TimeSeriesPaths: Projection-Based Explorative Analysis of Multivariate Time Series Data

Jürgen Bernard
Fraunhofer Institute for
Computer Graphics
Research, Darmstadt,
Germany
juergen.bernard
@igd.fraunhofer.de

Nils Wilhelm
Technische Universität
Darmstadt, Germany
nwilhelm
@rbg.informatik.tu-
darmstadt.de

Maximilian Scherer
Interactive Graphics
Systems Group, Technische
Universität Darmstadt,
Germany
maximilian.scherer
@gris.tu-darmstadt.de

Thorsten May
Fraunhofer Institute for
Computer Graphics
Research, Darmstadt,
Germany
thorsten.may
@igd.fraunhofer.de

Tobias Schreck
Data Analysis and
Visualization Group,
Universität Konstanz,
Germany
tobias.schreck
@uni-konstanz.de

ABSTRACT

The analysis of time-dependent data is an important problem in many application domains, and interactive visualization of time-series data can help in understanding patterns in large time series data. Many effective approaches already exist for visual analysis of *univariate* time series supporting tasks such as assessment of data quality, detection of outliers, or identification of periodically or frequently occurring patterns. However, much fewer approaches exist which support *multivariate* time series. The existence of multiple values per time stamp makes the analysis task per se harder, and existing visualization techniques often do not scale well.

We introduce an approach for visual analysis of large multivariate time-dependent data, based on the idea of projecting multivariate measurements to a 2D display, visualizing the time dimension by trajectories. We use visual data aggregation metaphors based on grouping of similar data elements to scale with multivariate time series. Aggregation procedures can either be based on statistical properties of the data or on data clustering routines. Appropriately defined user controls allow to navigate and explore the data and interactively steer the parameters of the data aggregation to enhance data analysis. We present an implementation of our approach and apply it on a comprehensive data set from the field of earth observation, demonstrating the applicability and usefulness of our approach.

Keywords: Multivariate Time Series, Visual Cluster Analysis, Exploratory Data Analysis, Data Projection, Data Aggregation

1 INTRODUCTION

Multivariate time series data are gathered in many domains including economics, experimental physics, computer vision, robotics, and earth observation. E.g., in the financial domain, large amounts of stock prices are tracked over time; in earth observation, daily temperatures and many additional parameters are observed at specific locations over time; time-dependent measurements also arise in monitoring traffic parameters on a communication network. Analysis of time series data can take many forms, including assumption-free exploration; correlation of time series with each other; or evaluation of specific generative models. Much work has been done focused on analyzing one-dimensional time series, and respective solutions are often applied to multivariate data by analyzing each dependent variable versus an independent one. However, for multivariate data the widely used *IID* assumption (independent and identically distributed) usually does not hold. Therefore there is a need to analyze all dimensions of such data at once.

In the context of data mining and visual analytics, multivariate time series analysis is a difficult problem, with solutions typically relying, in some form or the other, on dimensionality reduction, feature selection, projection, and glyph-based visualization. The task at hand often includes finding periodic or frequent patterns in
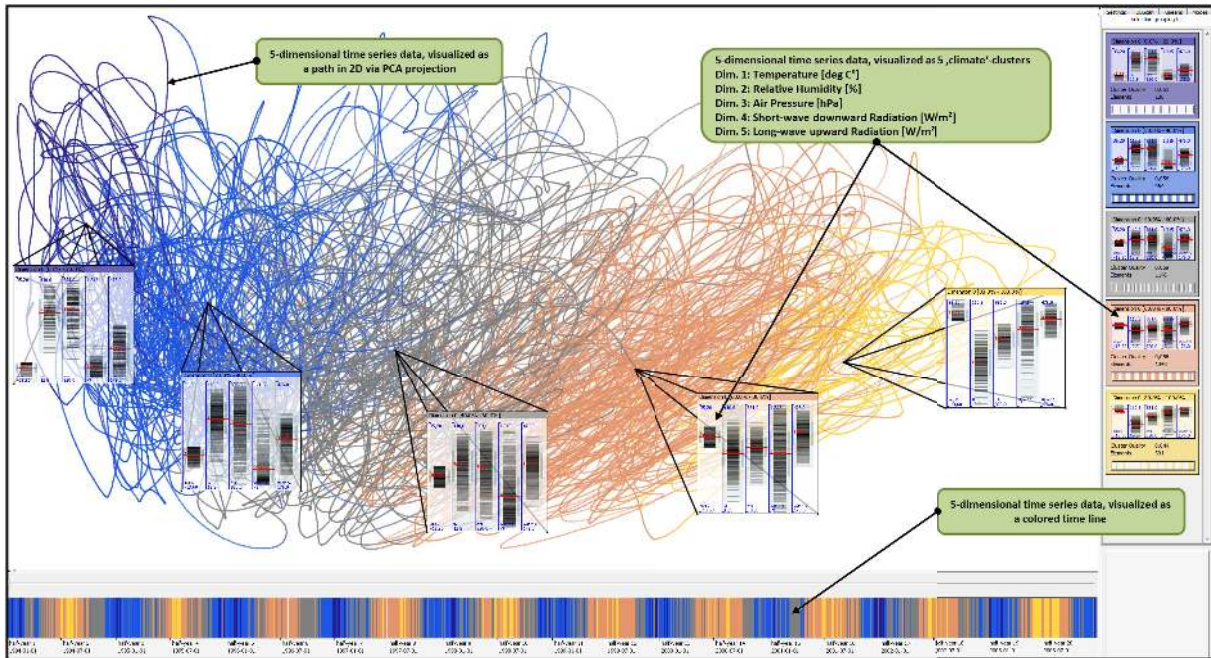
Figure 1: Main display: data analysis of a multivariate time series of 10 years length is always challenging due to overview problems and visual cluttering. This is the starting point of our data exploration. The Time Series Path System provides visual structures and interactive functionality to address the implied challenges. In this example, we aggregate a weather scenario by its temperature values and receive 5 well-distributed data clusters from cold (blue) on the left to warm (yellow) on the right. This is a qualified starting point for selection and filtering approaches to detect periodicity, dense data regions and outliers. Confer our case study in Section 5 for details about the 2D projection.

the data, relating multiple variables to each other, or detecting outliers or anomalies. Visual-interactive approaches can help to tackle these challenging tasks by closely involving the user in the exploration process, addressing the typically difficult parameter selection problem, which could be more complicated to solve relying on purely automatic methods.

