

ADetector: An interactive visual analytic debugger for time-series climate anomaly data detection with human

Tsung Tai Yeh*

Electrical Computer Engineering Department
Purdue University

Shuying Feng†

Electrical Computer Engineering Department
Purdue University

ABSTRACT

Data quality play an important on the data analysis results. However, finding and fixing anomaly data is time-consuming in big data. There should be a systematic approach to narrow down the data checking regions. In addition to the expected abnormal data, such as missing data, and extreme data, the time-series data also includes some unpredictable and latent erroneous data. The patterns of unpredictable data are difficult to predict and recognize. We present ADetector, a visual analysis tool for anomalies identification in complex and large time-series datasets. This tool applies the nonlinear regression model to fast locate potential anomalies in big time-series data, introduces a novel data representation of comic map grid for anomalies analysis to identify anomalies. ADetector exploits visual analytics interface and iterative anomaly detection pipeline to pick out possible latent abnormal data.

Keywords: time-series data, anomaly detection, visual analytics

1 INTRODUCTION

The precision of measurement plays an importance role in scientific experiments and engineering systems, since the mistakes of the measurement often cause disastrous consequences. For instance, the confusion about the value of the fuel led to the failure of NASA Mars Climate Orbiter in 1999, and the A Boeing 767 aircraft ran out of gasoline in the mid-flight[7]. However, errors embedded in the scientific measurements are inevitable, since the experimental uncertainties or malfunction of apparatus. For example, many inputs are either human-generated or unstructured data (e.g. user logs, tweets, comments on social networks), or assessment data from sensors (thermometers, scales, DNA sequencers) including measurement errors.

Error analysis aims to measure the quality of the scientific experiments and engineering systems, find out abnormal data hidden in a set of measurements and gets rid of these unexpected errors. However, it is not easy to figure out uncertain errors hidden in the big amount of data, and analyzing the complete set of data could be computationally expensive. The results from a thorough analysis of this big data may not be worth the cost for completeness. In addition, we are not able to know what the exact value is in the measurement. However, in fact, we can make sure the correct value exists in a specified likelihood[7]. Thus, an approximate clue from big data is supposed to be enough for us to obtain perfect answers to pinpoint errors similar to results from a thorough analysis.

Time-series analysis aims to understand the underlying time-series data structure, and then exploits the results of analysis to interpret and forecast the trends of events. Time-series data is used to be collected from monitoring procedures for an event, and problematic errors always mislead prediction results. Nonlinear regression

fitting curve methods, such as GaussNewton algorithm, and Levenberg Marquardt algorithm are often used in nonlinear regression analysis. However, the divergence problem and improper parameters often slow down their convergence speeds. In addition, Kalman filter is usually used in signal noise elimination. However, Kalman filter needs to include a predefined noise-free model. Thus, complexity, feasibility, and efficiency should be taken into consideration in nonlinear regression model design.

The great numbers of uncertainties in the measurement increase the challenge of error diagnosis, since the limited erroneous data knowledge base. In general, the taxonomy of anomaly in the scientific experiments includes random errors, systematic errors, and mistakes. Random errors impact on the precision of the measurement in which indicates the experimental results are close to the average one in the repeated measurements. For instance, the malfunctioned instrument often leads to the fundamental random noises, and cause the bad degree of precision. Furthermore, the systematic error means the measurement quantity is inaccurate, which is away from the accepted or predicted values. For example, the interference arises from environment often distract the measurement results from the expected one. The last one is the mistakes, which is often caused by some unpredictable conditions such as incorrect manual records, and is difficult to detect. These hatred errors could be solved by integrating visual representation, data analysis, and human interaction.

The least-square fitting methods, regression model and signal processing filters are often used to measure the relationship between two or more phenomena. The residual value among variables is usually being a baseline to estimate outliers included in a set of measurement. Furthermore, the previous research always pursue a best-fit model to predict or estimate the trends or patterns of data. However, the computational costs and convergence speeds of these curve fitting methods are getting to be a concern in big data analysis. In fact, the amount of noises or errors is usually less than the correct ones in various applications. In other words, the noise is not the principle characteristics of the measurement. In addition, we never obtain the exact measurement and the erroneous patterns are often shift away from the expected value. Thus, an approximate solution is enough to diagnose the experimental data. In this paper, we present an approximate nonlinear regression – matching pursuit, which integrate Fourier Transform and approximate least square method for anomaly inference.

