JOURNAL OF ENERGY SYSTEMS

2021, 5(3)

2602-2052 DOI: 10.30521/jes.878318 Research Article

Simple and effective descriptive analysis of missing data anomalies in smart home energy consumption readings

Kasaraneni Purna Prakash
School of Computer Science and Engineering, VIT-AP University, Amaravati-522237, Andhra Pradesh, India. kasaraneni.19phd7020@vitap.ac.in

Yellapragada Venkata Pavan Kumar*
School of Electronics Engineering, VIT-AP University, Amaravati-522237, Andhra Pradesh, India. pavankumar.yv@vitap.ac.in

Submitted: *10.02.2021*
Accepted: *08.08.2021*
Published: *30.09.2021*

* Corresponding Author: +917416285804

Abstract: Smart grids evolution is ramping up in the global energy scenario by offering deregulated markets, demand-side management, prosumer culture, demand response, contingency forecasting, outage management, etc., functionalities. These functionalities help to manage the grid effectively by taking informed decisions timely. Further, the progressive developments in information and communication technologies improve smartness in the power grids. Especially, smart homes are playing a key role, which possesses the communication between various devices/appliances and collect their functional data in terms of energy consumption readings, timestamp, etc. However, the availability of high-quality data is always desired to achieve superior benefits with respect to all the above-mentioned functionalities. But, the failures of communication networks, metering devices, server station issues, etc., create anomalies in the data collection. Hence, there is a dire need of identifying the ways of analyzing the smart home data to find the irregularities that occurred because of aforesaid failures. Especially, it has been a common problem to see missing data at some particular instants in the overall database captured. In this view, this paper proposes a simple and effective descriptive analysis to find missing data anomalies in smart home energy consumption data. A real-time dataset is used to execute the proposed method. For which, a clear enumeration of missing data is visualized using comprehensive simulation results. This helps to realize the actual problems that are hidden in the energy consumption data.

Keywords: *Data anomalies, Missing data, Smart home data analytics, Tracebase dataset, Visualization*

Cite this paper as: Purna Prakash, K., Pavan Kumar, Y.V., Simple and effective descriptive analysis of missing data anomalies in smart home energy consumption readings, *Journal of Energy Systems* 2021, *5*(3), 199-220, DOI: 10.30521/jes.878318

1. INTRODUCTION

1.1. Motivation – Role of Big Data in Smart Grids

The elements of the smart grid are always connected in a network and equipped with different sensors which continuously trace the data. This continuous stream of data produces voluminous data [1]. The participation of big data employs grid sustainability and efficiency [2]. The implementation of smart meters in the power distribution system provides the capturing of a large number of energy consumption readings of smart homes (or buildings). The information and communication technology (ICT) realm provides support for the power grids as well as leading to datasets with heterogeneity [3]. Several standards and techniques are being developed as the response of this field is rising [4]. The traditional electrical layer of the grid is added with a new information technology (IT) layer. This IT layer collects the data from the network of smart meters/sensors that are installed to capture data continuously [5], [6].

Big data concepts are discussed thoroughly in terms of theoretical and practical implementations and from the perspectives of energy sources. Also discussed the benefits of using big data models [7]. The generation of vast amounts of data is promoting to development of new tools and applications with respect to the current trends in power systems. In line with this, a new architecture is introduced to discuss the techniques such as collection, storage, analysis, control, and visualization of big data [8]. Further, to obtain better insights and maintain potentiality in energy big data, a smart energy model is referred in [9]. This model serves the smart grid from data generation to demand management.

1.2. Importance, Challenges, and Opportunities of the Data Analytics in Smart Grid Context

Data analytics can overcome the difficulties of traditional methods and techniques which affect processing and performance. It has a great impact on various aspects of the smart grid such as decision making, optimal storage systems, maintaining stability, monitoring of operations and control, etc. To cope up with the above challenges several data analytics tools are discussed in Ref. [1]. It can be applied to various domains of the smart grid such as generation, transmission, etc., [4]. The advent of data analytics and its applications can provide better service to customers. These include utility monitoring, integration of smart homes to maximize energy usage, utilization and generation management, etc., [5]. Further, the use of data analytics in smart grids will manage it effectively and increase the system resiliency by providing bidirectional power/information between utilities and customers [10], [11].

To implement data analytics for smart grids and visualize the benefits of modern analysis methods, one has to focus on different challenges. These challenges can be addressed by utilizing the available opportunities with respect to technology advancements. Some of these critical factors are given as follows:

- The data available in real-time is difficult for processing with respect to the data analytics tools [1].
- The load balancing issues of the grid. The implementation of the machine learning concepts on smart meter/grid data greatly helps in solving these issues [2].
- Embedding big data analytics into a smart grid is being considered a difficult task due to the volume, uncertainty, and computational complexity. Data analytics provide better decision-making over the planning and management of the grid. The gaps/challenges are discussed from the perspectives of utilities and industries [3].
- Big data tools and technologies and their implementation in smart grid technology are at the primary level [4].

- Improper standard format for data collection [5].
- Applying big data analytics to the whole data is not possible at once. The data compression techniques can be applied to get the data at the required level while implementing data mining functionalities [6].
- Security is considered as another important aspect to maintain data privacy and protection [8].

1.3. Analysis of State-of-the-art Literature Works

There were some works of literature that presented the importance of data analytics in smart homes as well as presented some analysis methods. All the literature works related to data analytics in the smart home application are segregated into general works [12]-[26] and works related to specific cases. To understand these specific cases, this paper investigates the research conducted on real-time datasets, viz., U.S.A. based Pecan Street power consumption dataset [27]-[35], Germany and Australia based Tracebase dataset [36]-[63], etc. Table 1, Table 2, and Table 3 elucidate all these state-of-the-art literature work respectively in terms of research works done on general energy consumption data as well as specific datasets.

Table 1. Data analytics based literature works executed on general energy consumption data

Ref.	Year	Dataset	Key Focus and Work Performed
[12]	2021	Phasor measurement units (PMUs) dataset	Presented various lessons learned about reducing barriers to adopting big data analytics on a real dataset. This dataset was captured with almost half-trillion records of 400 PMUs that were distributed through the North American power grid.
[13]	2021	Self-prepared smart home dataset	Proposed a design for smart city energy management by considering Internet of Things (IoT) devices' energy efficiency and data analysis. Along with energy management, this design involves service management and data processing. The energy-effective clustering, optimized scheduling, peak load shedding, and load-balancing concepts are used to realize efficient energy management. Rules and thresholds are used for service management and distributed framework is used for data processing.
[14]	2021	(No specific dataset is represented)	Various models were described as a guideline to design robust and efficient control strategies for smart grids. These models are useful to optimize and improve the whole system's performance against uncertain generation, loading, and communication constraints.
[15]	2020	Electricity customer behaviour trial data, Commission for energy regulation smart metering project, Archive of Irish Social Science data	A polar projection-based method using K-means clustering was proposed to minimize the computational complexity in smart meter data. To achieve this, the load profiles are generated and clustered according to peak and total consumption.
[16]	2020	Variety of datasets related to smart grid	Presented a systematic mapping study on different aspects of data analytics for smart grid applications. It is found that there are ten key aspects such as research aspects, techniques used, status of tool-support, replicability/ reproducibility levels, and research methodology types.
[17]	2020	Household smart energy meter data in UK	Proposed Gaussian-based model to group the customers using data mining techniques, viz., clustering and prediction.
[18]	2019	(No specific dataset is represented)	Focused on big data analytics and machine learning in the process industries. This promotes intelligent and innovative decision-making with no or less human intervention in the process industries.
[19]	2019	Household energy smart meter data by Scottish and Southern Energy (SSE) Supply Ltd.	This work concluded that there was very limited research related to big data analytics in the form of tools, methods, and systems to support data storage and management. To support this study, a system named "Smart Meter Analytics Scaled by Hadoop (SMASH)" was implemented.
[20]	2019	Smart data package in R	A new package "smartdata" was introduced with a common interface in R for preprocessing.
[21]	2018	Ph.D. Projects (Ex: Space weather, air pollution, lake water quality)	It was discussed that statistics and statisticians always had a major role in big data. They implement new techniques and methodologies to handle big data irrespective of its size.

[22]	2018	Google Trends on Analytics, Machine Learning and Artificial Intelligence	Discussed the data analytics importance of the applications like wind power forecasting, fault analysis, security management, and awareness to society. Focused on the stages of analytics and found the use of visualization of prediction.
[23]	2018	(No specific dataset is represented)	The emphasis was on the understanding of energy consumption behavior and load patterns of the end-users.
[24]	2018	Public datasets (Backblaze, StackExchange, NYC taxi and Stars)	A course was designed to enable the students to learn how to process and analyze large datasets. To achieve this, various software and tools such as Hadoop, MapReduce, Spark, etc. were used.
[25]	2018	(No specific dataset is represented)	The implementation of supervised/unsupervised learning methods, the characteristics of transparent power grids, and closed loops were discussed.
[26]	2017	Archive of Irish Social Science data	This work proposed a cloud-based platform called Infrastructure as a Service (IaaS) for the big data framework to analyze smart grid data. It was implemented on the scenarios such as visualization of grid status and enabling the demand response.

Table 2. Data analytics based literature works executed on the pecan street dataset

Ref.	Year	Key Focus and Work Performed
[27]	2020	Suggested the use of regression technique to find correlation and interdependencies among the variables such as cloud cover, wind speed, air pressure, temperature, humidity, etc. This helps to achieve better photovoltaic (PV) output.
[28]	2020	Considered the pecan street dataset as a case study and proposed a model “eDemand” to forecast energy consumption of one building at IIT Bombay.
[29]	2019	The application of heating, ventilation and air conditioning (HVAC) was discussed for regulating frequency. A model was proposed to generate data by monitoring the residential load with respect to frequency regulation.
[30]	2019	Focused on extracting new features (e.g.: energy consumption pattern, consumer behavior, etc.) from historical data of energy consumption to help in demand-response management.
[31]	2019	Proposed a method to detect base-load from daily load profiles, which would help in identifying consumers for energy efficient programs.
[32]	2019	A framework was proposed to schedule the household appliances that were controllable in the proximity of the grid by addressing the cooperative decision-making issue.
[33]	2019	This work presented a framework with a bottom-up method for generating realistic synthetic loads and proposed six energy services and some events with four demographic categories for distribution.
[34]	2018	Presented a co-simulation framework that was capable of capturing power and communication.
[35]	2015	Learning models were proposed to capture the annotations of appliances. These models understand the pattern of human activities to reduce energy consumption.

