

Exploration of Anomalies in Cyclic Multivariate Industrial Time Series Data for Condition Monitoring

Josef Suschnigg
Pro2Future, AVL List
Graz, Austria

Belgin Mutlu
Pro2Future, Know-Center GmbH
Graz, Austria

Anna Katharina Fuchs
AVL List
Graz, Austria

Vedran Sabol
Know-Center GmbH
Graz, Austria

Stefan Thalmann
University of Graz
Graz, Austria

Tobias Schreck
Graz University of Technology
Graz, Austria

ABSTRACT

Industrial product testing is frequently performed in cycles, resulting in cycle-dependent test data. Monitoring the condition of products under test involves analysis of large and complex test data sets. Main tasks are to detect anomalies and dependencies between observation variables, which appears to be challenging to engineers. In this paper, we present a flexible and extendable visual analytics approach for anomaly detection focusing on cycle-dependent data. It is based on a glyph representation to visualize anomaly scores of cycles with respect to interactively selected reference data. Our approach is built on a design study in collaboration with an industrial engineering corporation, and is demonstrated on real data from engines tested on automotive testbeds. Based on findings from evaluation results, we provide a discussion and an outlook for future work.

1 INTRODUCTION

The trend of digitization in industry (often synonym for so-called Industry 4.0) generates large amounts of data by sensors and data recordings from almost all machines and devices of the production process with the promise of creating new usage opportunities [39]. Over all stages of the industrial product life cycle (PLC), extracting valuable knowledge from generated data can lead to an improvement regarding costs, quality and increased flexibility (incl. safety, durability, reliability) [3][31]. However, for human perception, it can be overwhelming to observe and analyze large industrial data sets. Another important requirement of analyzing industrial data is that extensive professional and domain-specific knowledge of users is required [42]. Addressing those challenges, visual analytics research proposes tools supporting domain experts to explore large and complex data sets [26]. Further, to identify and address the industry's needs, our case study has been conducted in close collaboration with an industry partner from the automotive sector, focusing, among others, on engine testbeds. The main driver of the focus on the anomaly detection is the common occurrence of real-world problems in automotive condition monitoring. Due to the large amount of testbed data, data analysis is time-consuming and there is a risk that events which lead to the failure of an engine under test, may not be anticipated or overlooked by engineers. We hence research the development of visual analytics approaches to support test engineers in creating hypotheses and early warnings of potential failure cases.

As the first contribution of this paper we characterize testbed data and study how engineers achieve their data analysis goals. Based on the design study we propose an (1) extendable and (2) versatile glyph based visual analytics approach for anomaly detection in multivariate sensor data as our second contribution. The extendibility and versatility is achieved by making underlying data analysis methods exchangeable and flexible by a variable number of anomaly detectors. Additionally, we applied a technique enabling users to visually identify conspicuous sensor data by a matrix representation for drill-down and further comparative analysis. As a constraint, the concept of this work is only applicable for cyclic (also periodic or seasonal) data. Nevertheless, cyclic data is related to the repetitive behavior of many industrial applications. The approach has been designed for automotive testbed data with sensor-intensive technology, where anomalous or erroneous events should be anticipated by visual data analysis. The third contribution of this paper are results of the pair analytics evaluation [2], which has been conducted in collaboration with the target user group on the given use case data set. Results are encouraging and open promising directions for future work.

2 RELATED WORK

This section discusses the related work conducted in analyzing data using either algorithmic data analysis or visual analytics approaches. Furthermore, we detail the glyph representation as this technique has been proven to be an effective manner to represent time series.

2.1 Automated data analysis approaches for anomaly detection

One property of many typical industrial applications is the repetition of specific tasks. To give an example, Maier et al. [21] emphasized reoccurring processes (cycles) for automation and production. In theory, data generated within such cycles should be highly comparable for anomaly detection. Anomalies are generally understood to represent patterns in data that do not conform to a well-defined normal behavior [1]. The literature provides a comprehensive collection of algorithms for the detection of anomalies in multivariate time series data. Anomaly detection also often refers to the term novelty detection [23] or semi-supervised learning [6], whereas for those methods the definition of normal or rather reference data is needed. Anomaly detection in time series [16] found attention from industry for several applications, such as predictive maintenance [17], condition monitoring [40] or decision support systems [28]. Two groups of anomaly detection algorithms are used in this work. The first group are correlation-based approaches, which have been effectively applied on industrial sensor data [40]. The idea behind

this approach is, that changes of the bivariate correlation between two sensors can be interpreted as an anomaly. The second group are regression-based methods [12] or reconstruction-based novelty detection methods [23]. The basic idea of this type of anomaly detection is to build a regression model on reference (normal) data. Estimation errors of the model in comparison to measured data are rated as anomalies if they exceed a predefined threshold.

2.2 Visual analytics for industrial application

Recently, a survey on visualization and visual analytics applications for smart manufacturing has been published [42]. It reveals the diversity of several studies performed for industrial applications and the need of visual analytics. A few examples are available, on how to solve the problem of finding anomalies in multivariate time series data by visual analytics. An application for finding anomalies in the power consumption of buildings has been proposed by Janetzko et al. [18]. It suggests a model-based and a similarity-based anomaly score and visualizes them in several visualization techniques such as, recursive patterns [19], spiral graphs [34] and line charts. In the work of Wu et al. [35] anomalies are detected for condition monitoring by a model-based approach. The deviation of estimated and real values is visualized in a river plot view [9]. As an ongoing challenge, the authors outline the problem of analysts to trust and make use of the algorithms for condition monitoring. Many different algorithms are available for several applications and finding appropriate models and parameters is a hard task. Considering this problem, Xia et al. [36] proposed a visual analytics application to support users in finding the right model for dimensionality reduction. Another work addressing this challenge is presented by the EnsembleLens [38]. It is a visual analytics system to help data mining experts to evaluate, compare and select available anomaly detection algorithms.

