Survey: Visual Analysis Approaches to Time Series Prediction

Fabian Otto*

Technische Universität Darmstadt

1 Introduction

Making predictions about the future is a common problem in corporate scenarios as well as in everyones personal life. Corporations need predictions in order to determine e.g. the following day's demand on the market or which new product has the highest Return on Investment. This is also true for the personal live, when buying a new laptop or signing a contract. These decisions are easier if there is at least some certainty about future events.
According to Lu et al. [21] predictive analytics is concerned with the prediction of future outcomes and trends based past observations. It comprises an approach that includes pattern recognition, extrapolation and algorithmic modeling associated with domain knowledge. Contrary to that, descriptive or prescriptive systems are reactive and provide information after the decision was made or populate decision models respectively. Consequently, predictive analysis can be seen in between these two in the way of applying historical data to generate knowledge about the future. Time series are often highly complex, depend on different variables and cannot be easily predicted. Common models to analyze time series are often based on the Box-Jenkins method [7] or on regression [11]. These automatic approaches provide analysts with forecasts. However, they are often not able to provide analysts with an understanding of interesting phenomena. This means analysts cannot easily include their domain knowledge in the modeling process in order to create better predictive models. In combination with visual analytics, analysts have an interactive solution, which enables them to answer a wider range of research questions, such as:

(Q1) What are global trends within the time series?

(Q2) Which local patterns, seasonal trends and important events/periods can be found?

(Q3) Which model is the best to represent the time series?

(Q4) How can this information be used to detect anomalies and turning points, which may change the future direction of the time series?

(Q5) Which correlations between variables can be found?

(Q6) Which parameters are influencing my production process?

(Q7) How certain is my current prediction?

This survey will mainly focus on questions (Q1) and (Q2) in Sect. 2.1, on (Q3) in Sect. 2.2 and (Q4), (Q5), (Q6) in Sect. 2.3. (Q7) is addressed throughout all sections. Nevertheless, time series prediction is an ambiguous field and some overlap might occur.
Moreover, in some cases additional information about the location are available and have to be included in the pattern detection and prediction process. In spatiotemporal analysis, analysts search for regions with unusually high occurrences of events, also called hotspots. For cases where these hotspots are found, analysts would like to predict how these regions will develop and where new hotspots may occur in order to come up with better decisions. Prominent applications areas involve: detecting outbreaks of diseases as well as crime and terror developments.
Hence, the goal is to provide the reader with an entry point to select an appropriate approach according to her/his currently required area of application.

2 Abstract Time Series

Abstract time series only utilize the time dimension in order to predict future events or find patterns. This includes

2.1 Trend Detection

The most common problem in time series prediction is trend detection. This involves finding a overall global trend or a more specific local trend, e.g. seasonal changes in retail due to Christmas. In regression analysis this can expressed as finding a predictive function

$$y(\mathbf{x},\mathbf{w}) = \mathbf{w}_0 + \sum_{j=0}^{M-1} \mathbf{w}_j \phi_j(\mathbf{x})$$

where M represents the models parameters and $\phi(\cdot)$ a feature function to model the desired behavior. Unfortunately, these kind of models hardly allow the analyst to include domain knowledge in the prediction process.

An older Visual Analytics approaches can be found from Ichikawa et al. [17]. Their goal was to enable stock analysts to predict multiple daytime stock prices and simultaneously visualize a set of predictions from different simulation systems. This means their system mainly supports result exploration by efficiently visualizing a vast amount of predictions. As a consequence, the visualization includes multiple predictions for multiple stocks. Further, one major finding was, visualizing multivariate predictions in a 3-dimensional space creates high levels of occlusion, thus, it is not suitable to provide an easy and intuitive way of visualization. Instead, the system utilizes line charts with cluttering control and a color band display with level-of-detail control (Fig. 1).

The color band display reduces the complexity of the predictive time series as large amounts of simulations typically create occlusions within the line plot. This issue can also be found in TimeSearcher [9], which displays the mean instead. The color band is created by assigning similar predictions to the same cluster and consequently reducing the time series to a color band, i.e. its key elements and overall trend of each cluster. Therefore, the analyst is able to detect discrepancies between the clusters and compare specific predictions with the overall trend. For additional comparison, the system's workspace visualizes a set of predictions for different parameter ranges (e.g. sales organizations) as well as different stocks. Whereas the different parameter ranges could also support the analyst in answering (Q3). Consequently, the user can detect trends concerning the whole stock market and answer (Q1). However, due to the amount of simulations displayed, it might be hard to extract specific information. The color band display lowers the complexity significantly, hence this system is not able to answer (Q2) properly. Additionally, the system is not visualizing any information about the

