Visual Analytics System for Energy Data in Smart Cities and Buildings

Ânderson Pinto Alves, Alessandra M. P. Milani, Isabel H. Manssour
Pontifical Catholic University of Rio Grande do Sul, PUCRS. School of Technology. Porto Alegre, Brazil.
anderson.pinto@acad.pucrs.br, alessandra.paz@acad.pucrs.br, isabel.manssour@pucrs.br

***Abstract*—New sensors and devices are being incorporated in modern buildings and cities to facilitate the understanding of its dynamics and improve its efficiency. With this in mind, the combination of information technology and sensors capable of capturing and sharing energy data with other devices can help solve power consumption problems. However, large volumes of these data are collected and stored uninterrupted, becoming a challenge to analyze it to confirm trends, identify hidden patterns, and outliers to aid in decision-making. As an alternative to address the issue, this paper presents an interactive visual analytics system to understand energy data coming from a smart building or city. This system provides ways to analyze the data at different levels of time granularities for the identification of energy trends, patterns, and data outliers. Moreover, it combines different algorithms that allow fulfilling predictive analysis. An analysis with domain experts demonstrates the feasibility and advantages of using the system to monitor energy data.**

***Index Terms*—Visual Analytics, Energy Data, Energy System.**

I. Introduction

The concept of Smart Cities (SC) has been widespread in recent years [1]. It represents a multidisciplinary field characterized by technological advances to improve city infrastructure, citizens' quality of life, and urban development in several areas, such as health, economy, and mobility [2]. Also, it provides real-time information to deal with future challenges and to manage available resources more efficiently [3].

Since the technological advances provided many benefits to cities using integrated sensor systems, these resources began to be used in building construction and operation. Thus, Smart Buildings (SB) that combines automatic control devices for their technical and administrative systems started to receive visibility. The concept of SB emerged in the US in the 1980s, when security and lighting automation systems began to be connected [4]. A building is considered smart when it can offer a productive and cost-effective environment through optimizing systems, services, structure, and management [5].

It is a challenge for analysts working in the scope of SC and SB to analyze the generated data [6]. Mainly because these data are coming from several sources, and they are available in various formats. Most of these data correspond to time series data, which is data collected sequentially through time. Time-series data analysis performs an essential role in system investigation through the modeling, behavior prediction. It allows describing the data's dependence on time and predicts future values [7].

Energy data originated from different electrical appliances, or generated by clean energy sources, is an example of data usually collected. Understanding this data is one of the major tasks for assessing energy efficiency, which aims to reduce energy consumption and costs, maintaining the same quality of services [8]. Thus, the analysis of such data may generate information that can be used in strategic and tactical decisions in SC and SB.

In this context, interactive visualization techniques can help in the analysis of these collected data, communicating information clearly and effectively. The visual representations can contribute to the understanding of large datasets, enabling the analysts to observe issues and gain insight for helping decision-making [9]. Although exciting systems or approaches in the scope of visual analysis of energy data have been presented [10]–[12], it remains the opportunity to (i) explore visualization techniques associated with statistical analysis, (ii) provide forecast algorithms, and (iii) alternatives to explore the data with different granularity.

Thus, the main goal of this work is to present a visual analytics system for energy data to support decision-making in SC and SB. It allows the simultaneous analysis of meteorological data and consumption or generation sources of energy data, such as air conditioners and solar energy, respectively.

Our main contributions are:

- Visual design that allows the analysis of energy data of different sources over time and with different granularity, with the possibility of correlating meteorological data;
- Suitable interaction techniques based on the details on demand [13] integrated with Coordinated Multiple Views (CMV) [14] to support visual analysis;
- Statistical functions for data understanding;
- Prediction algorithms to foresee energy data based on past energy data patterns.

These contributions are directly associated to knowledge discovery, considering that several techniques are used to stimulate an active interaction of the end-user. The remainder of the paper is organized as follows. Section II presents some related work. The proposed system is described in Section III. Section IV details the results of a preliminary analysis of our system through the presentation of case studies and the feedback obtained with domain experts interview. A discussion about the obtained results and future work directions are presented in Section V. We end this paper with our conclusions.

978-1-7281-8294-0/20/$31.00 ©2020 IEEE

II. Related Work

To identify relevant and recent related work, we performed three main steps. First, we searched for papers in the visual analytics area related to the theme of energy consumption digital libraries such as IEEE and ACM, resulting in an initial set of 13 papers. Afterward, the snowballing search method [15] was applied, which resulted in 33 papers. Then, we applied the following five exclusion criteria for papers: (a) not available in the digital libraries that we have access; (b) not addressing visual analysis or predictive analysis and data visualization techniques; (c) not written in English; (d) with less than 4 pages; or (e) published earlier than 2017. The result of this third step led to the selection of 10 papers to proceed with a more in-depth analysis that is summarized below.

Nine of those papers support the visual representation of data with different levels of granularity (for example, annual, monthly, weekly, daily, or hourly) and allow the visualization of several energy data obtained from different sources [16]–[24]. Among these studies, Chin et al. [16] describe the development of a project to monitor and analyze efficiently wind energy data in real-time. To examine correlations between the energy generated and the wind speed, they use visualization techniques such as radar, parallel coordinates, and glyphs.

