Information-theoretic Analysis of Unsteady Data

Heike Jänicke

Swansea University
Swansea, Wales, UK
h.jaenicke@swansea.ac.uk

Abstract

The temporal evolution of scientific data is of high relevance in many fields of application. Understanding the dynamics over time is a crucial step in understanding the underlying system. The availability of large scale parallel computers has led to a finer and finer resolution of simulation data, which makes it difficult to detect all relevant changes of the system by watching a video or a set of snapshots. In recent years, algorithms for the automatic detection of coherent temporal structures have been developed that allow for an identification of interesting areas and time steps in unsteady data. With such techniques, the user can be guided to interesting subsets of the data or a video can be automatically created that does not occlude relevant aspects of the simulation. In this paper, we give an overview over the different techniques, show how their combination helps to gain deeper insight and look at different directions for further improvement. Two CFD simulations are used to illustrate the different techniques.

1998 ACM Subject Classification G.3 Time series analysis, E.4 Coding and information theory

Keywords and phrases Information theory, unsteady data

Digital Object Identifier 10.4230/DFU.Vol2.SciViz.2011.118

1 Introduction

One of the big challenges in visualization is the depiction of data that changes over time. Many interesting structures in fluid simulations owe their relevance to their temporal evolution and the effects they have on a wider region of the system. Wake vortices that occur behind aeroplanes, for example, originate from the aeroplanes wings and effect growing regions behind the plane with increasing time. One way to investigate such evolution over time is to watch a video of the data set or to take snapshots at relevant time steps. As many applications are inherently three-dimensional, it is difficult to find a good camera position to depict all relevant changes in an unsteady data set. A second problem is the fact that not all relevant properties are known a priori. Hence, the movie might be perfect to investigate large changing structures but smaller and more subtle phenomena might be overlooked or occluded.

One way to ensure that relevant structures are depicted in the video, is to automatically detect interesting phenomena beforehand using so called feature detection algorithms. Many such algorithms use a predefined mathematical description of the phenomenon and find corresponding structures in space and time. However, even for such basic structures as vortices there exists no unique detection criterion but rather a large variety of vortex measures. Structures so far unknown are likely to be missed as it is difficult to capture anomalies by mathematical descriptions. Such structures might easily evolve in simulations with conditions that did not exist before like abnormal weather patterns induced by climate change.

In order to detect all these different relevant patterns, several approaches have been recently proposed. Commonly, they identify coherent structures that are different from

© Heike Jänicke;
licensed under Creative Commons License NC-ND
Scientific Visualization: Interactions, Features, Metaphors. *Dagstuhl Follow-Ups, Vol. 2.*
Editor: Hans Hagen; pp. 118–128
Dagstuhl Publishing
Schloss Dagstuhl – Leibniz Zentrum für, Germany

what would be considered ordinary in the given data set. Most such techniques are based on concepts from information theory, a theory that is amongst other things concerned with measures that quantify how much information a subset of data contains about the entire system.

In the following we will give a brief overview over different techniques and use two different flow simulation data sets to show how they can be used in combination to explore time-dependent data. As the different techniques are related to a large variety of applications, we refer the interested reader to the related work sections of the papers cited for each technique.

2 Information-theoretic Data Analysis

Information theory is a vast field with a huge amount of measures to quantify information contents. However, only few of them are applicable to structured time-dependent spatial data. The concepts to come are based on the notion of causal states which can be thought of as stochastic spatio-temporal patterns capturing the dynamics of a local neighborhood. Hence, causal states are perfectly suited to investigate the dynamics in the applications under consideration. Afterwards, we will investigate measures to quantify which causal states feature unusual behavior, before we summarize the idea of ϵ-machines that can be used to investigate the evolution of entire systems over longer time intervals.

