

Empirical Voltage-current Signatures for Individual Household Consumers Obtained by Non-linear Regression*

Dadiana-Valeria CĂIMAN*, Toma-Leonida DRAGOMIR

Politehnica University Timișoara, Automation and Applied Informatics, 2 Vasile Pârva, Timișoara, Romania
dadiana.grando@aut.upt.ro (*Corresponding author), toma.dragomir@upt.ro

Abstract: The article is part of the approach of finding the signatures for the individual household consumers using the voltage-current trajectories. Four classes of signatures are presented, known as: discontinuous tangent, tangent, ellipse, and hybrid. The functions describing the signatures are defined using a method based on non-linear regression and genetic algorithms. The method is implemented, with concrete results, on consumers from each of the four classes, starting with pairs of voltage-current values measured at low frequency, of approximately 20 Hz. This represents an important aspect from a practical point of view. There are four examples of signatures determined for: a laptop, a vacuum cleaner, a TV, and a fridge.

Keywords: Voltage-current signature, Household consumers, Non-linear regression, Genetic algorithms.

1. Introduction

Saving energy has always been a worldwide priority. Over the last few years a lot of attention has been directed towards reducing the consumption of electricity in the household sector. A confirmation regarding the need for decreasing the domestic energy usage is proved by the high energy amount in the overall energy consumption: 29% in EU (EEA, 2018), 28.2% in Romania (Oprîș, 2018).

Hart (Hart, 1992) questions for the first time the identification of an electrical device out of the aggregate energy use based on a signature. The outcomes of this type of identification might be: reducing the electricity use as a result of the owners' awareness on consumption of each device, detecting the faulty devices, increasing the quality of the energy delivered, avoiding the spikes in consumption and, thus managing better the power grid.

In the identification of an electrical device out of an aggregate energy use, the defining elements are the type of the signature and the set of the

related parameters. For a given type of identifiable signature the focus is placed on establishing the set of parameters that will define the signature in the end. The signature uniquely identifies the paired device.

The measured signals in a household power grid are the voltage and the electrical current. The parameters that may define the signatures are calculated using time series of the voltage and current. According to the working regime chosen for measuring, i.e. a steady-state (quasi-stationary) regime or a transitory regime, two types of parameters may be computed parameters associated to the steady-state regime, and parameters associated to the transitory regime. In order to obtain the first type of parameters one can work with low acquisition frequency (Hz), while for the second type one needs kHz level frequencies so that the voltage/current time series can encompass the fast state features.

Initially, in literature, only one or two parameters were used as signature. Later on, several compound signature parameters had been considered.

The identification algorithm of an individual consumer from an aggregate consumer relies on defining the signatures as well on the usage of a decomposition algorithm. According to (Zhuang, Shahidehpour & Li, 2018) the main categories of the developed algorithms are based on optimisation methods, statistics probability methods, machine learning / pattern recognition, deep learning, and graph signal processing.

* This paper is based on two previous presentations in "The 12th International Symposium on Applied Computational Intelligence and Informatics (SACI 2018)", and "The 22nd International Conference on System Theory, Control and Computing (ICSTCC)" (see reference 3 and 4). This work integrates the previously presented methods in the category of regression methods, reduces the number of the parameters for the already known signatures and defines a new more complex class of signatures.

In (Azaza, & Wallin, 2017) the performances of using some identification algorithms related to different signatures are presented: power descriptors (active power, reactive power), current waveform, voltage-current (curve asymmetry, looping direction), current harmonics (power amplitude variation at different frequencies), transitory regime descriptors (variation of the power amplitude at switching on and off, oscillations duration, over increase power duration). Their work also presents the instance in which the signature incorporates the cumulating of the above-mentioned parameters. As a conclusion, the pattern-recognition failure depends not only on the algorithm but also on the type of the signature. There are algorithms for which the compound signature does not generate the minimal recognition error.

In (Ghorbanpour & Mallipeddi, 2018) the parameters normally used in voice recognition are proposed, namely the speed and acceleration of a descriptor.

In (Bouhouras, et al. 2019) the authors propose a signature that results through shaping the first three odd current harmonics into vectors that include phase shift and amplitude.

In (Huang, et al. 2017) the signature is obtained from the average active power time series calculated over time intervals ranging between 6s to 15 min. The article shows a non-linear performance reduction for various identification algorithms once the average active power calculation intervals have increased.

In (Held, et al., 2016) the signature resulted from a current matrix in which the number of rows represents the number of consecutive periods, and the columns are the samples from a period. The values in the matrix stand for the values of the sampled current at different time intervals. Before generating the matrix, the sampling operations of current and voltage are synchronised. This can be achieved by using the passing of voltage through zero, so that no information related to voltage-current phase shift is lost. This method, known as the Frequency Invariant Transformation of Periodic Signals (FIT-PS), enables the simplification of the calculation matrix.

The present article proposes a signatures development method based on parameters linked to some functions which approximates the voltage-current trajectory, $i(v)$, of the paired electrical device. The primary information necessary for building the signature consists in a series of values measured at a low sampling frequency (approximately 20 Hz). The same information obtained at a much higher frequency (15 – 30 kHz) is used in various methods. This is the case, of the paper published by Lam (Lam, Fung & Lee, 2007), where the signature is made of 7 from 8 proposed geometrical descriptors that characterise the voltage/current curve. In (Hassan, Javed, & Arshad, 2014), a new geometrical descriptor is proposed. In (Longjun et al., 2018) there are 10 geometrical descriptors of the trajectory used as signature. These include the 8 used in (Lam, Fung & Lee, 2007), as well as the current span and the variation of instantaneous admittance. In (De Baets, et al., 2018) the voltage-current trajectory is converted in a binary matrix.