Several works propose to visually analyze multivariate time-dependent data by dimensionality reduction [26, 12]. Multivariate data is visualized as two-dimensional time series paths obtained by dimensionality reduction (projection to 2D). While these works visually compare sections of labeled multivariate time-dependent data, they do not consider exploratory search in unknown data sets. Furthermore, these works do not focus on aggregation efforts to reduce over-plotting problems. To this end, we introduce interactively steerable data aggregation, supporting handling of multivariate time series data. In particular, the user is able to group data points according to data-specific characteristics like statistical calculations based on value and time, or clustering results.

Our approach supports an effective overview of frequent and infrequent *states* in multivariate time series data even in cases of very large data. Furthermore, users can interactively select meaningful path line subsets

for detailed exploration and for visual clutter reduction purposes. Understanding of aggregated data groups is supported by showing a comprehensive cluster glyph metaphor, wherever data aggregation visualization is required within the exploration process. We directly involve the user in the exploration process, combining data exploration with interactive steering of the automatic analysis methods, such as searching for appropriate clustering parameters, in particular.

We demonstrate the usefulness of our approach by an application to earth observation data. There, long time series of many parameters arise, and users want to understand periodicities, trends, and anomalies. We show how our set of interactive views allows for interactively exploring weather patterns of different lengths and parameters. Due to our data aggregations, domain users can explore multivariate weather data in a single display, giving an overview of all data aspects at once.

The remainder of this paper is structured as follows. In Section 2 we discuss related work in several areas. In Section 3 and 4 we motivate our approach, explain our system design and describe user interaction techniques. In Section 5 we apply our implementation to a real-world data set, demonstrating the usefulness of the approach. Finally, we summarize this paper and discuss future extensions in Sections and 6 and 7.
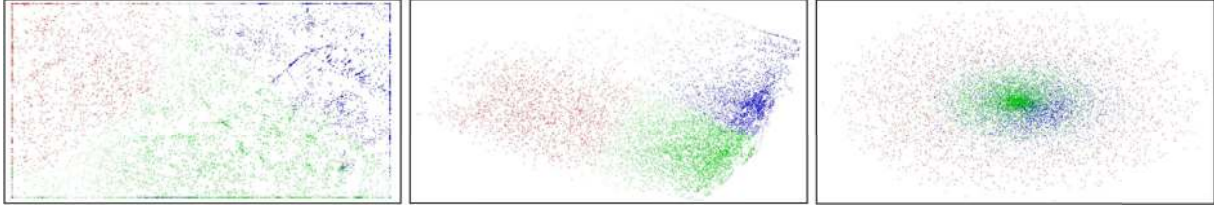
Figure 2: Visual comparison SOM, PCA and MDS projection technique. A k-means clustering result is shown.

2 RELATED WORK

Our work is related to analysis methods for time-dependent data and multivariate data. Time series analysis in general is conducted to understand the behavior of systems, to distinguish regular from extraordinary characteristics [14] and to predict future development [13].

Visualization of Time Series Data

The visualization of time series is often helpful for exploratory analysis. Traditionally, time series can be visualized by line charts [24]. However, using line charts is typically not effective for large time series data, as many and long time series lead to over-plotting if packed into a given display or would require excessive user navigation (cf. the problem specification in Figure 1). The *pixel paradigm* [2] for visualization of time series suggests to map the quantitative values of a time series to an appropriate color scale. Ultimately, each value can be represented by a single pixel. The Recursive Pattern technique [2] employs the pixel paradigm to arrange time series in a generic proximity-preserving way, allowing to arrange between row-by-row up to more complex patterns following space-filling curves. The comparison of many time series can be supported by rendering them next to each other in an appropriate display.

Besides side-by-side schemes, e.g., TreeMap-like layouts have been proposed [10]. An alternative to the pixel paradigm is to map the time axis to a spiral, effectively using more length, which is particularly useful for analysis of periodic data [27]. For domain-specific visualization tasks, e.g., atomistic simulation data, specialized techniques have been proposed [6]. An overview of time series visualization can be found in the textbook by Aigner et al. [1]

Automatic Support

Automatic analysis techniques are often used in time series visualization. E.g., the size of the data to be visualized may be reduced by aggregation [5] or dimensionality reduction [8].

In [25] prototypical time series patterns are found by cluster analysis, and linked to occurrence on the time scale by color-coding. In [17] a discretization approach is applied to support visual analysis of frequent subsequences in a node-link-diagram. Often, the similarity definition between time series or subsequences thereof is important to support exploratory search. In [28] so-called Perception Points of Interest are identified to sort a large number of time series for effective overviewing. Various other systems support the interactive retrieval of time series by defining appropriate similarity notions and query interfaces [9, 11, 3]. A visual-interactive approach to analyzing different families of functions is presented in [16]. Here, the authors allow the user to highlight data patterns of interest and provide linked views of the multidimensional data and the user-selected highlights.

Multivariate Time Series

The above methods mainly consider univariate time series. Yet, multivariate time series analysis is of importance in many domains. A number of approaches include small multiple displays for showing all variables over time next to each other. They may rely on line charts, pixel displays, or any other appropriate base technique. Also, automatic analysis methods for exploratory analysis in multivariate time series have been considered. E.g., in [19] a frequent-pattern-based approach is used to find interesting time series patterns along several levels of abstraction.