2.1 Causal States

A powerful concept from information theory is the notion of causal states as introduced by Crutchfield and Young [2]. A causal state is a stochastic spatio-temporal pattern that describes a position's or particle's past and future (compare figure 1). The past comprises all other positions that go directly or indirectly into the computation of the value at the central position. The future is the set of positions whose value is influenced by the value associated with the position or particle. A simulation's set of causal states is computed from the underlying data by dividing the occurring configurations into different classes according to their past and future. Each resulting causal state is a set of past configurations along with an associated distribution over possible futures that might occur after observing the past. For more details on the theory and computational aspects of causal states, we refer the interested reader to previous work [5]. In summary, the set of causal states summarizes the dynamics in the data set and gives an overview over the short-term temporal evolutions in it.

2.2 Eulerian vs. Lagrangian Frame of Reference

As shown in figure 1, the structural masks for past and future can be chosen differently. If we choose these structures such that they are analog to the finite difference model used in computational fluid dynamics (CFD), they have the shape of light cones as illustrated in figure 1 (top) labeled EulerianLSC. This structure supports the Eulerian description of motion, where the simulation domain is subdivided into cells and the evolution of values at different cell positions is recorded over time. An alternative frame of reference is the Lagrangian description of motion, where the flow is described using a number of particles that change their position according to the velocity field. In the Lagrangian frame of reference, the evolution of values associated with a certain particle is of interest. A simple example of the two frames of reference is the investigation of a river. We can either observe a fixed

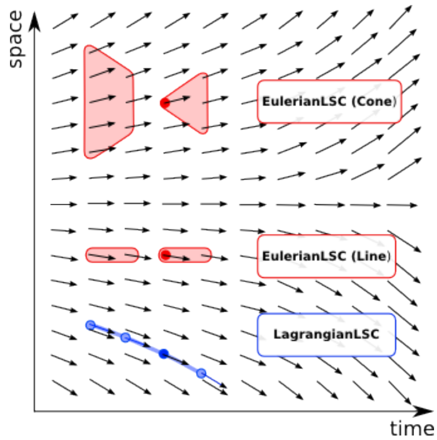


Figure 1 Past and future configurations for the different measures of complexity.

position and monitor the velocity, temperature or pressure at this particular position or we can put a small boat on the river and monitor the water properties beneath the boat that follows the current. Both frames of reference are of high importance in CFD research and can be supported by the causal states using different structures for past and future as illustrated in figure 1. The light cone structures at top and center support the Eulerian frame of reference and the one at the bottom the Lagrangian frame of reference. The cylinder shaped structure depicted in the middle is a computationally faster version of the EulerianLSC, which we found to give similar results as the full cone structure [4].

2.3 Local Statistical Complexity (LSC)

The goal of this work is to identify positions that form interesting structures in the data set. The causal states we have identified so far subdivide the data set into different classes of behavior. The next thing we need is a measure to quantify how interesting each class of coherent behavior is. Commonly, it is very difficult to say what a user might be interested in. If we assume that we are to show the user the unlikely events, the task becomes more feasible. Local statistical complexity [7, 3] is an information-theoretic measure that tells for a given causal state how much information from the past is required to predict the dynamics in the local future. If we only need little information and it is very easy to predict the dynamics, we found a pattern that is very common in the data set and the user probably knows that it is in there anyway. However, if we detect a pattern that is hard to predict and therefore unusual, it might be something that the user is interested in. Hence, by computing the local information contents, local statistical complexity assigns each causal state and thus each position in the (multivariate) data set a scalar value telling whether the local dynamics are common or not. The resulting time-dependent scalarfield can easily be used to guide the users attention or to compute a good position for the camera when generating a movie.

2.4 ϵ-Machines

In order to study the dynamics of the system as one, we need a depiction of the causal states that provides information on how they interact with each other. Such a visualization is provided by ϵ-machines [6]. ϵ-Machines can be thought of as directed graphs. The nodes are the causal states of the system and the edges depict the transitions between them.

Figure 2 Relevant structures in the delta wing data set: (left) Streamsurfaces indicating the six vortices above the wing. (center) ϵ-machine of this data set with colored substructures. (right) Physical positions corresponding to the colored subregions of the ϵ-machine.

