Eurographics Conference on Visualization (EuroVis) 2017
J. Heer, T. Ropinski and J. van Wijk
(Guest Editors)

Volume 36 (2017), Number 3

Cycle Plot Revisited: Multivariate Outlier Detection Using a Distance-Based Abstraction

M. Bögl[1], P. Filzmoser[1], T. Gschwandtner[1], T. Lammarsch[1], R. A. Leite[1], S. Miksch[1], and A. Rind[2]

[1]Vienna University of Technology, Austria [2] St. Pölten University of Applied Sciences, Austria

Abstract

The cycle plot is an established and effective visualization technique for identifying and comprehending patterns in periodic time series, like trends and seasonal cycles. It also allows to visually identify and contextualize extreme values and outliers from a different perspective. Unfortunately, it is limited to univariate data. For multivariate time series, patterns that exist across several dimensions are much harder or impossible to explore. We propose a modified cycle plot using a distance-based abstraction (Mahalanobis distance) to reduce multiple dimensions to one overview dimension and retain a representation similar to the original. Utilizing this distance-based cycle plot in an interactive exploration environment, we enhance the Visual Analytics capacity of cycle plots for multivariate outlier detection. To enable interactive exploration and interpretation of outliers, we employ coordinated multiple views that juxtapose a distance-based cycle plot with Cleveland's original cycle plots of the underlying dimensions. With our approach it is possible to judge the outlyingness regarding the seasonal cycle in multivariate periodic time series.

Categories and Subject Descriptors (according to ACM CCS): Mathematics of Computing [G.3]: Probability and Statistics—Time Series Analysis; Information Interfaces and Presentation [H.5.2]: User Interfaces—Graphical user interfaces

1. Introduction

In this paper we propose an interactive environment utilizing cycle plots to explore patterns and to detect multivariate as well as univariate outliers. For the construction of our distance-based cycle plot we use an abstraction based on a multivariate distance measure (Mahalanobis distance), to visualize patterns in multivariate seasonal time series, like trends and seasonal cycle. We build upon the established and effective *cycle plot* by Cleveland [Cle93], which is limited to univariate time series.

Time series often follow a periodically reoccurring pattern, called periodic or seasonal pattern. An example are monthly averages of temperatures over multiple years, with a yearly low, a yearly high, and smooth transitions in between. Such seasonal time series appear in various domains, like ecology, economics, or health. Examples of seasonal time series may be univariate, like number of influenza cases, but many real world examples are multivariate, like number of deaths caused by cardiovascular disease connected with air pollution data, or water quality measures [BD10].

An important objective in time series analysis is the detection of outliers, which in multivariate seasonal time series requires to consider seasonal pattern and trends, both for the several underlying variables and for the multivariate space. The cycle plot described by Cleveland [Cle93] is an effective visualization technique, which facilitates the identification of these seasonal pattern and trends in univariate data, and it allows for comparing data points within the same seasonal cycle (e.g., month of year) in close proximity. These subgroups within the seasonal cycle enable the detection of outliers and extreme values within the groups or whole groups that do not follow the behavior of the seasonal pattern. To achieve the same effect for multivariate seasonal time series, each variable can be represented by one original cycle plot. Although this allows the human analyst to analyze seasonal patterns, trends, and extreme/outlying values of each dimension, building an overview mentally by observing multiple cycle plots is a difficult, time-consuming, and unreliable task. Single data points may behave abnormal in just some or even none of these dimensions, but stand out in multivariate space. Furthermore considering multiple such cycle plots for each variable separately takes additional time and increases cognitive load to combine and transfer the individual dimensions in a multivariate mental model. For detecting anomalies like outliers in the multivariate space, the aid of further abstraction, introduced in Section 5, allows to ease this reasoning, as illustrated in Section 6 and discussed in Section 7. Empirical evidence for an increased performance of our approach is beyond the scope of this paper, but we consider comprehensive user studies of the task performance in future work. The main contributions of this paper are:

- The construction of a Mahalanobis-distance-based cycle plot that
  - involves an additional abstraction step based on generalized multivariate distances and

  – uses a modified visual encoding for these distances, but retains the idea of the original cycle plot.
- An interactive exploration environment of coordinated multiple views combining the distance-based cycle plot with original cycle plots to
  – identify outliers in multivariate time series considering the seasonality,
  – support the interpretation of multivariate outliers,
  – reduce the information loss inevitably accompanying the multivariate data abstraction.

2. Related Work

A variety of approaches have been proposed to visualize time-oriented multivariate data [AMST11]. Suitable approaches can be categorized into techniques, which provide (1) visualizations for multivariate data, mainly using projections and other aggregation methods, (2) visualizations, which take the structure of time into account, and (3) statistical methods for outlier detection.

Visualizations for Multivariate Data. A frequently used approach is using several line plots [Pla86] either in one coordinate system or as small multiples [Tuf83]. An alternative, already introduced by Playfair in 1786, is the stacked graph [Pla86, BW08]. Wu et al. [WWS*16] incorporate additional information in the stacked graph and discuss clustering and visual arrangement for detecting multivariate patterns. For using small multiples, space-efficient visualizations are well-suited, like Horizon Graphs [Rei08, Few08] and Qualizon Graph [FHR*14]. Javed et al. [JME10] compare the traditional line plot, small multiples, Horizon Graphs, and a new visualization called braided graphs. Another space-efficient visualization for multivariate data are CloudLines [KBK11], which are inspired by ideas of EventRiver [LYK*12]. Tominski et al. [TAS04] compare axes-based visualizations with radial layouts (for example, the time wheel) and discovered that all approaches are suitable for showing multiple variables at the same time and temporal trend detection, but are less appropriate for seasonal cycles. Another technique going back to Playfair et al. [Pla86] for the special case of bivariate data are connected scatter plots, for example Haroz et al. [HKF16]. A very similar concept called trajectories is used in small multiples by Schreck et al. [SBVLK09]. Visually similar are time curves [BSH*16] that project time series in a 2D space based on similarity measures. It is a strong visualization method for finding both regular and irregular temporal patterns. However, the projection makes it hard to compare the length of intervals, and there is no visual representation of the data underlying the distance measure.

The Structure of Time in Visualization. While the structure of time has many different aspects [AMST11], the aspects of granularities and cycles are the most important ones in the context of our work, as the cycle plot [Cle93, Cle94] supports them (see Section 3 for a detailed explanation). For pixel-based visualizations, the original work by Keim et al. [KKA95] (which also includes multivariate data), as well as related work by Van Wijk and van Selow [vWvS99], has been the basis for several further publications [SFdOL04, LAB*09, KJL14]. Borgo et al. [BPC*10] evaluate the performance of pixel-based visualizations according to the task complexity and cognitive load. Even though they only tested univariate data, we assume that including multiple variables is a task aspect that creates exactly a task complexity that worsens performance of pixel-based visualizations. Besides the visual representation of periodicity as in the approaches above, it is possible to isolate the seasonal component of time series and to perform further analysis such as detection of abnormal events [CTB*12] on residuals. Such seasonal time series models can as well be used for prediction [BAF*15, MHR*11].

Statistical Oriented Approaches. Outlier detection has been considered a foremost challenge in statistics for a long time and visual methods a possible solution. There are varying definitions of the term outlier in literature [Agg13, BL98, BG05]. They can be summarized by “an outlier is a data point which is significantly different from the remaining data” [Agg13]. A broad spectrum of methods for outlier detection in time series is available. Primarily, we refer to surveys, taxonomies, or other works which cover the breadth of the topic. Hodge and Austin [HA04] provide a good starting point for an overview on different types of methods, namely statistical models, neural networks, machine learning, and hybrid systems for outlier detection. The recent work by Aggarwal [Agg13], gives an in depth overview on outlier detection in general, with a particular part on outlier detection in time series. Ben-Gal [BG05] gives an overview and a taxonomy on statistical methods for outlier detection. The Mahalanobis distance is a distance-based outlier detection method in the class of parametric outlier detection methods [BG05]. It is an established method and commonly used in statistics to handle multivariate outliers [BG05, HA04, PnP01]. To avoid the influence of outliers on the estimation of the required variables, robust procedures have to be used to identify multivariate outliers [FGR05]. For analyzing and visualizing more than 3 dimensions with basic visualization methods, dimensionality reduction methods can be used, e.g., principal component analysis [Agg13] or multidimensional scaling [BBH11]. However, applying dimensionality reduction, the context of time, especially the periodicity, is lost, and the meaning of the principal components is difficult to interpret intuitively.

