# StreamVis^ND: Visualizing Relationships in Streaming Multivariate Data

Shenghui Cheng[1], Yue Wang[2], Dan Zhang[1], Zhifang Jiang[2] and Klaus Mueller[1]

[1]Visual Analytics and Imaging Lab, Computer Science Department, Stony Brook University and SUNY Korea

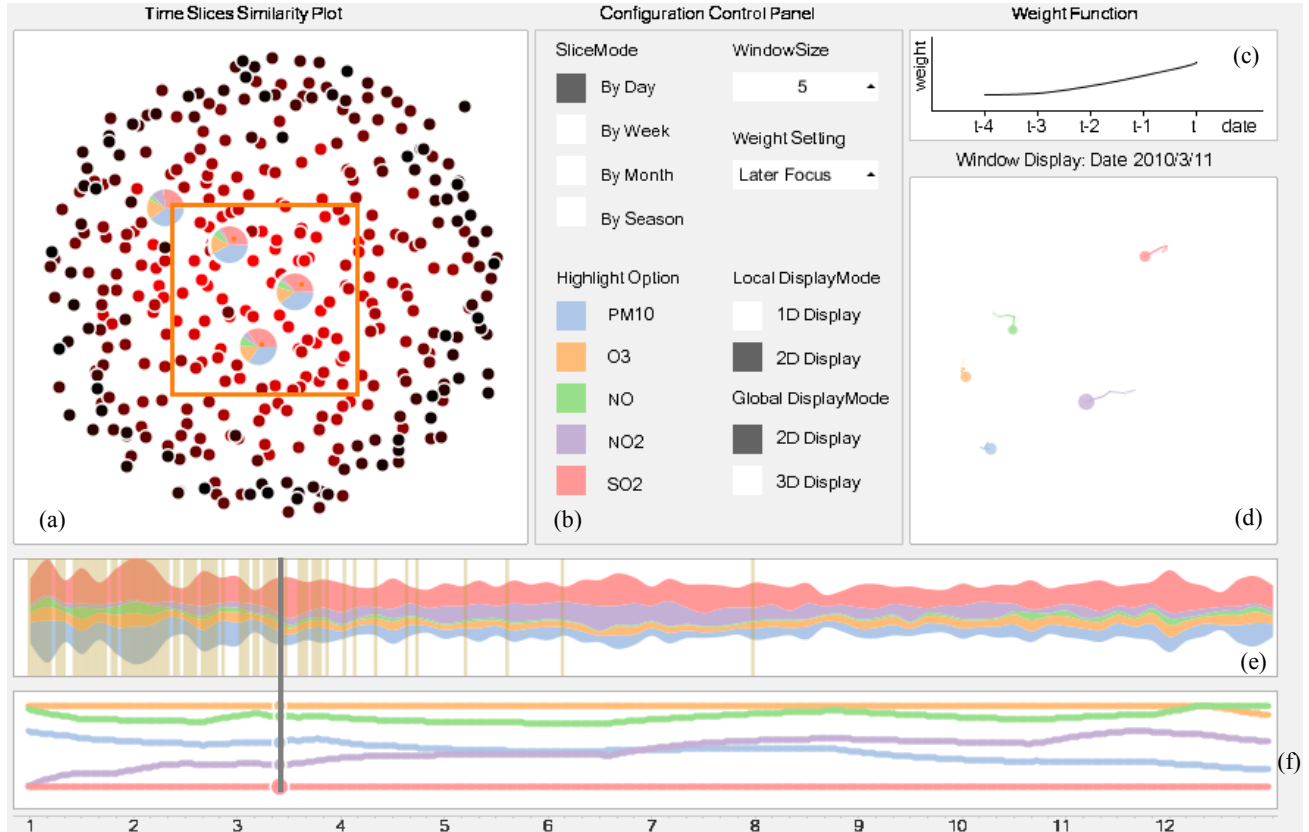[2]Department of Computer Science, Shandong University

Fig.1 Visualization of pollution data with StreamVis^ND: (a) The *Time Slice Similarity Plot* which shows the similarity of time slices over time (mapped to brightness), (b) the configuration control panel with pollutant color map; (c) the weight function, (d) the local change display of pollutant relations; (e) the Themeriver display of pollutant concentrations over time; and (f) the relations display which plots the relations of pollutant concentrations over time. Closer lines indicate higher temporal correlation in terms of the pollutant distribution.