Correlation is also mentioned by Blanco et al. [17] and Garcia et al. [22], both presented several ways to visualize correlations using histograms. Blanco et al. proposed a visual analytics approach based on data cube methods to provide several visualizations of energy consumption in a group of public buildings. Garcia et al. aimed to integrate the user's domain knowledge through a visual analysis web application using bar graphs, heatmaps, or calendars. In addition to having these characteristics in common, they still offer a way to load different datasets to be explored.

Almost all the related work has an interactive approach and seven of then provide several visualization techniques [16], [18], [19], [22]–[25]. Prouzeau et al. [19], e.g., describe an interactive Dashboard to view and compare energy consumption with HVAC (Heating, Ventilating, and Air Conditioning) data. Georgina et al. [23] and Nilsson et al. [18] also present a Dashboard, and Ahmed et al. [24] develop an interactive visualization tool for the energy consumption of their household devices and compare them with usage patterns of the past.

Only the work of Chou et al. [25] addressed the prediction algorithm and statistical functions as part of their solution. Other authors presented solutions only applied to predictive algorithms, such as Cui et al. [21] or just statistical methods, such as Georgina et al. [20].

Regarding the commercial systems available to analyse and monitor the energy data, we identified three: EnergyPlus [26], LoadProfileGenerator [27], and Tableau [28]. Although EnergyPlus and LoadProfileGenerator present essential features to simulate energy consumption, they do not offer the analysts options to visualize and interact with the data over time. Moreover, they do not support the inclusion of predicting algorithms. Tableau is not a specific tool for energy data analysis. Thus, although it contains features that serve this purpose, it does not allow an integrated and interactive visualization with different levels of data granularity over time.

Analyzing this related work, we identify the opportunity to offer new ways of analyzing energy data, with the possibility of exploring the data with different granularity, using prediction algorithms, providing statistical functions, and allowing to identify correlations between different energy data, or between energy and meteorological data.

III. Visual Analytics System

In this section, we present the requirements and functionalities of the proposed system, as well as its architecture, the input data format, and the prediction algorithms provided.

A. Requirements

To identify requirements for the development of the system, besides the analysis of related works, we applied informal interviews with eight specialists in the area of energy efficiency. Their knowledge and experience allowed us to identify opportunities for improvement to meet their expectations. So, we were able to identify the importance of exploratory knowledge discovery to understand energy data and to define the following key requirements (R):

- **R1** Offers a simple input data format allowing to load several datasets;
- **R2** Supports data visual representation with different levels of granularity (e.g., annual, monthly, weekly, daily or hourly);
- **R3** Creates an interactive approach;
- **R4** Provides different data visualization techniques;
- **R5** Allows the visualization of several energy data obtained from different sources (e.g., electric shower, air conditioning or appliances);
- **R6** Provides statistical analysis to facilitate data understanding;
- **R7** Allows the application of different prediction algorithms to foresee energy data.

Next we describe the visualization and interaction techniques, and algorithms used to meet these requirements.

B. Architectural Design

The system's architecture comprises five components, as shown in Figure 1. The view page (1) is responsible for displaying all visual information to the user. The Highcharts Javascript Charting Library (2) generates all charts in the system. The component of interpreting files (3) is responsible for reading all information from JSON files, while (4) corresponds to the API created in Python to generate the predictions. All components are integrated with (5), which was developed using JavaScript and PHP merged with the jQuery library. This architecture was designed to be extensible, allowing the inclusion of other types of charts and prediction algorithms, and enabling other file formats as input data. This project is available on Github[1].

[1] https://github.com/DAVINTLAB/VA-EnergyData/

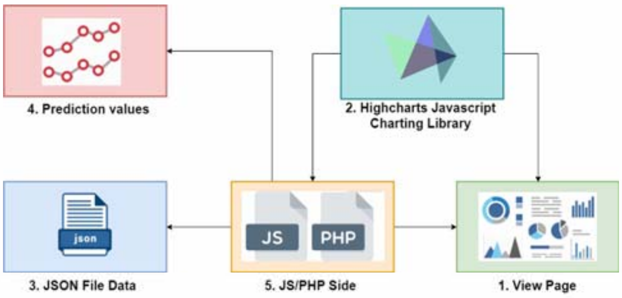

Fig. 1. System architecture and all components.

C. Input Data

Energy data could be collected from different sources and could be stored in different ways. It is essential to provide an input standard format and to normalize scalar values to allow a correct visual representation. Thus, we provide an efficient way of input data in our system and build a visual strategy design to simultaneously display multiple energy data values.

The energy data corresponds to time series data and needs to have an identification, date and hour, and the consumption or generation value of each source. This information is required and must be included in the structure inside the JSON file, which is the supported input format (related to **R1**).

Considering the provided visualization techniques (see Subsection III-E), it is recommended to use a dataset with information of at least one year. More than one year is required because the prediction algorithms use historical data to generate future values. Our system supports the visualization from different sources of energy data in different periods over time, such as annual, monthly, weekly, and daily. However, if the provided dataset only contains, e.g., data for a few months of the year, it will also be possible to visualize it.

