Definitions of Demand Flexibility for Aggregate Residential Loads

Intisar Ali Sajjad, *Member, IEEE*, Gianfranco Chicco, *Senior Member, IEEE*, and Roberto Napoli, *Member, IEEE*

***Abstract*—Nowadays, enhanced knowledge of the nature of the electricity demand is achieved through the progressively increasing deployment of smart meters and advanced data analysis techniques. One of the major challenges is to exploit this knowledge to support the introduction of strategies to modify the demand according to relevant objectives to be achieved, like users' participation in demand response programmes. A key point for facing this challenge is to characterize the demand flexibility. In spite of many discussions about the concept of flexibility, the few mathematical definitions of flexibility available do not address the variation in time of the overall demand aggregation. This paper starts from the analysis of time-variable patterns of aggregate residential customers, ending up with suitable definitions of expected flexibility for aggregate demand. These definitions are based on assessing positive and negative pattern variations and are identified from the analysis of the collective behavior of the aggregate users. A set of results is shown for different numbers of aggregate customers, by considering different values of the averaging time step for load pattern representation.**

***Index Terms*—Aggregate demand, binomial probability, customers, electrical load, demand flexibility, demand response, load variation pattern, maximum likelihood estimation.**

NOMENCLATURE

ADT	Acceptable Delay Time
AFI	Appliance Flexibility Index
CDF	Cumulative Distribution Function
CI	Confidence Interval
DR	Demand Response
FIAD	Flexibility Index of Aggregate Demand
MLE	Maximum Likelihood Estimation
PFL	Percentage Flexibility Level
RES	Renewable Energy Sources
a	Number of aggregate customers
k	Observation number
n_s	Total number of points in the load pattern data with time step duration Δt_s
$p^{(a)}_{k,x\Delta t_s}$	Aggregate demand at time instant $x\Delta t_s$ for observation number k
$\bar{p}^{(a)}_{x\Delta t_s}$	Average value of $\boldsymbol{p}^{(a)}_{x\Delta t_s}$
$u^{(a)}_{k,x\Delta t_s}$	Outcome of a Bernoulli trial
T	Time interval of observation (minutes)
$Z_{\alpha/2}$	Critical value of the normal distribution at significance level α
$\Delta p^{(a)}_{k,x\Delta t_s}$	Change in demand at time instant $x\Delta t_s$ between two successive time intervals
$\overline{\Delta p}^{(a)}_{x\Delta t_s}$	Average value of $\Delta \boldsymbol{p}^{(a)}_{x\Delta t_s}$
$^{+}\overline{\Delta p}^{(a)}_{x\Delta t_s}, {}^{-}\overline{\Delta p}^{(a)}_{x\Delta t_s}$	Mean load variations for increasing demand and non-increasing demand, respectively
Δt_s	Time step duration (minutes)
α	Defines CI width, i.e., $100(1-\alpha)\%$
$\omega^{(a)}_{x\Delta t_s}$	Probability of binomial discrete random variable
$\underline{\hat{\omega}^{(a)}_{x\Delta t_s}}$	Estimated value of $\omega^{(a)}_{x\Delta t_s}$ using MLE
$\hat{\omega}^{(a)}_{x\Delta t_s}, \underline{\hat{\omega}^{(a)}_{x\Delta t_s}}$	Upper and lower bounds of the Wilson Score Interval
$\hat{\omega}'^{(a)}_{x\Delta t_s}, \sigma'^{(a)}_{x\Delta t_s}$	Relocated mean and standard deviation for the Wilson Score Interval
$\boldsymbol{p}^{(a)}_{x\Delta t_s}$	Vector of aggregate demand for all k at time instant $x\Delta t_s$
$\bar{\boldsymbol{p}}^{(a)}_{\Delta t_s}$	Mean aggregate demand pattern
$\mathbf{u}^{(a)}_{x\Delta t_s}$	Binomial discrete random variable
$\Delta \boldsymbol{p}^{(a)}_{x\Delta t_s}$	Vector of change in demand for all k at time instant $x\Delta t_s$
$\boldsymbol{\psi}^{(a)}_{\Delta t_{s\%}}$	Percentage flexibility level (*PFL*)
$\boldsymbol{\varphi}^{(a)}_{\Delta t_s}$	Flexibility index of aggregate demand (*FIAD*)
$\overline{\boldsymbol{\varphi}^{(a)}_{\Delta t_s}}, \underline{\boldsymbol{\varphi}^{(a)}_{\Delta t_s}}$	Upper and lower bounds of *FIAD*.

Manuscript received December 11, 2014; revised March 18, 2015, June 28, 2015, and October 21, 2015; accepted January 25, 2016. Date of publication April 21, 2016; date of current version October 19, 2016. This work was supported in part by the European Union Seventh Framework Programme FP7/2007-2013 under Grant 309048, and in part by the Project Smart and Sustainable Insular Electricity Grids Under Large-Scale Renewable Integration (SiNGULAR). Paper no. TSG-01216-2014.

The authors are with the Power and Energy Systems Group, Energy Department, Politecnico di Torino, Turin 10129, Italy (e-mail: malik.sajjad@polito.it; gianfranco.chicco@polito.it; roberto.napoli@polito.it).

Digital Object Identifier 10.1109/TSG.2016.2522961

I. INTRODUCTION

THE INTEGRATION of renewable energy sources (RES) with intermittent nature in the electric power grid tends to introduce mismatch between demand and supply. Using different means, e.g., conventional generation and demand response (DR) resources including storage, can level this mismatch [1]. The use of conventional plants has different drawbacks over the DR alternatives, such as environmental effects, high generation costs and high ramp rates. On the other hand, the uncertainty associated with the operating points of conventional generation is lower. The DR resources have fast response time, but they are typically small and need to be aggregated to balance the mismatch. In addition, DR uncertainties are hard to model due to different factors like customers' behaviour, weather conditions, etc. Furthermore, the coordination of DR resources is challenging due to the lack of two-way communication with each individual load. These are some important challenges being addressed in different perspectives in the literature.

The current terminology has adopted the term *flexibility* to indicate the *capacity to adapt* across time, circumstances (foreseeable or not), intention (positive or negative reactions) and area of application [2]. For the applications to the electrical system, *flexibility* refers to the possibility of deploying the available resources to respond in an adequate and reliable way to the load and generation variations during time at acceptable costs.

One of the current challenges is to define and quantify flexibility in specific contexts. A number of recent contributions deal with obtaining flexibility from the generation side. Some examples include addressing generation variability through the insufficient ramping resource expectation metric [3] and quantifying the technical flexibility level of both individual generators and the whole generation system [4]. In other cases, both generation and loads are taken into account, e.g., quantifying operational flexibility by using power capacity, energy capacity and ramp-rate capacity [5], or applying the unit commitment optimization approach to compare flexibility from demand-side resources with the one from fast ramping generation [6]. Moreover, the decentralized participation of flexible demand from heat pumps and electric vehicles is addressed in [7], the balancing in time of heat and power demand in multiple areas in [8], and the use of a stochastic unit commitment model accounting for RES fluctuations and DR benefits to absorb these fluctuations in [9].

On the demand side, the definitions of flexibility depend on evaluations carried out at the level of individual appliances or for a load aggregation.

For *individual* appliances, definitions from the current literature include the consumers' Acceptable Delay Time (*ADT*) [10], that is, the maximum period of time to postpone the operation of an appliance without sacrificing the consumers' comfort, and the Appliance Flexibility Index (*AFI*) [11] measuring the adjustable range of time of the appliances. In both cases, the data needed to calculate these indices depend on the consumers' preferences and are gathered from questionnaires and surveys. The *ADT* is also used in [12] together with the penetration level of active consumers, depending on the number of active consumers (i.e., those able to modify the operation of controllable appliances on the basis of a signal received from the service provider) with respect to the total number of consumers. The model presented in [13] addresses delay-averse flexible loads by introducing cost-delay trade-offs and assessing the value of time flexibility.

For the *aggregate* load, various approaches have been followed, among which the use of sensitivity functions indicating each user's probability of shifting each device type usage by a certain time, given the reward in the new period of usage [14], an agent-based approach based on the Q-learning algorithm, obtaining flexibility factors used to simulate demand elasticity [15], and the application of flexibility criteria to partition the types of loads into sheddable, controllable and acceptable, in order to assess the total DR resource potential [4]. Furthermore, an approach with identification of the flexible loads and an optimal load control strategy based on a reference demand profile is presented in [16]. Other contributions specifically address the aggregate flexibility of thermostatically-controlled loads, to represent the dynamics in the collective response [17], the model and control of a comfort-constrained virtual generator [18], and the characterization of the load aggregation with a generalized battery model [19]. A recent work [20] refers to a population of appliances and introduces *load plasticity* as the potential of the load pattern of an appliance to be modified by control actions.

None of the above references address the quantification of the flexible amount of the aggregate demand by investigating on the uncertainty of the time-variable shape of the demand patterns belonging to residential customer groups. In particular, a practical formulation of flexibility depending on the collective behaviour of a population of consumers has not been provided yet. Introducing proper expressions of the aggregate demand flexibility is a key aspect to effectively assess the contribution of responsive demand, enabling better management of DR resources and most effective utilization of RES [1]. The uncertainty associated with the aggregate demand flexibility also depends on the aggregation size and the averaging time step with which the evolution in time of the average power is represented [21]–[23].

