**Figure 7:** *Top: Spaghetti plots of wind flow above the Indian Ocean based on the red seeding curves. Bottom: Uncertain path lines visualized based on the path lines from above resulting in tube-like structures.*

### 5.3. Red Sea

The Red Sea data set contains 50 ensemble members. Therefore the spaghetti plots and uncertain path lines are more cluttered than in the GEFS data set. We use an integration time $\tau$ of three days. We take a closer look at the Gulf of Aden as it is features multiple eddies. The former method in figure 8 and our method in figure 9 are seeded on a regular grid. The seeding curve for our algorithms is again the mean curve from the former method in order to create a comparable covering of the domain. Both spaghetti plots overlap strongly, so that it is almost impossible to assign a single path line to a seeding point. With more path lines along the seeding curve, tubes tend to be smaller with our method, making it easier to recognize eddies and their extent.

### 5.4. Performance

All algorithms were implemented in C++ using the VTK and CGAL libraries. As every data set consists of different numbers of ensembles and differences in the representation of the vector field, we measured the computation timings separately for each data set. The former method as well as our method were implemented with the use of pre-calculated flow maps. Table 1 shows the dimensions used for the Red Sea and Indian Sea Flows and their corresponding times to precalculate the flow maps. We do have to add, that our approach does need a second flow map for the back integration. Therefore our method takes double the effort and time to get the pre-calculated flow maps. With the pre-computed flow maps available, all visualizations can be computed in less than 5 ms. This allows for interactive visualizion of uncertain stream and path lines within the pre-calculated domain.

| Flow | $n$ | lat | lon | $k$ | $\tau$ | Time |
|---|---|---|---|---|---|---|
| Two Sinks with Saddle | 3 | 100 | 50 | 31 | 2 | 850 ms. |
| Red Sea | 50 | 23 | 35 | 15 | 72 h | 32 min. |
| Indian Sea | 21 | 100 | 50 | 15 | 72 h | 2 h 8 min. |

**Table 1:** *Flow maps pre-calculation time measurements for the Red Sea and Indian Sea. All measurements were taken on a Ryzen 9 3950x 16-core processor.*

## 6. Discussion

*Gaussian distribution:* Our approach assumes a Gaussian distribution of the stream lines. This is a popular standard assumption that is common and widely covered in the literature. This is the reason for our choice. We are aware that the assumption does not always hold, and alternative distributions may be more appropriate or even required. However, in our opinion a discussion of different distributions and adaptation of our model is beyond the scope of this work.

*Relation to existing approaches:* Our approach is related to existing approaches in several ways. On the one hand, we can modify our approach from a curve oriented one to a point oriented one by

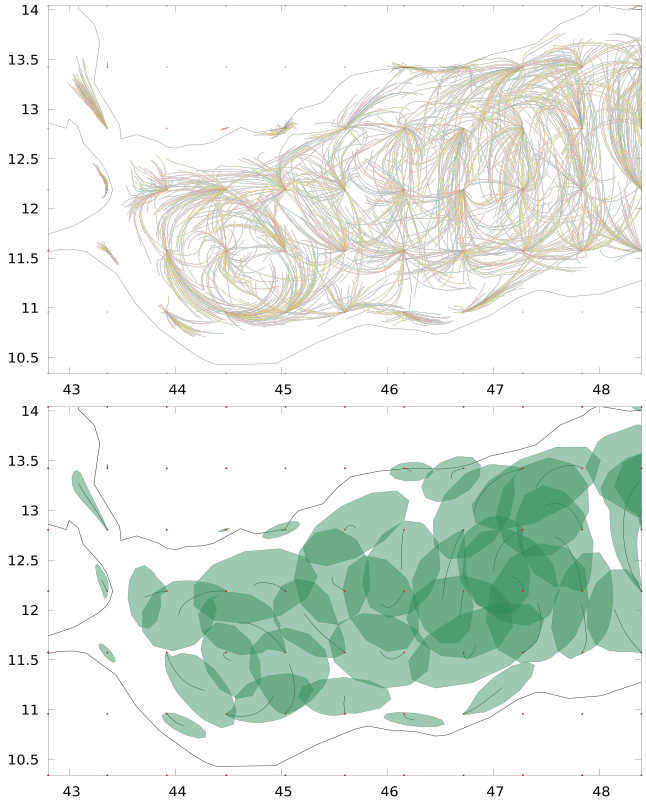


**Figure 8:** *Top: Spaghetti plots of ocean movements in the Gulf of Aden based on the red seeding points. Bottom: Uncertain path lines visualized based on the path lines from above resulting in cone-like structures.*