Table 3. Data analytics based literature works executed on tracebase dataset

Ref.	Year	Key Focus and Work Performed
[36]	2021	Proposed techniques to standardize and fill the gap between load modelling and non-intrusive load monitoring (NILM) in distributed energy networks.
[37]	2021	Implemented NILM algorithms to investigate the performance gap between actual aggregate signals and denoised aggregate signals to understand the effectiveness of aggregators in power system scenario.
[38]	2021	Proposed a deep learning (convolutional) based auto-encoder method for energy disaggregation of the home appliances to properly predict the targeted parameters.
[39]	2021	Performed a wide-ranging review on NILM datasets and highlighted the characteristics, strengths, and limitations of these datasets. This helps the researchers in evaluating the performance of new algorithms of NILM.
[40]	2020	Proposed heuristics to meet fairness criteria and mixed integer programming models to solve the issue of fair electric load shedding in developing countries.
[41]	2020	Introduced a methodology to automatically estimate the power consumption of the individual building through overhead imagery such as satellite, aerial, etc.
[42]	2020	Explored NILM algorithms comprehensively that would help in developing a precise and efficient NILM system to attain better energy management.
[43]	2020	Proposed an IoT-based and low-cost end-to-end solution for recognizing electrical appliances in real-time. This helps in effective building automation management.

[44]	2020	To reduce the adverse effect on Hall Effect sensors, a blind source separation algorithm was proposed. It reduces crosstalk noise made by adjacent wires of the circuit breakers.
[45]	2020	Focused on the preparation of building energy consumption datasets based on key features such as sampling rate, period of collection, location, timestamp, reading, number of appliances, etc. To achieve this, conducted a thorough study on the existing datasets and built a new energy consumption dataset named Qatar University dataset.
[46]	2019	Proposed a novel and effective method based on NILM and graph spectral clustering to predict the energy consumption in an individual house or a collection of houses.
[47]	2019	Discussed a comprehensive study of NILM based methods and compared these methods with machine learning based methods. Further, provided some future research directions.
[48]	2019	Implemented an extensively used data analytics scenario, i.e., event detection to evaluate the influence of sampling rate on load signature analysis.
[49]	2019	Proposed an auto-associative neural network based method to detect electrical appliances for NILM and segregate them based on their nature.
[50]	2019	Proposed an anomaly detection algorithm to distinguish the operating behavior of electrical appliances of residential homes by using their metered data.
[51]	2019	This work presented the challenges (re-sampling, data transformation, etc.,) of energy consumption datasets with heterogeneous data to understand their usefulness in nature.
[52]	2018	Developed an algorithm for NILM tasks in commercial buildings. It was identified that a limited number of public datasets are available on commercial buildings. Hence, a synthetic dataset was released to the public to motivate the researchers to work on commercial building case studies.
[53]	2017	Proposed a novel NILM method for electricity consumers who has a local solar energy generating plant. The proposed method determines the quantity of solar power influx.
[54]	2017	Proposed an advanced NILM concept for demand-side management by conducting an extensive exploration of the NILM applications and issues.
[55]	2017	Introduced an algorithm with the combination of reactive and active power in the additive factorial hidden Markov model for NILM.
[56]	2017	Proposed non-intrusive techniques to handle technical barriers, identify the potential consumers and their consumption behavior, and validate the response of consumers to the market prices.
[57]	2016	Presented a NILM method for detecting the energy consumption of residential appliances using low frequency uncorrelated spectral information.
[58]	2016	Proposed a lightweight and efficient privacy-preserving mechanism to safeguard the electricity consumers' personal information. This has shown guaranteed differential privacy for appliances.
[59]	2014	A framework was proposed with activity detection and energy advisor modules for detecting human activity to reduce the energy consumption of smart homes.
[60]	2013	This work proposed a model to capture the activity or status of appliances and designed it on top of the high-frequency measurements of voltage and current.
[61]	2012	This work emphasized tracing the average power consumption of an appliance in a day and further controlling (turn on/off) the appliances of households remotely.
[62]	2012	Discussed the importance of self-configuration to identify various appliances. A fingerprint mechanism based on a machine-learning model was implemented to match the appliance by using the energy consumption details.
[63]	2011	This work presented the implementation of wireless sensor nodes with low-power hardware components. These collect the energy consumption of each device at high resolution.

1.4. Article Justification, Contribution & Structure

It is required to have continuous data to do better analytics on smart home energy consumption in various aspects (e.g.: peak demand hours, slack hours, abnormality or fault detection studies, data forecasting for contingency readiness, demand-side management through peak load shaving/leveling/ curtailment, demand response, volt-ampere reactive (VAR)/voltage control, and outage management, energy efficiency enhancement, etc.). However, due to the failures of smart metering devices, communication channels, server stations, etc., missing data at some instants in the overall database has become a common problem.

Based on the preference given to data analytics in smart grids, and review conducted on state-of-the-art research, it is understood the importance of analyzing missing data anomalies. So, this is one such important factor to obtain perfect analytics to help various smart home energy consumption data aspects described above. With this motivation, this paper proposes a simple and effective descriptive analysis of smart home energy consumption readings, which identify the missing data instants and other associated anomalies. To achieve the proposed analysis, a real-time dataset is considered as described in Section 2.

As per the objective of this paper, various contents are deposited as follows: Section 2 presents the description of the real-time dataset considered for the analysis. Section 3 presents the proposed method and its execution procedure. Section 4 discusses the simulation results and identifies the clear problems related to smart home datasets. Finally, the summary of the paper observations and the concluding observations are given in Section 5.

2. TEST CASE DESCRIPTION: TRACEBASE DATASET

The Tracebase dataset consists of a set of files that includes energy consumption readings of the cities Darmstadt, Germany and Sydney, Australia [61]. The energy consumption readings of the city were collected in the year 2012 (which is a combination of consumption data of 2011 and 2012), and the latter is in 2013. This dataset is composed of three folders namely "complete", "incomplete", and "synthetic". The records in the dataset are called traces. The traces are collected as one sample per one second. All the traces of a particular day and a particular appliance are stored in the "comma-separated values (CSV)" file format. This collection of traces begins at midnight and ends at next day midnight. The traces are collected at each second. It follows the 24-hour clock. The trace is stored in a single column in the following format:

Day/Month/Year Hour:Minute:Second; Reading1; Reading 2. (e.g. 01/09/2011 00:46:42; 15; 13)

where Reading-1 and Reading-2 are the energy consumption readings. As mentioned above, the "complete" folder contains the full traces for the entire day, i.e., 24 hours. These traces are collected when the smart plug is not disconnected even the appliance is switched off. The "incomplete" folder contains some missing measurements in the collected data. This missing part is mainly due to wireless transmission/network failures. Here, no interpolation was applied to the recorded data. The "synthetic" folder contains the traces that contain consumption reading as 0. If any appliance is used for a period of time and then disconnected from the metering device, then there will be no readings which lead to "incomplete" traces. To avoid this situation, zeros are padded to the file to represent it as "complete" traces.

2.1. Challenges with the Dataset

The data collected cannot be analyzed as a whole at once. Parts of the dataset can be selected and this reduced size helps in better analysis and visualization of the energy consumption readings [17]. The Tracebase dataset's "complete" folder has 43 appliances. Each appliance, stored as a directory. Each directory contains multiple CSV files. Each of these files named with the combination of a device identification number and date [61]. Each appliance can be analyzed separately or the combination of various appliances can be considered. The main challenge in this considered database is the merging of multiple CSV files of different appliances into a single file. This results in more than one million records after merging, which is very tough to visualize.

3. DETAILED DESCRIPTION AND IMPLEMENTATION OF THE PROPOSED METHOD

The objective of this paper is to analyze a given real-time dataset and identify missing data values at different time instants. Out of the overall dataset given by the values of energy consumption readings of about one year, the complete data in a day (01/09/2011) are considered for the analysis.

The proposed descriptive analysis consists of three phases, viz., analysis, quantification, and correlation, as shown in Fig. 1. The first phase focuses on the analysis of data traces to understand the various levels of energy consumption readings. This helps in understanding the peak hours in a given day. The second phase focuses on the identification and quantification of the dataset to find available time instants (or readings) of the whole day. This analysis is the basis for the identification and quantification of missing time instants. The quantified data that are obtained in the second phase is used to observe the correlation between available and missing data which is performed in the third phase. Finally, a comprehensive report is produced which includes the details of all the possible anomalies with missing data, as an outcome of these three phases. The flow of executing all these three phases is given in Fig. 2, whose detailed description is provided in Table 4.

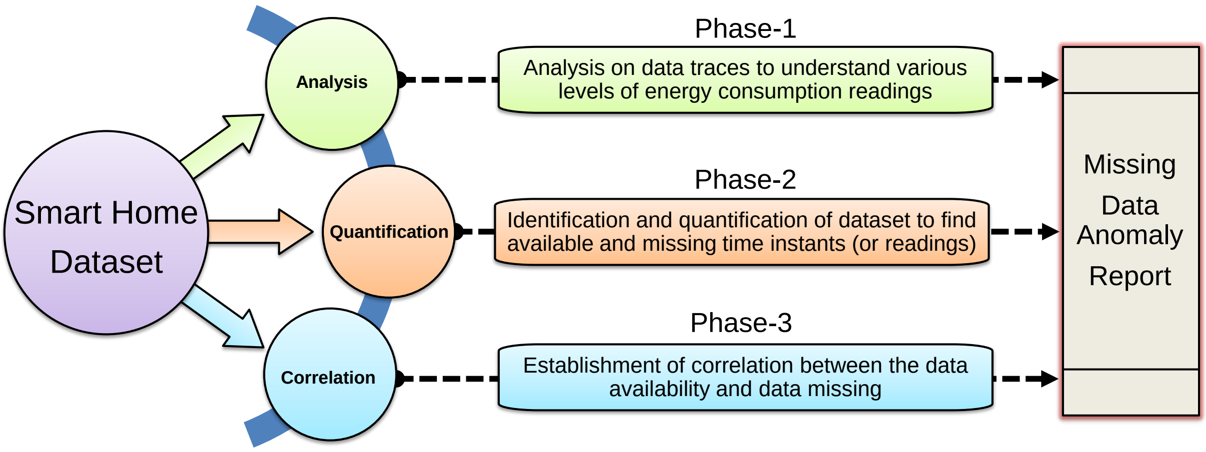


Figure 1. Various phases in the proposed descriptive analysis for finding missing data anomalies

