

# Design, Aggregation and Analysis of Power Consumption Data using the Jump Process

Yazid Hambally Yacouba<sup>1</sup>

Amadou Diabagaté<sup>2</sup>, Adama Coulibaly<sup>4</sup>

Training and Research Unit for Mathematics and Computer  
Science, Félix Houphouët Boigny University  
UPHB, Abidjan, Côte d'Ivoire

Michel Babri<sup>3</sup>

Training and Research Unit for Mathematics and Computer  
Science, INPHP, Yamoussoukro  
Côte d'Ivoire

**Abstract**—This work aims to seek a pragmatic approach to assess electricity consumption at the level of households, buildings and neighborhoods. The main concern consists in proposing aggregation methods based on jump process according to a customer environment that is intrinsically linked to the implementation of a centralized system. The aim of the approach is to present data aggregations that derive their basis from a data model in order to facilitate the processing of electricity data at different scales of analysis. Such a smart meter data management process merits the design of an aggregated database that can store data for a house, a building and a neighborhood. The advantage of this system lies in the facilitation of data interpretation and the ability to guide decision-makers in the management of electricity consumption. An analysis of the behavior of electricity consumption is also proposed based on the monitoring of the electricity consumption of the various devices connected to a smart meter.

**Keywords**—Design; aggregation; analysis; jump process; electricity consumption; smart meter

## I. INTRODUCTION

The systems for managing data from smart meters remain closed in the way they process data. However, these systems have the merit of presenting detailed output states. This work consists of:

- Analyzing the existing framework in favor of measuring methods for the smart meter in particular.
- Studying and adapting this research work to the documents and output reports made available by data managers of these smart meters.
- Initializing a data model for the management of smart metering data.
- Explaining the treatments of the central system including the various techniques applied and the presentation of data structures for facilitating data aggregation.
- Presenting the aggregation methods implemented using the jump process.

This work follows a previous work [1] related to the design of a multi-agent system for the management of data from smart meters. The advantage of this approach is to deepen the design

of the system by providing it with data aggregation methods that would lead to results for the benefit of decision-makers.

In [1], presents a formal framework for storing smart meter data and applying methods for analyzing electricity consumption data. For this, a database model for evaluating the electricity consumption data of a house, a building and a neighborhood is proposed. This work highlights a structured set of measurement data from smart meters. There are many elements to consider for the implementation of a data model [2, 3, 4]. Jaime Lloret and al. propose an integrated architecture based on the Internet of Things (IoT) for the deployment of smart meter networks [2]. This article [3] offers a large number of datasets as well as a web portal for visualizing this data. The work in this publication presents key technological solutions for better data analysis and energy efficiency research. In [4], reviews technologies oriented and adapted to the development of applications for the management of smart meter data. Yi Wang and al. provide insight into the future challenges of descriptive analytics including load analysis, predictive analytics, forecasting and prescriptive analytics for load management [4].

The processing of the data collected is carried out by means of aggregation methods. It is also about presenting data structures that will evolve progressively as needed. This work presents existing studies highlighting the importance of forecasts for a good management of electricity consumption, mainly at the level of end customers [5, 6]. In [5], highlights the need for detailed technical information on the devices used by households in order to measure their impacts on household behavior. Ya Wu and Li Zhang study the factors influencing energy saving in tiered electricity pricing and presents empirical results based on variables such as personal characteristics and living conditions [6].

A survey of studies similar to the approach conducted in this paper help to identify a few cases with different interests from each other. Xiaochen Zhang and al. propose a time-variant load model based on the exploration of a historical database of smart meter data [7]. A relational database management system is set up to leverage information that consumers and utilities can get from smart meters data [8]. The achievement of new energy efficiency services is highlighted by the use of smart meters with smartphones through an infrastructure and a set of algorithms [9]. Ming Dong and al. propose a method to help monitoring energy consumption and

identifying individual consumption from large household appliances [10]. Xiufeng Liu and al. [11] propose an innovative ICT solution based on a hybrid architecture, which proves the importance of smart meters in socio-economic surveys and proposes a rational approach to dealing with the complexity of their data. This solution concerns more areas of interest such as geographic locations, weather conditions and user information. Peng Xu and al. propose an ICT solution that aims to highlight the efficiency of smart meters and remote displays through real-time monitoring functions of energy consumption to achieve objectives in environmental and supply security [12]. Juan I. Guerrero and al. provide a method for integrating smart meter data from heterogeneous data sources and modeling the integrated information through an automatic data mining framework [13]. Jinsong Liu and al. [14] present the requirements of the fundamental models for the management of smart metering data and presents a concept of three-mode models as well as the design of the service architecture for accessing to generic models.

The review of work around the evaluation of electricity consumption and its opportunities has made it possible to identify other areas of interest, including studies relating to aggregations, forecasts, evaluation of electrical data and their impacts. There are many approaches of data aggregation that can be temporal, zone and by use. The large volume of these data has also led to the exploration of aggregation methods allowing this data to be represented in other forms while retaining the initial information contained in the raw data [15, 16, 17, 18, 19, 20, 21, 22]. Yi Wang, Qixin Chen and Chongqing Kang provide an overview of research related to the challenges of smart meter data analysis [15]. This work [16] presents a method based on Blockchain and Homomorphic Encryption (HE) technologies to ensure the confidentiality of smart meter data during the data aggregation phase. Mohamed Saleem Haja Nazmudeen and al. aim to adopt the fog computing architecture to reduce the amount of data collected in order to improve the performance of the process applicable to data within the central system [17]. In [18], highlights the need to implement a sub-metering of individual devices in a house to give a better interpretation to the data thanks to a proposed model called the Explicit-Duration Hidden Markov Model (EDHMM-diff). This paper [19] presents a model of data aggregation with fault tolerance, preventing data leakage and demonstrating robustness with negligible cost. Toshichika Shiobara and al. propose a method of reducing the cost of storing and using smart meter data through an aggregation method that reduces the total size of the data and the processing cost [20]. Accurate estimation of the energy [21] requirement forecast based on the amount of load growth and geographic space is carried out based on knowledge of small zone consumption, historical loads and weather conditions. The vector machine algorithm is used for data prediction, classification and analysis using a deep learning approach. Anastasia Ushakova and al. use a set of Gaussian models to aggregate time-series data from smart meters and to understand energy consumption based on the aggregation of data from the target population, particularly by inducing energy-efficient behavior [22].