*e-mail: fabian.otto@stud.tu-darmstadt.de

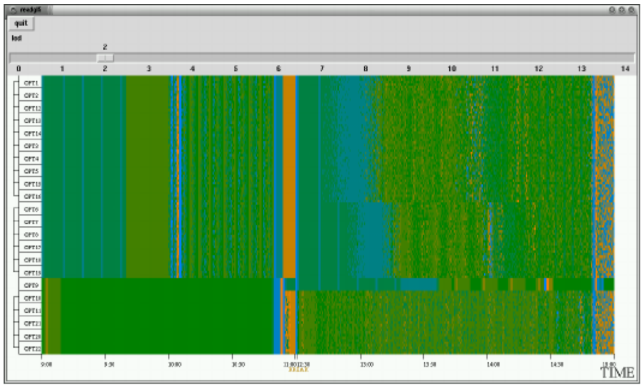

Figure 1: color band display [17]

simulations' certainty and makes it hard for an analyst to determine, which simulations are more important and fails to answer (Q7).

In contrast to Ichikawa [17] the system of Hao et al. [13, 14] explicitly focuses on time series prediction with peak preservation and is able to provide the analyst with better answers to (Q2). Identified peaks are explicitly included in the systems forecast. In their work they focus on cell based power consumption data in data centers, for which it is especially important to have deeper knowledge of peaks. They applied an automated peak preserving smoothing method in order to reduce noise and get a more reliable prediction as well as retain the seasonality of the data. Thereby, the analyst is able to determine the influence of more recent measurements, i.e. how far back in time seasonality is considered in the smoothing process. Further, the system provides a visualization of its prediction quality on the historical time series, which allows the analyst to judge the quality of the current model (Q7). It has to be mentioned here, that including a peak preservation is sometimes not wanted. For example when predicting the market's demand, the global trend is more important and including peaks may induce more uncertainty. A large drawback of those systems is their univariate focus, therefore they may be insufficient in a lot of practical applications.

Visualizing uncertainty was one of the analyst questions Ichikawa's [17] system was not able to answer. A popular choice to address this problem is based on ensemble visualizations, which were only partially applied, however the visualization was restricted to the color band display and the analyst was not presented with the certainty based on multiple predictions. The system from Köpp et al. [20] provides a heatmap-based option to visualize multiple time series predictions. Their interactive solution offers analysts an comprehensive analysis of the predictions certainties with e.g. quantiles, extrema and percentiles. Further, similar to Ichikawa, the analyst is able to detect diverging trends within the given ensembles. Another resembling feature to Ichikawa is the external source of simulations. Depending on the application area, this can be beneficial or disadvantages. Analysts can used their existing ensemble and visualize it, but it is more complicated to begin with if no such analysis environment is present. Other ensemble approaches incorporate models explicitly and help in general to answer (Q7), but this is beyond the scope of the survey and the reader is encouraged to use this work as a entry point for further research.

During the evaluation of TiMoVa [6], (further details about this tool in Sect. 2.2) they found that an actual prediction functionality would provide additional value for the analyst during the model selection. Therefor, Böegel et al. [5] included a result exploration and validation functionality. This allows the analyst to adjust different model parameters and see real time changes in the corresponding prediction visualizations. Consequently, the analyst gets feedback about the adequateness of the chosen model and can simultaneously use the provided visualizations after the selection process to search for trends to answer (Q1) and (Q2). The models certainty is incorporated with displaying confidence bands. Additionally, they specifically visualize the difference of true and predicted values for each data point as well as the direction (positive or negative). This gives the analyst a quick overview if the model is constantly over- or underestimating the time series as well as how long and often this occurs and is not provided by any of the other models.

2.2 Model Selection

2.3 Correlation Detection

The idea of peak preservation was inherited in an Motif detection approach [15], where the peak preservation is combined with an automated motif/pattern detection, which includes detecting overlapping patterns and extrapolating the patterns into the future. This later version, further addresses the univariate issue and supports multivariate analysis. By adding these additional features to the previously simple peak preserving prediction, the analysts can also answer (Q4), (Q5) and (Q6). A possible application presented in the paper is forecasting the ideal oil well flow pattern as well as analyzing how to recover from drops in flow (outages).

2.4 Modeling

An popular approach for general time series analysis is from Buono et al. [8] and focuses explicitly on automatically finding similar patterns compared to a user specified pattern. The system was built upon TimeSearcher proposed by Hochheiser and Shneiderman [16], which concentrates on high usability even for users without specialized skills such as in statistics. However, these two version of TimeSearcher are more interested in data exploration and consequently would be located either in the data preprocessing or as supporting tool in the feature engineering step. For modeling the updated version of TimeSearcher by Buono et al. [9] is of greater interest. For its prediction, it resorts on the similarity-based approach, which was used in the previous version [8] to detect time series with similar behavior. In order to provide the analyst with an better overall understanding of the selected subset of time series, the system offers a summarized view based on a river plot, which also incorporates confidence bands (bottom right of Fig. 2). The actual prediction is computed by extrapolating only those time series, which were identified as similar to the target time series. In order enable the analyst to select a better model, the simultaneous preview interface allows to compare multiple parameter choices as well as different modeling techniques, such as smoothing, in parallel (Fig. 2). Given the data-driven nature of the system, this requires the analyst to provide larger datasets compared to a model-driven approach. However, the simultaneous preview interface simplifies the modeling process and makes it accessible to untrained users and allows experienced analysts to get a better insight.