This paper provides new mathematical *definitions* of *flexibility* for an *aggregate demand*. These definitions are based on the statistical properties of the aggregate demand variations and originate from some studies carried out on the time-variable patterns of a number of aggregate residential customers. The rationale of the new flexibility definitions is based on assessing the positive and negative variations occurring during time in the aggregate load pattern. Flexibility is identified in terms of probability to change the collective behaviour of the aggregate users. The definitions are presented with reference to residential aggregate patterns, but may be generally valid for different types of load aggregations in which the individuals have a relatively similar size, providing a comparable impact on the aggregate pattern. The available data are the measurable load patterns characterizing the customers and their *aggregation* seen from their grid connection terminals. In this context, the information on the appliances located inside the households is not accessible, and the inclusion of entries referring to the real-time control of the specific appliances is not applicable.

The next sections of this paper are organized as follows. Section II illustrates the structure of the data used for the analysis. Section III introduces the categorical data analysis approach used to establish the statistical properties of the data set. Section IV defines the proposed flexibility indicators and illustrates different case study applications from the extra-urban residential sector with different aggregation levels and time step durations, to demonstrate the effectiveness of the proposed definitions. Section V provides further results

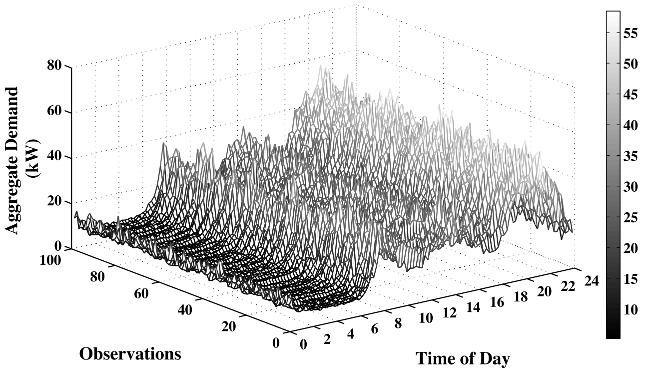


Fig. 1. Daily aggregate demand patterns for a group of 50 houses and time step 15 min, with 100 observations.

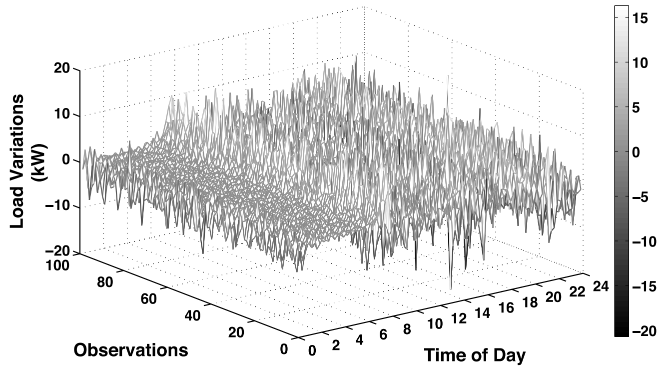


Fig. 2. Time evolution of the daily aggregate demand variations for 50 houses and time step 15 min, with 100 observations.

showing the variations of the flexibility indicators for different aggregations of users and duration of the time step used to represent the load patterns. The last section contains the concluding remarks.

II. Data Organization for Residential Customers

Measuring devices like smart meters log data about consumption in a discrete fashion with a given time step duration Δt_s. Let us consider a time interval T multiple of Δt_s, in which the total number of time steps is $n_s = T/\Delta t_s$.

Let us then consider a group of customers consisting of the aggregation of a number a of individual customers, with the number a identifying the customer aggregation level. For a given aggregation level and time step duration, let us assume that K observations are available, for example obtained from the Monte Carlo repetitions calculated on the basis of a bottom-up statistical model of aggregate residential loads [24]. Each observation provides an aggregate load pattern containing the sequence of average power values calculated for each successive time step.

The load pattern data used in this paper have been generated for extra-urban residential consumers by using Monte Carlo simulations, on the basis of information about the family composition and lifestyle, house characterization, usage of electrical appliances inside each type of house, directly collected from the residents [25], [26].

As an example, for the aggregation of $a = 50$ houses, Fig. 1 shows the daily aggregate demand patterns for a typical winter weekday ($T = 1440$ min) resulting from $K = 100$ Monte Carlo repetitions with time step $\Delta t_s = 10$ min.

For customer aggregation level a and time step duration Δt_s, the load pattern data is organized in matrix form as follows:

$$\boldsymbol{P}_{\Delta t_s}^{(a)} = \begin{bmatrix} p_{1,1\Delta t_s}^{(a)} & p_{1,2\Delta t_s}^{(a)} & \cdots & p_{1,n_s\Delta t_s}^{(a)} \\ p_{2,1\Delta t_s}^{(a)} & p_{2,2\Delta t_s}^{(a)} & \cdots & p_{2,n_s\Delta t_s}^{(a)} \\ \vdots & \vdots & \vdots & \vdots \\ p_{K,1\Delta t_s}^{(a)} & p_{K,2\Delta t_s}^{(a)} & \cdots & p_{K,n_s\Delta t_s}^{(a)} \end{bmatrix} \in \mathbb{R}^{K,n_s} \tag{1}$$

We further consider the load variations referring to load increase or decrease from one time step to the next one. For customer aggregation level a and time step duration Δt_s, the k^{th} load variation pattern, for $k = 1, \ldots, K$, is represented as:

$$\Delta p_{k,x\Delta t_s}^{(a)} = p_{k,x\Delta t_s}^{(a)} - p_{k,(x-1)\Delta t_s}^{(a)} \text{ for } x = 2, 3, \ldots, n_s \tag{2}$$

The load variation patterns are included in the rows of the matrix

$$\Delta\boldsymbol{P}_{\Delta t_s}^{(a)} = \begin{bmatrix} \Delta p_{1,2\Delta t_s}^{(a)} & \Delta p_{1,3\Delta t_s}^{(a)} & \cdots & \Delta p_{1,n_s\Delta t_s}^{(a)} \\ \Delta p_{2,2\Delta t_s}^{(a)} & \Delta p_{2,3\Delta t_s}^{(a)} & \cdots & \Delta p_{2,n_s\Delta t_s}^{(a)} \\ \vdots & \vdots & \vdots & \vdots \\ \Delta p_{K,2\Delta t_s}^{(a)} & \Delta p_{K,3\Delta t_s}^{(a)} & \cdots & \Delta p_{K,n_s\Delta t_s}^{(a)} \end{bmatrix} \in \mathbb{R}^{K,n_s-1} \tag{3}$$

Fig. 2 shows the time evolution of the demand variation during a day, for the aggregation of 50 houses and time step of 15 minutes indicated in Fig. 1.

Each column of (3) is a set of observations for a particular time step and is calculated using (2). We can represent the demand and its variation in terms of column vectors as:

$$\boldsymbol{P}_{\Delta t_s}^{(a)} = \begin{bmatrix} \boldsymbol{p}_{1\Delta t_s}^{(a)} & \boldsymbol{p}_{2\Delta t_s}^{(a)} & \cdots & \boldsymbol{p}_{n_s\Delta t_s}^{(a)} \end{bmatrix} \tag{4}$$

$$\Delta\boldsymbol{P}_{\Delta t_s}^{(a)} = \begin{bmatrix} \Delta\boldsymbol{p}_{2\Delta t_s}^{(a)} & \Delta\boldsymbol{p}_{3\Delta t_s}^{(a)} & \cdots & \Delta\boldsymbol{p}_{n_s\Delta t_s}^{(a)} \end{bmatrix}. \tag{5}$$

III. Categorical Data Analysis

Different methods are used in statistics to analyse data with different characteristics. Categorical data analysis is one of the statistical approaches to analyse data for data clustering, correlation analysis, system modelling etc., and has wide applications in different fields of science and technology [27]–[29]. The British statistician Pearson worked in this field around 1900, then very little development was noticed until 1960's. From 1960 till now, several studies have been done for method development related to categorical data analysis [30].

For the particular problem addressed in this paper, we consider that load variations may be positive or negative.[1]

A. Binomial Representation

There are three basic probability distributions used for categorical data analysis, i.e., Binomial, Poisson and Multinomial distributions. The binomial distribution is the special case of multinomial distribution with only two categories. The demand variations are modelled here using the binomial distribution, with two response variables:

1. Increasing demand.
2. Non-increasing demand, including demand decrease and stationary demand.

Each observation is a Bernoulli trial with only two outcomes. The total number of observations is fixed to K. Let us define the outcome of each Bernoulli trial for $k = 1, 2, \ldots, K$ at the particular time step $x\Delta t_s$ as:

$$u^{(a)}_{k,x\Delta t_s} = \begin{cases} 1, & \Delta p_{k,x\Delta t_s} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

For a particular aggregation level a and time step $x\Delta t_s$ with $x = 2, 3, \ldots, n_s$, let $\mathbf{u}^{(a)}_{x\Delta t_s}$ be a binomial discrete random variable, defined as:

$$\mathbf{u}^{(a)}_{x\Delta t_s} \sim Bin\left(K, \omega^{(a)}_{x\Delta t_s}\right) \tag{7}$$

$$\mathbf{u}^{(a)}_{x\Delta t_s} = \sum_{k=1}^{K} u^{(a)}_{k,x\Delta t_s} \text{ with } \mathbf{u}^{(a)}_{x\Delta t_s} \in \{0, 1, 2, \ldots, K\} \tag{8}$$

$$\omega^{(a)}_{x\Delta t_s} = prob\left(\mathbf{u} = \mathbf{u}^{(a)}_{x\Delta t_s}\right) \tag{9}$$

For example, if there are $K = 100$ observations for aggregation level a and time step $x\Delta t_s$, then $\omega^{(a)}_{x\Delta t_s}$ is the probability to get load increase $\Delta p_{k,x\Delta t_s} > 0$ for $\mathbf{u}^{(a)}_{x\Delta t_s}$ times.