Recently, a number of authors have considered the visualization of multivariate time series data based on projection. The basic idea is to project discrete points in time to a 2D display, which in turn allows for analysis of the time series for regularities and irregularities [23]. In [22, 12] multivariate observation measures from motion tracking are projected using the Self-Organizing Map (SOM) method [15]. Individual observations are connected by lines, and glyphs illustrating the particular configurations of the motion are shown. In [18] multivariate time series are extracted from text, by computation of certain text features for discrete intervals along the sequence of the text. A PCA-based display was used to assess the development of the text content, by analysis of feature trajectories observed in the display. In [26] the authors use PCA-based projection to explore the sequence of small fixed-size intervals (so-called *n-grams*) of long univariate time series data. The approach was applied to stock market data and shown to provide an informative overview over long time series
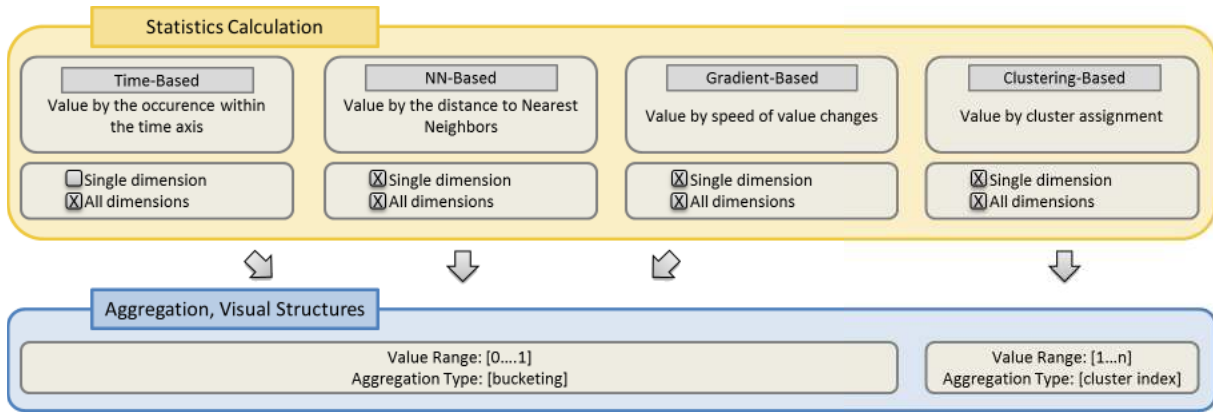
Figure 3: Aggregation of multivaritate time series data based on a variety of statistical data properties. Most of the functionality can either be performed on a single, or all dimensions.

data. In particular, the authors proposed their method to support the following analysis cases: Detection of cyclic behaviors; visual identification of base patterns and outliers; and analysis for trends.

3 SYSTEM DESIGN

In this work we present *TimeSeriesPaths*, a system for the analysis of multivariate time series data. The PCA projection algorithm is applied to arrange multivariate time-series on the (2D) display screen (the *Time Series Path Map*). We connect temporally adjacent data elements and receive a sequentially ordered set of points – a so called *time series path*. By default, such a visualization suffers from severe *over-plotting* and *overview problems*. In order to make such a visualization understandable for domain-experts and to counter the implied challenges, our approach comprises three contributions:

1. We apply semi-automatic data aggregation functionality, either derived from statistical data calculation, or from visual-interactive data clustering (cf. Subsection 3.2). This helps the user to get an *overview* to the dataset.

2. We present a cluster visualization technique that incorporates multiple information about the aggregated data (cf. Subsection 3.2). This supports *data interpretation and cluster comparison* approaches.

3. We propose a multi-view system with broad visual-interactive analysis functionality (cf. Subsection 4). Selection and highlighting modalities of data path subsets counter the challenge of *over-plotting* and allow for comprehensive *detail on demand* perspectives.

3.1 Visualizing Multivariate Time Series

Data Projection

We apply a projection technique to visualize multivariate time series data on 2D displays. An applicability

consideration between visualizations based on projection and the multiple linechart technique is given in Section 6.

A general requirement concerning projection is the preservation of data topology, by means that similar data in the multivariate input space is also arranged close to each other in the display space. Due to their popularity and their diversity in arithmetical manner we chose PCA, SOM and Multidimensional Scaling MDS [7] as promising candidates. After an evaluation of our requirement criteria and a visual comparison in Figure 2, we choose the PCA algorithm as a default for prospective projection needs in the TimeSeriesPaths system. The non-linear MDS proves to be rather unsuitable for our approach, solely because it has troubles in separating k-means clusters. The SOM algorithm suffers in respect to the calculation speed and a major difficult (fully automatic) parametrization. Yet the key benefit of PCA derives from the ability to project data points in a linear manner, by means that the projection results do not lack on local distortions and thus allow for a more straight forwarded interpretation. Furthermore, the visual comparison of the three projection techniques shows a good cluster separation by PCA. We accept that PCA does not exploit the complete display space as well as the SOM projection. However later in this section, we will present our cluster glyph and show how our glyph visualization mitigates this problem.

Visualizing Time Series Paths

The visualization of time series paths is provided by the Time Series Path Map in the center of the display. Based on our data projection method, we connect individual data points by their chronological order to form paths. The projection arranges similar data points close to each other and reflects the data point distances of the multivariate data input space. Accordingly, if path sequences are similar to each other, their possibly close
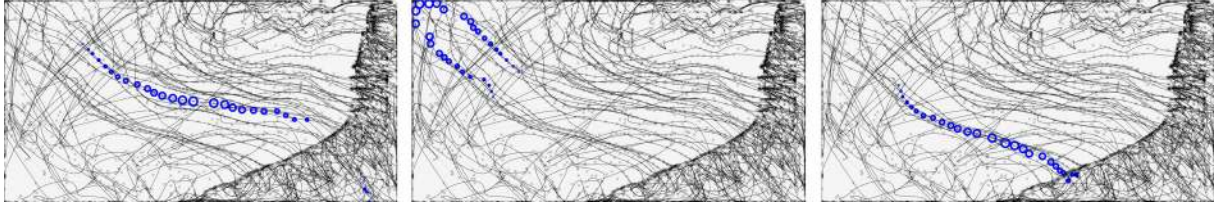
Figure 4: The "Rollercoaster Animation". By dragging the time slider, the user can explore the temporal development of the time series path. The cursor position and time series path neighbors are animated with circular shapes.

positions on the display space help the user with profound analysis approaches.