In summary, we could identify visualization approaches, which take into account the structure of time and different approaches for multivariate data. Moreover, we found several statistical approaches for multivariate outlier detection, specifically the well-established Mahalanobis distance. We could not find methods that can deal with the structure of time in relation within multivariate time series, neither visualize them in an intuitive and compact way, nor include both at the same time in a statistical oriented approach.

3. Background

Before we explain how to compute and construct the distance-based cycle plot, we briefly introduce the original cycle plot by Cleveland [Cle93] and define some variables.

3.1. Cycle Plot

The *cycle plot* is a representation described by Cleveland [Cle93] for time series that contains a reoccurring cycle, like a seasonal cycle, and a trend component, which often appear in time series. It

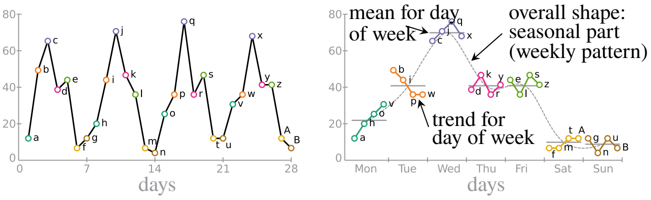

Figure 1: *Explanation of Cleveland's original cycle plot (adapted from Aigner et al. [AMST11]). Each individual day is labeled with the same letter in the conventional line plot (left) and the cycle plot (right). In the cycle plot the days within one group are connected to form a line. The average value for this day of week is indicated by a horizontal line.*

was presented as an alternative visualization for this type of data based on the seasonal subseries plot by Cleveland and Terpenning [CT82]. Cycle plots [AMST11, Cle93] are used to investigate the seasonal cycle and the trend along time *granularities*. The concept of granularities is explained in detail by Bettini et al. [BJW00]. Essentially, a granularity is a grouping of discrete points in time to larger units. For example, hours can be grouped into days. 'Day' is a granularity, while each specific day is called one *granule* of the granularity 'day'. The cycle plot inverts the order of grouping of two granularities: we illustrate this in Figure 1, using the granularities 'day' and 'week'. In the conventional line plot (see Figure 1, left) each granule of the granularity 'week' is used to create the tick marks on the horizontal axis. The data points are shown for each granule of each granularity, following the normal order of time. In the cycle plot (see Figure 1, right) the horizontal axis is grouped by day of week (Monday, Tuesday, etc.). Hence, the group 'Monday' contains all Mondays of these four weeks. All other days of the week are grouped accordingly. Aigner et al. state that the objective of the cycle plot is: "To make seasonal and trend components visually discernable", and the individual trends are shown "as line plots embedded within a plot that shows the seasonal pattern" [AMST11, p. 176]. In earlier work by Cleveland and Terpenning [CT82], the values of the subseries are plotted using vertical lines on the horizontal line representing the mean, in the later work by Cleveland [Cle93], a line is used for the subseries, like it is commonly known and used today. Yet Cleveland's original cycle plot represents univariate time series data only. In the following we use the term *original cycle plot*, whenever we want to explicitly refer to the original technique [Cle93] as described in this section.

3.2. Variable Specification

For the remaining part of the paper we specify variables and sets for the explanations. We will refer to p-dimensional time series data by $X = \{x_1, \ldots, x_n\}$ measured at time point $t_1, \ldots, t_n$. For simplification we use x_k to refer to the p-dimensional measurement at time point t_k, for $k = 1, \ldots, n$. Given a **seasonal length** (s), time points t_{i+j*s}, where $i = 1, \ldots, s$ and $j = 0, \ldots, \lfloor \frac{n-1}{s} \rfloor$ are in position i of the seasonal cycle and adding $j * s$ gives a time point in the same position, but j times further in the coarser granularity, such as the same month in different years. For instance, given monthly measurements x_k over 8 years, $s = 12$ represents the 12 months assembling a year. For $i = 1$ and $j = 0, \ldots, 7$, x_{i+j*s} would represent 8 January values: one measurement for each January of these 8 years. Likewise, the measures for month February, March, and April would be indexed by $i = 2, 3, 4$ respectively. In the following we refer to these bins as **groups** (X_i) within the seasonal cycle. By writing X_i we indicate the data points within one of the $i = 1, \ldots, s$ groups, where $X_i \subseteq X$, $\bigcap_{i=1}^{s} X_i = \emptyset$ and $\bigcup_{i=1}^{s} X_i = X$. In our example above, the groups represent the months of a year: $\{X_1 =$ January, $X_2 =$ February, $\ldots$, $X_{12} =$ December$\}$. For each of the groups we can define a group reference point $\mu_1, \ldots, \mu_s$, which can be the mean or median of the data points within the group and will be referred to as **group center** (μ_i). Moreover, we define a global reference point μ, named **global center** (μ) for mean or median of the whole dataset.

4. Task Abstraction and Requirements for Distance Measures

As a basis for discussing the design decisions and reasons for how to apply the additional distance-based abstraction, we first derive the tasks that are supported by the original cycle plot (Tasks T1-T5). Then we derive the tasks for outlier detection, going beyond the tasks supported by the original cycle plot (Tasks T6-T6). Theoretically our abstraction is independent from a specific distance measures, as long as it meets the specified requirements. For our prototypical implementation we apply the Mahalanobis distance, which is an example that meets these criteria.

4.1. Tasks

In the following we will use the terms *pattern* of the seasonal cycle and *behavior* within each group. By pattern of the seasonal cycle, we mean the shape formed by the group center perceived in the visual representation. For example in the cycle plot of Figure 1, the group of Mondays is generally a lower value followed by an steady increase until the peak on Wednesdays, saddling lower on Thursday and Friday with a drop to the lowest points on Saturday and Sunday. The behavior within a group means basically the pattern of the points or bars representing the data within the group.

The first set of tasks is derived from the tasks supported by the original cycle plot (Cleveland [Cle93], Cleveland and Terpenning [CT82]), as well as from our experience in applying the cycle plot [BFG*15].

T1: Identify the overall pattern of the seasonal cycle. Given a time series with a seasonal component, one wants to get an idea of the overall pattern of the seasonal cycle. This corresponds to an analysis of the finer granularity, e.g., patterns of the months' average over the year.

T2: Identify the behavior within each group. Beside the overall pattern of the seasonal cycle, one needs to assess the behavior of the subseries within each group. For univariate time series, this is often done to identify a larger trend corresponding to the coarser granularity, e.g., patterns within months over several years.

T3: Compare changes within each group to the seasonal cycle and across groups. Besides the individual behaviors of these aspects of the time series, one wants to know which of them drives the patterns of the whole time series and to which extent. It is also interesting how the behavior of one group compares to the behavior of another group.

T4: Detect extreme/outlying values within each group. The way the data is arranged in the cycle plot allows to identify extreme/outlying values with respect to data points within the same seasonal cycle. One wants to detect such extreme values and consider them as possible outliers.

T5: Identify whole groups that deviate from the seasonal cycle. When the overall pattern of the seasonal cycle (T1) is detected, one wants to identify groups within the seasonal cycle that deviate from this behavior.

Additionally, we specify tasks required for outlier detection in multivariate seasonal time series. These tasks are derived from domain knowledge about robust statistics and outlier detection both from literature and from the long-lasting experience of one of our co-authors [FGR05, FRGTA14], who is a statistician.