The energy data values must be standard scalar values. If a dataset has values between 0 to 100, we consider that 0 represents the lowest and 100 represents the highest energy data value. These values do not need to be 0 and 100, but it is required to be in a scalar interval because the numerical scalar interval is used to generate the color scale. However, if the dataset has an outlier, the analysis may be prejudiced. For instance, if a normalized scalar dataset has values between 0 and 500, but it also has one value 1,000, the color scale will vary from 0 to 1,000.

D. Interactive Exploration

Our system provides several interaction techniques to allow a large dataset to be analyzed and explored at distinct periods throughout the year and filtered using a *zoom out to zoom in* approach. It starts with the visualization of annual data and allows us to progressively decrease the data granularity, in months, weeks, days, and hours, as the lowest granularity (supporting **R2**).

To provide a way to connect two or more views of the same data, we decided to use the details on demand technique (described by Shneiderman [13] as "overview first, zoom and filter, then details on demand") integrated with CMV [29], which is used in many applications and visualization tools for interactive exploration [30]. Our system offers the possibility to offer several views of these data through a few clicks. Thus, every time a click event occurs in any of the provided visualizations, all associated charts are updated to display just the data related to the selection (supporting **R3**).

E. Visualization Techniques

Figure 2 presents the main interface of our system. The following subsections describe each visualization technique designed for it. The dashboard structure is divided into eight different blocks: a) Presents energy data in an annual view separated into months; b) Presents energy data in an annual view separated into days and hours; c) Displays temperature data for a city in an annual view separated into days and hours; d) Shows energy data in hours for a selected week and highlighting the energy peaks; e) Displays energy data for a selected day; f) Presents energy data for a specific hour throughout the week; g) Presents the predicted energy values; h) Displays the monthly energy data.

1) Calendar Visualization: As shown in Figure 2A, calendar visualization groups daily data and allows us to identify patterns across multiple time granularities simultaneously (supporting **R2**). This technique is interesting for rendering large datasets because it enables them to display thousands of values on a single screen. These data are on an annual basis showing the sum of the energy data of each day throughout the year. The red color represents higher energy consumption or generation. Below each month, the week of the year, and the average of the energy data of each week, are presented.

2) Dense Pixel Display: Pixel-oriented techniques, or dense pixel display, map each data value to individual pixels and create a filled polygon to represent each data dimension [31]. As calendar visualization, it allows showing thousands of values in a single screen, however, with a different approach: while the first one presents the sum of the consumption per day, dense pixel display presents the hourly consumption in all days of a year. Each data value controls the color for a single pixel, and new features and data patterns can be revealed according to data selection. Here, this technique is responsible for displaying information about both the energy data of any dataset collected in a time interval not greater than one year (Figure 2B), and meteorological data, such as temperature or air pressure (Figure 2C). Thus, 8,760 one-year records (or 8,784 for leap years) can be displayed, showing the energy consumed or generated of every hour per day for one year, with a vertical axis displaying the hours of a day and horizontal axis the months of the year. Due to the number of data records to be visualized, it became essential to use this technique, which takes full advantage of screen space.

3) Heatmap: The heatmap is a way to analyze the data with a weekly granularity (supporting **R4**), as shown in Figure 2D. It has the advantage to allow the identification of a pattern over a week selected by the user from the dense

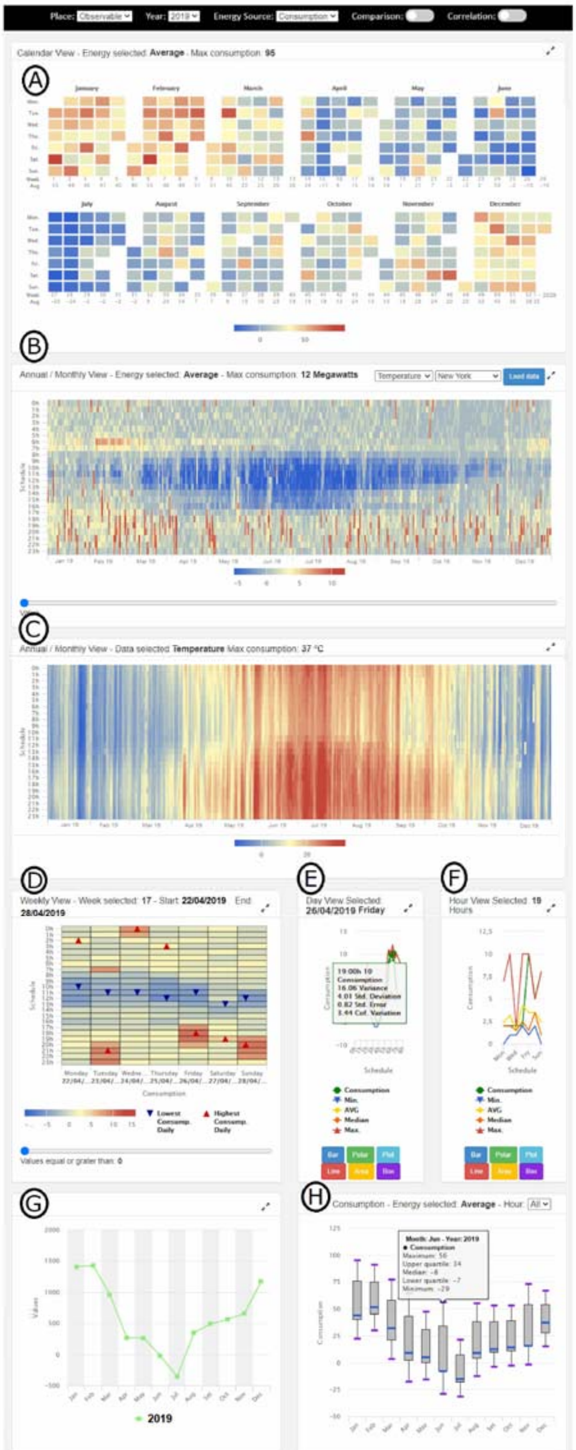

Fig. 2. The main interface displaying energy and weather data.

