Comparative Visual Analysis of Lagrangian Transport in CFD Ensembles

Abstract—Sets of simulation runs based on parameter and model variation, so-called ensembles, are increasingly used to model physical behaviors whose parameter space is too large or complex to be explored automatically. Visualization plays a key role in conveying important properties in ensembles, such as the degree to which members of the ensemble agree or disagree in their output. For ensembles of time-varying vector fields, there are numerous challenges for providing an expressive comparative visualization, among which is the requirement to relate the effect of individual flow divergence to joint transport characteristics of the ensemble. Yet, techniques developed for scalar ensembles are of little use in this context, as the notion of transport induced by a vector field cannot be modeled using such tools. We develop a Lagrangian framework for the comparison of flow fields in an ensemble. Our techniques evaluate individual and joint transport variance and introduce a classification space that facilitates incorporation of these properties into a common ensemble visualization. Individual and joint variances of Lagrangian neighborhoods are computed using pathline integration and Principal Components Analysis. This allows for an inclusion of uncertainty measurements into the visualization and analysis approach. Our results demonstrate the usefulness and expressiveness of the presented method on several practical examples.

Index Terms—Ensemble, flow field, time-varying, comparison, visualization, Lagrangian, variance, principal components analysis

1 INTRODUCTION

Recent increases in computation power and the availability of hardware-enabled high performance parallel computing have made it feasible to run numerical simulations repeatedly for large sets of different parameters in reasonable amounts of time. The purpose of such an approach to simulation is to overcome the complexity of real-world physical phenomena. The underlying assumption is that the insight gained from multiple alternative mathematical models allows for a more complete understanding of the real-world phenomenon and how it relates to computational models. Each simulation run in one of these so-called ensembles is a representation of such a single conceptual or computational model. Given this set of possible solutions, experts are able to perform a more efficient search for optimal simulation parameters or construct new models based on insights gained from the analysis of individual and joint simulation behaviors. In short, ensemble visualization techniques aim at expediting the exploration of parameter and model-space.

Ensembles of time-varying vector fields, as generated by Computational Fluid Dynamics (CFD) simulations, pose new challenges to the field of comparative visualization. For ensembles of (stationary) scalar fields, a localized Eulerian comparison can provide appropriate means for comparative analysis. In vector fields, however, the property of central interest, namely the transport of material, is a Lagrangian characteristic of the field. Here, local comparisons are complicated by the wide range of diverging and converging flow behaviors that may be present in individual fields. An analysis of this type of behavior therefore requires the development of novel approaches to efficient ensemble visualization.

In this paper, we introduce a Lagrangian approach to the comparison of vector-field ensembles. We develop a classification strategy that allows for the identification of common and contradictory trends in Lagrangian flow behavior across members of an ensemble. Our approach examines the behavior of Lagrangian neighborhoods in individual simulation runs and relates those to joint behaviors present in the ensemble. In both cases we make use of dense pathline integration techniques, as known from fields such as Finite Time Lyapunov Exponent (FTLE) analysis, and analyze properties of the advected neighborhood using Principal Components Analysis (PCA). We argue that only by connecting transport characteristics of individual Lagrangian neighborhoods to joint flow behavior present in the ensemble, is it possible to make reliable statements about the comparability of time-varying flow fields.

We provide results to show how our techniques enable a robust visual comparison of the transport behaviors present in flow field ensembles, and provide means to identify prevalent deterministic transport
behaviors. We furthermore demonstrate the effectiveness of interaction and querying techniques for such a comparison in a selection of 2D and 3D time-varying benchmark data sets. In summary, this work makes the following contributions

- **Computation of individual and joint vector field variance:** We provide a definition of generalized Lagrangian neighborhoods in flow fields that allows the computation of variance-based flow divergence for arbitrary scales. Furthermore our work establishes a notion of flow divergence for multiple fields. The proposed joint vector field variance can be seen as a form of multi-field FTLE.

- **Definition of Lagrangian comparability:** We propose a classification space for robust comparison of Lagrangian transport by connecting intra- and inter-vector field transport variance.

- **A Lagrangian framework for the comparison of time-varying vector fields in an ensemble:** Together with trends analysis, interaction, and comparability classification, this paper presents a framework for Lagrangian transport in CFD ensembles.

The manuscript proceeds with a presentation of related work and definitions in sections 2 and 3. We then detail Lagrangian properties and our approach to comparing transport in time varying flow fields in Section 4. Section 5 presents our work on comparative visualization of vector fields in an ensemble. Sections 6 and 7 present results of our work and conclusions.

2 RELATED WORK

Our work on the visual comparison of multiple fields is related to research in a number of topics such as uncertainty visualization, ensemble visualization, and general statistics. In the following we review closely related work from these fields. Statistical and stochastic descriptions of fields or data with numerical errors is often regarded as data with uncertainties [7] and visualized accordingly. Pang et al. [19] and Griethe and Schumann [8] give an overview of visualization techniques such as error-bars, pseudo-coloring, and geometry generation. Similarly, Johnson and Sanderson [10] investigate existing and future visualization techniques for uncertain data, suggesting improved interaction with statistical data.