B. Maximum Likelihood Estimation

For the binomial model presented in Section III-A, the probability $\omega^{(a)}_{x\Delta t_s}$ is unknown and needs to be determined using some suitable estimation technique. The Maximum likelihood estimation (MLE) method [31], [32] is used for this purpose, with the formulation indicated in [33] and recalled here in the Appendix.

The result of MLE becomes:

$$\hat{\omega}^{(a)}_{x\Delta t_s} = \mathbf{u}^{(a)}_{x\Delta t_s} \big/ K \tag{10}$$

The term $\hat{\omega}^{(a)}_{x\Delta t_s}$ is computable because all the parameters are known, and is also an unbiased estimator because the expected value is $E(\hat{\omega}^{(a)}_{x\Delta t_s}) = \omega^{(a)}_{x\Delta t_s}$ and the variance is $Var(\hat{\omega}^{(a)}_{x\Delta t_s}) = \omega^{(a)}_{x\Delta t_s}(1 - \omega^{(a)}_{x\Delta t_s})/K$.

[1] A third possibility is to have no load variation, meaning that the possible load variation in two consecutive demand values is lower than the amplitude resolution of the meter. This is more likely to occur when the meter resolution is relatively poor. This possibility is excluded in this paper, leading to the binomial representation introduced in Section III.A. In order to guarantee generality and preserve the nature of the binomial model, possible situations with no demand variation can be associated with either positive or negative variations. The latter case is used in this paper.

Using the results of MLE in (10) we can write, for all the numbers of points in the load pattern data with time step duration Δt_s, the probability vector in the following form:

$$\hat{\boldsymbol{\omega}}^{(a)}_{\Delta t_s} = \begin{bmatrix} \hat{\omega}^{(a)}_{2\Delta t_s} & \hat{\omega}^{(a)}_{3\Delta t_s} & \cdots & \hat{\omega}^{(a)}_{n_s\Delta t_s} \end{bmatrix} \in \mathbb{R}^{n_s - 1}. \tag{11}$$

C. Confidence Interval for Binomial Proportions

The entry $\hat{\omega}^{(a)}_{x\Delta t_s}$ is directly related to $\mathbf{u}^{(a)}_{x\Delta t_s}$ and is calculated based on the outcomes of each trial, $u_{k,x\Delta t_s}$. The probability of success is the same for each trial, and the trials are statistically independent of each other. If in one experiment $\omega_{x\Delta t_s}$ is equal to 0.6, it may happen that in a second experiment for the same environment this may be 0.61. We cannot predict the binomial parameters with 100% accuracy, because the calculations are not based on the whole population. For this reason, different methods are formulated to find the confidence intervals (CIs) for binomial parameters.

The CIs are very informative, because they indicate the level of uncertainty or randomness of the load increase or decrease. If in a given time period one scenario has lower CI in comparison with another, it means that the former scenario has a more regular trend about increasing or decreasing the load in that time period.

There are many established methods in the literature to calculate CIs [33]. The most simple and basic method is normal approximation using central limit theorem [34], [35]. This approximation fails when the trial entries are too low or $\hat{\omega}^{(a)}_{x\Delta t_s}$ is very close to 0 or 1. These bottlenecks were addressed by Bidwell Wilson who developed the Wilson score interval in 1927 [36]. For this method, the actual coverage probability of confidence interval is approximately equal to the nominal one, even for small number of trials or $\hat{\omega}^{(a)}_{x\Delta t_s}$ closer or equal to 0.1. This method has advantages in terms of good average coverage probability, less average expected length and smaller mean absolute error [33], [35], [37]. Using the modified version of the Wilson Score Interval method described in [35], the upper limit $\overline{\hat{\omega}^{(a)}_{x\Delta t_s}}$ and the lower limit $\underline{\hat{\omega}^{(a)}_{x\Delta t_s}}$ for our problem are calculated as:

$$\left(\overline{\hat{\omega}^{(a)}_{x\Delta t_s}}, \underline{\hat{\omega}^{(a)}_{x\Delta t_s}}\right) = \hat{\omega}'^{(a)}_{x\Delta t_s} \pm Z_{\alpha/2}.\sigma'^{(a)}_{x\Delta t_s} \tag{12}$$

where

$$\hat{\omega}'^{(a)}_{x\Delta t_s} = \left(\hat{\omega}^{(a)}_{x\Delta t_s} + \frac{Z^2_{\alpha/2}}{2K}\right) \Big/ \left(1 + \frac{Z^2_{\alpha/2}}{K}\right) \tag{13}$$

$$\sigma'^{(a)}_{x\Delta t_s} = \frac{\sqrt{\dfrac{\hat{\omega}^{(a)}_{x\Delta t_s}\left(1 - \hat{\omega}^{(a)}_{x\Delta t_s}\right)}{K} + \frac{Z^2_{\alpha/2}}{4K^2}}}{\left(1 + \dfrac{Z^2_{\alpha/2}}{K}\right)} \tag{14}$$

By using the results of Eq. (12) to (14), we can rewrite (11) as follows:

$$\hat{\boldsymbol{\omega}}'^{(a)}_{\Delta t_s} = \begin{bmatrix} \hat{\omega}'^{(a)}_{2\Delta t_s} & \hat{\omega}'^{(a)}_{3\Delta t_s} & \cdots & \hat{\omega}'^{(a)}_{n_s\Delta t_s} \end{bmatrix} \in \mathbb{R}^{n_s - 1} \tag{15}$$

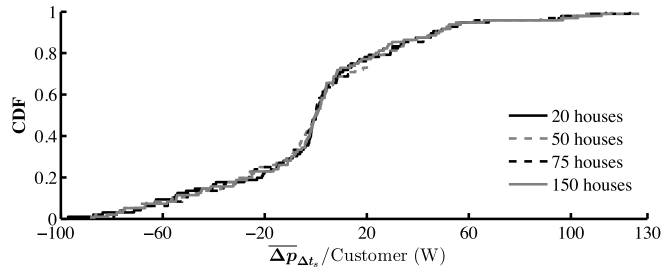


Fig. 3. Mean demand variations per customer for different aggregation levels.

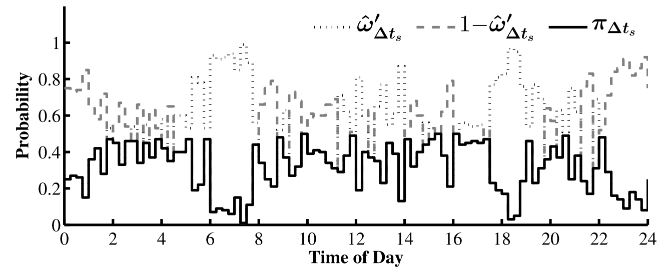


Fig. 4. Comparison of, $\hat{\boldsymbol{\omega}}'^{(a)}_{\Delta t_s}$, $1-\hat{\boldsymbol{\omega}}'^{(a)}_{\Delta t_s}$ and $\boldsymbol{\pi}^{(a)}_{\Delta t_s}$ with $a = 50$ houses and $\Delta t_s = 15$ minutes.

The vectors containing the upper and lower limits are, respectively:

$$\overline{\hat{\boldsymbol{\omega}}^{(a)}_{\Delta t_s}} = \left[\, \overline{\hat{\omega}^{(a)}_{2\Delta t_s}} \;\; \overline{\hat{\omega}^{(a)}_{3\Delta t_s}} \;\; \cdots \;\; \overline{\hat{\omega}^{(a)}_{n_s\Delta t_s}} \,\right] \in \mathbb{R}^{n_s-1} \tag{16}$$

$$\underline{\hat{\boldsymbol{\omega}}^{(a)}_{\Delta t_s}} = \left[\, \underline{\hat{\omega}^{(a)}_{2\Delta t_s}} \;\; \underline{\hat{\omega}^{(a)}_{3\Delta t_s}} \;\; \cdots \;\; \underline{\hat{\omega}^{(a)}_{n_s\Delta t_s}} \,\right] \in \mathbb{R}^{n_s-1} \tag{17}$$

The binomial probabilities calculated in this section are used in the next section to define the demand flexibility indicators.