pixel display. Likewise, selecting a cell in this chart allows the visualization of hourly energy consumption in another chart (see Subsection III-E4). The triangle icon identifies the lowest (blue and faced down) and the highest (red and faced up) peaks of energy consumption data, or energy generation data, each day of the week.

4) Charts to daily and hourly visualization: Line, polar, clock, area, and bar charts were also provided in the proposed system to allow a daily and hourly analysis. However, they can be shown only after the selection of a cell in the heatmap. This means that these charts present the energy consumption data or energy generation data that corresponds to the day of the week of the selected cell. An example of these charts is shown in Figures 2E and 2F.

5) Boxplot: The boxplot has been widely used as an exploratory or communicative tool for data analysis. It typically shows the median, mean, confidence intervals, and outliers of a population, and is especially useful for comparing distributions across groups. The boxplot is used in our system to allow the identification of patterns of monthly consumption, outliers, median, trends, and variability outside the upper and lower quartile, as illustrated in Figure 2H.

F. Statistical Analysis

Statistical analysis allows analyzing numerical data to enable to maximize its interpretation, understanding, and usage [32]. Due to its importance, we have included in our system the following statistical analysis: median, variance, standard deviation, standard error, and coefficient of variation (supporting **R6**). These functions are applied together with the daily and hourly charts, allowing to verify dispersion of data throughout the day, as shown in Figure 2E.

G. Prediction of Energy Data

To use a predictive analysis algorithm in our system (supporting **R7**), first, we investigated data availability, type, structure, context, and size. Thus, since we are dealing with time-series data, its analysis allows describing a relationship between the data on time and prediction values. Figure 2G illustrates the prediction values presentation in our system.

Then, we addressed three different predictive analysis algorithms in our system: Autoregressive [33], Holt-Winters [34], Sarima [35], and Prophet [36]. The main reason to use these algorithms is that they are suitable for time series data, allowing to define the dependence of the data respecting the time and to predict future values [7]. The application of these prediction algorithms in our system generated an interesting assertiveness in the predicted values.

To determine the assertiveness of the values generated by these models, we used the Mean Absolute Error (MAE) [37] and percentage error measurements. Figure 3 shows an example of the prediction values generated by these algorithms. Details about these predictions are described in Section IV-A.

H. Comparative Analysis

Compare different sources of energy consumption or energy generation data allows to carry out several analyzes. To analyze this data over time, it is desirable to have easy interaction between different visualizations. This resource was developed aiming to explore different data sets in different years and different periods throughout the year (supporting

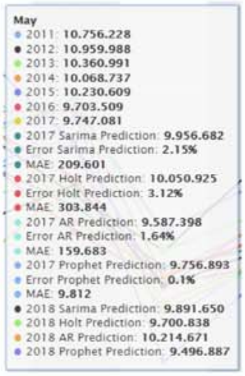

Fig. 3. Display of the month of may consumption for years 2011 to 2017 [38] and the forecast values for this month resulting from each prediction algorithm. MAE metrics and percentage errors are also exhibited.

requirement **R5**). The comparison feature of different energy sources was implemented through mobile pop-up windows, with the possibility to customize its size according to the user's preference and interact with the data in various ways. An example of analyzing multiple energy sources would be to select the energy sources of interest individually and open a window for each one by clicking on calendar visualization or dense pixel display. Since there is no limit for the creation of windows in the system, numerous pop-ups from different sources of energy can be created and placed side by side for comparison.

I. Correlation

The system allows us to do the correlation of energy data with climatic data. To view the correlation, it is necessary to load energy and weather data for the same year and then a line chart is plotted. Figure 4 shows an example of this chart. It shows the relationship between the temperature data and energy consumption data for the year 2004 of an SB located in the city of Fairbanks, AK. To measure the degree of correlation between these two variables (energy consumption and temperature) we applied the Pearson's correlation coefficient [39]. The value generated by the Pearson coefficient between these two variables is 0.2919, indicating a weak correlation [40].

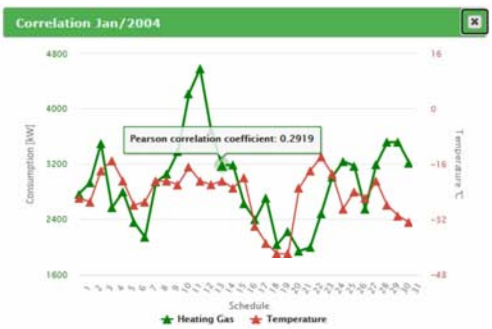

Fig. 4. The correlation between energy consumption (green line) and temperature (red line) in the value of 0.2919.

