Anomaly detection for visual analytics of power consumption data

Halldór Janetzko *, Florian Stoffel, Sebastian Mittelstädt, Daniel A. Keim

University of Konstanz, Germany

A B S T R A C T

Commercial buildings are significant consumers of electrical power. Also, energy expenses are an increasing cost factor. Many companies therefore want to save money and reduce their power usage. Building administrators have to first understand the power consumption behavior, before they can devise strategies to save energy. Second, sudden unexpected changes in power consumption may hint at device failures of critical technical infrastructure. The goal of our research is to enable the analyst to understand the power consumption behavior and to be aware of unexpected power consumption values. In this paper, we introduce a novel unsupervised anomaly detection algorithm and visualize the resulting anomaly scores to guide the analyst to important time points. Different possibilities for visualizing the power usage time series are presented, combined with a discussion of the design choices to encode the anomaly values. Our methods are applied to real world time series of power consumption, logged in a hierarchical sensor network.

1. Introduction

Commercial buildings consume a significant amount of electricity. According to the Energy Information Administration's 2010 statistics [1], the United States alone consumed an estimated 1.3 trillion kW. It is about 37% of the total electricity generated. How power is used in a commercial building has a large effect on energy efficiency strategies. The most important energy usage is lighting. Then heating and cooling are next in importance [2]. Current approaches for reducing the power consumption for example integrate motion detection sensors for each lamp switching them on and off.

There is a growing interest in understanding how energy is spent in the commercial buildings. Furthermore, building administrators want to know how to reduce the failure rate and detect anomalies. In addition, they want to know how to visualize large volumes of energy consumption data collected by power meters (sensors) in a building to find patterns, trends, and anomalies. In the end, our goal is to find how to automatically discover the anomaly, like unusual power consumption measurements highly differing from old observed patterns, and to reduce the energy cost of a building. For this task, anomalies are of special interest, because they can be caused either by faulty equipment or potentially misconfigured devices consuming significantly more or less energy than required for proper operation.

In this paper, we present an analytical and visual approach to support the building administrators in detecting anomalies and examining energy consumption data as shown in Fig. 1. Our input data consists of a tree of time series reflecting the hierarchical nature of the power meters, e.g., 1 m for the whole building and one for each power outlet. The analytical part is the automatic anomaly detection and is based on a time dependent energy consumption model. We have explored two different anomaly discovery methods. In the beginning, we estimate the error rate using prediction. Then, we use clustering based anomaly detection. Both methods have benefits and drawbacks and are complementing each other.

The last step in our pipeline is the visualization being capable of effectively displaying large amounts of data and, at the same time, allowing quick recognition of anomalous regions in the data. We integrated the three most common time series visualization techniques (line charts, spiral visualizations, and Recursive Patterns) presented in Aigner et al.'s book about time series [3]. Besides giving an appropriate overview of the data, the visualization is also able to support the administrator in a more detailed examination of the data, for example areas with unusual power consumptions by interaction facilities. In addition, the visualization is capable of showing the hierarchical nature of the data set. This is necessary, because commonly the energy consumption of different floors or buildings is independently monitored resulting in an inherent hierarchy in the recorded data.

Our methods rely purely on the recorded power consumption data, which we were not cleaning in any way as the data was in very good shape. There are many external influences to the power consumption, like the environmental conditions or the number of

* Corresponding author. Tel.: +49 7531884793.
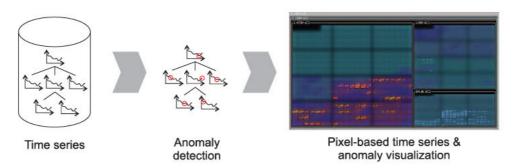 *E-mail address:* janetzko@dbvis.inf.uni-konstanz.de (H. Janetzko).

Fig. 1. The input set of hierarchical time series is processed by anomaly detection methods. The resulting anomaly values are visualized together with the time series values by pixel-based techniques. The visualization combines the raw time series with boosting techniques like highlighting and blurring for the anomaly scores.

people working in an office building. The large number and high complexity of external factors prohibit the fully automatic diag nosis of anomalies. Hence, a human subject matter expert is needed to validate found anomalies and investigate the interesting ones. Even though it is possible to think of extensions for an automatic analysis of anomalies like incorporating external factors as weather data and holidays.

It is important to note that our methods are applicable not only to power consumption time series data sets, although they have been developed with a particular application in mind. This is caused by the general nature of time series data and the generality of both, the analytical and the visual methods presented in this paper. The most application dependent part of this work is the anomaly detection being designed for daily patterns.

2. Related work

Reading energy consumption statistics shows that commercial buildings have a high energy usage, which motivates many research projects developed to improve power efficiency. Within the context of our work two main categories can be distinguished: analysis of power consumption data (detecting whether the energy consumption performs normally or abnormally over dif ferent locations and time) and visual analysis (visualizing simila rities and anomalies with appropriate interaction techniques).

2.1. Analysis of power consumption data

Applying data mining techniques for power consumption data is a known approach for identifying abnormal usage behavior. Agarwal et al. [4] examined 6 months of data from the UCSD campus, including aggregate power consumption of four buildings. Agarwal et al. focus more on the setup of power meters and provide only simple visualization methods like line charts. Catter son et al. [5] used an approach to monitor old power transformers. Their goal is to proactively search for abnormal behavior that may indicate the transformer is about to fail. Similarly, McArthur et al. [6] searched for anomalies to detect problems with power gen eration equipment. Jakkula and Cook [7] compared several outlier detection methods to find which is better at identifying abnormal power consumption. Seem [8] used outlier detection to determine if the energy consumption for a day is significantly different from previous days' energy consumption. This is a known approach for identifying abnormal system behavior.

The work conducted at Lawrence Berkeley National Laboratory [9] focuses on demand response. Mathieu et al. used a time of week and piecewise linear modeling approach to analyze com mercial and industrial electric load data. To our knowledge, the unsupervised anomaly detection algorithms from prediction and clustering described in this paper differ from the Mathieu et al.

method in two aspects: finer granularity and weighted by time distance (recent data weights more than old data).

