
This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/19251/

Link to published version: https://doi.org/10.1109/AFRCON.2017.8095675

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.
Residential Load Variability and Diversity at Different Sampling Time and Aggregation Scales

Andreas I. Elombo*, Thomas Morstyn, Dimitra Apostolopoulou, Malcolm D. McCulloch

Energy & Power Group: Department of Engineering Science
University of Oxford
Oxford, England, UK
*andreas.elombo@eng.ox.ac.uk

Abstract—The increasing use of large-scale intermittent distributed renewable energy resources on the electrical power system introduces uncertainties in both network planning and management. In addition to architectural changes to the power system, the applications of demand side response (DSR) also add a dimension of complexity – thereby converting the traditionally passive customers into active prosumers (customers that both produce and consume electricity). It has therefore become important to conduct detailed studies on system load profiles to uncover the nature of the system load. These studies could help distribution network operators (DNOs) to adopt relevant strategies that can accommodate new resources such as distributed generation and energy storage on the evolving distribution network and ensure updated design and management approaches. This paper investigates the relationship between both the system load diversity and variability when different customers are aggregated at different scales. Additionally, the implication of sampling time scales is investigated to capture its effect on load diversity and variability. The study looks at the diversity and variability that is observable from the viewpoint of higher power levels, when interconnecting different sized groupings of customers, at different sampling resolutions. The paper thus concludes that the per-customer capacity requirement of the network decreases as the size of customer groupings increases. The load variability also decreases as the aggregation level increases. For active network management, faster time scales are required at lower aggregation scales due to high load variability.

Keywords—After-diversity maximum demand; variance; aggregation; sampling resolution; system load

I. INTRODUCTION

The electrical power system is evolving into a more complex system that demands smarter and more efficient planning and operation techniques. This evolution is particularly fueled by the increasing use of distributed renewable energy resources, most of which are intermittent in nature. The increase in large scale connections of these energy resources is expected to continue into the future. In addition to architectural changes to the power system, the applications of demand side response (DSR) also add a dimension of complexity – thereby converting the traditionally passive customers into active prosumers (customers that both produce and consume electricity) [1]–[3]. For this reason, it has become important to conduct detailed studies on system load profiles to uncover the nature of the system load. This knowledge is crucial for distribution network operators (DNOs) to adopt relevant strategies that can accommodate new resources on the evolving distribution network and ensure updated design and management approaches [4], [5].

The residential sector makes up a significant proportion of the total system load, and it is commonly connected to the network 24 hours a day, all through the year [4]. In fact, the UK’s residential load demand represented about 35% of its total electricity demand in 2013 [6], [7]. This figure is expected to be higher today, given the recorded 1% year-on-year rise in the domestic load since 1970 [6]. Residential homes have widely varying electricity consumptions, and their upper-bound consumptions cannot be accurately forecasted. This creates an uncertainty in keeping a carefully managed balance between generation and demand on the system, especially at the distribution level, as well as in the design process of the distribution network. The use of data loggers (and more recently, smart meters) has gained popular use in collecting electricity consumption data at residential homes to study the patterns that may exist [8]–[11]. Such data serves as valuable input to load modelling and forecasting processes as demonstrated in [2], [12]–[14].

Traditionally, distribution networks were designed according to the expected peak demand, which determines the maximum capacity of the network infrastructure [1]. This characteristic is commonly known as the after-diversity maximum demand (ADMD) of the network, and it is defined as the maximum demand per customer as the number of connected customers, each with an individual maximum demand, approaches infinity [2], [15], [16]. The significant increase of stochastic distributed energy resources at the distribution level introduces the need for active control. This has stirred interest in understanding how the system net load (taking into account the distributed energy resources) varies by using data collected from the field.

Aggregating residential load data gives insight into the dynamic characteristics of the load at different levels of the distribution network, whilst a higher sampling resolution can provide information on the load variability at faster time scales [9]. For the ADMD, there is widespread knowledge that the maximum time-coincident demand per load entity, and the uncertainty of such occurrences, both decrease at increasing aggregation scales [4], [15]–[18]. A less explored aspect is the effect of sampling resolution at different aggregation scales. The
The remainder of the paper is organized as follows: Section II presents details of the proposed framework for data analysis and description of the data used in the study; Section III discusses the results obtained through the prosed framework; Section IV provides key conclusions and presents areas for future work.

II. PROPOSED FRAMEWORK

A. Description of Input Data

Three different data sets of residential load profiles were used to allow for a comparative study. The data was obtained from the following sources:

- **Centre for Renewable Energy Systems Technology (CREST) Model**: an Excel Workbook model by CREST of Loughborough University; generating a 24-hour period load profile within a single UK home at a 1-minute resolution. A total of 200 load profiles were generated. Detailed descriptions of the model are presented in [12]. This data is henceforth referred to as CREST data.