Visual analytics is used to reduce the size of data analysis, to audit the data, and to detect anomaly interactively. Even if statistical inference methods are helpful for the recognition of abnormal data, however, some kinds of spurious data which are hard to be recognized by numerical model. For instance, the small glitch, or a series stark unchanged measurement. The visualisation enables human-beings to extract additional characteristics via data comparison efficiently. Moreover, the statistical evidence and view highlights allow people to determining a specified region of data that is susceptible to errors to check and avoid the thorough analysis. We present an anomaly visual analytics pipeline, and a data comparison graphical interface – comic map to take insight into the complex measurement data.

*e-mail: yeh14@purdue.edu

†e-mail: feng98@purdue.edu

In sum, this paper includes the following contribution.

1. The approximate nonlinear fitting curve model: This model combines signal processing approach, and leverages fitting curve to predict abnormal data hidden in time-series climate data.
2. Time-series abnormal data diagnosis: The buggy region localization is a divide-and-conquer method to reduce the regions of data checking.
3. Visual analytics pipeline for anomaly detection: Showing the possible abnormal data in line graph comic map cells, and exploits various interactions such as zoom in/out, swapping, and merging to allow human-being recognize abnormal data patterns.

2 BACKGROUND

2.1 Anomaly data classification

The category of time-series anomaly is opt to the objective, and monitoring procedure of data collection. The time-series data we are focus on is weather data sets, which are collected by MesoWest weather data service settled by University of Utah. MesoWest weather data service collects numerous weather data such as temperature, humidity, and wind pressure etc.. at thousands of locations around the United States, and provides weather history records over 10 years at each measurement locations. According to the observation of the weather data from MesoWest service, we sum up time-series anomaly in weather data as follows.

a. Missing data results from unexpected status of climate sensors, and causes the value of collecting data becomes null. The problem of the missing data is to interrupt the numerical calculation in data analysis, and to induce the wrong analysis results. Thus, the missing data is often removed before the start of data analysis via data clean tool. However, the missing data occurrence frequency, and time record can be the inference of abnormal data, and evaluation of the status of measure instruments.

b. Extreme data is the outlier [16] in a set of time-series data. The extreme data can be undermined by the robust data analysis. Although the outlier is often looked as an anomaly in signal processing or image processing, since many environmental factors influence the change of the weather data patterns, the extreme data is not always an error in climate data sets. The spurious extreme data impacts on the accuracy of results of analysis. However, to visualize the extreme data is able to help for finding additional characteristics in the weather data sets. In addition, time-series visual comparison is also helpful for anomaly judgement.

c. Latent abnormal data indicates some unpredictable data and even the regression model is hard to discern this type of error. For instance, the small glitch or a set of unchanged measurement occurred in a continuous time interval. The latent abnormal data influence the precision of the measurement, and cause the incorrect trend estimation. The visualization is an efficient approach to understand the fluctuation of time-series data, and figure out more characteristic about latent abnormal data.

2.2 Time-series visual exploration

Considerable visual analytics research which is related to the time-series visual visualization, and exploration, and the relevant surveys include [18]. The basic line graph [20] is widely in temporal visualization. However, in recent year, in order to delve into the time-series, different visual representation techniques have been worked out [1], [14], [8], [19], [12], [11]. These graphical representation methods improves the efficiency and effectiveness for time-series data analysis. Furthermore, the massive time-series visual analytics has also become a consideration in ATLAS [2] system.

Query and exploration also plays a key role in time-series visual analytics. LifeLine [15] is an early visual system for personal history record exploration. TimeSearcher [6] is an interactive time-series query tool, and uses the timebox to explore the data sets. Pattern-Finder [5] includes an visual interface to query the medical records.

LiveRAC [14] designed an interactive visual interface to explore the data collected from computer systems. Data quality assessment is also getting to become an important topics in visual analytics research. Profiler [9], and RainMon [17] integrated statistical analysis and visual representation to find out abnormal hidden in database and computer systems.

The above projects concentrate mainly on the interactive visual explorations, and visual representation for time-series data. However, beyond the high-level visualization tasks, temporal visual exploration also requires additional attention to the pipeline of the data analytics for special properties. In particular, additional important time-series visual analytics tasks include (a) multi-variable relationship (b) dynamic time-series hierarchies (c) large-scale support [8]. Few of the above visualization tools support all these requirement, except Stack Zooming [8], and Live RAC [14].