T6: Detect multivariate and univariate outliers based on the specified boundary. One wants to specify a tolerance boundary and easily detect data points outside this boundary. This needs to be possible for univariate and multivariate outliers.

T7: Detect outliers that are univariate as well as multivariate outliers. Extending task T6, one needs to detect data points that constitute outliers in both, multivariate and univariate context.

T8: Detect multivariate outliers and explore the respective data points in the univariate space. One needs to make selections of data points, in order to explore multivariate outliers and analyze the corresponding values in the univariate plots.

T9: Detect univariate outliers and explore the corresponding data points in the other variables as well as in multivariate space. This task is similar to T8, but one wants to start the exploration with selecting a data point in one (univariate) dimension and see its position in other dimensions as well as its representation in multidimensional space.

T10: Adjust outlier-boundaries and track the resulting outlyingness of data points. The boundaries specify what separates normal from outlying data points. One wants to adjust these boundaries in order to detect borderline outliers.

4.2. Requirements for a Distance Measure

To visualize multivariate time series in a cycle plot, we need an additional abstraction step. We decided to use a distance measure, because they are easy to compute, applicable for multivariate data, and a well-known concept. Distance measures, also allows us to retain a visual representation similar to the original cycle plot. Moreover, distance measures are commonly used in outlier detection. To support the tasks described above, the distance measurement for the data abstraction needs to meet the following requirements:

R1: Applicable for multivariate data.
R2: Robust against outliers.
R3: Specific cut-off value exists.
R4: Incorporates the correlation of the data.

4.3. Distance Measure

A distance measure quantifies the distance between two points in multivariate space. For our prototypical implementation we decided to use the Mahalanobis distance [Mah36], a generalized multivariate distance, which is an established method in statistics for multivariate outlier detection [BG05, HA04, PnP01] and meets our requirements on a distance measured described above.

In contrast to a basic distance measure, like the Euclidean distance, the Mahalanobis distance considers also the correlation of the data, which meets our requirement R4. A covariance matrix specifies the covariance structure of the data, which involves the correlation and the spread of the p dimensions. In 2-dimensions the spread can be illustrated with ellipses, see the data points and ellipses in Figure 3a. Given a p-dimensional dataset with n observations, $X = \{x_1, \ldots, x_n\}$, with the data center μ, and a covariance matrix Σ, the Mahalanobis distance between points x_k, for $k = 1, \ldots, n$, and the center μ is defined as

$$MD(x_k, \mu, \Sigma) = \sqrt{(x_k - \mu)^{\mathrm{T}} \Sigma^{-1} (x_k - \mu)}. \tag{1}$$

The center μ and covariance matrix Σ need to be estimated based on the dataset X. To estimate them there are different methods, ranging from classical to robust estimation methods [BG05, FRGTA14]. Even though, the specific method is not relevant for the construction of the distance-based cycle plot, but if used for outlier detection, robust methods are required.

Our main reason for using the Mahalanobis distance is that it is an established distance measure in statistics and used in multivariate outlier detection. By definition, see Equation (1), the Mahalanobis distance is applicable for multivariate data, fulfilling our requirement R1.

According to Filzmoser et al. [FGR05], if estimated with robust procedures, the Mahalanobis distance can be used to identify multivariate outliers, using quantiles of the chi-squared distribution. In more detail, in case of multivariate normal distribution, the squared Mahalanobis distance of the data points to the center, with respect to the covariance matrix of the data, are approximately chisquare-distributed with p degrees of freedom, χ^2_p. Thus, a potential multivariate outlier has a higher squared Mahalanobis distance than a certain quantile, e.g., the quantile 0.975, of the χ^2_p. We can use this quantile as a boundary for deciding whether a data point is an outlier or not, which meets the requirement R3.

The center μ as well as the covariance matrix Σ required to calculate the Mahalanobis distances need to be estimated based on the data, and when using classical methods for the estimation, these methods are influenced by outliers. Thus, to avoid the influence of outliers, we use robust methods for the estimation of μ and Σ, which meet requirement R2. In the statistics literature, compare [BG05, FGR05, FRGTA14], the most commonly used methods are the *median* as a robust estimator for the center of the data and the *minimum covariance determinant (MCD)* estimator [Rou85] for estimating the covariance matrix.

5. Features of the Interactive Exploration Environment

The main element in our interactive exploration environment is the distance-based cycle plot, which shows an abstraction of a multivariate time series using distances. This section is aimed to ease the understanding of the construction of the distance-based cycle plot by first explaining the transformation of an original cycle plot to a distance-based cycle plot. We then illustrate its construction in a

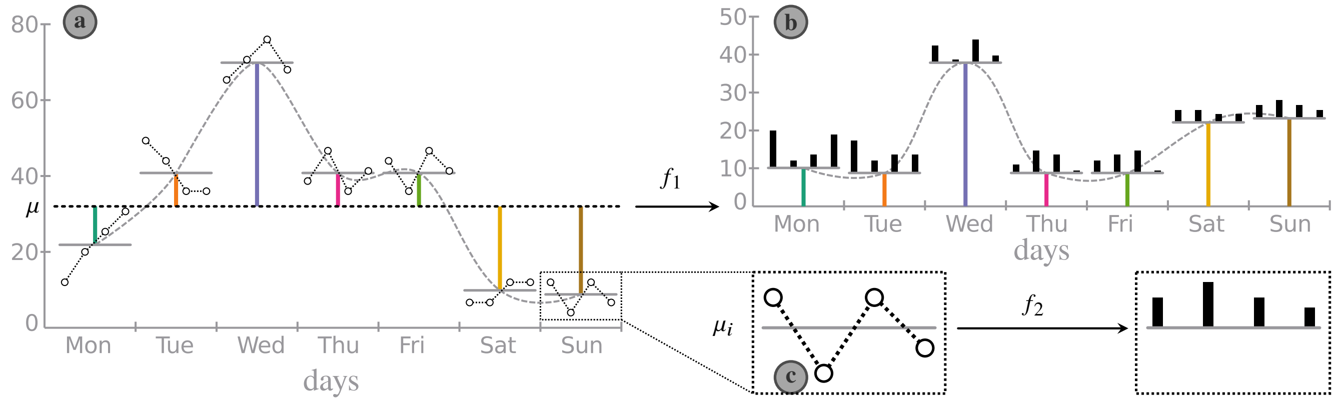

Figure 2: *Transformation of an original cycle plot to a distance-based cycle plot. Considering the group centers $\mu_1, \ldots, \mu_s$ as points forming a time series line plot and using the distance to the global center μ, we construct the base of the groups, transformation f_1, as described in Section 5.1. We do the same transformation f_2 for the pattern within each group. Both, the group centers and the points within each group, can form different patterns that are comparable to a seasonal pattern or trend in the original cycle plot.*

bivariate case. Finally it is generalized for the multivariate case and integrated into our interactive exploration environment.

5.1. Seasonal Cycle (Inter-Group Distance)

Our goal is to support the same tasks as the original cycle plot, but for multivariate seasonal time series, cf. Section 4.1. The original cycle plot shows the pattern of the reoccurring cycle, like the season over the year, which is required for tasks T1, T3, and T5 in the distance-based cycle plot. The identification of this overall pattern is supported by showing vertical lines that indicate the group centers $\mu_1, \ldots, \mu_s$, see Figure 1.

