# Ensemble Traces: Interactive Visualization of Ensemble Multivariate Time Series Data

***Swastik Singh, Song Zhang; Mississippi State University; Mississippi State, MS, USA***
***William Andrew Pruett, Robert Hester; University of Mississippi Medical Center; Jackson, MS, USA***

## Abstract

*This paper presents a simple yet effective approach to visualizing ensemble multivariate time series as 3D traces. Ensemble multivariate time series data are common in many areas. This type of data contains large amount of information which is often crucial to both knowledge discovery and decision making. Visualization can be employed to help the researchers quickly gain insight from the data. First, we project all multivariate data points to a 2D projection plane with a dimension reduction algorithm. Then we expand the data points of any ensemble member back into a trace in the 3D space spanned by the 2D projection plane and time. The resulting 3D ensemble traces provide a holistic and consistent view of the original ensemble multivariate time series. These traces are useful for revealing differences between ensembles, identifying groups and outliers, and catching temporal trends. In addition, we interactively link the ensemble traces to a panel of single variable plots. The combined visualization of raw data plots and multivariate ensemble traces provide a unique perspective to patterns and trends. We studied 3 different dimension reduction algorithms, i.e., t-Distributed Stochastic Neighbor Embedding (t-SNE), classic Multidimensional Scaling (MDS), and Locally Linear Embedding (LLE). We demonstrated our approach with two different datasets and evaluated our methods with domain experts.*

## Introduction

Time series data frequently arise in many areas. These data can have multiple attributes at each time point, known as multivariate time series data. Ensemble data are often simulated in scientific fields by varying parameters or initial conditions. As a result, ensemble multivariate time series, also referred to as doubly multivariate time series data [2], are increasingly more common. An example of such data is the physiological variables of multiple virtual patients over a period of time simulated with the HumMod model [3].

Another example would be emission produced by the electric power industry from various energy sources by states measured in a time period [4]. The different types of emission produced from different sources are the variables and such multivariate parameters are present for each state in the United States. The states form the ensemble. Such ensemble multivariate time series data can also be found in many other fields.

This work was originally motivated by the need to make sense of the ensemble physiological simulations produced by HumMod. However, the problem of visualizing ensemble multivariate time series is a generic one. We had several criteria in mind for visualization design (listed below). None of the available methods are particularly effective for our purposes.

- Intuitive time dimension representation. To identify temporal trends, a user must be able to follow the time dimension intuitively. That means little or no distortion or discontinuity in the time dimension. Methods that transform the time dimension or break it up will add to the cognitive difficulty.
- Unified multivariate representation. The wealth of information contained in the multivariate time series lies not only in individual variables, but also the relationship between the variables. Placing these variables in separate plots, or even in several lines of the same plot, will require the user to mentally integrate them which is cognitively challenging and will not scale well.
- Ensemble representation. On top of the multivariate time series, the ensemble data add to the complexity. Methods that work on a single multivariate time series might not work on an ensemble. An effective visualization must be able to distinguish between individual ensemble members and reveal the relationship between them at the same time.
- Visual clarity. This is a primary goal for all visualization and certainly applies here.

To date, line plot is the de facto standard for single variable time series visualization. It is clear, intuitive, and has less visual clutter. Therefore, we strive to find an analog to the time series line plot for ensemble multivariate time series. Ensemble traces are the result of this effort.

Our central goal in this paper is to visualize ensemble multivariate time series data. Given design criteria listed above, current methods from ensemble visualization, multivariate visualization, and time series visualization cannot be straightforwardly extended to ensemble multivariate time series. We have experimented with existing techniques, i.e., star plot [5] and TimeWheel [6], on the data. Neither of these techniques provides a good and intuitive representation of the ensemble data (see more details in the Results and Evaluation section). We have also animated multivariate ensemble data along the timeline. However, users still need to mentally integrate information from all time points into a holistic view. And this can be a challenging task. Hence, we want to achieve several objectives under the central goal. First, we want to show patterns, trends and outliers in the data. Second, we want to see the relationship between ensemble members and the groups they form. Third, we want to be able to intuitively interpret the patterns and relations. Fourth, we want to predict a new sample's behavior from existing data.

Our solution is a simple one. We project multivariate data points excluding the time dimension into 2D points using a dimension reduction technique. These 2D points are then reassembled for each ensemble member in the time dimension, forming a group of 3D traces. The resulting 3D traces incorporate the information from all the variables and provide a summary of the original data in a holistic manner. In addition, we also allow users to selectively browse ensemble line plots of the single variable time series in the data. A user can select a group of 3D traces or 2D time series with brushing and the selected data will be simultaneously highlighted in multiple views. This enables the user to link the patterns identified in integrated 3D traces to those in individual

variables, hence facilitating the discovery and interpretation of patterns and trends.

We experimented with three different dimension reduction techniques in this paper, i.e., t-Distributed Stochastic Neighbour Embedding (t-SNE) [7], classical Multidimensional Scaling (MDS) [8], and Locally Linear Embedding (LLE) [9]. We identified MDS as a suitable dimension method for our purpose.

We applied our approach to two dataset. The first dataset is the simulated hemorrhage data of virtual patients from HumMod [3] software. The second dataset is the state wise total electric power generation from different sources and total emissions of CO2, SOx and NOx released from different sources of electricity between 1999 and 2012. We pre-processed these data from the sources and made sure that there is no missing value [4].

To our knowledge, the ensemble trace representation of ensemble multivariate time series is novel. This method compared favorably to existing approaches for multivariate and ensemble time series visualization. First, the similarity in shape between the ensemble traces and the individual-variable line plots helps users intuitively understand and explore these traces. This similarity exists not only in their appearance, but also in their shapes. In fact, a user can control the similarity of the ensemble traces to the line plot of any variable by heavily weighting that variable (variable weighting is discussed in Weighted Dimension Reduction section). Second, the ensemble traces effectively incorporate the information from all variables, hence capable of exhibiting patterns and trends in any variable or combination of any group of variables. Third, in the linked view, the ensemble traces can serve as a summary of patterns from all variables, and the line plot panel can be effectively referenced from the ensemble traces by brushing, therefore facilitating the knowledge discovery and validation. Moreover, interactively weighting the variable provides another way to explore the data. Compared to alternative methods, the 3D traces' simplicity in shape and visual mapping allow them to scale well with the number of variables and the number of ensembles. The 3D perspective allows the users to easily identify the fine details of the 3D traces.

