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 life, when buying a new laptop or signing a contract. These decisions are easier if there is at least some certainty about the future.
According to Lu et al. [21] predictive analytics is concerned with the prediction of future outcomes and trends based on 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 forecasting process. In combination with Visual Analytics, analysts have an interactive solution, which enables them to better answer common 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 gathered variables and events can be found?

(Q6) Which parameters are influencing my (production) processes?

(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 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.
The goal of this survey is to provide the reader with an entry point to select an appropriate approach according to her/his currently required area of application.

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

2 Abstract Time Series

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 approach 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 (Fig. 1).

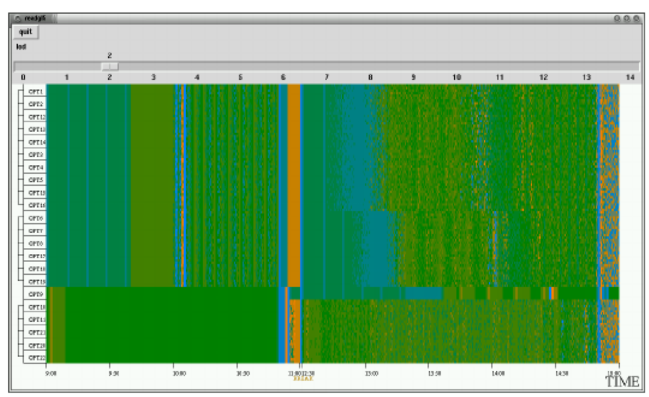

Figure 1: Color band display [17].

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 a river plot instead. The color band is created by assigning similar predictions to the same cluster, which reveals the overall trend for 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. Here, 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, in connection with 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 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 et al. [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 about 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 the system of Ichikawa et al. [17] was not able to answer. A popular choice to address this problem is based on ensemble visualizations, which were only partially applied, because 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. Another resembling feature to Ichikawa et al. [17] is the external source of simulations. Depending on the application area, this can be beneficial or disadvantageous. 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. Therefore, Bögel 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 methods.

2.2 Model Selection

Besides searching for trends with a Visual Analytics system directly, analyst are often interested in creating a model. Models have the advantage that they can be used in automatic systems as well as by unexperienced users once they are defined. However, as mentioned in the previous section, simply applying a model is restricting the analyst. Therefore, model selection approaches want to support analysts in finding the best model for their forecast, so that domain knowledge can be included as much as possible. More precisely, during model selection the systems support the analyst with e.g. subspace exploration, training set modification, parameter tuning [21].

A popular approach for pattern discovery in abstract time series analysis is from Buono et al. [8]. It focuses on automatically detecting similar behavior compared to a user specified pattern. The system was built upon TimeSearcher proposed by Hochheiser & 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 before the model selection step to provide the analyst with a better understanding of the data. To assist the model selection process, the updated version of TimeSearcher by Buono et al. [9] is providing additional features. These additional features include an actual prediction functionality and a preview interface with different parameter selection tools. For the prediction, TimeSearcher resorts on the similarity-based approach, which was used in the previous version [8, 16] 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) as in TiMoVA [5, 6]. 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 in parallel (Fig. 2). This idea of displaying different parameter outcomes is also included in the stock price prediction from Ichikawa et al. [17]. One drawback of the system is, it requires the analyst to provide large datasets in order to make the similarity-base approach work. Further, the actual prediction is only extrapolating a data subset's mean and thus might not be able to represent more complex behavior. 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 large amounts of clean data is often problem, which makes TimeSearcher [8, 9, 16] less valuable. Consequently, a model-driven system, which requires less data points, is preferred. The approach of Ichikawa et al. [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 a system for trend detection. On the other hand, TiMoVA [6] explicitly provides a model selection tool for traditionally successful models such as ARIMA, AR, MA, etc. As consequence, the system is designed after the Box-Jenkins method and is 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. Validation can be done, besides the interactive visualization of the prediction and its certainty, as described in Sect. 2.1, by different residual analysis plots and key figures. A big drawback of the system

