Pointwise Information Analysis for Multivariate Time-varying Feature Identification

Soumya Dutta, Xiaotong Liu, Ayan Biswas, Han-Wei Shen, Yifan Hu, James Giuliani, and Jenping Chen

Abstract— Identification of salient features from a multivariate time-varying system plays an important role in scientific data understanding. In this paper, we present an interactive analysis technique based on the pointwise mutual information (PMI) to identify the amount of information sharing between different value pairs from multiple variables. The various measures derived from PMI can be used to construct new scalar fields, which allow us to examine the combined and complementary information possessed by multiple variables. From a user selected time window, we can construct several time aggregated volumes based on different aggregation criteria, and use them to capture salient time-varying features. Since the pointwise mutual information gives us a way of quantifying the information shared among all possible combinations of scalar values for multiple variables over time, it can be used to identify temporally salient isocontour tuples. Simultaneous visualization of such isocontours depicts combined or complementary features in the data set and their evolution in time. To identify temporally salient variables and appropriate time step ranges for detailed exploration, a novel user interface is provided based on our information measures. Using the interface, users are able to visualize the shared information content and their temporal variation for all the variables, and can select temporally related variables accordingly for in depth analysis. Experiments on several scientific data sets show the effectiveness of our system in identifying salient features from time-varying data sets.

Index Terms—Information theory, framework, isosurface, multivariate time-varying data analysis

1 Introduction

Effective analysis and visualization of multivariate time-varying data sets is an important but challenging problem, as a thorough understanding of the complex relationships among the multiple variables is required. Existing methods that use different correlation metrics are mostly focused on studying the average behaviors of the variables, but little focus is on how the specific values of the variables interact with each other. Analysis of importance of scalar values from a single variable system [2, 10, 23, 6, 19] and multivariate domain [3] has been done in the past. However, a guideline to study the relationships of specific value combinations in multivariate time-varying data sets is still missing. In depth analysis of this kind, however, can guide users to explore relationships among variables in terms of their specific value occurrences, providing further insights which have not been exploited yet. Such analysis will also allow users to systematically explore time-varying relationships of multiple variables by enabling interaction with specific value combinations. By observing the co-occurrences of such value combinations, important regions in the data sets can be identified where the value combinations show joint features or features specific to a particular variable. Devising such analysis framework is a non-trivial task, because of the existence of large number of value combinations in the analysis space. The problem becomes even harder due to the dynamic nature of time-varying data.

To address the issues, in this paper an information-theoretic exploration framework is presented to help users conduct detailed analysis of multivariate time-varying data sets. The proposed framework analyzes variable relationships by measuring shared information for every specific value combinations. We make use of the pointwise mutual information (PMI) to quantify the information content for a specific value combination. This enables us to effectively analyze the interaction of variables by their values. To visualize the value combinations and their information content, we create a new scalar field, called PMI field, using the pointwise mutual information values at every spatial location. To identify salient features in the data, we segment these scalar fields into different regions to focus on regions where value combinations have higher co-occurrences or noticeably low co-occurrences. High co-occurrences in a region indicates the existence of a joint feature, and regions with lower co-occurrences are explored for any potential surprises. We further employ the idea of PMI fields into temporal domain to extend the pointwise study to time-varying data sets. Since time-varying data sets come with dynamic variable relationships, our framework aggregates several PMI fields into a single time aggregated PMI field to explore the temporal trends of variable relationships in the spatial domain. To provide specific insights about the individual value combinations of multiple variables, a time series for each value combination is created and their saliency is measured by the temporal variation of their PMI values. Important scalar value combinations are then selected based on their temporal information variation.

To facilitate the exploration in an effective and intuitive way, our framework provides an interactive interface based on several information-theoretic measures. Using our interface, temporally salient variables with an interesting time step range can be more easily identified so that further analysis can be performed. An interactive histogram and parallel coordinates interface is provided for the selection of specific value combinations. Using our interactive framework, users are able to identify temporally salient variables and perform pointwise analysis on them to explore multivariate time-varying data sets.

Our contributions in this work are thus threefold:

1. We present a pointwise study of variable relationships for multivariate time-varying data sets in terms of their specific value combinations by exploiting pointwise mutual information.

2. We construct a scalar field, called PMI field that preserves the spatial context during the analysis phase and identify salient spatial regions in the data. This provides us the basis for the creation of time aggregated PMI fields that in turn facilitates the identification of temporally salient value combinations based on their temporal variability.

- *Soumya Dutta is with The Ohio State University. E-mail: dutta.33@osu.edu.*
- *Xiaotong Liu is with The Ohio State University. E-mail: liu.1952@osu.edu.*
- *Ayan Biswas is with The Ohio State University. E-mail: biswas.36@osu.edu.*
- *Han-Wei Shen is with The Ohio State University. E-mail: hwshen@cse.ohio-state.edu.*
- *Yifan Hu is with Yahoo Labs. E-mail: yifanhu@yahoo.com.*
- *James Giuliani is with The Ohio State University. E-mail: giuliani.6@osu.edu.*
- *Jenping Chen is with The Ohio State University. E-mail: chen.1210@osu.edu.*

3. A new framework for multivariate time-varying data exploration is proposed which guides the users at each step of the exploration process by providing an intuitive and interactive interface.

This paper is organized as follows: in Section 2 we discuss the related research works to the topic of this paper. A brief overview of our framework is provided in Section 3. In Section 4 the proposed methodology has been discussed in detail. Section 5 discusses about the interfaces of our interactive system and in Section 6 we demonstrate the effectiveness of our system by showing results on some scientific data sets. In Section 7, we discuss about the feedback from the domain scientist, followed by Performance analysis in Section 8. Conclusions and future works are discussed in Section 9.

2 Related Works

The content of this paper is closely related to three broad research topics in the field of scientific data visualization : information theory, multivariate data analysis, and time-varying data analysis. Information theory [8] has been used extensively for solving many different problems in visualization. For polygonal scenes, Vázquez et al. [25] used the view point entropy for automatic computation of good viewing positions. Borodoloi and Shen [4] used entropy for selecting informative voxels to construct a view selection algorithm for volume rendering. Viola et al. [26] proposed another information theoretic approach to select the most expressive view point by maximizing the mutual information. Information theoretic approaches have also been used for light source placement for varying camera positions [14] and analysis of scene visibility and radiosity complexity [11]. Jänicke and Scheuermann [18] used information theory for the analysis of unsteady flow features by using ϵ-machine, a highly compressed abstract representation of the flow, where the causal states depict different dynamics within the data and the edges between them represent different transitions and likelihood. Haidacher et al. [15] used pointwise mutual information for modulation of opacity in their work.

Multivariate data analysis is a well researched area in the past for a wide range of applications. As presented by Oliveira et al. [9] in their survey of visual data exploration, multivariate data analysis can be broadly subdivided into four categories: geometric projection, pixel oriented techniques, hierarchical display, and applications of iconography. Wong and Bergeron presented an extensive survey of multivariate visualization [30]. Jänicke et al. [16] performed the analysis of multivariate data by transforming high dimensional data into more tractable 2D space. A high dimensional brushing for the multivariate data sets was proposed by Martin and Ward [20] which enabled user interaction in the exploration process. A Nugget Management System (NMS) was proposed by Yang et al. [33] where the nuggets are referred to as the data that are of interest to the user. Bramon et al. used specific mutual information for multi-modal data fusion [5]. In a recent work, Biswas et al. [3] used concepts of information theory to design an interactive interface for effective exploration of multivariate data sets. They used a mutual-information-based force directed graph for variable selection and applied specific information metrics to detect different degrees of uncertainty in the multivariate scenario.