For example, if a position in the data set changes from causal state A in time step t to causal state B in time step $t+1$, we add an edge linking causal states A and B in the ϵ-machine and assign it weight 1. This procedure is repeated for all transitions in the entire data set. The resulting structure can be thought of as a finite state machine capable of simulating the dynamics in the given data set. Now that we have a model representing the entire data, we can analyze its properties. For example, we can investigate strongly linked components to detect coherent structures in space, find subsets with high LSC to identify unusual formations or track features over time.

2.5 Areas of Application of the Different Techniques

In the preceding sections, we have summarized the concepts of causal states with their application to Lagrangian and Eulerian flows, local statistical complexity and ϵ-machines. Before we continue with the data analysis let us briefly look at the type of information the different concepts can extract. Causal states form the building blocks of all the techniques. They subdivide a given system into a set of stochastic spatio-temporal patterns. Based on the structural mask of the causal state, we can decide if we want to investigate the data from an Eulerian or Lagrangian point of view. The Eulerian view gives site specific information and relates different positions to one another. The Lagrangian view supports the analysis of particle evolution. Local statistical complexity can be used in both scenarios to identify either positions or particles that feature extraordinary local dynamics and hence, indicate interesting regions. When the dynamics of the entire system are of interest, ϵ-machines are a useful tool. They depict all causal states occurring in a system and show how they interact. Different enhancements can be used to focus on specific information such as the evolution of coherent structures, stability of different phenomena or the interaction between different structures in the domain.

3 Data Analysis

In the following we will apply the different analysis techniques to two different flow simulations and show how they can be used in combination to gain deeper insight into the data.

3.1 Delta Wing

The Delta Wing data-set represents the airflow around a delta wing at low speeds with an increasing angle of attack. Multiple vortex structures form on top of the wing due to the rolling-up of the viscous shear layers that separate from the upper surface. These

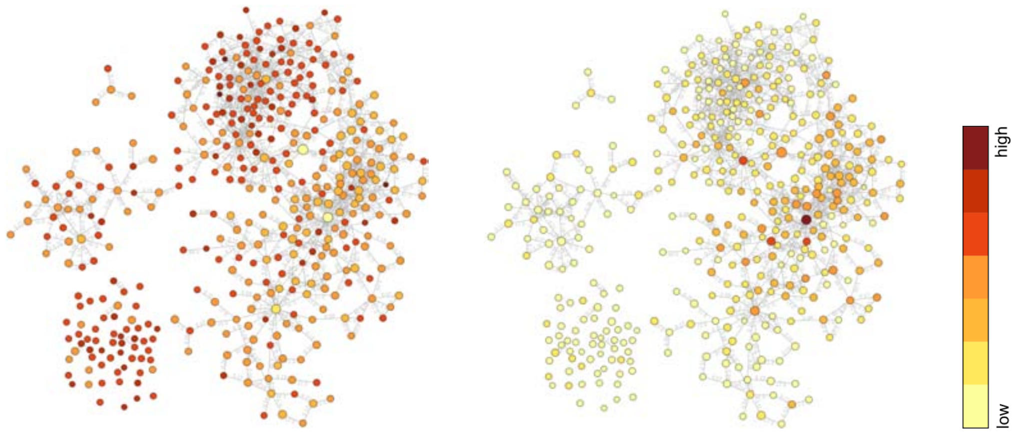
Figure 3 Evolution of the recirculating bubble in the delta wing data set: (left) ϵ-machine; Selected nodes are colored in blue. (center) Corresponding positions in time step 650. The pink frame indicates the area of the closeup. (right) Physical positions with corresponding causal states in time steps 650, 670, 690, and 710.

formations of three vortices can be observed on either side of the wing (Fig. 2(left)). With increasing angle of attack the intensity of the primary vortices (vortices nearest the symmetry axis) increases until in time-step 700 a vortex breakdown occurs (bubbles at the end of the vortices). The analysis of vortex breakdown is highly interesting, as it is one of the limiting factors of extreme flight maneuvers. The grid consists of approximately 3.1 million positions.