The rest of the paper is outlined as follows: we first discuss the related work in the following section. Then we describe our methodology. We then discuss our results and evaluate it. Finally, we conclude the paper.

## Related work

The multivariate time series data visualization is a challenging problem. One of the early works in this area is ThemeRiver [10], which is simple and can easily show trends and patterns across the time in the multivariate data. Kaleidomaps [11], CircleView [12], MultiComb[6], and TimeWheel [6] are visualizations for time-oriented data that are based on the radial axes layout. One of the major issues with these visualizations using radial layout is being unable to represent data effectively with an increasing number of variables. There are also 3D versions of TimeWheel [13] and MultiComb [13] but these visualizations have additional complexity and also share similar issues as their 2D counterparts. Other 3D visualizations that deal with multivariate time series data visualization include Time Tunnel [14] and Temporal star [15]. Both Time-tunnel and Temporal star are interactive visualization tools which use a central axis to represent time. However, these visualizations also have a radial layout arrangement and similar issues arise as mentioned above.

A convenient visualization method for ensemble data is the small multiples method [16, 17] associated with linking and brushing operations. Spaghetti plot [18] is another approach to visualize ensemble dataset. Potter et al. [16] introduced a framework Ensemble-Vis to visualize and explore ensemble dataset by leveraging multiple coordinated views. Sanyal et al. [19] developed a tool named Noodles to visualize ensemble uncertainty which was modeled with the standard deviation. Unfortunately, Noodles is feasible for only single variable ensemble data. Recently, research on this topic advances from visualization to analysis. Thomas et al. [20] presented an interactive system to study off-shore structures in ensemble ocean forecasting dataset. Gosink et al. [21] proposed a method to characterize different types of predictive uncertainty in ensemble dataset based on the Bayesian model averaging. Hummel et al. [22] developed a Lagrangian framework for visual analysis of ensemble flow fields.

Our approach visualizes ensemble multivariate time series data, which is more complicated as seen by the fact that little work has been performed in this area. Zhang et al. [23] used image plot to visualize this type of data, but problems and complications arise when the number of variables and ensembles increases. The data used by Dang et al. for visualization with TimeSeer [2] bears a close resemblance of the data type we have dealt with in our research. TimeSeer provides an interactive platform to visualize and organizing multiple multivariate time series data (doubly multivariate data series [2]) and it is good for dealing with high dimensional data. Also, it is good in detecting outliers in the time series data. However, TimeSeer has a sophisticated display with lots of details which may require user training for data exploration. Other works include techniques by Forlines et. al. [1] and Tominski et. al. [24], which uses space-time visualization. However, both of these techniques might not be able to instantly handle the data with higher number of variables and also it might be difficult for these techniques to identify groups of similar ensembles when the number of ensembles is high.

## Methodology

### *Datasets*

To test our method, we have chosen two different datasets from two different fields. The first dataset is simulated hemorrhage data of virtual patients from HumMod software [23] . The second dataset is the electricity data that consists of state wise total electric power generation and total emissions of CO2, SOx and NOx released from different sources of electricity from 1990 to 2012.

HumMod [3] is a human physiological simulation software. Using this software hemorrhage simulation was conducted in 399 physiologically viable virtual individuals at 3 fixed hemorrhage rates: 37.5, 75, and 150 mL/min over 40 minutes. The simulation recorded 19 physiological parameters during this time period. The measurements were taken every minute for every virtual patient. The result of the simulation is a time series multivariate data of an ensemble population. We also considered the Hemorrhage target rate as one of our parameters. Therefore, we have a total of 20 multivariate parameters. The measured physiological parameters during the simulation include SystemicArtys.SBP, SystemicArtys.Pressure, Breathing.RespRate, SystemicArtys.DBP, Breathing.TidalVolume, PO2Artys.Pressure, PO2Artys.Sat(%), Heart-Rate.Rate, CO2Veins.Pressure, BloodPhValues.ArtysPh, BloodPhValues.VeinsPh, BrainInsult.Effect, CardiacOutput.Flow, LeftAtrium.Pressure RightAtrium.Pressure, Brain-Ph.Ph, CardiacOutput.StrokeVolume, Brain-Flow.BloodFlow, Brain-Fuel.Adequacy, and Hemorrhage.TargetRate. The result of the simulation is a time series multivariate data of an ensemble

population. For the visualization purpose, we uniformly sampled 300 patients from 399 patient set. This selection reduces the number of ensembles for visualization, which eventually reduces clutter and also provides a good sample representation of the data. This also reduces the computation time of the dimension reduction algorithm.

Our second dataset is electricity data. This dataset consists of state wise total electric power generation from different sources from 1990 to 2012. These sources include coal, hydroelectric conventional, natural gas, petroleum, wind, wood and wood derived fuels, nuclear, other biomass, other gases, pumped storage, geothermal, solar thermal and photovoltaic, and other sources. Also, this dataset contains state wise total emission of CO2, SOx and NOx from different sources during electric power generation. These sources include: coal, natural gas, petroleum, geothermal, other biomass, wood and wood derived fuels, other gases and other sources. There are a total of 37 different variables in the data. The details of these data can be found in U.S. Energy Administration website [4].

### *Ensemble Trace*

The generation of ensemble traces is a two-step process. The first step is to project data points to lower dimensions and the second step is to reassemble all the data points of an ensemble member along its time line.

Assume our ensemble multivariate time series data have e × v × t values, e is the number of ensemble members, v is the number of variables, and t is the number of time points. We first disassemble all v dimensional data points from all ensemble members and all their time points. This will give us a n × v matrix D, where n= e × t. The n rows comprise of the data points and v is the dimension of each data point. For example, in the hemorrhage study, we have an ensemble of 300 members; each member has 40 time steps; in each time step 20 variables are recorded for each member. Therefore, the matrix D is $12000 \times 20$. 12000 data points come from 300 (number of ensemble members) multiplied by 40 (time steps).

The first step is to reduce the dimension of each of the n data points to some manageable dimensions for visualization. The dimensions will be feasible for visualization if they are between one and three. In the time series data, the time dimension has a special meaning, and we want it to be preserved in visualization for showing temporal trends rather than mixed together with other variables. Therefore, we reserve one visual dimension for time, and have to choose between one or two dimensions for all other variables. Projection to two dimensions has less projection error than to one dimension. Therefore, we chose to project the high dimensional points excluding time to two dimensions.[25]

The second step is to reassemble all the data points of an ensemble into a trace in a 3D space spanned by the 2D projection plane and the time dimension. We place the projected 2D points into the 3D volume based on their 2D locations and time stamp, and link the multiple data points of the same ensemble member with line segments between adjacent time points, resulting in a polyline we call an ensemble trace.