3.2 Multivariate Time Series Aggregation

Statistics Calculation and Aggregation

We integrate automatically generated statistical data information into the visualization to counter the overview problem and support the analysis process. So far, related approaches color-code data points for time-based and value-based changes or class labeling [12, 26]. Our approach generalizes this by a variety of statistical data measurements that provide additional important information, an overview is given in Figure 3. Altogether, our system provides four different properties of statistical data information for color coding:

- (a) occurrence within the time line (time-based)

- (b) nearest neighbor distance (NN-based)

- (c) speed of value change to adjacent time stamps (gradient-based)

- (d) cluster membership (clustering-based)

Except for (a), all statistical data information can either be calculated on a single dimension or on all dimensions of the data set. Thus, we are also able to perform domain-specific exploration tasks due to the level of detail in the aggregation setup. The number of data groups k can be specified by the user for all properties, (a)-(d).

Data color codings according to group affiliations are displayed on the Time Series Path Map, our time axis display at the bottom (called *Color Time Bar*), and the *Data Aggregation List* on the right, respectively. Showing multiple aspects of the data enables to find an appropriate aggregation level, to interpret groups of data and derive mutual characteristics, to detect outliers and to explore periodic behavior in the data.

In our case study (cf. Section 5), we will show that distributions of aggregated statistical data information on the Time Series Path Map and the Color Time Bar give valuable information about dense data regions, data anomalies and the periodicity of time series paths.

Generic Cluster Glyph

The aggregation of data into groups requires a meaningful cluster visualization method (cf. Figure 5). The main requirement is genericity in order to suit to a great variety of multivariate time series data. Additionally, averages, minima and maxima, variances, number of elements and cluster quality indices are needed. Each data dimension is displayed with an error bar chart glyph metaphor and labeled with the corresponding physical unit. Additionally, we include the distribution of time stamps on a time axis to monitor chronological data characteristics to detect periodic behavior or anomalies. Finally we demand the cluster glyph to show the cluster color coding for linking, and a headline for user-centered data labeling purposes.

Earlier we argued that PCA does not capitalize the entire border areas of the display space. We benefit from this instance due to the fact that we have free space remaining to position cluster glyphs for data aggregation operations. Four black concentrical lines connect the cluster glyph with the appropriate display coordinate without producing too much occlusion (see Figures 1, 6, 7 and 8).
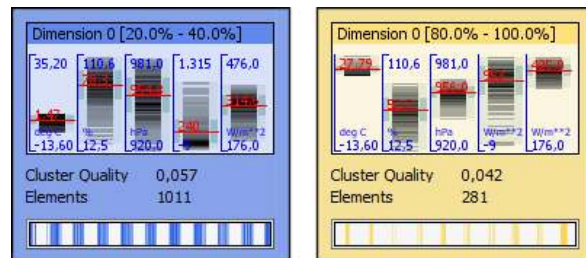

Figure 5: Generic Cluster Glyph. A boxplot-like visualization shows the distribution of data elements in each dimension of the dataset, transparency is used to show data frequency. Cluster centroid values are displayed as red bars, just like gray variance bands mapped laterally for each dimension. Statistical information about the data cluster is shown at the center, the data distribution on the global time axis is shown at the bottom.

4 INTERACTION TECHNIQUES

TimeSeriesPaths includes a set of linked user interaction modalities which work across the three different
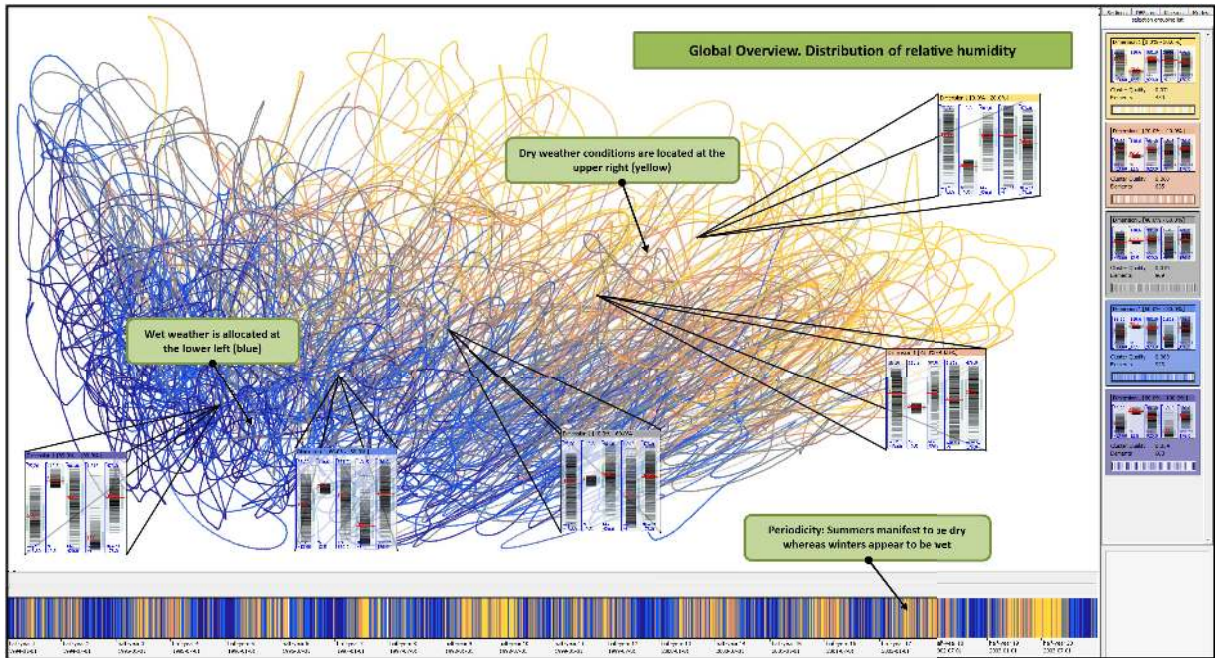
Figure 6: Data aggregation on single input data dimensions: Distribution of relative humidity values (color map: blue means wet, yellow means dry). We constitute rainy weather states to be located left on the Time Series Path Map. By exploring the Color Time Bar, we discover rainy weathers dominating the winter periods. Composing these two findings, we reason that (wet) winter climates are located on the left of the Time Series Path Map.

views. We give a short introduction to the major visual-interactive capabilities of the TimeSeriesPaths system.