Exploration of time-varying data sets to extract features has been a challenging task, and researchers have investigated this in the past. Woodring et al. [32] used wavelet transform on the time-varying data to generate a series of curves and cluster them to find time-varying trends. Akiba et al. [1] used time histogram for designing transfer functions to reveal features of time-varying data and provide feedback for simultaneous rendering. Younesy et al. [34] proposed the differential time histogram which allowed for efficient handling of queries for large time-varying data and also provided an error bound on the data visualization. Jänicke et al. [17] used the density-driven voronoi tessellation for dividing the unsteady data into classes that show different behaviors. Wang et al. [28] used information theory to extract importance curves from the blocks of data which allowed effective classification and visualization of features. In a recent work by Wang et al. [27], the authors used transfer entropy for selection of variables and visualize the information transfer that revealed the causal relationships in the data.

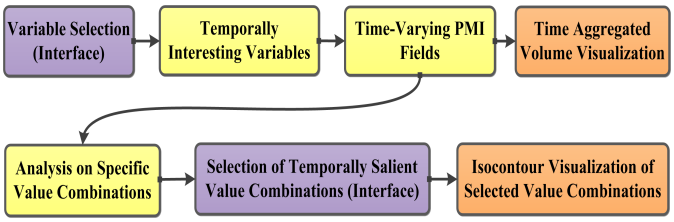


Fig. 1: A schematic representation of the proposed workflow.

3 System Overview

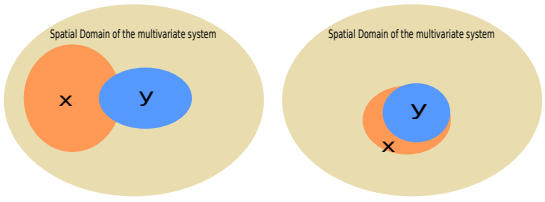
Our high level goal in this work is to analyze multivariate time-varying data using information theoretic measures and identify interesting regions in the data set. Figure 1 shows a schematic view of our complete framework. In this work, we propose to explore the relationships among multiple variables by analyzing the shared information content of their specific value combinations. We demonstrate that by performing pointwise mutual information (PMI) analysis, temporally salient value combinations can be identified. Given a multivariate time-varying data, we initially select multiple variables and identify an interesting range of time steps where they show high information overlap. Next, we quantify the information content for every value combinations of the selected variables using PMI. This enables us to construct a new scalar field, called PMI field that preserves the spatial context during the analysis phase. Interaction with this information field allows users to identify regions where variables show higher joint and individual activities. To extend the idea of PMI field into temporal domain, we aggregate several PMI fields from a sequence of time steps to construct a single scalar field. Such time aggregated scalar fields help us in detecting time-varying features and study their evolution over time. We further perform analysis on specific value combinations to identify temporally salient scalar values. Given the large number of possible value combinations, we measure the temporal variation of PMI values of each value combination and organize them from low to high variation value. Observing the temporal variation, temporally salient isocontours are detected that help in visual exploration of the data set in the spatial domain.

4 Detailed Description of Multivariate Temporal Analysis Framework

Below we discuss in detail our analysis framework, where information-theoretic measures are used to perform pointwise analysis of mutivariate time-varying data and explore important variable relationships from the perspective of specific value combinations.

4.1 Combined and Complementary Informativeness Characterization

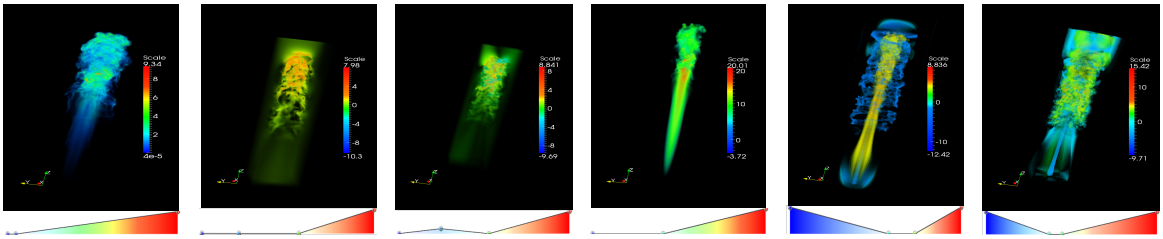
Study of interaction among multiple variables is of interest to the researchers of various fields. Numerous techniques have been developed to aid the researchers in the exploration process [30, 12, 20, 13, 22]. Still a robust technique, that can explore the multivariate relationship



(a) PMI of x and y is negative in this case since they have low joint probability and higher marginal probability.

(b) PMI of x and y is positive and high in this case since they have high joint probability.

Fig. 2: PMI value distribution at different scenarios.



(a) Volume of Gradient field in Plume data set.

(b) Volume of X-velocity in Plume data set.

(c) Volume of Y-velocity in Plume data set.

(d) Volume of Z-velocity in Plume data set.

(e) PMI field of Gradient and Z-velocity. High and low valued regions are emphasized.

(f) PMI field of X-velocity, Y-velocity, and Z-velocity. High and low valued regions are emphasized.

Fig. 3: Visualization of multivariate PMI fields of Plume data set.

through specific value combinations, is mostly lacking. Specific information measures were used in the past [3] to identify salient isocontours of one variable by observing the distribution of the other variable. In this work, we present a workflow that enables the users to employ more detailed analysis of a collection of variables by identifying their salient value combinations. Characterization of saliency for a specific value combination is a challenging task in a multivariate system. Given two variables and a pair of scalar values selected from them, the existence of a strong association between the value pairs can be concluded if they demonstrate high co-occurrence. The distribution of these value pairs in the spatial domain can represent a joint multivariate feature. Similarly, when the individual occurrences of the values dominates over their co-occurrence as a pair, the value pair tends to follow a complementary distribuion. Classification of scalar value pairs from this viewpoint can enhance the understanding of multivariate interaction significantly. Here we introduce the information measure that quantifies the shared information for a specific value combination. For two random variables X and Y, if x is a specific observation of X and y for Y, then the information shared between them can be expressed as

$$PMI(x,y) = \log \frac{p(x,y)}{p(x)p(y)} \tag{1}$$

This information measure is known as the pointwise mutual information (PMI), which was first introduced in the work of Church and Hanks [7] for the estimation of word association norms directly from computer readable corpora. When $p(x,y) > p(x)p(y)$, $PMI(x,y) > 0$ which means x and y have higher information sharing between them. If $p(x,y) < p(x)p(y)$, then $PMI(x,y) < 0$ indicating the two observations follow complementary distribution. When x and y do not have any significant information overlap then $p(x,y) \approx p(x)p(y)$ and $PMI(x,y) \approx 0$. In this case, x and y are considered as statistically independent.

Using PMI, pointwise analysis of two random variables can be performed. The sign and absolute value of this PMI measure enables the categorization of the variable interaction as mentioned above. Figure 2 explains the concept of PMI using the frequency of value occurrence between two variables. If x and y are two scalar values of two different variables and have low spatial overlap as shown in Figure 2a, the PMI value will be negative. This stems from the fact that they have higher individual occurrences and therefore high marginal distributions. However, Figure 2b produces high positive PMI value since they have high overlap in their joint occurrence. Identification of these two types of scalar value combinations in the data set provides a way of identifying value pairs with higher or lower co-occurrences. Hence, using PMI values, both associated and opposite or complementary regions in the data can be identified. So the regions that have opposite information will be the unique features in the data which is best represented by that particular variable only among the selected variables. Similarly, the regions with strong association highlights joint multivariate features characterized by high co-occurrence.

4.2 Multivariate PMI Fields

PMI, as discussed in Section 4.1, quantifies the information shared between every value pair for two variables. However, similar to most of the other information theoretic measures, PMI does not contain the spatial information such as where the value pair occurs. For an effective analysis of scientific data, it is essential to incorporate the spatial perspective in the analysis phase as the scientists are also interested in the locations of the features. Below we describe how to incorporate spatial information in our analysis framework.