**ABSTRACT**

In streaming acquisitions the data changes over time. ThemeRiver and line charts are common methods to display data over time. However, these methods can only show the values of the variables (or attributes) but not the relationships among them over time. We propose a framework we call StreamVis^ND that can display these types of streaming data relations. It first slices the data stream into different time slices, then it visualizes each slice with a sequence of multivariate 2D data layouts, and finally it flattens this series of displays into a parallel coordinate type display. Our framework is fully interactive and lends itself well to real-time displays.

Email: [1]{shecheng, zhang64, mueller}@cs.stonybrook.edu
[2] yuewang@mail.sdu.edu.cn
[2] zfjiang@sdu.edu.cn

## 1 INTRODUCTION

Numerous techniques have been developed to visualize streaming data. One prominent technique is ThemeRiver [1] which represents the values of the attributes (called *themes*) as vertical bars and joins them horizontally over time. This yields a display of layers – one layer per attribute – with time-varying cross-sections. While ThemeRiver presents streaming data continuously, it is difficult for users to assess the similarities of different time slices and the changing relations of the attributes over time. Let us assume a city A in 2013 with daily pollution data of PM10, O3, NO, NO2 and SO2. Environment experts would like to know which two days are similar, which days are outliers, what are the relations of the pollutants over time, etc. Multivariate visual analytics, such as PCA [2] and MDS [3], can address these issues, but so far there is no such technique that does this for time-varying data effectively. We have aimed for a display that can show these similarities but still preserves the time-varying context to some extent. This display is the subject of this paper.

## 2 INTERFACE OVERVIEW

In order to visualize streaming data in a more relationship-centric manner, we have designed an interface (see Fig 1) called StreamVis[ND]. It consists of six parts – time-slice similarity plot, configuration control panel, weight function, local change display, ThemeRiver plot, and relation display. The weight function and window display are used to set the window size and display the local relation changes. The configuration display is used to set the various parameters.

## 3 THE TIME SLICE SIMILARITY PLOT

This plot visualizes the similarity of the time slices in terms of the attribute values. As such, each time slice gives rise to an $A$-dimensional vector of attribute values ($A$ is the number of attributes) which we lay out via MDS, using correlation as the similarity metric. We further code the time stamps of the slices as brightness. Fig. 1a shows the resulting *Time Slice Similarity Plot* in a "by-day" slice mode -- we also provide "by week", "by month", and "by season". We observe that early days are similar and map to the display center, while later days are more dissimilar and map to the periphery.

We provide several interactions. Users can (1) select a time window and so restrict the number of points shown, (2) click on a point and see the pollutant's proportions as a pie chart directly inserted into the display, and (3) draw a rectangle and visualize the corresponding slices in the Themeriver display (Fig. 1e).

## 4 THE RELATION DISPLAY

This display shows the relation of the attributes over time. Here we subdivide the time series into equal-sized time intervals of length $T$, where $T$ is the number of discrete time steps of the slice. Then, for each slice $S_t$ and attribute $A_j$ we construct a slice-attribute vector $SA_{it}$ of length $T$ which holds the $T$ values $A_i$ has within the given $S_t$, ordered by time stamp. We again provide different *slice modes:* by-day, by-week, by-month, and by-season.