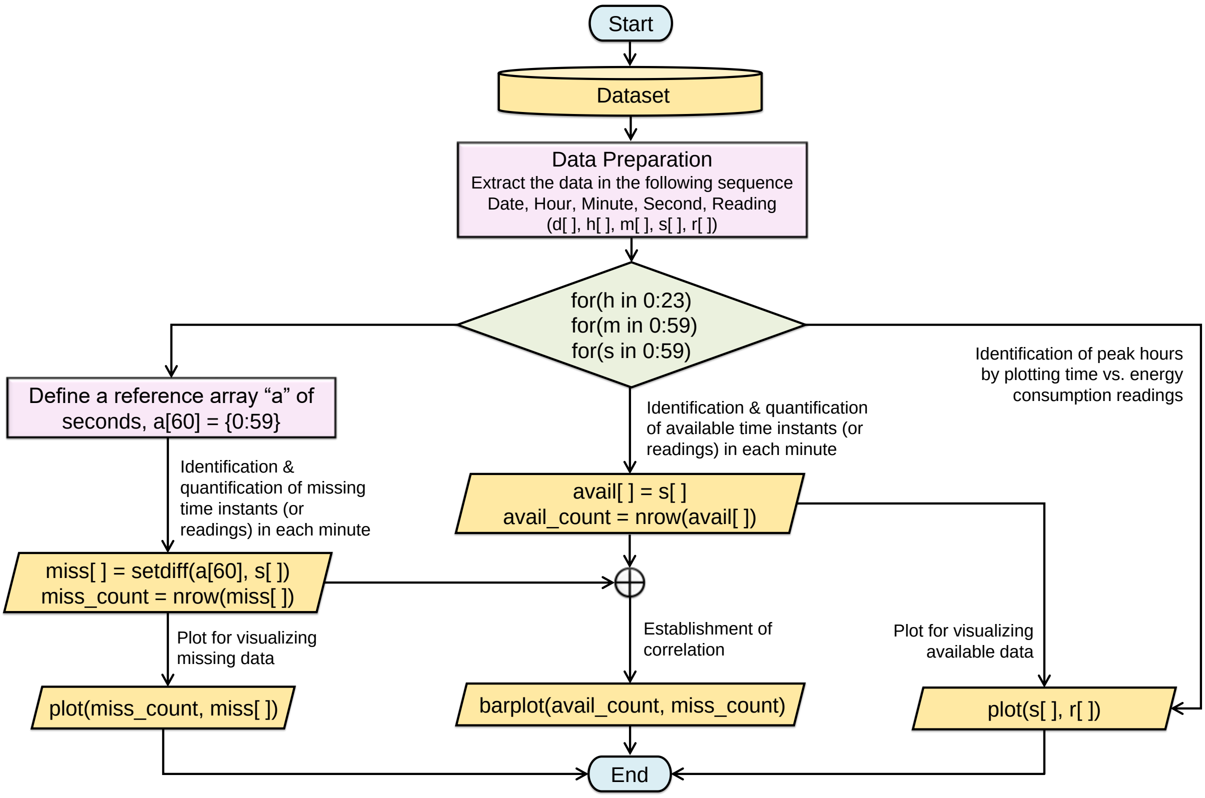


Figure 2. Implementation flow for proposed descriptive analysis

In order to use the proposed descriptive analysis, the input dataset should be prepared with the following considerations.

- Irrespective of the number of appliances connected in a smart home, the dataset should be available as a single merged file.
- This file should have columns of Date, Hour, Minute, Second, Reading in sequence.
- The data type of the values given in each of these columns should adhere.
- For e.g., the considered Tracebase dataset has a single column as described in Section-2 which is further split into multiple columns as desired to apply the proposed descriptive analysis.

Table 4. Various steps involved in the implementation of the proposed descriptive analysis

Phase	Operation	Description
Analysis	Detection of peak hours	The peak hours in a day are identified by plotting time vs. all the energy consumption readings in each hour by using a plot function “plot(s[ ], r[ ])”.
Quantification	Identification and quantification of available time instants (or readings)	Available time instants and corresponding readings are identified by extracting all the available seconds’ information from the considered dataset. This information is assigned as available data using “avail[ ] = s[ ]”. From this available data, the count of available readings is computed using “avail_count = nrow(avail[ ])”. This available data is visualized by plotting each minute using “plot(s[ ], r[ ])”.
	Identification and quantification of missing time instants (or readings)	Missing time instants and their readings are identified as per the below procedure. Initially, a reference array “a” of seconds is defined as “a[60] = {0:59}”. This reference array consists of all the seconds. So, this is used as a basis for the identification of missing seconds in the actual dataset. For this identification, the reference array “a[ ]” of seconds is compared row-wise with the actual array “s[ ]” of seconds of the considered dataset using “setdiff( )” functionality. This function treats each row of a[ ] and each row of s[ ] as single entities and returns the rows from a[ ] that are not in s[ ]. These remaining rows of a[ ] give the details of missing readings, which are stored into miss[ ]. Further, the count of miss[ ] is calculated and stored into miss_count. Finally, the missing data is visualized by plotting each minute using “plot(miss_count, miss[ ])”. The following statements perform the above discussed missing data analysis. miss[ ] = setdiff(a[ ], s[ ]) miss_count = nrow(miss[ ]) plot(miss_count, miss [ ])
Correlation	Correlation between available and missing data	The avail-count and miss_count obtained in the Quantification phase are plotted for understanding the correlation between them. This plot is generated using “barplot(avail_count, miss_count)”.

4. SIMULATION RESULTS AND ANALYSIS

To fulfill the objective and to validate the proposed concept of this paper, the results are organized into three subsections as described follows. To showcase this, some of the hours, viz., 2, 4, 6, 8, 19 and 23 of a day are considered. The simulation and analysis are carried out with the help of “R Programming and RStudio IDE”.

In Section 4.1, the represented visualizations give a basic idea of visualizing the missing instants in the acquired data. In Section 4.2, the represented results give the information of the actual available and missing instants of data and respective counts. In Section 4.3, the plots unveil the association among available and missing instants of data at every minute.

4.1. Results Related to Analysis Phase

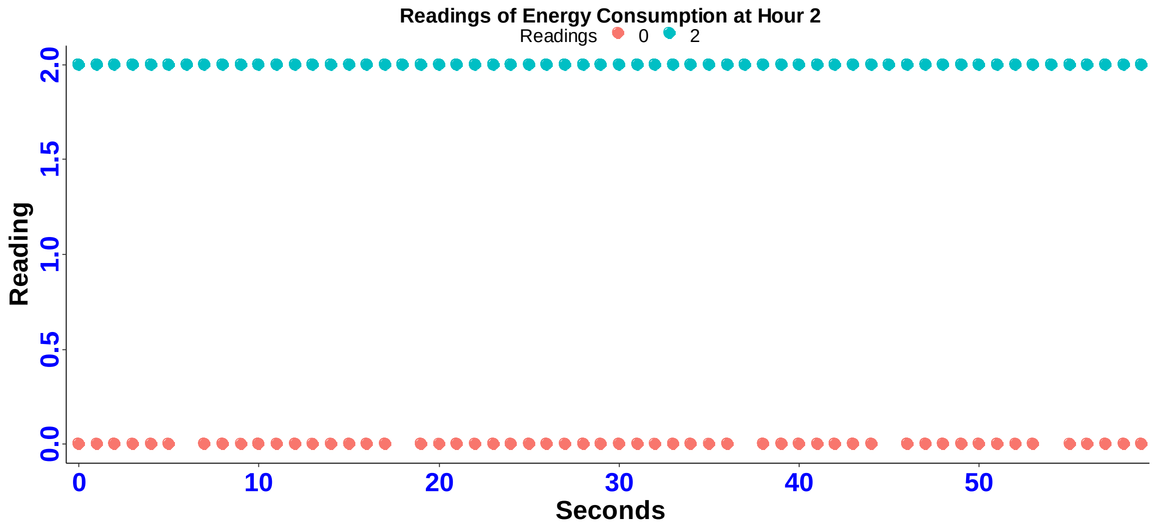
The objective of the analysis phase is to observe the given dataset for the possible identification of the missing data existence. For this, the available data values (energy consumption readings) at each second in different hours of a considered day are plotted as shown in Fig. 3(a) through Fig. 3(f).

From these plots, it can be seen that there are different levels of energy consumption readings available in different seconds in an hour. As shown in Fig. 3(a), Fig. 3(b), Fig. 3(d) and Fig. 3(e), the hours, such as hour-2, hour-4, hour-8, and hour-19 are having only two levels of readings. Whereas, in Fig. 3(c) and Fig. 3(f), the hours, such as hour-6 and hour-23 are having many different levels of readings.

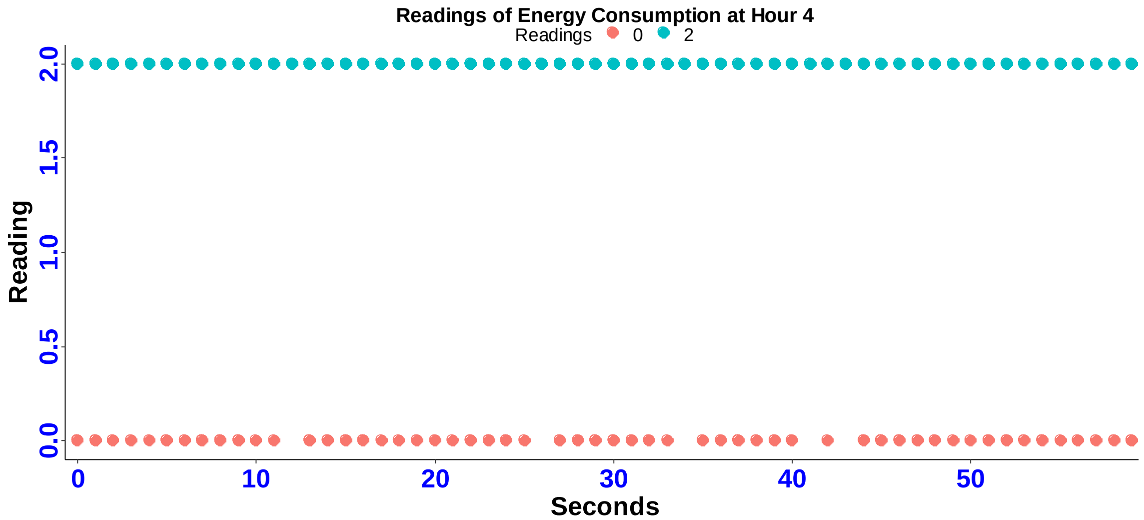
So, these plots help to visualize the existence of different levels of reading and consequently, their maximum and minimum levels in an hour. Hence, this analysis helps in identifying peak hours in a day based on the availability of maximum reading level in an hour.

However, it is important to note that these readings are plotted based on seconds in a given hour. Due to this, all these plots look like having no missing data instants, i.e., the readings (of any level) are available at each and every second throughout the axis, which is not actually true. Due to plotting of the readings with respect to seconds, even though there is missing reading at any one second in any one minute, the other reading may be available at the same second of another minute of the same hour. This means that the missing reading information at a second is covered by available reading at the same second of another minute. So, the existence of missing reading information cannot be seen using these plots.

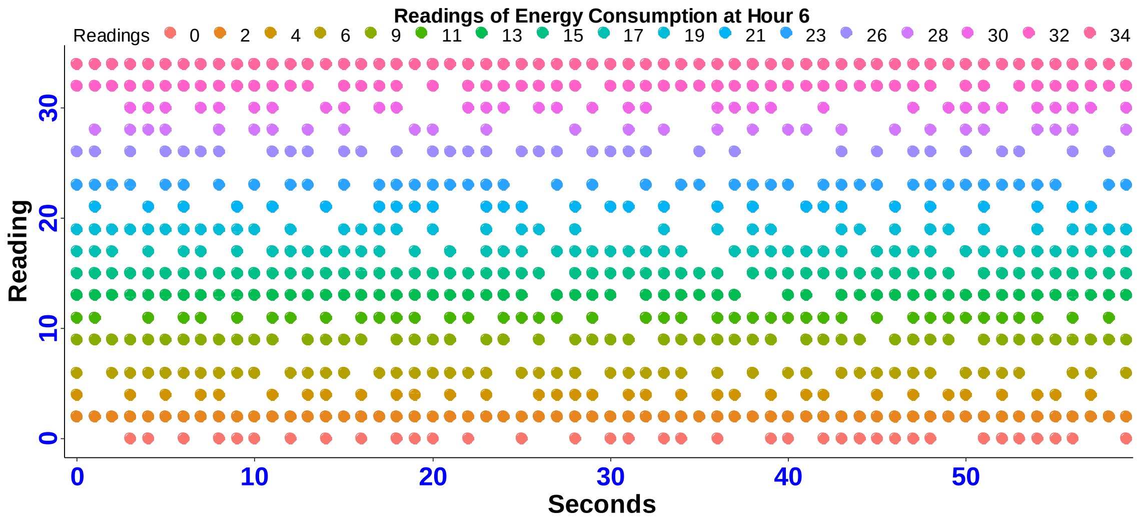
Hence, the analysis phase is moved to the next phase, i.e., the quantification phase as discussed in Section 4.2. Here, the analysis has been done at a minute level in a particular hour instead of the seconds level.