In the original cycle plot, each group center is a real number, where the absolute value can be considered as distance between this group center and the zero line. Instead of the zero line as central reference, we use by default the global center of all groups μ for constructing distance-based cycle plot. During the exploration process this global reference point can be changed interactively (see Section 6). In particular, we compute the Mahalanobis distance MD between each group center μ_i, $i = 1, \ldots, s$, and the global center μ. The result is one horizontal line for each group, serving as group base line. In Figure 2 we use the original cycle plot to illustrate how the distance-based cycle plot is constructed. In the original cycle plot (Figure 2a), the global center μ is indicated by a horizontal dashed line. We compute the distance of each group center $\mu_1, \ldots, \mu_s$ to this global center (indicated by the colored vertical lines). For the construction of this generalized distance based cycle plot, we apply these distances (group μ_i to global center μ) on the y-axis. Thus, the global center is represented by the x-axis itself (see transformation f_1 in Figure 2). This is due to the fact that these distance measures are always non-negative. The distinction between 'above' and 'below' the global center may be applicable in the univariate example given in Figure 2, but does not make sense in an actual multivariate scenario (as explained in Figure 3). We discuss this information loss due to the data abstraction Section 7 and describe how our interactive exploration environment allows to reduce this information loss.

In Figure 3 we show the construction in the bivariate case. We consider a small example of daily measurements of two variables: temperature and humidity. In this example the seasonal cycle $s = 2$ is grouping the data points into measurements during the *day* and measurements during the *night*. For each group – day and night – we calculate a bivariate average for temperature and humidity values combined, which is equivalent to the group centers μ_{day} and μ_{night}. In addition, we calculate the global center μ of the whole dataset, i.e., the bivariate (temperature and humidity) mean of all data points (8 days and 8 nights combined). In this bivariate example these two group centers and the global center are points in two-dimensional space (see Figure 3a). Given the group centers μ_{day} and μ_{night}, the global center μ, and the covariance matrix Σ (see Section 5.4), we calculate the Mahalanobis distance $MD(\mu_{\text{night}}, \mu, \Sigma)$ of the night-group center μ_{night} to the global center μ and $MD(\mu_{\text{day}} \mu, \Sigma)$ of the day-group center μ_{day} to the global center μ. This distance is used as the position of the group-base line on the y-axis shown by transformation g_2 and g_3 in Figure 3b.

In case of more than two dimensions, the group centers μ_i, the global center μ, and distances between them are calculated accordingly. By using these distance values, we are able to represent p-dimensional datasets in our distance-based cycle plot. For example, consider the monthly temperature (see Figure 4b'): there are low values in winter, high values in summer, and average values in spring and fall. This seasonal pattern is reflected in the position of the group center lines in a similar pattern like in the distance-based cycle plot (see Figure 4a). The winter months, like January as coldest month, and the summer months, like August as hottest month, have large distances to the global center and, therefore, appear as peaks in the distance-based cycle plot, whereas the average spring/fall months have small distances from the global center and therefore are closer to the x-axis.

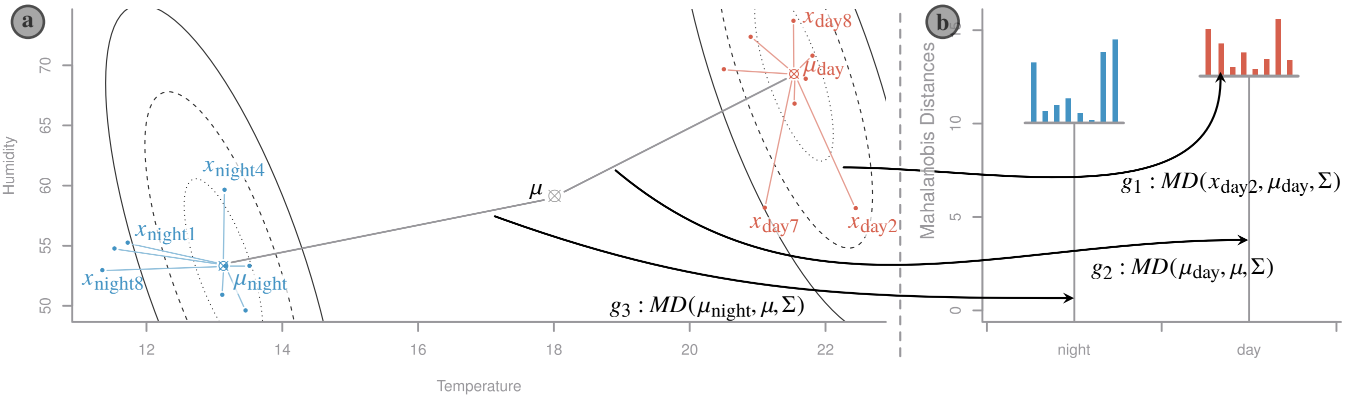

Figure 3: *Construction of the distance-based cycle plot with a bivariate example. The construction of the bivariate cycle plot on the right is based on Mahalanobis distances of the bivariate data points to the respective group center (e.g., transformation g_1) and on Mahalanobis distances of the group centers to the global center (transformation g_2 and g_3). The usage of distance measures that are applicable to p-dimensional space, is a key aspect applied in our data abstraction. The ellipses illustrate the spread of the data captured in the covariance matrix.*

5.2. Data Within Groups (Intra-Group Distance)

For representing the data points $x_1, \ldots, x_n$ within the groups $X_{i,\ldots,s}$, we apply a similar approach, like in the original cycle plot. In the original cycle plot, the data points are represented by points on a line following the order of the coarser granularity, and the position on the y-axis given by their values. One feature of this representation in the original cycle plot, is that it is possible to see the trend over the coarser granularity, for example over the years. For the distance-based cycle plot we compute the Mahalanobis distance of the data points in X_i to their group center μ_i. In contrast to the original cycle plot we represent distances instead of actual data points, and thus, we chose to use bars instead of connected points. This also picks up the original design by Cleveland and Terpenning [CT82], using vertical lines.

In Figure 1 we illustrate how the data points are binned and arranged in the original cycle plot. With univariate data the transformation to the distance-based cycle plot is similar to the construction of the horizontal lines for the group centers. Using the illustration from Figure 2, the data points in each group form a time series line plot with a horizontal line representing the group center. Computing the distance of each point to this group center allows to draw them as bar chart within each group, like we illustrate for one group in Figure 2c by transformation f_2.

In the bivariate example (Figure 3) we compute the Mahalanobis distance of each data point from the subset of measures during day $x_l \in X_{day}$ and night $x_m \in X_{night}$, where $l, m = 1, \ldots, 8$, to the respective group centers μ_{day}, μ_{night}, $MD(x_l, \mu_{day}, \Sigma)$ and $MD(x_m, \mu_{night}, \Sigma)$. These distances are ordered according to the coarser granularity, in our example calendar days, and represented as bars within their group, as shown by transformation g_1 in Figure 3b. As discussed before, the distance can be computed for any p-dimensional multivariate data set. However, in contrast to the original cycle plot, the seasonal pattern and trends need to be interpreted differently. In the following section we discuss the interpretation of these patterns in the distance-based cycle plot.

5.3. Design and Interactions

In Figure 4 we show the design of our interactive exploration environment. Our goal is to support users in exploring seasonal patterns and trends as well as detecting and exploring univariate and multivariate outliers (see tasks T1–T10 in Section 4.1). Using distances to construct the distance-based cycle plot (Figure 4a) allows for representing an arbitrary number of dimensions. To gain further insights into the multivariate dataset, we provide interactive exploration means that allow for switching back and forth between the distance-based visualization and the multiple underlying univariate representations. For visualizing the multiple variables of the multivariate time series, we provide the original cycle plot next to a time series line plot. This allows two perspectives on the same univariate dimension, like it is used by Bögl et al. [BFG*15] for time series containing missing values. In Figure 4, we show sparklines [Tuf06] for the univariate visualizations (b & c), to fit more variables on the screen. In the control panel (d), the user can switch between detailed view and sparkline view and adjust parameters. In the detailed view, the sparklines are replaced by more detailed line plots.

To explore patterns and outliers in the multivariate space and the underlying univariate dimensions, we provide multiple linked views with highlighting triggered by hovering and selection, including multiple selection. Highlighting and selection are supported in each of the visualizations. To allow the exploration of more dimensions, the univariate plots are scrollable.

