

Pointwise Information Guided Visual Analysis of Time-varying Multi-fields

Soumya Dutta

The Ohio State University
dutta.33@osu.edu

Xiaotong Liu

IBM Research-Almaden
xiaotong.liu@ibm.com

Ayan Biswas

Los Alamos National Laboratory
ayan@lanl.gov

Han-Wei Shen

The Ohio State University
shen.94@osu.edu

Jen-Ping Chen

The Ohio State University
chen.1210@osu.edu

ABSTRACT

Identification of salient features from a time-varying multivariate system plays an important role in scientific data understanding. In this work, we present a unified analysis framework based on mutual information and two of its decomposition: specific and pointwise mutual information to quantify the amount of information content between different value combinations from multiple variables over time. The pointwise mutual information (PMI), computed for each value combination, is used to construct informative scalar fields, which allow close examination of combined and complementary information possessed by multiple variables. Since PMI gives us a way of quantifying information shared among all combinations of scalar values for multiple variables, it is used to identify salient isovalue tuples. Visualization of isosurfaces on those selected tuples depicts combined or complementary relationships in the data. For intuitive interaction with the data, an interactive interface is designed based on the proposed information-theoretic measures. Finally, successful application of the proposed method on two time-varying data sets demonstrates the efficacy of the system.

CCS CONCEPTS

• Mathematics of computing → Information theory; • Human-centered computing → Scientific visualization; Visualization techniques; Visual analytics;

KEYWORDS

Information theory, pointwise mutual information, specific mutual information, time-varying multivariate data exploration.

1 INTRODUCTION

Effective feature exploration in time-varying multivariate data sets is challenging as a thorough understanding of the intricate relationships among multiple variables is involved. Oftentimes, instead of looking at the total correlation among variables, scientists search for specific value combinations of multi-variables which show positive or negative association to get in depth knowledge about the interaction of such variables. So, quantifying the importance of individual value combinations of multiple variables has gained significant importance in the recent years. Multi-field analysis based on their value combinations allows experts to understand how the total shared information among variables is distributed within all of its value combinations. It is to be noted that, a majority of the existing methods have mostly focused on studying the average behavior

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, 9, 17, 25, 34], and multi-field domain [3, 26] has been done in the past. But, a guideline to study the relationships of specific value combinations in time-varying multivariate data sets is still missing.

To address the aforementioned issues, an information-theoretic framework is presented in this work to help the scientists to conduct detailed analysis of time-varying multi-fields. It uses mutual information (MI) and two of its decomposition: (1) specific mutual information (SMI), and (2) pointwise mutual information (PMI) to quantify information of scalar values and value combinations. Time-varying study of multi-fields based on their value combinations using PMI constitutes the core of our framework, which is a novel contribution of this work. We observe the fact that, while decomposed hierarchically in a top-down fashion, MI conveys different facets of information, and by integrating all these information in a unified framework, in depth study of time-varying data sets based on their specific value combinations becomes possible.

The proposed framework enables the experts to select variable combinations based on their shared information content, and further guides them to identify interesting features from the data. We quantify information content of every spatial point by calculating its PMI. Using PMI values at each spatial location, a new scalar field is constructed, called *PMI field*, which are segmented into different regions with value combinations having strong co-occurrences or noticeably low association. High co-occurrence in a region indicates the existence of a joint feature, and regions with low co-occurrence are explored for any potential surprise. We employ this idea in the time domain to study the temporal evolution of such regions. By aggregating several PMI fields from consecutive time steps, we capture feature's temporal evolution. In order to identify temporally salient value combinations, we utilize the temporal information content of the value combinations following a refinement-based strategy. Positive feedback from the domain expert demonstrate the usefulness of our framework in time-varying data exploration.

2 RELATED WORKS

Information theory [15, 35, 38] has been used extensively for solving many problems in visualization. Detailed reviews about information theory in visualization can be found in [11, 12, 32, 33, 40].

Mutual information (MI) and Specific mutual information (SMI) in visualization. MI has been extensively used previously for registration and data fusion [14, 22, 23, 28, 30]. Identification

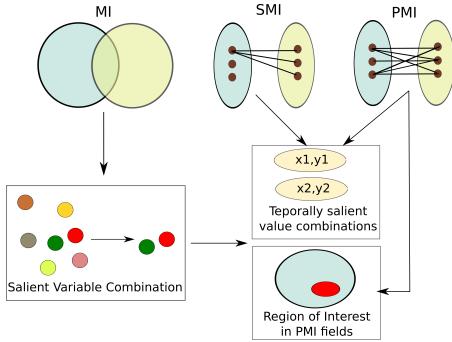


Figure 1: A schematic diagram of our framework.

of the best view and its smooth transition are achieved in [39] by using MI. By calculating MI between the intensity values of data and color pixels, Bramon et al. [6] measure the information transfer between input data and output pixels. Recently, researchers have used MI for the selection of streamlines based on coherent view directions [27, 36]. Bruckner et al. [8] have proposed similarity based exploration of isosurfaces of univariate data using MI. A level-set based method has been presented in [41] to analyze the representativeness of an isosurface of a volumetric data. Finally, MI has also been used for analyzing scene complexity [18], shape complexity [31], and mesh importance analysis [19]. Haidacher et al. [20] have extended MI based analysis in multi-modal domain for analyzing multi-modal surface similarities. In the recent years, researchers have sought after specific mutual information (SMI) because of its ability to quantify information of a scalar value given another variable. Bramon et al. [5] have used SMI to fuse multi-modal data sets. Biswas et al. [3] have created an information map based on SMI, which can be used to select isovalue of one variable which are either predictable or uncertain with respect to the other variable. SMI based transfer function design has also been introduced in [7].