The review of several prediction methods for power data performed by Zhao et al. in [10] investigates the effectivity and efficiency. Neural networks and Support Vector Machines were performing better than statistical approaches. We though decided to use the prediction technique developed in [11] as peak preservation is one of the main strengths of this technique.

2.2. Visual analysis

Visualization of building energy consumption has not yet been a major focus of research thus far. Most of the energy consumption visualizations have been time series line charts, scatter plots, and maps [12 15]. Recently, Many Eyes [16] allows analysts to choose a visualization type for analyzing public building electricity con sumption. The Google PowerMeter [17] recently provides a free energy monitoring tool for people to view home energy usage.

In addition to these existing tools, improving visualization techniques for time series data is ongoing research work. In SAVE [18], Shi et al. presented a sensor anomaly visualization engine that guides the user to diagnose sensor network failures and faults using multiple coordinated views. In this paper, we map multiple sensors' calendar time series in a single view to enable users to visually analyze energy usage and identify anomalies. In SAGA Dashboard [19], Buevich et al. provided a visual interface for interaction with the sensor network. They require the user to use a device that tracks and visualizes home energy usages. We extend the home energy consumption visual analysis to large commercial buildings with dozens of sensors. We therefore restricted ourself to space efficient visualizations like pixel based Recursive Patterns. Furthermore, no pre defined devices and sensor types in our methods are required. Another related work being capable of visualizing hierarchical time series data is the TimeEdgeTrees introduced by Burch and Weiskopf in [20]. The technique shows the time series as one dimensional, color coded timelines instead of drawing the graph edges. The hierarchy is preserved better by this approach while the space efficiency is worse compared to the pixel based approaches we use. We chose the pixel based techniques as periodic patterns are easier percei vable. Additional discussions on related work concerning anoma lies detection and boosting methods can be found in Sections 3 and 4.

2.3. Our contribution

To leverage the prior work and to support analysts in under standing power consumption data, we combine automated anom aly detection algorithms with interactive time series visualizations. The resulting anomaly score is used to highlight unusual power

usages in the time series visualizations. Our contributions in the visual analysis process of power consumption data are as follows:

1. In the anomaly detection process, we
   - detect power consumption anomalies based on either a clustering based approach or a time weighted prediction.
   - compare the prediction based method with a similarity based anomaly computation.
2. In the time series and anomaly visualization process, we
   - map the hierarchical time series onto a Treemap.
   - embed in each Treemap cell the corresponding meter's time series visualization.
   - provide different time series visualization techniques dependent on the analysis purpose.
   - visualize the anomaly score by visual boosting of the raw time series representation.

Furthermore, we have provided an advanced visual interface enabling the user to visually analyze the power usage. Histograms for viewing the frequency and power usages of important meters; visual queries for analyzing correlation and similarity; and various options on visualization types, Treemap layout, colormappings, and anomaly score computations enable the analysts to tailor the visualization to their needs.

3. Anomalies detection

Detecting and exploring of anomalies in time series is a very important aspect, especially when dealing with power consumption data of physical infrastructure. Saving cost and energy as well are one of the main motivations for observing and analyzing consumption data. But when dealing with infrastructure that may be even system critical the number of failures must be reduced to an absolute minimum. Early signs of failure should be visible in abnormal power usage patterns. In our main usage scenario abnormal behavior is defined as a difference from the expected daily pattern. Both methods described below assume a daily power usage pattern which, of course, can be different for each day of the week. Both techniques are not limited to daily patterns, but can be easily adapted to the periodicity of the underlying data set. The first described method is based on a weighted prediction, where recent measurements have a higher impact than older measurements. The latter approach is trans forming the observed daily pattern in the frequency domain and looking for dissimilarity in a transformed space.

3.1. Prediction based anomaly detection

The basis for prediction is an observed pattern and the assumption that it is reoccurring (with slight modifications) in the future. If this assumption does not hold true, the predicted values may be far off the measured values. Considering this fact the other way round, observed values far distant from the expected ones tell us that the model used does not explain the observed values. There might be two reasons, the first one is that the model quality is not good enough and the second one is that the values are really differing from the expected and explainable behavior. We assume that our data follows a regular underlying pattern and therefore also assume that the model describes the usual behavior well. Detecting anomalies using prediction follows this idea and is related to the statistical measure of residuals.

The prediction method used is crucial for the reliability and expressiveness of the computed anomaly scores. As already stated above we assumed daily patterns and include developments over time into the prediction process. We decided to use a prediction method developed and introduced in our earlier works [11]. Basically, this method predicts a value for each minute of the day by taking all previous measurement at the same time of the day. As an example, assume that we predict the value for a Tuesday at 11:05 am. We would now average all previous observed values of a Tuesday at 11:05 am. Taking just an average would have the disadvantage of neglecting recent developments in the time series. We therefore used a weighted averaging scheme with higher factors for recent values and linearly decreasing influence weights for older values. Further detailed explanations can be found in [11]. This prediction method works very well for weekly patterns and will neglect holidays or other external events. The prediction model will adjust to seasonal changes, but alternating behaviors cannot be modeled by this approach. Furthermore, power usage patterns randomly distributed over a day will negatively influence the prediction quality.

After predicting for each point in a time series the expected values based on all values occurring before this point in the time series, we can compute the difference between predicted and observed values. The difference is an indicator for the abnormality of the point in a time series but needs for higher expressiveness of some kind of normalization. From the choice and the design of the prediction method we are assuming a model which may not being applicable to all observed time series. We counterbalance for this fact by calculating the average fitting of our model. More in detail, we compute the average deviation from the predicted values for the whole time series. If a whole time series is highly unpredict able, the differences between predicted and actual values are less meaningful compared to a case when a time series follows perfect daily patterns with small deviations. Computation of the anomaly score is summarized by the following equation:

$$anomaly[time] = \frac{|predVal[time] \quad obsVal[time]|}{avg_{t \in Time}(|predVal[t] \quad obsVal[t]|)}$$

The variable time is the point in a time series for which the anomaly score is calculated. At this position the difference between the predicted and observed value is computed and afterwards normalized by the average deviation from the model.