Figure 2(center) shows the ϵ-machine of the data set. Several highly interconnected structures were brushed and positions corresponding to these causal states are depicted in figure 2(right) using isosurfaces of the same color. We can distinguish four different structures: the causal states indicated by number one and colored in dark blue can be found at the recirculating bubbles and at the tips of the wing, number two is the surrounding domain, number three comprises positions forming an outer hull around the major vortices and number four are the positions in the main vortex. If we follow the causal states in the dark blue selection number one over time as depicted in figure 3, we see that the outline perfectly follows the changing structures of the recirculating bubble. Hence, we are able to detect and follow features without giving an a priori mathematical definition of the structure. Moreover, as such structures consist of a set of connected causal states the description is more flexible and we can easily adopt it to features whose values change over time, by adapting the selection to strongly linked nodes in the graph.

Now that we have identified the different structures in the data set, we can look for the unusual ones. Therefore, we can either color the nodes in the ϵ-machine according to local statistical complexity as illustrated in figure 4(left) or we can render an isosurface highlighting all positions with a LSC value above a given threshold. The colored ϵ-machines reveals that the recirculating zones are the most complex structures. Additionally some interesting formations occur in all other subparts but it is rather difficult to isolate additional coherent structures. Looking at the isosurface in the LSC field (fig. 5), we can clearly distinguish the minor vortices that are part of subset 2 in the ϵ-machine. Color coding the ϵ-machine according to the number of edges per node, as shown in figure 4(right), reveals that positions belonging to the major vortices and their hulls are very dynamic and often change their state in a very unpredictable manner as there are many possible successors. To investigate this behavior more closely, we will look at the Lagrangian representation of the data.

Figures 6a and 6b show a number of pathlines started in front of the delta wing where color indicates the norm of velocity or pressure at the given position. Comparing the two color distributions, the images look very different. However, if we color the particle traces using the LSC of the respective variables, the results become quite similar. This means that particles feature an unusual evolution of values for velocity as well as for pressure. The two



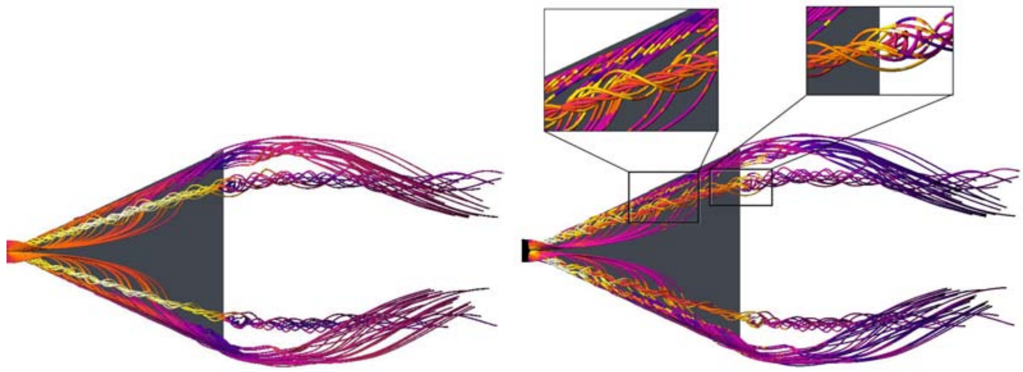
■ **Figure 4** ϵ-Machine of the delta wing using different color codings: (left) local statistical complexity of causal state and (right) number of edges per node.