Tooltipping

An important user requirement is detail on demand visualization. By hovering above data elements on the Time Series Path Map and the Color Time Bar, tooltips show the multivariate data information and the position of the respective data elements on the time axis (cf. Figures 9 and 10).

Selection, Interactive Grouping and Highlighting

The selection of data is supported in each of our three views. The user can (1) select single data points, (2) time series paths or subsequences thereof, (3) the selection of data within a distinct display region in the Time Series Path Map is possible (cf. Figure 7). The user sketches an arbitrarily polygonal shape on the map, and the surrounded data points will be selected.

Data selections can subsequently be added to the Data Aggregation List for additional information about the selection and for future re-selection. The respective selection is highlighted in all three views to allow the user the detection of interesting patterns. For example, when the user selects a data cluster from the Data Aggregation List (cf. Figures 9 and 8), respective data points are highlighted in the Time Series Path Map and the Color

Time Bar. Thus, the user has three different scopes for the exploration of the selected data: (a) the distribution of the data on the Time Series Path Map, (b) occurrences of data elements along the time line in the Color Time Bar and (c) cluster value distributions in the Data Aggregation List (cf. Figure 8).

By means of transparency and plotting size, the user can counter over-plotting on his own by reducing the visibility of elements that are not selected.

Rollercoaster Animation

The Color Time Bar also contains a *Time Slider* for animated time series analysis. We can drag the Time Slider to a specific point or interval in time, and corresponding subsequences are highlighted with circular shapes in real-time on the Time Series Path Map. A schematical demonstration of our so called "Rollercoaster Animation" is given in Figure 4, an application is shown in Figure 10. This interactive animation allows a detailed exploration of the distribution of projected values over time, and also to detect periodic patterns on the Time Series Path Map. The latter is especially helpful in case of over-plotted displays, where a large amount of data elements is visualized on the display.

5 CASE STUDY

We apply our system to a data set from earth observation research. Based on consultation with domain researchers, we explore weather phenomena hidden in the
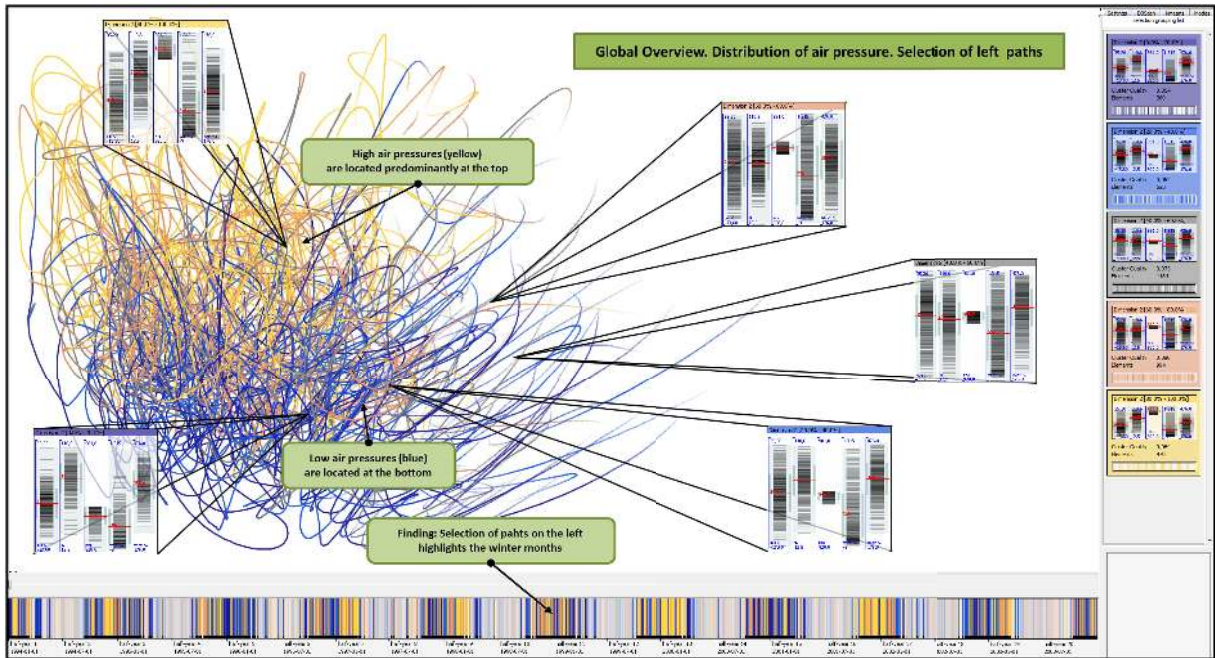
Figure 7: Data aggregation on single input data dimensions: Air pressure development. Selection of the left half of the paths (winter weathers). We discover a color gradient from high (top, yellow) to low (bottom, blue) air pressure values.

data like periodic patterns, frequent 'weather states' and abnormal behavior that can be found with our system.

5.1 Data Set and Application Domain

Our considered data set is acquired from the open data repository PANGAEA [21], operated by the Alfred Wegener Insitute (AWI) for Polar and Marine Research in Bremerhaven. PANGAEA archives and publishes geo-referenced scientific earth observation data in the research areas of water, ice, sediment and atmosphere. Our data set focuses on atmospheric weather measurements, gathered in the scope of the Baseline Surface Radiation Network (BSRN) [20] PANGAEA compartment. These measurements are multivariate atmospheric observations of radiation-related physical parameters) which were recorded every minute. We focus on a dataset of ten years duration, originated from the BSRN station in Payerne (Switzerland) in the time period of January 1st, 1994 to December 31th, 2003 [4]. Payerne is located in the center of the triangle Lyon, Milan and Fribourg at 491 meters above sea level. The climate of Payerne is temperate, semi-continental with average minimum temperatures at about -2°C in January and about 18°C in July. The average daily sunshine duration varies between 2 hours in January and 8 hours in July. Hence, the researchers affirm a yearly climate periodicity to the data that serves as ground truth and primary analysis goal. Beyond that, the so called "summer of the century" in 2003 produced temperature values up to 40°C and

motivates us finding this and yet other anomalies in the data set.