3.2. Clustering based anomaly detection

The second approach for detecting anomalies in time series data is similarity based. We assume often observed patterns to be the usual behavior and rarely occurring patterns to be abnormal. Following this idea, we first have to define and compute the similarity of patterns in order to detect whether a pattern occurs more than once. The approach described in this section is proposed and presented by Bellala et al. in [21,22]. The time series is first partitioned into days and afterwards transformed by a Fourier transformation into the frequency domain. Each day of the time series is resulting in a k dimensional vector in the frequency domain with k being a parameter of the transformation process. The next step described by Bellala et al. is a dimension reduction by multi dimensional scaling into a two dimensional space. The density distribution in the reduced MDS space is now interpreted as an anomaly score. Points (time series of a single day) being in a high density area with many (similar) neighbors are assumed to reflect the usual behavior. Outliers in the 2D space can be seen as days with unusual values and are assigned a high anomaly score. This technique only takes the frequency domain into account and does not integrate external effects like weather data or week of the day.

3.3. Comparison of anomaly detection methods

We previously described two methods for computing and detecting anomalies and both come with their advantages and

drawbacks. Comparing both methods the most obvious difference is the resolution of the anomaly score. The prediction based method computes for each point in a time series one anomaly score, whereas the clustering based method returns only one anomaly value per day. It is of course possible to extend Bellala's technique to cope with hours or even minutes of a day, but noise might influence the clustering approach. This behavior is inherited from the computation of the anomaly scores. The clustering based technique uses daily time series and uses them as one data item in the clustering process. An anomaly value is assigned to each data item based on the density distribution. Therefore, there is no possibility to assign different anomaly scores to temporal sub units of a day (i.e., minutes, hours).

The second essential difference can be seen in the complexity of the methods. Transforming the data set in the frequency domain and applying MDS results in a data space with axes that are hard to interpret. But the frequency domain is typically less prone to noise and induces some robustness to the observed time series. Though the transformed data space is complex, there exists the possibility to extract models of typical behavior by computing cluster representatives. Furthermore, the clustering approach allows supporting several typical 'behaviors' of a time series. Just assume a time series alternating between a day work and a night work pattern. The clustering approach will assign both low anomaly scores as both patterns are observed often, whereas the *prediction based* method will assign each day a very high anomaly score as averages are computed.

In Fig. 2, we present a visual comparison between both anomaly computation methods. The first column shows a visualization of the observed values for four consecutive Mondays. The exceptional behavior of the third Monday is obvious. This anomaly is reflected in all computed anomaly scores, while the higher temporal resolution for the prediction based method is visible. Altogether, the clustering based approach is good for cases when time series switches between different typical behaviors and the prediction based approach is good for cases when the behavior slightly changes over time following a (seasonal) trend.

4. Anomalies visualization

The anomaly scores computed in the previous section are used to highlight important time intervals of the input time series. The visualization for the time series is influencing the design possibilities depending on the visual variable encoding the numerical values. We implemented for comparison three well known, state of the art methods to visualize time series data: Recursive Patterns [23,24], spirals [25], and the traditional line chart. These techniques are configured to visualize only one time series.

We will discuss the different design alternatives and motivate our design decisions in the following sections. We focus hereby on the possibilities to encode the time series and the anomaly values simultaneously. We describe all state of the art techniques visualizing time series, namely Recursive Patterns, spiral visualizations, and line charts.

4.1. Recursive Pattern

The Recursive Pattern is a pixel technique using coloring to visually encode numerical values using a parameterized space filling layout. They are capable of displaying large amounts of time series data in a space efficient way. By setting proper parameters the resulting display can be calendar like pixel visualizations. The layout for 1 day is shown in Fig. 2 and the values of the series are clearly recognizable and furthermore patterns or exceptions are easy to identify.

In order to incorporate the anomaly score in the Recursive Patterns, we present a color boosting technique that highlights data points by manipulating the intensity of color values according to the anomaly score. The boosting techniques applied in this paper are a subset of the techniques presented by Oelke et al. in [26]. Out of the presented techniques, we used only color high lighting as the visualization will not be overcrowded with visual cues. Color boosting techniques bias the visual impression, which might lead to misinterpretations of the visualization. To keep these artifacts at a minimum, we created a perceptual uniform colormap
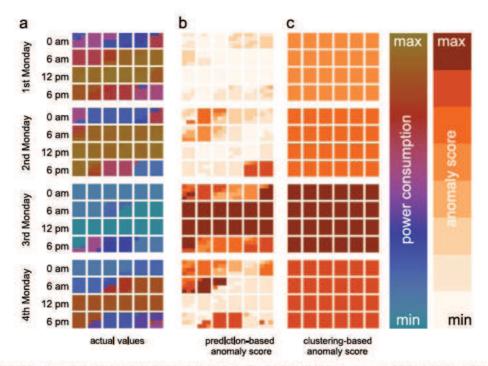

Fig. 2. Comparison of the resulting anomaly scores based on the proposed methods. The third Monday shows an unusual behavior being reflected in the anomaly scores.

that only varies over hues without variation in intensity. Since the change in intensity does only minimally shift the hue, the original color tone can be reconstructed mentally.

It is known that RGB and HSV are not perceptually uniform and that linear interpolations within these models do not produce color scales with equal or monotonically changing lightness [27]. CIE LUV and CIE LAB have already been proven useful in former visual analytics research [28,29]. By varying over the color opponents (a and b) but maintaining the same lightness value L, a perceptually uniform colormap can be created in the CIE LAB color space. However, interpolations in CIE LAB can lead to undefined RGB signals and thus, this color space cannot be used in the final application. Therefore, we use the HSI color space [30] for intensity manipulation.

This color space is an extension to the HSV color space that allows monotonical changes in lightness.

The two proposed color encodings for the anomaly values can be seen in Fig. 3. The first row depicts the original time series without any anomaly scores. We use different intensity levels to encode the anomaly scores and highlight important areas. The effect of the intensity boosting can be seen in the second row of Fig. 3. For further visual boosting we combined blurring and intensity highlighting shown in the last row of Fig. 3.