There are also studies on both long-term and short-term forecasts of electricity demand [23, 24, 25]. In [23], highlights a new method of probabilistic forecasting of electricity demand through a hierarchy with different levels of aggregation such as substations, cities and regions. This study [24] proposes the evaluation of seven algorithms from measurement data collected every fifteen minutes from sensors to predict the electricity consumption of residential buildings in the next hour. This work confirms the best results already obtained from commercial buildings data with methods based on the neural network, unlike the results obtained from residential data. However, the conclusion of this work also shows that the results with the data of residential buildings are better with the Least Squares Support Vector Machines. This article [25] shows the important issue of forecasting electricity consumption by proposing prediction methods based on Short-Term Load Forecast (STLF), the Self-Recurrent Wavelet Neural Network (SRWNN) and the Levenberg–Marquardt (LM) learning algorithm.

Upstream of the aggregation and forecasting of electricity consumption, there is first the need to formalize the data from smart meters in a database model and this is why a set of studies on the evaluation of electricity consumption data from smart meters is presented. The evaluation of electricity data is indeed a major concern and a prerequisite for understanding consumer behavior in order to guarantee energy efficiency at the level of producers, suppliers and customers alike [26, 27, 28, 29]. [26] explores the behavior of electricity customers on the use of heating and cooling. This work also shows the importance of using smart a meter data for understanding a consumer behavior with respect to thermal comfort, particularly in regions where automatic HVAC systems are virtually absent. Ilze Laicane and al. prove the importance of having up-to-date electricity data to determine the profile of households and reduce electricity consumption [27]. This work shows that the energy performance of households depends on the energy efficiency of equipment due to technological progress while highlighting an important part of the change in user behavior in the results observed [27]. In [28], is based on the PROBE and CarbonBuzz initiatives which illustrate that the concept of energy performance for more energy efficiency advocated by the construction industry does not live up to expectations and is biased by taking into account unrealistic factors such as occupancy behavior and facility management relating to the energy models used within these buildings. The study is based on the evaluation of data after occupancy of buildings to establish more realistic energy performance models. In [29], shows the impact of airflow on electricity consumption in computer data centers based on a comparative study of four different cities. Analysis of the cooling periods and energy saving periods yielded results in showing the importance of climatic conditions, energy prices and cooling technologies on cooling efficiency and costs of cooling exploitation.

Likewise, the impacts of the evaluation of smart meter data are numerous [30, 31, 32] and fully participate in encouraging consumers to save energy by identifying different criteria for reducing electricity consumption. This study [30] presents the mistaken motivations of a sample of US grid electricity

customers for choosing smart meters and associated technologies in residential homes and discusses the policy implications and risks perceived by customers. Jacopo Torriti shows that the demands of electricity customers using smart meters are 5.2 times lower than those of users of conventional meters using a comparative study carried out on three different floors of the same building in Italy [31]. Gordon Rausser, Wadim Strielkowski and Dalia Streimikien highlight relevant findings for the attention of stakeholders and policymakers on issues to consider in encouraging positive electricity consumption with smart meters [32].

The benefits of analyzing electricity consumption data from smart meters [5, 33, 34, 35, 36, 37] for consumers and for the environment in particular deserve to be highlighted. Iana Vassileva, Fredrik Wallin and Erik Dahlquist focus on identifying appropriate power saving measures from important data collected and analyzed monthly over a long period from identical buildings [13]. This work consists in defining different behavioral consumption profiles from technical data on electricity consumption, tenant characteristics, energy consumption behaviors as well as type and use of electrical devices [13]. In [33], is a review of the literature that assesses the effectiveness of smart metering in reducing energy consumption while explaining factors related to users behavior. This article [34] is an in-depth analysis of the environmental impacts resulting from the use of smart meters by considering all stages of the life cycle of these meters. In [35], illustrates the importance of electricity consumption data in studying users' consumption patterns in order to classify households according to predefined criteria. This study shows that a priori knowledge of certain criteria such as floor level or number of occupants can improve the accuracy of household classification. In [36], aims to provide industrial consumers with a methodological approach allowing the evaluation and choice of offers from electricity suppliers. The method used is based on characterization of electricity consumption, analysis of tariff offers and forecasts based on energy factors. The results of the study were drawn from the evaluation of fourteen different contracts of electricity suppliers. In [37], emphasizes the importance of long-term forecasting of electricity demand based on forecasts obtained from multivariate regression analysis of electricity consumption data.

Some works using the jump process among which [38, 39, 40, 41, 42] have been identified. The paper [38] determines the integral transforms of the joint distribution of the first-exit time from an interval and the value of a jump of a process over the boundary at exit time and the joint distribution of the supremum, infimum, and value of the process. Clifford A. Ball and Walter N. Torous show a simplified jump process for common stock returns using the jump process models information arrivals and, as such, stock price jumps [39]. The article [40] explores the use of the Markov jump process to model vehicular mobility at the macroscopic level. Lydia Chabane and al. study the fluctuations of systems modeled by Markov jump processes with periodic generators using large deviation theory; canonical biasing and generalized Doob transform [41]. They show that the asymptotic fluctuations process, called driven process, is the minimum under constraint

of the large deviation function for occupation and jumps [41]. Alexander Sikorski and al. present how augmenting the spatial information of the embedded Markov chain by the temporal information of the associated jump times [42]. The approach presented in this work is a very different application of the jump process in a context of electricity consumption management.

## II. PRINCIPLE FOR THE AGGREGATION METHODS OF THE SMART METERING SYSTEM

First, it should be noted that the data collection mechanisms both from smart meters and from the customer base make it possible to automatically provide data without duplication and by identifying proven or probable anomalies in the collection process in case of missing data or outliers through business analytics [1].

Data quality assessment should be subject to automated treatments. In addition, the storage of processed data must be possible over periods of time in accordance with the regulations in force in each country.

After data collection, it is necessary to aggregate the raw data as shown in Fig. 1.