To visualize the attribute relations we could create a *Time Slice Similarity Plot* for each of the $I$ time slices and compare them in a small multiples display. However, the disjoint nature of such a display makes it difficult to visually track individual points over time and it will also make it difficult to track the relations of specific pairs of points, both of which is our goal. We thus designed a different method for combining these multiple MDS displays that makes these types of comparison tasks easier.

The method we propose arranges the MDS plots $S_i$ into a horizontal stack ordered by time and connects points $A_{ij}$ in time slice $S_i$ with the corresponding points in the adjacent time slices $S_{i-1}$ and $S_{i+1}$ by straight lines. This gives rise to a 3D display (see sketch in Fig. 2) where the changing relations across time slices visualize as changes in the pairwise line configurations.

However, since conventional MDS randomizes the initial coordinates of the points and only preserves the relative (but not absolute) locations of the points the final layout van vary signifi-
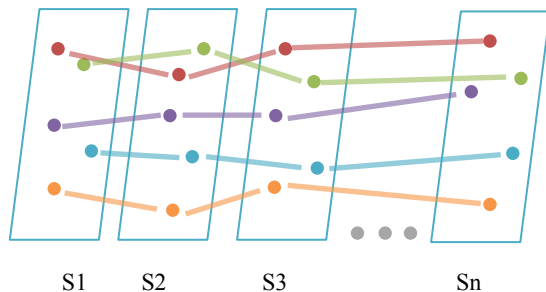


S1   S2   S3   Sn

Fig.2. The Relations Display.

cantly across the layout. Hence, the paths created between adjacent slices could be incoherent. We fix this by setting the initial coordinates of the points in a slice to the layout coordinates of the previous slice. In this way, the inter-slice paths can show the relation changes quite well. A remaining problem is that the paths are in 3D which suffers from occlusion problems. So our final step is to map these paths into 2D via another MDS step. In this way the user can appreciate any changes easily.

Fig. 1f shows this final *Relation Display* for the pollution dataset where the labels below the axes show the month. We can easily observe that PM10 and NO2 have a rather similar time-behaviour from the middle of March to the middle of June. We can also find that SO2 does not have a close temporal correlation with NO and O3.

## 5 THE LOCAL CHANGE DISPLAY

The "Relation Display" just presented focuses on the overall relation changes. However, for the streaming analysis the local changes can also be of interest. For this we need to focus on the local changes of different pollutants in a certain time slice. We developed a special display to show these changes.

### 5.1 Building the Weight Function

The local changes are typically related to close days (but this can be generalized). The window size selector (Fig. 1b) allows users to set the period they would like to monitor. A weight function allows users to set a preference for the days. We provide three types of weight functions: equal weights, previous focus, and later focus. "Equal weights" gives equal weights of the days in the windows. "Previous focus" gives higher weight for the previous days, while "later focus" emphasises the later days. Fig.1c shows the "later focus" weight function with a window size of 5.

### 5.2 Dynamic Display

The local (transient) changes can also be visualized via MDS, now by ways of a dynamic layout where the local changes of the points are visualized with streak lines. The plot of Fig. 1d shows a snapshot around March 10th where we observe that PM10, NO, O3 and SO2 move closer to each other.

## 6 CONCLUSION

We have presented a solution that can visualize changing inter-attribute relations within time varying multivariate data. Users specify the desired time slice granularity and then the system produces a graph that shows these temporally changing   similarity relationships across adjacent time slices using MDS. Future work will focus on user studies to refine the framework.

### REFERENCES

[1] S. Havre, E. Hetzler, L. Nowell: "ThemeRiver: Visualizing Theme Changes over Time," *Proc. IEEE InfoVis*, pp. 115-123, 2000.

[2] I. Jolliffe, "Principal Component Analysis," *Series: Springer Series in Statistics*, 2nd ed., Springer, NY, 2002, XXIX, 487 pp.28 illus. ISBN 978-0-387-95442-4.

[3] J. Kruskal. M. Wish, *Multidimensional Scaling*. Sage Publications, 1977.