We encode three types of outliers using color. We selected three distinguishable colors according to the L*a*b* color space and maximized the perceptual distance of the selected colors, compare [HSA*10]. The three types of outliers and respective colors are: (1) univariate outliers represented by cyan ■, (2) multivariate outliers represented by orange ■, and (3) outliers in univariate and multivariate represented by magenta ■.

Our interactive exploration environment is independent of a specific method for computing univariate or multivariate outliers. For doing so, a lot of methods exist in statistics literature, see

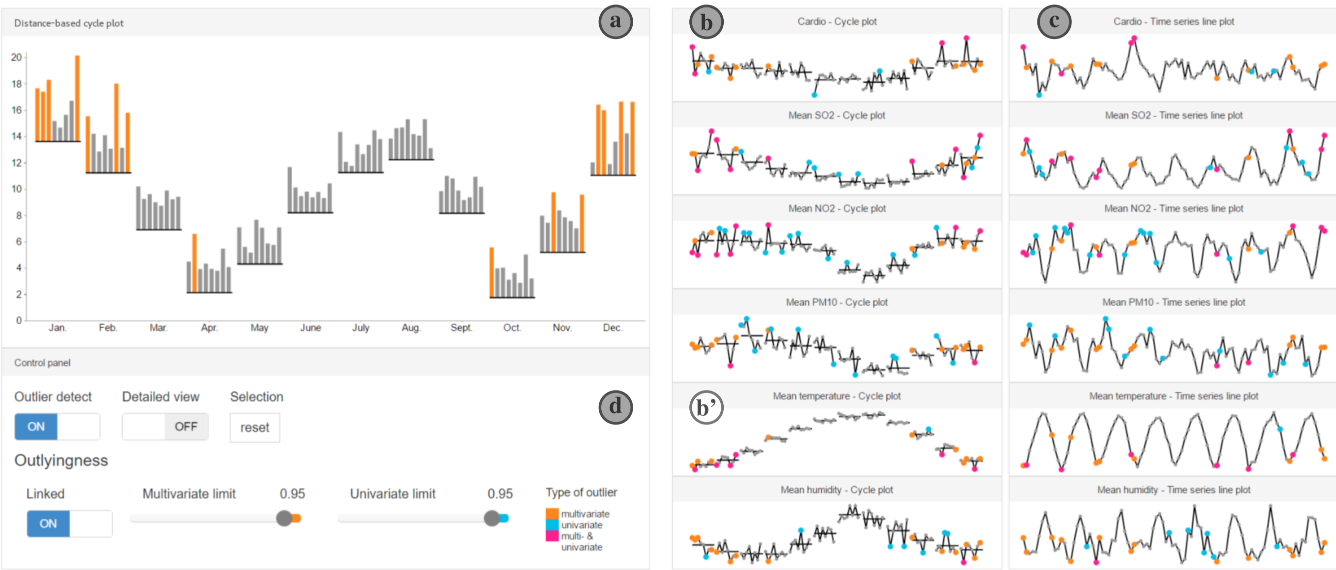

Figure 4: *Prototype implementation of our interactive exploration environment utilizing the Mahalanobis-distance-based cycle plot. The prototype employs coordinated multiple views with the distance-based cycle plot (a) next to the underlying univariate plots: an original cycle plot (b) followed by the univariate time series line plot (c). In this screenshot we use space-efficient sparkline representations, in order to provide a comprehensive overview. The small plots on the right side (b+c) can be changed to a more detailed view with a scroll bar for detailed exploration. The bottom left shows the control panel for interactive exploration (d). Color encodes the type of outlier and sliders are used to specify outlier-boundary values. (b') is the original cycle plot of variable temperature, used as example in Section 5.1.*

[BG05, HA04] for an overview. In our environment, we highlight the identified outliers in univariate as well as multivariate space and allow to adjust the parameters for outlier detection. For example, in Figure 4d the user can adjust the boundaries used for outlier detection. In the following section, we will give details on the calculations used for our prototype.

5.4. Robust Calculations

We use the median to compute the global center $\mu = \text{colMedian}(X)$, and the group centers $\mu_i = \text{colMedian}(X_i)$. For constructing the distance-based cycle plot, there are two possibilities to estimate the covariance matrix Σ. Either to estimate a separate covariance matrix for each of the groups Σ_i, or to center the data points on their group median and estimate a global covariance matrix Σ using the centered data points $\bar{X} = X_i - \mu_i$. Testing with some data sets showed an instability in the estimation of separate covariance matrices Σ_i. This is due to an often low number of data points in each group compared to the number of dimensions. Therefore, we apply the MCD method to compute the global covariance matrix $\Sigma = \text{covMcd}(\bar{X})$ with all centered data points $\bar{X}$. For a dataset with $\lfloor \frac{n-1}{s} \rfloor \gg p$ one can decide for any of these methods as needed.

For univariate outlier detection, we use a similar approach as described above. If the covariance matrix Σ is estimated robustly, the diagonal consists of robust estimates of the variance σ_r^2 for each variable $r = 1, \ldots, p$ in the p-dimensional dataset. Using the centered data points $\bar{X}$, the center of $\bar{X}$, $\mu_r = 0$, and the variance σ_r^2, we compute the outlier based on the selected quantiles of the underlying univariate distribution. In case the absolute value of a centered univariate data point is higher than a certain quantile, it is identified as univariate outlier. If the univariate data is normally distributed, we compute the quantile of $N(\mu, \sigma^2)$. Like for the multivariate boundary, we provide an interactive slider in our exploration environment for selecting the quantile, see Figure 4d.

Note that the assumed distributions (chi-square, normal) for the distances will most likely not be met because the observations are time-dependent, and thus not independent from each other. However, the quantiles of these distributions still serve as an indication of outlyingness of the data points. The goal of outlier detection is thus more in an exploratory context, namely to draw the attention of the user to these highlighted points.

6. Usage Scenario

We implemented the interactive exploration environment utilizing the distance-based cycle plot in a prototype and apply the prototype in a usage scenario. We use this usage scenario to illustrate how the distance-based cycle plot visualizes real data and how the interactive exploration environment advances the possibilities to explore patterns and outlying values in multivariate seasonal time series data. Throughout the usage scenario, we refer to the related tasks T1–T10 described in Section 4.1. We support the reader in following the usage scenario with additional figures provided in the supplementary material and refer to our prototype available online at `http://cycleplot.net`.

Mortality & Air Pollution Dataset. The dataset is about the mortality, air pollution, and meteorological data for major cities in South

Korea [LOK13]. It is available in the R project for statistical computing [R C16] as library named *HEAT* [LOK13]. The dataset contains several air quality indicators together with meteorological data as well as the number of deaths caused by cardiovascular diseases and respiratory diseases. The dataset consists of daily measurements for several years (2000-2007). For the illustration, we select a subset of 6 variables, cardio (deaths caused by cardiovascular diseases), SO_2 (Sulfur dioxide), NO_2 (Nitrogen dioxide), PM_{10} (particulate matter), temperature, and humidity from the city Seoul aggregated to monthly averages.

User. As a possible user is a public health official, who analyzes and explores the seasonal pattern, trends, and the outliers in the dataset described above.

Goal/Tasks. The overall goal is to get insights into the seasonal patterns, trends, extreme, and outlying values of the dataset. For details on the particular tasks to achieve this goal, we refer to the tasks T1–T10 described in Section 4.1.

As a proof of concept, we separated the preprocessing of the dataset and the prototype of the interactive exploration environment. The preprocessing of the dataset (HEAT library [LOK13]) was done in R [R C16]. The computation of global and group centers, Mahalanobis distances, and outlyingness values was done as described above (see Sections 4.3 and 5.4). The implementation of the interactive exploration environment was done as web application using *JavaScript*, where we imported the precomputed data file. For future work, one may combine the computations in R with the interactive visualization in a web application, by using appropriate libraries to connect them.