Given a multivariate time-varying data set, every spatial point in the domain has two or more scalar values associated with it, one from each variable. Since PMI measures the information content for every possible value pairs, we can use it to obtain the information content for the given spatial location. To facilitate this fine grained analysis of multivariate data that preserves the spatial context, a new scalar field can be created that we called the *PMI field*. The PMI field is defined as a scalar field where the spatial points contain the PMI values computed from the values of the variables at the point. If ζ is the scalar function that maps each spatial point to its PMI value, this multivariate interaction field can be expressed as $\zeta : P \mapsto PMI(P)$ where P is a spatial location.

Figure 3 shows one illustrative example where different velocity and gradient fields of the Solar Plume data set [21] are used to construct the multivariate PMI fields. Figure 3a, 3b, 3c and 3d show the scalar volumes of Gradient field, X-velocity, Y-velocity and Z-velocity respectively. Figure 3e shows the PMI fields of gradient and Z-velocity. It is evident from Figure 3e that both Z-velocity and gradient have stronger activity in the turbulent region of the plume, which indicates that the scalar value pairs in this region have higher co-occurrence resulting positive PMI values. However, around the turbulent region, Gradient field has unique activity that is missing in the Z-velocity field. In the PMI field, this region is identified as opposite information which is unique to the Gradient field.

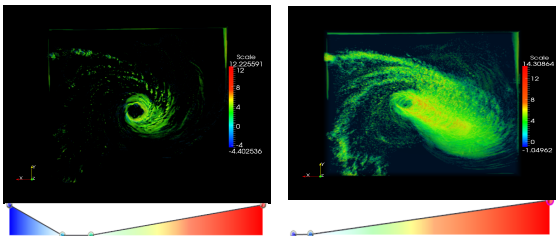
Equation 1 presented in the previous section deals with two variables at a time, however PMI can be generalized for n variables. Wantanabe proposed the concept of *Total correlation* in his work [29]. For a given set of n random variables $X_1, X_2, X_3, ...X_n$, the total correlation can be defined as

$$TC(X_1,X_2,X_3,..,X_n) = \sum_{\substack{x_1 \in X_1, \\ x_2 \in X_2, \\ \ldots, \\ x_n \in X_n}} \log \frac{p(x_1,x_2,..,x_n)}{p(x_1)p(x_1)...p(x_n)} \tag{2}$$

Using total correlation, information shared among a group of variables can be measured. $TC(X_1,X_2,X_3,..,X_n) \approx 0$ signifies the variables in the group are statistically independent and a high total correlation represents high information overlap among them. From this definition of total correlation Tim Van de Cruys [24] defined *Specific correlation* as

$$SC(x_1,x_2,x_3,..,x_n) = \log \frac{p(x_1,x_2,..,x_n)}{p(x_1)p(x_1)...p(x_n)} \tag{3}$$

The above equation is similar to equation 1, but can deal with more than two variables. Equation 3 has similar interpretations like Equation 1 as discussed earlier in Section 4.1 and we use this to construct multivariate PMI fields.



(a) PMI field of CLO and PRE at time step 30 of hurricane Isabel data set. High and low valued regions are emphasized.

(b) Time aggregated PMI field of CLO and PRE between time steps 21-47.

Fig. 4: PMI fields of Cloud (CLO) and Precipitation (PRE) variables of hurricane Isabel data set. Time steps between 21-47 are used for aggregation.

Exploration of multi-fields using the PMI field directly has unique advantages. It allows scientists to look at the data in spatial domain where the information content at each location has a meaningful interpretation. Since PMI values can be used to classify each spatial point into a specific category discussed in Section 4.1, regions with different PMI values can be easily visualized. By focusing on the regions with positive and larger values in the PMI field, users can identify regions where the selected variables are showing stronger association which is indicative of a possible existence of a joint feature. Also regions in the data that have negative PMI values can provide information about the unique characteristics of the selected variable. Figure 3f is the PMI field when three velocity fields are used together and the core of the turbulent region is highlighted as the combined activity region. Here, the trail of Z-velocity is identified as a unique feature that translates into a complementary region in this case.

4.3 Temporal Aggregation of Multivariate PMI Fields for Time-varying Feature Analysis

In this section, we discuss how we can extend the idea of PMI field, presented in Section 4.2 into temporal domain for time-varying feature analysis. Since PMI fields allow us to classify the spatial region into several segments depending on their information content, we can perform this individually in every time step. However, the dynamic behavior of the time-varying data poses unique challenges to data-analysis. Relationships among variables can change significantly as the time changes which further complicates the analysis process. In this scenario, instead of exploring the individual time steps, analysis of the global trend carries more importance and can provide useful insight.

To achieve this goal, several PMI fields are aggregated into a single scalar field using some aggregation functions. The aim for this aggregation is to combine information from a sequence of time steps into a single scalar filed for capturing time-varying phenomena. For example, if a time-varying feature is identified in the PMI field as a joint activity region and if the feature moves spatially over time, then at every individual time step the feature can be found by focusing on the regions of higher joint activity. To study the time-varying nature of the feature, it is useful to have a single representation where the overall trend of the feature becomes prominent. Our aggregation strategy is designed to facilitate this type of analysis which can leverage the usefulness of feature tracking techniques. Woodring and Shen [31] presented techniques for aggregating multiple time steps into a single volume in the univariate case. Our aggregation strategy however, takes into account multiple variable interactions while creating the combined field. To aggregate various types of information, criteria like max, min, mean etc. can be used. If we use max as our criterion, for example, then at every spatial location, PMI values for all the selected time steps are observed and the maximum value among them is picked. This is done for every spatial location over a user selected sequence of time steps. Formally, for a spatial location P, the aggregation value is calculated as:

$$AggPMI(P) = \phi(PMI_i(P)), \forall\, i = t_s, t_s+1, .., t_e \tag{4}$$

where t_s and t_e represents starting and ending time steps of the chosen time window, $PMI_i(P)$ is the value of pointwise mutual information of point P at time step i, and ϕ is the aggregation function.

If several variables consistently show stronger joint activity for a sequence of time steps, our aggregation method will be able to capture its temporal evolution. For each time step, the joint activity will be concentrated in the high valued regions of the PMI field. Hence, a suitable choice of aggregation function will be the max function in this case, which will select the maximum PMI value for every spatial location comparing the PMI values from all the time steps for that location.

In Figure 4 PMI fields of CLO and PRE variables from the hurricane Isabel data set has are shown for demonstration purposes. Time steps between 21-47 are selected for the analysis. Figure 4a shows the PMI field at time step 30 and Figure 4b is the time aggregated PMI field of time steps between 21-47. Max is used as the aggregation function in this example. Inspecting the individual PMI fields, it can be seen that CLO and PRE fields have higher joint activity around the eye of the storm and on the rain bands. From Figure 4b, which is the aggregated PMI field, it is clear that the aggregation field is capable of capturing the movement of hurricane and also the core of the storm.

4.4 Temporal Information Variation of Scalar Value Combinations

After scientists identify time-varying features using the joint temporal PMI fields, the next step is to identify salient scalar value combinations from multiple variables that construct joint or complementary features. This is an important but non-trivial task since identification of such value combinations provides a detailed understanding about the interaction of specific scalar values in multivariate time-varying domain. The problem of salient scalar value selection has been studied in the past for univariate systems and for multivariate domain as well. However, little study has been done in selection of scalar value combinations in multivariate temporal domain. In this section we discuss a new analysis technique which facilitates in the characterization of saliency of a value combination.