Proposed framework and pointwise mutual information. PMI is a measure in information theory which quantifies *informativeness* of any specific value combination of multi-variables. Based on PMI, the value combinations can be divided into two broad classes which show distinct statistical features in data. The proposed method extends this value combination based analysis in temporal domain and allows users to interactively select temporally salient value combinations which either show strong temporal association or complementary behavior. Even though Haidacher et al. [21] used PMI to retrieve the opposite information, use of PMI to perform time-varying analysis of multivariate data by exploiting the time-varying relationships of their specific value combinations is novel in our approach. It allows us to identify regions in the data set where scalar value combinations show strong or weak statistical association, and we explore how these regions evolve over time.

3 MULTIVARIATE TEMPORAL ANALYSIS FRAMEWORK

Given a time-varying multivariate data set, our analysis allows: (1) selection of several important variable combinations, and a suitable time range for detailed exploration; (2) efficient detection of regions with joint or complementary multivariate features over time; (3)

systematic identification of temporally salient scalar value combinations. Figure 1 shows a schematic view of our complete framework, where three hierarchical variations of MI are highlighted and how they are used to analyze multi-variate time-varying data is shown. In the following, we describe our framework in detail.

3.1 Defining Variable Interestingness

Typically in a time-varying multivariate domains, not all variables are relevant for analysis. Irrelevant variables can make the discovery of important features difficult. So, it is challenging to determine what variable combinations are salient and how the saliency of relationships evolves over time. Fortunately, domain experts usually have some prior knowledge about the data set and their first goal is to confirm those known facts using existing visualization techniques and then hypothesize new theories. Our variable selection technique exploits information theory to provide guidance for the scientists. When scientists start with any specific variable, we analyze the shared information of this reference variable with the other variables. Since this step is the beginning of our unified workflow and only requires us to quantify the total information shared between all the variables, MI seems a good choice for defining the information overlap between variables. It is defined as:

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (1)$$

where X and Y are two random variables, and $p(x), p(y)$ are the marginal probabilities of observation x of X , and y of Y , and $p(x, y)$ is their joint probability. Since MI considers all the possible values and quantifies the total information overlap between two random variables, we can only conclude about the degree of overlap of information between two random variables. For a time-varying data, the relationships among the variables change over time, therefore, variables can be identified based on their *temporal interestingness*. We quantify this by observing the change of MI over the time for selected variables which depicts how the information overlap of the these variables changes with time. Formally, this can be captured by calculating the temporal gradient of MI for a given variable as:

$$I'_t(X; Y) = \frac{dI(X; Y)}{dt} \quad (2)$$

where X and Y are two selected variables and $I'_t(X; Y)$ is the temporal gradient of MI between time steps t and $t + 1$.

3.2 Combined and Complementary Informativeness Characterization

Given a variable combination, in this work, we demonstrate a workflow that enables users to employ more detailed analysis over a pair of variables by identifying their salient value combinations. Given two variables and a pair of scalar values selected from them, the existence of a strong association between the value pair 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 dominate over their co-occurrence as a pair, the value pair tends to follow a complementary distribution. Here we introduce the information measure that quantifies the shared information for

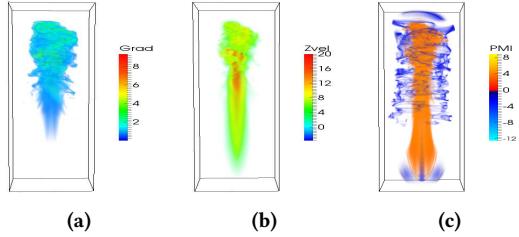


Figure 2: PMI fields of Plume data set. 2a Velocity gradient magnitude, 2b Zvel field, 2c PMI field of two variables.

a specific value combination. For two random variables X and Y , if x is an observation of X and y for Y , then the information content between them is expressed as:

$$PMI(x, y) = \log \frac{p(x, y)}{p(x)p(y)} \quad (3)$$

where $p(x)$ is the probability of a particular occurrence x of X , $p(y)$ is the probability of y of variable Y and, $p(x, y)$ is their joint probability. This information measure is known as the pointwise mutual information (PMI), which was first introduced in the works of Church and Hanks [13] 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. It is to be noted that, mutual information $I(X; Y)$ yields the expected PMI value over all possible instances of variable X and Y [37].

$$I(X; Y) = E_{(X, Y)}[PMI(x, y)] \quad (4)$$

The sign and absolute value of PMI enables the categorization of the variable interaction as mentioned above. Using PMI values, both statistically associated and opposite or complementary regions in the data can be identified. The regions that have opposite information will be the unique features in the data which is best represented by one particular variable among the selected variables. Similarly, the regions with strong association highlights joint multivariate features characterized by high co-occurrence.

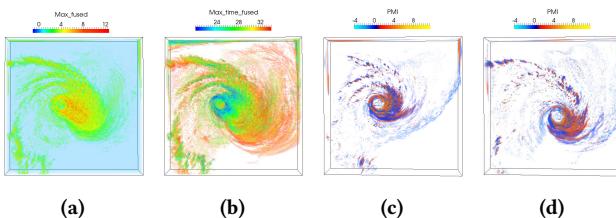


Figure 3: PMI fields of Cloud (CLO) and Precipitation (PRE) of Hurricane Isabel data set using time steps between 20-35. 3a time aggregated PMI field using max function, 3b time volume, 3c PMI field at T=20, and 3d PMI field at T=34.