Figure 1 shows an example of the ensemble traces for 300 patients in the hemorrhage simulation. We can observe that at the beginning of the simulation (time=0), all patients share the same parameters, and as the hemorrhage progresses the ensemble traces diverges into different paths, which we can clearly see in the figure. The diverging patterns are clearer when viewed in an interactive 3D display.

We allow users to assign weights to variables in the dimension reduction technique to emphasize the importance of certain variables. In addition, we employ the hierarchical clustering algorithm to cluster ensemble traces into distinct groups. And we allow partial ensemble traces to be embedded in the visualization for prediction.

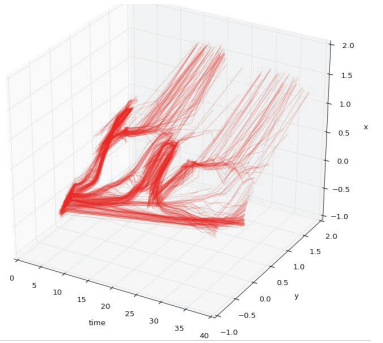



*Figure 1. Visualization of hemorrhage data of patients for dimension reduced normalized hemorrhage data for 300 patients using ensemble traces. (x,y) is the reduced dimension which is plotted against time.*

### *Dimension Reduction*

Dimension reduction technique maps v dimensional points of the data to a lower dimension k. There are numerous dimension reduction techniques. In this paper, we studied three different widely used dimension reduction methods, which are t-SNE, LLE and classical MDS, to find out a suitable method for our purpose. LLE and t-SNE are nonlinear methods, whereas classical MDS is a linear method. The overarching goal is to find out which of these techniques is able to capture the overall structure of the data and make sense in lower dimensional space. Apart from this, we also consider computational efficiency as a factor in the selection of dimension reduction algorithms. Before we apply the dimension reduction algorithm to our data points we normalize the values of all the variables in the range of 0 and 1 to bring them into the same scale.

t-SNE (t-distributed Stochastic Neighbor Embedding) [7] is a probabilistic technique used to visualize high dimensional data by shrinking it to lower dimensions (2D or 3D) based on SNE (Stochastic Neighbor Embedding) [26]. t-SNE tries to maintain the similarities between data points while composing low dimensional representation. This technique is capable of preserving local properties as well as global properties of the original high dimensional data. While computing low dimensional points of the data, t-SNE minimizes the Kullback-Leibler divergences of joint probabilities between high dimensional and low dimensional representation. The use of student's t-distribution and symmetric version of SNE cost function are two important features of t-SNE which sets it apart from SNE. Student's t-distribution is used to compute similarity between points in low dimension where t-SNE cost function has a simpler gradient, which is easier to optimize. [7]

LLE is a nonlinear dimension reduction technique which tries to preserve the local properties of the data in lower dimensions [27]. The LLE algorithm begins by stating the number of neighbors per data point k, which is the input of the algorithm. Then in the next step, after identifying k neighbors of the data points, optimal weights are computed by minimizing the reconstruction error given by a cost function, which are used for the reconstruction of the original data

points as accurately as possible. After this, the weights computed in second step are used to reconstruct the lower dimensional data points that best reconstructs the original data points by minimizing the embedding cost function [9, 28].

Classical MDS is a linear dimension reduction technique that maps v dimensional points of the data to a lower dimension k while trying to preserve the pairwise distances between the points [27]. The input of MDS includes a distance matrix and the target dimension. There are many choices for computing the distance matrix between n data points, including Euclidean, Manhattan, Canberra, etc. We empirically tested Euclidean, Manhattan and Canberra distances in our visualization. The resulting visualization using these distance matrix was similar. However, we found that Euclidean distance showed slightly better divergence of ensemble traces. Therefore, we decided to choose to compute the Euclidean distance between any two multivariate points.

In this study, to compute the dimension reduction in a reasonable time, we uniformly sample ensembles from the ensemble population.

### *Weighted Dimension Reduction*

By normalizing the values in all variables, we assume equal importance among variables. This is often not true. For example, in the hemorrhage data, a physiologist may be more interested in easily measurable variables in a clinical setting and their patterns, or they may place more emphasis on vital physiological parameters like heart rate. For this reason, we allow users to assign weights to the variables they deem important. The importance of the variable is determined by the users' prior knowledge and interests. We add weights to the data points after normalization and before computing the Euclidean distance matrix between the data points. After computing the Euclidean distance matrix we run a dimension reduction algorithm which will provide 2D representation of our original data points with user-defined weights on different variables.

### *Shepard Plots*

Shepard plot is a scatter plot that maps the distance between any points before dimension reduction to the distance after dimension reduction. If a dimension reduction is able to completely preserve the distances, then the points on the Shepard plot will be on the x=y line. More spreading of the points in the scatterplot can be interpreted as more error in the dimension reduction process. We use Shepard plot to show how good or bad the lower dimensional representation is of the higher dimensional points. The plots are provided to the users to make them aware of the projection error in the data exploration process. This method is suitable for visualizing the projection error in classical MDS but it may not be suitable for t-SNE and LLE. For t-SNE the lower dimensional coordinates are probabilities and not distances, so low dimensional coordinates obtained from t-SNE cannot be used to compute projection error [29]. LLE is a non-linear method and local approach which tries to preserve the local geometry of the data [9, 30, 31] and may not perserve global geometry. Hence Shepard plot may not work for LLE [25].

### *Clustering*

Users often want to find similar ensemble members, or very different ensemble members. Ensemble traces may run together in part or all of the time period, forming groups. A user can examine the grouping behavior of the ensemble traces directly in the 3D view. In addition, we also provide automatic clustering of the ensemble traces. We empirically select the hierarchical clustering algorithm with Ward's criterion [32] and we compute the distance matrix with the Euclidean distance. The resulting clusters can be viewed with different colors.

The decision to choose the hierarchical clustering method and the Euclidean distance is from our initial empirical testing results. We also experimented with the k-means clustering method, the Manhattan distance and Dynamic time wrapping distance.