IV. Definition of Demand Flexibility Indicators

A. Conceptual Deduction of the Demand Flexibility Index

Let $\bar{p}^{(a)}_{x\Delta t_s}$ and $\overline{\Delta p}^{(a)}_{x\Delta t_s}$ be the mean values of $\boldsymbol{p}^{(a)}_{x\Delta t_s}$ and $\Delta \boldsymbol{p}^{(a)}_{x\Delta t_s}$, respectively. Then, from the basic definition of load variations presented in Section III, we can rewrite (2) as:

$$\bar{p}^{(a)}_{x\Delta t_s} = \bar{p}^{(a)}_{(x-1)\Delta t_s} + \overline{\Delta p}^{(a)}_{x\Delta t_s} \tag{18}$$

The load variations for increase and decrease are separated for each $\Delta \boldsymbol{p}^{(a)}_{x\Delta t_s}$. Let ${}^{+}\overline{\Delta p}^{(a)}_{x\Delta t_s}$ and ${}^{-}\overline{\Delta p}^{(a)}_{x\Delta t_s}$ be the mean values of the load variations for increase in demand and non-increase in demand. Then, from the definition of weighted arithmetic mean, we can express $\overline{\Delta p}^{(a)}_{x\Delta t_s}$ in terms of the mean values of load increase or decrease:

$$\overline{\Delta p}^{(a)}_{x\Delta t_s} = \left(\frac{\mathbf{u}^{(a)}_{x\Delta t_s}}{K}\right) {}^{+}\overline{\Delta p}^{(a)}_{x\Delta t_s} + \left(\frac{1-\mathbf{u}^{(a)}_{x\Delta t_s}}{K}\right) {}^{-}\overline{\Delta p}^{(a)}_{x\Delta t_s} \tag{19}$$

By definition, from the MLE of binomial proportions presented in (10) and (13), $\mathbf{u}^{(a)}_{x\Delta t_s}/K$ is the estimated binomial probability of increase in demand $\hat{\omega}'^{(a)}_{x\Delta t_s}$. Eq. (19) can be rewritten as:

$$\overline{\Delta p}^{(a)}_{x\Delta t_s} = \left(\hat{\omega}'^{(a)}_{x\Delta t_s}\right) {}^{+}\overline{\Delta p}^{(a)}_{x\Delta t_s} + \left(1-\hat{\omega}'^{(a)}_{x\Delta t_s}\right) {}^{-}\overline{\Delta p}^{(a)}_{x\Delta t_s} \tag{20}$$

where $(1-\hat{\omega}'^{(a)}_{x\Delta t_s})$ is the binomial probability of non-increase in demand.

Let us now make the assumption that the mean behaviour for demand variations (increase and decrease) does not change with respect to $\hat{\omega}'^{(a)}_{x\Delta t_s}$. Then, $\overline{\Delta p}^{(a)}_{x\Delta t_s}$ only depends on $\hat{\omega}'^{(a)}_{x\Delta t_s}$. To support this assumption, Fig. 3 shows that the cumulative distribution function (CDF) of the mean demand variations (divided by the number of customers) is very similar for different sizes of customer aggregations.

The maximum and minimum ranges of variation of $\overline{\Delta p}^{(a)}_{x\Delta t_s}$ determine the flexibility margins of the actual demand $\bar{p}^{(a)}_{x\Delta t_s}$ at any time step $x\Delta t_s$. The second part of (20) is always negative or equal to zero, because the mean load variation for the binomial category of non-increase in demand is either negative or equal to zero. The actual mean demand $\bar{p}^{(a)}_{x\Delta t_s}$ can be reduced to its minimum possible value when $\hat{\omega}'^{(a)}_{x\Delta t_s}$ is zero and conversely can be increased at its maximum value when $\hat{\omega}'^{(a)}_{x\Delta t_s}$ is equal to 1.

Let $\bar{p}^{(a)*}_{x\Delta t_s}$ be the mean aggregate demand with binomial probability of increase in demand $\hat{\omega}'^{(a)*}_{x\Delta t_s}$. Then, the amount of flexible mean aggregate demand can be calculated by using (18) to (20) as:

$$\bar{p}^{(a)}_{x\Delta t_s} - \bar{p}^{(a)*}_{x\Delta t_s} = \left(\hat{\omega}'^{(a)}_{x\Delta t_s} - \hat{\omega}'^{(a)*}_{x\Delta t_s}\right) {}^{+}\overline{\Delta p}^{(a)}_{x\Delta t_s} + \left(\hat{\omega}'^{(a)*}_{x\Delta t_s} - \hat{\omega}'^{(a)}_{x\Delta t_s}\right) {}^{-}\overline{\Delta p}^{(a)}_{x\Delta t_s} \tag{21}$$

Let us now calculate the minimum between the current probability of demand increase and the complementary probability, in vector terms:

$$\boldsymbol{\pi}^{(a)}_{\Delta t_s} = \min_{\forall \hat{\omega}'^{(a)}_{x\Delta t_s}} \left(\hat{\boldsymbol{\omega}}'^{(a)}_{\Delta t_s}, 1-\hat{\boldsymbol{\omega}}'^{(a)}_{\Delta t_s}\right) \tag{22}$$

By definition, the value of each entry $\pi^{(a)}_{x\Delta t_s}$, for $x = 2, 3, \ldots, n_s$, belongs to the range [0, 0.5]. In fact, the minimum value of the complementary entries $\hat{\omega}'^{(a)}_{x\Delta t_s}$ and $(1-\hat{\omega}'^{(a)}_{x\Delta t_s})$ is equal to 0.5 for $\hat{\omega}'^{(a)}_{x\Delta t_s} = 0.5$, and is equal to zero when $\hat{\omega}'^{(a)}_{x\Delta t_s} = 0$ or $\hat{\omega}'^{(a)}_{x\Delta t_s} = 1$. In order to obtain the formulation of the proposed *flexibility index of aggregate demand* (*FIAD*) (denoted in vector form as $\boldsymbol{\varphi}^{(a)}_{\Delta t_s}$) whose entries are defined in the more intuitive range [0,1] in line with the probabilistic limits, the probabilities $\boldsymbol{\pi}^{(a)}_{\Delta t_s}$ are multiplied it by 2 (that is, the number of categories of the binomial probability distribution in this case), such that:

$$FIAD = \boldsymbol{\varphi}^{(a)}_{\Delta t_s} = 2 \times \boldsymbol{\pi}^{(a)}_{\Delta t_s}, \text{ with } \varphi^{(a)}_{x\Delta t_s} \in [0,1] \tag{23}$$

The binomial probabilities $\hat{\boldsymbol{\omega}}'^{(a)}_{\Delta t_s}$ and $1-\hat{\boldsymbol{\omega}}'^{(a)}_{\Delta t_s}$ with their minimum $\boldsymbol{\pi}^{(a)}_{\Delta t_s}$ for an aggregation of $a = 50$ houses and time step duration $\Delta t_s = 15$ minutes are shown in Fig. 4. It can be noted that $\boldsymbol{\pi}^{(a)}_{\Delta t_s}$ gives information about the possible probabilistic change to the nearest optimum (0 or 1) for each binomial category (demand increase or decrease). The entries $\pi^{(a)}_{x\Delta t_s}$ are by definition symmetrical with respect to the level 0.5. In other terms, any change in $\hat{\omega}'^{(a)}_{x\Delta t_s}$ determines opposite changes in $^{-}\overline{\Delta p}_{x\Delta t_s}^{(a)}$ and in $^{+}\overline{\Delta p}_{x\Delta t_s}^{(a)}$. The overall effect on changing the aggregate demand will be doubled, as $^{-}\overline{\Delta p}_{x\Delta t_s}^{(a)}$ is always negative or zero (Eq. (21)). For example, if $\hat{\omega}'^{(a)}_{x\Delta t_s} - \hat{\omega}'^{(a)*}_{x\Delta t_s}$ is equal to 0.2, the second term of (21) is positive because both $(\hat{\omega}'^{(a)*}_{x\Delta t_s} - \hat{\omega}'^{(a)}_{x\Delta t_s})$ and $^{-}\overline{\Delta p}_{x\Delta t_s}^{(a)}$ are negative; in this case, the overall effect will be 0.2 times the first term plus 0.2 times the second term. This is also true for the opposite case and provides a further justification to use the multiplier 2 in (23).

From Fig. 4, the region between $\boldsymbol{\pi}_{\Delta t_s}^{(a)}$ and 0 probability, and the region between $max(\hat{\boldsymbol{\omega}}'^{(a)}_{\Delta t_s}, 1 - \hat{\boldsymbol{\omega}}'^{(a)}_{\Delta t_s})$ and 1 probability are the regions in which flexibility of aggregate demand does exist.

From (23), if the value of $\varphi_{x\Delta t_s}^{(a)}$ is very close to 1, then in terms of load variations we can say that most of the time there is almost an equal probability of increase or decrease in demand. This equal probability can be due to two reasons. Firstly, if $\varphi_{x\Delta t_s}^{(a)}$ is very close to 1 and the aggregate mean load variations $^{+}\overline{\Delta p}_{x\Delta t_s}^{(a)}$ and $^{-}\overline{\Delta p}_{x\Delta t_s}^{(a)}$ are very small, then $\bar{p}_{x\Delta t_s}^{(a)} - \bar{p}_{x\Delta t_s}^{(a)*}$ will be very small and consequently possible DR benefits will also be very small. Secondly, if one of the terms $^{+}\overline{\Delta p}_{x\Delta t_s}^{(a)}$ and $^{-}\overline{\Delta p}_{x\Delta t_s}^{(a)}$ (or both of them) are reasonably high and *FIAD* is very close to 1, then this information indicates that the individual customers behave very randomly in the corresponding time step, and hence there is a chance to get a reasonable amount of DR benefits. This information is very helpful for an operator or aggregator, to improve the economic operation of the system by managing supply and demand side flexibilities. Further implications on how to pass from probability values to the amount of flexible demand are explained in the following subsections.

