Visual Analysis of Transport Similarity in 2D CFD Ensembles

Abstract—Currently, there are no methods of visual analysis for ensemble vector fields (EVF) that provide identification of flow trends and general flow similarity over the extent of Lagrangian transport across ensemble members. Finite-time Variance Analysis (FTVA) provides flow structure information only on particle distributions at the termination of streamline integration. In this paper, we first present a Lagrangian flow structure based on streamline clustering. Second, we discuss a method using streamline clustering to provide information of flow coherence at corresponding spatial regions in the EVF. We consider the regions where bifurcation in flow trends among the EVF members occur. We will also discuss how both methods can be used as a sequential framework for EVF analysis, by using the results of the scalar flow structure to find regions of member flow dissimilarity for further analysis.

Index Terms—Ensemble visualization, vector fields, flow visualization, feature clustering

1 INTRODUCTION

2 RELATED WORK

3 BACKGROUND

4 METHODS

5 EXPERIMENTS

5.1 Implementation

Our results were obtained from code written in Python, utilizing the SciPy package, Sci-kit Learn [4], and HDF5 [1] via H5py [2]. The PC system used an Intel Core i7-3930k with 32 GB of RAM, where six cores were utilized on each system. All Python scripts were run as single-threaded processes.

Table 1: Timings for flow maps and FTV A for the data sets in this study. Number of members reflects the members used in the computations and not necessarily the total available members. In cases where less members are used than available, those members used were randomly chosen from the available set. Compute times are dependent on number of ensemble members and field resolutions.

<table>
<thead>
<tr>
<th>resolution</th>
<th>members</th>
<th>time steps</th>
<th>flow map</th>
<th>FTVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lock-exchange</td>
<td>128x128</td>
<td>20</td>
<td>1100</td>
<td>0.0</td>
</tr>
<tr>
<td>Ocean</td>
<td>53x90</td>
<td>30</td>
<td>1100</td>
<td>0.0</td>
</tr>
<tr>
<td>Stirring</td>
<td>152x152</td>
<td>15</td>
<td>1100</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 2: Timings for pre-computation of clustering for terminal points and multiple streamline samples for the data sets in this study. Included is the total calculation time of the linear and angular entropy pre-computations. Compute times are dependent on number of ensemble members and field resolutions. Identical resolution and number of members for each data set are used for these timings and shown in table 1.

<table>
<thead>
<tr>
<th></th>
<th>terminal clustering</th>
<th>3-pt clustering</th>
<th>10-pt clustering</th>
<th>linear entropy</th>
<th>angular entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lock-exchange</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Ocean</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Stirring</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

5.2 Data Sets

Lock-exchange The initial conditions are heavy fluid on one side and light fluid on the other, separated by a barrier (the lock) [7]. At initial time, that barrier is removed, and the flow is allowed to evolve. See figure 1a. Initial uncertainty originates from not knowing the position of the interface between the two fluids. In other words, the volumes of heavy and light fluid on each side is not exactly known, and the initial barrier slides left and right accordingly. At the start of the simulation, the probability distribution of the position of the barrier is Gaussian. Therefore, after infinite time, it is expected that the barrier is characterized by a similar Gaussian distribution, but with the light fluid on top of the heavy one, and with the variance of distribution stretched if the size of the whole lock domain is not square. However, the probability distributions of the interface or the dominant dynamics in between this start and infinite time are not assumed Gaussian. The lock-exchange data has the following parameters: 128 x 128 grid with velocity measurements, 1000 realizations.

Ocean This data set covers a region of the Massachusetts Bay on the east coast of the United States of America [5, 6]. See figure 1b. The Massachusetts Bay volume in the study was divided into 53 x 90 grid with 16 depths. The depths at these 53 x 90 grid points vary significantly: depths as shallow as 90 meters and as deep as 196 meters were recorded. The important visualization concern for this data set is understanding where current flow splits into distinct paths between ensemble realizations. For example, flow patterns may deviate between the seed and the termination of a set of streamlines between realizations, but have similarly located terminal positions. See figure fig:oceanbifur.

Industrial Stirring The stirring data set is a set of 15 two-dimensional flow fields resulting from the simulation of mixing in a stirring apparatus [3]. See figure 1c. 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 to t = 10 (T = 5 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 646MB. The primary question for this data set regards the effectiveness 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.

5.3 Results and Analysis

This study does not compute individual member variances (FTLE) in the consideration of FTVA [3] but compares our new visualizations to
FTVA only. Using FTLE generalizes the application of FTVA to sensitivity between otherwise identical simulation runs (where variations due to numerical error and other noise-based variation is potentially present). Perhaps a more informative metric on FTVA, and streamline clustering in general, is streamline entropy, as discussed in section 4. Thus, our visualizations refer to both average linear and angular entropy maps, and FTVA maps for interpretation of streamline clustering and sampling frequency for individual streamlines.