We added another highlighting technique, in order to direct the analyst to the anomalous regions of the time series. This high lighting imitates the human perception regarding a focus and the context area, where usually the focus area is sharp and the context area is blurry. We used a similar approach to Kosara et al. and Giusti et al. in [31,32]. Since we have the anomaly score for every element of the visualization, we are capable to determine the important areas of the time series analytically corresponding to the focus area of the analyst. The implementation adapts locally the blurring according to the anomaly value of each element in the Recursive Pattern. Low anomaly values are more blurry than areas with a high anomaly score. This adaptive blurring technique utilizes the human depth intuition guiding the analyst to the interesting areas first in a pre attentive way, depicted in the bottom row of Fig. 3. The blurring will affect the visibility of pixel borders, and it influences the comparability between highlighted and non highlighted areas. We though believe that the preattentive focusing on anomalies helps the analyst in assessing interesting points in time at a glance.

4.2. Spiral visualization

The spiral visualization is a technique to display recurring time series data with a fixed periodicity. Our implementation is based on an Archimedean spiral, where the radius grows proportionally to the spiral angle, which leads to a uniform expansion of the spiral over time. In our implementation, each round of the spiral is used to display 1 day of data. The proportional growth of radius and spiral angle, combined with the absence of any border between each circle makes it possible to build a space efficient visualization. Comparing the value of the same time span on different days is possible, because these values are on a straight line going from the center of the spiral to the outermost part of the spiral. Each polygon along this line displays the same time span of different days.

To show the anomaly score of each of the displayed time spans, we apply the same color manipulations as described for the Recursive Pattern above. The right spiral in Fig. 4 shows the described color saturation and brightness adjustment to highlight the anomalous values of the time series. By comparing the left with the right spiral the highlight of the outer ring of the right spiral is clearly visible. There is a time range with unusual numerical values beginning after one fourth of the day and lasting for one quarter of a day. Besides that, some little colorful spots are visible in the right visualization, which were not that visible when applying only the brightness or saturation modification technique.

4.3. Line chart

The most common visualization of time series data is undoubtedly the line chart. The main difference to the Recursive Pattern or spiral based visualization can be found in the encoding of the actual time series value. In the latter two, the series value is shown by colored polygons, which have a spatial extent. In contrast, encoding the value in a line chart is done by the position on the y axis. The brightness and saturation based techniques to add the anomaly value into the visualization make no sense in such a positional encoding, having only a very small area available for the coloring. Coloring segments of the line and applying the same techniques to enrich the line with anomaly score information as before is not helpful as line segments are very hard to see. To use coloring a larger line stroke would be necessary, which would introduce high amount of over plotting and visual clutter. It may be fine for line chart displayed on a large screen, but as soon as several line charts are displayed the technique does not work anymore.

To show the anomaly value simultaneously with the time series values, we used the empty space in the background of the line chart as shown in Fig. 5. For each data point, we plot a red stripe in the background. The anomaly value is mapped to the opacity of the stripe in a way that for the lowest anomaly value it is completely transparent and therefore not visible. In contrast, the
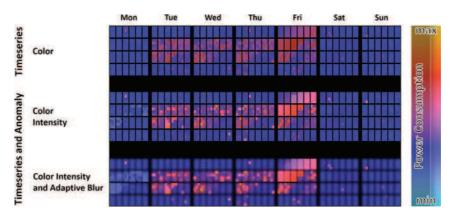

Fig. 3. Different methods to display the anomaly value. Top row: the time series values without anomaly values. Second row: the intensity of the color is adapted to the anomaly value. Third row: color intensity representing the anomaly score combined with adaptive Gaussian blurring. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)
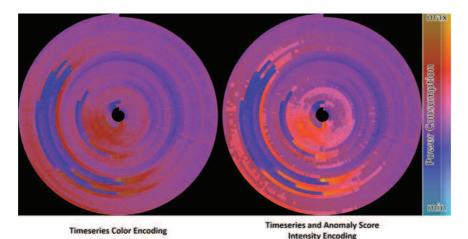
Fig. 4. Spiral visualization of time series. The left spiral shows the actual time series data, the right spiral shows the time series data with brightness and saturation value adapted to the anomaly score of the corresponding polygon.
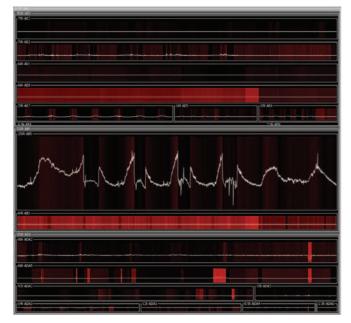

Fig. 5. The line chart visualization in a Tree Map with a horizontal strip layout using the value weighted anomaly to determine the cell size. The effect of the layout score is clearly visible enlarging the time series not with the highest anomaly score, but with the highest anomaly level regarding the time series values. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

highest anomaly score causes the stripe to have the highest opacity resulting in a clearly visible, red stripe.

To reduce the visual clutter introduced by coloring the background, we also support a minimized view. In this view, the anomaly stripes are only plotted above and below the line chart, which keeps the visualization distraction free, but still shows the anomaly values. A comparison of both anomaly visualization techniques for line charts can be seen in Fig. 6.

4.4. Treemap Integration

We integrated all visualizations in a Treemap display [33 36] (see Figs. 5 and 7). In that way, the hierarchical nature of our time series data set is reflected in the visualization. Treemaps are showing the leaves of each selected branch and the nesting depth by borders. The selection of visualized nodes can be achieved twofold, either by interactive roll up or drill down operations in the Treemap visualization or by an additional vertical tree repre sentation. Our design choice using Treemaps, though they visua lize only the leaves of each branch, was implied by the application needs. The analysts are mainly interested in finding the root causes of anomalies first and later on in analyzing the impacts by traversing the hierarchy to the root node. Further details concern ing the used time series can be seen in the application (Section 5).

Each cell of the Treemap contains the visualizations of the time series building one branch of the hierarchy. The border of each of the cells is furthermore drawn in white to allow a clear distinction in terms of the hierarchy. The caption of each Treemap cell is used to display the numerical value used for layout and the cell label.