In practical scenarios the requirement to have large amounts of data points can often be a problem, which makes the systems [8,9,16] less valuable. Consequently, a model-driven system, which requires less data points, is preferred. The approach of [17] could be seen as such system as it utilizes external simulations models. However, it is not supporting the modeling process and is therefore only seen as system for result exploration. On the other hand, TiMoVa a system, which explicitly provides model parameter selection was proposed from Bögel et al. [6]. The system is designed after the Box Jenkins method and is only supporting the model specification and selection process. For model specification, i.e. selection of an appropriate model type and its order, they provide autocorrelation function and partial autocorrelation function plots. These plots are also utilized by the analyst to select the model's parameters. A big drawback of the system are the assumptions about the preprocessed data. It requires a time series without missing values and only supports univariate analysis.

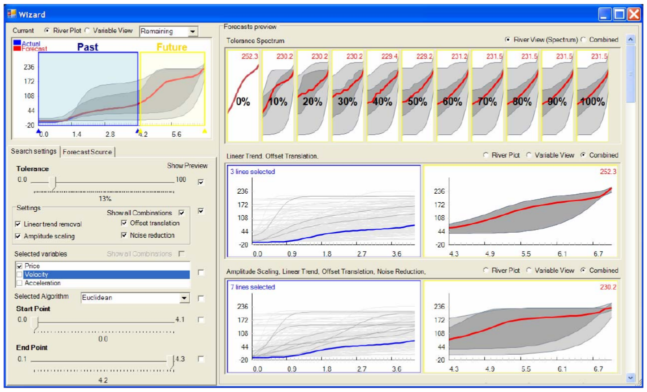

Figure 2: TimeSearcher's [9] simultaneous preview interface

2.5 Result Exploration and Validation

Subsequent to the creation of the model, the analyst wants to explore the results/predictions and eventually compare the performance between several models. Therefore, system are often highly interactive to provide the analyst as much freedom as possible. [21]
An older Visual Analytics approaches is from Ichikawa et al. [17]. They wanted to enable analysts to predict multiple daytime stock prices and simultaneously visualize a set of predictions from different simulation systems. This means, unlike other systems, their system only supports result exploration and/or model selection by efficiently visualizing a vast amount of predictions. As a consequence, the visualization includes multiple predictions for multiple stocks. One major finding was, visualizing multivariate predictions in a 3-dimensional space creates high levels of occlusion, thus, it is not suitable. Instead, the system utilizes line charts with cluttering control and color charts with level-of-detail control. The color chart tries to reduce the complexity of all simulations by clustering similar predictions and reducing the time series to its key elements, i.e. the overall trends of each cluster. This allows the analyst to detect discrepancies between the clusters. Further, he is able to compare specific predictions with the overall trend. For additional comparison, the systems workspace visualizes a set of predictions for different parameter ranges (e.g. sales organizations) as well as different stocks. Consequently, the user can detect trends concerning the whole stock market. However, due to the amount of simulations displayed, it can be hard to extract specific information.

During the evaluation of TiMoVa [6], they found that an actual prediction functionality would provide additional value for the analyst. Therefor, they included a result exploration and validation functionality. This allows the analyst to adjust different model parameters and see the changes in the corresponding prediction visualizations, which gives him feedback about the adequateness about the chosen model. Similar to [9], the prediction is incorporated with confidence bands to express the models certainty. Additionally, they specifically visualize the difference of true and predicted values for each data point as well as the direction (positive or negative). This gives the analyst a quick overview if the model is constantly over- or underestimating the time series as well as how long and often this occurs.