The time series data of the ensembles must be transformed and organized to compute the distance matrix with Euclidean distance. To illustrate this lets consider an i[th] ensemble $E_i$ with t points that form time series data. Each of these t points has a dimension d. We organize these data into a single vector $V_i$ in the following way.

$$V_i = (X_{i11}, X_{i12}, \dots X_{i1d}, X_{i21}, X_{i22}, \dots X_{i2d}, \dots \dots, X_{it1}, X_{it2}, \dots X_{itd}) \quad (1)$$

The Euclidean distance is then calculated as the Euclidean norm of the difference between any two vectors $V_i$ and $V_j$: $\|V_i - V_j\|$.

### *Partial Trace Embedding*

We provide the capability of embedding partial ensemble traces in the visualization. This is convenient when the data for only part of a time period are available for an ensemble member. For example, in the hemorrhage simulation, for a new virtual patient that needs to be diagnosed or triaged, only part of the 40 minutes simulation may be available. By embedding the new patient in the ensemble traces of other patients, a visual prediction can be made based on which group this patient belongs to. The physiological progression of this patient may be predicted by the behavior of other patients in the same group. To test this approach, we randomly select 4 patients from the hemorrhage data and we truncate their time series so that we have only the first 60% of the time points. We then embed these patients with our other patients with complete data and visualize all with ensemble traces. We can predict the path of a partial ensemble member with the group it coincides with partially. This technique was especially suitable for our hemorrhage data because ensembles with same bleeding rates clustered into different groups and these different groups can be visually identified. The results are shown in Results and Evaluation section.

### *Single Variable Plot*

The 3D ensemble traces provide a holistic view of the underlying data. But if the users want to go in detail and explore underlying reasons behind the patterns, a single variable time series plot provides the user the opportunity to view the raw data in addition to the ensemble traces. We allow the users to select the variables they want to view and place the plot of each selected variable in a panel. In each single variable plot, all ensemble members are shown as individual lines.

Figure 2 shows a panel of single variable plots for four selected variables. The data is same as in Figure 1.

Due to the nature of dimension reduction, there will almost always be distortions in the ensemble traces. In this case, the 2D view will play a crucial role in complementing the 3D plots because there is no distortion in 2D plot and all information is preserved, even only for a single variable.

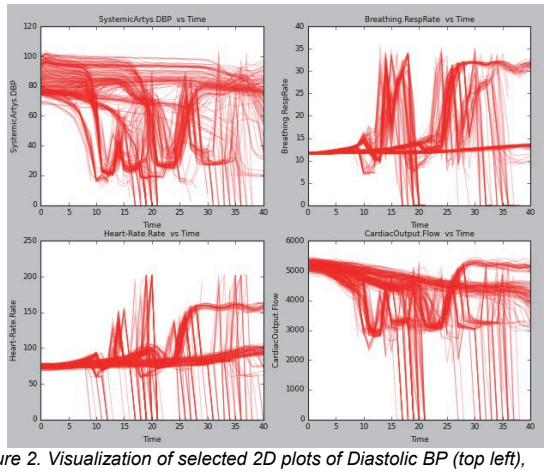



*Figure 2. Visualization of selected 2D plots of Diastolic BP (top left), Respiration rate (top right), Heart rate (bottom left) and Cardiac output flow (bottom right) for hemorrhage data.*

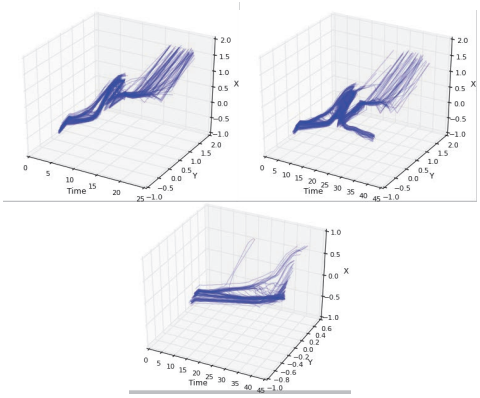


*Figure 3. Brushing produces three groups of ensemble traces generated using MDS that are separated by the hemorrhage rates.*

## Linked Interaction

To allow users to effectively explore the visualization, we provide standard 3D rotation and zooming interactions for ensemble traces. In addition, we provide the brushing interaction which allows the user to select the ensemble traces in the 3D view and time series in the 2D plots. Moreover, if both 2D and 3D windows are opened, then brushing in one view will be used to update all the visualizations, allowing users to highlight ensembles in both 2D and 3D views. This is known as linked interaction. To reduce visual clutter, our interface also provides the option to view only the highlighted polylines in both 2D and 3D plot, which allows the user to focus on the selected data.

The linked interaction becomes more effective with other features within our visualization such as zooming and rotation in 3D plots. They help the user to visualize ensemble traces in a convenient way by brushing the ensemble traces either from different angles or from enlarged view.

The linked interaction helps the user make sense of the 3D plot and ultimately understand the patterns in the underlying data. Figure 3 shows the three groups of ensemble traces identified with brushing. These 3 groups represent the three different ensemble

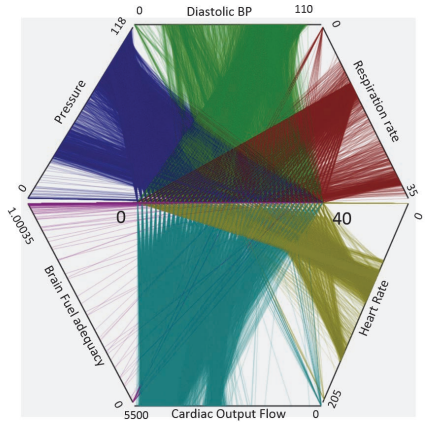



*Figure 4. Visualization of hemorrhage data using TimeWheel*

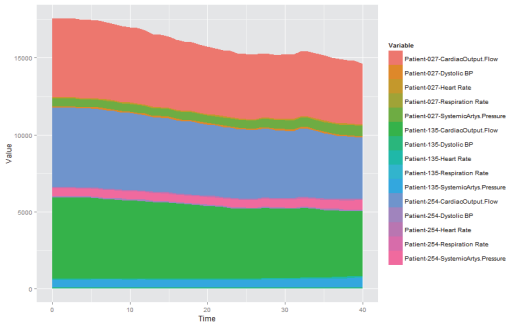



*Figure 5: Stacked area chart for 3 patients with 5 parameters from hemorrhage data*

populations having 3 different hemorrhage rates as shown in Figure 10.

## Comparison to Other Methods

Aigner et al. [33] and their TimeViz website (http://survey.timeviz.net/) provide a good survey of current methods for time oriented data. We have researched current methods and experimented with several of them. In this paper, we will discuss TimeWheel, star plot and Wakame[1] and compare them with our visualization.