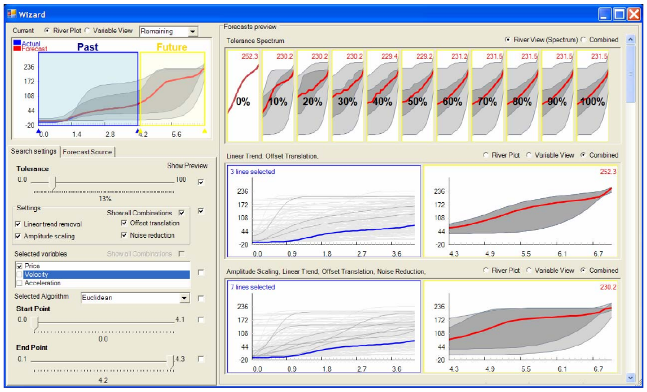

Figure 2: TimeSearcher simultaneous preview interface [9].

are its assumptions about the preprocessed data. It requires a time series without missing values and only supports univariate analysis.

Apart from Box-Jenkins approaches, regression models are a common choice. Further, with a growing social media use[1], those platforms may contain highly relevant information for predictions. Based on this, Lu et al. [22] proposed a model selection framework including mainly social media information. Their framework also offers other functionalities such as sentiment analysis, however, for this survey the focus is only on the time series prediction feature. Similar to TimeSearcher [8, 9, 16], they focus on users without prior knowledge in analysis and statistics and enable the analyst to validate the model on similar instances. Analogous to TiMoVA's extension [5], they offer visualizations to evaluate the degree the model is over- or underestimating the historical data of similar instances. Aside from the model selection, they incorporate an iterative feature selection process in the framework, which allows the analyst to trade-off between more training samples and more features. This can assist the analyst in creating improved models, e.g. the model may generalize better as only relevant features are included, as well as multiple different models. For unexperienced users, baseline models, created from known predictive features, are provided as entry point, which can then be modified by changing parameters, features, etc. As limitations the authors state that the final model is not able to detect cause-effect relationships, i.e. (Q4), (Q5) and (Q6) cannot be answered. Further, the application scenarios presented in the paper are only predictions of one time step into the future, however the framework has the capacity to predict further into the future. Similar to the system of Ichikawa et al. [17], uncertainty is not specifically visualized. Only validation on similar data sets is included.

2.3 Correlation Detection

Trends and models for time series are helpful in making predictions. However, analysts are often interested in predicting unusual future behavior, which allows them to react accordingly. Such anomalies can include fraud attempts, higher server loads or predictive maintenance.

The previously introduced TimeSearcher [8, 9, 16] is able to support analyst with this task. The system provides the ability to extract occurrences of patterns, which were specified by the analyst. Combined with the prediction capabilities of the system, the analyst is able to answer (Q4), (Q5) and (Q6). Equivalent to this, the idea of peak preservation [13, 14] was combined with in a Motif/Pattern detection approach [15]. Unlike TimeSearcher [8, 9, 16] overlapping patterns can be detected and the systems specifically extrapolates these patterns into the future and does not leave this to the analyst. Compared to the previous version [13, 14], it 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).

Apart from those general approaches, a more specialized Visual Analytics approach, called VAET, was proposed from Xie et al. [31]. The difference to other systems can be found in the application area of customer-to-customer e-transaction time series, where a time series consists out of transactions between a seller and a buyer. However, the commodities as well as the buyer can vary greatly. Predicting behavioral pattern into the future helps analysts to understand contextual connections between multiple transactions of one seller, which can be used for identifying fake transactions and fraud (Q4). For analysis the system employs a iterative process, which was also used in the model selection system for social media data [20]. An overview component proposes possible salient transactions based on an automatic saliency prediction. An detailed view is used to gain more insight in 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). This gives the analyst the opportunity to easily assess important information such as the amount of transactions, payment or relationships other time. Consequently, the analyst can identify contextual correlations (Q5) of transactions and find salient transactions and report this information back to the system. By making use of this process, analysts are able to iteratively narrow down their search to important temporal patterns, which enables them to detect fraud or similar attempts in advance. It is also possible, to adapt this idea to different application scenarios such as predictive maintenance as well as outage forecasts for data centers or oil platforms. Albeit, the current version of the system is highly optimized for the sales use case. Another drawback of the system is, that it requires annotated training data for its automated saliency proposals, i.e. analysts are required to annotate features of training transactions.