2.3 Glyph representation of cyclic time series data

Besides the major summaries and surveys [7] [25], recently a systematic review of experimental studies on data glyphs has been presented by Fuchs et al. [14]. To visualize multivariate time series data, glyphs are an appropriate choice and can enable quick visual comparison of data values over time [13]. In Ward and Lipchak [33], the visualization of a circular glyph for recognition of the evolution of a measurement of interest has been proposed. Another glyph-based design for outlier detection in social networks has been proposed by Cao et al. [10]. In their work, glyphs visualize suspicious behavior of users, based on the z-score of several attributes, in a star-glyph like design. The anomaly scores of entities are visualized by the intensity of the red color in the cyclic glyph center. As examples for glyph-based time series visualizations, a few techniques for glyph designs for comparison purposes are evaluated in the work of Fuchs et al. [13].

3 BACKGROUND ON AUTOMOTIVE TESTBEDS AND USE CASE

Our work has been conducted for automotive engines in the context the validation and verification phase of the industrial PLC [3] [30]. After an engine has been developed, its requirements are verified and validated in automotive testbed environments. Those requirements can be of functional nature, such as the

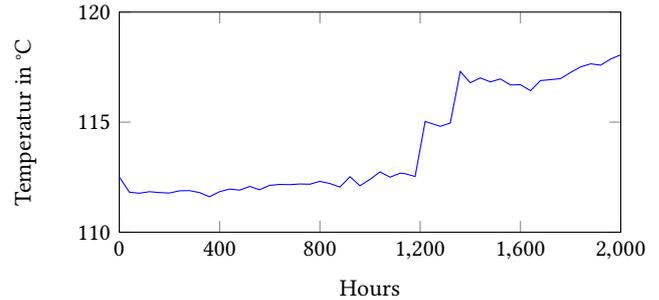


Figure 1: Temperature incident

Temperature of the critical component over all cycles. For each cycle over the whole durability test time the mean temperature has been calculated for trend analysis.

engine power density, speed and durability, or of legal nature, such as, along with others, fuel economy, noise pollution and exhaust gas emissions. For this research work, we analyzed data from a durability test of an internal combustion engine. The main goal of the test is to ensure the durability, reliability and life time expectations of the engine. Therefore, durability tests are conducted to let the engine undergo sufficiently high mechanical and thermal loads (stresses) and a sufficient number of fatigue cycles (e.g., hundreds of hours) [37]. During durability tests, a vast amount of data is collected by sensors which either are commonly build in modern vehicles and accessed through the engine control unit (ECU), or sensors which have been mounted on the engine and the testbed for testing purposes.

Throughout durability tests, engineers are observing the test, and are responsible for the performance and the condition of the testee. For that condition monitoring task, engineers generally monitor a few familiar sensors for threshold violations, manually selected and defined by their given domain knowledge or by the customer. However, for novel engine design, which has been recently developed, there is no knowledge on all sensors and their thresholds, and it is often not appropriate to apply earlier experiences. In fact, an important task in testing of novel engine designs is the comparison for differences with previous designs. Therefore, our work is motivated by making use of the time series data of all sensors and the information they might contain.

To give a practical example of the challenges engineers face during durability tests, in Figure 1 a line plot describing the problem of a use case is shown. After 1,200 hours of a 2,000 hours durability test a fatal error occurred, which increased the temperature of a critical component part of the engine by up to 8 °C over time. In the end, the durability test failed because of the temperature increase of the critical component. The failure could not be anticipated, because of the large amount of sensors. Consequently, it was a hard task to define rule-based thresholds or measurements of concern, indicating the failure upfront. However, in comparison to a simple rule-based anomaly detection approach, more complex data analysis and models which also take the interplay of sensor data into account may lead to better analysis results. Domain experts assume that it should be possible to anticipate such failures by advanced data analysis and visualization techniques. Therefore, to understand the current data analysis workflow of engineers, we carried out a design study as basis for our proposed visual analytics application. More details on the data and the design study will be given in the next section.

4 DESIGN STUDY

As the first contributions of this paper, we characterize testbed data and studied how engineers fulfill their condition monitoring task through data analysis. This section relates to Miksch and Aigners “design triangle” [22] and is generally based on the design study methodology of Sedlmair et al. [27]. For the tasks aspects of the design triangle we bridge from goals to tasks with the design study analysis report as proposed by Lam et al. [20].

4.1 Data

Input data used in our proposed visual analytics approach for anomaly detection is taken from an automotive engine testbed. One common task within engine development is to carry out durability tests. For such a durability test, a test cycle is specified to verify the durability of an engine. The test cycle, which is defined by a given engine speed and engine torque profile over time, is repeated in the period of several month until the target operating hours are reached. During a durability test, hundreds of sensor measurement signals are acquired and stored continuously, while the engine drives the given profile. In the automotive domain, those sensor measured time series are called channels, which we adopted throughout our research. Among others, channels mainly record several engine speed, engine torque, temperatures, pressures and exhaust gas measures.

One cycle is stored within one file and can be seen as a $N \times n$ -dimensional matrix, where N is the number of channels and n the length of the time series. All signals originally are recording numerical values in a frequency of 10Hz. Note that channels are aligned according to the given engine speed profile and therefore cycles of the same length can be extracted. The target data contains records of $c = 860$ cycles of a 2,000 hours’ durability test, in which each cycle has a duration of 140 minutes. Overall, the dataset has 860 cycles x 480 channels x 84,000 numerical values.

To conclude, the dataset can be characterized as a cyclic, piecewise equi-length segmented, multivariate streaming time series (accordingly to the characterization of Shurkhovetsky et al. [29]). For our research work data could be seen as stationary, since the durability has finished and the entire dataset was available for our work. Although, we kept a streaming time series scenario in mind where engineers use our proposed visual analytics approach throughout a durability test.