The user first wants to get an overview of the seasonal behavior of the time series in the multivariate space (compare Task T1). According to the group centers of the months, see Figure 4a, the user identifies that there are peaks in summer as well as in winter. This means that summer and winter months are on average more extreme than the global center, which basically represents an average month, e.g., spring (April) or autumn (October). Selecting one month center as the global reference point, e.g., January, shows that the other winter months are closer to January than the summer month (see the supplementary material for more details). Considering the transitions between high and low peaks of the season in the original cycle plot representation of each variable, the seasonal pattern of the Mahalanobis-distance-based abstraction follows a similar smooth behavior. The user then considers the behavior within the groups according to their position in the seasonal cycle (Tasks T2 & T3). He/She identifies a tendency in some of the peak months (Dec., Jan., & Feb.), that the data values within the group vary more than in others. Especially, when comparing to the other peak in summer, the user detects this additional variation with larger distances to the group center (Task T4). Next, the user compares the variations within the groups and across groups in more detail (Tasks T2 & T3). When looking at the months June and March, he/she spots distances with roughly the same length, except for the first year. These months seem to be quite stable months across all dimensions. Even without highlighting the user can easily identify extreme values by large bars, that may be possible outliers (Task T4). Amongst others, the user considers the last year in January, first in June, and several in December, as possible outliers. In our example, the user cannot find any group that deviates from the seasonal cycle (Task T5), but one can imagine one whole month, that stands out of the multivariate seasonal pattern.

The user activates the highlighting of outliers and selects a certain quantile for the univariate and multivariate boundaries to indicate outlyingness of the data points. The user selects the 0.95 quantile, shown in Figure 4, and gets an overview of patterns in the outlyingness that allows to detect multivariate as well as univariate outliers easily (T6). For example, the user considers interesting that there are multivariate outliers only in months Oct.–Apr., and an exceptionally large number in Dec.–Feb. Knowing that, the user detects the same pattern in the original cycle plots and recognizes that there are more data points in these winter months highlighted in magenta (T7), indicating outliers in both, uni- and multivariate space. The user immediately recognizes that the months Nov.–Feb. in the last year are all multivariate outliers. To further investigate the outlyingness in the univariate space, he/she selects the outliers (T8), which highlights the corresponding data points in the univariate plots. This exploration reveals that in some variables, e.g., cardio and PM_{10}, they are indicated as multivariate outliers only, yet in others, e.g., SO_2 and NO_2, they are highlighted as outliers in both, univariate and multivariate space. Looking at the original cycle plot for the variable cardio, the user detects two extreme data points in Nov. and Dec., highlighted in magenta. Selecting them shows that in the distance-based cycle plot, they can also be recognized as data points with large distance to the center (T9). The user also recognizes that besides being multivariate outliers, the variable cardio is also a univariate outlier in Nov. and Dec., but the variable temperature is a univariate outlier only in Nov. not in Dec. By changing the outlier boundary with the slider, the user can track the data points that are borderline and are indicated as outliers, when the boundary is decreased. For example, the first bar in month Mar. and Jun. in the distance-based cycle plot, see Figure 4a, are only highlighted as outliers, when changing the threshold from the 0.95 to the 0.9 quantile (T10). This allows to interactively get an impression about how extreme the outliers are.

In contrast to using only multiple original cycle plots, the user is able to explore the seasonal pattern and patterns within and across groups directly in the multivariate space. Obviously, it is required to also consult the underlying univariate visualizations, but combined in the interactive exploration environment, the distance-based cycle plot is vital for getting insight in the overall picture of the multivariate seasonal time series.

7. Discussion

So far, we introduced our interactive exploration environment for exploring patterns and outliers in multivariate seasonal time series and explained the construction of the utilized distance-based cycle plot. We abstracted the tasks relevant to do so together with the requirements for a distance measure in Section 4 and argued the construction of the visualization and the design of our environment, accordingly (Section 5). In these sections we briefly discussed the benefits and limitations of specific decisions for the construction and the design. In this section we will continue this discussion of benefits and limitations in more depth, cover the performance of our approach regarding the specified tasks, and give an outlook on future work.

7.1. Benefits and Limitations

One main benefit of the way we construct the distance-based cycle plot is the independence from the number of dimensions. This is achieved by constructing a distance based cycle plot using Mahalanobis distances (see Sections 4.3 and 5). This causes a different representation of patterns (i.e., the seasonal pattern and patterns within groups), and therefore, the distance-based cycle plot needs to be interpreted differently. A distance is a non-negative number by definition. As a consequence, all distances are represented above the group center lines (see Figure 2). We, thus, loose the information about the exact position of that data point, for the sake of being able to represent multiple dimensions. While the information about the actual position appears to be important for one-dimensional space (maybe even for two-dimensional, and three-dimensional space), it is very difficult to represent this in multivariate space. One commonly used method in this case is dimensionality reduction, like principal component analysis [Agg13] or multidimensional scaling [BBH11]. However, one limitation of this technique is, that it is difficult to interpret the meaning of the principal components, e.g., first and second for 2-dimensional visualizations. By applying dimensionality reduction, the context of time, especially the periodicity, is lost. Our approach, i.e. using the Mahalanobis distances, is also a type of abstraction from multivariate space to less dimensions. To reduce this information loss our interactive exploration environment, see Section 5.3, allows further investigations in both, the distance-based cycle plot and each of the single dimensions in multiple linked views. By taking the structure of time into account, and wisely selecting the granularity levels according to the seasonal time series, this abstraction still retains the temporal context.

While the idea of using intra-group vs. inter-group distances is a well-known strategy, we did not find this applied in cycle plots for multivariate time series (see Section 2). The full explorational power of our distance-based cycle plot for multivariate time series needs to be seen in context of the interactive exploration environment, where it is possible to connect the multivariate intra- and inter-group distances to the visualizations of the single dimensions in original cycle plots and line plots and thus, to investigate anomalies, like outliers.

Using distances, however, may lead to a setting where the global center has the same distance to multiple or even all of the group centers. This would result in a representation where all group base lines lie on the exact same position on the y-axis. In Figure 3 the global center μ is close to have the same Mahalanobis distance to both groups, and therefore, the base lines in Figure 3b are nearly on the same horizontal level. It is then difficult to judge if there is no seasonal pattern at all or if the group centers span a multi-dimensional sphere around the global center. The same drawback is true for original cycle plots. In case there is no seasonal pattern in univariate seasonal time series data, also the original cycle plot would show group centers aligned on a horizontal line. One way to tackle this problem is to relate the group centers to a different reference point. Instead of a mathematical global center we can define a global reference point. This reference point could be, for instance, the basis, the zero point, of the multivariate coordinate system. Another potential reference point would be choosing one representative group, i.e. a reference month whose group center line would then lie on the x-axis, and relate the other groups to this reference group. The same can be applied for group centers, by defining a group reference point instead.

For both issues mentioned above, we introduced the possibility to select any group center (μ_i) as global reference point and therefore investigate the relation of all other groups to the selected group in our interactive exploration environment. We use the multiple linked views to enable the selection of single or multiple points either in the distance-based cycle plot or in one of the univariate representations. Brushing and linking allows to further explore the relations of a point in the multidimensional space and the connection to the single dimensions. This feature helps to reduce the information loss introduced by the distance based abstraction. The flexibility of interactively adjustable global as well as group reference points, allows a further investigation of relations between data points within and even across groups. The exploration of different aspects of locality (in linear and periodic time, as well as across groups) is possible because of using Mahalanobis distances. There is relatively few work done considering this local aspects in outlier detection, see [FRGTA14].

Another possible limitation is demonstrated by a case in which there are very large distances between the global center and the group centers and only small variation of the data points within the groups. A common scale (y-axis) would thus lead to a distance-based cycle plot that shows mainly long vertical lines representing the distance of group centers to the global center. In consequence the small variations within the groups would not be visible. However, this can also happen in the original cycle plot. To tackle this problem, we propose using data transformations, such as a log scale. Another solution would be to use separate scales. Using one scale for the distance of group centers to the global center, and another scale for the distances within groups would allow for exploring the smaller variations within groups, while preserving the overall picture and comparison between groups.