(a) LSC > 10 **(b)** LSC > 14

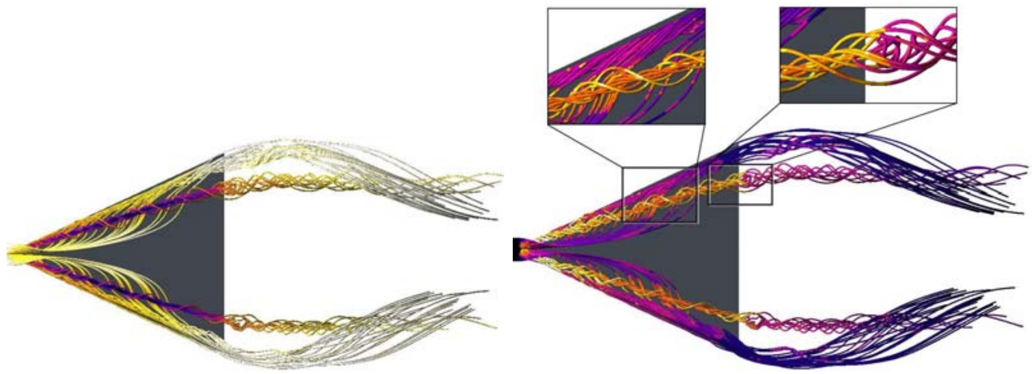
■ **Figure 5** Isosurfaces in the local statistical complexity field of the norm of velocity of the delta wing data set.

close-ups focus on two different features. The left image depicts an area where particles move from the area of influence of the major vortex into the one of the minor vortices. We can see that the particle traces that first go underneath the spiralling major vortex feature yellow color, i.e., high LSC, when close to the first vortex, go back to more usual temporal dynamics when between both vortex structures and change to more unusual dynamics again as they enter the minor vortices. The close-up on the right hand-side depicts particles that change from the major vortex into the recirculating bubble. Here we see that the dynamics are most unusual at the transition point and more predictable inside the structure.

In summary, we saw that ϵ-machines are well suited to identify major coherent structures in a data set and their evolution over time. Moreover, they give an overview over the distribution of different quantities such as LSC over the different structures. For a more detailed analysis, we used the depictions of the Eulerian- and the LagrangianLSC. While the EulerianLSC is good at giving holistic impressions of the relevant structure, LagrangianLSC is better suited to investigate more subtle phenomena. Hence, all three techniques are a powerful combination and can answer a large variety of questions.



(a) Norm of velocity



(b) Pressure

Figure 6 Pathlines in the delta wing data set color-coded using (a) (left) norm of velocity, (right) LagrangianLSC of norm of velocity and (b) (left) pressure and (right) LagrangianLSC of pressure.

3.2 Swirling Jet

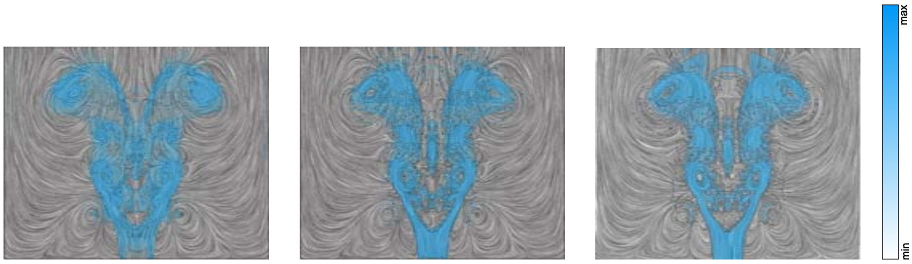
The development of a recirculation zone in a swirling flow is investigated by numerical simulation. This type of flow is relevant to several applications where residence time is important to enable mixing and chemical reactions.

The unsteady flow in a swirling jet is simulated with an accurate finite-difference method. The Navier-Stokes equations for an incompressible, Newtonian fluid are set up in cylindrical coordinates assuming axi-symmetry in terms of streamfunction and azimuthal vorticity. All equations are dimensionless containing the Reynolds number Re and the swirl number S as defined by Billant et al. [1]

$$Re \equiv \frac{v_z(0, z_0)D}{\nu} \qquad S \equiv \frac{2v_\theta(R/2, z_0)}{v_z(0, z_0)} \tag{1}$$

where $z_0 = 0.4D$, $D = 2R$ is the nozzle diameter and ν the kinematic viscosity, as dimensionless parameters.