4.2 Users

Users of our proposed visual analytics application are development engineers with mechanical engineering background, working with powertrains and engines on a regular basis. They have long-standing experience with engines in testbed environments and as front-line analyst, also practically analyzing data to achieve their analysis goals (i.e., condition monitoring). Three users collaborated in our project systematically by participating at the design study and the pair analytics evaluation (see section 4 and section 7). In general, the work with testbed data is essential for development engineers and offers the opportunity to measure indicators regarding functional or legal requirements and engine performance. During our research work, we also collaborated with data scientists who are daily working with testbeds and powertrains. They constantly provided informal feedback from a different view throughout our work.

4.3 Tasks

This design study is based on the domain question, if an engine is in a non-critical condition during a durability test. For that purpose, we define test cycles as the population unit (or entity, or unit of analysis) [20]. Furthermore, engineers have a high level of understanding of using cycles as a granularity level for their analysis. Due to the repetitive behavior of cycles, they are highly comparable, and therefore we define the data analysis goals of engineers as multiple population analysis. Consequently, we identified that engineers are pursuing all three multiple population goals defined in the design study analysis report framework [20]: (a) compare entities (b) explain differences and (c) evaluate hypothesis. In the following, we further investigate the characterization of those goals by their input, output and analysis steps.

4.3.1 Compare Entities. Engineers attempt to detect population differences as their top level analysis goal. In order to achieve that, engineers are observing trends of several familiar channels by calculating the mean values of channels per cycle and visually explore changes over cycles in a time-ordered line plot (see Figure 1). In trend charts, up to ten time series are compared either in juxta- and/or superpositioned lineplots [15]. The output of the compare entities analysis goal, is the observation of a conspicuous trend or anomaly, which is investigated in the explain differences goal.

4.3.2 Explain Differences. Regarding the domain question, the output of the explain differences goal can be either that the observation is not relevant for the engines condition, or as a hypothesis one specific component of the engine is in a bad condition. In the next analysis goal, the hypothesis needs to be evaluated.

4.3.3 Evaluate Hypothesis. Evidence that a component or part of the engine is in a bad condition needs to be evaluated by engineers. The analysis steps of that goal don’t differ from the preceding goal. Domain experts are exploring channel time series by their domain knowledge and attempt to find differences and interesting patterns by comparing different channel line plots of multiple populations. The final confirmation or rejection of hypotheses is made after evidence has been collected through data analysis and if required investigated directly on the engine.

Through that characterization of higher level analysis goals, we can derive lower level task definitions **T1 - T5** to address them in our visual analytics design considerations and automated data analysis:

T1 Identify population contrasts. Test cycles are the unit of analysis, or population, for engineers. As the first task, population contrasts or differences are explored. This is achieved by trend analysis of a few familiar channels. The problem of dealing with big channel amounts engineers face should be considered in the design by taking all channels into account.

T2 Application and visualization of semi-supervised anomaly detection methods. Engineers detect interesting patterns and anomalies mainly by visually exploring line plot trends of channels. Comparing past cycles to current cycles is related to a semi-supervised learning scenario and should be considered for the choice of automated data analysis and the visualization. First, the automation of data analysis to highlight interesting or conspicuous channels should be included into the visualization. Second, we assume that the combination of several anomaly detection algorithms leads

to more significant findings, for which reason an ensemble method [38] should be considered for the visualization.

T3 Examine conspicuous channels in multiple populations.

After conspicuous channels have been detected by trend analysis, engineers drill-down to examine and find differences between channel line plots of different populations. The comparison of interesting channels in different populations should be considered in the design.

T4 Detection of conspicuous channel relations.

With exceptions on visualizing multiple trend line plots in juxtaposition, engineers generally detect anomalies by univariate time series analysis of multiple channels. They, also compare line plots with the given engine speed and engine torque. Considering relations between channels at a broader scale should be considered for the automated data analysis and visualization.

T5 Reduce amount of data.

To address the large amount of channels, data reduction techniques should be taken into account for the visual analytics design. The main consideration for the visual analytics design is that interesting data should be highlighted to support engineers in their decision making.

Overall, data can only be analyzed by including extensive domain knowledge of users to the data analysis. Modern powertrains and engines are highly complex machines and therefore domain experts are necessary to interpret results of automated data analysis through a visual analytics approach. In the next section we introduce the anomaly detection methods we applied to the visual analytics approach.

5 AUTOMATED DATA ANALYSIS: ANOMALY DETECTION METHODS

In research, anomaly detection often refers to a two-class classification problem, in which data either is classified as an anomaly or not. In general, a model is built on normal data, considering that the model can calculate an anomaly score on unseen data sets (apart from unsupervised methods). If the anomaly score exceeds a predefined threshold, the data record or the entire set is classified as an anomaly. We consider the application of semi-supervised anomaly detection methods to our design T2. Most techniques are specific to different observational features, in consequence of which we assume that an ensemble-based approach obtains more robust anomaly scores [38]. Therefore, we propose to map results of different anomaly detection algorithms to a unified value for comparison purposes and describe two anomaly detection methods, used in the visual analytics approach.

5.1 Unified anomaly score

Test cycles are engineer's unit of analysis for which reasons we choose them as the granularity level for data analysis (T1). To make different anomaly detection methods comparable in an ensemble-based approach, we propose the following to map anomaly scores to unified values between 0 and 1: (1) Interactively select a reference cycle as input data for the training of the anomaly detection model. (2) A baseline cycle is selected to calculate a baseline anomaly score. (3) Define a threshold anomaly score, based on prior knowledge, domain knowledge or historical data. (4) Further, the anomaly score of cycles are calculated as the linear scaling from 0 (baseline) to 1 (threshold). Therefore, our approach needs the definition of a reference cycle for model training, a cycle for baseline definition and the definition of a threshold. The baseline anomaly score is used to consider a training error and therefore is taken as the lower limit of the unified

anomaly score calculation. Note, that the baseline anomaly score is calculated by a baseline cycle, which data should be recorded temporally close to the reference cycle. Hence, we specify the cycle subsequent to the reference as the baseline cycle. Defining the upper limit (threshold) is critical and can be changed interactively in the visualization. In the following, we discuss two methods, which have been applied on industrial sensor data in prior research. Another criterion for selecting those two methods is the capability of identifying conspicuous channels separately for further drill-down and comparative analysis (T3).