IV. Case Studies and User Evaluation

The following subsections present two case studies and a preliminary evaluation with domain experts to demonstrates the applicability of our approach.

A. Case Studies

To confirm the applicability of our system, we use it to analyze five different datasets. The first one corresponds to energy data collected from an American smart home [41], which has solar panels. These data were collected in the year 2019 and are presented in Figure 2. We can see that the highest energy consumption occurs in January, February, and December. Negative values correspond to the days when the generation of energy was higher than consumption, which occurred in the summer close to noon.

The second dataset was obtained at Kaggle [38] and contains over 5 years of hourly energy consumption data. It was provided by PJM Interconnection LLC, a regional broadcasting organization in the USA. Figures 3 and 4 were generated using this dataset. With 5-year data, we obtained good results with the prediction algorithms, as shown by the low error values in Figure 3.

A public dataset [42] with energy consumption data from commercial and residential buildings in the USA was the third one used in our system. It has hourly data for a year from several locations in the USA. Figure 5 illustrates the Dense Pixel Display of the year 2004 that includes several sources of energy consumption from one commercial building (e.g., water heater, electricity, and gas heating). Analyzing this visualization, it is possible to see that the highest energy consumption occurs from November to March, probably due to the harsh winter. Furthermore, the blue strip highlights the beginning and end of the daylight saving time and reveals a low energy consumption from 1 a.m. to 4 a.m..

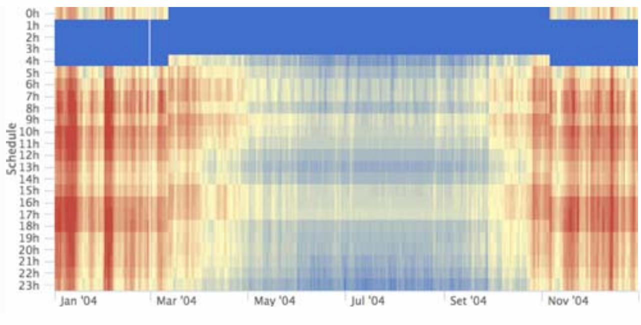

Fig. 5. Dense Pixel Display with energy consumption data from one commercial building [42].

The fourth dataset has high temporal resolution data from various climatic attributes, such as temperature, humidity, air pressure, and wind speed. This information is related to 30 cities in the USA, Canada, and Israel [43]. These data were shown in Figure 2C, acquired using the Weather API on the OpenWeatherMap website and are available under the license ODbL [44].

The climate data from Fairbanks (see Subsection III-I), was the last used dataset. These data were obtained through the OpenWeatherMap API [45].

B. User Evaluation

According to Boyce and Neale [46], in-depth interviews are a qualitative research technique that involves conducting intensive individual interviews with a small number of respondents to explore their perspectives on a particular idea, program, or situation. Thus, to evaluate the effectiveness of our approach, we conducted interviews with four domain specialists.

Each participant was interviewed in sessions that lasted from 60 to 90 minutes. Although the same environment configuration was used for all participants, one interview was face-to-face and three were online conversations. To start, we introduced our study, the procedure, and the consent form, in compliance with the Research Ethics Committee[2] process.

We developed a semi-structured questionnaire divided into three parts with a total of 27 questions to guide the interviews. Part one aimed to capture data about the participants' profile. Part two focused on their work practice and their perspective on how visualization can support data analysis. Finally, part three is considered a practical evaluation of our proposed system. The main insights obtained during the review of the answers are described in the next subsections.

1) Participants' Profile: The participants were selected because of their experience in the energy efficiency field, which ranged from two to 25 years. They hold positions in the industry as an energy efficiency analyst, energy specialist coordinator, and energy services coordinator. One of them works as a researcher and professor at a University. All of them had received a master's degree in engineering related fields and one has a degree in accounting.

2) Work practices: Energy Plus [26] and Excel were indicated as the most frequently used tools for data analysis. Other proprietary systems provided by sensor manufacturers are also mentioned, e.g., Notus Target, and GRAFMED system. Moreover, all participants reported using tabular datasets, managing file types such as XLSX, CSV, or TXT. The largest size of the dataset was indicated as close to 25 Gigabytes. The data are collected from different sources: public substations, private generators, and proprietary systems.

The main activity when performing an energy data analysis mentioned by the participants is to check the consumption curve, together with the consumption trend. However, two participants indicated the evaluation of the dataset looking for data quality issues, i.e., missing data and outliers, as one of the first activities, before analyzing the consumption.

One participant indicated that to verify any behavior outside the consumption curve or any deviation, the sensor manufacturer's systems are used to assist with the visual analysis. While another participant preferred to work exclusively with spreadsheets since they provide essential statistical analysis easy to manipulate, in his opinion, other tools lack this feature.

[2]http://plataformabrasil.saude.gov.br/ - CAAE: 26356419.6.0000.5336

As the main challenges, they mentioned missing data as a frequent data quality issue, due to sensors' problems, delays in the collection of information, and communication failures. One participant reported to work simultaneously on different projects, and he has difficulties in visualizing the project data as a whole. He also misses exploring energy data from various projects at the same time in interactive dashboards.