2.3 Support multi-focus comparison

The anomaly is the one that does not follow well defined normal behavior [3]. Anomaly is often unveiled via comparison with expected values or baseline model. Thus, we argue the temporal analysis is supposed to support the concept of multi-focus comparison. Multi-focus comparison framework is consist of the comparison tasks for the multiple focus, context, and distance awareness [8], [4]. This approach is supported by Stack Zooming [8], which leverages one-dimensional nature of the data to present hierarchies of visualization strips, and LiveRAC [14]. In this paper, we take advantage of different approach to multi-focus interaction compared to Stack Zooming and LiveRAC system. We leverage comic map which is composed of multiple small cells, and integrate with the types of anomaly and timestamps to present the relationship of anomaly and its occurrence time.

2.4 Support collaboration and data reduction

Kittur et al [10] claims crowdsourcing is a successful approach, which depends on the intelligence of a group of human-beings to solve some hard problems. Furthermore, the distributed cognition [13] has been showed the power to amplify sensemaking in the team-working analysis. The anomaly visual analytics tool is supposed to take the collaboration into consideration, since the mathematical model cannot understand all uncertainties embedded in time-series data. Thus, the participant in the team analysis can be a key role for anomaly exploration.

The scalability is getting to become a critical issue in big data visual analytics. It is hard to display all big data, because the overwhelming computational and memory overhead. The visual analysts can help to narrow down the size and range of data analysis based on the statistical view recommendation. In this paper, we work out a new visual analytics pipeline, which enable analysis participants join the iterative anomaly detection, and set up a possible abnormal time-series range to explore anomaly.

3 ADETECTOR

ADetector is an anomaly visual exploration tool. Since the overwhelming anomaly detection costs in the massive data sets, ADEtector combines an approximate statistical regression model, anomaly diagnostics, and visual analytics to narrow down the size of anomaly analysis, take insight into the complex structure of errors, and at last unveil abnormal data. Thus, ADEtector is consist of data clean, reduction and comparison widgets to tackle on the time-series weather anomaly detection, and convey possible anomaly data to visual analysts to figure out the patterns of anomaly collaboratively.

3.1 Anomaly diagnosis workflow

In order to take insight into the patterns of abnormal data. ADEtector takes advantage of an interactive visual anomaly detection

pipeline to infer unpredictable anomaly. Thus ADetector is composed of five steps – Data Table, Type inference, Anomaly region search, view recommendation, and interactive time-series navigation, and comparison.



Figure 1: The anomaly detection workflow of ADetector tool

Data table includes a series data manipulation options such as the selection of input and the type of weather data. Since ADetector tackles on missing, extreme, and latent anomaly, in the type inference step provides options to specify one of anomaly, then ADetector will use the specified analysis tool to analyze the selected one. Afterward, in the anomaly region search step, based on the selected anomaly, ADetector leverages statistical analysis to work out some statistical results that are related to the behavior of data to let users select a certain time interval (e.g. one year) to delve in to the micro-level. At last step, ADetector exploits regression model to carry out anomaly matching calculation, then show the possible anomaly region in multiple cells. The missing, extreme data statistics are showed in timeline widget, and users also can pick out data in one of time interval to view in cells in comic cells. Altogether, ADetector workflow is composed of selection, reduction, and comparison steps.

3.2 The composition of ADetector

3.2.1 Data clean tool

Data clean tool aims to find out all missing data in the data sets and show their timestamp. Users are able to figure out the occurrence frequency and time from this data clean tool. Furthermore, users also can remove the missing records from this data clean tool. Altogether, the data clean tool gives users an overview of the information about missing data, and allows users to remove them.

3.2.2 Abnormal region reduction tool

The purpose of possible abnormal region reduction tool is to narrow down the data checking. This tool is composed of several histogram, and show the proportion of multiple statistical analysis results in each year. ADetector give three view recommendations including the number of missing data, the average of extreme data, and the residual between raw data, and fitting curve model to help users determinate which year is supposed to pay attention to. Afterward, users can delve into one specific year, and take a look at the monthly distribution. In the meantime, according to view recommendations. Users can choose one month to proceeds micro-analysis.