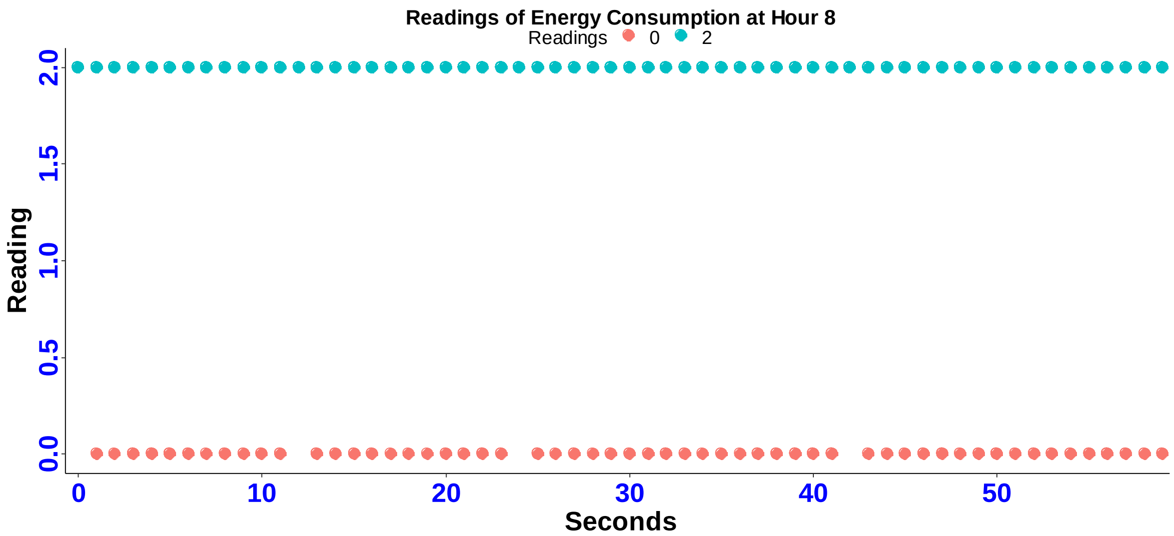
(a) Plot of data traces in hour 2



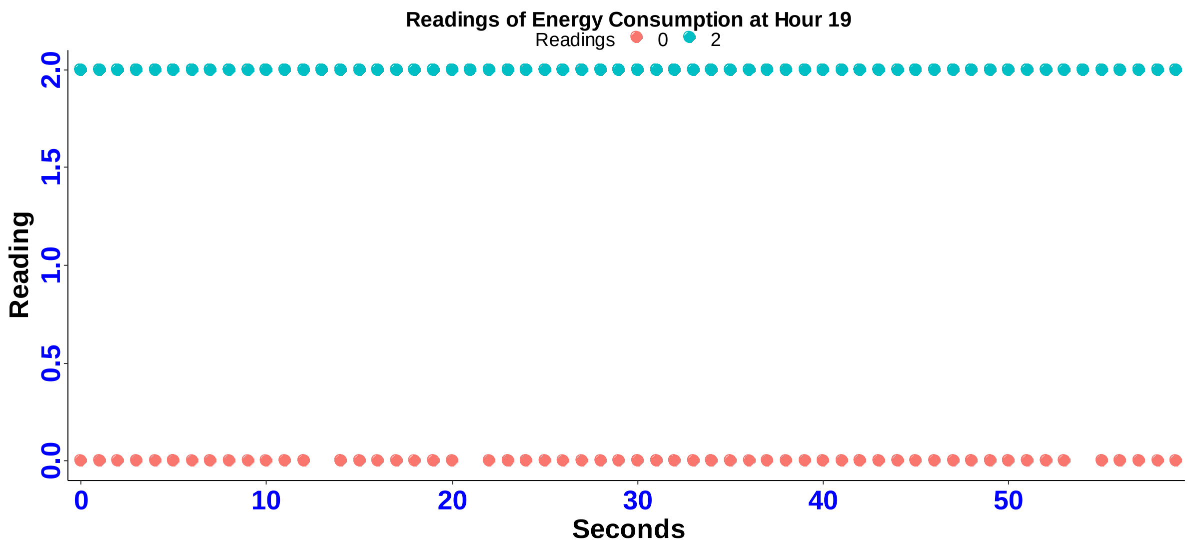
(b) Plot of data traces in hour 4



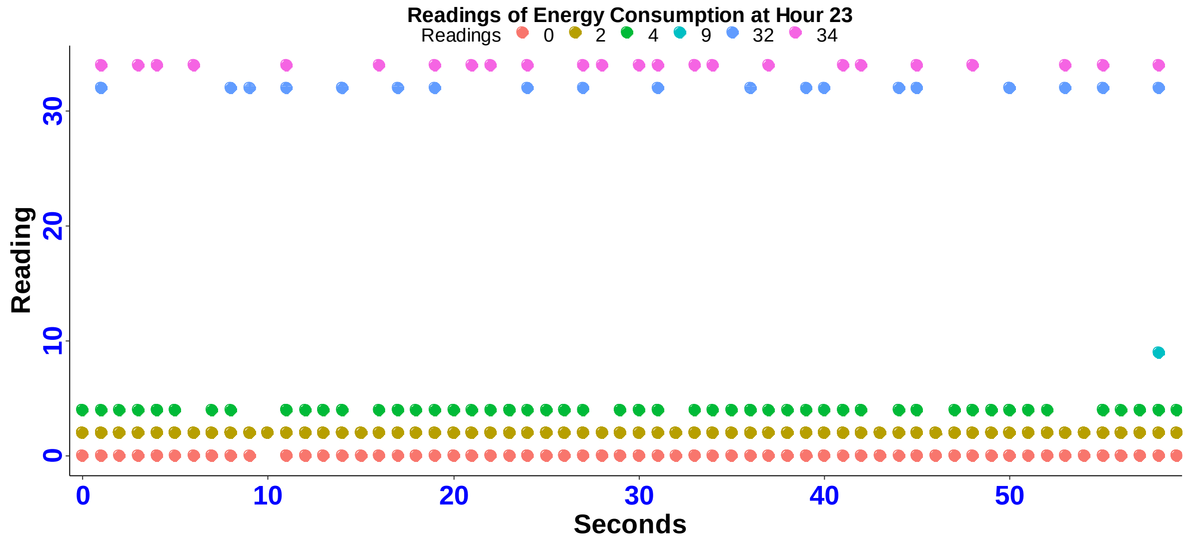
(c) Plot of data traces in hour 6



(d) Plot of data traces in hour 8



(e) Plot of data traces in hour 19

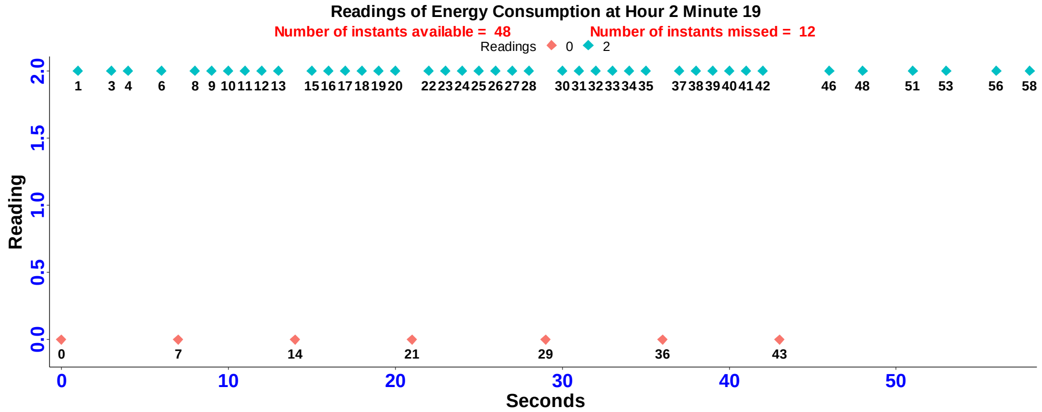


(f) Plot of data traces in hour 23

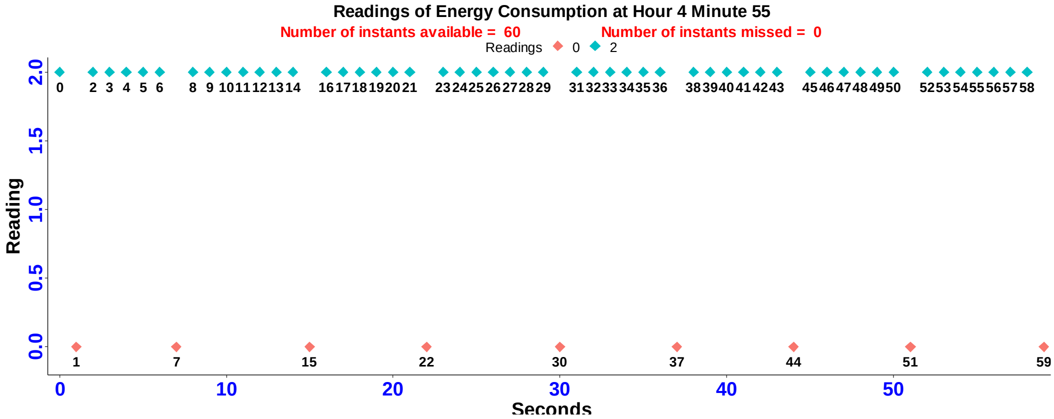
Figure 3. Plots of readings captured at each second of different hours in a day