B. Confidence Bounds for the FIAD

Since the binomial probabilities are calculated from a limited number of observations, there is an uncertainty associated with the binomial probabilities, as calculated in Section III-C. The calculated confidence limits also introduce uncertainty in the *FIAD*. This subsection explains the CIs that are associated with *FIAD*.

Let $\overline{\boldsymbol{\pi}_{\Delta t_s}^{(a)}}$ and $\underline{\boldsymbol{\pi}_{\Delta t_s}^{(a)}}$ be the upper and lower bounds of the CIs for the minimum selected in (22). Then, the confidence limits associated with the *FIAD*, $\boldsymbol{\varphi}_{\Delta t_s}^{(a)}$, can be calculated by using (24) and (25).

$$\overline{\boldsymbol{\varphi}_{\Delta t_s}^{(a)}} = 2 \times \min\left(0.5, \overline{\boldsymbol{\pi}_{\Delta t_s}^{(a)}}\right) \tag{24}$$

$$\underline{\boldsymbol{\varphi}_{\Delta t_s}^{(a)}} = 2 \times \underline{\boldsymbol{\pi}_{\Delta t_s}^{(a)}} \tag{25}$$

The reason for using the multiplier 2 in (24) and (25) is the same as the one mentioned for (23). The upper limits of one or more entries of $\overline{\boldsymbol{\pi}_{\Delta t_s}^{(a)}}$ may exceed 0.5, and in that case the corresponding lower bound of $\max(\hat{\boldsymbol{\omega}}'^{(a)}_{\Delta t_s}, 1 - \hat{\boldsymbol{\omega}}'^{(a)}_{\Delta t_s})$ becomes a minimum. This is the reason why 0.5 is used in (24) as the maximum possible upper limit. This kind of situation can be observed from Fig. 5.

At the time step corresponding with hour 9:00, the minimum $\pi_{\Delta t_s}^{(a)}$ is $1 - \hat{\omega}'^{(a)}_{\Delta t_s}$ but the lower bound of $\hat{\omega}'^{(a)}_{\Delta t_s}$ is lower than the

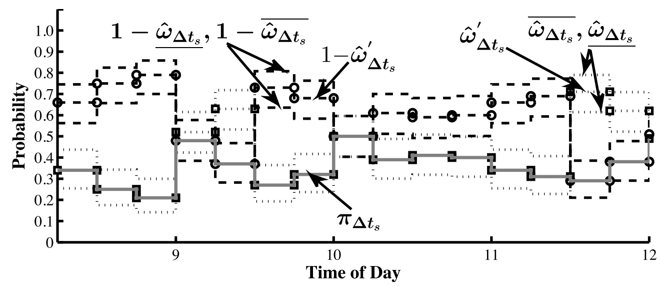


Fig. 5. Zoomed view of $\hat{\boldsymbol{\omega}}'^{(a)}_{\Delta t_s}$, $1 - \hat{\boldsymbol{\omega}}'^{(a)}_{\Delta t_s}$ and $\boldsymbol{\pi}_{\Delta t_s}^{(a)}$ with their confidence limits between hour 8:00 and hour 12:00.

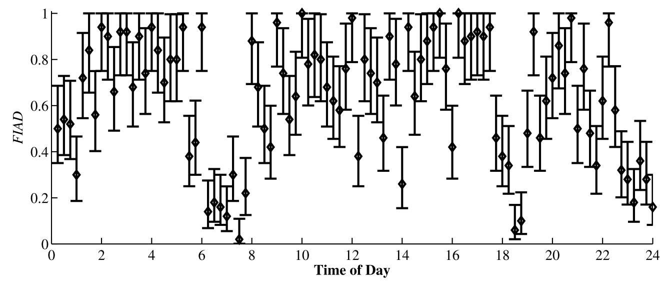


Fig. 6. *FIAD* with its confidence limits $(\overline{\boldsymbol{\varphi}_{\Delta t_s}^{(a)}}, \underline{\boldsymbol{\varphi}_{\Delta t_s}^{(a)}})$ with $a = 50$ houses and $\Delta t_s = 15$ minutes.

upper bound of $1 - \hat{\omega}'^{(a)}_{\Delta t_s}$, so there is a possibility that $\hat{\boldsymbol{\omega}}'^{(a)}_{\Delta t_s}$ may become a minimum. On the other hand, if $1 - \hat{\boldsymbol{\omega}}'^{(a)}_{\Delta t_s}$ or $\hat{\boldsymbol{\omega}}'^{(a)}_{\Delta t_s}$ is the minimum for (22), then its lower bound always holds.

The flexibility index *FIAD* with its confidence limits for an aggregation of $a = 50$ houses with $\Delta t_s = 15$ minutes is shown in Fig. 6. It can be noted that the lower *FIAD* values appear during the morning ramp-up and during the evening peak, that is, in the periods in which there is a consistent collective behaviour and the aggregate demand becomes more "rigid" to accept changes.

The above indications give a further input to interpret the *FIAD* indicator. The flexibility meant by *FIAD* is not a quantitative margin (expressed for instance in kW), but has a behavioural interpretation in terms of *collective trend* of the load aggregation. This definition incorporates both the possibility of increasing or decreasing the aggregate load, accepting variations on the basis of the confidence limits reported in Fig. 6.

For example, a *FIAD* number close to 100% means that in the corresponding time period the customers are behaving in a very random way, so that no collective trend emerges, and the flexibility to change is high because any external input to change behaviour could find the consumers "free" to accept changes without specific conditioning. Conversely, low flexibility values mean that the collective trend is biased enough to limit the possibility to induce changes in the collective consumer's behaviour.

C. Percentage Flexibility Level (PFL)

Let us now introduce a new indicator called percentage flexibility level (*PFL*) and denoted as $\boldsymbol{\psi}_{\Delta t_{s\%}}^{(a)}$, expressing the percentage of flexible demand associated with the demand flexibility index and defined as:

$$PFL = \boldsymbol{\psi}^{(a)}_{\Delta t_{s\%}} = \frac{{}^{+}\overline{\Delta \boldsymbol{p}}^{(a)}_{\Delta t_s} - {}^{-}\overline{\Delta \boldsymbol{p}}^{(a)}_{\Delta t_s}}{\bar{\boldsymbol{p}}^{(a)}_{\Delta t_s}} \left(\frac{\boldsymbol{\varphi}^{(a)}_{\Delta t_s}}{2} \right) \times 100 \qquad (26)$$

This indicator represents what percentage of the aggregate demand can be reduced or increased without affecting the average *change* in demand of the group of customers.

The *PFL* can be increased if all the customers change their behaviour from increase in demand towards decrease in demand or vice versa. This is very difficult when aggregate customers follow a trend (i.e., $\hat{\boldsymbol{\omega}}'^{(a)}_{\Delta t_s}$ is close to 0 or 1) and the probability to obtain a high *PFL* is very low. For example, if an entry of $\hat{\boldsymbol{\omega}}'^{(a)}_{\Delta t_s}$ is equal to 0.9 it is very difficult to turn it into 0; on the other hand, if it is equal to 0.1 it would be easier to reduce it to about 0. This has been also the reason to define flexibility as the minimum of the two binomial probabilities in (22).

The definition of *PFL* is applicable by taking into account increase and decrease in demand together. In order to obtain separate information for demand increase and decrease, it is possible to consider the maximum *PFL* for increase in aggregate demand (i.e., $\overline{\Delta p}^{(a)}_{x\Delta t_s} = {}^{+}\overline{\Delta p}^{(a)}_{x\Delta t_s}$) or decrease in aggregate demand (i.e., $\overline{\Delta p}^{(a)}_{x\Delta t_s} = {}^{-}\overline{\Delta p}^{(a)}_{x\Delta t_s}$, which can only be achieved with $\hat{\omega}'^{(a)*}_{x\Delta t_s} = 1$ or 0, respectively (see Eq. (20)). Otherwise, in order to achieve $\overline{\Delta p}^{(a)}_{x\Delta t_s} = {}^{+}\overline{\Delta p}^{(a)}_{x\Delta t_s}$ the probability $\hat{\omega}'^{(a)}_{x\Delta t_s}$ should be increased to 1, with the change equal to $1 - \hat{\omega}'^{(a)}_{x\Delta t_s}$. On the other hand, to achieve $\overline{\Delta p}^{(a)}_{x\Delta t_s} = {}^{-}\overline{\Delta p}^{(a)}_{x\Delta t_s}$ the probability $\hat{\omega}'^{(a)}_{x\Delta t_s}$ should decrease to 0. The change required is simply $\hat{\omega}'^{(a)}_{x\Delta t_s}$. The corresponding values of change in probability can replace $\boldsymbol{\varphi}^{(a)}_{\Delta t_s}/2$ in (26) to get maximum flexibility levels with respect to the binomial categories.