3.3 Multivariate PMI Fields

Given a multivariate time-varying data set, every spatial point in the domain has several scalar values associated with it, one from each variable. Since PMI measures the information content for any value pair, we can use it to obtain the information content for any location. To facilitate this fine grained analysis by preserving the spatial context, we create a new scalar field, called *PMI field*. In PMI field, 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 formally expressed as: $\zeta : P \mapsto PMI(P)$, where P is a spatial location and $PMI(P)$ is the PMI value at P . The probabilities for the computation of PMI values are estimated from the histograms of the scalar values obtained from all the grid points.

Figure 2 shows one illustrative example of a PMI field where Z-velocity and velocity gradient magnitude field of the Solar Plume data set [29] are used to construct the PMI field. Figure 2c depicts the PMI field constructed using these two variables. It is evident from Figure 2c that, both Z-velocity and gradient magnitude field have strong statistical association 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 magnitude has unique activity that is missing in the Z-velocity. In the PMI field, this region is considered to contain complementary information which is unique to the gradient magnitude.

3.4 Time-varying PMI Fields

Next, we aggregate several PMI fields into a single scalar field using an aggregation function. The aim of this aggregation is to combine information from a set of time steps into a single scalar field for capturing time-varying patterns. For example, if a time-varying feature is identified as a joint activity and if the feature moves spatially over time, then at every time step, the feature can be located by focusing on the regions with higher joint activity. Since positive and high PMI values reflect joint activity, and relatively low co-activity regions have negative PMI values, we use *max* and *min* functions for the aggregation. If we use *max* as our aggregation criterion, then at every spatial location, PMI values for all the selected time steps are observed and the maximum value among them is selected to construct a *time aggregated PMI field*. Formally, for a spatial location P , the aggregation value is calculated as:

$$AggPMI(P) = \Psi(PMI_i(P)), \forall i = t_s, t_s + 1, \dots, t_e \quad (5)$$

where t_s and t_e represent starting and ending time steps, $PMI_i(P)$ is the value of PMI of point P at time step i , and $\Psi(\cdot)$ is the aggregation function. We also create another scalar field where at every grid point we put the time step number from which the PMI value is selected. We call it the *time volume* which presents the temporal trace of the feature.

In Figure 3, we demonstrate the usefulness of this idea with an example where Cloud (CLO) and Precipitation (PRE) variables from Hurricane Isabel data set are used. Given the range of time steps between 20 – 35, we construct the time aggregated PMI field using max aggregation function. Figure 3a shows the aggregated PMI field and 3b shows the associated time volume. From Figure 3b we

can visualize how the feature has moved by looking at the change of color in Figure 3b. Note that, in Figure 3b the color signifies time steps and it varies from blue to red as time increases.

On demand inspection of the feature at specific time steps.

To allow analysis of the identified feature at specific time steps, we incorporate a threshold based visualization. Initially users select a threshold for both high and low PMI values which highlight their regions of interest at the first selected time step. Then we apply this threshold to all the other time steps to extract the regions that show either strong or weak statistical association. For maintaining consistency, before the threshold is applied, all the PMI fields from the selected time range are normalized so that the PMI values are scaled consistently over the selected time window. Figure 3c and 3d show the results of the thresholding where snapshots of time steps 20 and 34 are displayed respectively. The *continuation* of the downward rotational movement of the cloud structures and precipitation bands are visible from these images.

3.5 Identification of Temporally Salient Scalar Value Combinations

Acknowledging the fact that the total number of value combinations can be significantly large, we present a refinement based strategy which aims at grouping value combinations with similar behavior. The proposed method exploits both SMI and the PMI to devise a top-down approach. In our work, SMI measure *predictability* is used which was introduced in [16]. Formally, given a value x of the variable X , its predictability is defined as:

$$SMI(x; Y) = H(Y) - H(Y|x) \quad (6)$$

where Y is the other selected variable, $H(Y)$ is the entropy of Y , and $H(Y|x)$ is the entropy of Y given observation x . $SMI(x; Y)$ is called predictability because based on the value of $SMI(x; Y)$, it can be inferred, how well the observation x can predict the behavior of Y . Higher and positive values of $SMI(x; Y)$ reflects higher predictability, whereas, negative values of $SMI(x; Y)$ signifies increased uncertainty about Y after x is observed. So, using $SMI(x; Y)$, the scalar values of X can be divided into two groups: (1) scalars with positive $SMI(x; Y)$ i.e. the predictable scalars, and (2) scalars with negative $SMI(x; Y)$ containing the uncertain scalars of variable X .

After this initial grouping using PMI, the value combinations are further classified into two groups: (1) combinations with positive PMI values, and (2) combinations that have negative PMI values. Here we are focusing on the combined and complementary features of multi-variables, so, we do not consider the combinations with PMI value 0. Since our goal is to quantify the temporal trends of the value combinations, we observe the PMI value of each value combination for all the selected time steps and group them separately if the value combinations have always positive or always negative values throughout the specified time range.