The numerical value is used by the layout manager to compute the final Treemap layout and directly influences the size of a single Treemap cell. The computation of the numerical values is critical for the expressiveness of the visualization since the size of a cell has a large influence on the perception. The size of a Treemap cell can be computed by different measures. Given the interest of an analyst to quickly recognize unusual or highly anomalous time series, the Treemap layout can be adjusted to support these tasks by computing the layout score in different ways. For example, the analyst can choose between the statistical variance, sum, or the arithmetic mean of the anomaly score. To incorporate the level of the anomaly, there is also the possibility to compute the layout based on the product of the anomaly score and the time series value. In addition to anomaly score based layouts, the sum and the statistical variance of the time series values can be used to compute the layout. Having those choices, the visualization can be adapted to the priorities of the analyst independent of the visualization technique. We also added the possibility to assign the same importance value to each node resulting in a regular layout enabling easy comparisons. Besides the general layout the actual width and height (the aspect ratio) of a single cell is an important factor when using different time series visualization techniques. For that reason, we implemented different layout algorithms for the previously described visualization methods.

A Recursive Pattern has a rectangular shape and, therefore, a squarified layout [37] is applied to the Treemap. This layout algorithm results in a square like cell, which obviously leads to an efficient space usage of the overall display. In addition, we framed the Treemap cells to improve the overall structure percep tion of the Treemap and the hierarchical representation.

The circular shape of the spiral graphs combined with the squarified Treemap layout leads to the best readability and space efficiency. We hereby maximize the size of visualization and at the same time use as much space of the Treemap cell as possible.
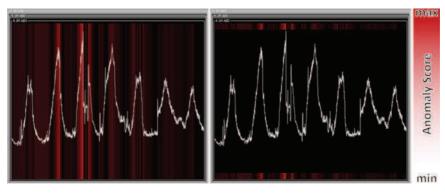
Fig. 6. Comparison of the anomaly visualization technique for line charts. On the left, the whole background is used to show the anomaly scores, whereas on the right, only a small stripe on the top and bottom of the chart background is used to display the anomaly score, which reduces the clutter from the background coloring. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)
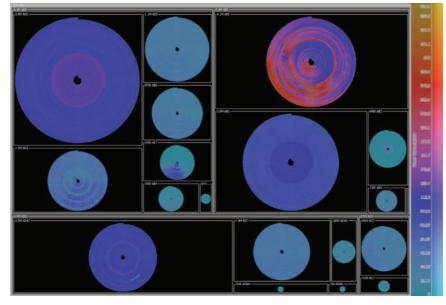

Fig. 7. Treemap visualization of 19 time series, each time series has 4 weeks of data. Interesting spots or patterns in the data are highlighted and can be therefore easily detected.

Creating the layout for Treemaps containing line charts comes with a fundamental difference to the Recursive Pattern and the spirals: the width of a line chart is much larger than its height as our observed time span is quite long. Consequently, this leads to the conclusion that a squarified layout is not the best choice. Instead, we implemented a so called strip layout [38], which makes sure that the line charts are getting more space on the horizontal axis than on the vertical (see Fig. 5). Otherwise, the line charts would be very hard to interpret and this would be an unfair comparison to the pixel based techniques. Note that the size of each Treemap cell still reflects the numerical value used for layout.

4.5. Comparison of anomaly visualizations

We have presented three different state of the art visualization approaches for time series and visual extensions to show time series and anomaly score simultaneously. All techniques have their own advantages and disadvantages. The Recursive Patterns presented first have the ability to visualize large amounts of data in a very compact and space efficient way. Regardless of the shown time range, lasting from weeks and months to years, the Recursive Patterns are always capable of showing the data in a readable fashion revealing patterns. The visualization is designed in such a way that the value representation by color enables the analyst to

easily spot interesting areas or regular patterns, nearly indepen dent of the actual size of the visualization. In Fig. 7, patterns and outstanding time spans are visible, even in the compact Treemap representation of 19 different time series. Having spotted regular patterns Recursive Patterns enable also the cross comparison in different time series, since the relative position of one point in time is well aligned. Using Recursive Pattern in Treemap is more difficult to compare the same hour of a day, for example, as the position of the same hour varies through the visualization.

Comparing the same hour is an advantage of the spiral visualization as the periodicity was set to daily patterns. The angular encoding of the time of a day enables these comparisons as a straight line from the spiral center to the outer spiral connects these data values. With such visualizations, it is easy to explore the value of the time series over time. In addition, comparing time ranges and/or spot longer lasting trends is a simple task, since the analyst has to only follow the continuous spiral over time. This is an advantage compared to the non continuous time display of the Recursive Patterns, where layout breaks are needed, as with any space filling curve. Line charts are great for detailed visual explorations of continuous data for single time series. For the usage scenario of anomaly visualization, there exist only a small number of application possibilities, since condensed visualizations are needed as limited screen space is an issue. The low space

efficiency of line charts leads to our proposed solution to re use the empty space in the background to visually encode the anomaly value. We avoid the arising visual clutter by applying the stripe based anomaly visualization, which keeps the anomaly informa tion but reduces the colored area distracting the analyst.

5. Applications

The prototype integrating all the presented analytic and visual techniques focuses especially on the detection of anomalies and their temporal occurrence. With this task in mind, two general use cases can be identified. First, general browsing and exploration of the data is important to get an overall impression of the power usage. All different visualization techniques presented above can be applied to gain from their individual strengths. The second task is the examination of a specific issue, like unusually high or low power consumption. Our system can provide the analytical and visual insights necessary to find the source of the unusual energy consumption. All visualizations are integrated in the same analy tical framework, but use different methods of displaying the power consumption and the anomaly values.

5.1. Analytical framework

Our prototype consists basically out of three parts reflecting the different dimensions in the data set and can be seen in Fig. 8. The left panel allows the navigation through the hierarchy of the sensor graph by selecting the nodes being visualized. The visua lization panel in the center consists of the Treemap visualization together with a colormap legend. The panel at the bottom of the window allows to navigate in time and select the time range that should be visualized. This timeline visualization shows the total amount of power usage over time in order to give the analysts additional hints.