- **Low Voltage Network Solutions (LVNS) Tier-1 Project by Electricity North West Limited (ENWL) and University of Manchester**: a total of 200 load profiles were adopted; 24-hour period load profiles at a 5-minute resolution. Full details of the load profiles are provided in [21]. This data is henceforth referred to as LVNS data.

- **Belgian Low Voltage (LV) Feeder Smart Meter (SM) Data**: 2-year long daily load profiles obtained from smart meters at a 15-minute resolution, from which a total of 200, 24-hour period load profiles were adapted. It is henceforth referred to as LVSM data. This is real data sourced from an ongoing research project being conducted by the University of Mons, in Belgium.

The first two are synthetic data; and the third data set is constructed based on real measurements of the Belgian network.

B. The Analysis Framework

The analyses of diversity and variability in relation to both aggregation and sampling resolution scales were done according to the framework illustrated in Fig. 1. The framework was based on the Monte Carlo approach, which is a method of propagating uncertainty from the input to the output sources of a system. In particular, multiple randomized experiments are executed where a number of random inputs are fed to the model to compute its outputs [22]. This approach is necessary due to the large number of possible combinations of homes that occur at each aggregation scale. The idea of using repeated experiments was to estimate the true values of the parameters of interest, and it is expected that as the number of the Monte Carlo trials increase, the observed parameters will tend towards their true values [23].

The random inputs to the model were the 24-hour period load profiles from the respective data sets, and the output parameters of interest were the mean, ADMD, and variance (denoted as VAR). For each sampling resolution, aggregation scales from 1 to 60 homes were considered, and a total of 100 Monte Carlo trials were carried out to compute the model outputs for each of the 60 aggregation scales.

To include the impact of time scale on diversity and variability, it is necessary to consider data with various sampling resolutions. The sampling resolutions (in minutes) for the three respective data sets considered in the study were:

\[
\tau_{\text{CREST}} = \{1, 6, 10, 12, 15, 20, 30, 60\} \quad (1)
\]

\[
\tau_{\text{LVNS}} = \{5, 10, 15, 20, 30, 60\} \quad (2)
\]

\[
\tau_{\text{LVSM}} = \{15, 30, 60\} \quad (3)
\]

These were selected so that they were integer multiples of the underlying data sampling resolutions, and divided evenly into the number minutes during the day.

For each sampling interval \(\tau\), the pool of 24-hour load profiles for the 200 homes in each dataset (from which inputs to the model are randomly sampled) can be represented as follows:

\[
P_{t,m}^{\tau}, t \in \{1, ..., T_\tau\}, m \in \{1, ..., 200\}, T_\tau = \frac{24 \times 60}{\tau} \quad (4)
\]

where \(P_{t,m}^{\tau}\) is the load for home \(m\) at the time interval \(t\), for the time-series load profiles sampled at a given resolution of \(\tau\) minutes.

Aggregation scales from 1 to 60 homes were used. For each aggregation scale \(\alpha = \{1, ..., 60\}\), 100 Monte Carlo trials are...
completed. For the $k^{th}$ Monte Carlo trial at aggregation scale $\alpha$, a set of $\alpha$ houses, $H_{\alpha}$, is randomly obtained by sampling homes without replacement. Then, the column vector of the average powers at each time of day for the aggregated group of homes at a sampling resolution $\tau$ is given by:

$$
\mathbf{\bar{P}}^\tau_{\tau,\alpha,k} = \left[ \bar{P}^\tau_{1,\alpha,k}, \ldots, \bar{P}^\tau_{T,\alpha,k} \right] = \frac{1}{\alpha} \sum_{n=1}^{\alpha} \bar{P}^\tau_{1,n,\alpha,k}, \ldots, \frac{1}{\alpha} \sum_{n=1}^{\alpha} \bar{P}^\tau_{T,n,\alpha,k}
$$

(5)

The mean, ADMD and variance are thus computed as follows:

$$
MEAN(\tau, \alpha) = \frac{1}{100} \sum_{k=1}^{100} \mathbf{\bar{P}}^\tau_{\tau,\alpha,k}^T \mathbf{\bar{P}}^\tau_{\tau,\alpha,k}
$$

(6)

$$
VAR(\tau, \alpha) = \frac{1}{100} \sum_{k=1}^{100} \mathbf{\bar{P}}^\tau_{\tau,\alpha,k}^T (\mathbf{\bar{P}}^\tau_{\tau,\alpha,k} - MEAN(\tau, \alpha) \mathbf{1})^2
$$

(7)

$$
ADMD(\tau, \alpha) = \max_{1 \leq k \leq 100} \max_{1 \leq \tau \leq T} \left( \mathbf{\bar{P}}^\tau_{\tau,\alpha,k} \right)
$$

(8)

C. Results Representation

According to the proposed framework, we thus obtain results in the format represented in TABLE I. These results were obtained for each data set in accordance with the corresponding sampling resolutions as defined in (1)-(3).