Flow Visualization For fluid flows, dense texture-based visualization techniques of uncertainties in flow fields have proven to be especially suitable for 3D vector fields [1]. Combined with overlaying approaches, they can be used to compare small numbers of flow and scalar fields [26]. Alternatively, geometric features, such as vector glyphs [30] can be extended to encode uncertain information, such as variability in direction and magnitude. Trajectory based comparison techniques [12] allow for the analysis of streamlines generated by different numerical integration schemes. Verma and Pang [27] extend this work by performing pairwise comparisons of streamlines from two different vector fields. They note that distinguishing between errors caused by numerical integration and differences present in flow fields is crucial for effective comparison. The approach presented in this work verifies this observation and generalizes the notion of integration reliability or uncertainty and field differences by distinguishing between joint and individual transport variance in ensembles of time-varying vector fields.

Stationary vector fields may also be compared based on their topological skeletons as demonstrated by Otto et al. [17, 18]. For time varying data sets, however, no definition of a topological decomposition exists. Instead, separating stable and unstable material lines may be extracted over finite amounts of time, by extracting the so called **Finite Time Lyapunov Exponent** (FTLE) field [9]. Extremal values of this field indicate that flow trajectories seeded within an infinitesimal neighborhood diverge strongly over time. Semantically, such an FTLE field represents a comparison measure for fluid transport within a spatial neighborhood. In visualization, this type of divergence measure has been extracted from dense sets of trajectories [5], as well as in a localized fashion [11, 15]. Recently, Schneider et al. [24] proposed the use of PCA to compute variance based flow divergence in uncertain vector fields. Our work makes use of the notion of FTLE as a representation of integration reliability and extends PCA based variance estimation to neighborhoods of particles within a single flow field as well as across members of an ensemble.

**Uncertainty and Ensembles** Uncertain data often consists of data samples drawn from a set of simulation runs, and as such is closely related to ensemble data sets, a data type that has recently come into focus of the visualization community. A prime application area for ensemble visualization is climate research, where multiple climate prediction models are used to produce climate forecasts. Potter et al. [21, 22] investigate visual analysis of such ensembles through trends and contour or Spaghetti Plots and color-mapped per-point means and averages [29]. Generalizations of such contour plots are used by Sanyal et al. [23] alongside uncertainty glyphs to analyze scalar-valued weather forecasts. In simulation ensembles, where clear trends in qualitative outcome are detectable, an ensemble clustering step can be beneficial before performing statistical data averaging, as demonstrated by Smith et al. [25]. Our work employs per-point averaging and trends analysis techniques within the context of Lagrangian flow trajectories. As common in ensemble visualization, we also make use of summary statistic plots. A complete comparison of members of an ensemble is a challenging and complex task that requires the implementation of linked components for data comparisons on multiple levels. For 2D scalar field ensembles in powertrain system development, such a complex system may consist of domain, feature, and member-based comparison components [20].

In our work we focus on the comparison of a specific property of time-varying flow fields, namely material transport. Since this property is not directly encoded in the form of scalar fields, we are confronted with the challenge to compare sets of flow trajectories across CFD ensembles. As a consequence, our work extends the state-of-the-art in ensemble visualization to 2D and 3D ensembles of time-varying flow fields.

3 BACKGROUND

The visual analysis of ensemble data sets has gained in popularity over recent years [29]. In this work we focus on the visual comparison of transport behaviors present in an ensemble of vector fields. We presently recapitulate formal definitions of such data and motivate the desire to perform a visual analysis of flow ensembles.

3.1 Vector Field Ensembles A d-dimensional time-varying flow field

\[ v : \Omega \times I \rightarrow \mathbb{R}^d \]

(1)

is a vector-valued function defined over the spatial domain \( \Omega \subseteq \mathbb{R}^d \) and over an interval in time \( I \subseteq \mathbb{R} \). In this paper a flow field ensemble \( E \) is defined as a collection of \( m \) such vector fields defined over a common domain, \( \Omega_E = \Omega_1 \cap \ldots \cap \Omega_m \) and \( I_E = I_1 \cap \ldots \cap I_m \):

\[ E : \{1, \ldots, m\} \times \Omega_E \times I_E \rightarrow \mathbb{R}^d. \]

(2)

We call the first (discrete) parameter dimension of this mapping the \emph{ensemble dimension}. In order to establish meaningful semantics for a vector field comparison, all \( v_i \in E(i, \ldots) \) are assumed to correspond to solutions of models for the same physical phenomenon. Because this work does not aim at explicit parameter optimization, our definition of an ensemble does not require that the simulations used to create individual \( v_i \) share a common parameter space.

3.2 Material Transport

Numerical simulation models for complex physical phenomena involve large sets of simulation parameters. For fluid simulations, typical parameters include physical properties such as boundary conditions (e.g., pressure or temperature), viscosity, stiffness, or other internal forces as well as simulation parameters, such as simulation accuracy or resolution. In order to reproduce a physical behavior, this
parameter space needs to be explored (cf. [2]) until simulation results converge to experimental observations. As simulation behavior in this parameter-space is highly nonlinear, unguided manual exploration of parameter space requires tedious and repeated comparisons between simulation outputs. In order to expedite this process, domain experts increasingly make use of ensemble techniques.