- The aggregation of measurement data can be done according to a time step (hourly, daily, monthly, etc.) regardless of any consideration of the zone and use to which the measurements are linked. This temporal aggregation process can be done in the data collection transaction from the smart meters or in a separate transaction from the data collection process. It is sufficient to rely on the arrival date "ARRIVAL\_DATE" of each data in the central system for the need of ordering the data of a source [1]. The collection of data from each smart meter requires the storage of the date "LAST\_COLLECTION\_DATE" of the last measurement data recorded in the source [1]. This mechanism makes it possible to ensure the collection of all data of a smart meter within the source to which it is attached. Several smart meters could also be associated with a single source. However, it is essential for the aggregation mechanism to save in a transaction, the "LAST\_AGGREGATION\_DATE" parameter for each source data. The "SOURCE" table has therefore been modified by adding the "LAST\_AGGREGATION\_DATE" field. "LAST\_AGGREGATION\_DATE" corresponds to the arrival date in the central system of the last aggregated data for each source. Aggregation is therefore done in order of arrival of data from one or all the smart meters associated with a source as represented in Fig. 2.
- The aggregations are then carried out on the sub-zones of the lowest level towards the sub-zones of the highest level. The aggregation of the data of a bounding zone is conditioned by the aggregation of all its sub-zones.
- Data aggregations are also possible by use. Indeed, the aggregation of an use can be applied to all the measurement data of this use for a given zone.

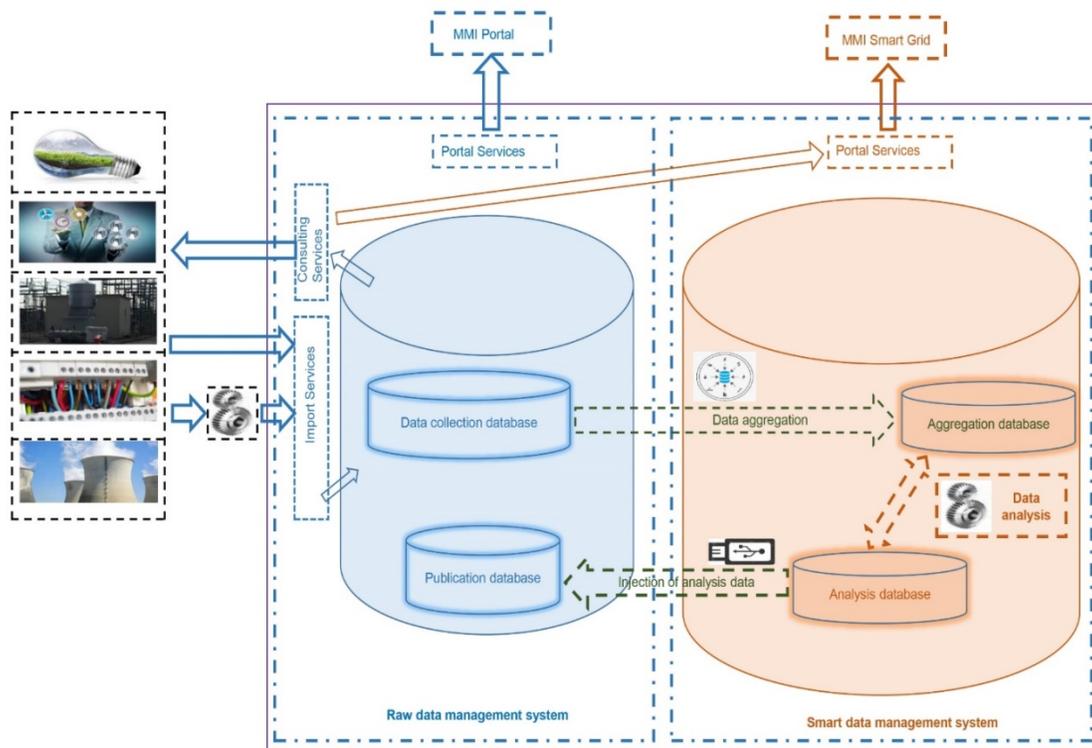


Fig. 1. Functional Architecture of the Electricity System Including Aggregations and Interactions with other Actors in the Electricity Value Chain.

SOURCE		
Feature	Format	Description
ID_SOURCE	Integer [14]	Unique identifier of the source representing the combination of information on the meter number and the customer reference.
REF_CUSTOMER	String[14]	Unique reference allowing identification of the customer regardless of the number of his meter.
LIST_REF_SMART_METER->	List<String[14]>	List of unique references of smart meters.
LAST_COLLECTION_DATE	AAAA-MM-JJThh:mm:ss+hh:mm	Last date of data collected.
LAST_AGGREGATION_DATE	AAAA-MM-JJThh:mm:ss+hh:mm	Last date of date aggregated.
LIST_MEASUREMENTS	List<MEASURE->	List of measurements.

Fig. 2. Representation of a Source.

- It will no longer be possible to integrate data for a validated (procedurally completed) or non-initiated billing period in the aggregated database. This condition helps to limit and avoid the overlap in the collection and aggregation of electricity consumption data over time. This will also make it possible to identify functional and technical irregularities, in particular cases of attempted fraud.

Let's present a view of the measurement data table regardless of its relationship to other entities in the master data model as indicated below in Fig. 3.

MEASURE		
Feature	Format	Description
ID_SOURCE	Integer [14]	Unique identifier of the source representing the combination of information on the meter number and the customer reference.
ID_POINT	Integer	Unique identifier of the measurement point.
MEASURE_TYPE	String	Type of measure
TIMESTAMP	AAAA-MM-JJThh:mm:ss+hh:mm	Specifies the date and time of the measurement from the smart meter
VALUE	Integer	Value of the measure
ARRIVAL_DATE	AAAA-MM-JJThh:mm:ss+hh:mm	Specifies the date and time of receipt of the measurement at the central data processing system

Fig. 3. Representation of the Electricity Consumption of a Measurement.

The analysis of this table highlights the correlation between the volume of data and the time step for collecting the data values for each measurement. The metering data is voluminous and therefore difficult to use.

Reducing complexity involves moving to a more understandable data presentation scale by aggregating data over hourly, daily and monthly ranges. This aggregation consists of the accumulation of consumption data collected over the various intervals for each measurement.