Figure 4 shows the visualizations of the hemorrhage data with TimeWheel. For TimeWheel, we plotted 6 variables for 300 patients for all time points. The resulting visualization is cluttered with many lines intersecting each other. The relationship between variables and individual time series are difficult to discern. Moreover, the visualization fails to accommodate all parameters easily as the number of parameters increases. This is true for other visualizations that uses radial layout for visualization of multivariate data, e.g. Star plot shown in Figure 6.

Figure 5 shows the stacked area chart which is similar to ThemeRiver [10]. The ThemeRiver type of visualization is effective for showing aggregated effects of multiple variables over time, but the relationship between variables and individual time series is difficult to discern.

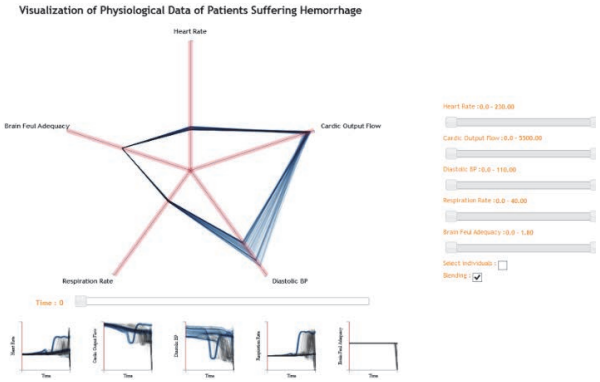



*Figure 6. Visualization of hemorrhage data of patients using star plot*

Figure 6 shows one our experimental studies of multivariate time series data visualization of ensembles using a star plot. The star plot visualizes the time series data of multiple patients. The time slider is used to navigate through time and study the temporal changes. The starplot displays the variables value at one instance of time. We found that the star plot's radial layout and the use of time slider is not effective in visualizing our data. The radial layout does not scale well as the number of variables increases. The use of time slider was also not very effective because there are many data points in our data and it is difficult to follow all the data points with the time slider [34].

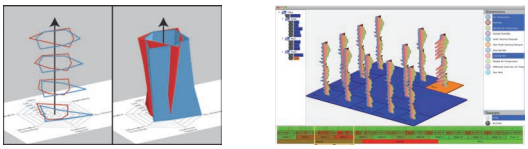


*Figure 7. (a) Wakame visualization design [1] (b) Wakame visualizations of sensor data's recorded for 11 rooms [1]*

Figure 7 shows an example of Wakame visualization design with an example of a prototype system taken from Forlines et. al. [1] paper. Wakame is a 3D object formed by stacking radar chart that grows up vertically with time and each Wakame can represent multivariate time series data. Figure 7b shows a prototype system that uses Wakame visualizations to display multiple multivariate time series data, i.e., sensor data's recorded for different rooms. The weakness of this visualization lies in the difficulty in comparison and identification of similar ensembles as their number increases. It is also difficult to accommodate parameters as its number increases because its design is based on radar chart.

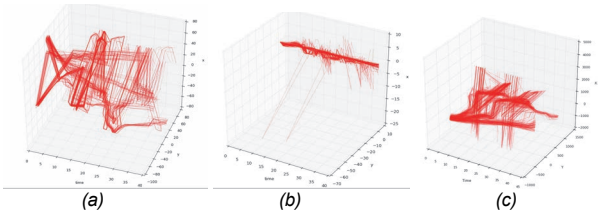
Ensemble traces compare favorably to the available methods, including the three shown in Figure 4, Figure 5, Figure 6 and Figure 7. One particularly important advantage of ensemble trace is its simplicity, and the resulting easy interpretation.

## Results and Evaluation

We applied our method to the two datasets described in the Dataset section. Two physiologists provided empirical evaluation for the hemorrhage data results.

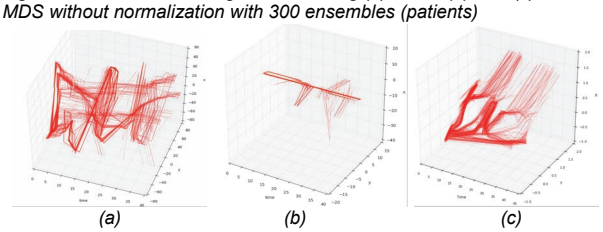
As shown earlier in Figure 1, the ensemble traces for the 300 patients simulated in a hemorrhage experiment were generated. We notice the branching of the polylines in a three dimensional space. In this image, the time-dynamic interplay between heart rate and mean arterial blood pressure are shown. Physiologically, these factors reflect the integration of the baroreceptor reflex loop that uses heart rate and contractility to counter low blood pressure. In the acute setting, this is the primary mechanism by which humans maintain their blood pressure. By tracking both heart rate and blood pressure against one another, one can determine the effectiveness of the baroreflex system. If heart rate is rising while blood pressure is falling, that suggests that the compensatory limits of the baroreflex have been exceeded and cardiac decompensation is approaching. While it is typical, in an emergent situation to show each of these variables against time, the picture is clarified by increasing the dimensionality of the data. In this example, it is clear that the baroreceptor reserve is disappearing as the curves make their way to their peak.

Figure 2 shows a panel of user-selected single variable line plots for the same data. These images more closely represent the current standard of data visualization for a working physician. In this case, the physician must commit effort into integrating the three traces. This takes away effort that could instead be devoted to managing the patient's care.



(a) (b) (c)

*Figure 8. Ensemble traces generated using (a) t-SNE (b) LLE (c) classic MDS without normalization with 300 ensembles (patients)*



(a) (b) (c)

*Figure 9. Ensemble traces generated using (a) t-SNE (b) LLE (c) classic MDS with normalization with 300 ensembles (patients)*

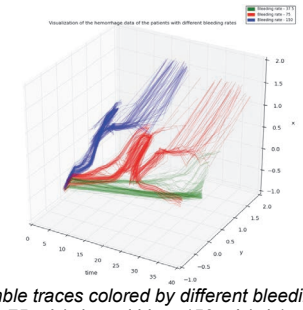


*Figure 10. Ensemble traces colored by different bleeding rates (green-37.5 mL/min, red -75 mL/min and blue -150 mL/min)*