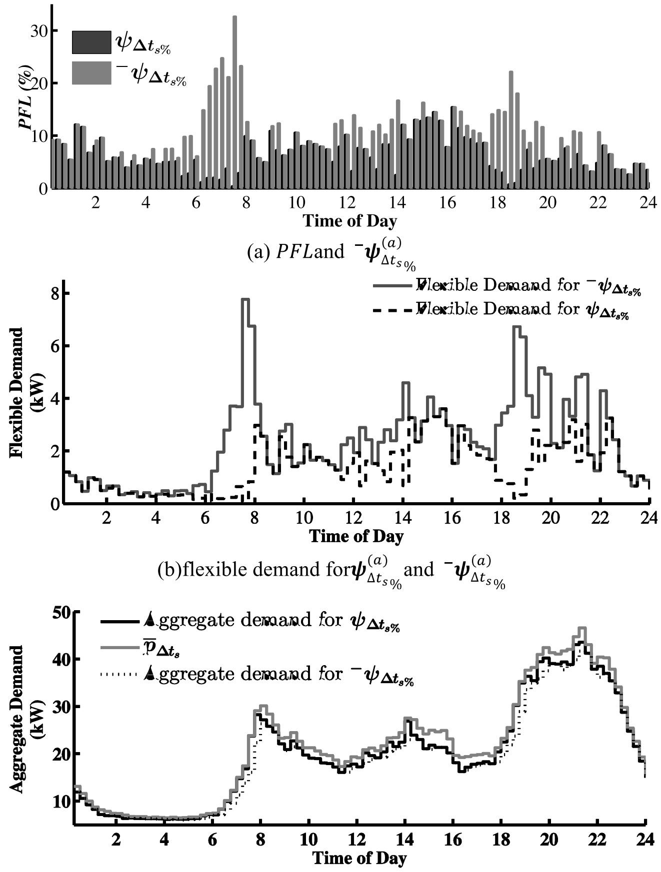
Let ${}^{+}\boldsymbol{\psi}^{(a)}_{\Delta t_{s\%}}$ and ${}^{-}\boldsymbol{\psi}^{(a)}_{\Delta t_{s\%}}$ be the maximum percentage flexibility levels for increase and decrease in aggregate demand, respectively. These indicators are defined as shown in (27) and (28).

$${}^{+}\boldsymbol{\psi}^{(a)}_{\Delta t_{s\%}} = \frac{{}^{+}\overline{\Delta \boldsymbol{p}}^{(a)}_{\Delta t_s} - {}^{-}\overline{\Delta \boldsymbol{p}}^{(a)}_{\Delta t_s}}{\bar{\boldsymbol{p}}^{(a)}_{\Delta t_s}} \left(1 - \hat{\boldsymbol{\omega}}'^{(a)}_{\Delta t_s} \right) \times 100 \qquad (27)$$

$${}^{-}\boldsymbol{\psi}^{(a)}_{\Delta t_{s\%}} = \frac{{}^{+}\overline{\Delta \boldsymbol{p}}^{(a)}_{\Delta t_s} - {}^{-}\overline{\Delta \boldsymbol{p}}^{(a)}_{\Delta t_s}}{\bar{\boldsymbol{p}}^{(a)}_{\Delta t_s}} \left(\hat{\boldsymbol{\omega}}'^{(a)}_{\Delta t_s} \right) \times 100 \qquad (28)$$

These indicators represent the maximum demand variation (in per cent) that may be obtained in the ideal case in which all the increasing demand changes to decreasing demand, and vice versa. However, this information refers to load variability but not to the flexibility that can be obtained from the aggregate load due to the collective behaviour of the consumers.

A comparison of $\boldsymbol{\psi}^{(a)}_{\Delta t_{s\%}}$ and ${}^{-}\boldsymbol{\psi}^{(a)}_{\Delta t_{s\%}}$ with other related indicators is shown in Fig. 7a and Fig. 7b. The effect of these indicators for aggregate demand can be seen in Fig. 7c. In the morning, between hour 6:00 to hour 8:00, there is a large amount of possible load decrease (since most of the load is increasing), but this does not correspond to high flexibility because of the "rigid" collective trend. As such, the *FIAD* is close to zero (Fig. 6).



(a) $PFL$and ${}^{-}\boldsymbol{\psi}^{(a)}_{\Delta t_{s\%}}$

(b)flexible demand for$\boldsymbol{\psi}^{(a)}_{\Delta t_{s\%}}$and ${}^{-}\boldsymbol{\psi}^{(a)}_{\Delta t_{s\%}}$

(c) aggregate demand pattern with and without demand flexibility

Fig. 7. *PFL*, flexible demand and aggregate demand comparisons with $a = 50$ houses and $\Delta t_s = 15$ minutes.

The same effect can be observed between hour 18:00 and hour 19:00. The *PFL* and the amount of flexible demand during these time slots are very small and can be seen in Fig. 7. From Fig. 6, the values of *FIAD* during the day between hour 9:00 and hour 17:00 are varying above 0.5, so a certain flexible demand is available (Fig. 7b).

During the night, between hour 2:00 and hour 5:00, the *FIAD* values are relatively high. These *FIAD* values are induced by the load *diversity* mainly due to the non-synchronized duty-cycles of the refrigerators that create load variations during the night period. This effect becomes less relevant when the averaging time step increases, as the patterns during the night become smoother (see Section V).

D. Confidence Bounds for PFL, ${}^{+}\boldsymbol{\psi}^{(a)}_{\Delta t_{s\%}}$ and ${}^{-}\boldsymbol{\psi}^{(a)}_{\Delta t_{s\%}}$

Since there are confidence limits associated with $\boldsymbol{\varphi}^{(a)}_{\Delta t_s}$, the uncertainty associated with $\boldsymbol{\psi}^{(a)}_{\Delta t_{s\%}}$ ${}^{+}\boldsymbol{\psi}^{(a)}_{\Delta t_{s\%}}$ and ${}^{-}\boldsymbol{\psi}^{(a)}_{\Delta t_{s\%}}$ can be calculated by substituting the confidence limits of $\boldsymbol{\varphi}^{(a)}_{\Delta t_s}$, defined in (24) and (25), into (26) to (28).

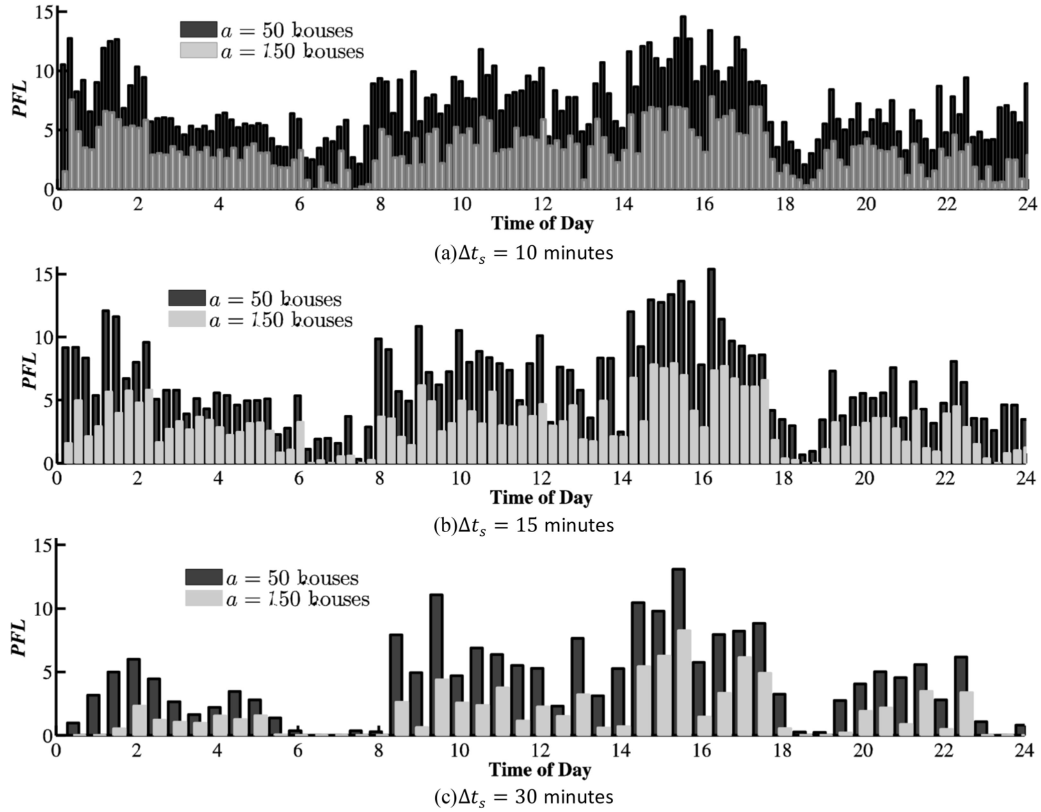


Fig. 8. Comparison of percentage flexibility level $\boldsymbol{\psi}^{(a)}_{\Delta t_{s_{\%}}}$ for $a = 50$ and 150 houses.