Based on the above discussion, the value combinations of variable X and Y are grouped into 4 distinct classes:

- (1) $\{(x_i, y_j) \mid \forall i, j \text{ where } PMI(x_i, y_j) > 0 \text{ & } SMI(x_i; Y) < 0\}$
- (2) $\{(x_i, y_j) \mid \forall i, j \text{ where } PMI(x_i, y_j) < 0 \text{ & } SMI(x_i; Y) < 0\}$
- (3) $\{(x_i, y_j) \mid \forall i, j \text{ where } PMI(x_i, y_j) > 0 \text{ & } SMI(x_i; Y) > 0\}$
- (4) $\{(x_i, y_j) \mid \forall i, j \text{ where } PMI(x_i, y_j) < 0 \text{ & } SMI(x_i; Y) > 0\}$

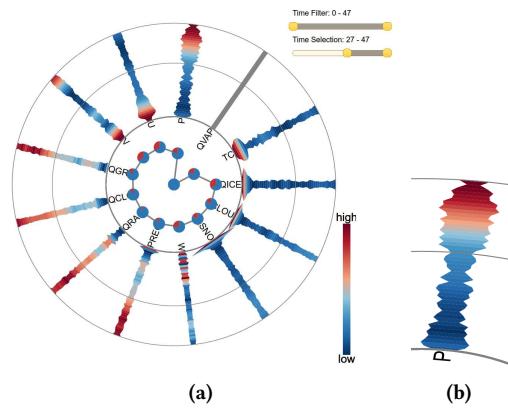


Figure 4: 4a Variable selection for Isabel data set when Qvapor (QVA) variable is selected. 4b Zoomed in Pressure axis.

Given any class, for each value combination in it, we construct a time series using its PMI values and its temporal saliency is measured by the variation of the PMI values. Formally, the variation for a value combination is measured as:

$$Var(TS_i) = \sqrt{\sum_{j=t1}^{t2-1} |PMI_{i,j} - PMI_{i,j+1}|^2} \quad (7)$$

where, TS_i is the time series of i th value combination. $PMI_{i,j}$ is the PMI value of series TS_i at j th time step and the selected time step range is $t1 - t2$. A high variation value indicates that the value combination has weaker association among them and their occurrences are not consistent temporally. In contrast, the time series with low variation are likely to reveal a region that has higher statistical association. With this classification strategy, the complexity in the relationships among the large number of value combinations is reduced significantly.

4 INTERACTIVE WORKFLOW

4.1 Identifying Temporally Related Variables

Interface Design. As domain scientists often have some prior knowledge about which variable is more interesting to initiate the exploration process, we are interested in seeing the relationships between the reference (initially picked by the user) variable and the other variables. Figure 4a shows our variable selection interface where Hurricane Isabel data set is used for demonstration. The reference variable (QVapor) is selected by the user and placed in the center, while the other variables are placed such that one is closer to the center if it is strongly related to the reference variable. The layout has several benefits: (1) it emphasizes the stronger relationships as the central part of the visualization, which is more active in the user's visual field; (2) it preserves the user's mental map. The time-varying one-to-many relationships form a time series for each variable pair. We visualize these in multiple time plots, which are aligned with variables in the radar plot. Such a design is usually more efficient than shared space techniques [24] and results in an overview + detail visualization.

Visual Encoding. While the location of a variable in the radar plot shows the strength of the corresponding relationship, each

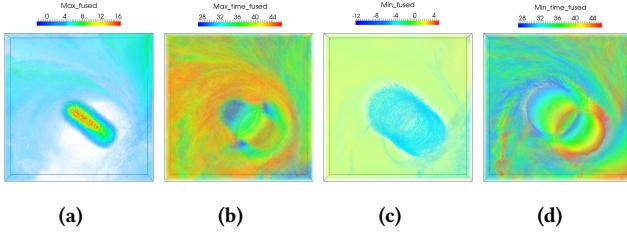
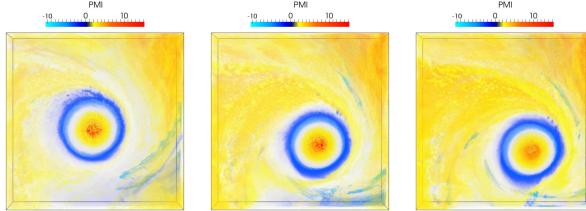


Figure 5: Aggregated PMI fields of P and QVA of Isabel data between time steps 27-47. 5a Aggregated PMI field using max function and 5b its Time volume; 5c Aggregated PMI field using min function, and 5d its Time volume.



(a) PMI field at T=30. (b) PMI field at T=40. (c) PMI field at T=45.

Figure 6: Time-varying PMI fields of Pressure (P) and Qvapor (QVA) of Hurricane Isabel data set between time steps 27-47.

variable itself is visualized as a pie chart to depict the percentage of the positive (blue) and negative (red) PMI values for each variable pair. The segments in a time plot in Figure 4a is colored by the strength of the relationship, mutual information in this case, while the width of each segment is modulated by the temporal gradient of the relationship's strength. In Figure 4a we also show the selection of time steps where the gray circles show the time range selected.

Guidelines for Variable Selection. With our design, once the user has selected a reference variable (QVA in this case), they can select another variable and an appropriate time window based on how the information is shared between it and the reference variable:

(1) *Varying information overlap*: a variable showing a rapid change of colors with a wide time axis. (Selected time steps of P in Figure 4a, highlighted by the two gray circular rings), (2) *Constantly high information overlap*: a variable that has mostly red regions for a sequence of time steps. (Later time steps of PRE, QRA, QCL etc. in Figure 4a), (3) *Low information overlap with high variation*: a variable containing blue regions with a relatively wide time axis (later time steps of U in Figure 4a), (4) *Constantly low information overlap*: a variable with mostly blue regions with a narrow time axis. (Majority parts of the time axis of QSN, QIC, CLO etc. in Figure 4a).

4.2 Analysis using PMI Fields

After a pair of variables are identified, we construct PMI fields for each selected time step using those variables and aggregate them using both max and min functions. This PMI field based visualizations allow scientists to directly interact with the information in spatial domain. After the informative regions and temporal trends are analyzed using the aggregated volumes, we allow users to interact with the specific value combinations so that the scalar values creating such joint temporal features can be specifically identified.