A more specialized data-driven visual analytics approach was proposed from Xie et al. [30]. The difference to other approaches can be found in the application area of customer-to-customer e-transaction time series. The time series consists of transactions between the a seller and a buyer. However, the commodities as well as the buyer can vary greatly. Finding patterns in such time series helps analysts to understand temporal and contextual connections between multiple transactions of one seller, e.g. high number of transactions with same seller/ with same commodity. This can be extended to gain an understanding of the overall selling process or identifying fake transactions. For analysis the system employs a iterative process consisting of two main visual analytics components, an overview component and a detail component. The overview component proposes possible salient transactions. Therefor, a decision tree learner calculates a saliency score for each transaction and a certain analysis task. This learner addresses basic features (e.g. commodity, order amount), textual features (e.g. sensitive words in comments) and temporal features (e.g. transaction amount of seller within on time period). The latter is difficult to address by a decision tree, hence the systems harnesses the transaction frequency of a seller in a user defined time interval. Consequently, the saliency scores are displayed in a time-of-saliency map with an user adjustable level-of-detail. The detailed view is used to gain more insight on specific transactions, which were selected by the user in the overview. For visualization, they introduce a musical notation inspired visual metaphor, called KnotLines (Fig. 3). The amount of transactions in one section (e.g. books) defines the size of the corresponding note head, where missing values are explicitly marked to catch the attention of an analyst. For each time interval the sections' heads are placed on the same note stem, where the length represents the total payment per seller and per time interval. In order to capture the relationship over multiple time intervals, the stems of one seller are connected by beams. This view enables the analyst to observe multiple attributes as well as temporal and contextual correlations of transactions and find salient transactions. The identified salient transactions are fed back to decision tree learner, which changes the predictions of the unknown transactions. A big drawback of the initial computation of saliency values is the need for training data. Analysts are required to annotate features of training transactions, unlike most other approaches introduced in this survey, which do not require this additional step.

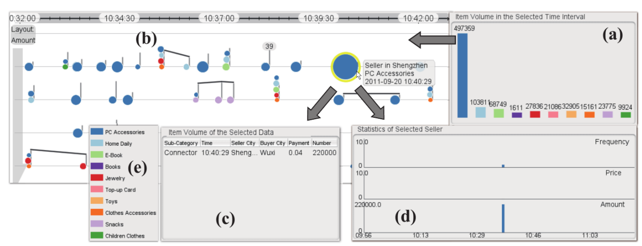

Figure 3: KnotLine view from [30]. (a) Bar chart for the number of commodities in different categories. (b) The big knot indicates an unusually large number of commodities in a transaction. (c) Detailed information. (d) Statistical information. (e) Sales category legend.

Similar to [30], another specialized system was recently presented by Steed et al. [28]. Their approach is focused on understanding patterns in log and imagery data collected by 3D printers, which is highly irregular, includes missing values and has a high complexity in general. Unlike the other systems, this visual analytics tool is designed from an manufacturing standpoint to discover patterns related to defects and system performance issues, optimizations to avoid defects and increasing production efficiency. The visualization of the system is based on the visual information seeking strategy. Therefor, the system provides different charts for each variable with individual level-of-detail control and filtering including an interactive statistical view (Fig. 4). They also provide a new visualization technique, called waterfall visualization, to combine overview and detailed view, which is similar to the idea of color charts by [17] Further, a comparative view is provided to analyze different variables of multiple configurations, which also includes pattern matching capabilities. With the application of information scent, the system encodes quantitative values from similarity and statistical methods in the visualization in order to highlight relationships and reduce the search space. From an operational point of view, the system offers, similar to [30], user driven analysis and helps to detect and highlight univariate and multivariate patterns from different angles. Technologically, the system is bin based, each bin contains a small subset of data points and the corresponding descriptive statistics. Through the aggregation of bins, a higher level-of-detail abstraction can be achieved. The above parts of the system can be applies well to other problem ares, however they also included visualizations specifically for the analysis of 3D printer data. One of those is a segmented time series view, which partitions a selected time series of a variable and also visualizes the images of the printing process next to it. By providing a reference time series, the system computes the similarity/dissimilarity of both series. Another feature specifically for 3D printers is the segmentation based on the build height as well as porosity detection. Whereas, these specific features provide important information for this use case, they make the system less generic and applicable to other areas. Further, the system currently does not support provenance information and the similarity/dissimilarity methods are limited to the segmented time series view.

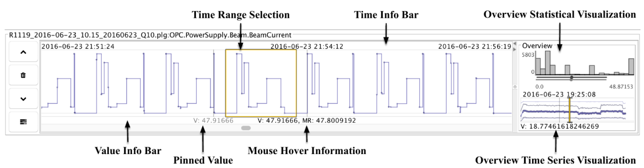

Figure 4: Time series visualization of one variable from [28]

Other application specific approaches involve pattern detection within patient treatment plans [12] or hypothesis generation for climate research [19]. Whereas, the latter can also be seen as spatiotemporal as it tries to detect regions in the atmosphere, which indicate climate change. Similar to [30], Janetzko et al. [18] focuses specifically on anomaly detection. However, the underlying time series is based on server power consumption. In the same application context is the system of McLachlan et al. [27]. Their analytics tool was developed for system management and is also viable for large time series, i.e. hundreds of parameters across thousands of network devices. Other multivariate systems, such as [8, 9, 17], are often not able to present these amounts clearly. The application of [28] is one of the few to provide similar capacity, although a different application area. Time series prediction with peak preservation poses a similar problem as pattern and anomaly detection. The identified patterns/peaks need to be included in the systems forecast. An systems which also includes this additional step in the prediction process is from Hao et al. [13, 14]. They applied an automated peak preserving smoothing method in order to reduce noise and get a more reliable prediction as well as retain the seasonality of the data. Thereby, the analyst is able to determine the influence of more recent measurements, i.e. how far back in time seasonality is considered in the smoothing process. Otherwise, the system is comparable to [9]. This idea is inherited in [15], where the peak preservation is combined with an automated motif/pattern detection, which includes detecting overlapping patterns. Similar to [9], depending on the user's preference, the detected patterns can be aggregated to increase their significance.