For vector fields, the output of fluid simulations, the most characteristic property is their transport behavior. This transport behavior describes where material is transported, defines (topological) boundaries or transport barriers, regions of flow divergence and convergence, and the deformation and displacement of material elements in general. Consequently, a natural question arising in the visualization of vector field ensembles is, how to extract and visualize similarities and differences in Lagrangian transport. Figure 2 illustrates transport behavior in multiple vector fields and exemplifies the task of Lagrangian comparison. This illustration reveals immediate similarities between transport behavior analysis in an ensemble and divergence analysis for neighborhoods in individual fields. In the following sections we make use of such divergence measures to analyze differences and commonalities of transport behavior in vector fields of an ensemble and present a framework for visual comparison of Lagrangian transport behaviors.

Note that ensemble comparison strategies can generally take two different approaches. The most common approach is data-oriented, and operates on merging or overlapping the domains of individual runs, in combination with a per-point comparison in data space. An alternate approach is to compare members of an ensemble in feature space. Hereby comparisons are not performed on data at a specific location, but over a set of features present in the flow field. Similarities can then be defined by the presence or absence of common feature types, shapes, or properties (see for example [28, 16] for feature based similarity metrics in flow fields). The techniques presented in this paper represent a hybrid method – comparisons are performed in data-space, while relying on the transport property computed from flow trajectories.

3.3 FTLE and Flow Map

The observation that multi-field comparisons of Lagrangian transport are related to divergence measures in single fields is a important insight for the techniques developed in this work. For this reason, the following sections analyze the notion of fluid transport in an individual flow field and existing means for comparison and establish a basis for the subsequent presentation of novel multi-field ensemble comparison techniques.

A time-varying flow field \( \nu \) describes the trajectories \( x \) of infinitesimal particles according to the following differential equation

\[
\frac{dx(t)}{dt} = \nu(x(t), t)
\]

with initial particle position \( x(t_0) = x_0 \). In the following we denote an individual particle or trajectory simply as \( x \) or \( x_i \). Numerical integration of Eq. (3), e.g. using 4th-order Runge-Kutta techniques, allows the explicit computation of particle trajectories. In a Lagrangian setting, flow field properties such as pressure, speed, and particle neighborhoods are explicitly transported along with the flow and stored at such moving particle locations, rather than being position-bound in the form of a Eulerian representation. In the following, we recognize transport as a central Lagrangian property of flow fields.

For a single flow, analysis of transport behavior can be achieved by evaluating such Lagrangian properties over a spatial neighborhood around a flow particle. These comparisons allow for the analysis of differences in particle behavior and are detailed in the following sections.

The best-studied property of Lagrangian neighborhoods is the tendency of particle neighborhoods to diverge or converge over time. A measure of this characteristic is given by the Finite Time Lyapunov Exponent (FTLE), which measures exponential stretching of infinitesimal particle neighborhoods over finite intervals in time. This measure can be computed directly from the gradient of a displacement map, the so-called flow map \( \phi \), of a vector field. Given such a displacement mapping between initial and advected positions

\[
\phi(x(t); T) = x(t+T),
\]

the FTLE value at \( x(t) \) for advection time \( T \in \mathbb{R} \) is given as the logarithm of the largest singular value of \( \nabla \phi \) (which is identical to the square root of the largest eigenvalue of the Cauchy Green tensor \( \nabla \phi^T \nabla \phi \)):

\[
FTLE\left(x(t), T\right) = \frac{1}{T} \log \sqrt{\lambda_{\text{max}}(\nabla \phi(x(t); T)^T \nabla \phi(x(t); T))}.
\]

In summary, during FTLE computation magnitudes and directions of maximal deformation of an infinitesimal neighborhood can be obtained directly from the eigenvalues and eigenvectors of the left Cauchy Green tensor of \( \nabla \phi \) [15]. Thus, the FTLE is a suitable measure to compute relative stretching and divergence of an infinitesimal neighborhood in a single flow field; however, a generalization of this approach to ensembles is not straightforward.

4 LAGRANGIAN COMPARISON OF FLOW ENSEMBLES

A possible avenue for comparative transport analysis in flow ensembles is a generalization of Lagrangian measures over particles neighborhoods, which we examine in the following.

4.1 Divergence of Generalized Neighborhoods

Let the domain of \( v \) be densely sampled by a set of \( n \) flow particles \( x_i \in \{x_1, \ldots, x_n\} \). Let \( x \) be a specific such flow particle. At any point in time \( t \in \mathbb{R} \), we can define a generalized neighborhood for \( x \) as follows:

\[
N_x(t) = \{x_i \in \{x_1, \ldots, x_n\} \mid P(x(t), x_i(t)) = true\}
\]

where \( P \) is a neighborhood predicate that indicates local adjacency. An example of a spatially isotropic neighborhood predicate \( P \) with radius \( r \in \mathbb{R} \) is

\[
P(x(t), x_i(t)) = \begin{cases} 
1 & \text{if } \|x(t) - x_i(t)\| \leq r \\
0 & \text{otherwise}
\end{cases}
\]

Note that these types of spatial neighborhood predicates allow the discussion of an implied neighborhood scale based on the choice of \( r \). In Section 4.2, we give neighborhood predicates for ensembles that also include non-spatial dimensions. In this generalized definition, the Lagrangian neighborhood is no longer limited to particles that are immediate spatial neighbors of \( x \), but may be a scattered across larger regions in space.