We consulted researchers from BSRN to select a suitable parameter subset for detecting interesting weather scenarios. Besides *temperature*, *relative humidity* and *air pressure*, we incorporate the *short-wave downward radiation (SWD)* and the *long-wave downward radiation (LWD)*. The SWD is well suited to give statements about cloud occurrences. Most radiation is measured at the so called clear-sky condition, even when there are no clouds in the sky. It is used for climate research in general and in applied sciences, e.g., in land surface assimilation models, surface energy budget models, and ocean assimilation models. In agriculture, the short-wave downward radiation is used as an input for crop modeling and the solar industry applies it for estimations where to build solar power plants. The LWD is another important factor in the energetic exchange between atmosphere and the earth surface. While the solar dependent short-wave downward radiation is near zero at night, the long-wave downward radiation can be measured all night long. The long-wave downward radiation is higher when the sky is clear. By applying these five measurements as our data set, we are able to make statements about different weather states that possibly change within a seasonal cycle.

Due to the long time period of ten years, we determine each single day as one data point, periodic behaviors within single days are also discovered in the data set and possible to analyze with our system, but not in the focus in this case study. In order to remain on a uni-
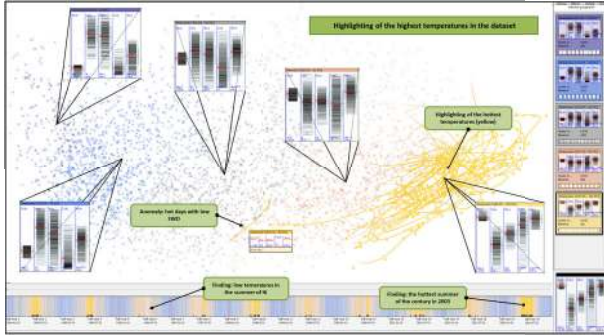
Figure 8: Advanced exploration of Figure 1. We have no problems in identifying the periodic appearance of hot temperatures in the summers in the Color Time Bar. Furthermore, the "summer of the century" anomaly in 2003 stands out with a lot of hot days.

versally accepted approach, we use a generic moving average routine to produce meaningful daily representatives, climate specific daily data aggregation procedures are not applied. Missing value periods of less than 6 hours are linearly interpolated, longer data gaps are ignored. We want to point out that other preprocessing approaches are possible and merely implicate for us the effort of reconfiguring parameters or, if necessary, add a domain-specific preprocessing routine.

5.2 Obtaining a Global Overview

We primarily obtain a global overview to the Time Series Path Map and the data, respectively (cf. Figures 1 and 6). This is crucial due to the described problems in dealing with large multivariate data and projection-based approaches (cf. Section 3). The Color Time Bar indicates a meaningful periodicity with in the seasonal cycle. We constitute Payernes climate to be warm in the summer and cold in the winter period (cf. Figure 1). The overview is completed with Figure 6, where the relative humidity appears to be high (rainy) on the left and low (dry) on the right. At least since the Time Color Bar shows summers to be dry and winters to be wet, we can constitute that the left half of the Time Series Path Map depicts the winter period whereas the summer time is allocated at the right of the display. We prove this hypothesis in Figure 7 by selecting the left half of the time series paths and obtain a meaningful segmentation on the Color Time Bar between summer and winter. Taking the cluster glyphs of the three discussed images into account, we assess correlations between dimension 1 (temperature), 4 (SWD) and 5 (LWD) and thus register another finding in the data set.

After we have received a global overview to the data and our views (some findings may appear evident to the reader so far), we now proceed our case study and focus on the exploration of more particular findings.

5.3 Findings in the Data Set

We now focus on abnormal behavior and anomalies in the data set. We try to discover the "summer of the century" of 2003 as a first finding. We use the view shown in Figure 1 and select the hottest data cluster (yellow); the result is shown in Figure 8. Besides, we discover the coldest summer of the data set in the year 1996 as a new finding. Together with the researchers from AWI, we find our final data exploration goal in the detection of thunderstorms and intense low-pressure systems. Besides the researchers expertise, we consult Internet forums, weather history literature, and insurance damage reports to verify our findings. Figure 9 displays our course of exploration. We focus on the air pressure dimension and apply our *gradient-based* statistical property that measures value changes over time. An aggregation to six clusters produces one group of about 200 highlighted data points that manifest extremely decreasing air pressure gradients. We tooltip a collection of five proven hurricanes and chose the most prominent and devastating hurricane *Lothar* for a detail on demand exploration. Figure 10 details about the air pressure development over 10 days in december 1999. The Rollercoaster Animation helps us navigating through a clearly arranged display, released from visual clutter and overview problems.

6 DISCUSSION

One of the most traditional visualization types for time series data are line charts. In case of multivariate time series, multiple parallel line charts can be used for data visualization. Eventually, projection-techniques such as studied in this paper need to be compared against alternative time series visualization options. While we have not yet done a formal comparison, we provide a conceptual discussion to point at possible advantages and disadvantages of the projection-based approaches vs. line chart approaches.