Every value combination has a specific PMI value for each time step. If a sequence of time steps have been selected for analysis, a particular value combination will have PMI values at every time step. Using this PMI values, a time series for each value combination is constructed. If a time series has always positive PMI values in a user selected time window, it can be said that the value combinations which created this time series has higher joint temporal activity and they tend to co-occur more in the spatial domain over time. Similarly, the value combinations which have always negative PMI value for all the time steps can be identified. These time series have opposite information throughout the selected time window. We initially divide all the time series into two groups. One group containing time series with all positive values and the other with all negative. Then, inside each group we observe the fluctuation of PMI values for each time series. A time series with high variation of PMI values over time indicates that the co-occurrences of the scalar values in this particular value combination is not consistent over time. Similarly a value combination with lower temporal PMI variation indicates that their co-occurrences in the temporal domain is more consistent and frequent. We measure the variation of one time series as:

$$Var(TS_i) = \sqrt{\sum_{j=0}^{t-1} |TS_{i,j} - TS_{i,j+1}|^2} \tag{5}$$

Here, TS_i is the ith time series, and t is the number of time steps within the selected time window. A high variation value indicates

that the value combination for that time series has weaker association among them and their occurrences are not significant over the selected time step range. Similarly the time series which has low variation are likely to reveal a region where the value combination have higher association among them. The isocontours on such region will have higher overlap and show stronger joint temporal activity.

4.5 Variable Interestingness in Multivariate Temporal Interaction

While exploring a multivariate system that contains a large number of variables, typically the first task is to select initial variables that can be processed further. When little prior knowledge is available regarding a system of variables, it is desirable to provide a guidance to the users to help them make the selection by defining *variable interestingness*. We hypothesize that an interesting variable should have high information overlap with the other variables, i.e. knowing this variable, the users would gain information regarding the other variables of the system. Also, this initial selection will allow us to further analyze using the previously mentioned PMI measure. To achieve this, the mutual information (MI) measure can be used which is interpreted as the aggregation of PMI values over all the scalars. Mutual information is defined as:

$$MI(X,Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} = E_{(X,Y)}[PMI(x,y)] \quad (6)$$

where X and Y are two random variables. Since different variables can have varying amounts of information overlap with other variables, we define *variable interestingness* as the averaged MI of a variable calculated based on all the other variables. Thus, for n random variables, the variable interestingness of variable X_i at time step t is given as:

$$I_t(X_i) = \sum_{j\in n, j\neq i}\frac{MI_t(X_i,X_j)}{n-1}. \quad (7)$$

For a static multivariate system, *variable interestingness* is defined for a given time step. For a multivariate time varying data set, where the relationships among the variables change over time, variables can be identified that are *temporally interesting*. We conclude that, for a variable to be interesting in the time domain, its relationship should be changing over the time given the other variables of the system. This is quantified by observing the change of MI over the time for a selected variables which depicts how the information overlap (or the correlation) of the this variable changes with time. Formally, this can be captured by calculating the temporal gradient of MI for a given variable as:

$$I_t^{'}(X_i) = \frac{I_{t+1}(X_i) - I_{t-1}(X_i)}{2} \quad (8)$$

Specifically, $I_0^{'} = 0$ and $I_T^{'} = 0$ (T is the last time step) for all the variables. Following this, variables can be categorized according to their interestingness and selected for detailed analysis.

5 Interactive Interface Design

Section 4 presents an exploration method which allows users to study multivariate time-varying relationships from the perspective of specific value combination interactions. This type of analysis enables the scientists to directly interact with the multivariate data measured at each spatial location. In order for the scientists to gain sufficient insights about the specific data values, it is imperative to provide them a suitable and intuitive interface for data and variable selection. The goal of such interface is to add the domain scientists into the analysis loop, where they can create or refine hypotheses and confirm that with appropriate visualizations. Next we discuss the individual components of our interface.

5.1 Identifying Temporally Related Variables.

In the initial phase of analysis, if scientists are not sure about what variables to look at first, it becomes cumbersome for them to continue

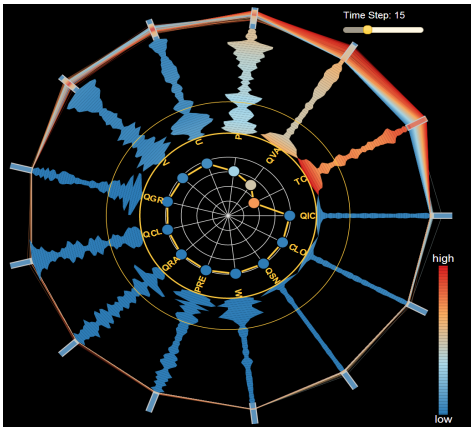
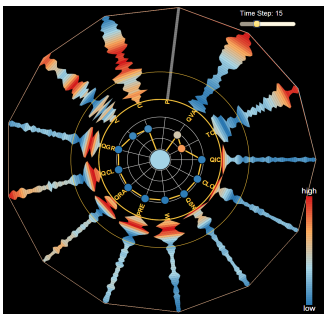
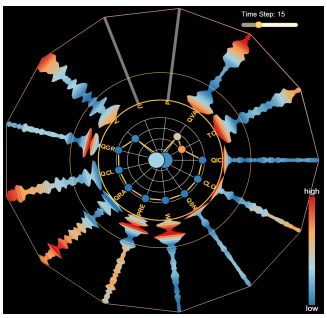


Fig. 5: Variable selection interface with hurricane Isabel data set. Initial stage when no variable is selected.



(a) Variable selection of hurricane Isabel data set when variable P is selected.



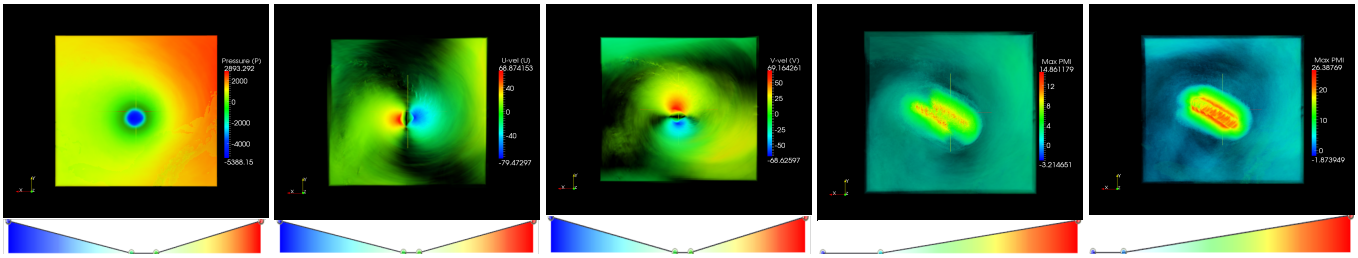
(b) Variable selection of hurricane Isabel data set when both P and U are selected.

Fig. 6: Variable selection for hurricane Isabel data set

further analysis in depth. To facilitate this requirement, an information theory guided novel variable selection interface is designed. Our variable selection interface presents the information content of the data in an interactive layout. Figure 5 shows such interface where hurricane Isabel data set is used for demonstration purposes. All the variables are arranged at the center using a *radar plot*, where multivariate interestingness (discussed previously in Section 4.5) for each variable is represented by their distances to the center. This multivariate interestingness is inversely proportional to the length of the radius, so if the interestingness value is high, then that variable will be closer to the center. Using this radar plot, users can visually inspect the multivariate interestingness of all the variables at a specific time step and select multiple variables for in depth analysis. Once a variable is selected, it is then highlighted and placed at the center as can be seen in Figure 6a.