Figure 8 and 9 shows the ensemble traces generated using t-SNE, LLE and classic MDS using normalization and without normalization of the hemorrhage data. The results with normalization show the divergence of traces more prominently in the visualization as we can see from Figure 9c. LLE fails to display ensemble traces effectively both in the case of with normalization

and without normalization with 300 ensembles as seen in figure 8b and 9b. In the case of t-SNE, we see curves more scattered in space (Figure 8a and 9a) compared to the other two methods. However, we found that the running time of t-SNE is longer compared to other methods on hemorrhage data with 300 ensembles in our case. The ensemble traces generated by classic MDS with normalization have more pronounced divergence and clearer ensemble paths than without normalization as seen from Figure 9c and Figure 8c respectively. In our case of hemorrhage data, the MDS results revealed the patient groups well and was quicker to compute compared to t-SNE. The resulting ensemble traces generated using classic MDS were more organized and meaningful in our case than compared to t-SNE. Moreover, the Shepard plot for hemorrhage data for MDS from Figure 15 reveals that lower 2D and 3D representations of the data points are comparable with a reasonable

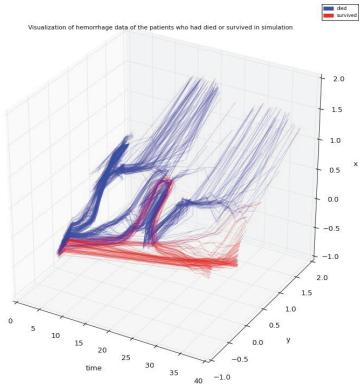



*Figure 11. Ensemble traces showing surviving patients and dying patients at the end of simulation*

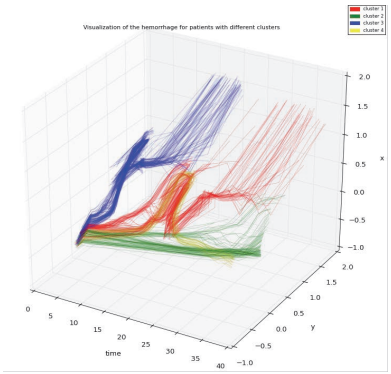
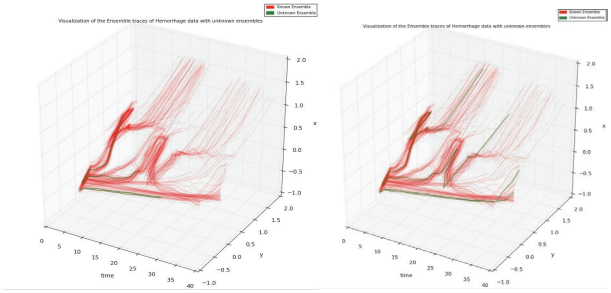



*Figure 12. Ensemble traces showing 4 clusters of ensemble traces*



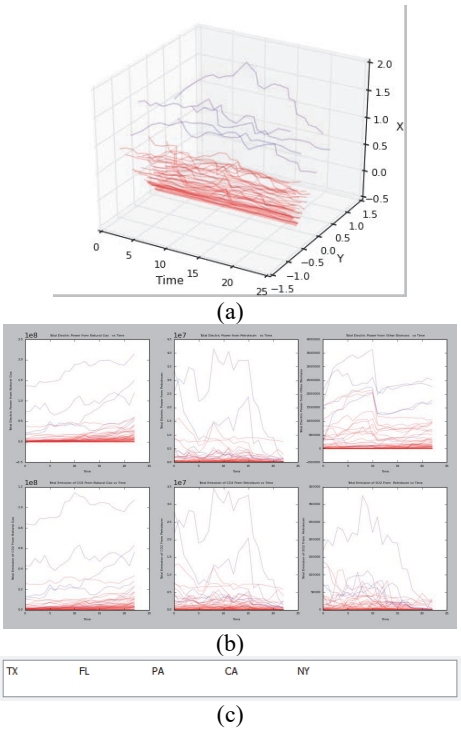

(a) (b)

*Figure 13. Prediction of path of unknown ensemble Traces (Red-Known ensemble and Green – Unknown ensemble) (a) partial path with incomplete data (b) actual path with complete data*

projection error. Comparing all the resulting visualizations, we conclude that classic MDS is more effective and suitable in our case than LLE and t-SNE. Also, we see that normalization plays a crucial role in dimension reduction.

We studied the ensemble traces by coloring it on the basis of hemorrhage rate and survival/death of the patients at the end of the simulation. We see that figure 10 and 11 show that some inherently different groups of the patients are visually separated by the ensemble traces. Figure 10 shows the ensemble traces colored by the hemorrhage rate. The three different colors represent three different bleeding rates as shown in the figure. It is clear that the ensemble traces form bundles separated by different hemorrhage rates. Note that, especially in the 75 ml/min case, patient behavior bifurcates into two dissimilar bands. This reflects the effects of different physiologies on the decompensation process, and reflects the type of awareness that a physician must keep when treating patients. Essentially, early decompensators must be recognized and treated quickly, while late decompensators can have their care safely postponed. Figure 11 shows the patients who died (blue traces) and the patients who survived (red traces). Most of these two groups of patients are separated in their regions. However, there is a group of patients with similar physiological responses in the first half of the simulation that bifurcate in the second half of the simulation. This bifurcation pattern is particularly important in a clinical setting since these patients need to be further evaluated for treatment.

Figure 12 shows the hierarchical clustering results with 4 clusters. Note the similarity between the patterns in Figure 10 and Figure 12. In this case, unsupervised clustering reveals the main




(a) (b) (c)

*Figure 14. Ensemble traces showing state wise total electric production and Total emission (CO2, SO2 and NOx) from different sources (a) Selected ensemble outliers in 3D plot (b) 2D plots for the ensemble with outliers selected in 3D plot (2D plots is for Total Electric power generated from coal and petroleum; and CO2, SO2 and NOx generated from power station using coal and petroleum as fuel] (c) Outlier States displayed in the visualization interface*

reason behind the three different groups of responses among patients.

Figure 13 demonstrates the utility of the ensemble trace for temporal prediction. Figure 13a shows three patients with partial time simulations (green traces) and all other patients with full time simulations. It is clear that each of the three green traces belongs to its own group. Prediction can be made that each trace will likely follow the trends within its group in the future. Figure 13b shows all the patients with full time simulations. It confirms that the three patients indicated by green indeed follow the patterns of their respective groups.