3 Spatial Time Series

The previous section (Sect. 2) presented an overview of systems for temporal time series analysis. However, in other application areas such as crime prevention as well as emergency and epidemic intelligence, not only the temporal information is valuable, but also the spatial information. Typically, spatial information fits into a hierarchical categorization structure, which can be filtered. Further, the data categories are processed either as aggregated time series over a spatial location (e.g. county, zip code, collection station) or represent a spatial snapshot of a small time aggregate (e.g., day, week). [2, 3] also found that geospatial analysis has a higher complexity and automated methods cannot adequately solve this task. It requires the knowledge an human analyst in order to solve these problems comprehensively. However, due to the high dimensionality of data, an analyst needs support from computational/visual analytics systems.

An system, which can be seen between temporal and spatiotemporal is from Andrienko et al. [4]. They can use spatial dimensions, but put their focus on it compared to the following approaches. Similar to [13–15], they utilize a method for peak preservation in order ensure local specialties are represented in the prediction.

Maciejewski et al. [22], on the other hand, focuses on spatiotemporal prediction with a model-driven approach. Thereby, they are proceeding from their previous work [24, 25] and center around categorical geospatiotemporal event data, where events consist of locations in time and/or space and each event fits into a hierarchical categorization structure. Specifically, they used a data set for detecting adverse health events using pre-diagnosis information from emergency departments. The system itself provides a line chart with certainty bands and a colored geospatial window, which shows the percentage of events in a certain area, e.g. patients at an emergency department, which where classified with respiratory syndromes (Fig. 5). Further, the user is able to apply filtering on a fine and coarse-grained level. The systems also differentiates between the time series and the geospatial prediction. The time series prediction is achieved by cumulative summation or a Seasonal Trend decomposition based on locally weighted regression. For multivariate data, each event category is modeled as a separate time series signal. Equivalent to the time series prediction, the granularity/level of aggregation (e.g. state, county, etc.) for geospatial predictions can be adjusted by the user. Further, the system utilizes a kernel density estimation, which allows to display the spatiotemporal distribution on a fine-grained level. In order to detect anomalies, e.g. outbreaks, (also see **??**) the system calculates the difference between the predicted and the actual values and highlights areas above a user specified threshold (yellow diamonds in Fig. 5).

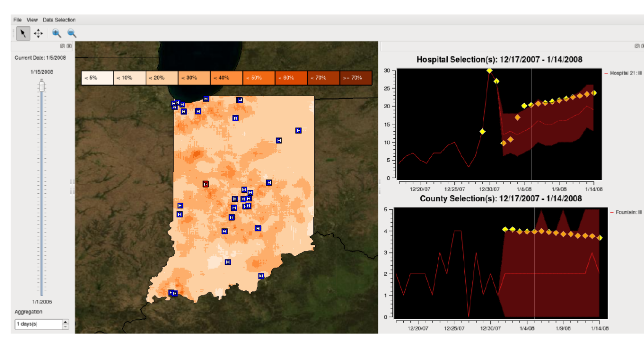

Figure 5: Interactive visual analytics environment from [22]. This is shows an analysis of respiratory syndrome counts in Indiana with county level aggregation. The yellow diamonds indicate temporal alerts, the white line the current day and the transparent polygon the prediction's confidence.

Another system also from Maciejewski et al. [23] follows a similar approach. They target multivariate data with high signal to noise ratio and a degree of uncertainty. Equivalent to [22] the system provides a linked environment of geospatial data and time series graphs and allows users to filter data. Additionally, the system also uses cumulated summation and Kernel Density Estimation. The major difference is, this work focuses on finding and understanding the pattern, rather than only predicting them. Therefor, the system establishes temporal contour maps, which are overlaid contour maps over a period of time. This allows the user to view shifting hotspots across time and analyze the movements of trends and patterns over this period. Further, the system allows to search for correlations between multiple variables via overlaying contour maps, heat maps and/or including height. However, one issue with this visualization is, it only works with three different variables and cannot help to find larger correlations. Moreover, aggregating too many data points, may yield to largely exaggerated hotspots or a uniform surface, through too many hotspots.