At first glance, collecting the measurements of a house can generate a significant volume of data. The amount of data from a building, or even a neighborhood might be gigantic. Temporal aggregation of data is necessary not only to compress the volume of data but also to facilitate data search, analysis and visualization.

### III. MODELIZATION OF THE TEMPORAL, ZONE AND USE AGGREGATIONS OF THE DATA FOR THE SMART METERING SYSTEM USING THE JUMP PROCESS

#### A. Fundamental Principle of Modelizing the System Aggregations by the Jump Process

The jump process can be used to model the data aggregation of the smart metering system. Indeed, consumption data are recorded at successive and regular times. In addition, the current state of a smart meter is independent of the previous state at each time step of the arrival of electricity consumption data. The jump process thus models the data aggregations as follows:

1) Calculation of the electricity consumption of a smart meter at a time  $t$

$$Z_t = \sum_{n=1}^{\infty} X_n \mathbb{1}_{[T_{n-1}, T_n[}(t), \text{ electricity consumption at time } t.$$

$n$  the number of data records.

$X_n$  electricity consumption between  $T_{n-1}$  and  $T_n$  at time  $t$ .

$$\forall n, 0 = t_1 < t_2 \dots < t_n.$$

$\forall n, X_{t_{n-1}}$  and  $X_{t_n}$  are independent.

$$\mathbb{1}_A(t) = \begin{cases} 1 & \text{if } t \in A \\ 0 & \text{otherwise} \end{cases}$$

$T_n - T_{n-1} = H$ ,  $H$  being the consumption data recording time step for a given meter. In the case of the jump process  $T_{n-1} - T_n$  may not be constant unlike the special case of smart meters.

2) Calculation of the cumulative electricity consumption of a smart meter so far  $t$

$$Y_t = \sum_{n=1}^{\infty} X_n \mathbb{1}_{T_n < t}$$

$Y_t$ , the total electricity consumption up to time  $t$ .

$Y_t$  is a jump process,  $X_n$  being the different values of the smart metering system.

#### B. Temporal Aggregation of the Measurement

1) Case of temporal aggregation of data from smart meters.

a) Temporal aggregation of data from a set of smart meters in the metering system at time  $t$ .

Let  $k$  be the index of any smart meter and  $Z_t^k$  be the electricity consumption at time  $t$ .

$Z_t^k = \sum_{n=1}^{\infty} X_n^k \mathbb{1}_{[T_{n-1}^k, T_n^k[}(t)$ ,  $X_n^k$  being the electricity consumption of the smart meter  $k$  between  $T_{n-1}^k$  and  $T_n^k$  at time  $t$ .

Let  $N$  be the number of smart meters in the metering system and  $S_N$  be the total electricity consumption for all these meters at time  $t$ .

$S_N = \sum_{k=1}^N Z_t^k$  represents the data aggregation for all  $N$  smart meters at time  $t$ .

b) Temporal aggregation of data from a set of smart meters in the metering system up to the instant  $t$ .

Let  $k$  be the index of any smart meter and  $Y_t^k$  be the total electricity consumption up to time  $t$ .

$$Y_t^k = \sum_{n=1}^{\infty} X_n^k \mathbb{1}_{T_n^k < t}$$

Let  $N$  be the number of smart meters in the metering system and  $R_N$  be the total electricity consumption for all of these meters at time  $t$ .

$R_N = \sum_{k=1}^N Y_t^k$  represents the aggregation of data from all  $N$  smart meters up to the moment  $t$ .

#### C. Aggregation by Zone of Measurements

This section is devoted to data aggregations in relation to zones. This step assumes that the temporal aggregations of the measurement data have already been carried out for a specific time step.

The basic data model does not allow a zone aggregation. For that, the notion of zone is necessary. The zone being a space in which one or more measurements can be counted. A zone is attached to a meter that also makes it possible to determine the customer.

It is also necessary to define the notion of parent zone. This notion is essential in the implementation of a neighborhood data model. It is also possible to locate a zone by its geographic coordinates. The difference between a zone of a residence and that of a neighborhood is made by adding the notion of parent zone for a neighborhood. Indeed, the basic data model is suitable for evaluating the power consumption of a residential zone (house, apartment, building, district, etc.).

The "ZONE" table takes into account the characteristics of a building and a district and is presented as follows in Fig. 4:

ZONE		
Feature	Format	Description
ID	Integer [14]	Unique identifier of the zone.
ID_PARENT	Integer [14]	Identifier of the parent zone.
ID_SOURCE	Integer [14]	Identifier of the source.
NESTED_LEVEL	Integer [3]	Indicates the nesting level of a zone (1 for the first base level, 2 for the second level including the base level, 3 for the third level including level 2, etc.).
CODE	String[100]	The unique identification code for the zone.
LABEL	String[200]	The label of the zone.
TYPE	String[1..50]	Takes the values "Building", "Apartment", "Studio", "Bedroom", "Living room" or "Kitchen", and so on.

Fig. 4. Representation of a Zone.

In evaluating the consumption of a building, it is necessary to cumulate the consumption of all the zones of that building. All of the building zones are made up of building zones that do not have a parent zone. Indeed, these zones already take into account the sub-zones, which compose them.

Similarly, the consumption of a district is determined by the cumulative consumption of all the zones of this district that do not have a parent zone (level 0 zone) then come the zones with only one parent (level 1 zone) And so on.

The notion of the "SOURCE" table is also essential for the transition from the basic data model to the district data model.

1) *Presentation of zone aggregation structures:* Below is the presentation of the "CONSUMP\_ZONE" table as shown in Fig. 5. Calculating the consumption of all the zones will facilitate the restitution of the consumption of a building or even a district. Indeed, a building and a district can be considered as a set of independent or combined zones through which the measurements are distributed.

The identifier "ID\_SOURCE" has been added to the table "CONSUMP\_ZONE" to both take into account the size of the building and the district but also to avoid duplicating zones with the same identifiers and belonging to different sources.

CONSUMP_ZONE		
Feature	Format	Description
ID_ZONE	Integer [14]	Unique identifier of the zone.
ID_SOURCE	Integer [14]	Unique identifier of the source.
TIMESTAMP	AAAA-MM-JJThh:mm:ss+hh:mm	Corresponds to the time slot of the consumption.
VAL_CONSUMP	Integer	Total value of consumption for the time slot.