It is well known to trauma physicians that there are no statistical differences in heart rate, blood pressure, or oxygen saturation status in patients who are about to enter hemorrhagic shock and patients that are not [35]. Various systems exist that assign an index to a patient's hemodynamic status in order to inform physicians of their current level of risk. These indices generally cannot be deconvoluted into specific information about the interacting systems, and therefore cannot aid treatment. In situations where patients require close monitoring, physicians follow two dimensional plots of variable versus time, with a series of decision trees that determine their next action. For instance, low pressure and high heart rate is bad, low pressure and low heart rate is good, etc. By visualizing more variables simultaneously, the "ands" and "ors" of the decision process can be reduced to observing the trends on the ensemble traces. By reducing the cognitive load of the physician in observing patient status, working memory is freed for the task of treating the patient, or for managing more patients. This will also flatten the learning curve by reducing the number of observations required for understanding patient status by trainee physicians.

Figure 14 shows the ensemble traces and several individual variable plots for the electricity data. From the ensemble traces, one can easily identify several traces that are far away from the majority of the traces. These are likely outliers. After brushing them with blue color, we can further examine the behavior of the individual variables of the blue traces. We use a text box in our visualization which displays the states of the highlighted ensemble traces in the visualization. This is shown in Figure 14c. We see that New York,

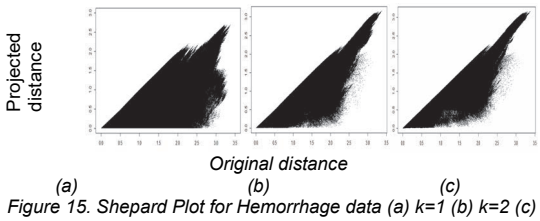



*Figure 15. Shepard Plot for Hemorrhage data (a) k=1 (b) k=2 (c) k=3 using MDS*

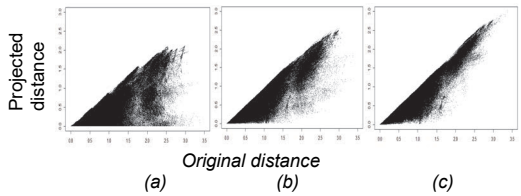



*Figure 16. Shepard Plot for State wise electric data [Total electric production and Total emission (CO2, SO2 and NOx) from different sources] (a) k=1 (b) k=2 (c) k=3 using MDS*

Texas, Florida, Pennsylvania and California are the outlier states indicated by ensemble traces highlighted in blue in Figure 14a. Upon further examination in single variable plot as shown in Figure 14b, we find that they are indeed outliers in a number of variables, but not in others. For example, in visualization shown in Figure 14, the outlier states have higher total electric power generation from natural gas as well as total $CO_2$ emissions from natural gas. This case shows the effectiveness of the linked view for not only discovering knowledge but also finding out the underlying causes.

Figure 15 and 16 shows the Shepard plots for the classic MDS results of the hemorrhage data, and the electricity data respectively. There are three Shepard plots generated for each data with the number of dimensions set to 1, 2, and 3 respectively. Not surprisingly, the plots are increasingly less scattered with higher projection dimensions, indicating less projection errors. We note that there is a very noticeable difference between 1D and 2D, while the difference between 2D and 3D is less noticeable. These plots prove the advantage of generating ensemble traces in 3D instead of in a 2D plot.

Figure 17 shows the ensemble traces from the MDS of the weighted variables. 17a to 17e show the results of weight changes

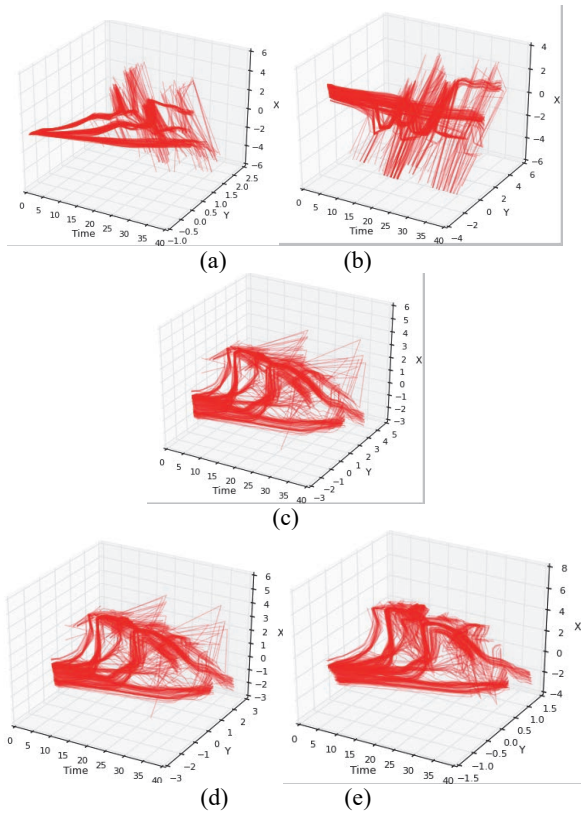



*Figure 17. Visualization of weighted traces from hemorrhage data (a) Heart Rate=100 (b) Heart Rate=75, Diastolic BP=25 (c) Heart Rate=50, Diastolic BP=50 (d) Heart Rate=25, Diastolic BP=75 (e) Diastolic BP=100*

from 100 to 1 for heart rate, and from 1 to 100 for diastolic blood pressure. All other variables are weighted by the default 1. From these images we can see a gradual change of shape in the ensemble traces influenced by the weights on the variables. The ensemble traces will likely exhibit patterns from the more heavily weighted variables, and less likely from the lightly weighted variables. Variable weighting gives the users an important tool for exploring the multivariate aspect of the data.

## Conclusion

This paper presents a simple yet effective method for visualizing ensemble multivariate time series data. Our method utilizes classical MDS to project all multivariate data points from all time steps and all ensembles to 2D points, and then reassemble these points along the time dimension to form a group of ensemble traces. We applied this method to two datasets and demonstrated the effectiveness of the method. Two domain experts confirmed the effectiveness of the ensemble traces in visualizing ensemble multivariate time series data.

## References


[1] C. Forlines and K. Wittenburg, "Wakame: sense making of multi-dimensional spatial-temporal data," in Proceedings of the International Conference on Advanced Visual Interfaces, 2010, pp. 33-40.

[2] T. N. Dang, A. Anand, and L. Wilkinson, "Timeseer: Scagnostics for high-dimensional time series," Visualization and Computer Graphics, IEEE Transactions on, vol. 19, pp. 470-483, 2013.

[3] R. L. Hester, A. J. Brown, L. Husband, R. Iliescu, D. Pruett, R. Summers, et al., "HumMod: a modeling environment for the simulation of integrative human physiology," Frontiers in physiology, vol. 2, 2011.