4.2. Results Related to Quantification Phase: Identification of Counts of Available and Missing Data Instants

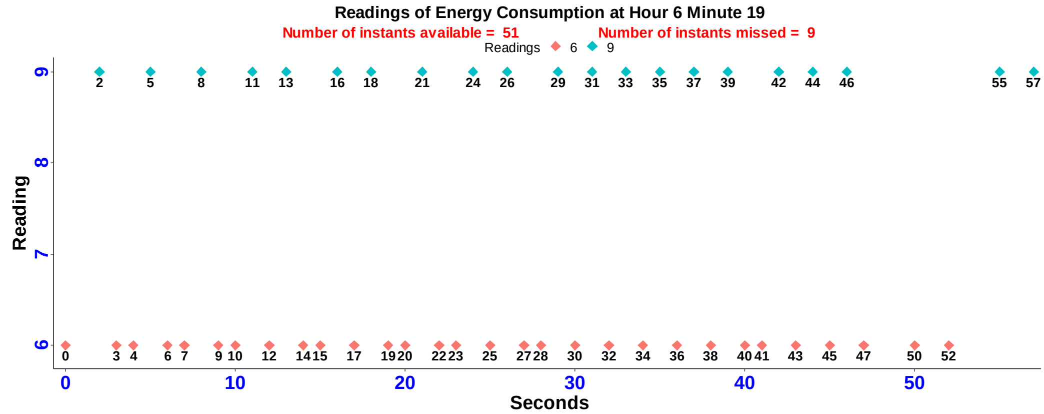
The available instants and their total count at each minute and hour are visualized as shown in Fig. 4(a) through Fig. 4(f). For the sake of explanation, only a few hours and their minutes of the day such as minute 19 at hour 2, minute 55 at hour 4, minute 19 at hour 6, minute 37 at hour 8, minute 50 at hour 19, and minute 55 at hour 23, are shown in these plots. Similarly, the missing instants and their total count at each minute and hour are visualized as shown in Fig. 5(a) through Fig. 5(f). For the sake of explanation, only a few hours and their minutes of the day such as minute 19 at hour 2, minute 55 at hour 4, minute 19 at hour 6, minute 28 at hour 8, minute 37 at hour 8, and minute 55 of hour 23, are shown in these plots. Unlike the previous plots given in Section 4.1, these present plots help to clearly identify available data instants with their total count, and thereby, missing data instants with their total count in a particular minute. Further, a critical anomaly is identified with respect to duplicate traces at some seconds, whose consequences are explained as follows.



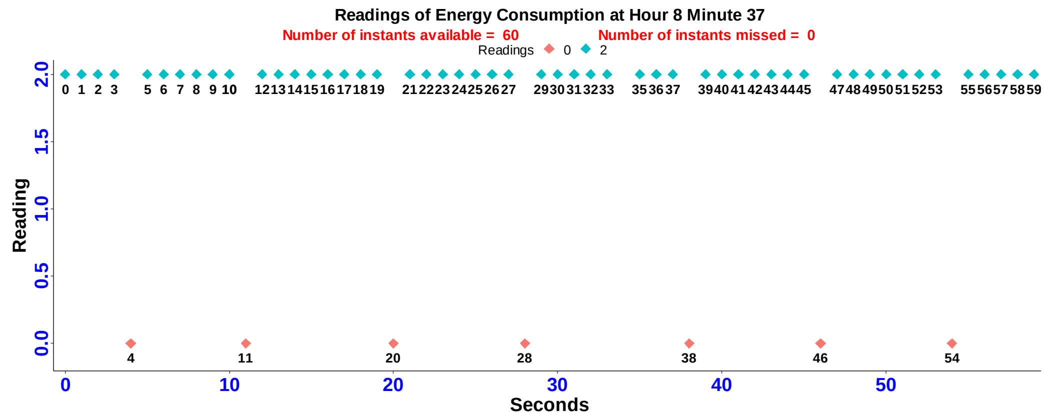
(a) Available instants and their count at Hour 2 Minute 19



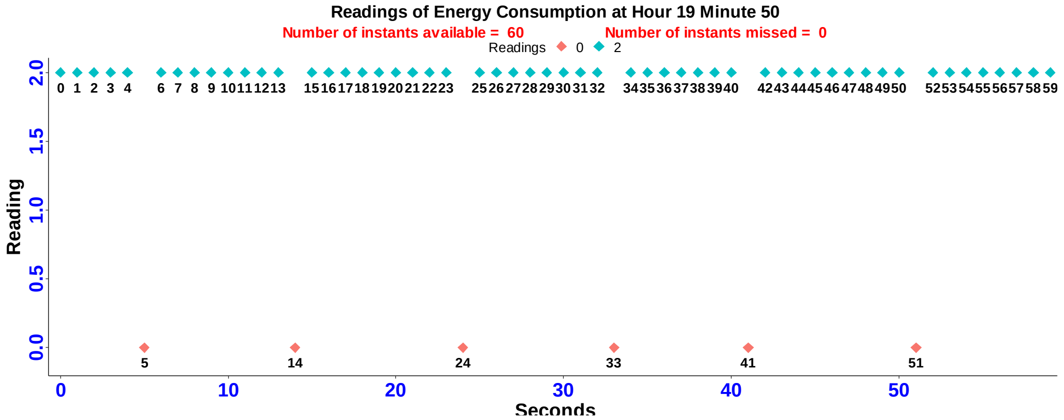
(b) Available instants and their count at Hour 4 Minute 55



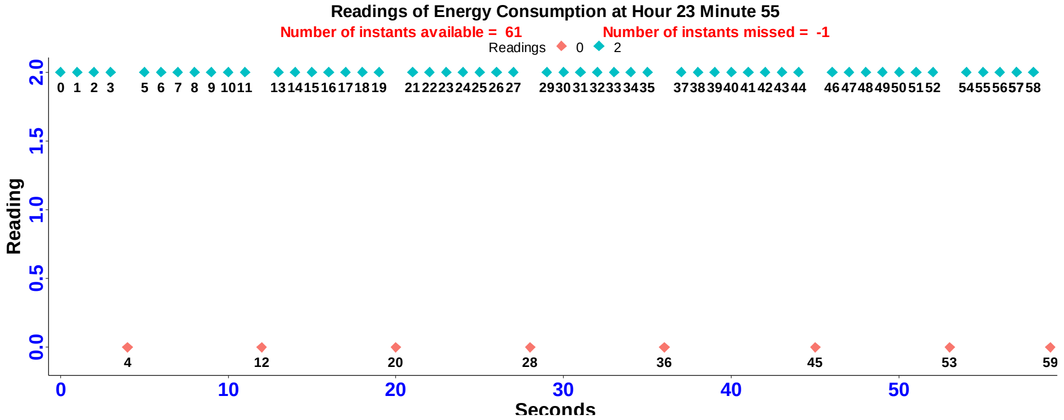
(c) Available instants and their count at Hour 6 Minute 19



(d) Available instants and their count at Hour 8 Minute 37

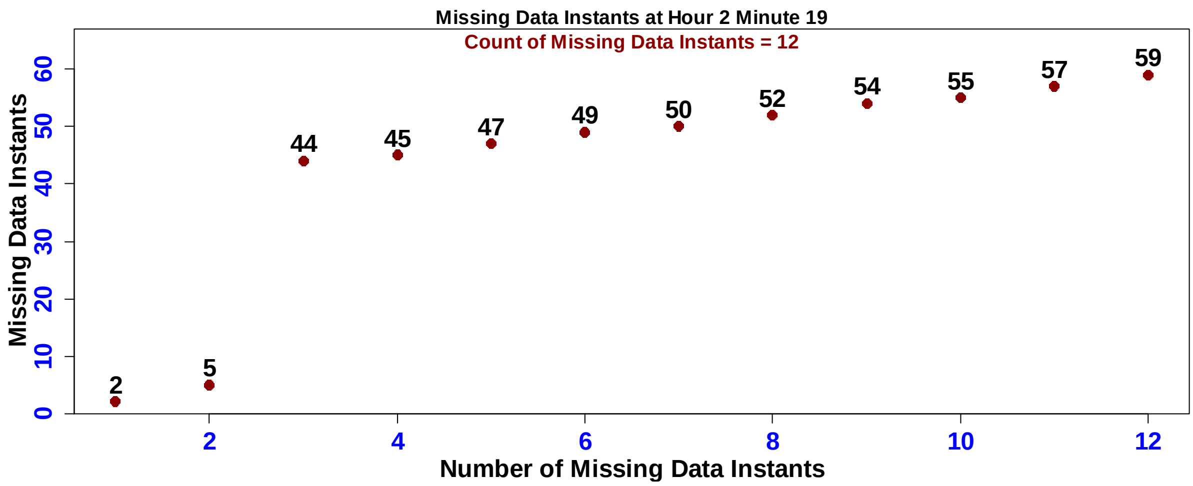


(e) Available instants and their count at Hour 19 Minute 50

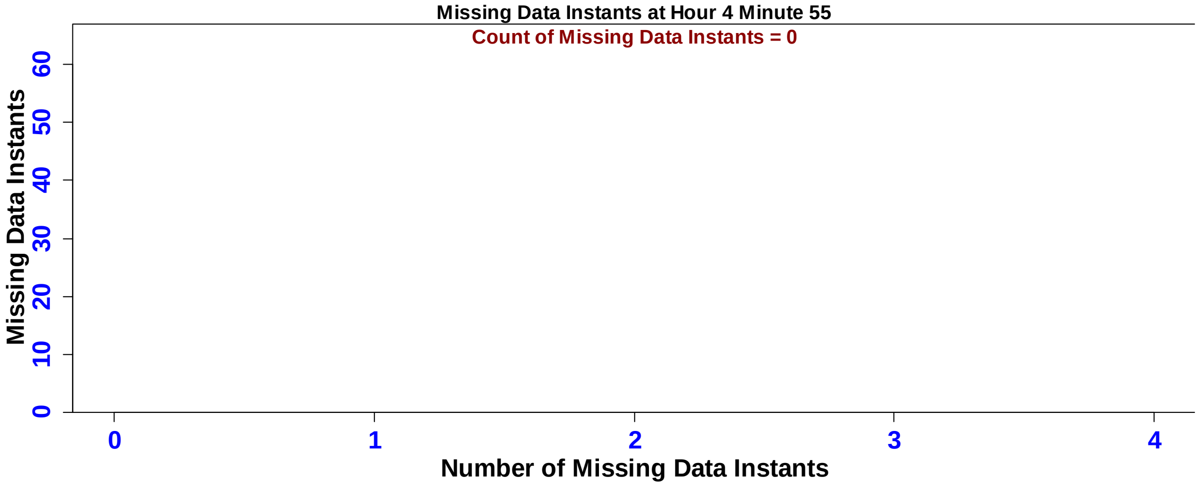


(f) Available instants and their count at Hour 23 Minute 55

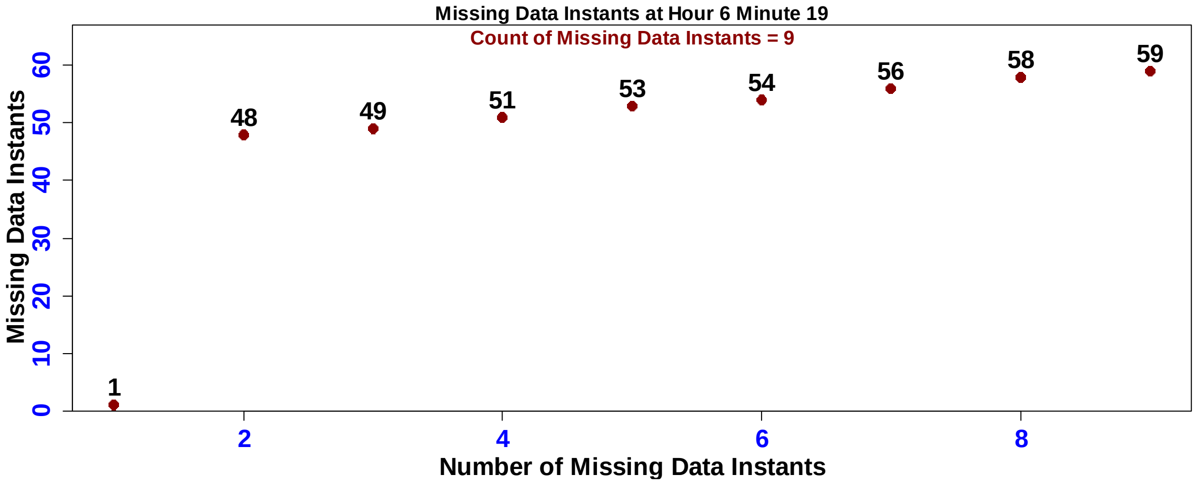
Figure 4. Plots for identification of available instants of data at different minutes and hours