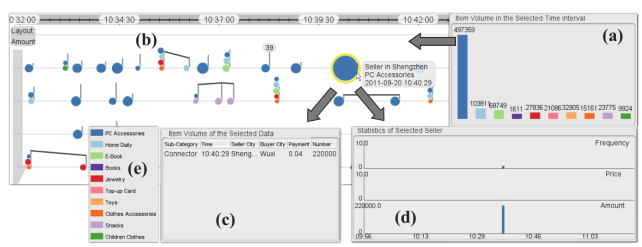

Figure 3: KnotLine view of VAET [31].

Comparable to VAET [31], Falcon was recently presented by Steed et al. [29]. Their system is focused on detecting correlative 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 a manufacturing standpoint to discover defects and system performance issues or increasing production efficiency. The system provides different line plots for each variable with individual level-of-detail control and filtering including an interactive statistical view (Fig. 4). Consequently, multiple different variables can be displayed and examined by the analyst in order to find correlations (Q5). The system also offers a new visualization technique, called waterfall visualization, to combine overview and detailed view (Q1). This idea can be seen similar to the color charts by Ichikawa et al. [17]. From an operational point of view, the system offers, similar to VAET [31], user driven analysis and helps to detect and highlight

[1] https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

univariate and multivariate patterns from different angles. Steed et al. were supporting a universal approach for Falcon to make it applicable in different domains. However, to support their specific analysis task, they included some non-generalizing functionalities. Analogous to the model selection framework from Lu et al. [22], they enable the analyst to compare the time series to a historical/user-defined time series. This comparison makes it easier to distinct between normal behavior and anomalies (Q6). As a consequence, the system grants analyst the ability for predictive maintenance, e.g. detecting a heat development pattern, which signals failure of the printer head in the near future. Currently the system does not support provenance information and the similarity/dissimilarity methods are limited to a single view.

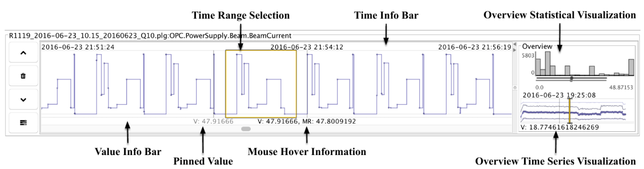

Figure 4: Time series visualization of one variable from Falcon [29].

It has to be noted here, that unlike TimeSearcher [8, 9, 16] or the peak preservation approach [15], VAET [31] and Falcon [29] do not incorporate an automatic prediction functionality. They only assist analysts in examining time series, so that they can extract information about future behavior. Nevertheless, this allows them to provide better support to answer (Q4), (Q5), (Q6) and detect cause-effect relationships, which is typically not possible with other systems.

Other application, which specifically focus on pattern detection can be found for patient treatment plans [12] or climate research [19]. Whereas, the latter can also be interpreted as spatiotemporal as it detects regions in the atmosphere, which indicate climate change. Related to the focus of Hao et al. [13, 14], Janetzko et al. [18] focuses on power consumption in data centers, more specifically on anomaly detection of the corresponding time series. In the same application context is LiveRAC from McLachlan et al. [28], which is supporting, similar to Falcon [29], large scale time series, in particular it was developed for large scale systems management of network devices. Scalability is an issue, which all other systems are not able to handle properly. Falcon [29] is the only other system in this survey to provide similar capacity, although in a different application area. These system were just described shortly as they only touch the topic of the survey peripherally, but should provide the reader with a entry point for further research.

3 Spatial Time Series

The previous section presented an overview of systems for abstract time series analysis. However, for other application areas, such as crime prevention as well as emergency and epidemic intelligence, not only temporal information are valuable, but also spatial information. Typically, spatial information have 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). Andrienko et al. [2, 3] also found that geospatial analysis has a higher complexity and automated methods cannot adequately solve this task. It requires the domain knowledge of a human analyst in order to solve these problems comprehensively. However, as a result of the high dimensionality of data, an analyst needs support from Visual Analytics systems.

Andrienko et al. [4] proposed ab approach, which is in between temporal and spatiotemporal. They include spatial dimensions, but their analysis is largely based on abstract time series compared to the following approaches. Similar to the peak preserving approach from Hao et al. [13–15], they ensure local specialties are represented in the prediction.