Recent work from Malik [26] focuses, like [8, 9, 16, 30], on a data-driven visual analytics approach that provides domain experts (i.e. non statistic experts) a proactive and predictive environment, which allows them to utilize their domain expertise. Similar to [22] the system applies seasonal trend decomposition based on locally weighted regression and Kernel Density Estimation for its predictions. One issue they identified in [22] was that domain experts need additional guidance in order to improve their analysis. This motivation is equivalent to the idea behind TimeSearcher [8,9,16] in the temporal analysis. Hence, they provide geospatial and temporal scale templates, this presents the users with a starting point and avoids searching over the complete parameter space. For the geospatial templates, the system separates the space in subregions and filters for regions, which show a high predicted activity and provide sufficient data. Further, the system allows the user to interactively change the initial template to include e.g. police beats and avoid zero counts with no predictive statistical value. In order to compensate for small amounts of data points in some regions, the system can make use of demographically similar neighborhoods within a certain radius by averaging their prediction. Temporal templates follow the idea of peak preservation as trends can occur on different scales of time. Therefore, the system provides a clock view to highlight high activity on a hourly granularity. Moreover, it enables the user to filter on daily and monthly basis to detect such patterns.

In contrast to the previous approaches Andrienko et al. [2] did not use Seasonal Trend decomposition and kernel density estimation, but applied self-organizing maps (*SOM*) for spatial and temporal prediction. SOMs can be seen as a combination of clustering and dimensionality reduction based on the similarity of space and time. The SOM method is applied to spatial situations that occur in different time units and to local temporal variations that occur in different places. This idea can also be found in an earlier approach from Andrienko [1]. The visualization of the more recent approach is implemented with feature images and index images (Fig. 6 top and bottom of each cell). Feature images do not represent detailed information, they solely display the relative magnitude of the attribute, i.e. different attributes can have different scales, therefore they cannot be used for attribute comparison and only give an overview. The implementation is done by either maps and multi-attribute mosaics for spatial data or temporal mosaics for temporal data (Fig. 6 left and right). Temporal mosaics can be seen as similar to the color charts of [17] and the waterfall visualization of [28]. Index images show the temporal or spatial positions of the objects included in the SOM matrix cells. More precisely, a black cell or territory compartment is contributing to the visualization of the above feature image. In order to detect pattern, multiple cells including different feature images and index images for either spatial or temporal data are displayed in a matrix. Additionally, a second visualization aggregates the index images to display the respective other dimension. This allows the analyst to determine at which time a certain spatial cell is present or at which location a certain temporal cell occurs. The evaluation of the system shows that it is applicable for detecting expected patterns as well as unexpected patterns. This allows analysts to either find evidence, which supports their previous formed hypothesis or discover new patterns, which give them a better understanding of their data.

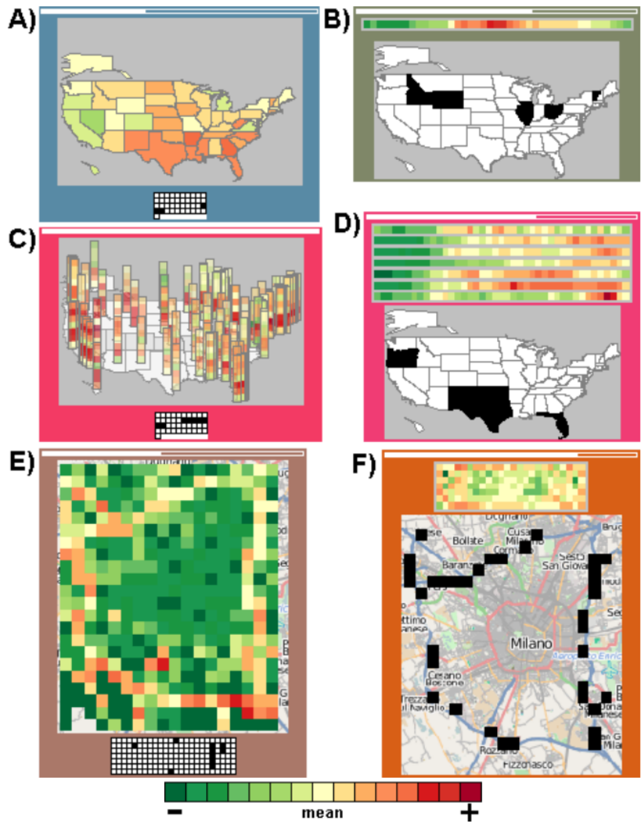

Figure 6: SOM approach from [2] Left column: grouping of spatial situations. Right: grouping of places according to temporal variations of attribute values. A,B: single attribute. C,D: multivariate data with 7 attributes. E,F: one attribute with higher temporal and spatial detail. The upper image in each cell is the feature image, the lower im- age is the index image.

Aside from medical and crime data sets, spatiotemporal data can also be helpful in other domains. Similar to the first approaches, seasonal trend decomposition can be used in context of social media anomaly detection [10, 29]. This can help to increase situational awareness of local events as well as provide insight for investigations and incidents, their severity and consequences.