First, we expect the traditional line chart approach to have scalability problems with respect to very long time series, and to a high number of dimensions. The projection-based approach for the visualization of time series data aims at improving scalability with respect to (1) the time axis (long time series) and (2) the potentially high number of dimensions. Considering (1), information loss occurs for line charts as soon as the number of time stamps becomes larger than the number of available pixels on the x-axsis of the line chart display. Basically, three observations can be made:

1. Drawing multiple data points per pixel coordinate leads to visual artifacts and information loss.

2. Downsampling the number of time stamps reduces the amount of information of the visualization.
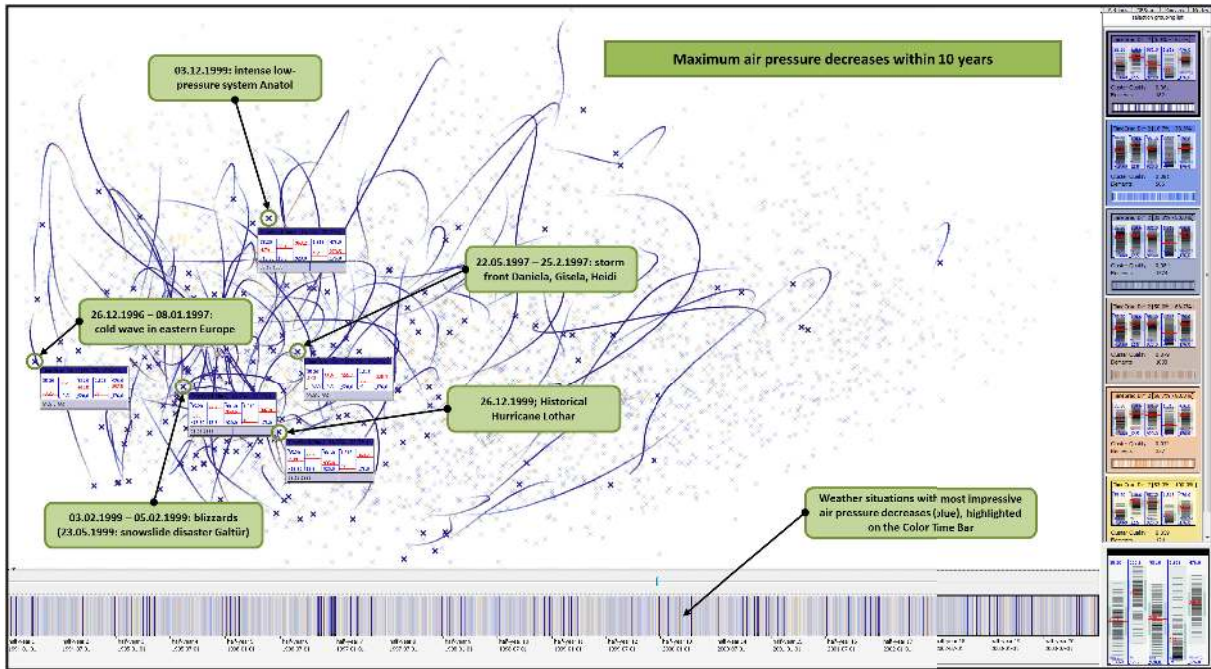
Figure 9: Detection of historic thunderstorms by highlighting most crucial air pressure decreases (blue).

3. Scrolling facilities allow to navigate large time series, yet can be cumbersome for very large time series and may lead to loss of user overview.

Using a 2D projection has the potential to show longer time series in a given display space, if an appropriate projection can be found. On the other hand, interpretation of the projected time series paths may become harder, as directions and distances cannot be read as straightforward as in a line chart.

Considering (2), a high number of dimensions may constitute a problem for multiple line charts. At some point, the available display space is exhausted when too many dimensions are combined in a multiple line chart visualization. In projection, dense data point regions are not only visual clutter. These regions represent dense regions in the input data space and offer potential starting points for further user search and filtering.

The second distinction between multiple line charts and projection concerns the number of data attributes to show. The projection condenses the information of all dimensions in one time series path, providing dimensionality reduction. In general, projection of multivariate data brings up questions about the application-dependent choice of the projection variant (cf. Subsection 3.1) and the preservation of information hidden in the input data. As future work, we need to compare the information preservation of multiple line charts (considering problems for large data or many dimensions) and projection-based time series visualization approaches. One first idea is to define a benchmark data set with periodic behavior that is compared

in multiple line charts and in projection-based visualization. At present, we depict that the first two main components of the PCA-based 2D projection approach preserve 78% of the chosen 5D input data information in our weather data case study. Thus, we may assume that the amount of used information is rather high. Yet, more precise evaluation and comparison of the information contents and usage in parallel line charts and in projection-based approaches is needed.

7 CONCLUSION

We presented a system for the analysis of multivariate time-series data. The identification of relations between multiple attributes is an intrinsically difficult problem, even with a viable projection to 2D-space. In order to make such a visualization understandable for domain-experts, our system provides methods for statistical aggregation and clusterings, which can be steered by the user in a very flexible way. Beyond just showing cluster IDs we propose a new glyph-based visualization. This glyph shows the multivariate profiles of the clusters and allows for an effective comparison and interpretation of their properties. The system provides linked views to relate different perspectives of the data to each other. In cooperation with earth observation researchers, we tested the usefulness of the approach with a dataset for atmospheric weather measurements over a ten-years time frame.

We believe that the approach presented in this paper is easily applicable to time-series of different domains. In future projects we will apply and test this system with consumption data of the electric power grid. We used
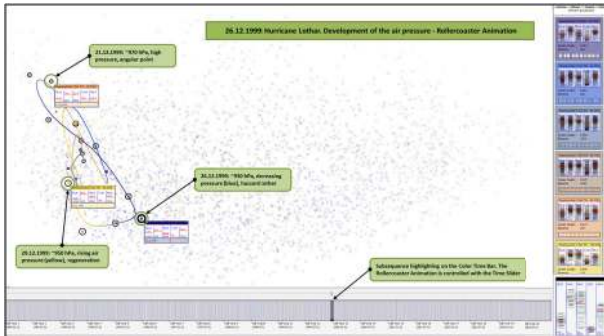
Figure 10: Rollercoaster Animation on hurricane Lothar. Air pressure coloring (blue means low).

projection techniques as an overview because of their popularity as a method for multivariate analysis. However, the methods to calculate, steer and explore the clusters are not restricted to a specific type of overview. In future, we will extend the linked views by other visualizations for multivariate time-series to test for the most effective combination of domain, overview and aggregation methods.

ACKNOWLEDGMENTS

8 REFERENCES

[1] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of Time-Oriented Data*. Springer-Verlag New York Inc, 2011.