5.2 Correlation-based anomaly score

Inspired by the approach presented by Zhao et al. [40], we assume that the change of linear correlations between two sensors over time refers to an anomaly. During our research work, we investigated the application of correlation-based anomaly detection on testbed data. Despite its limitations, the method has also strengths that are beneficial in solving the tasks defined in the design study. First, we highlight the main limitation in detecting anomalies by the change of linear sensor correlations, if throughout the durability test no linear correlation between two specific sensors exist. Nevertheless, we examined that in testbed data many linear correlations between sensors exist. For example, data from several temperature channels are likely to correlate. Experiments demonstrated that this method can detect anomalies in testbed data, and therefore has been applied to our visual analytics approach.

The basis for the correlation-based anomaly score is the correlation difference matrix, which represents the deviation of linear channel relations between two testbed cycles. The correlation matrix for each sensor combination of the reference cycle, the baseline cycle and the unseen cycle are calculated by using Pearson's correlation coefficient. Then, the correlation matrix of the unseen cycle is subtracted by the correlation matrix of the reference cycle, which results in the correlation difference matrix. As a result, the anomaly score is calculated as the average of all values in the difference matrix and is mapped to the unified anomaly score accordingly to the method explained before.

5.3 Regression-based anomaly score

As the second anomaly score, we make use of regression models for regression-based anomaly detection [12]. For this research work, we train regression models to estimate a time series. Consequently, the model is applied to an unseen data set, in which the difference between the estimation and the real values (residuals) can be interpreted as anomalies. Considering this method for T1 and T2, an anomaly score between populations or cycles needs to be calculated in a semi-supervised manner. Therefore, regression models with data of a user-defined reference cycle are trained for all channels separately. To make those channel regression models comparable, it is necessary to standardize data first, i.e. standardization of the entire time series to values between 0 and 1. The regression models can now be used to estimate all channel time series for unseen cycles and the anomaly score of one cycle can be calculated by the average mean average error over all channels of a cycle. In the following it can be mapped to a unified anomaly score accordingly to the method explained above.

We chose Random Forest regressor, as suggested by Breiman [8], considering that this model has been proven to perform well

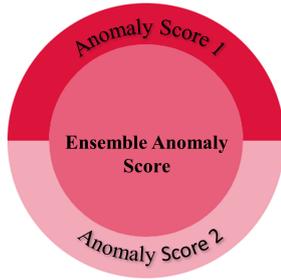


Figure 2: Proposed glyph design

The glyph visualizes two anomaly scores and their ensemble (aggregate). Anomaly Score 1 visualizes a value of 0.8, Anomaly Score 2 visualizes a value of 0.2, whereas the equally weighted center visualized the Ensemble Anomaly Score of 0.5.

in many domains [41]. As input data for training channel regression models, engine speed and engine torque are chosen, since those two channels are given by the test and strongly relate to the majority of the channels (T4). Also, sliding window features for those two channels are extracted, whereas sliding windows contain differences and mean values of three seconds into the past. We assume that in this time frame the most relevant information can be extracted for our models. The aim of this approach is not to estimate each channel as accurate as possible, but to detect change of anomaly scores between populations. As the correlation-based method, we are aware of the limitation that this method may not return a decent estimation for all channels, but it may be effective for some types of anomalies. This consideration should also emphasize the choice of an ensemble method.

6 VISUAL ENCODING AND CONSIDERATIONS

This section explains how we use two anomaly scores for a glyph-based visualization. Also, an example on how to identify conspicuous channels within a cycle either in a matrix representation and a ranked channel list is given. The visual considerations explained in this section will be brought together in the prototype, describing the visual analytics approach by the prototype implementation.

6.1 Cycle anomaly glyph

The proposed glyph in Figure 2 is flexible and independent of the underlying analytical methods for anomaly detection, as long as it implements the framework for calculating unified anomaly scores between 0 and 1 (subsection 5.1). Anomaly scores are visualized in the outer circular segments of the glyph representation as the opacity value of the red background color. Our aim, when designing the glyph was that no algorithm can detect all kind of anomalies, relevant for different applications. As a result, we choose an extensible glyph design, achieved by its circular shape, which offers the capability of adding and removing anomaly scores in their according circular segments. The main visual focus of the glyph stays at the center circle, which represents an equally weighted average of anomaly scores combined, labeled as the ensemble anomaly score.

During our work, the main concern of visualization experts regarding the presented glyph design was the benefit compared to simpler visualizations, such as line plots. As stated in the design study, line plots are a well-known visualization type and

comprehensible to the target group. However, our approach has advantages over line-plot-based visualizations. In general, testbed cycles as granularity level are highly comprehensible for engineers. Therefore, cycles are visualized as individual and complete entities, whereas the glyph design offers the following opportunities: (1) As a visual entity it can be clearly selected by users for further exploration, reasoning and drill-down. Also, glyphs can be selected interactively to be defined as reference for the underlying semi-supervised learning algorithms to identify contrasts between populations (T1, T2). (2) The glyph design can be extended with several anomaly detection algorithms by adding additional outer circular segments. (3) Similar glyphs can be aggregated, clustered and arranged for further interpretation by domain experts and to save screen space. (4) The glyph can be visualized on its own as a quick overview of an engine's condition. This is also related to the idea of involved engineers of having a simple "traffic light like" system, which also encouraged us, developing the presented glyph-based approach.