FTLE The applicability of FTLE techniques is directly influenced by properties of this neighborhood. FTLE computations are immediately dependent on the local linearization of the deformation map, i.e., the computation of the flow map gradient \( \nabla \phi \), or the velocity gradient tensor [15]. For this reason, the Cauchy Green tensor derived during FTLE computations is only accurate, if the deformation of the Lagrangian neighborhood can be linearized with sufficient accuracy. For generalized neighborhoods, i.e., particle neighborhoods that extend over a set of simulations or over regions in space, the function...
describing fluid deformations of space tends to become increasingly nonlinear and therefore unsuitable for the direct application of FTLE techniques, as shown in Figure 3. Note also that while the flow map may be averaged over larger regions of space to obtain FTLE values for spatially large neighborhoods, flow map gradients cannot be computed across multiple vector fields due to a lack of a metric/ordering for the ensemble dimension.

PCA An alternative to computing the linearized deformation or shape change of a neighborhood is to measure geometric or statistical properties of the neighborhood after deformation. Note that these properties are in general absolute measurements of neighborhood properties and stand in contrast to the relative notion of deformation given by FTLE. However, if the statistical and geometric properties of the neighborhood definition in Eq. (6) are constant over the complete data, these properties can provide a notion of certain relative neighborhood changes. Let in the following $\phi(N_t(t); T) = \{x_i(t+T) \mid x_i \in N_t(t)\}$ denote the set of particle positions obtained by advecting all particles in a neighborhood $N_t(t)$ for a time of $T$.

Principal Components Analysis allows the extraction of extremal directions of this set of displaced point positions. PCA computes the eigenvectors and eigenvalues, i.e., directional variances, of the covariance matrix

$$\Sigma(\phi(N_t(t); T))_{ij} = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij}(t+T) - \mu_i) \cdot (x_{ij}(t+T) - \mu_j)$$

where $n = |N_t(t)|$, $x_{ij}$, and $x_{jk}$ are the $i$th and $j$th component of a vector indicating particle position. The vector $\mu$ represents the mean of all positions in $\phi(N_t(t); T)$. By construction, eigenvectors of $\Sigma$ correspond to directions that minimize the squared distance to all points in the point set. Additionally, eigenvalue magnitudes correspond to directional variances. PCA can be viewed as an approximation technique for a multidimensional Gaussian distribution and is therefore a purely statistical measure. This implies that there are several conceptual differences between PCA and Cauchy Green tensor based spectral decompositions: Rather than representing a relative change in displacement, it measures characteristics of the final neighborhood distribution. PCA is therefore identical for arbitrary permutations of particle positions. Note that there are special cases, where outliers in $\phi(N_t(t); T)$ can distort PCA values (see for example [14]). We make use of PCA in the remainder of this work for its capability to give a measure of flow divergence and represent statistical variance of advected point clouds at the same time.

We propose to combine the presented definitions of generalized Langrangian neighborhoods and PCA into a novel technique for the Lagrangian comparison of time-varying flow fields in an ensemble $E$. In the following we do not distinguish steady from unsteady flow fields but note that flow field comparison for steady fields can generally rely on simpler techniques, such as a comparison of topological graphs.

4.2 Lagrangian Neighborhoods for Ensembles

As indicated previously, defining a comparison technique that incorporates transport behaviors of multiple flow fields requires the extension of classic Langrangian neighborhoods across multiple domains. We propose to use the following neighborhood predicate as a neighborhood definition for a complete ensemble

$$P_E(x(t), x_i(t)) = \begin{cases} 1 & \text{if } \|x^E(t) - x^{E_i}(t)\| \leq r \\ 0 & \text{otherwise} \end{cases}$$

as required by Eq. (6). The main difference to Eq. (7) is that particles $x_i$ and $x$ may belong to different ensemble members. The positions $x^E(t)$ correspond to particle positions projected into the common ensemble domain. This essentially defines a radial neighborhood spanning across all domains $\{\Omega_1, \ldots, \Omega_m\}$, treating particles $x_i$ and $x$ as if they originated from the same simulation run. Thus, the predicate results in neighborhoods containing exactly $m$ particles (one for each run) for a radius of zero. We denote a Lagrangian ensemble neighborhood as $N_E$.

This neighborhood definition treats all ensemble members equally by collapsing the ensemble dimension.

4.3 Naive Transport Comparison

Given our ensemble neighborhood definition together with PCA-based variance estimation, we are able to make a direct comparison of transport properties of multiple flow fields. Analysis of Lagrangian flow field comparison based on an ensemble neighborhood with radius $r = 0$, a point-based comparison, reveals the following insights: For a given advection time $T$ and a specific point $x$ in $\Omega$, PCA of $\Sigma(N_{E}(x); T)$ measures the variance in advection across all ensemble members. A large maximal eigenvalue of $\Sigma$ indicates that the compared vector fields disagree in the final advected point position. As a consequence, the agreement of ensemble members in a neighborhood $N_{E}(x)$ is measured by its joint variance $Var_E(N_{E}(x); T) = _{\Sigma}(\Sigma(\phi(N_{E}(x); T)))$. While this is a valid measure of inter-field variance, there are multiple concerns with respect to expressiveness and utility.

1. Outliers and grouping/trends are not detected.
2. Susceptibility w.r.t. simulation noise.
3. Susceptibility w.r.t. accuracy of integration technique.

The last two points are mainly due to the fact that intra-field variance is neglected during comparison. The following example illustrates this problem:

Let $v_i$ be a set of vector fields originating from the same simulation model and with identical parameter settings. Let there be minimal noise in the data due to differences in numerical precision, varying hardware environments, or due to indeterministics caused by programming errors. Let further $x = (x_0, t_0)$ be the location of the point considered for Lagrangian comparison.