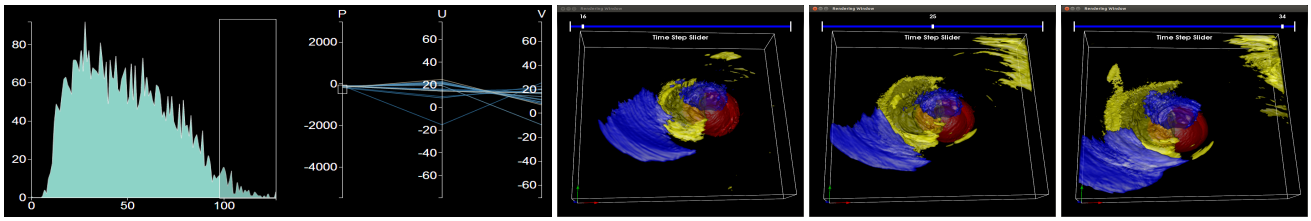
To incorporate the temporal information into the variable selection interface, a *time plot* has been used along with the radar plot. A time plot associates a time axis with every variable, placed radially around the radar plot as can be seen in Figure 5. Time progresses from the center, radially outwards. To provide the information about variable's temporal interestingness, the time axes are divided into multiple small segments, one for each time step. The segments are colored by the interestingness value (red for high value and blue for low value) of that time step and the width of the segment is modulated by the temporal gradient of the interestingness values (defined similarly as in (8)). Given this configuration, scientists can have four different criteria for variable selection:

1. Select the variable that has higher red regions in its time axis and also the axis width is non-uniform. This signifies that, such a variable has high information overlap with all the other variables at multiple time steps, hence the red color, but the infor-



(a) Pressure (P) field of hurricane Isabel data set at time step 30. (b) Uvel (U) field of hurricane Isabel data set at time step 30. (c) Vvel (V) field of hurricane Isabel data set at time step 30. (d) Aggregated PMI field of P and U between time steps 15-35. (e) Aggregated PMI field of P, U and V between time steps 15-35.

Fig. 7: Visualization of hurricane Isabel data set.



(a) Selection of value combination of P, U and V. Value combinations with high temporal PMI variation have been selected in this case. (b) Simultaneous isocontour visualization at P = -80.8013, U = -15.1759 and V = 25.397 at time step 16. (c) Simultaneous isocontour visualization at P = -80.8013, U = -15.1759 and V = 25.397 at time step 25. (d) Simultaneous isocontour visualization at P = -80.8013, U = -15.1759 and V = 25.397 at time step 34.

Fig. 8: Temorally salient isocontour identification of hurricane Isabel data set using P, U and V field.

mation overlap changes constantly over time, indicated by the non-uniform width.

2. Select the variable that shows higher red regions in its axis, but the width is uniform. In this case the variable has consistent high information overlap with others.

3. Select the variable that has blue regions in its axis, however the width is non-uniform which signifies low information overlap but high temporal variation.

4. The final choice can be the variable that shows more blue regions and a uniform time axis. These variables have low information overlap and low information variation as well.

Once a variable is selected, the view of the interface changes and provides information about each variable with the reference variable, which is already been selected. Figure 6a shows the interface when variable P is selected and placed at the center. The color of the time axis for the selected variable becomes gray to indicate that it is currently selected for exploration. Once a variable is selected, the information presented along the time axes change to pairwise mutual information. So in Figure 6a the color in the time axes is the mutual information between P and every other variable. Similarly, the width of the time axes now becomes temporal gradient of pairwise mutual information. This allows users to investigate information overlap between every pair of variables and their temporal variation which facilitates in further variable selection.

In addition, radial coordinates plot is created to display the multivariate data distributions for the user to select a data subspace of interest through interactive brushing. Initially a subsample of data is randomly selected to show the overall pattern of data distributions. The length of the axes in the radial coordinates plot can be adjusted by reducing the length of time axes when users want to study the data distributions using it. A video of our interactive framework has been included with the submission for further demonstration.

5.2 In Depth Analysis on Selected Variables Using Multivariate Time-varying PMI fields.

Using the proposed variable selection interface, multiple variables can be identified with a time window where they have high information overlap/variation. PMI fields using such variables are created for identification of informative regions in spatial domain. Figure 4 shows such PMI fields constructed using CLO and PRE variable of hurricane Isabel data set. Users can interactively select appropriate opacity map to suppress unwanted values and focus on interesting regions. From Figure 4 it can be seen that regions around the eye of the hurricane and the rain bands are identified as informative regions. To capture time-varying features, several PMI fields are combined using different aggregation functions. The rightmost image in Figure 4 shows one time aggregated field where the overall movement of the hurricane with time is seen. After the informative regions and temporal trends are identified using the PMI fields and aggregated volumes, we allow users to interact with specific value combinations such that the scalar values creating such joint temporal features can be specifically identified.

5.3 Temporal Variation Based Identification of Salient Isocontours.

Since the possible number of value combinations increases significantly with the addition of every variable in analysis space, an effective interactive tool is required to facilitate the selection of specific value combinations. In Section 4.4 we show how the variation of a value combination can be measured. Using these variation values, a probability distribution of all possible value combinations is created. Such a distribution is presented in Figure 8a. This distribution allows us to group value combinations with similar variation as they fall in the same histogram bins. This makes the huge number of value combinations tractable. By brushing a few bins of such a histogram, users can select a subset of value combinations with high or low variation. A Parallel Coordinates Plot (PCP) is attached with the histogram so that the specific data values that are selected by brushing the histogram can be readily visualized. The color in the PCP specifies the degree of variation for each value combination, that provides additional visual cues to the users. Interacting with the PCP interface, scientists can select

specific value combinations and interactive isocontour visualization is provided for inspecting the behavior of such value combination in temporal domain.

6 RESULTS

The experiments were conducted on a Linux machine with an Intel core i7-2600 CPU, 16 GB of RAM and an NVIDIA Geforce GTX 660 GPU with 2GB texture memory. For the calculation of information-theoretic measures, 256 histogram bins were used. As the number of bins effects the accuracy of the information metric calculation, several bin numbers 64, 128, 256, and 512 were tested. It was observed from the experiments that the general pattern remains the same with minor variations.

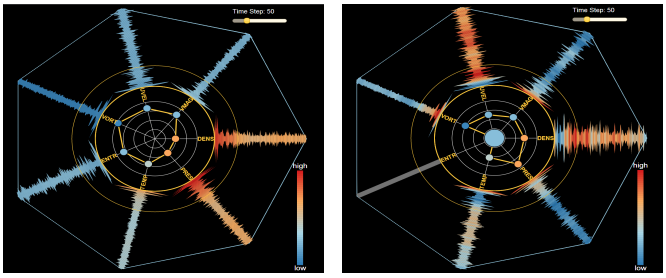
6.1 Hurricane Isabel Data set

The Hurricane Isabel data is a multivariate time-varying data consisting of 13 scalar fields: Pressure (P), Total cloud moisture mixing ratio (CLO), Cloud moisture mixing ratio (QCL), Graupel mixing ratio (QGR), Snow mixing ratio (QSN), Rain mixing ratio (QRA), Water vapor mixing ratio (QVA), Precipitation (PRE), Temperature (TC), X wind speed (U), Y wind speed (V), and Z wind speed (W). The data set is a courtesy of NCAR and the U.S. National Science Foundation (NSF), and was created using the Weather Research and Forecast (WRF) model. The resolution of the grid for each time step is $250 \times 250 \times 50$ and there are total 48 time steps.

Figure 5 represents the initial layout of the variable selection interface for the Isabel data set. From the layout, it is observed that variable P has a larger information variation with all the other variables around time step 15. This is identified by the width of P axis and observing the amount of information sharing between this variable and other variables. As time progresses, however, the variation gradually decreases and then again starts to increase beyond time step 30. This prompts the selection of variable P as our initial choice. Figure 6a presents the scenario when P is selected. Now, the interaction of P with rest of the variables can be analyzed. Variable U shows a high information overlap with P in the time step range 15-35. So the next variable of choice is U and the time step range for analysis is 15-35. In Figure 6b, the state of the interface is presented where both P and U are selected. Following our variable selection method, after P and U have been selected, next possible choice can be V or QRA since both have high information variation with P and U in the selected time step range. For this case study, variable V is selected as the third variable for demonstration purposes.