3.2.3 Data comparison tool

The data comparison tool consists of comic map, which includes multiple line graph cells. Importing the selected monthly data from susceptible abnormal region reduction tool, ADetector proceeds residual calculation toward selected data, and show a number of possible abnormal region in each comic map cells. In addition, this tool also supports cell switch and multiple line graph merge functionalities. From this tool, users are able to compare the fluctuation of each comic map cell. Furthermore, ADApps data comparison tool also includes stack zooming functionality. Users can select one certain time interval to zoom in and zoom out.

4 COMIC MAP VISUAL REPRESENTATION

After getting the potential anomalies from the computer, users can step in to analyse the potential anomaly in order to identify if it is actually an anomaly or to find out the reason for the anomaly. In time-series climate data, for example, some abnormal temperature may imply some useful climate phenomena in a specific time frame. In order to find out the pattern or trend behind these anomalies, it is useful to provide the users some method to compare the anomalies. Also an overview of the relationship of these anomalies with the whole datasets will help the users establish an overall understanding as well. Here we introduce a novel representation of anomalies comic map.

Comic map is composed of several grid cells. Each shows detail information of each anomaly (Figure 1). With this layout, user can swap and merge cells freely to compare the anomalies with simple drag-and-drop functionality. Moreover, An overview timeline representation of all anomalies is also established corresponding the comic map grid to give the user more knowledge of the anomaly based on the overall timeline (Figure 2)

4.1 Comic map cells

The basic component of the comic map is every comic map cell. For time series data, a timeline-based representation is more intuitive for user to receive. And also in order to better understand the anomaly, the data surrounding this anomaly is needed for the user to get a picture of the trend of this anomaly, like for extreme anomaly, it appears much higher than other data around. So in each cell, anomaly data is shown along with some more data before and after it in a time sequence visualized as line graphs. In addition, the anomaly data is marked with a red circle as shown in Figure 3.

Other than the basic line graph of anomaly, some additional information is provided as plain text shown in the upper left corner of the cell, including standard deviation, mean value, start date of the data in this grid and end data of this grid cell. There is no axis along with timestamp labels is shown due to the space limit of the grid cell. Then the start date and end date is given to provide the user the information about the time. All these information are trying to help analyst get a better understanding of the dataset in the cell.

4.2 Layout

Cells of comic map are laid out as comic book grid (Figure 1). In this way, each cell has three (corner cell), five (edge cell) or eight neighbours around it. This kind of layout is more helpful for comparison. Since usually users prefer to set objects side by side to compare it. In our layout, each cell can be easily compared with their neighbours and the number of neighbours is bigger than the traditional horizontal or vertical layout. This is efficient since analyst can do more comparison in comic map layout with less cells moving around.

The comic map tries to maintain a layout more like a square, which is not only efficient of using the visual space, but also guarantees the aggregation of cells for comparison mentioned before. For example, if there are 4 anomalies in total, a 2x2 comic map will be shown, and if the anomalies number is 9, the comic map is a 3x3 grid. More specific, comic map will be chosen as the closest square based on anomalies number. For example, anomalies number between 10 and 16 results in a 4x4 grid.

Moreover, in order to reduce the work of swap, our backend model sorts all anomalies by putting anomalies with similar pattern besides each other. During analysis process, analyst can also rearrange cells for comparison by simply swapping cells. Our layout is very flexible, which provides the analyst a lot of freedom.

4.3 Timeline overview

For each cell, some information related to the specific anomaly can be obtained, but analyst still lacking information of this anomaly related to the whole dataset. So we introduce a timeline overview corresponding to the comic map grid (Figure 2). Since the dataset is a time-series climate data, time information should be an understandable choice for an overall representation. In the timeline overview, an axis of timestamp is used as the main visualization. Some timestamps are shown to give analyst an overview of time. On the time axis, anomalies are marked corresponding to their happening time. Specifically, different type of anomalies are marked using different colours. Through this overview visualization, analyst gains information of anomalies based on the whole dataset, like the frequency of anomaly in a specific timeframe, the occurrence of different type of anomalies, etc. The comic map cells are also tightly connected with this timeline overview. When a comic map cell is selected, the corresponding timeframe of data in the cell will be highlighted in the overview. More information can be obtained for analyst to analyse anomaly in the cell (Figure 4).

4.4 Comparison interaction

As mentioned before, comparison plays an important role in anomaly analysis. Our comic map visual representation provides advanced interactions to support comparison, which gives the analyst more freedom of doing comparison and increases the efficiency of analysing process.