E. Operational Implications of the PFL Definition

For the system operator, the aggregate residential load is just one of the contributors to the overall load pattern of a distribution feeder or substation. The possibility of changing the shape of the overall load pattern depends on how flexible the different contributors can be. The indicators defined in this paper for aggregate residential demand may be useful for the operator's decision-making to establish whether selected time periods are feasible to initiate DR programmes involving residential customers. These data are also useful to identify the statistical properties of baseline patterns considered for testing the effectiveness of demand response programmes.[2]

For example, higher values of *PFL* in some time periods suggest that a reasonable reduction or increment in aggregate demand can be achieved in these time periods. To improve the decision-making effectiveness, the operator may consider the *PFL* indicator together with other inputs (e.g., electricity price). If during a specific time period the electricity price is high and the *PFL* is comparatively high with respect to other time periods with the same electricity price, then initiating a DR programme to reduce the demand in that time period may be effective. Likewise, higher values of *PFL* in time periods with low electricity price may suggest initiating a DR programme for increasing demand.

[2]The formulation of demand response programmes is outside the scope of this paper.

V. Effects of Aggregation Level and Time Step Duration

In the case study applications, $K = 100$ observations have been executed for different combinations of a and Δt_s. For space reasons, some results are presented here with reference to load pattern data sets of two different aggregations ($a = 50$ and $a = 150$ houses), with three different time step durations ($\Delta t_s = 10, 15, 30$ minutes) for each aggregation level a. The load variations are calculated for each data set by using (2) to (5). The binomial probability model explained in Section III-A is used to calculate the random variable $\mathbf{u}^{(a)}_{x\Delta t_s}$ for each combination of a and Δt_s by using the equations from (6) to (8). The binomial probabilities $\hat{\omega}^{(a)}_{x\Delta t_s}$ are estimated by using the MLE estimator explained in Section III-B. To overcome the limiting cases described in Section III-C, the relocated mean for binomial proportions and their CIs are calculated using the methodology presented in [35] by using (12) to (17).

For the different combinations of a and Δt_s, the indicators *FIAD* and *PFL* are calculated by using (23) and (26), respectively. The *PFL* provides information about the available amount of percentage flexible power with respect to the aggregate demand. The comparison of *PFL* with different aggregation levels and averaging time step durations is shown in Fig. 8. As discussed in Section IV-C, during the night the random variations in demand with small amplitudes are due for example to the non-synchronized duty-cycles of the refrigerators, and by increasing the averaging time step duration their effect becomes less prominent. From Fig. 8, for the same aggregation level, with the increase in time step duration a noticeable reduction in *PFL* can be seen during the time slot between hour 00:00 and hour 6:00. Conversely, from hour 08:00 to hour 18:00 the change in the *PFL* level is generally much lower. A remarkable reduction can be observed in particular between hour 06:00 and hour 08:00 and between hour 18:00 and 19:00, when the aggregate demand follows a strict trend towards increase or decrease in demand (see Fig. 4).

VI. Concluding Remarks

The focus of this paper has been set on the aggregate demand representing the collective behaviour of the customers, in order to quantify the flexibility achievable from the aggregate load in different time periods. Two novel demand flexibility indicators referring to residential demand aggregations have been formulated by using the binomial probability model of demand variations. The *FIAD* indicates the flexibility of aggregate customers in terms of probability of demand increase and decrease, and the *PFL* quantifies the per cent amount of flexible demand available for DSM purposes. These indicators extract *information* from demand variations and are useful for the system operator to select suitable time slots of the day to initiate DR programmes.

Using aggregate load patterns, the data handled do not require the knowledge of individual user details. This is a specific feature of the proposed approach, in which the calculations can be carried out by aggregating the patterns to discover their collective behaviour from statistical analysis, without interacting directly with the individual consumer and the related data. Thereby, in this approach privacy concerns are not an issue.

Furthermore, the proposed framework does not operate in real time. As such, possible effects of controls taking place at a given moment in time on specific appliances are not immediately affecting the outcomes of the statistical analysis that will consider the overall demand pattern resulting in the time period of analysis.

A specific analysis has been carried out by considering different averaging time step durations and different aggregation levels. The smaller averaging time step duration provides more granular information about the flexibility in change in demand behaviour but, particularly for the time slots with very low aggregate demand, this information is affected by the inherent randomness in the operation of appliances with non-synchronous duty cycles. Using variable averaging time step duration in different time slots of the day (e.g., with longer duration in the night period) can be a reasonable solution to mitigate this issue.

The effect of the aggregation level is also significant. When the aggregation level increases, the aggregate demand pattern represents more generalized system trend towards change in demand and is affected by the compensation of the load variations among the individual users. This compensation makes the overall demand appearing as more regular in time (with fewer variations), thus reducing the *PFL*.

The *PFL* indicates that the flexibility of the aggregate residential customers studied range from few per cents to about 7% for 150 houses. The maximum value of *PFL* for demand decrease could ideally represent the maximum level of flexibility that can be obtained by introducing incentives for demand reduction, e.g., within a DR programme. However, during the time steps when *FIAD* is close to zero, that is, when the aggregate demand follows a strict trend of increase or decrease, the difference between the values of $^{-}\psi^{(a)}_{\Delta t_{s\%}}$ and $\psi^{(a)}_{\Delta t_{s\%}}$ is relatively high. Hence, in practice when *FIAD* is close to zero it is unlikely to succeed in proposing relatively large demand reductions.

The approach followed in this paper is useful for the operator to identify how flexible can be the aggregate residential load in different periods of time. In this way, the operator can decide whether or not it can be viable to incentivize residential customers to change their demand patterns, taking into account the flexibility information identified in this paper to represent the collective behaviour of the residential customer aggregation. On the basis of the flexibility indices, the expected customer response to the incentives can be higher in some time periods and lower in other time periods, so that it may be useless for the operator to propose incentives to residential customers in time periods in which their aggregate demand is poorly flexible. In particular, in the time periods in which the values of the *FIAD* and *PFL* indicators are low, the proposal of actions aiming to re-shape the *aggregate* demand, even through specific incentives, could be poorly effective, for example because most consumers would be unavailable to change their lifestyle in these time periods. This fact limits the overall demand flexibility. The same actions proposed in other time periods in which the *FIAD* and *PFL* values are higher (that is, the collective behaviour of the consumers indicates no clear trend towards changing the demand in the same direction) could find in the set of consumers more candidates available to accept changes, leading to better ability to re-shape the aggregate demand under appropriate incentives.

The results presented in this paper have been found by applying a statistical model of the customers. However, today the companies managing the customer data have the capability of gathering simultaneously a sufficient amount of real residential load data to reproduce data sets similar to the ones indicated in this paper. For example, let us refer to Fig. 1, drawn for 100 aggregations of 50 customers each. The same situation can be constructed by gathering 5,000 simultaneous real load patterns at 15 min time step duration.

The definition of the flexibility indicators can be applied to other types of customer aggregations, in particular when the individual customers within the aggregation have a similar size, in order to maintain a similar meaning among the entries used to form the binomial distributions.

The way to exploit flexibility depends of many factors referring to the customers (including delays, energy payback, economic incentives, etc.). The indicators developed can be directly recalculated in case of changing behaviour of the consumers. The variation of these indicators can be seen as a further input for specific analyses referring to demand response. Further studies are in progress to assess the formulation of DR strategies that may use the information given by the flexibility indicators introduced in this paper.

APPENDIX
DEDUCTION OF FORMULA (10)

The MLE for the binomial proportions is calculated by using the following expression, based on the likelihood function $L(.)$:

$$\hat{\omega}^{(a)}_{x\Delta t_s} = \arg\ \max\left[L\left(prob\left(\mathbf{u} = \mathbf{u}^{(a)}_{x\Delta t_s}\right)\right)\right] \tag{29}$$

Eq. (29) can be rewritten by using (9):

$$\hat{\omega}^{(a)}_{x\Delta t_s} = \arg\ \max\left[L\left(\omega^{(a)}_{x\Delta t_s}\right)\right] \tag{30}$$

where the term $L(\omega^{(a)}_{x\Delta t_s})$ is the likelihood function for $\omega^{(a)}_{x\Delta t_s}$ and for binomial probability distributions it can be defined as:

$$L\left(\omega^{(a)}_{x\Delta t_s}\right) = \binom{K}{\mathbf{u}^{(a)}_{x\Delta t_s}}\left(\omega^{(a)}_{x\Delta t_s}\right)^{\mathbf{u}^{(a)}_{x\Delta t_s}}\left(1 - \omega^{(a)}_{x\Delta t_s}\right)^{K-\mathbf{u}^{(a)}_{x\Delta t_s}} \tag{31}$$

The log likelihood function is used to estimate $\omega^{(a)}_{x\Delta t_s}$. Then, the right hand side of (30) can be written as:

$$\frac{d}{d\omega^{(a)}_{x\Delta t_s}}\left[\ln\left(L\left(\omega^{(a)}_{x\Delta t_s}\right)\right)\right] = 0 \tag{32}$$

By solving (32):

$$\frac{d}{d\omega^{(a)}_{x\Delta t_s}}\left[\mathbf{u}^{(a)}_{x\Delta t_s}\ln\left(\omega^{(a)}_{x\Delta t_s}\right) + \left(K - \mathbf{u}^{(a)}_{x\Delta t_s}\right)\ln\left(1 - \omega^{(a)}_{x\Delta t_s}\right)\right] = 0 \tag{33}$$

the derivative yields:

$$\frac{\mathbf{u}^{(a)}_{x\Delta t_s}}{\omega^{(a)}_{x\Delta t_s}} - \frac{K - \mathbf{u}^{(a)}_{x\Delta t_s}}{1 - \omega^{(a)}_{x\Delta t_s}} = 0 \tag{34}$$

Finally, solving (34) with respect to $\omega^{(a)}_{x\Delta t_s}$ gives the estimated probability $\hat{\omega}^{(a)}_{x\Delta t_s}$ in equation (10).