Fig. 5. Representation of Total Electricity Consumption over a Slot Time for a Zone.

2) Case of data aggregation by zone from smart meters.

a) Aggregation by zone of data from a set of smart meters in the metering system at the instant  $t$

Let  $k$  be the index of any smart meter and  $Z_t^k(i)$  the electricity consumption at time  $t$  in zone  $i$ .

$$Z_t^k(i) = \sum_{n=1}^{\infty} X_n^k(i) \mathbb{1}_{[T_{n-1}^k, T_n^k[} (t) \quad , \quad X_n^k(i) \text{ being the electricity consumption of the smart meter } k \text{ between } T_{n-1}^k \text{ and } T_n^k \text{ at time } t \text{ in zone } i.$$

Let  $N$  be the number of smart meters in the metering system and  $S_N(i)$  the total electricity consumption for all these meters at time  $t$  in zone  $i$ .

$S_N(i) = \sum_{k=1}^N Z_t^k(i)$  represents the aggregation of data for all  $N$  smart meters at time  $t$  in zone  $i$ .

b) Aggregation by zone of data from a set of smart meters in the metering system up to time  $t$ .

Let  $k$  be the index of any smart meter and  $Y_t^k(i)$  the total electricity consumption up to time  $t$  in zone  $i$ .

$$Y_t^k(i) = \sum_{n=1}^{\infty} X_n^k(i) \mathbb{1}_{T_n^k < t}$$

Let  $N$  be the number of smart meters in the metering system and  $R_N(i)$  the total electricity consumption for all of these meters at time  $t$  in zone  $i$ .

$R_N(i) = \sum_{k=1}^N Y_t^k(i)$  represents the aggregation of data from all  $N$  smart meters up to time  $t$  in zone  $i$ .

D. Aggregation by use of Measurements

In this section, the presentation of aggregations in relation to uses is highlighted. These aggregations are limited in space as they are attached to a zone, which can be a building or even a neighborhood. Indeed a district can be described as a set of nested zones or not.

1) *Presentation of aggregation structures for an use;* Measurement data can be matched to uses. Indeed, the consumption data from each measurement comes from a well-defined type of measure (lighting, air conditioning, a socket, etc).

Each source has a set of uses associated with it as displayed in Fig. 6. The "USE" table therefore takes a reference to the "SOURCE" table.

USE		
Feature	Format	Description
ID_USE	Integer [14]	Unique identifier of the use.
ID_SOURCE	Integer [14]	Unique identifier of the source.
CODE	String[100]	Unique identification code for the zone.
LABEL	String[200]	Label of the zone.
TYPE	String[1..50]	Takes the values "Lighting", "Air conditioning" or "Socket", etc.

Fig. 6. Representation for an use of Electricity.

Each measurement is associated with an use. A reference to the "SOURCE" table has therefore been added to the "USE" table, which makes it possible to identify the use for all measurements.

The use information is only define at the source level but not at the smart meter level. Aggregation by use for a building or a district requires the information of the field "ID\_SOURCE" at the level of the "USE" table to link each measurement to its source of correspondence as observed in Fig. 7. The "ID\_SOURCE" field also makes it possible to resolve conflicts in the case of identical identifiers of measurements belonging to different sources.

CONSUMP_USE		
Feature	Format	Description
ID_SOURCE	Integer [14]	Unique identifier of the source.
ID_USE	Integer	Unique identifier of the use.
TIMESTAMP	AAAA-MM- JJThh:mm:ss+hh:mm	Corresponds to the time slot of the consumption.
VAL_CONSUMP	Integer	Total value of consumption for the time slot.

Fig. 7. Representation of Total Electricity Consumption over a Slot Time for an use.

2) Case of data aggregation by use from smart meters.

a) Aggregation by use of the data of a set of smart meters in the metering system at time  $t$ .

Let  $k$  be the index of any smart meter and  $Z_t^k(j)$  the electricity consumption at time  $t$  for use  $j$ .

$Z_t^k(j) = \sum_{n=1}^{\infty} X_n^k(j) \mathbb{1}_{[T_{n-1}^k, T_n^k[}(t)$ ,  $X_n^k(j)$  being the electricity consumption of the smart meter  $k$  between  $T_{n-1}^k$  and  $T_n^k$  at time  $t$  for use  $j$ .

Let  $N$  be the number of smart meters in the metering system and  $S_N(j)$  the total electricity consumption for all of these meters at time  $t$  for use.

$S_N(j) = \sum_{k=1}^N Z_t^k(j)$  represents the data aggregation of all  $N$  smart meters at time  $t$  for use  $j$ .

b) Aggregation by use of the data of a set of smart meters in the metering system up to time  $t$ .

Let  $k$  be the index of any smart meter and  $Y_t^k(j)$  the total electricity consumption up to time  $t$  for use  $j$ .

$$Y_t^k(j) = \sum_{n=1}^{\infty} X_n^k(j) \mathbb{1}_{T_n^k < t}$$

Let  $N$  be the number of smart meters in the metering system and  $R_N(j)$  the total electricity consumption for all of these meters at time  $t$  for use  $j$ .

$R_N(j) = \sum_{k=1}^N Y_t^k(j)$  represents the aggregation of data from all  $N$  smart meters up to time  $t$  for use  $j$ .

#### E. Aggregation by Zone and by use of the Measurements

A set of uses constitute the link between a zone and its measurement data. For this purpose, the definition of the ZONE\_USE table is necessary. This table contains the references of all uses and measurements attached to each zone.

1) Presentation of aggregation structures for a zone and an use: In a zone containing sub-zones, it is necessary to group together all the uses according to the types of uses and sub-zones. This will make it possible to present the consumption of uses for a given zone as a whole but also to highlight the detail of the consumption of uses for a given zone according to its sub-zones. Fig. 8 shows the link between a zone and an use.

ZONE_USE		
Feature	Format	Description
ID_ZONE	Integer [14]	Unique identifier of the zone.
ID_USE	Integer [14]	Unique identifier of the use.
ID_SOURCE	Integer [14]	Unique identifier of the source.

Fig. 8. Representation of the Relationship for a Zone according to an Electricity use.