3) Evaluation: We requested the participants to explore two different datasets in our system. The first one has the energy consumption of an American electric utility for a period of six years (from 2011 to 2016) [38]. They were supposed to: a) View the data on an annual basis; b) Identify the month with the highest and lowest monthly consumption; c) Compare consumption between two months; d) Compare consumption between two weeks; e) Analyze the predictions generated. The other dataset has energy consumption data from one year from a commercial building [42]. They were supposed to: a) View the data on an annual basis; b) Compare consumption between two energy sources in the same period; c) Identify the highest and lowest consumption in a given week of the year for two sources of energy.

All the participants were able to view, compare, and analyze the data, and they indicated our proposed system as easy to understand and manipulate. Although two participants mentioned that our system could be considered a complete solution since it offers the visualizations and features that meet all their expectations, when we invited the participants to indicate the items that were challenging or if any improvement opportunities could be noted, they mentioned the following.

One participant suggested that it would be of great value if our system could support non-standard consumption detection algorithms, indicating if one device is consuming more energy than another or if a particular consumption is outside the normal range. Moreover, two participants suggested an indication of the best prediction algorithm. The system displays data from four prediction algorithms but does not display which algorithm has the lowest average monthly or annual error rate.

Regarding the visualizations, one participant considered challenging to understand the Boxplot chart because she was not very familiar with this technique. Another participant was not familiar with the Dense Pixel Display visualization and got confused in the first view since it provides a lot of information.

Finally, one participant suggested that our system could provide an association of energy consumption with financial cost data. He suggested that the system could identify the monetary value of kilowatts per hour, and thus be able to compare or correlate these values with energy consumption.

During the participants' evaluation process, they highlighted features favorable to the adoption of our system.

- It was easy to identify the consumption trends throughout the year and consumption patterns in different seasons, e.g., viewing the highest and lowest energy consumption.
- The use of prediction algorithms. One participant reported that mainly due to the error rate, as it helps to predict both future consumptions, as the purchase and sale of energy. When analyzing the graph with the predictions, another

participant informed that it would have an application in his company, making it possible to estimate the operational costs of buying and selling energy.

- The application of statistical functions, mainly, to offer options they use spreadsheets to compute.
- The interactive dashboard was reported as a significant differential since our system offers, for instance, time granularity in which the data can be analyzed.
- The display over time is instructive, as the interactions and possibilities of visualization in different time granularities facilitate the refinement and identification of energy consumption.
- The possibility of comparing data in different datasets.

Based on the feedback received with the interviews, our system proved to be applicable in different contexts of use. All participants manifested interest in using it during their daily activities. As the main advantages for its adoption, they cited the subsequent refinement of time, interactivity features to manipulate the data, the possibility of comparing different data, and the prediction algorithms. Lastly, it demonstrated to be adequate and efficient to visualize energy and meteorological data and support the decision making of the energy analysts.

V. Discussion

In this section, we address the lessons learned, future work, and some experiments and limitations. We point out a few difficulties, the feedback obtained with the interviewees, and what are the advantages of our proposal. We also discuss system limitations and some directions for future research.

A. Lessons learned

The first lesson learned with this work is there was an apparent lack of a Visual Analytics System with different interactive visualization techniques integrated with statistical analysis and prediction algorithms. About system architecture, we consider it positive to develop a web system using the Highcharts Javascript Charting Library. In addition to facilitating its use, this library allows dynamically and interactively generation of charts. Moreover, considering data input, it is desirable to allow loading data in XLSX or CSV format, something that we intend to do now.

The related work analysis, as well as the initial informal conversations with expert users, helped us to identify some requirements. Unifying the requirements raised through these conversations with the opportunities identified was a major challenge. For system validation, in addition to interviews with domain experts, we used our system with different data sets through some case studies. Thus, we could confirm that the system offers several benefits such as the identification of patterns and behaviors through the visualization with different levels of granularity, analyzing different datasets, and predicting future behaviors.

B. Future work, Experiments and Limitations

To identify the amount of input data that our system supports to manage, we did the following experiment. We developed an algorithm to simulate energy data and generate five datasets contained 1,140 sources of energy consumption, randomly informing consumption values for each hour throughout the year. We generated a total of more than ten million records for each created file, totaling more than fifty million records.

We loaded this large dataset in our system and we first calculate the overall average of all energy sources contained in these datasets to make sure we had access to all of them. After, we tried to visualize and interact with it. Then, we noticed that it was possible to visual analyze it with a response time of approximately 30 seconds (only for this specific case considering the amount of data loaded) using a 2.2GHz MacBook Pro notebook, with an I7 processor and 16GB. We have tried to increase this data, but then the system did not behave properly. Thus, we can say that it supports a dataset with approximately fifty million values.

One limitation we faced during the entire development of this study was to obtain different datasets for case studies, both for energy and climatic data. This difficulty is associated with both, public and private companies that have information security contracts, preventing the availability of these data for the general public.

As future work, we want to test the system with different datasets. We also aim to apply other prediction algorithms for this specific data type and evaluate the inclusion of other correlation techniques. Another improvement is the inclusion of frameworks such as Elastic Stack [3] to enable receiving data in real-time. Also, we will analyze whether there is a need to provide other statistical functions and visualization techniques.