4 Summary

For time series prediction we can separate two major directions: temporal time series and spatiotemporal. Temporal time series analysis mainly follows data-driven interactive approaches, which include the user in the pattern finding process by providing supportive features such as the simultaneous preview interface [9], the segmented time series view [28] or the KnotLines [30]. They are all enabling the analyst to better understand the data, generate hypothesis and predict future developments or anomaly occurances. Spatiotemporal time series add another important dimension, which makes automated analysis more complicated. Visual analytics systems approach this problem, unlike exclusive temporal time series, often with a model-driven approach. Further, the systems have to enable the user to select spatial and temporal resolution. The system of [26] even suggests different templates to makes the systems more accessible to domain experts.

One important drawback all systems, besides [27, 28], have is often the missing scalability to large amount of parameters and different time series at once. Although, this can be seen as common problem for multi-product companies, customer-to-customer online platforms, etc. On the other hand the current visual analytics system offer a broad range of application areas including manufacturing, medical, environmental and social media oriented.

References

[1] G. Andrienko and N. Andrienko. Visual exploration of the spatial distribution of temporal behaviors. In *Ninth International Conference on Information Visualisation (IV'05)*, pp. 799–806, July 2005. doi: 10.1109/IV.2005.135

[2] G. Andrienko, N. Andrienko, S. Bremm, T. Schreck, T. von Landesberger, P. Bak, and D. Keim. Space-in-time and time-in-space self-organizing maps for exploring spatiotemporal patterns. In *Proceedings of the 12th Eurographics / IEEE - VGTC Conference on Visualization*, EuroVis'10, pp. 913–922. The Eurographs Association and John Wiley and Sons, Ltd., Chichester, UK, 2010. doi: 10.1111/j.1467-8659.2009.01664.x

[3] G. Andrienko, N. Andrienko, J. Dykes, S. I. Fabrikant, and M. Wachowicz. Geovisualization of dynamics, movement and change: Key issues and developing approaches in visualization research. *Information Visualization*, 7(3):173–180, June 2008. doi: 10.1057/ivs.2008.23

[4] G. Andrienko, N. Andrienko, M. Mladenov, M. Mock, and C. Poelitz. Extracting events from spatial time series. In *2010 14th International Conference Information Visualisation*, pp. 48–53, July 2010.

[5] M. Bögl, W. Aigner, P. Filzmoser, T. Gschwandtner, T. Lammarsch, S. Miksch, and A. Rind. Visual analytics methods to guide diagnostics for time series model predictions. In *IEEE VIS 2014 Workshop Visualization for Predictive Analytics, VPA*, 2014.

[6] M. Bögl, W. Aigner, P. Filzmoser, T. Lammarsch, S. Miksch, and A. Rind. Visual analytics for model selection in time series analysis. *IEEE Transactions on Visualization and Computer Graphics, Special Issue "VIS 2013"*, 19:2237 – 2246, 12 2013. doi: 10.1109/TVCG.2013.222

[7] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. *Time series analysis: forecasting and control*. John Wiley and Sons, 2015.

[8] P. Buono, A. Aris, C. Plaisant, A. Khella, and B. Shneiderman. Interactive pattern search in time series. In *Visualization and Data Analysis 2005*, vol. 5669, pp. 175–187. International Society for Optics and Photonics, 2005.

[9] P. Buono, C. Plaisant, A. Simeone, A. Aris, G. Shmueli, and W. Jank. Similarity-based forecasting with simultaneous previews: A river plot interface for time series forecasting. In *Information Visualization, 2007. IV'07. 11th International Conference*, pp. 191–196. IEEE, 2007.

[10] J. Chae, D. Thom, H. Bosch, Y. Jang, R. Maciejewski, D. S. Ebert, and T. Ertl. Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 143–152, Oct 2012. doi: 10.1109/VAST.2012.6400557

[11] N. R. Draper and H. Smith. *Applied regression analysis*, vol. 326. John Wiley & Sons, 2014.

[12] T. Gschwandtner, W. Aigner, K. Kaiser, S. Miksch, and A. Seyfang. Carecruiser: Exploring and visualizing plans, events, and effects interactively. In *2011 IEEE Pacific Visualization Symposium*, pp. 43–50, March 2011. doi: 10.1109/PACIFICVIS.2011.5742371

[13] M. C. Hao, H. Janetzko, S. Mittelstädt, W. Hill, U. Dayal, D. A. Keim, M. Marwah, and R. K. Sharma. A visual analytics approach for peak-preserving prediction of large seasonal time series. In *Proceedings of the 13th Eurographics / IEEE - VGTC Conference on Visualization*, EuroVis'11, pp. 691–700. The Eurographs Association and John Wiley and Sons, Ltd., Chichester, UK, 2011. doi: 10.1111/j.1467-8659.2011.01918.x