4.4.1 Swap

The first interaction for comic map is swap. Analyst can swap any two cells by simply drag and drop. By using this interaction, analyst can put any cells beside to do a comparison. Also, analyst can rearrange the comic map to put similar anomalies together (Figure 5).

4.4.2 Merge

The second interaction for comic map is merge. Comparison can be done by putting objects side by side, but can also be done shown two datasets in one view. The introduction of merge interaction enables the ability of doing comparison in one cell. In order to merge two cells into one cell, analyst can just pick up one cell and drag-and-drop it onto the cell that he wants to merge with. For example, if the analyst wants to merge first, second and third cell, he can drag-and-drop the first cell to the third cell, and then drag-and-drop the second cell to the third cell. A merged view is automatically generated in the third cell (Figure 6).

The merged view is also revertible. There is a cross sign on the upper-right corner of the merged view. Simply by clicking the cross sign, the merged view will be reverted back to the original dataset view. For example in our former example, after clicking the cross sign, the third cell will show the original anomaly data in that cell (Figure 7).

4.4.3 Stack zooming

The third interaction for comic map is stack zooming. Analyst often needs to see closer to the anomaly by zooming. The way of doing zooming in our comic map representation is stack zooming.

In order to zoom in a specific subset of data, analyst select the region (stack) he wants to zoom in (Figure 7). Then the cell will be automatically updated showing that portion of data. The same as merge interaction, stack zooming is also revertible by just clicking on the cross sign on the upper-right corner. Then, the original whole dataset of anomaly will be shown.

5 MATCHING PURSUIT – REGRESSIVE ANOMALY DETECTOR

Regression analysis is used to estimate the trends among variables with the change of the time. Regression analysis model is often used to predict or interpret observed data. Matching is one nonlinear regression model used in ADetector to figure out possible anomaly. The role of regression model is being a baseline to account for the data, and distinguish anomaly from normal one. Least-square fitting curve aims to work out a best-fit formula to describe the distribution or behavior of data. Linear regression model such as ordinary least squares, and weighted least square. However, these linear models sometimes are not applied in nonlinear time-series data, such as financial data, and weather data. In order to find out anomaly from massive data, the nonlinear regression model is supposed to meet scalable, low complexity, and feasible requirements.

5.1 Approximate nonlinear regression model

Fourier transform is a well-known tool in engineering and science. The amplitude in the frequency domain is able to represent the characteristic of the data in the time domain. Supposed that the noise is not the principle feature in the data, so the amplitude of noise is low. In addition, Fourier transform can be looked as a noise filter, since in the Fourier frequency domain, high amplitudes usually gather in the initial frequency. Thus, our objective is to leverage the top k big amplitude to reconstruct a fitting curve to estimate the anomaly.

Matching pursuit is composed of the following five steps. The first step is to use fourier transform to convert time-series input data to frequency domain. The second one is to obtain the frequency index of top k high amplitude to be the training data. Since the estimate data is consist of the scalar A , and the Fourier Transform equation, but A is unknown. Thus, in order to get the A scalar, the least square method is used to get the scalar of A . After getting the number of A , the next step is to use scalar A_1 to A_k to work out the number of k eigensolutions. At last, superpositing these k eigen-solutions and carry out an approximate model that is closed to the original one.

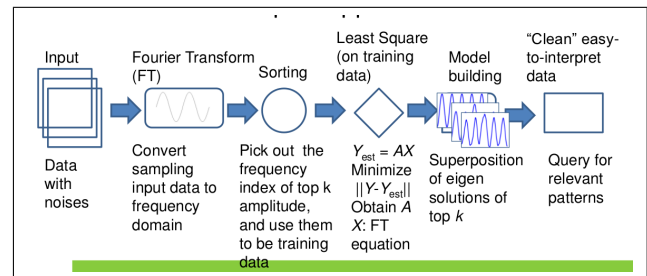


Figure 2: The workflow of matching pursuit regression model

The advantages of matching pursuit regression model is it uses the number of frequency to build up the nonlinear fitting curve, and reduce the computational analysis costs. Furthermore, matching pursuit is non-parameter or unsupervised learning algorithm, there is no any training data requirement. At last, matching pursuit algorithm is also scalable, because it can be run in divide-and-conquer approach and in parallel way. Matching pursuit regression model provides us a rapid method to estimate the anomaly in massive time-series data.