Maciejewski et al. [23] are proceeding from their previous work [25, 26] and focus on categorical medical event data. Events consist out of locations in time and/or space, whereas each event can be placed into a hierarchical categorization structure. Similar to TimeSearcher [8, 9, 16], TiMoVA [5, 6] and peak preservation [13–15] confidence bands are established in order to answer (Q7). A geospatial view gives an overview of the percentage of events in a certain area, e.g. patients at an emergency department, which where classified with respiratory syndromes (Fig. 5). One important part is, the systems differentiates between the time series and the geospatial prediction. For multivariate time series data, each event category is modeled as a separate time series signal and consequently forecast. For geospatial predictions, the system utilizes density estimation to determine the spatial distribution of the time series prediction. In order to predict anomalies, e.g. outbreaks of diseases, 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). This is also sent to the analyst as alert, so that he can analyze the occurrence further.

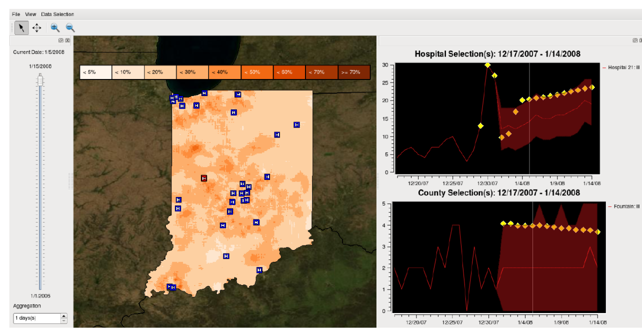

Figure 5: Interactive Visual Analytics environment from Malik et al. [23]. Analysis of respiratory syndrome counts in Indiana on county level. The yellow diamonds indicate temporal alerts, the white line the current day and the transparent polygon the prediction's confidence.

Another system also from Maciejewski et al. [24] is based on the same method . They target multivariate data with high signal to noise ratio and a degree of uncertainty. Equivalent to the previous work [23], the system provides a linked environment of geospatial data and time series graphs as well as separates time series and spatial forecast. The major difference is, this work focuses on finding and understanding patterns, rather than only predicting them. Therefore, the system establishes temporal contour maps, which are overlaid contour maps over a period of time. This allows the analyst 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 in a three dimensional variable space and cannot help to find correlations in a higher dimensional space. Moreover, in both systems aggregating too many data points, may yield to largely exaggerated hotspots or a uniform surface, as a result of too many hotspots.

Recent work from Malik et al. [27] focuses, comparable to TimeSearcher [8, 9, 16] and the model selection framework [22], on a Visual Analytics approach that provides non statistic experts a proactive and predictive environment, which allows them to utilize their domain expertise. Analogous to Maciejewski et al. [23, 24], the systems is using the same approach in order to predict time and space. One issue which was identified by Malik et al. in earlier work from Maciejewski et al. [23] was that domain experts need additional guidance in order to improve their analysis. This motivation is equivalent to the preview interface in TimeSearcher [8, 9, 16] or the baseline models from the model selection framework [22]. Geospatial and temporal scale templates present the analyst with a starting point and avoid searching through the complete parameter space. Further, the system allows the user to interactively change the initial templates to include e.g. police beats. In order to compensate for insufficient data points, the system can make use of demographically similar neighborhoods within a certain radius around that area. Temporal templates follow the idea of peak preservation, such as in the work of Hao et al. [13–15]. An additional improvement to the previous systems includes the easier access to observe trends on different scales of time, i.e. on hourly, daily and monthly basis.

In contrast to all the previous spatiotemporal approaches Andrienko et al. [2] did not use the same prediction approach, but applied self-organizing maps (*SOM*) for either spatial or temporal prediction. SOMs can be seen as a combination of clustering and dimensionality reduction based on the similarity of space and time, therefore this can be seen at least alike to the similarity-based approach of TimeSearcher [8, 9, 16]. 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 earlier work from Andrienko & Andrienko [1]. The more recent approach is applying feature images and index images (Fig. 6) for the spatial and temporal dimension. Feature images are used in order to analyze the magnitude of attributes. Spatial data is represented as map and temporal data as temporal mosaic, which has a large correspondence to the color charts of Ichikawa et al. [17] and the waterfall visualization of Falcon [29]. Index images show either the temporal or spatial elements, which provide the data source for one SOM matrix cell, depending on the analysis context. In order to enable the analyst to find correlations between time and space, multiple SOM cells with different feature images and index images 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 temporal aggregation shows again a similar structure as the color charts of Ichikawa et al. [17] and the waterfall visualization of Falcon [29]. The evaluation of the system shows that it is applicable for detecting expected as well as unexpected results. This allows analysts to either find evidence, which supports their previous formed hypothesis or discover new connections, which enables them to make assumptions about future events.