[4] United States Energy Information Administration website, http://www.eia.gov/electricity/data/state/, accessed March 21.

[5] M. J. Saary, "Radar plots: a useful way for presenting multivariate health care data," Journal of clinical epidemiology, vol. 61, pp. 311-317, 2008.

[6] C. Tominski, J. Abello, and H. Schumann, "Axes-based visualizations with radial layouts," in Proceedings of the 2004 ACM symposium on Applied computing, 2004, pp. 1242-1247.

[7] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," Journal of Machine Learning Research, vol. 9, p. 85, 2008.

[8] A. Buja, D. F. Swayne, M. L. Littman, N. Dean, H. Hofmann, and L. Chen, "Data visualization with multidimensional scaling," Journal of Computational and Graphical Statistics, vol. 17, pp. 444-472, 2008.

[9] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," Science, vol. 290, pp. 2323-2326, 2000.

[10] S. Havre, B. Hetzler, and L. Nowell, "ThemeRiver: Visualizing theme changes over time," in Information Visualization, 2000. InfoVis 2000. IEEE Symposium on, 2000, pp. 115-123.

[11] K. Bale, P. Chapman, N. Barraclough, J. Purdy, N. Aydin, and P. Dark, "Kaleidomaps: a new technique for the visualization of multivariate time-series data," Information Visualization, vol. 6, pp. 155-167, 2007.

[12] D. A. Keim, J. Schneidewind, and M. Sips, "CircleView: a new approach for visualizing time-related multidimensional data sets," in Proceedings of the working conference on Advanced visual interfaces, 2004, pp. 179-182.

[13] C. Tominski, J. Abello, and H. Schumann, "3D Axes-Based Visualizations for Time Series Data," in Proc. of the Ninth International Conference on Information Visualization (IV'05), 2005.

[14] H. Notsu, Y. Okada, M. Akaishi, and K. Niijima, "Time-tunnel: Visual analysis tool for time-series numerical data and its extension toward parallel coordinates," in Computer Graphics, Imaging and Vision: New Trends, 2005. International Conference on, 2005, pp. 167-172.

[15] M. Noirhomme-Fraiture, "Visualization of large data sets: the zoom star solution," International Electronic Journal of Symbolic Data Analysis, pp. 26-39, 2002.

[16] K. Potter, A. Wilson, P.-T. Bremer, D. Williams, C. Doutriaux, V. Pascucci, et al., "Ensemble-vis: A framework for the statistical visualization of ensemble data," in Data Mining Workshops, 2009. ICDMW'09. IEEE International Conference on, 2009, pp. 233-240.

[17] A. T. Wilson and K. C. Potter, "Toward visual analysis of ensemble data sets," in Proceedings of the 2009 Workshop on Ultrascale Visualization, 2009, pp. 48-53.

[18] P. Diggle, P. Heagerty, K.-Y. Liang, and S. Zeger, Analysis of longitudinal data: Oxford University Press, 2002.

[19] J. Sanyal, S. Zhang, J. Dyer, A. Mercer, P. Amburn, and R. J. Moorhead, "Noodles: A tool for visualization of numerical weather model ensemble uncertainty," Visualization and Computer Graphics, IEEE Transactions on, vol. 16, pp. 1421-1430, 2010.

[20] T. Hollt, A. Magdy, G. Chen, G. Gopalakrishnan, I. Hoteit, C. D. Hansen, et al., "Visual analysis of uncertainties in ocean forecasts for planning and operation of off-shore structures," in Visualization Symposium (PacificVis), 2013 IEEE Pacific, 2013, pp. 185-192.

[21] L. Gosink, K. Bensema, T. Pulsipher, H. Obermaier, M. Henry, H. Childs, et al., "Characterizing and visualizing predictive uncertainty in numerical ensembles through bayesian model averaging," Visualization and Computer Graphics, IEEE Transactions on, vol. 19, pp. 2703-2712, 2013.

[22] M. Hummel, H. Obermaier, C. Garth, and K. Joy, "Comparative visual analysis of lagrangian transport in cfd ensembles," Visualization and Computer Graphics, IEEE Transactions on, vol. 19, pp. 2743-2752, 2013.

[23] S. Zhang, W. A. Pruett, and R. Hester, "Visualization and Classification of Physiological Failure Modes in Ensemble Hemorrhage Simulation."

[24] C. Tominski, P. Schulze-Wollgast, and H. Schumann, "3d information visualization for time dependent data on maps," in Information Visualisation, 2005. Proceedings. Ninth International Conference on, 2005, pp. 175-181.

[25] F. Wickelmaier, "An introduction to MDS," Sound Quality Research Unit, Aalborg University, Denmark, 2003.

[26] G. E. Hinton and S. T. Roweis, "Stochastic neighbor embedding," in Advances in neural information processing systems, 2002, pp. 833-840.

[27] L. J. van der Maaten, E. O. Postma, and H. J. van den Herik, "Dimensionality reduction: A comparative review," Journal of Machine Learning Research, vol. 10, pp. 66-71, 2009.

[28] L. K. Saul and S. T. Roweis, "An introduction to locally linear embedding," http://www.cs.toronto.edu/~roweis/lle/publications.html, accessed August 2015., 2000.

[29] L. Van der Maaten, t-SNE – Laurens van der Maaten website, https://lvdmaaten.github.io/tsne/, accessed November 2015.

[30] M. Polito and P. Perona, "Grouping and dimensionality reduction by locally linear embedding," 2002.

[31] V. D. Silva and J. B. Tenenbaum, "Global versus local methods in nonlinear dimensionality reduction," in Advances in neural information processing systems, 2002, pp. 705-712.

[32] F. Murtagh and P. Legendre, "Ward's hierarchical agglomerative clustering method: Which algorithms implement ward's criterion?," Journal of Classification, vol. 31, pp. 274-295, 2014.

[33] W. Aigner, S. Miksch, H. Schumann, and C. Tominski, Visualization of time-oriented data: Springer Science & Business Media, 2011.

[34] G. Robertson, R. Fernandez, D. Fisher, B. Lee, and J. Stasko, "Effectiveness of animation in trend visualization," Visualization and Computer Graphics, IEEE Transactions on, vol. 14, pp. 1325-1332, 2008.

[35] V. A. Convertino, G. Grudic, J. Mulligan, and S. Moulton, "Estimation of individual-specific progression to impending cardiovascular instability using arterial waveforms," Journal of Applied Physiology, vol. 115, pp. 1196-1202, 2013.