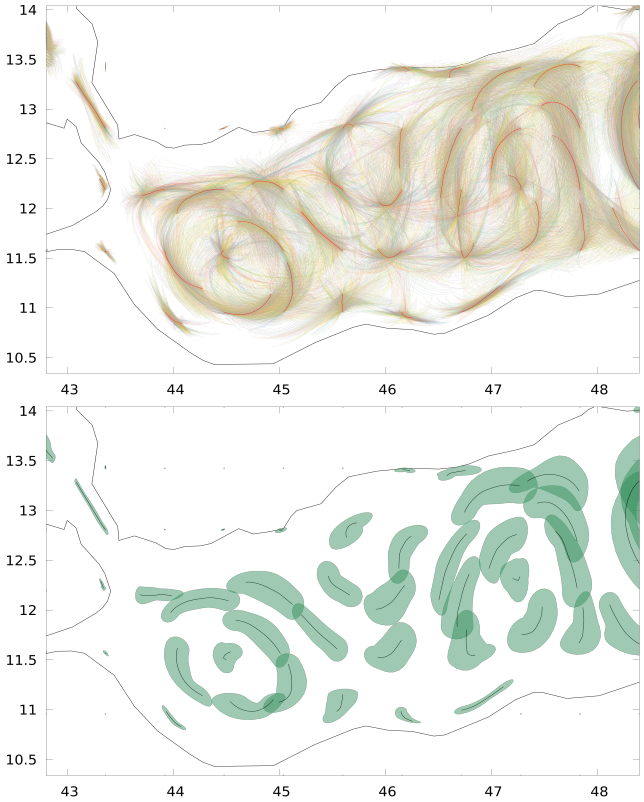


**Figure 9:** *Top: Spaghetti plots of ocean movements in the Gulf of Aden based on the red seeding curves. Bottom: Uncertain path lines visualized based on the path lines from above resulting in tube-like structures.*

considering a special field of interest: for $\mathbf{w}(\mathbf{x}) = \mathbf{0}$, all curves of interest are actually points of interest , resulting in $\mathbf{d}_i(t,\tau) = \phi_i(\mathbf{x},-\tau)$, e.g., a stream line of $\mathbf{v}_i$ starting at $\mathbf{x}$. With this, $\mathbf{m}(\mathbf{x})$ , $\mathbf{C}(\mathbf{x})$ collect information only from all stream lines starting from $\mathbf{x}$, similar to point-oriented techniques. On the other hand, most existing approaches that we classified as point oriented are generic in the sense that they require a set of curves in a common parametrization as input. Instead of considering only stream lines starting from a common point (as done in [MWK14], [FBW16], [WHLS18]), one may feed the algorithms with spaghetti plots constructed by our approach.

*Alternative design decisions:* Instead of designing a global approach as done here, an alternative but similar approach is a more local and thus faster: instead of integrating stream lines from $\mathbf{x}(t)$, only the flux of the ensemble members across $\mathbf{x}(t)$ is considered. With this, the result would only depend on $\mathbf{v}_i$ on the stream line of $\mathbf{w}$ through $\mathbf{x}$. At first glance, this seems attractive. However, we provide a simple counter-example to show that this can give different information.

We consider another simple example consisting of two families of three ensemble members with constant parameters $\alpha$ and $p$:

$$\mathbf{v}_1 = \begin{pmatrix} \cos\alpha \\ \sin\alpha\,\cos\pi y \end{pmatrix},\ \mathbf{v}_2 = \begin{pmatrix} \cos\alpha \\ -\sin\alpha\cos\pi y \end{pmatrix},\ \mathbf{v}_3 = \begin{pmatrix} p \\ 0 \end{pmatrix} \quad (12)$$

and

$$\overline{\mathbf{v}}_1 = \begin{pmatrix} \cos\alpha \\ \sin\alpha\,\sqrt{e^{\pi|y|}} \end{pmatrix},\ \overline{\mathbf{v}}_2 = \begin{pmatrix} \cos\alpha \\ -\sin\alpha\,\sqrt{e^{\pi|y|}} \end{pmatrix},\ \overline{\mathbf{v}}_3 = \begin{pmatrix} p \\ 0 \end{pmatrix} \quad (13)$$

For both ensembles in (12) and 13) we analyze the curve of interest

$$\mathbf{x}(t) = \begin{pmatrix} \frac{1}{3}\,(p+2\cos\alpha) \\ 0 \end{pmatrix}, \quad (14)$$

which is a stream line of both $\frac{1}{3}(\mathbf{v}_1+\mathbf{v}_2+\mathbf{v}_3)$ and $\frac{1}{3}(\overline{\mathbf{v}}_1+\overline{\mathbf{v}}_2+\overline{\mathbf{v}}_3)$. For this example, the flux of $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3$ over $\mathbf{x}(t)$ is identical to the flux of $\overline{\mathbf{v}}_1,\overline{\mathbf{v}}_2,\overline{\mathbf{v}}_3$, respectively. Our approach gives different results for the ensembles in (12) and (13).