The flow domain is the meridional plane $\mathcal{D} = \{(r, z) : 0 \leq r \leq R, 0 \leq z \leq L\}$ with $R = 5D$, $L = 8D$ and D denoting the nozzle diameter at the entrance boundary. The flow domain is mapped onto the unit rectangle which is discretized with constant spacing. The mapping is separable and allows to a limited extent crowding of grid points in regions of interest. The present simulation uses $n_r = 91$ and $n_z = 175$ grid points in radial and axial directions. The boundary conditions are of Dirichlet type at the entrance section and the



(a) LIC + EulerianLSC (Cone) **(b)** LIC + EulerianLSC (Line) **(c)** LIC + LagrangianLSC

Figure 7 Swirling jet: Relevant structures extracted using Eulerian- and LagrangianLSC.

outer boundary and at the exit convective conditions are imposed for the azimuthal vorticity. The initial conditions are stagnant flow and the entrance conditions are smoothly ramped up to their asymptotic values within four time units.

The depiction of time-dependent two-dimensional data is much easier than the visualization of 3D data as it is free of occlusion. All structures can directly be presented using texture-based techniques such as line integral convolution (LIC) and the movie approach. However, these approaches have two disadvantages. Texture-based approaches do not distinguish between high relevance features and those that are induced by noise. Hence, subtle irrelevant structures are overemphasized and it is difficult to distinguish between structures and noise. An additional color-coding based on the norm of velocity or another field might provide assistance in this direction, but the user has to know the relevant value ranges beforehand. Similar problems occur in the temporal domain, where no assistance on what is relevant and what is not is provided. In the following, we will show how the previously introduced methods can help detect interesting phenomena in space and time.

Figure 7 depicts the three different complexity measures applied to the swirling jet data. In general the structures look very similar. This is due to the fact that they change little over time and that the vortex structures remain at more or less fixed locations. Minor differences are present at the central part of the jet and the region at the top of the image where water leaves the domain. While the dynamics in the center are more unusual from an Eulerian point of view, those at the upper parts feature higher LagrangianLSC. EulerianLSC (Cone) marks large areas of the flow and the complexity slowly decreases at the boundaries. This happens as the cones consider a large area of influence that gradually moves out of the relevant structures at the boundaries. This more extended spatial pattern has its strengths when it comes to the analysis of structures that are in itself not very interesting such as vortex center, where the velocity is zero, but are relevant due to the surrounding flow. LSC is perfectly able to capture vortex corelines as it takes neighboring values into account. Hence, the unusual pattern is a stagnant center surrounded by stronger current. The two line-based measures, EulerianLSC (Line) and LagrangianLSC are not able to capture such structures, but nevertheless highlight the vortices as the surrounding flow is very unusual. The more crisp boundaries of the last two measures are, however, better suited to detect sharp boundaries than the conical shear region. Here the boundaries do not get washed out as with EulerianLSC (Cone) but precisely mark the outer and inner boundary. In summary we can say that all three techniques were able to detect the relevant structures and using a combined presentation of LIC and LSC gives a better impression of the dynamics in the data as relevant structures are immediately visible. Spatial patterns in the lower part of the images that are close to noise are not highlighted by the LSC and hence, attract less attention.