Assuming for simplicity that $v_i$ are truly identical ($v_i = v$ for all $i$), we can model this noise by applying a random infinitesimal offset vector $e_i$ to $x_0$ in every simulation. The displacement of every $(x_0 + e_i, t_0) \in N_{E}(x)$ with respect to $(x_0, t_0)$ is then

$$\phi((x_0 + e_i, t_0); T) = \phi(x_i(t_0); T) + \nabla \phi(x_0(t_0); T) \cdot e_i$$

Here, the maximal value is used in the computation of FTLE$(((x_0, t_0), T))$. Consequently, the eigenvalue magnitudes of $\Sigma(\phi(N_{E}(x); T))$, the joint ensemble variances, are directly dependent on FTLE$(((x_0, t_0), T))$, or the variances of individual fields, $\Sigma(\phi(N_{E}(x); T))$.

We conclude that there is an immediate relationship between single field divergence/variance and joint ensemble variance, when numerical inaccuracies, as caused by noise or inaccurate particle tracing [13], come into play. In summary, large joint variances may not accurately describe fundamentally different transport behaviors in the field, if high individual variances are present. This insight is central to the development of our joint variance classification scheme developed in this paper and similar to observations made by Verma and Pang [27].
4.4 Variance Classification

We conclude from these observations that a one-dimensional space, e.g., one based on joint variance values, is insufficient for reliable comparison and instead utilize a two-dimensional classification space.

To enable a complete comparison of transport behaviors in a vector field ensemble, we propose to combine individual and joint variance analysis into a complete variance classification framework. Instead of exclusively computing \( \text{Var}_v(N_x(t); T) \) for comparison of a Lagrangian neighborhood \( N_x(t) \), we express every particle \( x \) with associated ensemble neighborhood in a new classification space \( C \) comprised of individual and joint variances:

\[
C = \text{Var}(N_{ix}(t);T) \times ... \times \text{Var}(N_{mx}(t);T) \times \text{Var}_v(N_x(t);T)
\]

(11)

where \( N_{ix}(t) \) is a respective neighborhood in \( \Omega \). In this space, every neighborhood is expressed as a \( m+1 \)-dimensional point. Note that dimensions of this classification space are constant for fixed advection time \( T \). An illustration of this space is given in Figure 4.

Because the ensemble dimension is not ordered and possesses no metric, we collapse/project this classification space to two dimensions, as illustrated. A single ensemble neighborhood may then, for example, be represented as an error bar graph.

4.5 Classification Space

In summary, every particle neighborhood \( \phi(N_{Ex}(t); T) \) has a single joint variance and \( m \) individual variances. Using the average of these individual variances, we are able to represent variance for each point \( x \) in the ensemble as a single point in the collapsed classification space:

\[
C' = \text{Var}(N_{i}(t);T) \times \text{Var}_v(N_{Ex}(t);T)
\]

(12)

with

\[
\text{Var}(N_{i}(t);T) = \frac{1}{m} \sum_{i=1}^{m} \text{Var}(N_{ix}(t);T).
\]

(13)

This location is identical to the mean location of the error-bar representation in Figure 4. The position of a point \( x \), when mapped into the classification space \( C' \), can provide insights into the comparability of the flow fields’ transport behavior. More specifically, we can distinguish first by average individual variance and then consider joint variance to obtain four distinct cases:

- **Low Joint Variance**: The vector fields in the ensemble show similar and stable transport behavior. Flow trajectories transport fluid to locations that are spatially close. Furthermore, the computed individual flow trajectories are reliable representatives of flow behavior due to low individual variances. If the ensemble is used to model a stochastic process, these features correspond to deterministic flow transport.

- **High Joint Variance**: Vector fields show strongly dissimilar transport behavior. Paths of individual flow trajectories are in regions with low divergence and thus reliable representatives of the underlying flow in the region, but at the same time trajectories in different flow fields transport fluid matter to spatially separate regions of the flow field. A trend analysis (see Section 4.6) can aid in further analysis.

In contrast, **high average individual variance** (upper region of \( C' \)) increases the difficulty of drawing reliable conclusions about flow comparison, since noise and numerical errors have a strong influence on the displacement of fluid matter. If the neighborhood radii used to compute individual and joint variances are identical, joint variance values tend to be higher than the largest individual variance (only case 2 below applies). The reason for this is that the joint particle neighborhood used for PCA corresponds to a union of the individual displaced particle neighborhoods. Otherwise, the following distinction can be made:

- **Low Joint Variance**: All fields might show similar transport behavior. All fields transport particles to spatially close regions in the ensemble. Due to numerically unstable trajectories in individual fields, drawing a conclusion about comparability needs further investigation of the fields.

- **High Joint Variance**: The fields might show dissimilar transport behavior. Due to highly divergent individual trajectories, the difference between transport across the ensemble may be caused by noise or inaccurate integration. Without further investigation, transport behaviors must be regarded as incomparable.

Figure 5 illustrates two of the four described cases. In order to allow the investigation of transport behavior in regions with a high joint variance, we implement a transport trends analysis based on particle clustering (Section 4.6) and provide means for interactive trajectory seeding (Section 5).