Figure 7a, 7b and 7c show the individual volumes for the selected variables at time step 30. Figure 7d is the aggregated PMI field of P and U, and Figure 7e shows the aggregated PMI field of P, U and V. The *Max* function is used as the aggregation function to highlight the joint activity regions in this illustration. From Figure 7d and 7e, it can be seen that the hurricane eye is well represented by the combined activity of P, U and V. Our proposed aggregation strategy enables us to readily identify the path of the storm over time. Tracking of the hurricane eye is an important task for this type of data set, and the results presented here demonstrate the ability of the proposed system in capturing such time-varying features using time aggregated PMI fields.

Next we identify salient value combinations of P, U and V that show such joint features in this data set. Figure 8a shows the distribution of value combinations that possesses positive PMI value for every time step in the selected time range. Time series that have high variation are selected by brushing the histogram, and the scalar value combinations from the selected ranges are shown in the PCP. Simultaneous isosurface visualization is presented in Figure 8b, 8c, and 8d for time steps 16, 25 and 34 respectively. The transparency is applied to reveal the inner structures of the overlapping isosurfaces. The Red isosurface corresponds to P, yellow isosurface to U and blue isosurface corresponds to V. It can be seen that, at P = -80.8013, U = -15.1759 and V = 25.397 the isosurfaces of P, U and V show joint activity in representing the storm structure of hurricane Isabel. A time slider has been provided so that, users can interactively see the temporal evolution of such isosurfaces within the selected time range.



(a) Variable selection interface of aerodynamic data set.

(b) Variable selection interface of aerodynamic data set when Entropy is selected.

Fig. 13: Variable selection interface for the aerodynamic data.

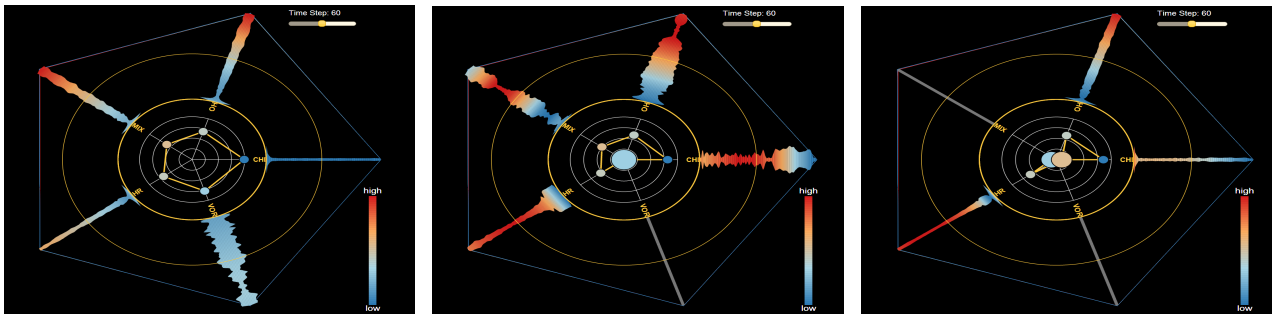
6.2 Combustion Data set

The Combustion data set is a multivariate time-varying turbulent simulation data set. It has five variables, Mixture Fraction (MIX), Vorticity (VOR), Mass Fraction of Hydroxyl (OH) radical, Heat Release Rate (HR), and Scalar Dissipation Rate (CHI) in turbulent flames. The resolution at every time step for each variable is $240 \times 360 \times 60$. The data set was made available by Dr. Jacqueline Chen at Sandia Laboratories through US Department of Energy′s SciDAC Institute for Ultra-scale Visualization.

Figure 9a shows the initial state of our variable selection interface. Since VOR shows a high information variation with rest of the variables over time, we initially select this variable to start the exploration process. Figure 9b shows an update of the interface when VOR is selected. From Figure 9b, we can conclude that between time steps 60-100, MIX has a high amount of shared information with VOR and the variation of shared information initially decreases and later increases. To analyze this in more details, we select MIX as our second variable. Figure 9c displays the state of our variable selection interface when both VOR and MIX have been selected. Since OH variable also shows high information overlap in this time step range, OH becomes our third variable of choice for further analysis.

Figure 10a, 10b, and 10c show VOR, MIX and OH fields at time step 65. In Figure 10d, the aggregated volume of these three fields between time steps 60-100 is presented. The *Min* function was used as the aggregation function in this case. In Figure 10d, low valued regions indicate the locations where these three variables do not show any joint feature. Consequently, most of the points in that region contain negative PMI values. The region which shows positive values are also highlighted by appropriate opacity function and it identifies the region where the three variables act together. This is the region where the fast chemical reactions takes place and is of interest to the scientists.

Figure 11b, 11c and 11d show the simultaneous isosurfaces of the selected variables at VOR = 100947, MIX = 0.372632 and OH = 0.000610 for the time steps 65, 80 and 95 respectively. These particular value combinations were obtained by brushing the high variation region of the histogram as shown in Figure 11a. The red isosurface corresponds to VOR, yellow for MIX and blue isosurface is the contour of OH. From their simultaneous visualization, it is seen that these three specific values of VOR, MIX and OH tend to co-occur consistently over time. Isocontour of MIX = 0.42, the stoichiometric MIX, represents the flame structure in this data set. Around the region of flame, the chemical reaction rate is generally higher than the turbulence generation, so lower turbulence values are observed in these regions. These regions also develop high temperature and radical concentration. The value combination resulted from our analysis indicates similar trends and it shows that such regions have higher information variation, which might be due to the fast chemical reaction.

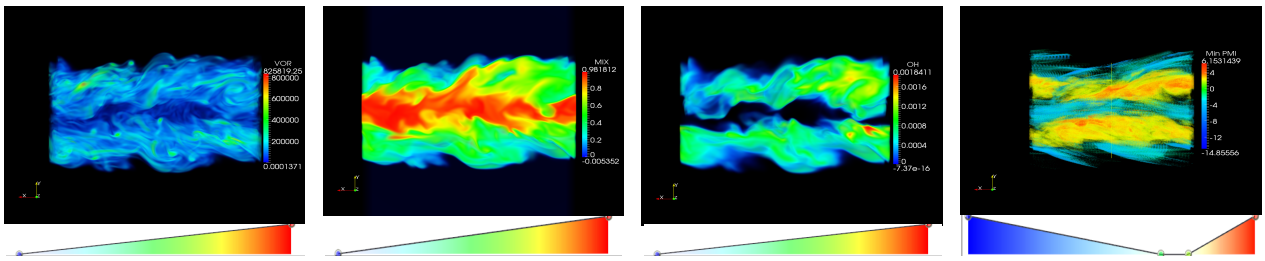


(a) Variable selection interface loaded with Combustion data set. Initial stage when no variable is selected.

(b) Variable selection interface of Combustion data set when VOR is selected.

(c) Variable selection interface of Combustion data set when VOR and MIX are selected.

Fig. 9: Variable selection interface for Combustion data set



(a) Vorticity (VOR) field of Combustion data set at time step 65.

(b) Mixture fraction (MIX) field of Combustion data set at time step 65.

(c) OH Radical (OH) field of Combustion data set at time step 65.

(d) Aggregated PMI field of VOR, MIX and OH between time steps 60-100.

Fig. 10: Visualization of Combustion data set.

Table 1: CPU Time performance (in seconds) for different computation components.