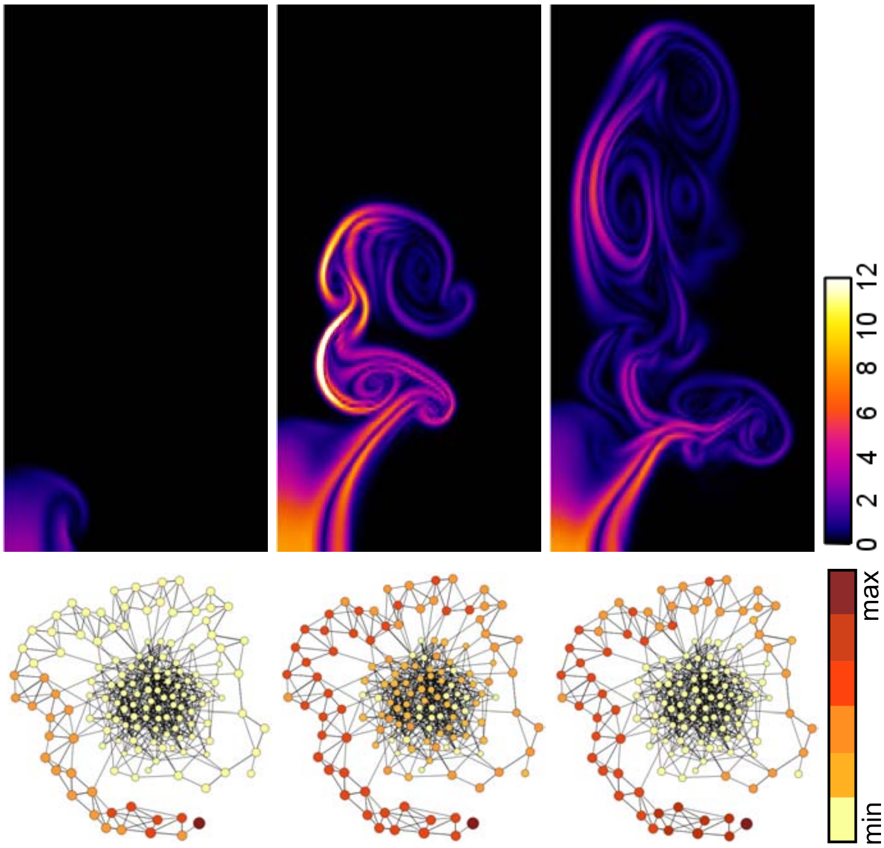


Figure 8 Evolution of the ϵ-machine of the swirling jet (norm of vorticity): Time steps (0), 2, (8), 56, 100.

Now that we have identified the unusual structures in space, we will focus on the temporal component. To investigate the longterm evolution, we look at the ϵ-machine of a related simulation as depicted in figure 8. The corresponding ϵ-machine depicted below looks like a curling up snail-shell. We see a well pronounced outer arc and a lump-like inner set of nodes that is highly interlinked. The color-coding of the nodes indicates the number of positions per causal state in the different time steps.

An easier overview over the temporal dynamics can be gained when representing these structures in a matrix view as depicted in figure 9. Nodes are aligned along the x-axis, the y-axis encodes time and color indicates the number of physical positions that are in a certain causal state at a given time. Two different features can be observed: First, some of the nodes are present throughout almost the entire run and have quite constant intensity, i.e., a stable number of corresponding positions. Second, around time steps 18 and 56 strong fluctuations occur and many of the otherwise unused causal states appear.

When we look at the rendering of the data set at these noticeable time steps, we can relate certain formations in the matrix to physical events. The flow starts with a resting fluid and all positions feature a single causal state, the dark red node at the spirals end in the machine and the corresponding dark red line in the matrix view. With increasing time, the velocity increases and larger parts of the domain leave the steady causal state. During this process (first eight time steps), the nodes in the outer arc of the machine become used more frequently. In the matrix view, this corresponds to the lower part of the matrix where

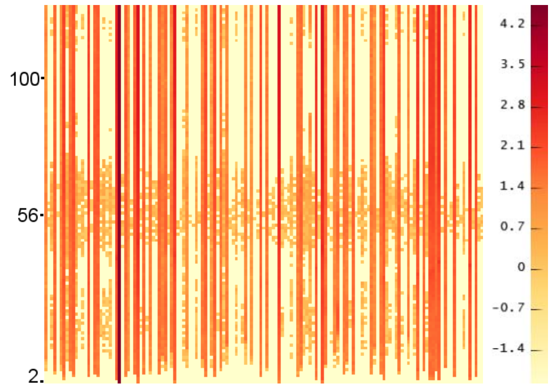


Figure 9 Transition probability in matrix view with logarithmic scaling.