4.6 Outliers and Trends

We use the Principal Components Analysis as a variance based description of particle distributions after advection. As such, PCA suffers from the same limitations as other statistical analysis techniques that fit normal distributions to arbitrarily distributed data. While its

![Figure 4](image1.png)

![Figure 5](image2.png)
ability to provide a quantitative measure of the spreading of flow particles is suitable for the identification of commonalities and differences in flow transport, it cannot model trends or outliers in the set of flow neighborhood positions. These, however, are important measures for the comparison of ensemble data. For this reason we extend the purely variance-based analysis with a capability to highlight trends and outliers.

For the purpose of this work we rely on a single abstract definition that covers trends and outliers. **Trends or outliers** in transport behavior are present in a particle neighborhood, if \( \phi(N_{k}(t); T) \) contains one (outlier) or multiple (trend) flow neighborhood locations \( S \) that are (jointly) separate from the rest of the advected neighborhood, \( \phi(N_{k}(t); T) \setminus S \). In practice, we detect and quantify trends by performing a Minimum Spanning Tree (MST) cluster analysis of \( \phi(N_{k}(t); T) \).

Trends and outliers become relevant for locations in the ensemble that show a large joint variance. In these cases we perform a trends analysis as follows:

1. Construct one MST for every set of particles in \( \phi(N_{k}(t); T) \) that originates from the same run.
2. From these individual MSTs, construct one joint MST, such that it connects all particles in \( \phi(N_{k}(t); T) \).
3. Remove edges created by the joint MST that are longer than \( \Delta sep \).
4. Count connected components in the resulting graph.

Since the MST is cycle free, the number of connected components is one higher than the number of edges created by the joint MST whose lengths exceeds \( \Delta sep \). We chose \( \Delta sep \) empirically to be a fraction of the average trajectory lengths of particles in \( \phi(N_{k}(t); T) \). Figure 6 provides an illustration of our trends and outlier detection. Components that contain only particles from one member of the ensemble represent a run whose transport behavior is an outlier. A graph component with particles from two or more runs constitutes a trend. The number of connected components may be used as simple visual indicator of transport trends and can enrich purely variance based visual comparison.

### 5 Visualization

Our visualization system consists of two components: a statistical representation of variance distributions in the form of a scatterplot, along with a rendering of the flow domain. Both representations are semantically linked with regard to interaction.

#### 5.1 Classification Space

The collapsed classification space \( C \) can be visualized directly in the form of a scatterplot. Each point \( x \) of the ensemble domain \( \Omega_{E} \) is mapped to its position \( (\text{Var}(N(t); T), \text{Varg}(N_{E}(t); T)) \).

Section 4.5 described four cases of the variance configuration and their associated interpretation with regard to ensemble transport behavior. To obtain a first impression of ensemble behavior, a colormap is constructed using four colors and used to depict the ensemble’s behavior, as exemplified in Figure 7. While any choice of colors is technically possible, in this work, green is used for the lower left corner (low variances), red for the lower right corner (low individual variance, high joint variance), blue for the upper left corner (high individual variance, low joint variance), and white for the upper right corner (high variances). Note that red and green regions indicate comparable behavior across the ensemble, while other regions are difficult to compare and exhibit tendencies towards similarity (white) or dissimilarity (blue).

A second color-coding variant is aimed at visualizing the results of the trend analysis from Section 4.6. Here, we choose color based on the number of identified trends (e.g. blue, gray and yellow for one, two, or three or more trends, respectively), (cf. Figure 7)

An impression of individual variance distributions can additionally be provided by extending the scatterplot points into error bars along average individual variance visualization axis (cf. Figure 4).

#### 5.2 Ensemble Domain

For two-dimensional fields \( (\Omega_{E} \subset \mathbb{R}^{2}) \), a straight-forward visual representation can be produced immediately in the form of a color-mapped planar representation according to the color maps described above. To facilitate transferring insights gained by visualizing the classification space back to the ensemble domain, we can directly assign colors in the ensemble domain \( \Omega_{E} \), by assigning to each position its corresponding classification color, (see e.g. Figure 8)

Since the green areas (low variances) indicate similar behavior across the entire ensemble, in these areas traditional vector field visualization methods such as Line Integral Convolution (LIC) [3] can be applied to a representative field to jointly visualize the flow, e.g. from one of the ensemble members. This is achieved by assigning a lower opacity to the green color and then blending the visualization over a LIC-image. An example is provided in Figure 8. Note that LIC images are representations of stationary flow and as such have a limited expressiveness in highly time-varying flows. Alternative time-varying options include e.g. UFLIC, which was demonstrated to work well in combined visualizations [6].

For three-dimensional fields \( (\Omega_{E} \subset \mathbb{R}^{3}) \), occlusion renders such a straight-forward method impossible. Instead, we visualize the ensemble domain using volume rendering. The quality and usefulness of volume rendering depends strongly on the choice of a transfer function, that maps the visualized data values to color and opacity [4]. For this reason we define opacity in two steps: Analogously to the two-dimensional case, the low-variance, green areas are assigned a lower opacity than high-variance regions. In a second step, this opacity value is used as input for a user-definable, bump-shaped opacity transfer function that specifies the final opacity of a voxel in the rendering. This yields a depiction of structures that exhibit either high individual variance, high joint variance, or both, which are of most interest in comparative flow ensemble analysis (cf. Figure 9).