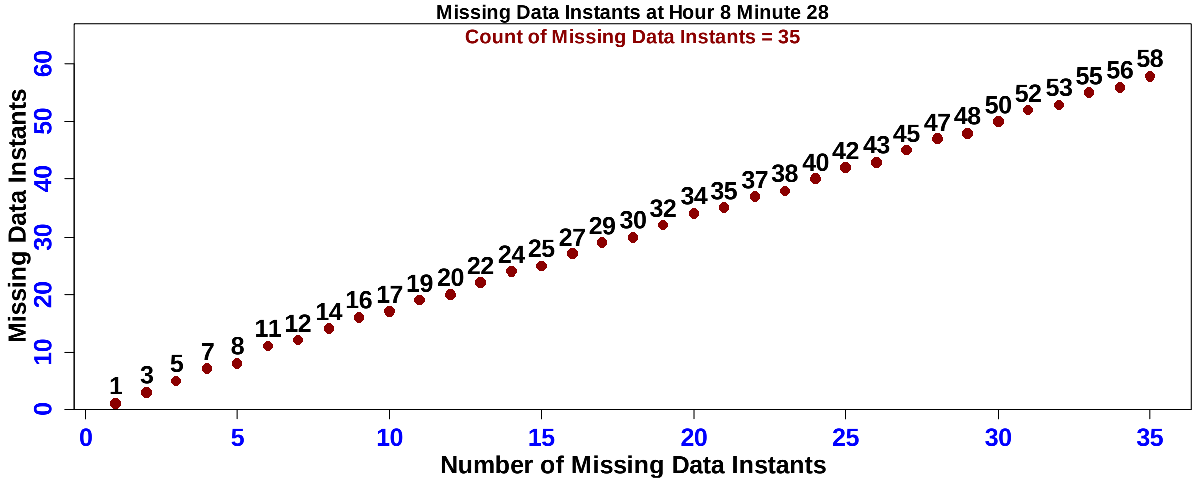
(a) Missing instants and their count at Hour 2 Minute 19



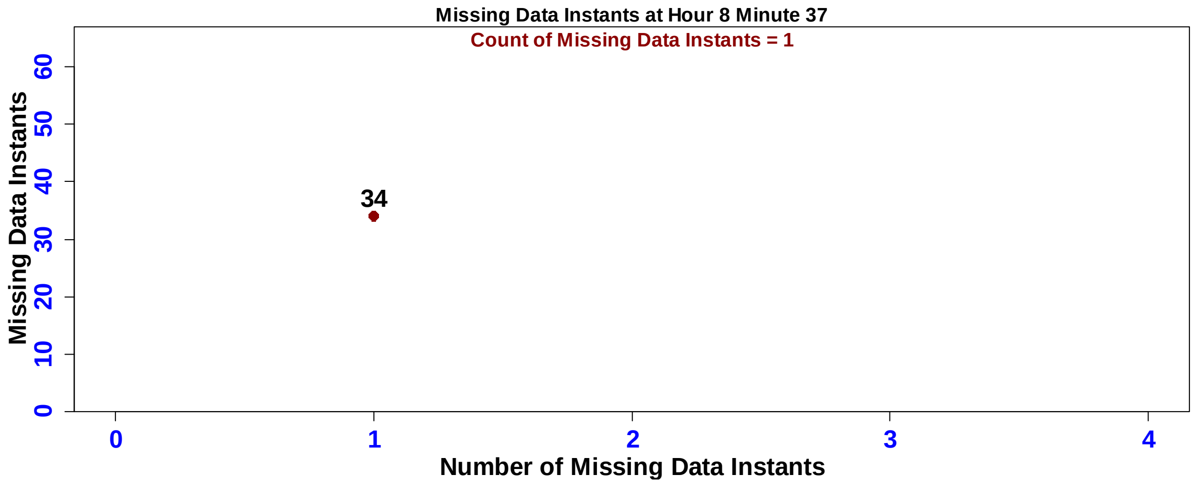
(b) Missing instants and their count at Hour 4 Minute 55



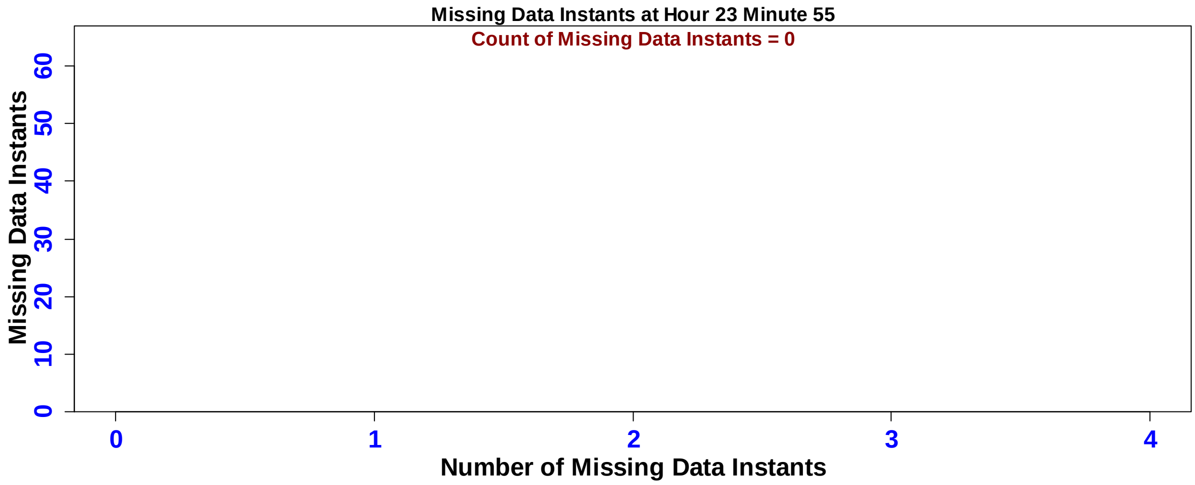
(c) Missing instants and their count at Hour 6 Minute 19



(d) Missing instants and their count at Hour 8 Minute 28



(e) Missing instants and their count at Hour 8 Minute 37



(f) Missing instants and their count at Hour 23 Minute 55

Figure 5. Plots for identification of missing instants of data at different minutes and hours

As shown in Fig. 6, the occurrence of duplicate traces can be seen at second 10 in minute 37 of hour 8 and at second 57 in minute 55 of hour 23. Similarly, the occurrence of the missing trace can be seen at second 34 in minute 37 of hour 8.

From Fig. 4(d), it can be observed that all the reading instants are available in minute 37 of hour 8. So, the plot is showing the total available instants as “60” and the total missing instants are “0”. But, based on Fig. 6, there is a repetition of a trace at second 10 and a missing trace at second 34 of minute 37 in hour 8. Due to this duplicate trace, it is shown that the number of total instants is 60 instead of 59 in Fig. 4(d). The actual count of missing is “1” and is occurring at second 34, which is elucidated in Fig. 5(e).

In general, as there will be a maximum of “60" seconds in a minute, the expected reading count is “60”. So, the best possible case is expected as the total count of available instants is “60” and missing instants is “0”. However, “61” total available instants and “-1” total missing instants are observed in minute 55 of hour 23 as shown in Fig. 4(f). This discrepancy revealed that there is a duplicate trace at second 57 in minute 55 of hour 23, which is clearly shown in Fig. 6.

Further, from Fig. 5(b) and Fig. 5(f), it can be seen that there are “no” missing instants in minute 55 at hour 4 and minute 55 at hour 23 respectively. Similarly, from Fig. 5(d), it is observed that there are “more” missing instants in minute 28 at hour 8 than the other considered minutes and hours in the same day.

Besides, another critical anomaly is identified. The count of missing instants is identified as “0” in Fig. 4(d), whereas, the same is identified as “1” in Fig. 5(e) corresponding to minute 37 at hour 8. Similarly, the count of missing instants in minute 55 at hour 23 is identified as “-1” in Fig. 4(f), whereas, the same is identified as “0” shown in Fig. 5(f).

All the abovementioned anomalies that are identified from phase-2 of the proposed descriptive analysis are very critical. This creates ambiguity in the calculation of the total count of missing and available instants of data. This ambiguity affects various smart grid functionalities that are explained in the above sections if not addressed properly.

```
> H8 <- filter(dfe3, REC_DATE=='01/09/2011', REC_HOUR==8, REC_MINUTE==37)
> H8[7:12,]
# Showing the occurrence of duplicate trace at Hour 8 Minute 37 Second 10
```

	REC_DATE	REC_HOUR	REC_MINUTE	REC_SECOND	READING_1	READING_2
7	01/09/2011	8	37	6	2	0
8	01/09/2011	8	37	7	2	0
9	01/09/2011	8	37	8	2	0
10	01/09/2011	8	37	9	2	0
11	01/09/2011	8	37	10	2	0
12	01/09/2011	8	37	10	2	0

```
> H8 <- filter(dfe3, REC_DATE=='01/09/2011', REC_HOUR==8, REC_MINUTE==37)
> H8[33:38,]
# Showing the missing of trace at Hour 8 Minute 37 Second 34
```

	REC_DATE	REC_HOUR	REC_MINUTE	REC_SECOND	READING_1	READING_2
33	01/09/2011	8	37	31	2	0
34	01/09/2011	8	37	32	2	0
35	01/09/2011	8	37	33	2	0
36	01/09/2011	8	37	35	2	0
37	01/09/2011	8	37	36	2	2
38	01/09/2011	8	37	37	2	2

```
> H23 <- filter(dfe3, REC_DATE=='01/09/2011', REC_HOUR==23, REC_MINUTE==55)
> H23[56:61,]
# Showing the occurrence of duplicate trace at Hour 23 Minute 55 Second 57
```