We implemented besides animation also interaction techni ques like dragging the selected time range (blue rectangle in the timeline visualization) left and right causing immediate updates to

the visualization. The visualization allows basically three interac tion possibilities. The first is a tooltip allowing to inspect the underlying data values invoked by mouse hovering. We further more directly support drill down and roll up operations in the Treemap visualization, allowing the analyst to keep his focus on the visualization during traversing the sensor graph. Finally, the analyst is able to select a region in the visualization and query the system for similar time series sharing the selected behavior by means of distance or correlation calculations. Switching the visualization technique, colormap, value normalization, anomaly calculation, or the weights for the Treemap layout is possible by selecting the respective option.

5.2. Visual inspection of anomalies

In this use case, the building administrator gets the informa tion, that in February 2012 the overall power consumption and energy costs of a building were higher than expected. The investigation starts by getting an overview and some contextual information about the general energy consumption of the build ing. Undoubtedly, the most suitable visualization for this task is the Recursive Pattern visualization, which can be seen in Fig. 9. The blurring approach at the right side highlights the anomalies further compared to the left figure, where we visualized anomalies only by color intensity. The resulting visualization points directly to one time series, which can be seen in Fig. 10 on the right. Both, the left and the right visualizations show the power consumption data beginning on 6 February 2012. Each of the bigger rectangles contains the data from 1 day, starting with Monday on the left. In total, there are 4 weeks of data visible, starting on 6th February and ending on 4th March.

In the visualization, there are some single, outstanding spots. Those look relatively random and last only 1 pixel, which stands for a time span of 5 min. Although the color is quite intense and reddish, they are far too few and do not last long enough to have a large influence on the power consumption. Besides these spots, an area in the fifth column of the third row stands out. The intensity seems to increase from pixel to pixel over a long time. Having in
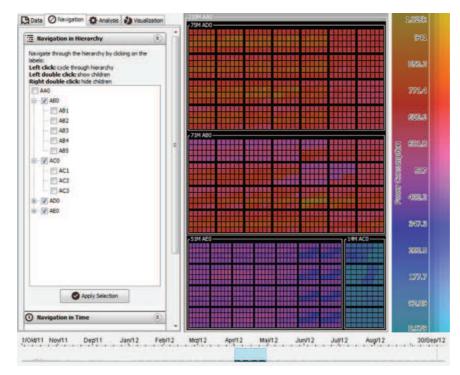

Fig. 8. Screenshot of our prototype showing the hierarchical and temporal selection possibilities together with the visualization panel.
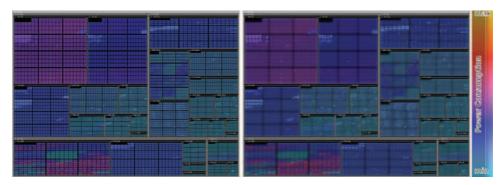
Fig. 9. Overview of the power consumption data from 28 sensors during 48 weeks. Despite the huge amount of data, patterns are still clearly visible. On the right, the same visualization with adaptive blurring highlighting unusual power consumptions can be seen.
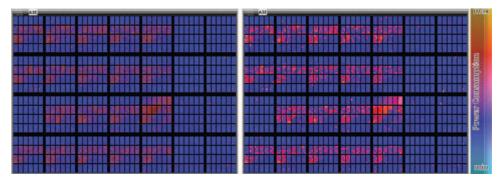

Fig. 10. Sensor readings of sensor AE3 from 6th February to 4th March 2012. On the left, the power consumption is visualized. On the right, the intensity of the colors reflects the anomaly score. Due to the high intensity, an area in the fifth column of the third row stands out. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

mind that one small black framed rectangle of the Recursive Pattern stands for 1 h the anomaly score seems to increase over 10 h, until suddenly the anomaly score drops again. Due to the long duration of the anomaly and the intense red color, the actual energy consumption in this time frame is very high. This makes this anomaly a candidate for the cause of the higher energy costs in February.

The building administrator found an anomaly in the given time frame with the Recursive Pattern visualization. To identify potentially correlated time series, our prototype implements a top n time series similarity search. The query can be created by clicking on a part of the visualization and selecting the query area with the mouse. After wards, the desired similarity measure can be selected. The system supports the standard Euclidean distance and positive, negative, and unsigned Pearson Correlation for different analysis tasks. In this case, selecting the positive Pearson Correlation and the Euclidean Distance is appropriate. The result of the query can be seen in Fig. 11.

The query results show three very similar series: AE4, AE5, and AE6. All three sensors are part of the same subtree of the sensor hierarchy. This means they are located in the same building as sensor AE3, which logged the time series identified as anomalous by the Recursive Pattern visualization before. With this additional knowledge, the building administrator can conclude that the anomaly affected not only one, but at least four parts of the building, where the sensors have been installed.

The quality of the conclusions drawn from the visualizations and analytical methods depends heavily on the sensor deploy ment. If each of the sensors monitors a single machine or office, the building administrator has a concrete subject of further examination. When they are deployed in a more general way, for example per building floor or even per building, the shown analysis allows narrowing down the investigation of power con sumption to the affected units.

6. Evaluation

We showed the applicability of our proposed technique in the previous application section, but it is very important that real expert users rate our approach to be effective and helpful. We therefore presented our approach to the target user group in a big company. We had contact to two analysts and interviewed them first about their state of the art technology. The company develops sensor networks measuring the power consumption for large buildings and is experienced with power management. The current state of the art technology they are using is a line chart based visualization. They are able to select arbitrary time frames and inspect the temporal power distribution. Further analysis steps are yet impossible to perform. In later meetings we explained our approach to the experts and afterwards let them interact with our system and investigate the time series data. We asked them to describe their typical way of analyzing data and furthermore to comment on our proposed technique by thinking aloud using our prototype. We got very valuable and interesting feedback from the experts regarding the benefits and room for improvement.