the different vertical lines start, each representing a node in the outer arc of the ϵ-machine. After this initial period, the flow becomes more dynamic and nodes in the center of the machine are activated as well. Strongest dynamics can be observed around time step 56 (fig. 8(center)) when almost the entire inner part of the machine is active. Up to this time step, a strong conical shear region has formed and ring-like vortex structures evolved. Both the matrix view and the ϵ-machine show that the dynamics during this period are hard to predict. The matrix view contains many fluctuations and in the ϵ-machine the inner lump is active that is highly connected with transitions of very low probabilities. In time step 75 the flow becomes more predictable again. The inner lump is no longer active and only stable patterns with few transitions in the outer arc occur in the flow. This stable evolution lasts until the end of the simulation and time step 100 is given exemplarily in figure 8(right).

The LSC analysis in the first part revealed the different relevant structures in the data set, namely the conical shear region and the ring-like vortex structures. However, the measures provided little information about the temporal distribution of the structures and their evolution. This part can be better analyzed using the ϵ-machine and the representation of the causal states. Thus, we can easily go to time steps in the simulation that features an unusual formation in the matrix view of the causal state distribution, render the data using a standard texture-based visualization algorithm such as LIC and highlight relevant components using LSC. If the features are more dynamic, a similar analysis as used with the delta wing based on the LagrangianLSC is possible to investigate the local temporal evolution more closely. However, in this example we found the application of the ϵ-machine combined with any of the LSC measures sufficient to extract all relevant structures over time.

4 Conclusion

In this paper we gave a summarizing overview over recently introduced visualization techniques that aim at the extraction of coherent structures based on information theory. So far, causal states, local statistical complexity and ϵ-machines have always been introduced as independent techniques. While LSC is a measure to extract coherent structures in a time-dependent data set, ϵ-machines are applied to study the longterm behavior in unsteady dynamics. In this paper we showed how the different techniques can be used in combination to gain deeper insight into the data than can be provided by the individual approaches.

In the future we would like to extend the set of methods towards better investigation concepts for long-term correlations and effects. Moreover, the feature tracking ability of the information-theoretic concepts requires more attention to uncover its full ability.

Acknowledgements

The authors would like to thank Wolfgang Kollmann for the swirling jet data-set and Markus Rütten (DLR) for the delta wing data-set.

References

1 P. Billant, J.M. Chomaz, and P. Huerre. Experimental study of vortex breakdown in swirling jets. *JFM*, 376:183–219, 1999.

2 James P. Crutchfield and Karl Young. Inferring statistical complexity. *Phys. Rev. Lett.*, 63(2):105–108, July 1989.

3 Heike Jänicke, Michael Böttinger, Xavier Tricoche, and Gerik Scheuermann. Automatic Detection and Visualization of Distinctive Structures in 3D Unsteady Multi-Fields. *Computer Graphics Forum*, 27(3):767–774, 2008.

4 Heike Jänicke and Gerik Scheuermann. Measuring complexity in lagrangian and eulerian flows. *Under review*, 2009.

5 Heike Jänicke and Gerik Scheuermann. Steady visualization of the dynamics in fluids using ϵ-machines. *Computers and Graphics*, 2009.

6 Heike Jänicke and Gerik Scheuermann. Visual analysis of flow features using information theory. *Conditionally accepted for the special issue on knowledge-assisted visualization in Computer Graphics and Applications*, 2009.

7 Heike Jänicke, Alexander Wiebel, Gerik Scheuermann, and Wolfgang Kollmann. Multifield Visualization Using Local Statistical Complexity. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1384–1391, 2007.