## 7. Conclusions and Future Work

We introduced an approach to representing uncertain stream lines in ensemble of vector fields by considering curves of interest instead of points of interest. This way, we obtain global representations with the same dimensionality as "certain" stream lines. There are different roads for future research:

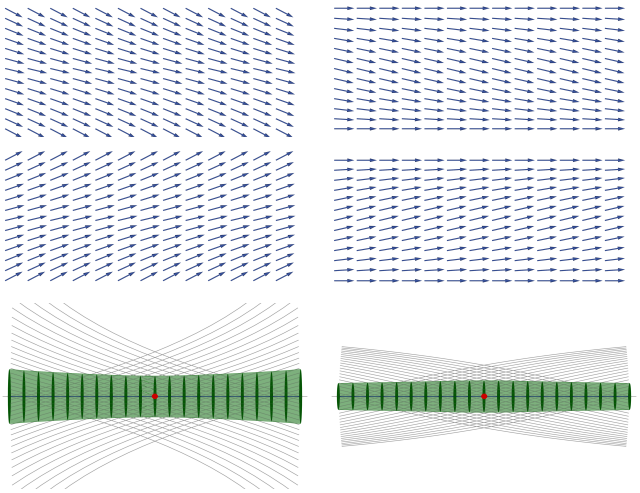

**Figure 10:** *Counter-example: A more local approach would not work. Top left row: ensemble members (13). Top right row: ensemble members (12). Bottom left: uncertain stream line for 13. Bottom right: uncertain stream line for 12.*

- *Non-Gaussian distributions:* Instead of a Gaussian distribution, more involved distributions may be included.
- *Alternative representations of sweeping covariance matrices:* Instead of a sweeping ellipsoid representation for $\mathbf{C}(\phi(\mathbf{x},\tau))$, alternative representations such as superquadrics [Kin04] are possible.
- *Uncertain stream surfaces:* While the extension of our approach to uncertain stream surfaces requires the solution of a number of problems, we do not see a fundamental reason that hinders the extension of our approach.

## Acknowledgments

The work was partially supported by DFG grant TH 692/17-1.

## References

[BHJ*14] BONNEAU G.-P., HEGE H.-C., JOHNSON C. R., OLIVEIRA M. M., POTTER K., RHEINGANS P., SCHULTZ T.: Overview and state-of-the-art of uncertainty visualization. In *Scientific Visualization: Uncertainty, Multifield, Biomedical, and Scalable Visualization*, Mathematics and Visualization. Springer, 2014, pp. 1–25. 2

[BWE05] BOTCHEN R. P., WEISKOPF D., ERTL T.: Texture-based visualization of uncertainty in flow fields. In *IEEE Visualization* (2005), pp. 647–654. 2

[BWE06] BOTCHEN R. P., WEISKOPF D., ERTL T.: Interactive visualization of uncertainty in flow fields using texture-based techniques. In *Electronic Proc. Intl. Symp. on Flow Visualization* (2006). 2

[CR05] CHLAN E. B., RHEINGANS P.: Multivariate glyphs for multi-object clusters. In *Proceedings of InfoVis '05* (2005), pp. 141–148. 2

[ESG97] EHLSCHLAEGER C. R., SHORTRIDGE A. M., GOODCHILD M. F.: Visualizing spatial data uncertainty using animation. *Computers in GeoSciences 23*, 4 (1997), 387–395. 2

[FBW16] FERSTL F., BÜRGER K., WESTERMANN R.: Streamline variability plots for characterizing the uncertainty in vector field ensembles. *IEEE Transactions on Visualization and Computer Graphics 22* (2016), 767–776. 2, 7

[GR04] GRIGORYAN G., RHEINGANS P.: Point-based probabilistic surfaces to show surface uncertainty. *IEEE Transactions on Visualization and Computer Graphics 10*, 5 (September/October 2004), 546–573. 2

[GS06] GRIETHE H., SCHUMANN H.: The visualization of uncertain data: Methods and problems. In *Proceedings of SimVis '06* (2006), SCS Publishing House, pp. 143–156. 2