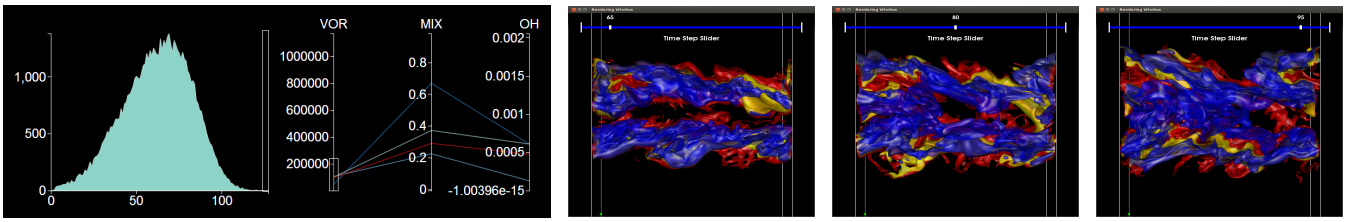
Data Sets	Histogram	All pair MI	AVG Histogram	AVG All pair MI	PMI	PMI Aggregation
Isabel ($250 \times 250 \times 50 \times 48$)	1098.288	4.3056	22.881	0.0897	16.962	4.3101
Combustion ($240 \times 360 \times 60 \times 122$)	1418.372	1.8666	11.626	0.0153	98.3036	14.232

7 Aerodynamic Data Set and Expert Feedback

In this section, we present our results on an Aerodynamic data set and discuss the feedback we received from the experts. To start the process, initially the overall idea and the purpose of each component were explained to the scientists. Then we discussed in detail about our system with examples. After the scientists became familiar with our framework, our system was used to generate results using their data, which were generated from a simulation of a high-speed engine compressor stage undergoing a rotating stall process. This simulation was conducted using a time-accurate Navier-Stokes turbomachinery flow solver. The compressor stage consists of 36 rotor blades, which are followed by 46 stator blades. Included in the data are five fundamental aerodynamic variables: density, velocity momentum in x/y/z direction, and total energy. These five fundamental variables can be used to derive variables such as pressure, temperature, and entropy. The data includes 192 time instants starting from about a quarter revolution before the stall inception to one revolution after.

To conduct the experiment, we used the Entropy (ENTR), Pressure (PRES), Uvelocity (UVEL), Velocity Magnitude (V_MA), Temperature (TEMP), and Vorticity (VORT) fields. The domain scientists first visualized some pre-generated images, and then with our help, started to explore the interaction of the UVEL and ENTR variables during an engine pre-stall phase in their simulation. Figure 13a shows the initial state of our variable selection interface and Figure 13b shows the the interface when ENTR is selected. From Figure 13a it is observed that ENTR and UVEL have high variation in information throughout the time scale. Following our framework, these two variables were identified as the possible choices as the initial variable. When ENTR was selected, from Figure 13b, it can be seen that UVEL shows high information overlap (Red color on the UVEL axis) as well as high variation with ENTR within time step sequence 50-100. Therefore, time steps between 50-100 were selected for further analysis. Figure 12b displays the aggregated volume of PMI fields across time steps 50-100. The *Min* function was used as the aggregation criteria in this experiment. From the Figure 12b, it is apparent that the inner regions of the stage has low combined activity between ENTR and UVEL, so the value combinations have mostly negative PMI in this region. However, around the circumference, ENTR and UVEL show strong combined activity.

Next we analyze specific value combinations for salient scalar value selection. Figure 12a shows the selection of value pairs using the histogram and PCP interface. Simultaneous isocontour visualization at ENTR = 1.08251 and UVEL = -0.00939441 is presented in Figure 12c and 12d for time steps 60 and 100 respectively. The yellow surface indicates the isocontour of UVEL and the green surface corresponds to ENTR. Since we choose the value pairs from the regions where UVEL and ENTR have always positive PMI values and low variation of PMI, this suggests that, in those values, the occurrences of ENTR and UVEL is consistently high within the selected time window. Domain scientists provided further explanation on the obtained results. According to the expert, one of the main reasons that causes stall in this type of engine, is the flow separation that can be identified by negative UVEL regions. Negative UVEL represents the regions where the flow reverses its normal direction and this can initiate turbulence. As the negative UVEL spreads over a wide region as time progresses, ENTR starts to show high values in those regions. The co-occurrences of negative UVEL and high ENTR values increase as the engine proceeds towards a stall which can be explained from the results we received. At a critical situation, the engine finally stalls.



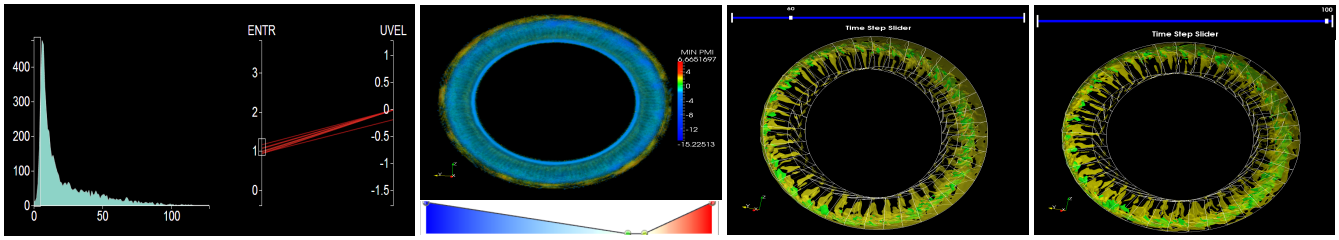
(a) Selection of salient value combination of VOR, MIX and OH. Scalar values with high temporal PMI variation have been selected in this case.

(b) Simultaneous isosurface visualization at VOR = 100947, MIX = 0.372632, and OH = 0.000610 at time step 65.

(c) Simultaneous isosurface visualization at VOR = 100947, MIX = 0.372632, and OH = 0.000610 at time step 80.

(d) Simultaneous isosurface visualization at VOR = 100947, MIX = 0.372632, and OH = 0.000610 at time step 95.

Fig. 11: Temorally salient isosurface identification of Combustion data using VOR and MIX field.



(a) Selection of value combination of Entropy and Uvel. Value combinations with low temporal PMI variation have been selected in this case.

(b) Time aggregated PMI field of ENTR and UVEL in selected time step range 50-100. The circumference region show higher joint activity between ENTR and UVEL.

(c) Simultaneous isosurface visualization of ENTR = 1.08251 and UVEL = -0.00939441 at time step 60.

(d) Simultaneous isosurface visualization of ENTR = 1.08251 and UVEL = -0.00939441 at time step 100.

Fig. 12: Time aggregated PMI field visualization and temporally salient isosurface identification of aerodynamic data set using ENTR and UVEL fields. Salient isosurfaces indicate rotating stall inception in the turbine stage.

Results presented in this case study show that the negative UVEL and higher ENTR values tend to co-occur more in the circumference of the engine stage, which indicates joint activity of ENTR and UVEL and can be used as a precursor to stall inception.

We asked the domain scientists to provide general suggestions about our overall approach. The domain scientists showed interest in our variable selection interface and agreed that, the ability to integrate correlations of multiple variables of the entire history of a time-accurate simulation in one central place as shown in Figure 13, and histogram in Figure 12a are powerful tools to analyze unsteady data. These visualizations allow users to systematically interrogate unsteady data in real-time and identify interesting correlations that otherwise would be lost in the sea of a large data set. The experts have mentioned that, the ability of the centralized tools to quickly link to the spatial representation of the data as demonstrated in Figure 12 allows users to retrieve detailed flow data in real time for close examination. The domain experts also encouraged us to extend our current system for analyzing big data such as the compressor stall simulation by using data reduction schemes to significantly reduce the size of unsteady data and devise efficient ways to reconstruct individual solutions.

8 Performance

In this section, we discuss the performance of different computation components in our system, as shown in Table 1. In the preprocessing stage, histograms are constructed to compute the mutual information between variables. The histogram computation time (shown in the second column) depends on the total number of data points for all the variables, while the mutual information computation (shown in the third column) relies on the number of variables in the system. The fourth and fifth column show the averaged time for the two computation components per time step, respectively. In column six and seven, the time for PMI calculation and PMI field aggregation is shown. These two quantities depend on the number of time steps and number of variables that have been selected for analysis. For the Isabel data set, recorded time was obtained for analysis on 20 time steps and 3 variables, and for the Combustion data set the time reported here was obtained for analysis on 40 time steps and 3 variables. Although we have not fully optimized the computation during preprocessing, it can be further reduced through parallel computing for individual time steps, since there is no dependency in the computation at different time steps. The computation time of variable interestingness and gradient of interestingness is not listed here as it is in the order of milliseconds and can be done in real-time.