[2] Mihael Ankerst, Daniel A. Keim, and Hans-Peter Kriegel. Recursive pattern: A technique for visualizing very large amounts of data. In *Proceedings of Visualization '95, Atlanta, GA*, pages 279–286, 1995.

[3] Jürgen Bernard, Jan Brase, Dieter W. Fellner, Oliver Koepler, Jörn Kohlhammer, Tobias Ruppert, Tobias Schreck, and Irina Sens. A visual digital library approach for time-oriented scientific primary data. In *Proc. European Conference on Digital Libraries*, pages 352–363, 2010.

[4] Jürgen Bernard, Nils Wilhelm, Maximilian Scherer, Thorsten May, and Tobias Schreck. Reference list of 120 datasets from time series station payerne used for exploratory search. doi:10.1594/pangaea.783598, 2012.

[5] Lior Berry and Tamara Munzner. Binx: Dynamic exploration of time series datasets across aggregation levels. In *Proc. IEEE Symposium on Information Visualization*, 2004.

[6] D. Bhattarai and B.B. Karki. Visualization of atomistic simulation data for spatio-temporal information. In *The 14th Int'l. Conf. on Central Europe in Computer Graphics, Visualization and Computer Vision (WSCG'06)*, 2006.

[7] Trevor F. Cox and M.A.A. Cox. *Multidimensional Scaling, Second Edition*. Chapman and Hall/CRC, 2 edition, 2000.

[8] Tim Dwyer and David R. Gallagher. Visualising changes in fund manager holdings in two and a half-dimensions. *Information Visualization*, 3:227–244, December 2004.

[9] Ming C. Hao, Umeshwar Dayal, Daniel A. Keim, Dominik Morent, and Jörn Schneidewind. Intelligent visual analytics queries. In *IEEE Symposium on Visual Analytics Science and Technology*, pages 91–98, 2007.

[10] Ming C. Hao, Umeshwar Dayal, Daniel A. Keim, and Tobias Schreck. Importance driven visualization layouts for large time-series data. In *Proc. IEEE Symposium on Information Visualization*. IEEE Computer Society, 2005.

[11] Harry Hochheiser and Ben Shneiderman. Dynamic query tools for time series data sets: Timebox widgets for interactive exploration. *Information Visualization*, 3(1):1–18, 2004.

[12] Yueqi Hu, Shuangyuan Wu, Shihong Xia, Jinghua Fu, and Wei Chen 0001. Motion track: Visualizing variations of human motion data. In *PacificVis*, pages 153–160, 2010.

[13] N.K. Kasabov and Q. Song. Denfis: Dynamic evolving neural-fuzzy inference system and its application for time-series prediction. *Fuzzy Systems, IEEE Transactions on*, 2002.

[14] E. Keogh, J. Lin, and A. Fu. Hot sax: Efficiently finding the most unusual time series subsequence. In *Data Mining, Fifth IEEE International Conference on*, pages 226 – 233. Ieee, 2005.

[15] Teuvo Kohonen. *Self-Organizing Maps*. Springer, Berlin, 3rd edition, 2001.

[16] Zoltan Konyha, Kresimir Matkovic, Denis Gracanin, Mario Jelovic, and Helwig Hauser. Interactive visual analysis of families of function graphs. *IEEE Transactions on Visualization and Computer Graphics*, 12(6):1373–1385, November 2006.

[17] J. Lin, E. Keogh, S. Lonardi, J.P. Lankford, and D.M. Nystrom. VizTree: a tool for visually mining and monitoring massive time series databases. In *Proc. of the int. conf. on Very Large Data Bases*, pages 1269–1272. VLDB Endowment, 2004.

[18] Yi Mao, Joshua Dillon, and Guy Lebanon. Sequential document visualization. *IEEE Transactions on Visualization and Computer Graphics*, 13:1208–1215, 2007.

[19] Fabian Mörchen and Alfred Ultsch. Efficient mining of understandable patterns from multivariate interval time series. *Data Min. Knowl. Discov.*, 15(2):181–215, 2007.

[20] A. Ohmura, E. G. Dutton, B. Forgan, C. Fröhlich, H. Gilgen, H. Hegner, A. Heimo, G. König-Langlo, B. mcarthur, G. Müller, R. Philipona, R. Pinker, C. H. Whitlock, K. Dehne, and M. Wild. Baseline surface radiation network (BSRN/WCRP): New precision radiometry for climate research. *Bull. Amer. Met. Soc.*, 79:2115–2136, 1998.

[21] PANGAEA - Data Publisher for Earth and Environmental Science. http://www.pangaea.de/. Last accessed on April 5, 2012.

[22] Y. Sakamoto, S. Kuriyama, and T. Kaneko. Motion map: image-based retrieval and segmentation of motion data. In *Proc. 2004 ACM SIGGRAPH/Eurographics symposium on computer animation*. Eurographics Association, 2004.

[23] Geoffroy Simon, Amaury Lendasse, Marie Cottrell, and Université Paris. Long-term time series forecasting using self-organizing maps: the double vector quantization method, 2003.

[24] Edward R. Tufte. *The visual display of quantitative information*. Graphics Press, Cheshire, CT, USA, 1986.

[25] Jarke J. Van Wijk and Edward R. Van Selow. Cluster and calendar based visualization of time series data. In *Proc. IEEE Symposium on Information Visualization*, pages 4–9. IEEE Computer Society, 1999.

[26] Matthew O. Ward and Zhenyu Guo. Visual exploration of time-series data with shape space projections. *Eurographics / IEEE Symposium on Visualization (EuroVis)*, 30(3), 2011.

[27] M. Weber, M. Alexa, and W. Müller. Visualizing time-series on spirals. In *proceedings of the IEEE Symposium on Information Visualization*, pages 7 – 13, 2001.

[28] H. Ziegler, M. Jenny, T. Gruse, and D.A. Keim. Visual market sector analysis for financial time series data. In *Visual Analytics Science and Technology (VAST), IEEE Symposium on*, pages 83–90, 2010.