First of all, they validated the temporal patterns shown in the pixel oriented visualization techniques with their knowledge of typical power consumption patterns. Their proof of concept was that the daily periodic patterns were visible at a glance, at the same time reflecting their expectations for the time series. After they found the patterns like low power consumption at nighttime and weekends they started to look for anomalies using our visual boosting techniques. At first, obvious patterns like holidays or the Christmas vacation have been found. Afterwards, less obvious patterns have been investigated. During their analysis, we asked the experts to comment on our techniques and give feedback related to visualization and analysis methods.
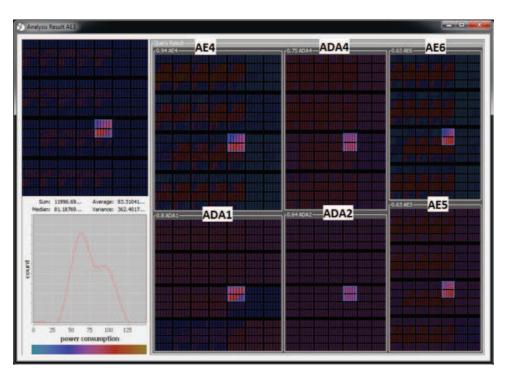
Fig. 11. The time series query result window. On the top left, the query time series is displayed, on the right the top-n query results are shown. The query range is highlighted.

The first point they commented on was the helpfulness of the overview visualization in the form of the Recursive Patterns. Compared to line chart based visualizations they are very familiar with, the calendar like representation of the power consumption was highly appreciated. Furthermore, the possibility to interactively change the visualization type helped them a lot to get familiar with the pixel oriented techniques. The coloring of pixels was intuitive to them and they could interpret the visualization easily.

From an analysis point of view, a very interesting point was their comment on our prediction based anomaly computation. They agreed with our definition of anomaly: "The anomalous day is likely to deviate from the daily pattern in some way." As shown above, our anomaly method is very fine grained, but to the experts a single time spot with a high anomaly score is not important. They were more interested in longer periods of unusual behavior, starting at approximately 1 h duration. On the other hand the related anomaly computation method based on days was too coarse grained for them for this kind of analyses. An aggregation of the anomaly values might help to let the analyst focus on the severe anomalies. The visualization of the anomaly scores together with the time series was mentioned very positive, especially with respect to the Recursive Pattern. The overview calendar like visualization with intensity highlighting and adaptive blurring let them focus on the interesting spots. They had the impression that their attention was guided to the anomalies, while the unimportant, common daily patterns were pushed in the back ground. As soon as they found some unexpected anomalies they applied further analysis techniques.

The experts very much appreciated the possibility to select a region in the time series and query for other similar time series. When they selected a leaf in the hierarchy of time series they would look for the impacts of the anomaly on the parent nodes. The other way around, querying for anomalies on higher levels would show the root causes for the unusual power consumption.

A possibility for improvement mentioned by them is the integration of external events into the application. Sometimes managers know in advance of extraordinary events that will cause unusual power consumptions. It should be possible to include this information whenever available and to reflect the additional events in the visualization. Overall they found the integration of different time series visualization techniques combined with an anomaly representation very helpful and wanted to integrate our techniques in their management tools.

7. Conclusion

Analyzing and interpreting unusual patterns in time series data is a very important task. In this paper, we applied novel analysis and proven visualization techniques to a system, which supports analysts finding those patterns in a visual way. We supported the analysis process by computing anomaly scores of the given time series data with an anomaly detection algorithm which produces very fine grained results. This also allows the creation of detailed visualizations resulting in a fine grained pixel based date representation. Furthermore, the algorithm is very efficient in terms of required computing power, because neither does it require expensive transformations nor does it rely on elaborate analysis of the time series data.

Having the anomaly scores, different visualizations can be used to get deep insight into the time series and the anomaly scores, depending on the task to fulfill. Recursive Patterns generate overviews of large time spans and large amounts of data. Spiral views provide the possibility to quickly detect and analyze periodic patterns. If the actual data values are of interest, the classical line charts are also available for further investigations of the data set.

The double encoding of time series values and anomaly scores is solved in different ways. The novel adaptive blurring, which generates a focus and a context area by blurring the visualization according to the anomaly scores, guides the analyst directly to interesting spots of the visualization. This makes the technique a particular advantage in overview of visualizations, where

irrelevant areas of the time series are losing their level of detail by a strong blur, whereas interesting, high anomalous areas are clearly visible and attract the focus of the human eye. To support the display of multiple visualizations, the well known Treemap approach is extended by layouts based on space efficiency and specific visual properties of the visualization. Since the anomaly scores determining the layout can be selected depending on the analysis task, the resulting Treemap layout also supports further analyses.

The use case of power consumption data shows the applicability of the methods shown in this paper. The general nature of the analysis and visualization methods makes it possible to apply these techniques to time series not only from the application domain of power consumption data. In the future, we want to integrate external knowledge like known events influencing the time series like weather information. It would be also interesting to automatically determine the visualization method, colormap, and possible enhancements like the adaptive blurring based on the displayed data.