Fig. 9 is a view of the table containing consumption by zone and by use:

CONSUMP_ZONE_USE		
Feature	Format	Description
ID_ZONE	Integer [14]	Unique identifier of the measurement zone.
ID_USE	Integer	Unique identifier of the use.
ID_SOURCE	Integer [14]	Unique identifier of the source.
TIMESTAMP	AAAA-MM- JJThh:mm:ss+hh:mm	Corresponds to the time slot of the consumption.
VAL_CONSUMP	Integer	Total value of consumption for the time slot.

Fig. 9. Representation of the Relationship of Total Electricity Consumption over a Period for a Zone according to an Electricity use.

2) Case of data aggregation by zone and by use.

a) Aggregation by zone and by use of the data of a set of smart meters in the metering system at time  $t$ .

Let  $k$  be the index of any smart meter and  $Z_t^k(i, j)$  the electricity consumption at time  $t$  in zone  $i$  and for use  $j$ .

$Z_t^k(i, j) = \sum_{n=1}^{\infty} X_n^k(i, j) \mathbb{1}_{[T_{n-1}^k, T_n^k[}(t)$ ,  $X_n^k(j)$  being the electricity consumption of the smart meter  $k$  between  $T_{n-1}^k$  and  $T_n^k$  at time  $t$  in zone  $i$  and for use  $j$ .

Let  $N$  be the number of smart meters in the metering system and  $S_N(i, j)$  the total electricity consumption for all of these meters at time  $t$  in zone  $i$  and for use  $j$ .

$S_N(i, j) = \sum_{k=1}^N Z_t^k(i, j)$  represents the aggregation of data from all  $N$  smart meters at time  $t$  in zone  $i$  and for use  $j$ .

b) Aggregation by zone and by use of data from a set of smart meters in the metering system up to time  $t$ .

Let  $k$  be the index of any smart meter and  $Y_t^k(i, j)$  the total electricity consumption up to time  $t$  in zone  $i$  and for use  $j$ .

$$Y_t^k(i, j) = \sum_{n=1}^{\infty} X_n^k(i, j) \mathbb{1}_{T_n^k < t}$$

Let  $N$  be the number of smart meters in the metering system and  $R_N(i, j)$  the total electricity consumption for all of these meters at time  $t$  in zone  $i$  and for use  $j$ .

$R_N(i, j) = \sum_{k=1}^N Y_t^k(i, j)$  represents the aggregation of data from all  $N$  smart meters up to time  $t$  in zone  $i$  and for use  $j$ .

#### IV. ANALYSIS OF USER CONSUMPTION BEHAVIOR

This section focuses on the analysis of the behavior of electricity consumption by users. To do this, A study is carried out on the effects of consumer's daily actions on the use of each device connected to its smart meter.

Let  $M$  be the number of devices connected to a given smart meter with index  $k$ . Then consider  $l$  the index of any device connected to this smart meter  $k$ .

Let  $T_l^n$  be the duration of consumption for device  $l$  in the interval  $[T_{n-1}, T_n]$  and let  $C_l$  be the consumption per unit of time for this device.

The consumption of the device  $l$  in the interval  $[T_{n-1}, T_n]$  can then be represented by  $C_l^n = C_l \times T_l^n$ .

The duration of consumption  $T_l^n$  is random because the interval  $[T_{n-1}, T_n]$  may or may not occur, which makes it possible to deduce that  $T_l^n$  therefore follows a continuous law. In addition, the duration which separates two consecutive realizations in a series of independent realizations is modeled by an exponential law of parameter  $\theta$ .  $T_l^n$  is therefore an exponential variable.

The consumption of electricity between  $T_{n-1}$  and  $T_n$  is independent of the consumption before  $T_{n-1}$  and that after  $T_n$ .

The duration between any two realizations in a series of independent realizations follows a gamma law with parameter  $(a, \theta)$ ,  $a$  designates the number of intervals between the first realization and the last realization; and  $\theta$  denotes the average number of realizations per unit of time.  $\theta$  is estimated over time by following the consumption history linked to a smart meter.

If the duration of consumption  $T_l^n$  is performed only once in the interval  $[T_{n-1}, T_n]$ , this duration of consumption  $T_l^n$  follows an exponential law.

If the duration of consumption  $T_l^n$  is carried out several times in the interval  $[T_{n-1}, T_n]$ ,  $T_l^n$  follows a gamma law.

$X_n^k = \sum_{l=1}^M C_l^{n,k} = \sum_{l=1}^M C_l^k \times T_l^n$  with  $C_l^{n,k}$  the consumption read by the meter  $k$  for the device  $l$  in the interval  $[T_{n-1}, T_n]$  [and  $C_l^k$  the consumption per unit of time for the device  $l$  relatively to the characteristics of the smart meter  $k$ .

#### V. DISCUSSION

The present work is part of an approach to design a formal framework for storing data from smart meters. This study also aims to present an approach for resizing large data from smart meters in order to facilitate their use. Likewise, the processing methods applied to data are presented.

This work shows the importance of data aggregation in the management of smart meters data and proposes a method of aggregating this data, based on the jump process. In addition, the storage structures for raw data and aggregated data are described as well as the mechanisms for recording and updating this data. An analysis of the behavior of electricity consumption is also carried out based on the actions of electricity consumers and the characteristics of the various electrical devices connected to their smart meters.

The research carried out did not make it possible to identify similar work on the aggregation of data from smart meters, on an attempt to present the structures of their databases and even less on the details of the processing applied to their data. It is indeed all these reasons that motivated and led to the completion of this work. This publication aims to promote an understanding of how smart meters work and to provide a basis for further research.

However, there is a great deal of research on the analysis of the behavior inherent in the consumption of electricity from data compiled by existing systems [8, 43]. The publication [8] provides an overview on algorithms and applications applicable to smart meter data. The article [43] shows that the individual predictability of user consumption can be determined with a high degree through time models of energy demand analysis. Likewise, the analysis of electricity consumption in the present publication is distinct from those encountered in similar studies including [8, 10, 12]. It is indeed atypical and is based on the determination of the mathematical law likely to explain the phenomenon of electricity consumption.