REFERENCES

[1] B. V. Mathiesen and H. Lund, "Comparative analyses of seven technologies to facilitate the integration of fluctuating renewable energy sources," *IET Renew. Power Gener.*, vol. 3, no. 2, pp. 190–204, Jun. 2009.

[2] W. Golden and P. Powell, "Towards a definition of flexibility: In search of the Holy Grail?" *Omega*, vol. 28, no. 4, pp. 373–384, 2000.

[3] E. Lannoye, D. Flynn, and M. O'Malley, "Evaluation of power system flexibility," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 922–931, May 2012.

[4] J. Ma, V. Silva, R. Belhomme, D. S. Kirschen, and L. F. Ochoa, "Evaluating and planning flexibility in sustainable power systems," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PES)*, Vancouver, BC, Canada, 2013, pp. 1–11.

[5] A. Ulbig and G. Andersson, "On operational flexibility in power systems," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, San Diego, CA, USA, 2012, pp. 1–8.

[6] D. S. Kirschen, A. Rosso, J. Ma, and L. F. Ochoa, "Flexibility from the demand side," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, San Diego, CA, USA, 2012, pp. 1–6.

[7] D. Papadaskalopoulos and G. Strbac, "Decentralized participation of flexible demand in electricity markets—Part I: Market mechanism," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 3658–3666, Nov. 2013.

[8] X. Chen *et al.*, "Increasing the flexibility of combined heat and power for wind power integration in China: Modeling and implications," *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 1848–1857, Jul. 2015.

[9] A. Papavasiliou and S. S. Oren, "Large-scale integration of deferrable demand and renewable energy sources," *IEEE Trans. Power Syst.*, vol. 29, no. 1, pp. 489–499, Jan. 2014.

[10] R. Stamminger *et al.*, "Synergy potential of smart appliances," Fed. Ministry Environ., Nat. Conserv., Nuclear Safety, Bonn, Germany, Tech. Rep. Project Smart-A 0327668/0327668A, 2008.

[11] C. Vivekananthan, Y. Mishra, G. Ledwich, and F. Li, "Demand response for residential appliances via customer reward scheme," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 809–820, Mar. 2014.

[12] A. Safdarian, M. Fotuhi-Firuzabad, and M. Lehtonen, "Benefits of demand response on operation of distribution networks: A case study," *IEEE Syst. J.*, vol. 10, no. 1, pp. 189–197, Mar. 2016.

[13] M. Kefayati and R. Baldick, "On energy delivery to delay-averse flexible loads: Optimal algorithm, consumer value and network level impacts," in *Proc. CDC*, Maui, HI, USA, 2012, pp. 2401–2408.

[14] C. Joe-Wong, S. Sen, S. Ha, and M. Chiang, "Optimized day-ahead pricing for smart grids with device-specific scheduling flexibility," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 6, pp. 1075–1085, Jul. 2012.

[15] B. Kladnik, A. Gubina, G. Artač, K. Nagode, and I. Kockar, "Agent-based modeling of the demand-side flexibility," in *Proc. IEEE Power Energy Soc. Gen. Meeting,* San Diego, CA, USA, 2011, pp. 1–8.

[16] J. A. F. Moreno, A. M. Garcia, A. G. Marin, E. G. Lázaro, and C. A. Bel, "An integrated tool for assessing the demand profile flexibility," *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 668–675, Feb. 2004.

[17] W. Zhang, K. Kalsi, J. Fuller, M. Elizondo, and D. Chassin, "Aggregate model for heterogeneous thermostatically controlled loads with demand response," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, San Diego, CA, USA, 2012, pp. 1–8.

[18] D. Wang *et al.*, "Performance evaluation of controlling thermostatically controlled appliances as virtual generators using comfort-constrained state-queueing models," *IET Gener. Transm. Distrib.*, vol. 8, no. 4, pp. 591–599, Apr. 2014.

[19] H. Hao, B. M. Sanandaji, K. Poolla, and T. L. Vincent, "Aggregate flexibility of thermostatically controlled loads," *IEEE Trans. Power Syst.*, vol. 30, no. 1, pp. 189–198, Jan. 2015.

[20] M. Alizadeh, A. Scaglione, A. Applebaum, G. Kesidis, and K. Levitt, "Reduced-order load models for large populations of flexible appliances," *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 1758–1774, Jul. 2015.

[21] J. L. Mathieu, M. G. Vayá, and G. Andersson, "Uncertainty in the flexibility of aggregations of demand response resources," in *Proc. 39th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Vienna, Austria, 2013, pp. 8052–8057.

[22] I. A. Sajjad, G. Chicco, and R. Napoli, "Effect of aggregation level and sampling time on load variation profile—A statistical analysis," in *Proc. 17th IEEE Mediterr. Electrotech. Conf. (MELECON)*, Beirut, Lebanon, 2014, pp. 208–212.

[23] I. A. Sajjad, G. Chicco, and R. Napoli, "A probabilistic approach to study the load variations in aggregated residential load patterns," in *Proc. 18th Power Syst. Comput. Conf. (PSCC)*, Wrocław, Poland, 2014, pp. 1–7.

[24] A. Capasso, W. Grattieri, R. Lamedica, and A. Prudenzi, "A bottom-up approach to residential load modeling," *IEEE Trans. Power Syst.*, vol. 9, no. 2, pp. 957–964, May 1994.

[25] A. Cagni, E. Carpaneto, G. Chicco, and R. Napoli, "Characterisation of the aggregated load patterns for extraurban residential customer groups," in *Proc. 12th IEEE Mediterr. Electrotech. Conf. (MELECON)*, Dubrovnik, Croatia, 2004, pp. 951–954.

[26] E. Carpaneto and G. Chicco, "Probabilistic characterisation of the aggregated residential load patterns," *IET Gener. Transm. Distrib.*, vol. 2, no. 3, pp. 373–382, May 2008.

[27] G. C. Simbahan and A. Dobermann, "An algorithm for spatially constrained classification of categorical and continuous soil properties," *Geoderma*, vol. 136, nos. 3–4, pp. 504–523, 2006.

[28] N. Sourial *et al.*, "Correspondence analysis is a useful tool to uncover the relationships among categorical variables," *J. Clin. Epidemiol.*, vol. 63, no. 6, pp. 638–646, 2010.

[29] B. Xu, X. Feng, and R. D. Burdine, "Categorical data analysis in experimental biology," *Develop. Biol.*, vol. 348, no. 1, pp. 3–11, 2010.

[30] A. Agresti, *An Introduction to Categorical Data Analysis* (Series in Probability and Mathematical Statistics). Hoboken, NJ, USA: Wiley, 2007.

[31] A. DasGupta and H. Rubin, "Estimation of binomial parameters when both n, p are unknown," *J. Stat. Plan. Infer.*, vol. 130, nos. 1–2, pp. 391–404, 2005.

[32] B. J. Oommen, S.-W. Kim, and G. Horn, "On the estimation of independent binomial random variables using occurrence and sequential information," *Pattern Recognit.*, vol. 40, no. 11, pp. 3263–3276, 2007.

[33] R. G. Newcombe, "Two-sided confidence intervals for the single proportion: Comparison of seven methods," *Stat. Med.*, vol. 17, no. 8, pp. 857–872, 1998.

[34] I. Bárány and V. Vu, "Central limit theorems for Gaussian polytopes," *Ann. Probab.*, vol. 35, no. 4, pp. 1593–1621, 2007.
[35] L. D. Brown, T. T. Cai, and A. DasGupta, "Interval estimation for a binomial proportion," *Stat. Sci.*, vol. 16, no. 2, pp. 101–117, 2001.
[36] E. B. Wilson, "Probable inference, the law of succession, and statistical inference," *J. Amer. Stat. Assoc.*, vol. 22, no. 158, pp. 209–212, 1927.
[37] Y. Guan, "A generalized score confidence interval for a binomial proportion," *J. Stat. Plan. Infer.*, vol. 142, no. 4, pp. 785–793, 2012.

Intisar Ali Sajjad (S'06–M'16) received the Ph.D. degree in electrical engineering from Politecnico di Torino, Turin, Italy, in 2015. He is currently an Assistant Professor with the Electrical Engineering Department, University of Engineering and Technology, Taxila, Pakistan. His current research interests include smart buildings, power system analysis, and load management.

Gianfranco Chicco (M'98–SM'08) received the Ph.D. degree in electrotechnics engineering from Politecnico di Torino (PdT), Turin, Italy, in 1992. He is currently a Professor of Electrical Energy Systems, Energy Department, PdT. His current research interests include power system and distribution system analysis, energy efficiency, multi-generation, load management, artificial intelligence applications, and power quality. He is a member of the Italian Association of Electrical, Electronics, and Telecommunications Engineers.

Roberto Napoli (M'74) received the master's degree in electrical engineering from Politecnico di Torino (PdT), Turin, Italy, in 1969. He is a Full Professor of Power Systems with PdT. His research interests include operation, planning, economics, and security of electric energy systems, domotics, energy efficiency controls, and electrical safety.