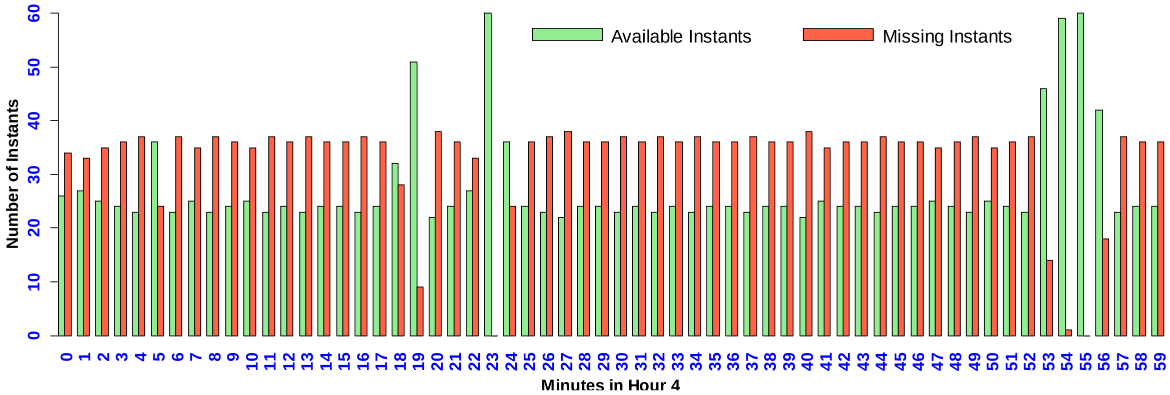
	REC_DATE	REC_HOUR	REC_MINUTE	REC_SECOND	READING_1	READING_2
56	01/09/2011	23	55	55	2	0
57	01/09/2011	23	55	56	2	0
58	01/09/2011	23	55	57	2	0
59	01/09/2011	23	55	57	2	0
60	01/09/2011	23	55	58	2	0
61	01/09/2011	23	55	59	0	0

Figure 6. Occurrence of missing traces and duplicate traces at particular instants of time

4.3. Results Related to Correlation Phase: Comparison of Available and Missing Data Counts

The correlation between the available data traces and missing data traces at each minute is visualized as shown in Fig. 7(a) through Fig. 7(f) to understand the proportionality between them. There are full traces available at some instants, for example, minute 23 and minute 55 of hour 4, minute 37 of hour 8, minute 50 of hour 19, and minute 52 of hour 23 as referred in Fig. 7(b), Fig. 7(d), Fig. 7(e) and Fig. 7(f) respectively. Further, it is observed that there is one negative value at minute 55 of hour 23 as shown in Fig. 7(f). This is due to the occurrence of 61 instants at minute 55 of hour 23 as depicted in Fig. 4(f) and Fig. 6. From these plots, it is observed that an average of 30 missing data instants is identified in each minute.

(a) Count of available data instants vs. count of missing data instants at Hour 2



(b) Count of available data instants vs. count of missing data instants at Hour 4

(c) Count of available data instants vs. count of missing data instants at Hour 6

(d) Count of available data instants vs. count of missing data instants at Hour 8

(e) Count of available data instants vs. count of missing data instants at Hour 19

(f) Count of available data instants vs. count of missing data instants at Hour 23

Figure 7. Correlation of total available data instants and missing data instants in different hours of a day

5. CONCLUSION

This paper proposes a simple and effective descriptive analysis of the smart home energy consumption dataset to find various missing data-related anomalies. All such possible anomalies are visualized using simulation results. The presented visualizations help the user to get a quick grasping of available and missing instants of data which creates a discontinuity in the dataset.

This discontinuity may create the problem in achieving various smart grid functionalities, viz., monitoring, operation and control, load shedding, peak load shaving, etc., in real-time. The key observations of this proposed analysis conducted on real-time smart home energy consumption data are given as follows.

From Fig. 3, peak hours in a day based on the availability of maximum reading level in an hour is found. For the considered day (01/09/2011) in this paper, the 6th hour is identified as a peak hour with a consumption of 34 kilowatt-hour (34 kWh or 34 units).

From Fig. 4, the available instants of data are found. This helps in the identification of missing data instants as depicted in Fig. 5. From these results, it is observed that the number of missing instants varies in different minutes and hours. Moreover, it is seen that this number is varying disproportionately, which can misguide the smart grid operations and control.

From Fig. 6, it is clearly seen that there is an occurrence of duplicate traces while capturing the data from the meters. This duplication may have an adverse effect on the integrity of data.

From Fig. 7, i.e., the correlation between available and missing data instants, it is found that an average of 50% data instants is missed in an hour.

Hence, the presented descriptive analysis in this paper is useful to identify missing data anomalies that occur while capturing the energy consumption data. This helps the smart home operations to have some predetermined assumptions about these issues, thereby can take better-informed decisions.

REFERENCES

[1] Moharm, K. State of the art in big data applications in microgrid: A review. *Advanced Engineering Informatics* 2019; *42*: 100945, DOI: https://doi.org/10.1016/j.aei.2019.100945

[2] Wang, Y, Qixin, C. Review of smart meter data analytics: Applications, methodologies, and challenges. *IEEE Transactions on Smart Grid* 2019; *10*: 3125-3148, DOI: https://doi.org/10.1109/TSG.2018.2818167

[3] Bhattarai, PB, Paudyal, S, Luo, Y, Manish, M, Cheung, K, Reinaldo, T, Rob, H, Kurt, SM, Rui, Z, Power, Z, Milos, M, Song, Z, Xiaping, Z. Big data analytics in smart grids: State-of-the-art, challenges, opportunities, and future directions. *IET Smart Grid* 2019; *2*: 141-154, DOI: https://doi.org/10.1049/iet-stg.2018.0261

[4] Ghorbanian, M, Dolatabadi, SH, Pierluigi, S. Big data issues in smart grids: A survey. *IEEE Systems Journal* 2019; *13*: 4158-4168, DOI: https://doi.org/10.1109/JSYST.2019.2931879

[5] Zhang, Y, Tao, H, Bompard, EF. Big data analytics in smart grids: A review. *Energy Informatics* 2019; *8*, DOI: https://doi.org/10.1186/s42162-018-0007-5

[6] Wen, L, Zhou K, Shanlin, Y, Li, L. Compression of smart meter big data: A survey. *Renewable and Sustainable Energy Reviews* 2018; *91*: 59-69, DOI: https://doi.org/10.1016/j.rser.2018.03.088

[7] Tu, C, Xi, H, Shuai, Z, Fie, J. Big data issues in smart grid – A review. *Renewable and Sustainable Energy Reviews* 2017; *79*: 1099-1107, DOI: https://doi.org/10.1016/j.rser.2017.05.134

[8] Jiang, H, Kun, W, Wang, Y, Gao, M, Zhang, Y. Energy big data: A survey. *IEEE Access* 2016; *4*: 3844-3861, DOI: https://doi.org/10.1109/ACCESS.2016.2580581

[9] Kaile, Z, Fu, C, Yang, S. Big data driven smart energy management: From big data to big insights. *Renewable and Sustainable Energy Reviews* 2016; *56*: 215-225, DOI: https://doi.org/10.1016/j.rser.2015.11.050

[10] Vasilica, SO, Bara, A, George, BT, Maria, IC, Botezatu, AM. Insights into demand-side management with big data analytics in electricity consumers' behaviour. *Computers and Electrical Engineering* 2021; 8*9*: 106902, DOI: https://doi.org/10.1016/j.compeleceng.2020.106902

[11] Syed, D, Zainab, A, Refaat, SS, Haitham, AR, Bouhali, O. Smart grid big data analytics: Survey of technologies, techniques, and applications. *IEEE Access* 2020; *9*: 59564-59585, DOI: https://doi.org/10.1109/ACCESS.2020.3041178

[12]Kumar, SV, Tianyi, W, Aggour, SK, Wang, P, Hart, JP, Yan, W. Big data analysis of massive PMU datasets: A data platform perspective. In: 2021 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT); 16-18 February 2021: IEEE, Washington, DC, USA: pp. 1-5.
[13]Babar, M, Khattak, SA, Jan, MA, Tariq, UM. Energy aware smart city management system using data analytics and Internet of Things. *Sustainable Energy Technologies and Assessments* 2021; *44*: 100992, DOI: https://doi.org/10.1016/j.seta.2021.100992
[14]Kumar, DP, Saptarshi, D. Smart grid architecture model for control, optimization and data analytics of future power networks with more renewable energy. *Journal of Cleaner Production* 2021; *301*: 126877, DOI: https://doi.org/10.1016/j.jclepro.2021.126877
[15]Mohajeri, M, Ghassemi, A, Gulliver, TA. Fast big data analytics for smart meter data. *IEEE Open Journal of the Communications Society* 2020; *1*: 1864-1871, DOI: https://doi.org/10.1109/OJCOMS.2020.3038590
[16]Rossi, B, Chren, S. Smart grids data analysis: A systematic mapping study. *IEEE Transactions on Industrial Informatics* 2020; *16*: 3619-3639, DOI: https://doi.org/10.1109/TII.2019.2954098
[17]Ushakova, A, Mikhaylov, JS. Big data to the rescue? Challenges in analysing granular household electricity consumption in the United Kingdom. *Energy Research & Social Science* 2020; *64*: 101428, DOI: https://doi.org/10.1016/j.erss.2020.101428
[18] Shang, C, Fengqi, Y. Data analytics and machine learning for smart process manufacturing: Recent advances and perspectives in the big data era. *Engineering* 2019; *5*: 1010-1016, DOI: https://doi.org/10.1016/j.eng.2019.01.019
[19] Tom, W, Jin, N, Falch, P, Joshua, T. A big data platform for smart meter data analytics. *Computers in Industry* 2019; *105*: 250-259, DOI: https://doi.org/10.1016/j.compind.2018.12.010
[20] Cordon, I, Luengo, J, Garcia, S, Herrera, F, Francisco, C. Smartdata: Data preprocessing to achieve smart data in R. *Neurocomputing* 2019; *360*: 1-13, DOI: https://doi.org/10.1016/j.neucom.2019.06.006
[21] Scott, ME. The role of statistics in the era of big data: Crucial, critical and under-valued. *Statistics and Probability Letters* 2018; *136*: 20-24, DOI: https://doi.org/10.1016/j.spl.2018.02.050
[22] Tao, H. Big data analytics making the smart grid smarter. *IEEE Power & Energy Magazine* 2018; *16*: 12-16, DOI: https://doi.org/10.1109/MPE.2018.2801440
[23] Kang, C, Yi, W, Xue, Y, Mu, G, Liao, R. Big data analytics in China’s electric power industry: Modern information, communication technologies, and millions of smart meters. *IEEE Power & Energy Magazine* 2018; *16*: 54-65, DOI: https://doi.org/10.1109/MPE.2018.2790819
[24] Eckroth, J. A course on big data analytics. *Journal of Parallel Distrib. Comput.* 2018; 118: 166-176, DOI: https://doi.org/10.1016/j.jpdc.2018.02.019
[25] Kunjin, C, He, Z, Shan, XW, Jun, H, Li, L, Jinliang, H. Learning-based data analytics: Moving towards transparent power grids. *CSEE Journal of Power and Energy Systems* 2018; *4*: 67-82, DOI: https://doi.org/10.17775/CSEEJPES.2017.01070
[26] Munshi, AA, Yasser, ARIM. Big data framework for analytics in smart grids. *Electric Power Systems Research* 2017; *151*: 369-380, DOI: https://doi.org/10.1016/j.epsr.2017.06.006
[27] Tarek, A, Dev, S, Visser, L, Murhaf, H, Van, SW. A systematic analysis of meteorological variables for PV output power estimation. *Renewable Energy* 2020; *153*: 12-22, DOI: https://doi.org/10.1016/j.renene.2020.01.150
[28] Somu, N, Raman, GMR, Krithi, R. A hybrid model for building energy consumption forecasting using long short term memory networks. *Applied Energy* 2020; 261: 114131, DOI: https://doi.org/10.1016/j.apenergy.2019.114131
[29] Abbas, A, Chowdhury, B. A data-driven approach for providing frequency regulation with aggregated residential HVAC units. In: 2019 North American Power Symposium (NAPS); 13-15 October 2019: IEEE, Wichita, KS, USA: pp. 1-6.
[30] Dai, B, Ran, W, Kun, Z, Hao, J, Ping, W. A demand response scheme in smart grid with clustering of residential customers. In: 2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm); 21-23 October 2019: IEEE, Beijing, China: pp. 1-6.
[31] Huishi, L, Ma, J, Sun, R, Du, Y. A data-driven approach for targeting residential customers for energy efficiency programs. *IEEE Transactions on Smart Grid* 2019; *11*: 1229-1238, DOI: https://doi.org/10.1109/TSG.2019.2933704
[32] Chi, S, Sanmukh, RK, Xiong, C, Kannan, R, Prasanna, VK. A cooperative multi-agent deep reinforcement learning framework for real-time residential load scheduling. In: International Conference on Internet of Things Design and Implementation (IoTDI’19); 15-18 April 2019: Association for Computing Machinery (ACM), Montreal, QC, Canada: pp. 59-69.
[33] Mammoli, A, Matthew, R, Victor, A, Manel, MR, Chien-fei, C, Joana, MA. A behavior-centered framework for real-time control and load-shedding using aggregated residential energy resources in