#### 5.3 Interaction and Exploration

Beyond the semi-automated color mapping outlined above, it is of course necessary to allow user control of color and opacity mapping. We provide a number of interaction modes of increasing complexity.

First, it is of course not immediately obvious which variance values should be considered “low” or “high”, especially since the maxima of joint and average variance are usually represented by outliers. Also, such a classification may depend on the data set, application domain, and features of interest. To steer the corresponding mapping, we allow a user to select a rectangular area within the visualization of the classification space – initially set to the extent of the classification space – that is used as the basis of the above color mapping. The four corner values are interpolated smoothly within the rectangle, see Figure 7. For data points outside the rectangle, the color of the closest point on the rectangle is used. Further, the inside of the rectangle defines a transitional area to cover variance ranges for which a clear classification cannot be made.

The techniques described so far are capable of providing an overview of the ensembles’ joint transport behavior. We provide two further tools for a more specific investigation. First, in classification space, we enable a user to manually highlight specific regions using
standard brushing tools. The visualization of the ensemble domain is instantly updated to reflect the change of the colormap. This allows a user to closely examine features such as classification-space clusters, that may be apparent in the scatterplot.

Moreover, in the ensemble domain, we support the placement of path lines to visualize the flow locally, as shown in Figure 7. For a given seed point, a pathline is started in each ensemble member and then displayed in the ensemble domain. While this method is not suitable for visualizing the flow globally, it is highly useful to provide additional insight into the local transport behavior. When applied to regions of low joint variance, the result is a tight bundle of pathlines that indicates the flow direction in all ensemble members. Outliers can be identified immediately as they diverge from the bundle.

When a pathline is seeded in an area of high joint variance, the curves tend to fan out in different directions. Trends can be identified as clusters of the pathline endpoints. This is especially useful in connection with a visualization of the number of trends, since each trend can be explicitly identified.

6 Experiments

6.1 Implementation

The implementation of our visual analysis framework is comprised of tasks that require different amounts of processing time and memory. For the construction of advected neighborhoods \( \phi(N_s(t);T) \), we sample the flow fields uniformly and compute a corresponding flow map \( \phi(:;T) \) for every field of the ensemble. This pre-computation is identical to the flow map computations required for classic FTLE techniques. Depending on the underlying grid, interpolation scheme, flow map resolution, and \( T \), this can be computationally expensive. In the examples shown here, the computation time of a complete set of flow maps for all members of an ensemble stayed well below 10 minutes in total on an Intel Core i7 workstation, leveraging basic OpenMP parallelization with 8 threads.

With these sampled flow maps, individual and joint variance computation for all points in the ensemble is reduced to performing look ups in \( \phi(:;T) \) for particles in \( N_s(t) \) followed by a PCA of \( \phi(N_s(t);T) \). These computations are significantly faster than the flow map computation and were also performed with multiple threads, but are not real time. The same holds for the computation of the minimum spanning trees for trends analysis.

After this pre-processing has been performed, the proposed interaction and exploration techniques are work in real-time using standard graphics hardware.

6.2 Data Sets

We applied our approach to the following three different scenarios.

Convection The convection datasets are a simulation of fluid around a hot, cylindrical object. Material at rest is heated around the cylinder, begins to rise, and forms a plume. By slightly perturbing the initial velocity conditions at the bottom of the domain, a two-dimensional ensemble of 30 flow fields and a three-dimensional ensemble of 10 flow fields were generated. The convection process splits the domain into two regions of upwards streaming material. A key feature to investigate is how fluid transport on the two sides interacts in the middle of the data set, which structures are formed during the convection process, and how robust they are under perturbation.

We examined a two-dimensional convection dataset for a time period from \( t = 0 \) up to \( t = 5 \), comprising 100 time steps in each ensemble, for a total ensemble size of 786 MB. A three-dimensional dataset also consists of 100 time steps with a combined size of 6 GB; we considered a time period from \( t = 2.5 \) to \( t = 3 \).

Industrial Stirring The stirring dataset is a set of 20 two-dimensional flow fields resulting from the simulation of mixing in a stirring apparatus. The device consists of two counter-rotating pairs of mixing rods that stir a medium in a cylindrical tank. The observed time range of the simulation \((t = 5 \text{ to } t = 10 (T = 5 \text{ corresponds to 50 time-steps with a step size of 0.1}))\) covers approximately one complete revolution of the stirring rods. The ensemble was generated by slightly varying the viscosity of the fluid to investigate mixing quality of the device for a range of different fluids, and totals 646 MB. The key question for this data set regards the efficiency of the stirring process. An ensemble visualization is expected to be able to identify regions where the mixing quality is high or low throughout the ensemble.

Rayleigh-Taylor Instability The Rayleigh-Taylor instability is the flow resulting from the mixing of a heavy fluid placed on top of a fluid with lower density. As gravity displaces the heavy fluid into the light fluid, unstable/chaotic behavior occurs. We use an ensemble of 8 three-dimensional simulation runs created for slightly varying density values, with each member comprised of 50 time steps and a total data size of 1.4 GB. The chaotic transport behavior present in the data set suggests a highly indeterministic process, challenging the expressiveness of similarity based ensemble visualization techniques. We observed the dataset from \( t = 0 \) to \( t = 1 \).