4.3 Identification of Temporally Salient Scalar Value Combinations.

In section 3.5 we have described how the value combinations can be grouped based on their informativeness. Next, we create a histogram of all the value combinations in each group using their PMI variation values. We allow brushing in the *variation histogram* so that users can select bins with high or low information variation. A parallel coordinates plot (PCP) is attached with the histogram, so that the selected value combinations can be easily visualized. Finally, users can brush the PCP to select specific value combinations and while visualizing isosurfaces of those selected value combinations, their PMI time series are also displayed. Users can change the time steps to inspect the temporal changes of such isosurfaces and also observe how their PMI values change. Figure 7a shows a variation histogram where the analysis is done using QVA and P variables of the Isabel data and the value combinations from the group 1 (described in Section 3.5) is chosen. The light yellow highlighted region shows the user selected bins and in Figure 7b and Figure 7c the corresponding PCP and the PMI time series are shown.

5 CASE STUDIES

The experiments were done 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. The visualizations were generated using D3 library [4] and ParaView [1].

5.1 Hurricane Isabel Data Set

Hurricane Isabel data is a multivariate time-varying data consisting of 13 scalar fields. 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 is $250 \times 250 \times 50$, and there are total 48 time steps. From Figure 4a, we see that Qvapor (QVA) is selected as the reference variable for this study. Given QVA, following our variable selection interface, Pressure (P) is selected as the second variable since it shows varying information overlap between time steps 27 – 47.

Figure 5a and 5b show the aggregated PMI field and its time volume when max is used for aggregation. We see that the Hurricane eye has strong co-activity. Similarly, Figure 5c and 5d depict the aggregated PMI field and its time volume when min is used for aggregation. The eye wall of the storm is visible as a complementary feature whose temporal trend is observed from the associated time volume in Figure 5d. To facilitate exploration of identified regions at specific time steps, Figure 6 presents 3 selected PMI fields where we can visualize the combined and complementary regions at three individual time steps. Figure 6a, 6b, and 6c depict the temporal changes of the regions of interest at time steps 30, 40, and 45 respectively where reddish yellow regions signify regions with stronger co-activity and the light blue region which shows the eye wall of the storm is identified as the complementary informative region.

Figure 7a shows the variation histogram of QVA and P representing the value combinations when the group with all positive PMI and negative SMI values of QVA are considered. Brushing some low variation bins (yellow highlighted region) yields the PCP in Figure 7b, from which a specific value combination is selected. In Figure 7c

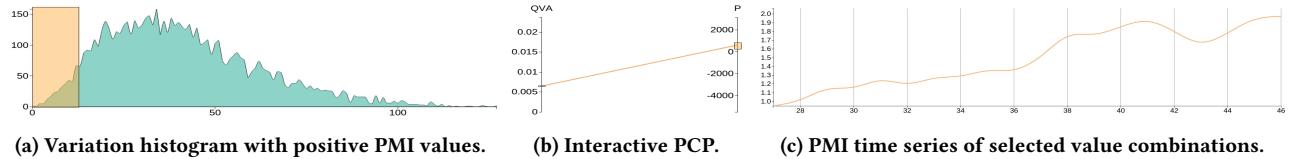


Figure 7: Selection of salient scalar value combinations of Pressure (P) and Qvapor (QVA) variables of Hurricane Isabel data set between time steps 27- 47. Selected value combinations reflect combined activity of the selected variables.

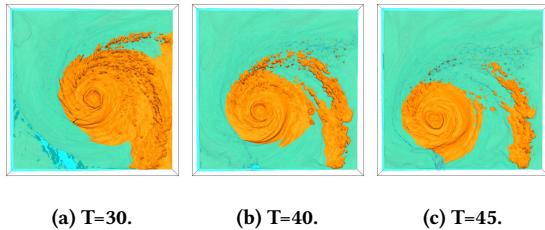


Figure 8: Temporally salient isosurface visualization of Pressure (P) = 617.04 (blue) and Qvapor (QVA) = 0.00647598 (orange) of Hurricane Isabel data set.

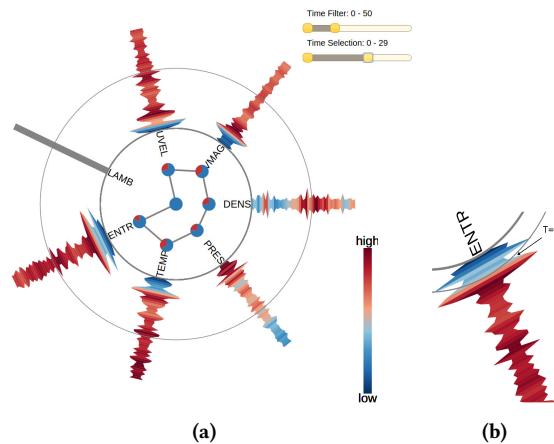


Figure 9: 9a Variable selection for Turbine data set when λ_2 (LAMB) variable is selected, 9b Zoomed in ENTR axis.

the PMI series is depicted. The trend shows that the magnitude of the PMI values increase over time which is reflected in the isosurfaces in Figure 8. In Figure 8a, 8b, and 8c isosurfaces of $P = 617.04$ (blue) and $QVA = 0.00647598$ (orange) are shown for time steps 30, 40 and, 45 respectively. From these, we observe that, as the PMI values increase with time, the degree of the statistical association between the value pair also strengthens which is revealed by their increased spatial overlap over time.

