# 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

## Introduction

In streaming acquisitions the data change 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 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.

## Interface Overview

We provide three types of weight functions: equal weights, previous focus, and later focus. The window size selector allows users to set the period they would like to monitor. Here we show the "later focus" weight function with a window size of 5.

This plot visualizes the similarity of the time slices in terms of the attribute values.

We further code the time stamps of the slices as brightness.

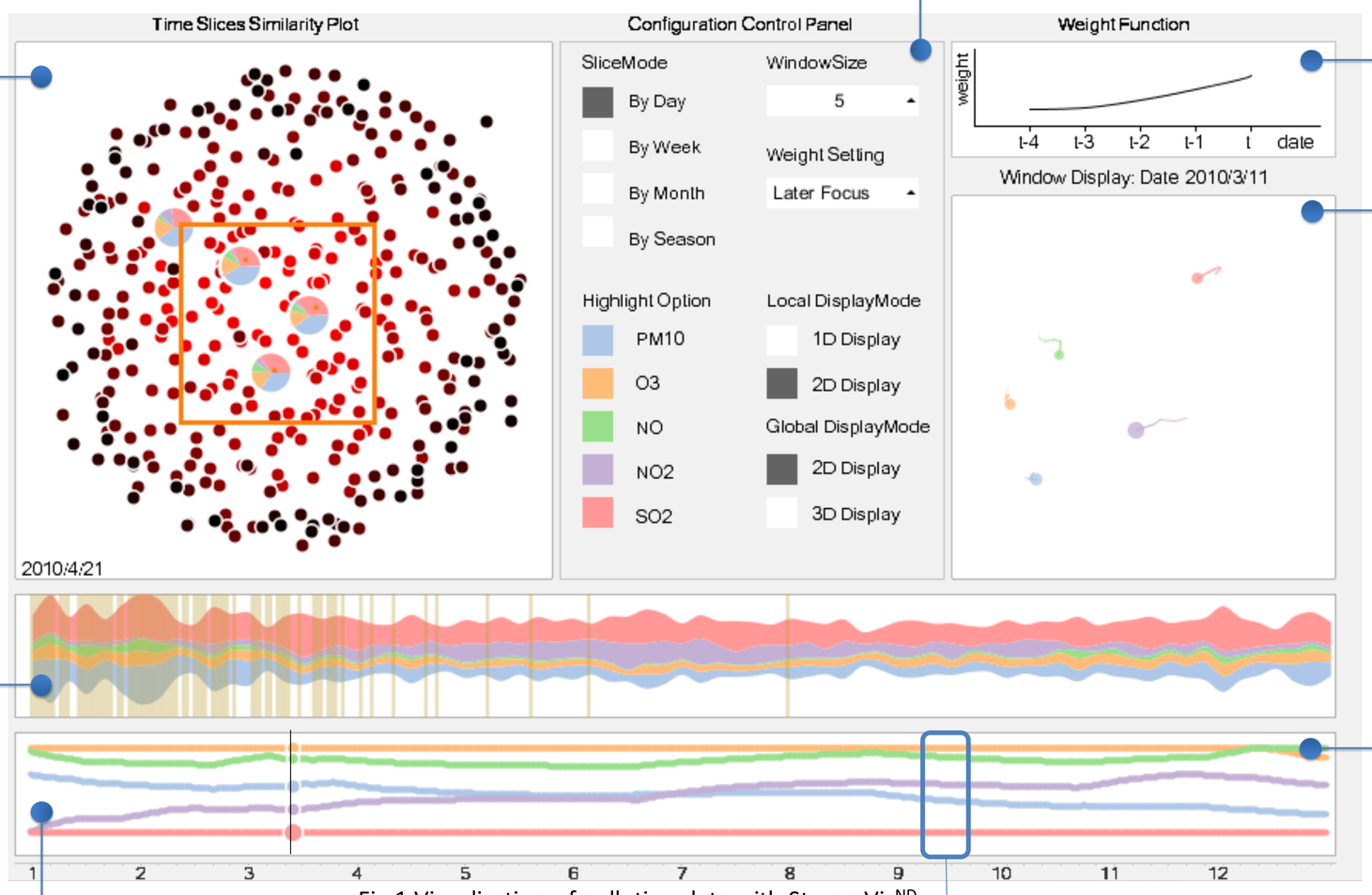


Fig.1 Visualization of pollution data with StreamVis$^{ND}$

A weight function allows users to set a preference for these days.

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.

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 can vary significantly 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.

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,
(3) draw a rectangle and visualize the corresponding slices in the Themeriver 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.

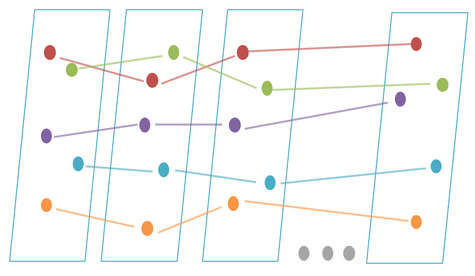

Fig.2. The Relations Display.

## Case Study

Let's assume a city A in 2013 with daily pollution data of PM10, $O_3$, NO, $NO_2$ and $SO_2$. Environmental experts would like to know which two days are similar, which days are outliers, what are the relations of the pollutants over time, etc. Through the Time Slice Similarity Plot, they can observe that early days are similar and map to the display center, while later days are more dissimilar and map to the periphery. In the Relation Display they can easily observe that $PM_{10}$ and $NO_2$ have a rather similar time-behaviour from the middle of March to the middle of June. It can also be seen that $SO_2$ does not have a close temporal correlation with NO and $O_3$. The plot of Window Display shows a snapshot around April 11$^{th}$ where it can be observed that $PM_{10}$, NO, $O_3$ and $SO_2$ move closer to each other but $NO_2$ moves further apart.

This work was supported by the MSIP, Korea, under the "ICT Consilience Creative Program" (IITP-2015-R0346-15-1007).

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