6.2 Identification of anomalous channels

As stated above, we choose two anomaly scores by their capability to further explore single anomalous channels. After an anomalous channel has been identified in the glyph representation, users are interested in the cause of that anomaly. Therefore, we visually represent anomalies for both anomaly scores, as follows:

Matrix-based identification of anomalous channels. The correlation deviation matrix calculated for the correlation-based anomaly score is shown in step (c) of Figure 3. Basically, in this symmetrical matrix, deviations of correlations of channels within a given cycle with respect to the selected reference cycle are visualized. More specifically, we compute the difference of the correlation matrices of these two cycles, and show the result by color-coding the cells of the difference matrix. Hence, levels of red representing the anomaly score of channel correlations. This matrix representation supports the analysis goal to determine and quantify visual patterns for pattern-driven visual exploration. Together with appropriate matrix reordering methods, we can use this display to search for typical patterns in matrix visualizations, including line patterns and block patterns [4]. Most importantly, if one sensor shows an anomalous behavior, its correlation difference values to many or all other sensors will be rather large, leading to line patterns. Such visual patterns attract the attention of the analyst and are a starting point for drilling-down into the respective sensor data (Figure 3 (e2)).

Ranked mean average error list. The regression-based anomaly score can be explored by the ranked mean average error list as proposed in Figure 3 (d). Channels that deviate from the reference are listed and ranked by their anomaly score. This enables a guided approach for exploring anomalies and simplifies data analysis. By clicking on channel names users can explore the reference and the anomalous channel time series by visually comparison in juxtaposition for hypothesis generation (Figure 3 (e1)).

6.3 Prototype

The workflow of the approach, applied to data of the given use case, is exhibited and briefly described in Figure 3. It shows screenshots of the implemented prototype, whereas further explanations are given in the following: In (a) glyphs are placed in a grid, with each cell representing a test cycle in chronological

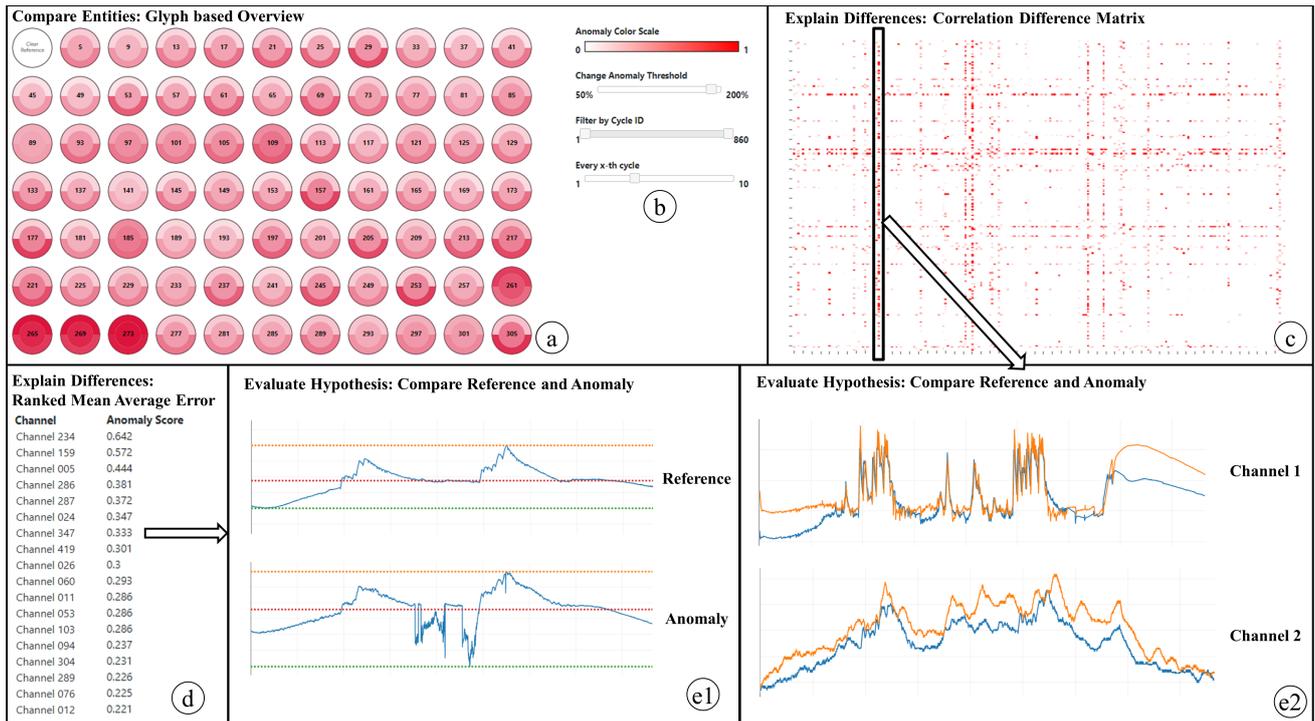


Figure 3: Proposed visual analytics tool

(a) Differences between the selected reference cycle and other cycles can be explored, whereas glyphs are positioned in a time ordered grid. (b) Besides some filter capabilities, the anomaly score threshold can be interactively changed by a ruler. (c) Anomalies found by the correlation-based anomaly score can be explored in the matrix representation (d) Anomalous channels found by the regression-based anomaly score can be explored by the ranked mean average error list. (e) Hypothesis can be evaluated by comparing channel time series in the reference cycle and the cycle of interest.

order (from top left to bottom right, inspired by the calendar-based view [32]). Note, that the reference cycle is interactively selectable and represented by a white circle, as visible in the top left, or first, glyph. In (b) three interaction possibilities are visible: (i) to add flexibility to the visual comparison of glyphs the user can interactively change the anomaly score threshold to values between 50% - 200% of the original value (ii) the user can change the amount of displayed glyphs by filtering them by a 'from - to' range slider (iii) glyphs can be filtered by the definition to visualize every x^{th} glyph only. Both anomaly scores of interesting glyphs can be selected for further exploration by a drill-down in (c) and (d): In (c) a drill-down example to inspect and identify one or more conspicuous channels within the selected cycle by a matrix representation visualizing the correlation-based anomaly score is shown. An example for a visual perceptive line pattern is outlined, representing a possibly conspicuous channel. Further, the conspicuous channel can be selected in the matrix for exploration and comparative analysis with the reference line plot in (e2). In (d) an example of the ranked mean average error representation of the regression-based anomaly scores is given, in which its drill-down capabilities are visible in (e1). In general, drill-down information needs to be investigated and interpreted by domain experts. However, our approach supports users in the identification of interesting data by visually highlighting deviating cycles and sensors. As a side note, interactive line plots and heatmaps in the prototype have been created with the JavaScript visualization library Plotly.js [24] and are anonymized in Figure 3 screenshots.