5.2 Expert Evaluation Using Turbine Data Set

The Turbine data set is generated by a flow simulation TURBO, where the compressor is undergoing rotating stall [10]. It is a multi-block data consisting of 36 blade passages. The resolution of each passage/block is $151 \times 71 \times 56$ and it has five variables: Density (DENS), Velocity momentum in x/y/z direction (MOM), and Total

energy (TOTENR). These five fundamental variables are used to derive other variables. The simulation was run for 192 time steps.

In a stable state, the tip region of each blade develops a vortex known as *tip vortex*. Detailed study of this tip vortex during stall inception and identification of the associated variable interactions are essential for the scientist. In order to facilitate domain experts with a better understanding on the behavior of the tip vortex, we have computed vortex criterion λ_2 (LAMB). In Figure 9a we show our variable selection interface when λ_2 (LAMB) is selected as the reference variable by the expert. It is observed that variable entropy (ENTR) displays the highest information variation during the initial time steps and at time step 6 the shared information between these two variables become high which can be seen by the change of color of ENTR axis from initial blue to red in Figure 9b. To investigate this, ENTR becomes a suitable second variable for analysis and time steps between 1 – 29 is selected for detailed investigation. For highlighting this region, in Figure 9a we only show time steps between 0 – 50 using the time step filter slider.

Figure 10a and 10b show the max and min time volumes of the selected variables. In Figure 10a, we observe that the regions away from the tip show stronger statistical association during the later time steps, hence more red regions are located away from the tip region. The min time volume displayed in Figure 10b, show the opposite trend. Here the tip regions show more opposite activity during later time steps identified by the reddish yellow regions. Figure 10c, 10d, and 10e present the PMI fields of three selected time steps, 5, 15 and 25 respectively. We can see that as time progresses, more regions along the tip with a complementary activity appear. In this case study, we see that the complementary region of interest grows over time which signifies the increase in opposite information between these two selected variables.

Figure 11a depicts the variation histogram with all positive PMI and negative SMI values for LAMB. Figure 11b shows when a specific value combination is picked by brushing the histogram first, and then filtering from the PCP. Figure 12 presents simultaneous isosurfaces of the picked value combination ($LAMB = -5784.25$ (orange) and $ENTR = 1.16768$ (blue)) for time steps 1, 6, 7, and 15 respectively. Note that, the isosurface of LAMB we visualize here shows vortices (tip vortex in this case). We find that from time step 6 on-wards, the LAMB isosurface becomes fragmented which represents the *breakdown of the tip vortices*. This is an indication of the stall inception. In the variable selection interface (Figure 9a), we observe that time steps 5 – 7 cause sudden change of MI between these two variables when the tip vortices breakdown. Also, after the tip vortices break down, the degree of co-occurrence of this value combination is reduced which can be seen from Figure 11c, where the magnitude of PMI value gradually decreases.

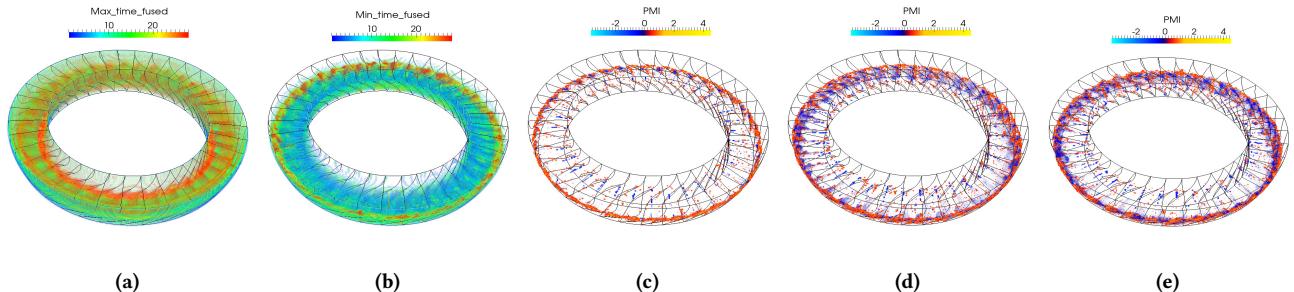


Figure 10: Visualization of PMI fields of Turbine data when variables $\lambda 2$ (LAMB) and Entropy (ENTR) are selected between time steps 1 - 29. 10a Time volume with max function, 10b Time volume with min function, 10c PMI field at T=5, 10d PMI field at T=15, and 10e PMI field at T=25



(a) Variation histogram with all positive PMI values. (b) Interactive PCP. (c) PMI time series of selected value combinations.

Figure 11: Selection of salient scalar value combinations of λ_2 (LAMB) and Entropy (ENTR) variables of Turbine data set between time steps 1-29. Selected value combinations reflect combined activity of the selected variables.

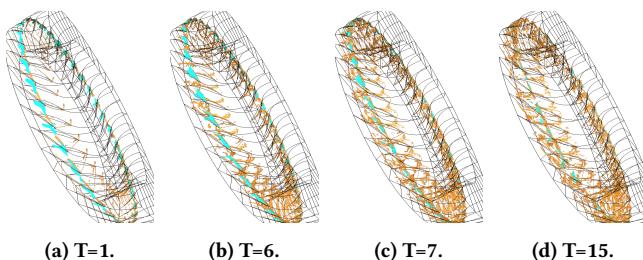


Figure 12: Temporally salient isosurface visualization of variables λ_2 (LAMB) = -5784.25 (orange) and Entropy (ENTR) = 1.16768 (blue) of Turbine data set.