It is therefore necessary to underline that the work carried out within the framework of this publication covers several aspects of smart meters data management and provides contributions that can be explored in future publications.

#### VI. CONCLUSION

This work made it possible to define the framework for using data from smart meters. It was used to design a data model based on a comparative study of the functionalities of smart meters. The specificity of the data in a metering system has led to the implementation of data aggregation methods to facilitate data processing and analysis. The complexity of the data management of smart metering systems is highlighted by the proposal for a comprehensive approach that includes a concrete case of implementation.

The limitations of this work lie in the lack of an implementation of the metering system. Likewise, this does not make it possible to test the performance of the various algorithms put in place. However, issues relating to the implementation of an intelligent metering system as well as that of data processing are presented. The proposed design makes it possible to process the data collected at the level of a house and a neighborhood.

The originality of this work lies in the presentation of a mechanism for processing centralized data from smart meters, the aggregation of this data by the jump process and the determination of the law determining the behavior of the users on electricity consumption.

The lack of information on smart metering systems makes any comparative study difficult. Indeed, the implementation of these systems remains closed even if the functionalities are well known to the users. The future work will address the issues of factors influencing electricity consumption such as weather conditions, communication, billing, data correction and incident management on smart meters.

REFERENCES

- [1] Yazid Hambally Yacouba, Amadou Diabagaté, Abdou Maiga, Adama Coulibaly. Multi-agent system for management of data from electrical smart meters.
- [2] Jaime Lloret, Jesus Tomas, Alejandro Canovas, Lorena Parra. An Integrated IoT Architecture for Smart Metering. Instituto de Investigación para la Gestión Integrada de zonas Costeras. Universidad Politécnica de Valencia, Spain.
- [3] Xiufeng Liu, Per Sieverts Nielsen. A Hybrid ICT-Solution for Smart Meter Data Analytics. Energy, Volume 115, Part 3, 15 November 2016, Pages 1710-1722.
- [4] Yi Wang, Qixin Chen, Tao Hong, Chongqing Kang. Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges. IEEE Transactions on Smart Grid ( Volume: 10 , Issue: 3 , May 2019 ).
- [5] Iana Vassileva, Fredrik Wallin, Erik Dahlquist. Analytical comparison between electricity consumption and behavioral characteristics of Swedish households in rented apartments. Applied Energy Volume 90, Issue 1, February 2012, Pages 182-188.
- [6] Ya Wu, Li Zhang. Evaluation of energy saving effects of tiered electricity pricing and investigation of the energy saving willingness of residents. Energy Policy Volume 109, October 2017, Pages 208-217.
- [7] Xiaochen Zhang; Santiago Grijalva; Matthew J. Reno. A Time-Variant Load Model Based on Smart Meter Data Mining. 2014 IEEE PES General Meeting | Conference & Exposition, DOI: 10.1109/PESGM.2014.6939365, Electronic ISBN: 978-1-4799-6415-4.
- [8] Xiufeng Liu; Lukasz Golab; Ihab F. Ilyas. SMAS: A smart meter data analytics system. 2015 IEEE 31st International Conference on Data Engineering, DOI: 10.1109/ICDE.2015.7113405, Electronic ISBN:978-1-4799-7964-6.
- [9] Markus Weiss; Adrian Helfenstein; Friedemann Mattern; Thorsten Staake. Leveraging smart meter data to recognize home appliances. 2012 IEEE International Conference on Pervasive Computing and Communications, DOI: 10.1109/PerCom.2012.6199866.
- [10] Ming Dong; Paulo C. M. Meira; Wilsun Xu; Walimir Freitas. An Event Window Based Load Monitoring Technique for Smart Meters. IEEE Transactions on Smart Grid ( Volume: 3, Issue: 2, June 2012), DOI: 10.1109/TSG.2012.2185522.
- [11] Xiufeng Liu, Per Sieverts Nielsen. A Hybrid ICT-Solution for Smart Meter Data Analytics. Energy, Volume 115, Part 3, 15 November 2016, Pages 1710-1722, <https://doi.org/10.1016/j.energy.2016.05.068>.
- [12] Peng Xu, Jingchun Shen, Xingxing Zhang, Xudong Zhao , Yingchu Qian. Case Study of Smart Meter and In-home Display for Residential Behavior Change in Shanghai, China. Energy Procedia Volume 75, August 2015, Pages 2694-2699.
- [13] Juan I. Guerrero, Antonio García, Enrique Personal, Joaquín Luque, Carlos León. Heterogeneous data source integration for smart grid ecosystems based on metadata mining. Expert Systems with Applications Volume 79, 15 August 2017, Pages 254-268.
- [14] Jinsong Liu, Xiaolu Li, Member, IEEE, Dong Liu, Member, IEEE, Hesen Liu, and Peng Mao. Study on Data Management of Fundamental Model in Control Center for Smart Grid Operation. IEEE Transactions on Smart Grid ( Volume: 2, Issue: 4, Dec. 2011), DOI: 10.1109/TSG.2011.2160571.
- [15] Yi Wang, Qixin Chen, Chongqing Kang. Overview of Smart Meter Data Analytics. Springer Link, 25 February 2020.
- [16] Yuxuan Wang ; Fengji Luo ; Zhaoyang Dong ; Ziyuan Tong ; Yichen Qiao. Distributed meter data aggregation framework based on Blockchain and homomorphic encryption. IET Cyber-Physical Systems: Theory & Applications ( Volume: 4 , Issue: 1 , 3 2019 ), DOI: 10.1049/iet-cps.2018.5054.
- [17] Mohamed Saleem Haja Nazmudeen ; Au Thien Wan ; Seyed M. Buhari. Improved throughput for Power Line Communication (PLC) for smart meters using fog computing based data aggregation approach. 2016 IEEE International Smart Cities Conference (ISC2), DOI: 10.1109/ISC2.2016.7580841.
- [18] Zhenyu Guo ; Z. Jane Wang ; Ali Kashani. Home Appliance Load Modeling From Aggregated Smart Meter Data. IEEE Transactions on Power Systems ( Volume: 30 , Issue: 1 , Jan. 2015 ).
- [19] Xiaodi Wang, Yining Liu, Kim-Kwang Raymond Choo. Fault Tolerant Multi-subset Aggregation Scheme for Smart Grid. IEEE Transactions on Industrial Informatics, 05 August 2020, DOI: 10.1109/TII.2020.3014401.
- [20] Toshichika Shiobara ; Peter Palensky ; Hiroaki Nishi. Effective metering data aggregation for smart grid communication infrastructure. IECON 2015 - 41st Annual Conference of the IEEE Industrial Electronics Society.
- [21] J. Shanmugasundaram, K. Indhumathi, S. Sivaranjani. Deep Learning Approach cum Aggregated Smart Meter Data Based Residential Energy Load Modeling. Iconic Research and Engineering journals, february 2020, volume 3, issue 8, issn: 2456-8880.
- [22] Anastasia Ushakova, Slava Jankin Mikhaylov. Big data to the rescue? Challenges in analysing granular household electricity consumption in the United Kingdom. Energy Research & Social Science, Volume 64, June 2020, 101428.
- [23] Souhaib Ben Taieb, James W. Taylor & Rob J. Hyndman. Hierarchical Probabilistic Forecasting of Electricity Demand With Smart Meter Data. Journal of the American Statistical Association, 30 Mar 2020.
- [24] Predicting future hourly residential electrical consumption: A machine learning case study. Energy and Buildings, Volume 49, June 2012, Pages 591-603.
- [25] Hamed Chitsaz, Hamid Shaker, Hamidreza Zareipour, David Wood, Nima Amjadi. Short-term electricity load forecasting of buildings in microgrids. Energy and Buildings, Volume 99, 15 July 2015, Pages 50-60.
- [26] Pedro Gouveia, Júlia Seixas, Ana Mestre. Daily electricity consumption profiles from smart meters - Proxies of behavior for space heating and cooling. Energy, Volume 141, 15 December 2017, Pages 108-122.
- [27] Ilze Laicane, Dagnija Blumberga, Andra Blumberga, Marika Rosa. Evaluation of Household Electricity Savings. Analysis of Household Electricity Demand Profile and User Activities. Energy Procedia, Volume 72, June 2015, Pages 285-292.
- [28] Anna Carolina Menezes, Andrew Cripps, Dino Bouchlaghem, Richard Buswell. Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap. Applied Energy Volume 97, September 2012, Pages 355-364.
- [29] Z. Song, X. Zhang, C. Eriksson. Data Center Energy and Cost Saving Evaluation. Energy Procedia 75 (2015) 1255 – 1260.
- [30] Tamar Krishnamurti, Daniel Schwartz, Alexander Davis, Baruch Fischhoff, Wändi Bruine de Bruin, Lester Lave, Jack Wang. Preparing for smart grid technologies: A behavioral decision research approach to understanding consumer expectations about smart meters. Energy Policy, Volume 41, February 2012, Pages 790-797.
- [31] Jacopo Torriti. People or machines? Assessing the impacts of smart meters and load controllers in Italian office spaces. Energy for Sustainable Development Volume 20, June 2014, Pages 86-91.
- [32] Gordon Rausser, Wadim Strielkowski and Dalia Streimikien. Smart meters and household electricity consumption: A case study in Ireland. Energy & Environment, November 22, 2017, DOI: 10.1177/0958305X17741385.
- [33] I. Laicâne, A. Blumberga, M. Rosa and D. Blumberga. Assessment of changes in households' electricity consumption. Agronomy Research 11 (2), 335-346, 2013.
- [34] Slavisa Aleksic, Vedad Mujan. Exergy-based life cycle assessment of smart meters. IEEE, 14 July 2016, DOI: 10.1109/ELEKTRO.2016.7512075.
- [35] Christian Beckel, Leyna Sadamori, Silvia Santini. Automatic socio-economic classification of households using electricity consumption data. e-Energy '13: Proceedings of the fourth international conference on Future energy systems, May 2013, Pages 75-86, <https://doi.org/10.1145/2487166.2487175>.
- [36] Bruna Di Silvio, Vittorio Cesarotti, Vito Introna . Evaluation of electricity rates through characterization and forecasting of energy consumption: A case study of an Italian industrial eligible customer. International Journal of Energy Sector Management, ISSN: 1750-6220, Publication date: 22 May 2007.