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, 30]. 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 and Conclusion

Time series prediction has two major directions: abstract and spatiotemporal. Abstract time series analysis mainly follows interactive approaches, which support the analyst in trend detection, model selection or correlation detection. Supportive features in this context are e.g. the simultaneous preview interface [9], the comparative variable view [29] or KnotLines [31]. They enable the analyst gain more knowledge about the data, generate hypothesis and predict future developments or anomaly occurrences. Spatiotemporal time series add another important dimension, which makes automated analysis more complicated. Those systems have to enable the user

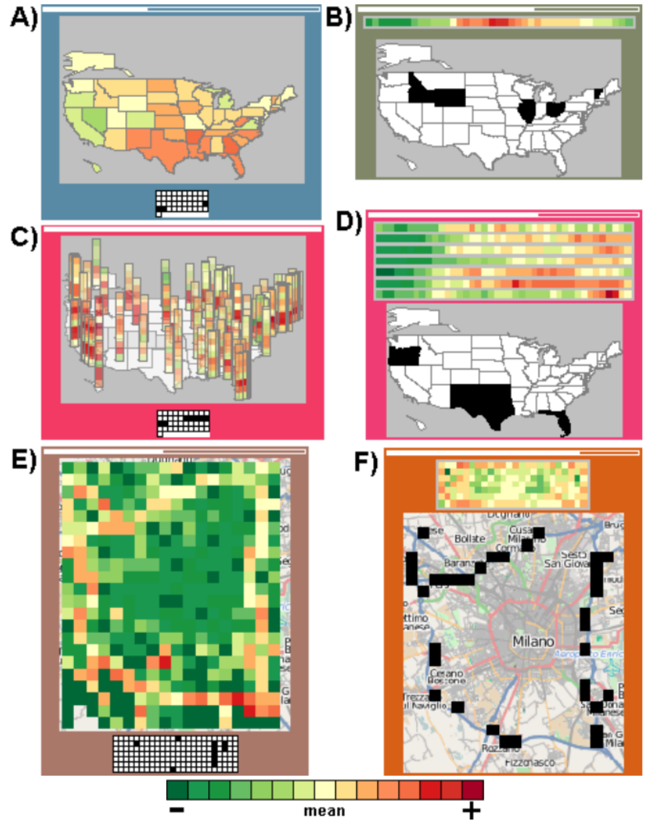

Figure 6: Feature images (top) and index images (bottom) grouped by space (left column) or time (right column) [2].

to select spatial and temporal resolution. The system of Malik et al. [27] even suggests different templates to makes the systems more accessible to domain experts.

One important drawback most systems, besides Falcon [29] and LiveRAC [28], have, is the missing scalability to large scale data. Further, none of in this survey presented approaches was able to provide a solution high dimensional outputs, e.g. predicting the demand for multi-product companies, customer-to-customer online platforms, etc. In contrast to this, having sparse data was only addressed by Malik et al. [27] using neighboring areas. On the other hand, current Visual Analytics system offer a broad range of application areas including manufacturing, medical, environmental and social media oriented forecasts. Additionally, different systems address different levels of the analyst's competence, hence show completely different application scenarios. Uncertainty was addresses by some of the systems, but only the ensemble approach [20] offered multiple analysis options. Thus, future work should include more combined work with time series prediction and ensemble methods, which additionally might provide solutions to deal with sparse and dense data sets.

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] Y. Lu, R. Krüger, D. Thom, F. Wang, S. Koch, T. Ertl, and R. Maciejewski. Integrating predictive analytics and social media. In *Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on*, pp. 193–202. IEEE, 2014.

[23] 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

[24] 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.

[25] 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

[26] 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.

[27] 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.

[28] 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

[29] 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.

[30] 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

[31] 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