7 EVALUATION

We conducted a pair analytics evaluation [2] with three subject matter experts (SME), who represent the target user group identified in subsection 4.2, and the dataset described in subsection 4.1. The main target was to evaluate either the comprehensibility of the different views and the underlying automated data analysis, along with the capabilities and limitations in supporting users with their daily condition monitoring analysis goals. According to the pair analytics protocol, the evaluation is done by a human-to-human interaction of one SME and one visual analytics expert (VAE), in which the SME acts as the navigator and the VAE as the driver (operator) of the visual analytics tool. In general, all three SME participants stated that the visual analytics tool can be of great benefit to support them in their daily work for two reasons: The visual analytics tool supports engineers in analyzing testbed data (1) more efficiently by highlighting interesting data on different granularity levels and (2) more effectively by enabling the analysis of the entire dataset and not only a subset of well-known channels. To give evidence to that statement, we connect participants comments and actions during the pair analytics evaluation to the task definition of subsection 4.3 in the following:

Each evaluation session started by an introduction to the visual analytics approach and a short demonstration of the prototype. It is notable, that all three participants (P1, P2, P3) gained a quick understanding of the concept for two reasons: First, we conducted the design study with the same engineers and connected findings of the study with explanations of our visual analytics tool. Second, the design study clearly identifies tasks and goals of engineers,

therefore the visual analytics prototype accurately addresses the needs of engineers. In general, participants appreciated our effort in developing a decision support system supporting engineers in handling the big amount of data for their condition monitoring tasks.

The actual pair analytics evaluation sessions started by defining a reference cycle in the glyph-based overview (T2). P1 and P3 appreciate the capability of selecting the reference cycle interactively in the visualization. However, P2 questioned the necessity of interactively selecting the reference cycle, because the testee is likely to be in a good condition before the first cycle, considering that the testee runs through an extensive health check at the beginning of the entire durability test. We are aware of the fact that selecting the first cycle may be an appropriate default choice, but we wanted to keep the analysis more flexible.

After the reference cycle has been selected, other glyphs in the overview turned red regarding their anomaly scores (see Figure 1 (a)). All participants were immediately curious in exploring those anomalies by the visualization and easily identified cycles that appear interesting to them (T1). P2 and P3 pointed at cycles that had a more intensive color of red than the majority of all cycles, whereas P1 mentioned that all glyphs that visualize at least a small anomaly score are interesting. However, in a productive use scenario the exploring strategy may differ since not all cycles are available from the beginning, and new data would be explored incrementally on a regular basis as it becomes available. We emphasize that at this stage of the visual data analysis we successfully reduced the amount of data (T5) and enabled the further exploration of anomalies in the succeeding views.

T3 and T4 are both achieved by exploring one of the two anomaly scores of a specific cycle: (1) The correlation-based anomaly score and the correlation difference visualization in Figure 1 (b) were comprehensible to the participants as they were able to identify conspicuous channels. However, participants articulated the need of a more guided approach to engage engineers using the matrix visualization, because it appears overloaded and thus overwhelming to engineers. (2) The regression-based anomaly score was also highly comprehensible to participants, since they have a general understanding of regression models. On the other hand, we avoided explaining the actual regression model Random Forest to participants in detail. In comparison, to the correlation difference matrix, participants commented that the exploration of conspicuous channels is easier by the ranked mean average error list (Figure 2 (c)). Also, they expressed their interest of additional guided approaches in the other views, considering that such rankings represent a clear order on what channels to focus on, especially if they are short of time during their analysis.

As the last step of the visual analysis, participants evaluated hypothesis of channels being anomalous by comparing anomalous line plots with their reference cycle equivalents (see Figure 1 (e1 + e2)). From a data perspective, engineers approved that all explored anomalies are interesting, because they highlight a significant difference to the reference. From a domain perspective, some of the anomalies were interesting, but others were explainable and irrelevant for the condition monitoring task. Another type of anomaly, that has been detected during evaluation are defect or unconnected sensors. Line plots of those anomalies visualize a constant or noisy signal. Therefore, we characterize three types of anomalies that have been found during evaluation: (1) Domain irrelevant (2) Domain relevant and (3) Defect sensors.

Overall, we evaluated that the visual analytics prototype receive acceptance from all participants. They confirm the benefit

of the proposed visual analytics approach and are interested in using the prototype in a productive scenario. In the next section, we discuss some of the aspects of the evaluation in greater detail and also consider generalizability and future work.

8 DISCUSSION AND FUTURE WORK

Our visual analytics approach has been designed and evaluated on testbed data, but we emphasize that it is not limited to the automotive domain. At least the glyph-based overview should be applicable on any other cyclic multivariate data set, as long as the underlying automated data analysis methods and dependent visualization techniques are adapted to the specific domain. The visual analytics prototype has been evaluated to be useful for collaborators and they clearly identified advantages in terms of efficiency and effectiveness in comparison to their current workflow. Also, the evaluation opened up many directions for future work: Analyzing up to a thousand of cycles can be critical regarding the screen space. Applied filtering techniques in Figure 3 (b) can be improved by a more scalable solution and clearly needs further attention. For example, similar glyphs can be aggregated to save space on the display. We also address the scalability for the matrix-representation for future work. As a generic abstraction of anomalies in the glyph-based overview, the calculations of anomaly scores are exchangeable and more extensive research on additional available anomaly detection methods for the use case needs to be done. The evaluation demonstrated that anomalies can be characterized in three manners. Hence, engineers should be able to provide feedback on their findings, i.e., to classify the relevance of anomalies consequently for analyzing data in further iterations more efficiently. Even if the target users are not data mining experts, we experienced that they gain a quick understanding of the proposed workflow. However, for future work we will address guidance for visual analytics [11] to reduce system complexity from the user perspective and support users to further achieve their analysis goals. One promising venue we see to this end is the application of visual interestingness measures [5] to automatically select cycle pairs from the database showing significant visual patterns.