[HCLS16] HE W., CHEN C.-M., LIU X., SHEN H.-W.: A Bayesian approach for probabilistic streamline computation in uncertain flows. In *Proc. of Pacific Visualization* (04 2016), pp. 214–218. 2

[HKR*18] HANSER K., KLEIN O., RIECK B., WIEBE B., SELZ T., PIATKOWSKI M., SAGRISTÀ A., ZHENG B., LUKÁCOVÁ-MEDVIDOVÁ M., CRAIG G., LEITTE H., SADLO F.: Visualization of parameter sensitivity of 2d time-dependent flow. In *Advances in Visual Computing*. Springer International Publishing, 2018, pp. 359–370. doi:10.1007/978-3-030-03801-4_32. 2

[HP20] HOLLISTER B. E., PANG A.: Uncertainty rank for streamline ensembles. *Journal of Imaging Science and Technology 64(2)* (2020). 2

[Jon03] JONES D. K.: Determining and visualizing uncertainty in estimates of fiber orientation from diffusion tensor mri. *Magnetic Resonance in Medicine 49* (2003), 7–12. 2

[Kin04] KINDLMANN G.: Superquadric tensor glyphs. In *Eurographics / IEEE VGTC Symposium on Visualization* (2004), Eurographics Association, pp. 147–154. 8

[KLDP02] KAO D., LUO A., DUNGAN J. L., PANG A.: Visualizing spatially varying distribution data. In *Proceedings of the Sixth International Conference on Information Visualisation, 2002* (2002), pp. 219–225. 2

[LFC02] LODHA S. K., FAALAND N. M., CHARANIYA A. P.: Visualization of uncertain particle movement. In *Proceedings of the Computer Graphics and Imaging Conference* (2002), pp. 226–232. 2

[LKP03] LUO A., KAO D., PANG A.: Visualizing spatial distribution data sets. In *VISSYM '03: Proceedings of the Symposium on Data visualisation 2003* (2003), pp. 29–38. 2

[MWK14] MIRZARGAR M., WHITAKER R. T., KIRBY R. M.: Curve boxplot: Generalization of boxplot for ensembles of curves. *IEEE Transactions on Visualization and Computer Graphics 20* (2014), 2654–2663. 2, 7

[OGHT10] OTTO M., GERMER T., HEGE H.-C., THEISEL H.: Uncertain 2d vector field topology. In *Computer Graphics Forum (Proceedings of Eurographics 2010)* (Norrköping, Sweden, 5 2010). 2

[OGT11] OTTO M., GERMER T., THEISEL H.: Uncertain topology of 3d vector fields. In *Proceedings of 4th IEEE Pacific Visualization Symposium (PacificVis 2011)* (Hong Kong, China, March 2011), pp. 67–74. 2

[PP97] PETERNELL M., POTTMANN H.: Computing rational parametrizations of canal surfaces. *J. Symb. Comput. 23*, 2/3 (1997), 255–266. 4

[RT12] RÖSSL C., THEISEL H.: Streamline embedding for 3d vector field exploration. *IEEE Transactions on Visualiztion and Computer Graphics 18-3* (2012), 407–420. fast track TVCG from IEEE Visualization 2010. 1

[SJK04] SANDERSON A. R., JOHNSON C. R., KIRBY R. M.: Display of vector fields using a reaction-diffusion model. In *IEEE Visualization* (2004), pp. 115–122. 2

[TWHS05] THEISEL H., WEINKAUF T., HEGE H.-C., SEIDEL H.-P.: Topological methods for 2d time-dependent vector fields based on stream lines and path lines. *IEEE Transactions on Visualization and Computer Graphics 11*, 4 (Julyi-August 2005), 383–394. 4

[TZG*17] TOYE H., ZHAN P., GOPALAKRISHNAN G., KARTADIKARIA A. R., HUANG H., KNIO O., HOTEIT I.: Ensemble data assimilation in the red sea: sensitivity to ensemble selection and atmospheric forcing. *Ocean Dynamics 67*, 7 (2017), 915–933. 5

[WHLS18] WANG J., HAZARIKA S., LI C., SHEN H.-W.: Visualization and visual analysis of ensemble data: A survey. *IEEE Transactions on Visualization and Computer Graphics 25* (2018), 2853–2872. 2, 7

[ZDG*08] ZUK T., DOWNTON J., GRAY D., CARPENDALE S., LIANG J.: Exploration of uncertainty in bidirectional vector fields. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* (2008), vol. 6809, p. 68090B. published online. 2