Expert feedback. The domain expert who evaluated our system has more than 25 years of experience in the computational fluid dynamics simulations and is one of the developers of the TURBO simulation code which we used for this study. The feedback were collected through several meetings with the expert during which we explained our system to the expert in detail. According to the expert, identification of the breakdown of tip vortices was helpful for a detailed study of stall. Comparing our tools with the existing tools such as *FieldView*, the expert pointed that, our tool was able to provide the information-theoretic guidance when little knowledge was available. Also, the expert confirmed that the PMI field based identification of salient regions allowed to locate interesting regions where the variables show strong or weak statistical association. The expert also mentioned that the time series based analysis of value combinations provided a new tool to perform detailed multivariate temporal study on the scalar values. This helped in identifying salient value combinations which were either strongly correlated or showed weak association.

Performance evaluation. Here we report the timings of the different components of our system. The average computation time

Table 1: Timings for computing PMI fields and aggregation.

Data Set	Avg. time per PMI field creation (secs.)	Avg. time for aggregation (secs.)
Hurricane Isabel	0.461	0.210
Turbine	3.817	0.325

for the PMI field creation and their aggregation is shown in Table 1. The computation of all pair MI for our variable selection interface was done as a pre-processing step. Note that, since creation of PMI fields are independent for each time step, they can be generated in parallel which will further improve the performance.

6 CONCLUSION AND FUTURE WORKS

we have presented an information-theoretic framework for analysis of multivariate time-varying data sets using their specific value combinations. We use PMI to measure the information content of all value combinations and further construct PMI fields allowing us to analyze multivariate relationships. To capture temporal evolution of the features, we aggregate several PMI fields into a single field. For identifying temporally salient value pairs, we measure the temporal PMI variation of each value combination. Interactive selection of salient value pairs and their isosurfaces provide detailed information about the temporal interaction of the value pairs. In the future, we would like to extend our framework for different data types such as vector fields, and ensemble data for feature analysis.

REFERENCES

- [1] Utkarsh Ayachit. 2015. *The ParaView Guide: A Parallel Visualization Application* (4.3 ed.). Kitware Inc. <http://www.paraview.org/paraview-guide/> ISBN 978-1-930934-30-6.
 - [2] C.L. Bajaj, V. Pascucci, and D.R. Schikore. 1997. The contour spectrum. In *Visualization '97, Proceedings*. 167–173.
 - [3] Ayan Biswas, Soumya Dutta, Han-Wei Shen, and Jonathan Woodring. 2013. An Information-Aware Framework for Exploring Multivariate Data Sets. *IEEE*

- Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2683–2692. <https://doi.org/10.1109/TVCG.2013.133>
- [4] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D3 Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (Dec. 2011), 2301–2309. <https://doi.org/10.1109/TVCG.2011.185>
- [5] R. Bramon, I. Boada, A. Bardera, J. Rodriguez, M. Feixas, J. Puig, and M. Sbert. 2012. Multimodal Data Fusion Based on Mutual Information. *Visualization and Computer Graphics, IEEE Transactions on* 18, 9 (sept. 2012), 1574 –1587. <https://doi.org/10.1109/TVCG.2011.280>
- [6] Roger Bramon, Marc Ruiz, Anton Bardera, Imma Boada, Miquel Feixas, and Mateu Sbert. 2013. An Information-Theoretic Observation Channel for Volume Visualization. *Comput. Graph. Forum* 32, 3 (2013), 411–420.
- [7] R. Bramon, M. Ruiz, A. Bardera, I. Boada, M. Feixas, and M. Sbert. 2013. Information Theory-Based Automatic Multimodal Transfer Function Design. *Biomedical and Health Informatics, IEEE Journal of* 17, 4 (July 2013), 870–880. <https://doi.org/10.1109/JBHI.2013.2263227>
- [8] Stefan Bruckner and Torsten Möller. 2010. Isosurface Similarity Maps. *Computer Graphics Forum* 29 (2010), 773–782.
- [9] Hamish Carr, Jack Snoeyink, and Ulrike Axen. 2000. Computing Contour Trees in All Dimensions. In *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '00)*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 918–926. <http://dl.acm.org/citation.cfm?id=338219.338659>
- [10] Jen-Ping Chen, Michael D. Hathaway, and Gregory P. Herrick. 2008. Prestall Behavior of a Transonic Axial Compressor Stage via Time-Accurate Numerical Simulation. *Journal of Turbomachinery* 130, 4 (2008), 041014. <https://doi.org/10.1115/1.2812968>
- [11] Min Chen, Miquel Feixas, Ivan Viola, Anton Bardera, Han-Wei Shen, and Mateu Sbert. August 25, 2016. *Information Theory Tools for Visualization*. A K Peters/CRC Press.
- [12] M. Chen and H. Jänicke. 2010. An Information-theoretic Framework for Visualization. *Visualization and Computer Graphics, IEEE Transactions on* 16, 6 (2010), 1206–1215. <https://doi.org/10.1109/TVCG.2010.132>
- [13] Kenneth Ward Church and Patrick Hanks. 1989. Word association norms, mutual information, and lexicography. In *Proceedings of the 27th annual meeting on Association for Computational Linguistics (ACL '89)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 76–83. <https://doi.org/10.3115/981623.981633>
- [14] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. 1995. Automated multi-modality image registration based on information theory. In *Information Processing in Medical Imaging*.
- [15] Thomas M. Cover and Joy A. Thomas. 2006. *Elements of Information Theory 2nd Edition* (Wiley Series in Telecommunications and Signal Processing). Wiley-Interscience.
- [16] M. R. Deweese and M. Meister. 1999. How to measure the information gained from one symbol. *Network: Computation in Neural Systems* 4 (nov 1999), 325–340.
- [17] B. Duffy, H. Carr, and T. Moller. 2013. Integrating Isosurface Statistics and Histograms. *Visualization and Computer Graphics, IEEE Transactions on* 19, 2 (Feb 2013), 263–277.
- [18] Miquel Feixas, Esteve Del Acebo, Philippe Bekaert, and Mateu Sbert. 1999. An Information Theory Framework for the Analysis of Scene Complexity. (1999).
- [19] Miquel Feixas, Mateu Sbert, and Francisco González. 2006. A unified information-theoretic framework for viewpoint selection and mesh saliency. (2006).
- [20] Martin Haidacher, Stefan Bruckner, and Meister Eduard Gröller. 2011. Volume Analysis Using Multimodal Surface Similarity. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (Oct. 2011), 1969–1978.
- [21] Martin Haidacher, Stefan Bruckner, Armin Kanitsar, and Meister Eduard Gröller. 2008. Information-based Transfer Functions for Multimodal Visualization. In *VCBM*, W.J. Niessen C.P. Botha, G. Kindlmann and B. Preim (Eds.). Eurographics Association, 101–108.
- [22] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. 2001. Medical image registration. *Physics in Medicine and Biology* 46, 3 (2001), R1.
- [23] William M. Wells III, Paul Viola, Hideki Atsumi, Shin Nakajima, and Ron Kikin. 1996. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis* 1, 1 (1996), 35 – 51. [https://doi.org/10.1016/S1361-8415\(01\)80004-9](https://doi.org/10.1016/S1361-8415(01)80004-9)
- [24] Waqas Javed, Bryan McDonnel, and Niklas Elmquist. 2010. Graphical Perception of Multiple Time Series. *IEEE Transactions on Visualization and Computer Graphics* (2010).
- [25] M. Khouri and R. Wenger. 2010. On the Fractal Dimension of Isosurfaces. *Visualization and Computer Graphics, IEEE Transactions on* 16, 6 (Nov 2010), 1198–1205.
- [26] X. Liu and H. W. Shen. 2016. Association Analysis for Visual Exploration of Multivariate Scientific Data Sets. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (Jan 2016), 955–964. <https://doi.org/10.1109/TVCG.2015.2467431>
- [27] Jun Ma, Chaoli Wang, and Ching-Kuang Shene. 2013. Coherent view-dependent streamline selection for importance-driven flow visualization. (2013), 865407–865407-15 pages.
- [28] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. 1997. Multi-modality image registration by maximization of mutual information. *Medical Imaging, IEEE Transactions on* 16, 2 (April 1997), 187–198. <https://doi.org/10.1109/42.563664>
- [29] High Altitude Observatory. 1998. Compressible plume dynamics and stability. 369 (1998), 125–149.
- [30] Josien P. W. Pluim, J. B. Antoine Maintz, and Max A. Viergever. 2003. Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging* (2003), 986–1004.
- [31] J. Rigau, M. Feixas, and M. Sbert. 2005. Shape complexity based on mutual information. In *Shape Modeling and Applications, 2005 International Conference*. 355–360. <https://doi.org/10.1109/SMI.2005.42>
- [32] J. Rigau, M. Feixas, and M. Sbert. 2008. Informational Aesthetics Measures. *Computer Graphics and Applications, IEEE* 28, 2 (2008), 24–34.
- [33] Mateu Sbert, Miquel Feixas, Jaume Rigau, Miguel Chover, and Ivan Viola. 2009. *Information Theory Tools for Computer Graphics*. Morgan and Claypool Publishers Colorado. <http://dx.doi.org/10.2200/S00208ED1V01Y200909CGR012>
- [34] Carlos E. Scheidegger, John M. Schreiner, Brian Duffy, Hamish Carr, and Claudio T. Silva. 2008. Revisiting Histograms and Isosurface Statistics. *IEEE Transactions on Visualization and Computer Graphics* 14, 6 (2008), 1659–1666. <https://doi.org/10.1109/TVCG.2008.160>
- [35] C. E. Shannon. 2001. A Mathematical Theory of Communication. *SIGMOBILE Mob. Comput. Commun. Rev.* 5, 1 (Jan. 2001), 3–55. <https://doi.org/10.1145/584093>
- [36] Jun Tao, Jun Ma, Chaoli Wang, and Ching-Kuang Shene. 2013. A Unified Approach to Streamline Selection and Viewpoint Selection for 3D Flow Visualization. *IEEE Transactions on Visualization and Computer Graphics* 19, 3 (March 2013), 393–406.
- [37] Tim Van de Cruys. 2011. Two multivariate generalizations of pointwise mutual information. In *Proceedings of the Workshop on Distributional Semantics and Compositionality (DiSCo '11)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 16–20. <http://dl.acm.org/citation.cfm?id=2043121.2043124>
- [38] S. Verdú. 1998. Fifty years of Shannon theory. *Information Theory, IEEE Transactions on* 44, 6 (Oct 1998), 2057–2078. <https://doi.org/10.1109/18.720531>
- [39] I. Viola, M. Feixas, M. Sbert, and M.E. Groller. 2006. Importance-Driven Focus of Attention. *Visualization and Computer Graphics, IEEE Transactions on* 12, 5 (2006), 933–940. <https://doi.org/10.1109/TVCG.2006.152>
- [40] Chaoli Wang and Han-Wei Shen. 2011. Information Theory in Scientific Visualization. *Entropy* 13, 1 (2011), 254–273. <https://doi.org/10.3390/e13010254>
- [41] Tzu-Hsuan Wei, Teng-Yok Lee, and Han-Wei Shen. 2013. Evaluating Isosurfaces with Level-set-based Information Maps. In *Proceedings of the 15th Eurographics Conference on Visualization (EuroVis '13)*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 1–10. <https://doi.org/10.1111/cgf.12087>