A problem related to the previous one, would be a very strong trend within the data (e.g., monthly data over several years). A steep trend over the years would distort the patterns within groups of months. This, again, is a problem that affects the original cycle plot as well. In time series decomposition the time series is split up into trend-, seasonal-, and the error- or irregular-component [BD10]. These can then be analyzed separately with appropriate visualizations, e.g., using our distance-based cycle plot for the seasonal-component. Seasonal adjustment is usually applied to remove the seasonal component in order to analyze the other components. There are also recent approaches to seasonal adjustment for multivariate time series (see [GM13]), that can also be used to separate the seasonal component for a more detailed analysis, which again could be supported by using our distance-based cycle plot.

7.2. Task Performance

A formal cost-benefit analysis of our approach on the basis of Chen and Golan [CG16] is beyond the scope of this paper, but the following discussion should better explain the benefit of our approach when performing the task outlined in Section 4.

Using only an original cycle plot for each variable (only the plots in column (b) of Figure 4), one can easily identify overall patterns of the seasonal cycle (Task T1), identify the behavior within each group (Task T2), and compare changes within each group to the seasonal cycle and across groups (Task T3), but only for each of the variables separately. For picturing these patterns and behaviors within the multivariate space, one can do this to a certain extend by mental aggregation. If there are very similar and smooth transitions and patterns in each of the variable, one can imagine or derive mentally a similar pattern in multivariate space. On the other hand, when using only the distance-based cycle plot (Figure 4a) for these tasks, it is possible to identify the patterns and behavior of the abstraction only. One can identify peaks and transitions as well as intra- and inter-group similarities in the multivariate space, but it is not possible to break down any information for individual variables. While the tasks of detecting extreme/outlying values within each group (Task T4) and the identification of whole groups that deviate from the seasonal cycle (Task T5) can be done for each variable separately, multivariate outliers are possibly not outliers in any of the variables or outliers just in single variables. Also by summarizing mentally the individual variables, data points that are outliers in single or many individual variables, may or may not be multivariate outliers (see Section 6). For these tasks (Tasks T1-T5) a combination the distance-based cycle plot (Figure 4a) and the several original cycle plots (Figure 4b) is necessary. The support of interaction is also beneficial to keep track of single points, when switching the focus between the single variables and the distance-based cycle plot.

If we consider the Tasks T6-T10 for multivariate and univariate outlier detection that are going beyond the detection of extreme/outlying values only within groups (Task T4) and identification of whole groups that deviate from the seasonal cycle (Task T5), one also requires the classic line plot representation of each variable (Figure 4c) to also identify peaks and anomalies, like possible outliers, and how they relate to each other in linear time, e.g. the larger values within the last year of variable NO_2 and SO_2 in Figure 4b and c. This side-by-side presentation of the same data in different representations, cycle plot and traditional line chart, adds an additional perspective that allows to investigate the connections of outliers in each of the variables. Highlighting univariate outliers in original cycle plots and line plots only, does not allow for investigating data points that are no outliers in any of the single variables but are multivariate outliers. The distance-based cycle plot allows to investigate the data points relations with different aspects of locality, e.g. to data points in the same position of the periodicity, but also the distances compared to other groups as well as interactively changing global and local (group) reference points. Only the possibility to select and highlight the data points in these coordinated multiple views of different perspectives, allows to easily switch from the abstracted multivariate data to the single dimensions, which eases the external memorization of the analysis. This affects all the tasks for all combinations of only multivariate or only univariate outliers in one or more dimensions, as well as outliers that are both, multivariate and univariate outliers in single or multiple dimensions.

The main advantage of the distance-based cycle plot is the aggregated overview of all dimensions combined, to show directly the patterns and anomalies like outliers in a condensed view. The full exploration power is only achieved by the interactions, highlighting, and coordinated multiple views in combination with the twofold visualization of each variable in a classical cycle plot and a line plot. Another benefit of this combination of different representations and the additional abstraction of the distance-based cycle plot is, that it enables the investigation of each data point. First, locally according to the normal linear time scale for each variable, second, locally according to the periodicity in the several original cycle plots for each variable, and third, locally in context of the distance-based abstraction from the multivariate data, again according to the periodicity in the distance-based cycle plot.

7.3. Future Work

There are several questions with regard to the usability of our approach that remain open. First, the distance-based cycle plot, although encoding an abstraction using a multivariate distance, uses a very similar design to the original cycle plot. This may cause confusion and needs to be learned by the user. Furthermore, we do not know about the performance regarding the abstracted tasks and whether the combination of different cycle plots discussed above helps users to identify multivariate outliers. For multivariate time series involving periodicity, this remains an open question. Only future evaluation with real users can answer these questions. We plan to tackle this issue first by providing additional use case examples, and by formal user studies focusing on the correct interpretation of the plots and task performance. Due to the number of tasks and their complexity, this will require more than one study. For the time being, we demonstrate the applicability of our approach by the walk-through in the usage scenario (Section 6), the comprehensive discussion above, the provided supplementary material, and the possibility to test the prototype in an online demo.

8. Conclusion

Our interactive exploration environment utilizes a distance-based cycle plot for identifying seasonal patterns and outliers in multivariate seasonal time series. It revisits and retains a visual representation similar to the original cycle plot by Cleveland [Cle94]. The construction of the distance-based cycle plot includes an additional abstraction step using the Mahalanobis distances, which enables the generalization to an arbitrary number of dimensions. With our interactive exploration environment we combine statistics and visualization techniques and balance their benefits and limitations for visually analyzing patterns and outliers in multivariate seasonal time series, with respect to the structure of time and the relations among multiple dimensions.

Acknowledgements The authors like to thank R. Maciejewski, J. Bernard, C. Turkay, and the anonymous reviewers of InfoVis 2016 & EuroVis 2017 for the discussions, critical questions, feedback, and suggestions. This work was supported by: Austrian Federal Ministry of Science, Research, and Economy via CVAST (#822746), a Laura Bassi Centre of Excellence; Austrian Science Fund (FWF) via VISSECT (#I2850-N31) and KAVA-Time (#P25489-N23).

References

[Agg13] Aggarwal C. C.: *Outlier Analysis*. Springer New York, 2013. 2, 9

[AMST11] Aigner W., Miksch S., Schumann H., Tominski C.: *Visualization of Time-Oriented Data*. Springer, London, UK, 2011. 2, 3

[BAF*15] Bögl M., Aigner W., Filzmoser P., Gschwandtner T., Lammarsch T., Miksch S., Rind A.: Integrating predictions in time series model selection. In *Proceedings of the EuroVis Workshop on Visual Analytic, EuroVA* (2015), Bertini E., Roberts J. C., (Eds.), pp. 73–77. 2

[BBH11] Bunte K., Biehl M., Hammer B.: A general framework for dimensionality-reducing data visualization mapping. *Neural Computation 24*, 3 (2011), 771–804. 2, 9

[BD10] Barnett A. G., Dobson A. J.: *Analysing Seasonal Health Data*. Statistics for Biology and Health. Springer Berlin Heidelberg, 2010. 1, 9

[BFG*15] Bögl M., Filzmoser P., Gschwandtner T., Miksch S., Aigner W., Rind A., Lammarsch T.: Visually and statistically guided imputation of missing values in univariate seasonal time series. In *Proc. IEEE Conf. Visual Analytics Science and Technology, VAST* (Oct 2015), pp. 189–190. 3, 6

[BG05] Ben-Gal I.: Outlier detection. In *Data Mining and Knowledge Discovery Handbook*, Maimon O., Rokach L., (Eds.). Springer US, 2005, pp. 131–146. 2, 4, 7