VI. Conclusion

This paper aimed to describe our visual analytics system for energy and meteorological data for SB and SC. The proposed system assists the decision making related to energy consumption data or energy generation data. It allows you to visualize, explore, predict, compare, and correlate data at different levels of granularity over time.

Our main contributions are: Analysis and comparison of meteorological data and energy data over time and with different granularities; Interaction techniques that support visual analysis; Availability of statistical analysis and prediction algorithms; Domain expert evaluation demonstrating the system advantages. Through the proposed system it is possible to identify anomalies, trends, and behavior patterns. Lessons learned and future research directions presented in Section VI demonstrate that this is a relevant topic for SC and SB.

Acknowledgment

This work was achieved in cooperation with Hewlett Packard Brasil LTDA. using incentives of Brazilian Informatics Law (Law N 8.2.48 of 1991). Also, we would like to thank OpenWeatherMap API to kindly provided us with data to validate this paper.

[3]https://www.elastic.co/elastic-stack

REFERENCES

[1] H. Chourabi, T. Nam, S. Walker, J. R. Gil-Garcia, S. Mellouli, K. Nahon, T. A. Pardo, and H. J. Scholl, “Understanding smart cities: An integrative framework,” in *2012 45th Hawaii International Conference on System Sciences*, Jan 2012, pp. 2289–2297.

[2] C. S. Lai, Y. Jia, Z. Dong, D. Wang, Y. Tao, Q. H. Lai, R. T. Wong, A. F. Zobaa, R. Wu, and L. L. Lai, “A review of technical standards for smart cities,” *Clean Technologies*, vol. 2, no. 3, pp. 290–310, 2020.

[3] V. Albino, U. Berardi, and R. Dangelico, “Smart cities: Definitions, dimensions, performance, and initiatives,” vol. 22, p. 2015, 02 2015.

[4] B. Morvaj, L. Lugaric, and S. Krajcar, “Demonstrating smart buildings and smart grid features in a smart energy city,” in *Proceedings of the 2011 3rd International Youth Conference on Energetics (IYCE)*, July 2011, pp. 1–8.

[5] S. Katipamula, “Smart buildings can help smart grid: Transactive controls,” in *2012 IEEE PES Innovative Smart Grid Technologies (ISGT)*, Jan 2012, pp. 1–1.

[6] L. Piovano, D. Garrido, R. Silva, and I. Galloso, “What (smart) data visualizations can offer to smart city science,” *Communications & Strategies*, no. 96, p. 89, 2014.

[7] M. H. Becker, “Modelos para previsão em séries temporais: uma aplicação para a taxa de desemprego na região metropolitana de porto alegre,” 2010.

[8] B. Ang, W. Choong, and T. Ng, “Energy security: Definitions, dimensions and indexes,” *Renewable and Sustainable Energy Reviews*, vol. 42, pp. 1077 – 1093, 2015.

[9] S. Card, J. Mackinlay, and B. Shneiderman, “Information visualization,” *Human-computer interaction: Design issues, solutions, and applications*, vol. 181, 2009.

[10] S. Barns, “Smart cities and urban data platforms: Designing interfaces for smart governance,” *City, Culture and Society*, 2017.

[11] M. Brehmer, J. Ng, K. Tate, and T. Munzner, “Matches, mismatches, and methods: Multiple-view workflows for energy portfolio analysis,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, pp. 449–458, Jan 2016.

[12] S. Goodwin, J. Dykes, S. Jones, I. Dillingham, G. Dove, A. Duffy, A. Kachkaev, A. Slingsby, and J. Wood, “Creative user-centered visualization design for energy analysts and modelers,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, pp. 2516–2525, 2013.

[13] B. Shneiderman, “The eyes have it: a task by data type taxonomy for information visualizations,” in *Proceedings 1996 IEEE Symposium on Visual Languages*, Sep. 1996, pp. 336–343.

[14] J. C. Roberts, “On encouraging multiple views for visualization,” in *Information Visualization, 1998. Proceedings. 1998 IEEE Conference on*. IEEE, 1998, pp. 8–14.

[15] C. Wohlin, “Guidelines for snowballing in systematic literature studies and a replication in software engineering,” in *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, New York, USA, 2014, pp. 1–10.

[16] G. Chin, Y. Chen, E. Fitzhenry, B. McGary, M. Pirrung, J. Bruce, and S. Winner, “A visual analytics platform and advanced visualization tools for interpreting and analyzing wind energy time-series data,” *IFAC-PapersOnLine*, vol. 51, no. 28, pp. 480 – 485, 2018.

[17] I. D. Blanco, A. A. C. Vega, D. P. López, M. D. González, S. A. Castro, and M. Á. P. Medrano, “Energy analytics in public buildings using interactive histograms,” *Energy and Buildings*, vol. 134, pp. 94–104, 2017.

[18] A. Nilsson, M. Wester, D. Lazarevic, and N. Brandt, “Smart homes, home energy management systems and real-time feedback: Lessons for influencing household energy consumption from a swedish field study,” *Energy and Buildings*, vol. 179, pp. 15–25, Nov 2018.

[19] A. Prouzeau, M. Dharshini, M. Balasubramaniam, J. Henry, N. Hoang, and T. Dwyer, “Visual analytics for energy monitoring in the context of building management,” in *Proceedings of the International Symposium on Big Data Visual and Immersive Analytics*, Konstanz, Germany, 2018, pp. 1–9.