9 CONCLUSION

We propose a visual analytics approach to improve the engineer's daily work by the glyph-based visual analytics workflow. We have found promising results on the given use case, but the concept still needs to be proven with other datasets. We contribute to the need of visual analytics approaches for condition monitoring or anomaly detection in cyclic time series data. Further, our approach devises a methodology to reduce large amounts of industrial data, by drawing attention to anomalous cycles on a higher granularity level to increase efficiency and effectiveness of engineers' data analysis work.

ACKNOWLEDGMENTS

This research work is done by Pro²Future and AVL List. Pro²Future is funded within the Austrian COMET Program-Competence Centers for Excellent Technologies- under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry for Digital and Economic Affairs and of the Provinces of Upper Austria and Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

REFERENCES

- [1] Shikha Agrawal and Jitendra Agrawal. 2015. Survey on Anomaly Detection using Data Mining Techniques. In *Procedia Computer Science*, Vol. 60. Elsevier, 708–713.
- [2] Richard Arias-Hernandez, Linda T Kaastra, Tera M Green, and Brian Fisher. 2011. Pair Analytics: Capturing Reasoning Processes in Collaborative Visual Analytics. In *44th Hawaii International Conference on System Sciences*. IEEE, 1–10.
- [3] Eric Armengaud. 2017. Industry 4.0 as Digitalization over the Entire Product Lifecycle: Opportunities in the Automotive Domain. In *Systems, Software and Services Process Improvement, 24th European Conference, EuroSPI 2017*. Springer, 334–351.
- [4] Michael Behrisch, Benjamin Bach, Nathalie Henry Riche, Tobias Schreck, and Jean-Daniel Fekete. 2016. Matrix Reordering Methods for Table and Network Visualization. *Computer Graphics Forum* 35, 3 (2016), 693–716.
- [5] Michael Behrisch, Benjamin Bach, Michael Hund, Michael Delz, Laura von Räden, Jean-Daniel Fekete, and Tobias Schreck. 2017. Magnostics: Image-Based Search of Interesting Matrix Views for Guided Network Exploration. *Transactions on Visualization and Computer Graphics* 23, 1 (2017), 31–40.
- [6] Gilles Blanchard, Gyemin Lee, and Clayton Scott. 2010. Semi-supervised novelty detection. *Journal of Machine Learning Research* 11 (2010), 2973–3009.
- [7] Rita Borgo, Johannes Kehler, David H. S. Chung, Eamonn Maguire, Robert S. Laramée, Helwig Hauser, Matthew Ward, and Min Chen. 2013. Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications. *Eurographics 2013 - State of the Art Reports* (2013), 39–63.
- [8] Leo Breiman. 2001. Random Forrest. *Machine Learning* 45 (2001), 5–32.
- [9] P. Buono, C. Plaisant, A. Simeone, A. Aris, B. Shneiderman, G. Shmueli, and W. Jank. 2007. Similarity-Based Forecasting with Simultaneous Previews: A River Plot Interface for Time Series Forecasting. In *11th International Conference Information Visualization*. IEEE, 191–196.
- [10] Nan Cao, Conglei Shi, Sabrina Lin, Jie Lu, Yu Ru Lin, and Ching Yung Lin. 2016. TargetVue: Visual Analysis of Anomalous User Behaviors in Online Communication Systems. *IEEE Transactions on Visualization and Computer Graphics* 22 (2016), 280–289.
- [11] Davide Ceneda, Theresia Gschwandtner, Thorsten May, Silvia Miksch, Hans-Jörg Schulz, Marc Streit, and Christian Tominski. 2017. Characterizing Guidance in Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics* 23 (2017), 111–120.
- [12] Varun Chandola, Arindam Banerjee, and Vipin Kumar. 2009. Anomaly Detection: A Survey. *ACM Computer Surveys* 41, 3, Article 15 (2009), 58 pages.
- [13] Johannes Fuchs, Fabian Fischer, Florian Mansmann, Enrico Bertini, and Petra Isenber. 2013. Evaluation of Alternative Glyph Designs for Time Series Data in a Small Multiple Setting. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, 3237–3246.
- [14] Johannes Fuchs, Petra Isenber, Anastasia Bezerianos, and Daniel Keim. 2017. A Systematic Review of Experimental Studies on Data Glyphs. *IEEE Transactions on Visualization and Computer Graphics* 23 (2017), 1863–1879.
- [15] Michael Gleicher. 2018. Considerations for Visualizing Comparison. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 413–423.
- [16] Manish Gupta, Jing Gao, Charu C Aggarwal, and Jiawei Han. 2014. Outlier Detection for Temporal Data: A Survey. *IEEE Transactions on Knowledge and Data Engineering* 26, 9 (2014), 2250–2267.
- [17] Clemens Gutsch, Nikolaus Furian, Josef Suschnigg, Dietmar Neubacher, and Siegfried Voessner. 2019. Log-based predictive maintenance in discrete parts manufacturing. In *12th CIRP Conference on Intelligent Computation in Manufacturing Engineering*, Vol. 79. Elsevier, 528–533.
- [18] Halldór Janetzko, Florian Stoffel, Sebastian Mittelstädt, and Daniel A. Keim. 2014. Anomaly detection for visual analytics of power consumption data. *Computers & Graphics* 38 (2014), 27–37.
- [19] Daniel Keim, H.-P Kriegel, and Mihael Ankerst. 1995. Recursive pattern: A technique for visualizing very large amounts of data. In *Proceedings of the IEEE Visualization Conference*. 279–286.