9 Conclusion and Future Work

In this paper, we present an information theoretic framework for exploration of multivariate time-varying data sets. We use pointwise mutual information to measure the information content for specific value combinations of multiple variables and further use such information to construct PMI fields. The PMI field allows us to analyze variable relationships using the pointwise mutual information keeping the spatial perspective intact. For identification of time-varying features, we extend the PMI field into temporal domain by aggregating several PMI fields using various aggregation criteria. To identify salient value combinations, we measure temporal information variation of each value combination and grouped the value combinations using their variation values. Interactive visualizations are provided for selection of salient value combinations and isocontours are used to visually inspect the saliency of selected value combinations in spatial domain.

In the future, we wish to use our framework to select key time steps. We will also use our framework for different data types such as ensemble data. Also, we would like to make suitable adjustments to our current framework to apply our analysis technique for In-Situ data analysis.

References

[1] H. Akiba, N. Fout, and K.-L. Ma. Simultaneous classification of time-varying volume data based on the time histogram. In *Proceedings of the Eighth Joint Eurographics / IEEE VGTC conference on Visualization*, pages 171–178, 2006.

[2] C. Bajaj, V. Pascucci, and D. Schikore. The contour spectrum. In *Visualization '97., Proceedings*, pages 167–173, Oct 1997.

[3] A. Biswas, S. Dutta, H.-W. Shen, and J. Woodring. An information-aware framework for exploring multivariate data sets. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2683–2692, 2013.

[4] U. Bordoloi and H.-W. Shen. View selection for volume rendering. In *Visualization, 2005. VIS 05. IEEE*, pages 487–494, 2005.

[5] R. Bramon, I. Boada, A. Bardera, J. Rodriguez, M. Feixas, J. Puig, and M. Sbert. Multimodal data fusion based on mutual information. *Visualization and Computer Graphics, IEEE Transactions on*, 18(9):1574 – 1587, sept. 2012.

[6] H. Carr, J. Snoeyink, and U. Axen. Computing contour trees in all dimensions. In *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '00, pages 918–926, Philadelphia, PA, USA, 2000. Society for Industrial and Applied Mathematics.

[7] K. W. Church and P. Hanks. Word association norms, mutual information, and lexicography. In *Proceedings of the 27th annual meeting on Association for Computational Linguistics*, ACL '89, pages 76–83, Stroudsburg, PA, USA, 1989. Association for Computational Linguistics.

[8] T. M. Cover and J. A. Thomas. *Elements of Information Theory 2nd Edition (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[9] M. de Oliveira and H. Levkowitz. From visual data exploration to visual data mining: a survey. *Visualization and Computer Graphics, IEEE Transactions on*, 9(3):378–394, 2003.

[10] B. Duffy, H. Carr, and T. Moller. Integrating isosurface statistics and histograms. *Visualization and Computer Graphics, IEEE Transactions on*, 19(2):263–277, Feb 2013.

[11] M. Feixas, E. D. Acebo, P. Bekaert, and M. Sbert. An information theory framework for the analysis of scene complexity, 1999.

[12] R. Fuchs and H. Hauser. Visualization of multi-variate scientific data. *Computer Graphics Forum*, pages 1670–1690, 2009.

[13] L. Gosink, C. Garth, J. Anderson, E. Bethel, and K. Joy. An application of multivariate statistical analysis for query-driven visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 17(3):264–275, March 2011.

[14] S. Gumhold. Maximum entropy light source placement. In *Visualization, 2002. VIS 2002. IEEE*, pages 275–282, 2002.

[15] M. Haidacher, S. Bruckner, A. Kanitsar, and M. E. Gröller. Information-based transfer functions for multimodal visualization. In W. N. C.P Botha, G. Kindlmann and B. Preim, editors, *VCBM*, pages 101–108. Eurographics Association, Oct. 2008.

[16] H. Jänicke, M. Bottinger, and G. Scheuermann. Brushing of attribute clouds for the visualization of multivariate data. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1459–1466, 2008.

[17] H. Jänicke, M. Böttinger, X. Tricoche, and G. Scheuermann. Automatic detection and visualization of distinctive structures in 3d unsteady multifields. *Comput. Graph. Forum*, 27(3):767–774, 2008.

[18] H. Jänicke and G. Scheuermann. Visual analysis of flow features using information theory. *Computer Graphics and Applications, IEEE*, 30(1):40–49, 2010.

[19] M. Khoury and R. Wenger. On the fractal dimension of isosurfaces. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1198–1205, Nov 2010.

[20] A. Martin and M. Ward. High dimensional brushing for interactive exploration of multivariate data. In *Visualization, 1995. Visualization '95. Proceedings., IEEE Conference on*, pages 271–, 1995.

[21] H. A. Observatory. Compressible plume dynamics and stability. 369:125–149, 1998.

[22] N. Sauber, H. Theisel, and H.-P. Seidel. Multifield-graphs: An approach to visualizing correlations in multifield scalar data. *IEEE Transactions on Visualization and Computer Graphics*, 12(5), Sept. 2006.

[23] C. E. Scheidegger, J. M. Schreiner, B. Duffy, H. Carr, and C. T. Silva. Revisiting histograms and isosurface statistics. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1659–1666, 2008.

[24] T. Van de Cruys. Two multivariate generalizations of pointwise mutual information. In *Proceedings of the Workshop on Distributional Semantics and Compositionality*, DiSCo '11, pages 16–20, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.

[25] P.-P. Vázquez, M. Feixas, M. Sbert, and W. Heidrich. Automatic View Selection Using Viewpoint Entropy and its Application to Image-Based Modelling. *Computer Graphics Forum*, 22(4):689–700, 2003.

[26] I. Viola, M. Feixas, M. Sbert, and M. Groller. Importance-driven focus of attention. *Visualization and Computer Graphics, IEEE Transactions on*, 12(5):933–940, 2006.

[27] C. Wang, H. Yu, R. Grout, K.-L. Ma, and J. Chen. Analyzing information transfer in time-varying multivariate data. In *Pacific Visualization Symposium (PacificVis), 2011 IEEE*, pages 99–106, 2011.

[28] C. Wang, H. Yu, and K.-L. Ma. Importance-driven time-varying data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1547–1554, nov 2008.

[29] S. Watanabe. Information theoretical analysis of multivariate correlation. *IBM J. Res. Dev.*, 4(1):66–82, Jan. 1960.

[30] P. C. Wong and R. D. Bergeron. 30 years of multidimensional multivariate visualization. In *Scientific Visualization, Overviews, Methodologies, and Techniques*, pages 3–33, Washington, DC, USA, 1997. IEEE Computer Society.

[31] J. Woodring and H.-W. Shen. Chronovolumes: A direct rendering technique for visualizing time-varying data. In *Volume Graphics*, volume 45 of *ACM International Conference Proceeding Series*, pages 27–34. Eurographics Association, 2003.

[32] J. Woodring and H.-W. Shen. Multiscale time activity data exploration via temporal clustering visualization spreadsheet. *IEEE Trans. Vis. Comput. Graph.*, 15(1):123–137, 2009.

[33] D. Yang, E. Rundensteiner, and M. Ward. Analysis guided visual exploration of multivariate data. In *Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on*, pages 83–90, 2007.

[34] H. Younesy, T. Mller, and H. Carr. Visualization of time-varying volumetric data using differential time-histogram table. In *Proceedings of the Fourth Eurographics / IEEE VGTC Workshop on Volume Graphics 2005, Stony Brook, NY, June 20-21, 2005*, pages 21–29. Eurographics Association, 2005.