We now discuss streamline clustering for the two-dimensional ocean data set. Figures 3a and 3b, show FTVA for the surface level currents. Figure 3a is for forward integration from the seed and 3b is for backward integration. The primary variance occurs in the central region of the simulation domain for both integration directions. This is somewhat intuitive, since streamlines seeded there have the potential to cover a larger area and thus their terminal positions to differ over greater distances. The trend/clustering analysis for terminal points is shown in figures 3c and 3d, for forward and backward integration respectively.

Our streamline clustering method provides a much higher sensitivity for visualizing trends in the streamlines from more conventional FTVA. Figure 3g shows streamline clusters for streamlines both forward and backward integrated, using the minimum of three sample locations. Even between figure 3g and either figures 3a or 3b, we see in a single visualization the trends over more of the data set. Figure 3h shows the same method using ten sampled locations. The number of clusters increase on the map for the higher sampling frequency. This is due to detecting more variation on the streamlines and seeing a higher resolution of the trends. Figures 3a through 3d are all deficient in expressing much of the flow behavior that occurs near the upper coastal region and the particular types of separate flow trends present there, i.e. flow bundles between the realizations that separate along the intermediate positions of the streamlines but often have similar positions farther into their flow trajectories. (See figure XXX for an example.)

Figures 3e and 3f both show linear and angular entropy maps, respectively. In figure 3e, it can be seen that streamlines seeded near the edges of the domain generally have high average linear entropy, i.e. a property proportional to average streamline length. The higher sampling rate of figure 3h versus 3g, does not reflect more streamline clusters however. Figure 3f exhibits high average angular entropies near the middle of the domain. This appears proportional to the number of streamline clusters for those same regions of the domain. Thus, higher average angular entropy, in this data set in particular, provides evidence for the rational of increased streamline sampling in those regions.

Intuitively, the greater that streamlines accumulate curvature (i.e. a property proportional to angular entropy) there is more opportunity to miss such similar flow behavior when sparsely sampling. This is especially true in the simplest case of terminal particle evaluation. It should be noted that the highest average angular entropies appear near the coastline, but little to no trends are present at those locations in the domain.

Figure 4 applies the same method to the industrial mixing simulation ensemble. As discuss in [3], the design of the stirring machinery shows needed improvement due to the low variance in much of the domain via FTVA. This is corroborated and repeated here in figures 4a and 4b. The trend analysis from [3] additionally shows much of the domain possessing at least two clusters of terminal particle positions for both the forward and backward integrations. Our method shown in figures 4g and 4h, sharply contrasts the previous analysis. Due to higher sampling rates, not clusters are found along the paths of the stirring apparatus itself. Indeed, the motions of the mixing apparatus clearly shows little fluid movement, even at advanced time, of the fluid touched directly by the paddles. This suggests that a potential geometrical or material design may be implemented to prevent lack of agitation at the fluid and paddle points of contact. This flow behavior is also seen readily in both entropy maps, i.e. figures 4e and 4f. Once again, to contrast our method over FTVA, even though there is noticeable variance at some locations in the field, those same regions possess low streamline entropy and little if none clustering. Thus, the consideration of only the relative variance for those regions is missing evidence for the rational of increased streamline per member over the region) and a single flow bundle of representative streamlines among the ensemble members.

6 Conclusion

References

Fig. 3: Massachusetts Bay data set at surface level. (a) Forward FTV A. (b) Backward FTV A. (c) Number of trend clusters in forward integration. (d) Number of trend clusters in backward integration. (d) Map of average linear streamline entropies for ensemble. (e) Map of average angular streamline entropies for ensemble. (f) Streamline clusters sampled at three points per streamline. (g) Streamline clusters sampled at ten additional points per streamline.
Fig. 4: Industrial stirring data set. (a) Forward FTVA. (b) Backward FTVA. (c) Number of trend clusters in forward integration. (d) Number of trend clusters in backward integration. (d) Map of average linear streamline entropies for ensemble. (e) Map of average angular streamline entropies for ensemble. (f) Streamline clusters sampled at three points per streamline. (g) Streamline clusters sampled at ten additional points per streamline.

Fig. 5: Streamline clusters for an incoherent flow region in lock-exchange data set. (a) All streamlines from a single member. (b) First cluster from (a) with representative streamline. (c) Second cluster from (a) with representative streamline. Representative streamlines are highlighted in red for (a) - (c). (d) Plot of representative streamlines for 20 members, each a random color.