- [20] Heidi Lam, Melanie Tory, and Tamara Munzner. 2018. Bridging from Goals to Tasks with Design Study Analysis Reports. *IEEE Transactions on Visualization and Computer Graphics* 24 (2018), 435–445.
- [21] Alexander Maier, Tim Tack, and Oliver Niggemann. 2012. Visual Anomaly Detection in Production Plants. In *Proceedings of the 9th International Conference on Informatics in Control, Automation and Robotics*. SciTePress, 67–75.
- [22] Silvia Miksch and Wolfgang Aigner. 2014. A matter of time: Applying a data–users–tasks design triangle to visual analytics of time-oriented data. *Computers & Graphics* 38 (2014), 286–290.
- [23] Marco AF Pimentel, David A Clifton, Lei Clifton, and Lionel Tarassenko. 2014. A review of novelty detection. *Signal Processing* 99 (2014), 215–249.
- [24] Plotly Technologies Inc. 2015. *Collaborative data science*. Montreal, QC. <https://plot.ly>
- [25] Timo Ropinski, Steffen Oeltze, and Bernhard Preim. 2011. Survey of glyph-based visualization techniques for spatial multivariate medical data. *Computers & Graphics* 35 (2011), 392–401.
- [26] Dominik Sacha, Andreas Stoffel, Florian Stoffel, Bum Chul Kwon, Geoffrey Ellis, and Daniel A. Keim. 2014. Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics* 20 (2014), 1604–1613.
- [27] Michael Sedlmair, Miriah Meyer, and Tamara Munzner. 2012. Design Study Methodology: Reflections from the Trenches and the Stacks. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2431–2440.
- [28] Maxim Shcherbakov, Adriaan Brebels, N.L. Shcherbakova, V.A. Kamaev, O.M. Gerget, and D. Devyatykh. 2017. Outlier detection and classification in sensor data streams for proactive decision support systems. *Journal of Physics: Conference Series* 803 (2017).
- [29] Georgiy Shurkhovetskiy, N Andrienko, G Andrienko, and Georg Fuchs. 2018. Data Abstraction for Visualizing Large Time Series. *Computer Graphics Forum* 37, 1 (2018), 125–144.
- [30] Stefan Thalmann, Gursch Heimo, Josef Suschnigg, Milot Gashi, Helmut Ennsbrunner, Anna Katharina Fuchs, Tobias Schreck, Belgin Mutlu, Jürgen Mangler, Gerti Kappel, Christian Huemer, and Stefanie Lindstaedt. 2019. Cognitive Decision Support for Industrial Product Life Cycles: A Position Paper. In *Cognitive 2019*, Vol. 11. IARIA, 3–9.
- [31] Stefan Thalmann, Juergen Mangler, Tobias Schreck, Christian Huemer, Marc Streit, Florian Pauker, Georg Weichhart, Stefan Schulte, Christian Kittl, Christoph Pollak, Matej Vukovic, Gerti Kappel, Milot Gashi, Stefanie Rinderle-Ma, Josef Suschnigg, Nikolina Jekic, and Stefanie N. Lindstaedt. 2018. Data Analytics for Industrial Process Improvement A Vision Paper. In *IEEE 20th Conference on Business Informatics (CBI)*, Vol. 02. 92–96.
- [32] Jarke J Van Wijk and Edward R Van Selow. 1999. Cluster and calendar based visualization of time series data. In *Proceedings IEEE Symposium on Information Visualization*. 4–9.
- [33] Matthew O. Ward and Benjamin N. Lipchak. 2000. A visualization tool for exploratory analysis of cyclic multivariate data. *Metrika* 51 (2000), 27–37.
- [34] Marc Weber, Marc Alexa, and Wolfgang Müller. 2001. Visualizing time-series on spirals. In *IEEE Symposium on Information Visualization, 2001*. 7–13.
- [35] Wencho Wu, Yixian Zheng, Kaiyuan Chen, Xiangyu Wang, and Nan Cao. 2018. A Visual Analytics Approach for Equipment Condition Monitoring in Smart Factories of Process Industry. In *IEEE Pacific Visualization Symposium*. 140–149.
- [36] Jiazhi Xia, Fenjin Ye, Wei Chen, Yusi Wang, Weifeng Chen, Yuxin Ma, and Anthony K.H. Tung. 2018. LDSScanner: Exploratory Analysis of Low-Dimensional Structures in High-Dimensional Datasets. *IEEE Transactions on Visualization and Computer Graphics* 24 (2018), 236–245.
- [37] Qianfan Xin. 2013. 2 - Durability and reliability in diesel engine system design. In *Diesel Engine System Design*, Qianfan Xin (Ed.). Woodhead Publishing, 113–202.
- [38] Ke Xu, Meng Xia, Xing Mu, Yun Wang, and Nan Cao. 2019. EnsembleLens: Ensemble-based Visual Exploration of Anomaly Detection Algorithms with Multidimensional Data. *IEEE Transactions on Visualization and Computer Graphics* 25 (2019), 109–119.
- [39] Shen Yin and Okyay Kaynak. 2015. Big Data for Modern Industry: Challenges and Trends [Point of View]. *Proc. IEEE* 103, 2 (2015), 143–146.
- [40] Pushe Zhao, Masaru Kurihara, Junichi Tanaka, Tojiro Noda, Shigeyoshi Chikuma, and Tadashi Suzuki. 2017. Advanced correlation-based anomaly detection method for predictive maintenance. In *IEEE International Conference on Prognostics and Health Management*. 78–83.
- [41] Xun Zhao, Yanhong Wu, Dik Lun Lee, and Weiwei Cui. 2018. iforest: Interpreting random forests via visual analytics. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 407–416.
- [42] Fangfang Zhou, Xiaoru Lin, Chang Liu, Ying Zhao, Panpan Xu, Liu Ren, Tingmin Xue, and Lei Ren. 2019. A survey of visualization for smart manufacturing. *Journal of Visualization* 22 (2019), 419–435.