[20] E. Pickering, M. Hossain, R. French, and A. Abramson, “Building electricity consumption: Data analytics of building operations with classical time series decomposition and case based subsetting,” *Energy and Buildings*, vol. 177, pp. 184–196, Oct 2018.

[21] W. Cui and H. Wang, “Anomaly detection and visualization of school electricity consumption data,” in *Proceedings of the IEEE 2nd International Conference on Big Data Analysis*, Beijing, China, 2017, pp. 606–611.

[22] D. García, I. Díaz, D. Pérez, A. A. Cuadrado, M. Domínguez, and A. Morán, “Interactive visualization for nilm in large buildings using non-negative matrix factorization,” *Energy and Buildings*, vol. 176, pp. 95 – 108, 2018.

[23] G. Wood, R. Day, E. Creamer, D. Horst, A. Hussain, S. Liu, A. Shukla, O. Iweka, M. Gaterell, P. Petridis, N. Adams, and V. Brown, “Sensors, sense-making and sensitivities: Uk household experiences with a feedback display on energy consumption and indoor environmental conditions,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 55, pp. 93–105, Sep 2019.

[24] N. Ahmed and K. Mueller, “Energyscout: A consumer oriented dashboard for smart meter data analytics,” *ArXiv Preprint*, vol. 1, pp. 1–23, Nov 2019.

[25] J. Chou, A. Telaga, W. Chong, and G. Gibson, “Early-warning application for real-time detection of energy consumption anomalies in buildings,” *Journal of Cleaner Production*, vol. 149, pp. 711–722, 2017.

[26] EnergyPlus, “EnergyPlus,” June 2020. [Online]. Available: https://energyplus.net/

[27] N. Pflugradt, “LoadProfileGenerator,” June 2020. [Online]. Available: https://www.loadprofilegenerator.de/

[28] Tableau, “Tableau,” June 2020. [Online]. Available: http://www.tableau.com/

[29] M. Q. Wang Baldonado, A. Woodruff, and A. Kuchinsky, “Guidelines for using multiple views in information visualization,” in *Proceedings of the working conference on Advanced visual interfaces*. ACM, 2000, pp. 110–119.

[30] P. Koytek, C. Perin, J. Vermeulen, E. André, and S. Carpendale, “Mybrush: Brushing and linking with personal agency,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 605–615, Jan 2018.

[31] D. A. Keim, “Designing pixel-oriented visualization techniques: theory and applications,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 6, no. 1, pp. 59–78, Jan 2000.

[32] R. B. Brown and M. P. Saunders, *Dealing with statistics: What you need to know*. McGraw-Hill Education (UK), 2007.

[33] J. Wang and T. Zhang, “Degradation prediction method by use of autoregressive algorithm,” in *2008 IEEE International Conference on Industrial Technology*, April 2008, pp. 1–6.

[34] P. R. Winters, “Forecasting sales by exponentially weighted moving averages,” *Management Science*, vol. 6, no. 3, pp. 324–342, 1960.

[35] K.-Y. Chen and C.-H. Wang, “A hybrid sarima and support vector machines in forecasting the production values of the machinery industry in taiwan,” *Expert Systems with Applications*, vol. 32, no. 1, pp. 254 – 264, 2007.

[36] S. J. Taylor and B. Letham, “Forecasting at scale,” *The American Statistician*, vol. 72, pp. 37–45, abr 2018.

[37] T. Chai and R. R. Draxler, “Root mean square error (rmse) or mean absolute error (mae)?–arguments against avoiding rmse in the literature,” *Geoscientific model development*, vol. 7, no. 3, pp. 1247–1250, 2014.

[38] kaggle, “kaggle,” June 2020. [Online]. Available: https://www.kaggle.com/robikscube/hourly-energy-consumption/

[39] J. Benesty, J. Chen, Y. Huang, and I. Cohen, *Pearson Correlation Coefficient*. Berlin, Heidelberg: Springer, 2009, pp. 1–4.

[40] H. H. H. Renate Buber, *Qualitative Marktforschun*. Universität Dortmund: Gabler, 2009.

[41] M. Bostock, “Electricity usage, 2019,” June 2020. [Online]. Available: https://observablehq.com/@mbostock/electric-usage-2019

[42] O. of Energy Efficiency Renewable Energy (EERE), “Commercial and residential hourly load profiles for all tmy3 locations in the united states,” June 2020. [Online]. Available: https://tinyurl.com/y5avfedo

[43] D. Beniaguev, “Historical hourly weather data 2012-2017,” June 2020. [Online]. Available: https://www.kaggle.com/selfishgene/historical-hourly-weather-data

[44] O. K. Foundation, “Open data commons open database license (odbl),” June 2020. [Online]. Available: https://opendatacommons.org/licenses/odbl/index.html

[45] OpenWeather, “Openweather,” June 2020. [Online]. Available: https://openweathermap.org/

[46] C. Boyce, P. Neale, and P. International, *Conducting In-depth Interviews: A Guide for Designing and Conducting In-depth Interviews for Evaluation Input*, ser. Pathfinder International tool series: Monitoring and evaluation. Pathfinder International, 2006.