Acknowledgments

References

[1] United states energy information administration. Annual energy review, 2010.
[2] US department of energy. Energy efficiency trends in residential and commercial buildings; 2008. ⟨http://apps1.eere.energy.gov/buildings/publications/pdfs/corporate/bt_stateindustry.pdf⟩.
[3] Aigner W, Miksch S, Schumann H, Tominski C. Time & time-oriented data. Human–computer interaction series. London: Springer; 2011 ISBN 978-0-85729-078-6.
[4] Agarwal Y, Weng T, Gupta RK. The energy dashboard: improving the visibility of energy consumption at a campus-wide scale. In: Proceedings of the first ACM workshop on embedded sensing systems for energy-efficiency in buildings. ACM; 2009. p. 55–60.
[5] Catterson VM, McArthur SD, Moss G. Online conditional anomaly detection in multivariate data for transformer monitoring. IEEE Trans Power Deliv 2010;25(4):2556–64.
[6] McArthur SD, Booth CD, McDonald J, McFadyen IT. An agent-based anomaly detection architecture for condition monitoring. IEEE Trans Power Syst 2005;20(4):1675–82.
[7] Jakkula V, Cook D. Outlier detection in smart environment structured power datasets. In: 6th International conference on intelligent environments (IE), 2010. IEEE; 2010. p. 29–33.
[8] Seem JE. Using intelligent data analysis to detect abnormal energy consumption in buildings. Energy Build 2007;39(1):52–8.
[9] Mathieu JL, Price PN, Kiliccote S, Piette MA. Quantifying changes in building electricity use, with application to demand response. IEEE Trans Smart Grid 2011;2(3):507–18.
[10] Zhao Hx, Magoulès F. A review on the prediction of building energy consumption. Renew Sustain Energy Rev 2012;16(6):3586–92.
[11] Hao MC, Janetzko H, Mittelstädt S, Hill W, Dayal U, Keim DA, et al. A visual analytics approach for peak-preserving prediction of large seasonal time series. Comput Graph Forum 2011;30(3):691–700.
[12] ⟨http://en.wikipedia.org/wiki/Energy_in_the_United_States⟩ [viewed 17.06.2013].
[13] IBM TRIRIGA Energy Optimization: integrated software solution for improving buildings management and facilities operation. ⟨http://pic.dhe.ibm.com/infocenter/tivihelp/v57r1/topic/com.ibm.iteo.doc/infocenter.pdf⟩. [viewed 17.06.2013].
[14] UCEI (University of California Energy Institute), Berkeley, California, CA. New york city building energy map; 2007. ⟨http://www.visualizing.org/visualizations/new-york-city-building-energy-map⟩.
[15] Granderson J, Piette MA, Ghatikar G, Price P. Building energy information systems: state of the technology and user case studies. In: Handbook of web based energy information and control systems, 2009.
[16] IBM Research and the IBM Cognos Software Group. Many eyes: public building energy consumptions, 2013.
[17] Google PowerMeter. ⟨http://www.google.com/powermeter/about/⟩ [viewed 17.06.2013].
[18] Shi L, Liao Q, He Y, Li R, Striegel A, Su Z. SAVE: sensor anomaly visualization engine. In: 2011 IEEE conference on visual analytics science and technology (VAST). IEEE; 2011. p. 201–10.
[19] Buevich M, Rowe A, Rajkumar R. SAGA: tracking and visualization of building energy. In: 2011 IEEE 17th international conference on embedded and real-time computing systems and applications (RTCSA), vol. 2. IEEE; 2011. p. 31–6.
[20] Burch M, Weiskopf D. Visualizing dynamic quantitative data in hierarchies. In: Proceedings of international conference on information visualization theory and applications; 2011. p. 177–86.
[21] Bellala G, Marwah M, Arlitt M, Lyon G, Bash C. Following the electrons: methods for power management in commercial buildings. In: Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining. ACM; 2012. p. 994–1002.
[22] Bellala G, Marwah M, Arlitt M, Lyon G, Bash CE. Towards an understanding of campus-scale power consumption. In: Proceedings of the 3rd ACM workshop on embedded sensing systems for energy-efficiency in buildings. ACM; 2011. p. 73–8.
[23] Keim DA, Ankerst M, Kriegel HP. Recursive pattern: a technique for visualizing very large amounts of data. In: Proceedings of the 6th conference on visualization'95. IEEE Computer Society; 1995. p. 279–86.
[24] Lammarsch T, Aigner W, Bertone A, Gartner J, Mayr E, Miksch S, et al. Hierarchical temporal patterns and interactive aggregated views for pixel-based visualizations. In: 2009 13th International conference on information visualisation. IEEE; 2009. p. 44–50.
[25] Weber M, Alexa M, Müller W. Visualizing time-series on spirals. In: Proceedings of the IEEE symposium on information visualization, 2001. p. 7–13.
[26] Oelke D, Janetzko H, Simon S, Neuhaus K, Keim DA. Visual boosting in pixel-based visualizations. Comput Graph Forum 2011;30(3):871–80.
[27] Keim DA. Designing pixel-oriented visualization techniques: theory and applications. IEEE Trans Vis Comput Graph 2000;6(1):59–78.
[28] Healey CG. Choosing effective colours for data visualization. In: Proceedings of Visualization'96. IEEE; 1996. p. 263–70.
[29] Wang L, Kaufman A. Importance driven automatic color design for direct volume rendering. Comput Graph Forum 2012;31(3pt4):1305–14.
[30] Keim DA, Kriegel HP. Issues in visualizing large databases. In: Proceedings of conference on visual database systems (VDB'95), Lausanne, Schweiz, 1995. p. 203–14.
[31] Kosara R, Miksch S, Hauser H. Semantic depth of field. In: Proceedings of IEEE symposium on information visualization 2001 (INFOVIS 2001). IEEE Computer Society Press; 2001. p. 97–104.
[32] Giusti A, Taddei P, Corani G, Gambardella L, Magli C, Gianaroli L. Artificial defocus for displaying markers in microscopy z-stacks. IEEE Trans Vis Comput Graph 2011;17(12):1757–64.
[33] Johnson B, Shneiderman B. Tree-maps: a space-filling approach to the visualization of hierarchical information structures. In: Proceedings of IEEE conference on visualization, 1991 (Visualization'91). IEEE; 1991. p. 284–91.
[34] Shneiderman B. Tree visualization with tree-maps: 2-d space-filling approach. ACM Trans Graph 1992;11(1):92–9.
[35] Schreck T, Keim DA, Mansmann F. Regular treemap layouts for visual analysis of hierarchical data. In: Proceedings of the spring conference on computer graphics (SCCG'2006). Casta Papiernicka, Slovak Republic: ACM Siggraph; 2006. p. 184–91.
[36] Hao MC, Dayal U, Keim DA, Schreck, T. Importance-driven visualization layouts for large time series data. In: IEEE symposium on information visualization, 2005 (INFOVIS 2005). IEEE; 2005. p. 203–10.
[37] Bruls M, Huizing K, van Wijk J. Squarified treemaps. In: Proceedings of the joint eurographics and IEEE TCVG symposium on visualization (VisSym'00). Eurographics Association; 2000. p. 33–42.
[38] Baker MJ, Eick SG. Space-filling software visualization. J Vis Lang Comput 1995;6(2):119–33.