[14] M. C. Hao, H. Janetzko, R. K. Sharma, U. Dayal, D. A. Keim, and M. Castellanos. Poster: Visual prediction of time series. In *2009 IEEE Symposium on Visual Analytics Science and Technology*, pp. 229–230, Oct 2009. doi: 10.1109/VAST.2009.5333420

[15] M. C. Hao, M. Marwah, H. Janetzko, U. Dayal, D. A. Keim, D. Patnaik, N. Ramakrishnan, and R. K. Sharma. Visual exploration of frequent patterns in multivariate time series. *Information Visualization*, 11(1):71–83, Jan. 2012. doi: 10.1177/1473871611430769

[16] H. Hochheiser and B. Shneiderman. Dynamic query tools for time series data sets: Timebox widgets for interactive exploration. *Information Visualization*, 3(1):1–18, Mar. 2004. doi: 10.1145/993176.993177

[17] Y. Ichikawa, T. Tsunawaki, I. Fujishiro, and H. Yoon. A visualization environment for multiple daytime stock price predictions. In *Proceedings of the 2nd VIIP International Conferences on Visualization, Imaging and Image Processing, Malaga, Spain*, 2002.

[18] H. Janetzko, F. Stoffel, S. Mittelstädt, and D. A. Keim. Anomaly detection for visual analytics of power consumption data. *Computers and Graphics*, 38:27–37, 2014.

[19] J. Kehrer, F. Ladstdter, P. Muigg, H. Doleisch, A. Steiner, and H. Hauser. Hypothesis generation in climate research with interactive visual data exploration. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1579–1586, Nov 2008. doi: 10.1109/TVCG.2008.139

[20] C. Köpp, H.-J. von Mettenheim, and M. H. Breitner. Decision analytics mit heatmap-visualisierung von mehrschrittigen ensembledaten. *Wirtschaftsinformatik*, 56(3):147–157, 2014.

[21] Y. Lu, R. Garcia, B. Hansen, M. Gleicher, and R. Maciejewski. The state-of-the-art in predictive visual analytics. *Comput. Graph. Forum*, 36(3):539–562, June 2017. doi: 10.1111/cgf.13210

[22] R. Maciejewski, R. Hafen, S. Rudolph, S. G. Larew, M. A. Mitchell, W. S. Cleveland, and D. S. Ebert. Forecasting hotspots: A predictive analytics approach. *IEEE Transactions on Visualization and Computer Graphics*, 17(4):440–453, April 2011. doi: 10.1109/TVCG.2010.82

[23] R. Maciejewski, S. Rudolph, R. Hafen, A. Abusalah, M. Yakout, M. Ouzzani, W. S. Cleveland, S. J. Grannis, and D. S. Ebert. A visual analytics approach to understanding spatiotemporal hotspots. *IEEE Transactions on Visualization and Computer Graphics*, 16(2):205–220, 2010.

[24] R. Maciejewski, S. Rudolph, R. Hafen, A. Abusalah, M. Yakout, M. Ouzzani, W. S. Cleveland, S. J. Grannis, M. Wade, and D. S. Ebert. Understanding syndromic hotspots - a visual analytics approach. In *2008 IEEE Symposium on Visual Analytics Science and Technology*, pp. 35–42, Oct 2008. doi: 10.1109/VAST.2008.4677354

[25] R. Maciejewski, B. Tyner, Y. Jang, C. Zheng, R. V. Nehme, D. S. Ebert, W. S. Cleveland, M. Ouzzani, S. J. Grannis, and L. T. Glickman. Lahva: Linked animal-human health visual analytics. In *2007 IEEE Symposium on Visual Analytics Science and Technology*, pp. 27–34, Oct 2007.

[26] A. Malik, R. Maciejewski, S. Towers, S. McCullough, and D. S. Ebert. Proactive spatiotemporal resource allocation and predictive visual analytics for community policing and law enforcement. *IEEE transactions on visualization and computer graphics*, 20(12):1863–1872, 2014.

[27] P. McLachlan, T. Munzner, E. Koutsofios, and S. North. Liverac: Interactive visual exploration of system management time-series data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pp. 1483–1492. ACM, New York, NY, USA, 2008. doi: 10.1145/1357054.1357286

[28] C. A. Steed, W. Halsey, R. Dehoff, S. L. Yoder, V. Paquit, and S. Powers. Falcon: Visual analysis of large, irregularly sampled, and multivariate time series data in additive manufacturing. *Computers and Graphics*, 63:50–64, 2017.

[29] D. Thom, H. Bosch, S. Koch, M. Wrner, and T. Ertl. Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages. In *2012 IEEE Pacific Visualization Symposium*, pp. 41–48, Feb 2012. doi: 10.1109/PacificVis.2012.6183572

[30] C. Xie, W. Chen, X. Huang, Y. Hu, S. Barlowe, and J. Yang. Vaet: A visual analytics approach for e-transactions time-series. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1743–1752, Dec 2014. doi: 10.1109/TVCG.2014.2346913