5.2 Anomaly diagnostic logics

Only matching pursuit regressin model is not enough to figure out the location of anomaly, since matching pursuit is like a ruler, there is supposed to be a method to figure out how to use this ruler to measure.

(a) Systematic error: the systematic error is a bias that is shift away from the mean. According to the residual formula, we can calculate studentized residual between real data, and model one to figure out the outliers. ADetector use residual obtained from the matching procedure, and determine possible anomaly.

$$res_i = \sigma^2 \left(1 - \frac{1}{n} \frac{((x_j - \bar{x})^2)}{\sum_i (x_j - \bar{x})^2} \right) \quad (1)$$

(b) Latant anomaly data: Latent anomaly data is easy to unaware of because the characteristic of this anomaly is not revalent. In order to look for this anomaly, we calculate heteroskedasticity, and use cook's distance to understand the influential relation between real data and model, and visualize them in comic map cells.

6 RESULTS

We have conducted the evaluation of A-Dector on the time-series weather data collected from Meso weather service website. Meso time-series weather data service keeps over 17 years at different locations around the United States. We now want to validate the relationship between the degree of approximate and the quality of matching pursuit nonlinear regression model. In addition, we also want to figure out the analysis performance of this model, and the effectiveness of this model.

6.1 The performance validataion

In this validation, we want to see the performance of the anomaly detection in different size of data sets.

Mean square error (MSE) is often used to account for the difference between two variables. In addition, Pearson correlation coefficient is used to measure the linear correlation between two variable. In this experiment, we input 3 year time-series weather data (the total number is 14600), and we change the number of k (the number of iteration) in the model construction, then measure the variations of MSE and Pearson correlation coefficient.

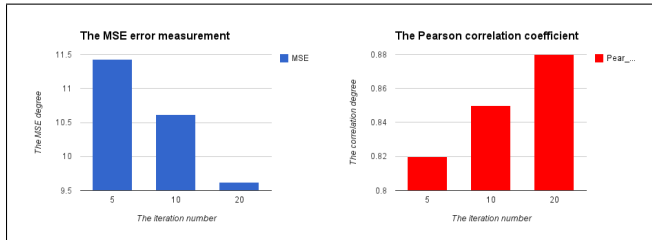


Figure 3: The MSE and Pearson correlation coefficient

At first, we can find the correlation in each experiemnt result reveal positive correlation, which means the model data is correlated to the real data. Furthermore, we also find the correlation coefficient is high when we only iterate the regresion algorithm 5 times. In other words, even the number of iteration is small, the degree of correlation is high positive. Second, we also conduct the MSE measurement, and we found the MSE decrease with the number of iteration, but the completion time increases with the growth of iteration.

6.2 The performance validation

In this validation, we want to see the performance of the anomaly detection in different size of data sets. We can see A-Detector can complete the anomaly data less than one minute.

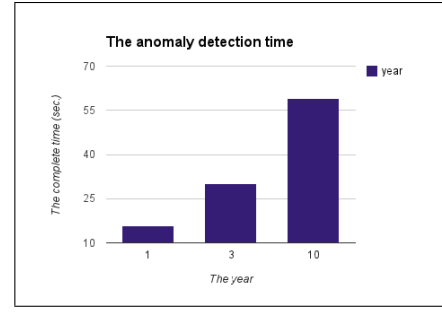


Figure 4: The workflow of matching pursuit regression model

7 FUTURE WORKS

In future work, we plan to support adaptive model tuning based on the feedback from users in the anomaly detection process. Currently in our tool, the anomaly detection model is totally processed in backend and users dont have any control on it. Users can only compare the anomalies detected by the backend model, determinate whether it is really a anomaly by comparison and finally decide to delete it or not. So in order to give uses the ability to refine the backend model, we can let user add new parameters or change the value of parameters of the model (like correct data range, slope, etc) after finding out anomalies from visual graph.

Moreover, for the data representation in the current version of our tool, we only support data aggregation in the overview window. For example, the overview of year shows first, if user clicks on a year data, detailed monthly data shows up. But in analyze workspace, an overview map only shows up for monthly view. We may continue investigate a suitable aggregation representation in the analyze workspace map representation as well, which can give the user a better view of the whole structure for analyzing anomalies data as well.