- [37] A K Intiaz; Norman B Mariun ; M M R Amran ; M Saleem ; N I A Wahab ; Mohibullah. Evaluation and Forecasting of Long Term Electricity Consumption Demand for Malaysia by Statistical Analysis. 2006 IEEE International Power and Energy Conference, DOI: 10.1109/PECON.2006.346658.
- [38] V. F. Kadankov and T. V. Kadankova. on the Distribution of the Time of the First Exit From an Interval and the Value of a Jump Over the Boundary for Processes with Independent Increments and Random Walks. Ukrainian Mathematical Journal, Vol. 57, No. 10, 2005.
- [39] Clifford A. Ball and Walter N. Torous. A Simplified Jump Process for Common Stock Returns. The Journal of Financial and Quantitative Analysis , Mar., 1983, Vol. 18, No. 1 (Mar., 1983), pp. 53-65.
- [40] Yong Li, Depeng Jin, Zhaocheng Wang, Pan Hui, Lieguang Zeng, and Sheng Chen. A Markov Jump Process Model for Urban Vehicular Mobility: Modeling and Applications. IEEE TRANSACTIONS ON MOBILE COMPUTING, VOL. 13, NO. 9, SEPTEMBER 2014.
- [41] Lydia Chabane, Raphaël Chérite and Gatien Verley. Periodically driven jump processes conditioned on large deviations. Laboratoire de Physique Théorique (UMR8627), CNRS, Univ. Paris-Sud, Université Paris-Saclay, 91405 Orsay, France, Laboratoire J A Dieudonné, UMR CNRS 7351, Université de Nice Sophia Antipolis, Nice 06108, France (Dated: October 30, 2019).
- [42] Alexander Sikorski, Marcus Weber, Christof Schütte. The Augmented Jump Chain - a sparse representation of time-dependent Markov jump processes. Zuse Institute Berlin, November 2020.
- [43] Adrian Albert, Ram Rajagopal. Smart Meter Driven Segmentation: What Your Consumption Says About You. IEEE Transactions on Power Systems ( Volume: 28, Issue: 4, Nov. 2013).