[BJW00] Bettini C., Jajodia S., Wang S.: *Time Granularities in Databases, Data Mining and Temporal Reasoning*. Springer, 2000. 3

[BL98] Barnett V., Lewis T.: *Outliers in statistical data*, 3rd ed. Wiley, 1998. 2

[BPC*10] Borgo R., Proctor K., Chen M., Janicke H., Murray T., Thornton I.: Evaluating the impact of task demands and block resolution on the effectiveness of pixel-based visualization. *IEEE Trans. Visualization and Computer Graphics 16*, 6 (2010), 963–972. 2

[BSH*16] Bach B., Shi C., Heulot N., Madhyastha T., Grabowski T., Dragicevic P.: Time Curves: Folding time to visualize patterns of temporal evolution in data. *IEEE Trans. Visualization and Computer Graphics 22*, 1 (Jan. 2016), 559–568. 2

[BW08] Byron L., Wattenberg M.: Stacked graphs – geometry & aesthetics. *IEEE Trans. Visualization and Computer Graphics 14*, 6 (Nov. 2008), 1245–1252. 2

[CG16] Chen M., Golan A.: What may visualization processes optimize? *IEEE Trans. Visualization and Computer Graphics 22*, 12 (2016), 2619–2632. 9

[Cle93] Cleveland W.: *Visualizing data*. Hobart Press, Summit, USA, 1993. 1, 2, 3

[Cle94] Cleveland W. S.: *The elements of graphing data*, rev. ed. Hobart Press, Summit, NJ, USA, 1994. 2, 10

[CT82] Cleveland W. S., Terpenning I. J.: Graphical methods for seasonal adjustment. *Journal of the American Statistical Association 77*, 377 (1982), 52–62. 3, 6

[CTB*12] Chae J., Thom D., Bosch H., Jang Y., Maciejewski R., Ebert D. S., Ertl T.: Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In *Proc. IEEE Conf. Visual Analytics Science and Technology, VAST* (2012), pp. 143–152. 2

[Few08] Few S.: Time on the horizon. Visual Business Intelligence Newsletter, `http://www.perceptualedge.com/articles/visual_business_intelligence/time_on_the_horizon.pdf`, cited Mar 20, 2016, June 2008. 2

[FGR05] Filzmoser P., Garrett R. G., Reimann C.: Multivariate outlier detection in exploration geochemistry. *Computers & Geosciences 31*, 5 (2005), 579–587. 2, 4

[FHR*14] Federico P., Hoffmann S., Rind A., Aigner W., Miksch S.: Qualizon graphs: Space-efficient time-series visualization with qualitative abstractions. In *Proc. of the 12th International Working Conference on Advanced Visual Interfaces, AVI* (2014), ACM, pp. 273–280. 2

[FRGTA14] Filzmoser P., Ruiz-Gazen A., Thomas-Agnan C.: Identification of local multivariate outliers. *Statistical Papers 55*, 1 (2014), 29–47. 4, 9

[GM13] Greenaway-McGrevy R.: *A Multivariate Approach to Seasonal Adjustment*. Tech. rep., Bureau of Economic Analysis, 2013. 9

[HA04] Hodge V. J., Austin J.: A survey of outlier detection methodologies. *Artificial Intelligence Review 22*, 2 (2004), 85–126. 2, 4, 7

[HKF16] Haroz S., Kosara R., Franconeri S.: The connected scatterplot for presenting paired time series. *IEEE Trans. Visualization and Computer Graphics 22*, 9 (2016), 2174–2186. 2

[HSA*10] Hurter C., Serrurier M., Alonso R., Tabart G., Vinot J.-L.: An automatic generation of schematic maps to display flight routes for air traffic controllers: Structure and color optimization. In *Proc. of AVI'10* (2010), pp. 233–240. 6

[JME10] Javed W., McDonnel B., Elmqvist N.: Graphical perception of multiple time series. *IEEE Trans. Visualization and Computer Graphics 16*, 6 (Dec. 2010), 927–934. 2

[KBK11] Krstajić M., Bertini E., Keim D. A.: CloudLines: Compact display of event episodes in multiple time-series. *IEEE Trans. Visualization and Computer Graphics 17*, 12 (2011), 2432–2439. 2

[KJL14] Kerren A., Jusufi I., Liu J.: Multi-scale trend visualization of long-term temperature data sets. In *Proc. SIGRAD* (2014). 2

[KKA95] Keim D., Kriegel H.-P., Ankerst M.: Recursive pattern: A technique for visualizing very large amounts of data. In *Proc. IEEE Visualization, Vis95* (1995), pp. 279–286. 2

[LAB*09] Lammarsch T., Aigner W., Bertone A., Gärtner J., Mayr E., Miksch S., Smuc M.: Hierarchical temporal patterns and interactive aggregated views for pixel-based visualizations. In *Proc. 13th Int. Conf. Information Visualisation, IV 2009* (2009), IEEE, pp. 44–50. 2

[LOK13] Lim Y.-H., Ohn I.-S., Kim H.: *HEAT: Health Effects of Air Pollution and Temperature (HEAT)*, 2013. R package version 1.2. 8

[LYK*12] Luo D., Yang J., Krstajic M., Ribarsky W., Keim D.: EvenRriver: Visually exploring text collections with temporal references. *IEEE Trans. Visualization and Computer Graphics 18*, 1 (2012), 93–105. 2

[Mah36] Mahalanobis P. C.: On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (Calcutta) 2* (1936), 49–55. 4

[MHR*11] Maciejewski R., Hafen R., Rudolph S., Larew S. G., Mitchell M. A., Cleveland W. S., Ebert D. S.: Forecasting hotspots—a predictive analytics approach. *IEEE Trans. Visualization and Computer Graphics 17*, 4 (2011), 440–453. 2

[Pla86] Playfair W.: *The Commercial and Political Atlas and Statistical Breviary*, 2005 ed. edited and introduced by Wainer H. and Spence I. Cambridge University Press, 1786. 2

[PnP01] Peña D., Prieto F. J.: Multivariate outlier detection and robust covariance matrix estimation. *Technometrics 43*, 3 (2001), 286–310. 2, 4

[R C16] R Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. 8

[Rei08] Reijner H.: The development of the horizon graph. In *Proc. Vis08 Workshop From Theory to Practice: Design, Vision and Visualization* (2008), Bartram L., Stone M., Gromala D., (Eds.). 2

[Rou85] Rousseeuw P. J.: Multivariate estimation with high breakdown point. *Mathematical Statistics and Applications 8* (1985), 283–297. 4

[SBVLK09] Schreck T., Bernard J., Von Landesberger T., Kohlhammer J.: Visual cluster analysis of trajectory data with interactive Kohonen maps. *SAGE InfoVis 8*, 1 (2009), 14–29. 2

[SFdOL04] Shimabukuro M. H., Flores E. F., de Oliveira M. C. F., Levkowitz H.: Coordinated Views to Assist Exploration of Spatio-Temporal Data: A Case Study. In *Proc. 2nd Int. Conf. Coordinated and Multiple Views in Exploratory Visualization, CMV04* (2004), IEEE, pp. 107–117. 2

[TAS04] Tominski C., Abello J., Schumann H.: Axes-based visualizations with radial layouts. In *Proc. ACM Symp. Applied computing, SAC'04* (2004), ACM, pp. 1242–1247. 2

[Tuf83] Tufte E. R.: *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT, 1983. 2

[Tuf06] Tufte E. R.: *Beautiful Evidence*. Graphics Press, Cheshire, CT, 2006. 6

[vWvS99] van Wijk J. J., van Selow E. R.: Cluster and calendar based visualization of time series data. In *Proc. IEEE Symp. Information Visualization, InfoVis* (1999), pp. 4–9, 140. 2

[WWS*16] Wu T., Wu Y., Shi C., Qu H., Cui W.: PieceStack: Toward better understanding of stacked graphs. *IEEE Trans. Visualization and Computer Graphics 22*, 6 (2016), 1640–1651. 2