6.3 Results and Analysis

We begin by analyzing the two-dimensional convection data set. Figure 7 shows an application of our techniques to that ensemble. As indicated, our classification improves expressiveness over a naive joint variance-based visualization, by including regions where unstable trajectories do not allow for reliable comparisons. Exploration in classification space and the ensemble domain facilitate goal driven comparison of the underlying transport behaviors. By applying our techniques we find that the lateral regions of the data set correspond to extremely similar flow behavior and only the middle section above the heating cylinder contributes to transport variance in the ensemble. Interestingly, regions with high variance show some correlation with out trends analysis. As a consequence, two distinct trends are visible, where flow fields in the ensemble disagree reliably. This is an interesting observation, as it indicates, that the oscillation patterns forming behind the convection cylinder are to a large extent the result of a combination of two distinct behaviors caused by slightly uneven heating of the two lateral sides.

Figure 8 applies the same techniques to the stirring simulation ensemble. Because green regions indicate areas where average variance and joint variance are low, these parts of the ensemble represent significantly less mixing than the rest of the data set. There are two clearly visible such regions at the top and bottom of the dataset, where all data sets show similar low variance behavior and little to no mixing occurs. For the design of a stirring apparatus, such regions are to be avoided to improve overall mixing performance. Our observations strongly suggest modifications to the design.

In the 3D Rayleigh-Taylor instability ensemble (see Figure 9), high values of both individual and joint variance confirm the expectation that transport behavior near the mixing zone in the center varies greatly across the ensemble. Chaotic behavior leads to incomparable flow and indeterministic structures, as expected. These regions are successfully identified by our method. With our visual analysis techniques the stochastic process of chaotic transport in Rayleigh-Taylor flow can be separated into two distinct regions: Incomparable, indeterministic flow at the interface and similar deterministic behavior at the upper and lower regions.

Results generated with the three-dimensional version of the convection data set are presented in Figure 10. The system setup is similar to the 2D case, with the exception that the diameter of the heating cylinder varies along the depth axis. The resulting transport behavior analysis reveals very similar structures to the 2D field. An interesting observation is how the flow structure revealed by our joint classification strategy resembles a union of individual and joint variance structures.

7 Conclusions and Future Work

The presented work extends comparative visual analysis of ensemble data to time varying flow fields. More specifically, we have presented variance-based comparison techniques for Lagrangian fluid transport in 2D and 3D vector field ensembles. To reduce the limitations of exclusively variance-based analysis, we have extended the proposed technique with automatic trend detection and interactive exploration.
Fig. 7. Visual comparison of transport in the 2D convection data set. Top row: Modifying the variance classification changes the notion of high- and low-variances. The resulting color map facilitates easy distinction of regions with similar and dissimilar transport behavior. Bottom row, from left to right: The naive visualization based exclusively on joint variance values (red shading) does not indicate comparison quality or certainty. Our joint classification scheme indicates regions, where a comparison is not reliable without further exploration (white and blue). Our visualization indicates that both regions interact in an indeterministic fashion. Pathlines seeded in red regions - regions where the data sets show low average individual variance and high joint variance reveal the presence of strong transport variance, as expected. Furthermore, two distinct transport trends are visible. Automatic trends analysis using MST shows that the data set consists of regions with one prevalent trend (blue) and smaller areas, where two trends are visible (gray). Interactive brushing in classification space can help in providing a semantic link between classification space and the ensemble.

capabilities. Our experiments have shown that deterministic as well as incomparable transport behaviors in ensembles can be identified reliably. We believe that the presented techniques are a first step towards the comparison of Lagrangian transport in ensembles and may be extended in several directions, for example defining advanced trajectory similarity metrics, or including directional information of divergence during a similarity evaluation.

In the future we plan to investigate the comparison of LCS structures in flow field ensembles. Similar to a comparing vector field topologies, this has the potential to allow for feature based comparison of transport properties.

REFERENCES
Fig. 8. Visual analysis of an ensemble generated by varying fluid viscosities in a stirring apparatus simulation. Outer regions and especially the upper and lower parts of the data set reveal very similar and therefore deterministic transport behavior with low variances, indicating little to no fluid mixing. Varying the classification space color map allows to distinguish between reliable disagreement (red) and uncertain disagreement (white), which is not visible in the naive joint variance rendering (bottom right). A trends analysis allows for segmentation of the data set into regions that show one, two, or more different transport behaviors. Interactive pathline seeding reveals that trends in this data set are generally not caused by a grouping of different ensemble runs, but by outlier behavior present in the data set.

Fig. 9. Volume rendering of the classified Rayleigh-Taylor Instability. Left: A rendering of variance and joint variance based clustering. Transport behavior in the middle of the data set is detected to be strongly dissimilar. Middle: Deterministic structures (in green) are only present at the upper and lower parts of the data set, where no chaotic transport has occurred in the specified time range. Right: The scatter plot visualization is enhanced to show variance of the individual variance distribution. The resulting vertical lines give a good indication as how to specify classification boundaries. For this data set, a clean color segmentation along the y-axis is only suitable in few areas of the classification space.

Fig. 10. Volume rendering of the classified 3D convection data set. From left to right: A rendering of a single individual variance field lacks in impressiveness but indicates a large region of unstable trajectories (in blue). A rendering of the joint variance field reveals basic information about transport comparability in the flow field (in red). The classification procedure proposed in this work reveals strong disagreement between ensemble members along lateral sides of the ensemble. Manual pathline seeding within the volume rendering conveys flow transport in one point of the ensemble.