TABLE I. THE REPRESENTATION OF THE MODEL OUTPUTS FOR EACH DATASET

<table>
<thead>
<tr>
<th>Output Matrix</th>
<th>[\text{MEAN}(\tau, \alpha), [\text{kW}]]</th>
<th>[\text{ADMD}(\tau, \alpha), [\text{kW}]]</th>
<th>[\text{VAR}(\tau, \alpha), [\text{kW}^2]]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_1$</td>
<td>$\text{MEAN}(\tau_1, \alpha_1)$</td>
<td>$\text{ADMD}(\tau_1, \alpha_1)$</td>
<td>$\text{VAR}(\tau_1, \alpha_1)$</td>
</tr>
<tr>
<td>$\tau_2$</td>
<td>$\text{MEAN}(\tau_2, \alpha_2)$</td>
<td>$\text{ADMD}(\tau_2, \alpha_2)$</td>
<td>$\text{VAR}(\tau_2, \alpha_2)$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$\tau_T$</td>
<td>$\text{MEAN}(\tau_T, \alpha_T)$</td>
<td>$\text{ADMD}(\tau_T, \alpha_T)$</td>
<td>$\text{VAR}(\tau_T, \alpha_T)$</td>
</tr>
</tbody>
</table>

$n$ denotes the number of sampling intervals considered.

$n_{\text{CRES}} = 12$, $n_{\text{LVNS}} = 6$, $n_{\text{LVSM}} = 3$

In this section, we present several case studies, for the three data sets, to demonstrate the capabilities of the proposed methodology. We followed the procedure as discussed in Section II, and observed that the average daily demand for the different home grouping sizes tends toward the population mean as the aggregation scale increases. This is true for all the three data sets. The CREST, LVNS and LVSM settled off at about 0.52kW, 0.41kW and 0.83kW, respectively. It was also observed that the mean is the same for all sampling resolution scales across all aggregation scales, as expected. The following
sections discuss how the ADMD and variance values differ for each of the three data sets.

A. After-diversity Maximum Demand

The time-coincident maximum demand was found to decrease as the home grouping size increases. As expected, the highest ADMD was recorded for an aggregation scale of one home for all the data sets, as illustrated in Fig. 2 for CREST. The ADMD then continues to drop before it starts to stabilize at a settling ADMD level when the aggregation level continues to increase. A settling value between 1kW and 2kW was recorded for all the data sets. This observation corresponds with the findings reported in [15].

![ADMD variations versus aggregation scales at different sampling resolutions](image1.png)

Fig. 2. CREST: The ADMD variations versus aggregation scales at different sampling resolutions

In Fig. 3, we show how the ADMD level changes versus both the aggregation level and sampling resolutions, and it can be observed that higher ADMD values were obtained at higher sampling resolutions. This is more evident for CREST and LVNS data when compared to the LVSM data. Note that the LVSM data was collected at 15 minutes sampling interval.

![A summary of how the ADMD changes when loads of different home grouping sizes are aggregated at different sampling resolutions](image2.png)

Fig. 3. A summary of how the ADMD changes when loads of different home grouping sizes are aggregated at different sampling resolutions

B. Variance

The highest variance was obtained for an aggregation level of one home grouping. As it can be seen in Fig. 4 for CREST, the variance decreases as the level of aggregation increases for all the data sets. This observation is true for all the data sets as depicted in Fig. 5. The CREST data (originally 1 minute resolution) exhibits higher variance levels when compared to the LVNS data (originally 5 minutes resolution). The variance stabilizes at a settling value for all the data sets. The settling value for CREST is the highest (0.07kW²), whereas the settling value for LVSM is the lowest (0.04kW²).
The effect of aggregation and sampling resolution scales on the variance is summarized in Fig. 5. The variance decreases as the sampling resolution increases.

IV. CONCLUDING REMARKS

In this paper, we have investigated the relationship between system load diversity and variability at different sampling time and aggregation scales. The analysis is relevant for understanding both capacity and operational requirements when interconnecting different sized groupings of residential customers. The per-customer capacity requirement, and load variability of a network decreases as the size of customer groupings increase. However, both capacity requirements and load variability tend to saturate at a finite level as the size of customer groupings go to infinity. For active network management, faster time scales are required at lower aggregation scales due to high load variability.

A promising area for future work would be to apply the methods developed for analysis in this paper to determine the optimal placement and sizing of distributed generation and energy storage systems, and to determine the optimal control time scales for active energy management of the network.

ACKNOWLEDGMENT

We are grateful to the University of Mons, Belgium, for granting permission to use their data in this study.

REFERENCES