8 CONCLUSION

In this paper, we introduced a theoretical regression model, and a practical implementation of multi-focus comparison technique called comic map. The quality of data play a critical role in scientific and engineering systems, because errors often cause catastrophic consequence. However, in order to meet the requirement of a mountain of data, we should rethink the anomaly detection methodology. ADetector combine signal processing techniques, and visual analytics to work out an anomaly visual exploration pipeline. ADetector is able to complete analysis of over 10 years time-series weather data, and pick out some possible anomaly less than one minute. Furthermore, ADetector also supports interactive actions such as timeline navigation, swap, multi-variable comparison, and collaboration. ADetector enable the workflow of time-series anomaly detection become efficient and interactive.

ACKNOWLEDGEMENTS

The authors wish to thank A, B, C. This work was supported in part by a grant from XYZ.

REFERENCES

- [1] S. Bremm, T. von Landesberger, M. Heß, T. Schreck, P. Weil, and K. Hamacherk. Interactive visual comparison of multiple trees. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*, pages 31–40. IEEE, 2011.
- [2] S.-M. Chan, L. Xiao, J. Gerth, and P. Hanrahan. Maintaining interactivity while exploring massive time series. In *Visual Analytics Science and Technology, 2008. VAST'08. IEEE Symposium on*, pages 59–66. IEEE, 2008.

- [3] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. In *ACM Computing Surveys (CSUR)*, volume 41, 2009.
- [4] N. Elmqvist, N. Henry, Y. Riche, and J.-D. Fekete. Melange: space folding for multi-focus interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1333–1342. ACM, 2008.
- [5] J. A. Fails, A. Karlson, L. Shahamat, and B. Shneiderman. A visual interface for multivariate temporal data: Finding patterns of events across multiple histories. In *Visual Analytics Science And Technology, 2006 IEEE Symposium On*, pages 167–174. IEEE, 2006.
- [6] H. Hochheiser and B. Shneiderman. Dynamic query tools for time series data sets: timebox widgets for interactive exploration. *Information Visualization*, 3(1):1–18, 2004.
- [7] I. G. Huges and T. P. Hase. *Measurements and their Uncertainties – A practical Guide to Modern Error Analysis*, volume 1. Oxford, 2012.
- [8] W. Javed and N. Elmqvist. Stack zooming for multi-focus interaction in time-series data visualization. In *Pacific Visualization Symposium (PacificVis), 2010 IEEE*, pages 33–40. IEEE, 2010.
- [9] e. a. Kandel, Sean. Profiler: Integrated statistical analysis and visualization for data quality assessment. 2012.
- [10] A. Kittur, E. H. Chi, and B. Suh. Crowdsourcing user studies with mechanical turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 453–456. ACM, 2008.
- [11] e. a. Liao, Qi. Visualizing anomalies in sensor networks. 2011.
- [12] Y. Y. Liao, Zicheng and B. Chen. Anomaly detection in gps data based on visual analytics. 2010.
- [13] Z. Liu, N. Nersessian, and J. Stasko. Distributed cognition as a theoretical framework for information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1173–1180, 2008.
- [14] P. McLachlan, T. Munzner, E. Koutsofios, and S. North. Liverac: interactive visual exploration of system management time-series data. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 1483–1492. ACM, 2008.
- [15] C. Plaisant, B. Milash, A. Rose, S. Widoff, and B. Shneiderman. Lifelines: visualizing personal histories. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 221–227. ACM, 1996.
- [16] P. J. Rousseeuw and A. M. Leroy. *Robust regression and outlier detection*, volume 589. Wiley. com, 2005.
- [17] I. Shafer, K. Ren, V. N. Boddeti, Y. Abe, G. R. Ganger, and C. Faloutsos. Rainmon: an integrated approach to mining bursty timeseries monitoring data. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1158–1166. ACM, 2012.
- [18] S. F. Silva and T. Catarci. Visualization of linear time-oriented data: a survey. In *Web Information Systems Engineering, 2000. Proceedings of the First International Conference on*, volume 1, pages 310–319. IEEE, 2000.
- [19] e. a. Sips, Mike. A visual analytics approach to multiscale exploration of environmental time series. 2012.
- [20] E. R. Tufte and P. Graves-Morris. *The visual display of quantitative information*, volume 2. Graphics press Cheshire, CT, 1983.