distribution microgrids. *Energy & Buildings* 2019; *198*: 275-290, DOI: https://doi.org/10.1016/j.enbuild.2019.06.021
[34] Awad, A, Peter, B, German, R. A house appliances-level co-simulation framework for smart grid applications. In: Puliafito A, Trivedi K (Eds.), Systems Modeling: Methodologies and Tools, Cham: Springer, 2018, pp. 303-317.
[35] Cao, HA, Tri, KW, Karl, A, Nuno, N. A collaborative framework for annotating energy datasets. In: 2015 IEEE International Conference on Big Data (Big Data); 29 October-1 November 2015: IEEE Computer Society, Santa Clara, CA, USA: pp. 2716-2725.
[36] Moreno, AFJ, David, LM, Morrow, DJ, Jesús, MdR, Aoife, MF. Load modelling and non-intrusive load monitoring to integrate distributed energy resources in low and medium voltage networks. *Renewable Energy* 2021; *179*: 445-466, DOI: https://doi.org/10.1016/j.renene.2021.07.056
[37] Klemenjak, C, Stephen, M, Wilfried, E. Investigating the performance gap between testing on real and denoised aggregates in non-intrusive load monitoring. *Energy Informatics* 2021; *4*: 1-15, DOI: https://doi.org/10.1186/s42162-021-00137-9
[38] Huan, C, Yue-Hsien, W, Chun-Hung, F. A convolutional autoencoder-based approach with batch normalization for energy disaggregation. *The Journal of Supercomputing* 2021; *77*: 2961-2978, DOI: https://doi.org/10.1007/s11227-020-03375-y
[39] Khurram, HI, Farhan, HM, Muhammad, A, Muhammad, AQ, Nawaz, AM, Chishti, AR. A critical review of state-of-the-art non-intrusive load monitoring datasets. *Electric Power Systems Research* 2021; *192*: 106921, DOI: https://doi.org/10.1016/j.epsr.2020.106921
[40] Ifeoluwa, OO, Obaid, M, Zhang, J, Dyanand, SR. Solving the fair electric load shedding problem in developing countries. *Autonomous Agents and Multi-Agent Systems* 2020; *34*: 1-35, DOI: https://doi.org/10.1007/s10458-019-09428-8
[41] Artem, S, Malof, MJ, Bohao, H, Bradbury, K. Estimating residential building energy consumption using overhead imagery. *Applied Energy* 2020; *280*: 116018, DOI: https://doi.org/10.1016/j.apenergy.2020.116018
[42] Gopinath, R, Mukesh, K, Chandra, CPJ, Kota, S. Energy management using non-intrusive load monitoring techniques – State-of-the-art and future research directions. *Sustainable Cities and Society* 2020; *62*: 102411, DOI: https://doi.org/10.1016/j.scs.2020.102411
[43] Radu-Casian, M, Hurtig, D, Charlie, O. End-to-end anytime solution for appliance recognition based on high-resolution current sensing with few-shot learning. *Internet of Things* 2020; *11*: 100263, DOI: https://doi.org/10.1016/j.iot.2020.100263
[44] Langevin, A, Ghyslain, G, Cheriet, M. Crosstalk suppression in semi-intrusive load monitoring systems using Hall Effect sensors. *IEEE Transactions on Smart Grid* 2020; *11*: 5019-5027, DOI: https://doi.org/10.1109/TSG.2020.3002668
[45] Himeur, Y, Abdullah, A, Faycal, B, Abbes, A. Building power consumption datasets: Survey, taxonomy and future directions. *Energy & Buildings* 2020; *227*: 110404, DOI: https://doi.org/10.1016/j.enbuild.2020.110404
[46] Chinthaka, D, Stephen, M, Bajić, IV. Residential power forecasting using load identification and graph spectral clustering. *IEEE Transactions on Circuits and Systems II: Express Briefs* 2019; *66*: 1900-1904, DOI: https://doi.org/10.1109/TCSII.2019.2891704
[47] Nalmpantis, C, Dimitris, V. Machine learning approaches for non-intrusive load monitoring: from qualitative to quantitative comparation. *Artif Intell Rev* 2019; *52*: 217-243, DOI: https://doi.org/10.1007/s10462-018-9613-7
[48] Jana, H, Andreas, R. A study on the impact of data sampling rates on load signature event detection. *Energy Inform* 2019; *2*: 1-12, DOI: https://doi.org/10.1186/s42162-019-0096-9
[49] Lorena, MR, Adriana, RCG. Competitive autoassociative neural networks for electrical appliance identification for non-Intrusive load monitoring. *IEEE Access* 2019; *7*: 111746-111755, DOI: https://doi.org/10.1109/ACCESS.2019.2934019
[50] Rashid, H, Pushpendra, S, Stankovic, V, Lina, S. Can non-intrusive load monitoring be used for identifying an appliance's anomalous behaviour?. *Applied Energy* 2019; *238*: 796-805, DOI: https://doi.org/10.1016/j.apenergy.2019.01.061
[51] Christoph, K, Andreas, R, Pereira, L, Stephen, M, Mario, B, Elmenreich, W. Electricity consumption data sets: Pitfalls and opportunities. In: Proceedings of the 6th ACM International Conference on Systems of Energy-Efficient Buildings, Cities, and Transportation (BuildSys'19); 13-14 November 2019: Association for Computing Machinery (ACM), New York, NY, USA: pp. 159-162.
[52] Henriet, S, Umut, Ş, Fuentes, B, Gaël, R. A generative model for non-Intrusive load monitoring in commercial buildings. *Energy & Buildings* 2018; *177*: 268-278, DOI: https://doi.org/10.1016/j.enbuild.2018.07.060

[53] Dinesh, C, Shirantha, W, Liyanage, Y, Mervyn, PBE, Roshan, IG, Ekanayake, J. Non-intrusive load monitoring under residential solar power influx. *Applied Energy* 2017; *205*: 1068-1080, DOI: http://dx.doi.org/10.1016/j.apenergy.2017.08.094

[54] Saeed, SH, Kodjo, A, Kelouwani, S, Alben, C. Non-intrusive load monitoring through home energy management systems: A comprehensive review. *Renewable and Sustainable Energy Reviews* 2017; *79*: 1266-1274, DOI: http://dx.doi.org/10.1016/j.rser.2017.05.096

[55] Bonfigli, R, Emanuele, P, Marco, F, Severini, M, Stefano, S, Francesco, P. Non-intrusive load monitoring by using active and reactive power in additive Factorial Hidden Markov Models. *Applied Energy* 2017; *208*: 1590-1607, DOI: http://dx.doi.org/10.1016/j.apenergy.2017.08.203

[56] Antonio, G, Molina, R, Alejandro, M, Sergio, V, Carlos, A. Residential end-uses disaggregation and demand response evaluation using integral transforms. *J. Mod. Power Syst. Clean Energy* 2017; *5*: 91-104, DOI: 10.1007/s40565-016-0258-8

[57] Chinthaka, D, Buddhika, WN, Indika, GR, Mervyn, PBE, Janaka, E, Janaka, VW. Residential appliance identification based on spectral information of low frequency smart meter measurements. *IEEE Transactions on Smart Grid* 2016; *7*: 2781-2792, DOI: https://doi.org/10.1109/TSG.2015.2484258

[58] Pedro, B, Brito, A, Hyggo, A. A Technique to provide differential privacy for appliance usage in smart metering. *Information Sciences* 2016; *370-371*: 355-367, DOI: http://dx.doi.org/10.1016/j.ins.2016.08.011

[59] Alaa, A, Felix, R, Andreas, R, Englert, F, Daniel, B, Doreen, B, Gottron, C, Ralf, S. SMARTENERGY.KOM: An intelligent system for energy saving in smart home. In: 39th Annual IEEE Conference on Local Computer Networks Workshops; 8-11 September 2014: IEEE, Edmonton, AB, Canada: pp. 685-692.

[60] Frank, E, Till, S, Sebastian, K, Andreas, R, Ralf, S. How to auto-configure your smart home? High-resolution power measurements to the rescue. In: Proceedings of the Fourth International Conference on Future Energy Systems (e-Energy'13); 21-24 May 2013: Association for Computing Machinery (ACM), New York, NY, USA: pp. 215-224.

[61] Andreas, R, Paul, B, Burgstahler, D, Matthias, H, Hristo, C, Werner, M, Ralf, S. On the accuracy of appliance identification based on distributed load metering data. In: 2012 Sustainable Internet and ICT for Sustainability (SustainIT); 4-5 October 2012: IEEE, Pisa, Italy: pp. 1-9.

[62] Andreas, R, Dominic, B, Manzil, Z, Ralf, S. Electric appliance classification based on distributed high resolution current sensing. In: 37th Annual IEEE Conference on Local Computer Networks - Workshops; 22-25 October 2012: IEEE, Clearwater, FL, USA: pp. 999-1005.

[63] Andreas, R, Dominic, B, Parag, SM, Manzil, Z, Ralf, S. SmartMeter.KOM: A low-cost wireless sensor for distributed power metering. In: 2011 IEEE 36th Conference on Local Computer Networks; 4-7 October 2011: IEEE, Bonn, Germany: pp. 1032-1039.