In the subsequent sections the article proposes regression functions as signatures for individual consumers in order to approximate the voltage-current trajectory (as series of pairs of voltage-current values), with determined parameters using genetic algorithm. The genetic algorithms are a known tool used for obtaining mathematical models that include regression functions (Minghua, et al., 2017). The regression analysis is used in various fields. E.g., in (Kialashaki & Reisel, 2013), the linear regressions approach is used to model the required energy for the residential sector.

Although the above mentioned methods make use of complex calculation methods and algorithms, they do not provide as a result analytical expressions for the signatures. In the mentioned references the signatures appear in numerical form (vector of parameters, number of harmonics, etc.) or in graphic form (current-voltage trajectories or other trajectories derived from them). In this context, in the papers (Căiman & Dragomir 2018a), (Căiman & Dragomir 2018b), the authors proposed parameterizable analytic expressions as signatures based on measurements. These signatures offer the possibility to identify individually different consumers. This work extends the methods and

the results of the two mentioned papers. A unitary terminology is also introduced.

The paper is thus structured as follows: in section 2 the regression functions of four classes of consumers and their associated signatures are presented; in section 3 the method used for determining the signature using genetic algorithms is described; section 4 refers to case studies; section 5 includes the conclusions and some possible future research directions.

2. Signatures of Some Individual Consumers

Basically the signature of any consumer represents all the pairs of points $M = \{(v, i)\}$ measured by sampling in quasi-stationary regime over a long time span. Due to the practical working conditions and measurement uncertainties, the points are spread in a point cloud around a closed curve, periodically covered (see Figures 1-4). In this context it is natural to consider this closed curve and its equation as the signature of the consumer. The curve is essentially the graphic representation of a function $i(v)$ associated with the processing of the experimentally gathered data $\{(v, i)\}$.

Under the assumption that the function $i(v)$ is „a realistic approximation” of the correlation between the two variables of the process that takes place in the consumer, voltage v and current i , this function is a regression function (Vlada, 2012). The computing of the regression function, like empirical model, is necessary because due to the nonlinearities, the dependency between v and i cannot be analytically and accurately established for one consumer or the other.

The regression function expresses the dependency between the two variables through a set of parameters, i.e. regression function parameters (coefficients for linear regression), denoted in the following with Π .

If the regression function is a linear expression of Π , one can refer to it as a linear regression.

The regression is non-linear otherwise (Vlada, 2012).

Based on the shape of the closed curve, the signature of an individual consumer, characterised by the regression function, shall be composed of a descending branch and an ascending branch of equations $i_a = S_{C_a}(v)$ and $i_d = S_{C_d}(v)$ respectively. The set of these functions is denoted by $S_{C_*} : [V_{min}, V_{max}] \subset R \rightarrow R$. Each function is determined from one set of measured points, M_a , respectively M_d , extracted from set M .

The method through which the signatures are obtained relies on two working hypotheses. First hypothesis: the regression function is selected from a set of available patterns that can be parameterized using two sets of parameters: set Π_a , for ascending branch, and set Π_d , for the descending branch. Second hypothesis: the values of the parameters from the two sets are estimated and obtained using genetic algorithms. In order to determine the values of the parameters, they are restricted to closed intervals. The sets of these intervals are denoted by I_{Π_a} and I_{Π_d} . For example, if the ascending branch has 3 parameters, $\Pi_a = \{p_1, p_2, p_3\}$, then $I_{\Pi_a} = \{[p_{1min}, p_{1max}], [p_{2min}, p_{2max}], [p_{3min}, p_{3max}]\}$.

To simplify the presentation, the term *class* was used to denote the form of the regression function, i.e. the form of the pair of functions $i_a = S_{C_a}(v), i_d = S_{C_d}(v)$, and the term *signature was used* for denoting the concrete functions associated with the values of the parameters (Π_a, Π_d). The classes are named after the analytical functions used.

In this paper 4 classes are presented, namely: discontinuous tangent (C_{TgDisc}), continuous tangent (C_{Tg}), ellipse (C_{ell}) and hybrid (C_h) classes. The first 3 classes had been introduced in (Căiman, & Dragomir, 2018a) and (Căiman, & Dragomir, 2018b). The present article summarizes, refines and supplements the previous presentations by introducing conditions that had led to the decrease of the number of parameters. For the discontinuous tangent class the number of the parameters is reduced from a set of 16 to a set of only 12 parameters, for the continuous tangent class from 14 to a set of only 10 parameters, and for the ellipse class from a set of 5 to a set of 4 parameters.

2.1 Discontinuous Tangent Class

When the point cloud resulted from the measurements appears as illustrated in Figure 1 one estimate that the signatures belong to the discontinuous tangent class.

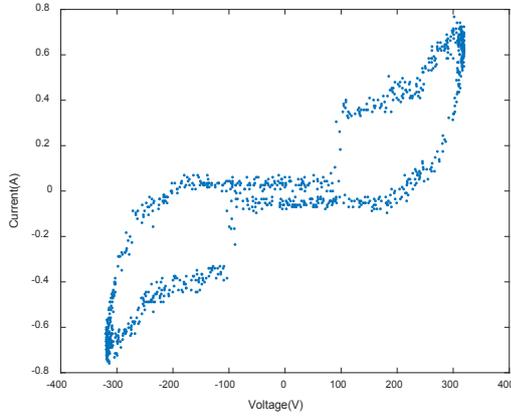


Figure 1. Point cloud for discontinuous tangent class

For this class the support function has the form (1) in which „ \ast ” can be a (ascending branch) or d (descending branch):

$$S_{TgDisc\ast} : \left[-c_l \frac{\pi}{2} - d_l + \varepsilon, -c_r \frac{\pi}{2} - d_r + \varepsilon \right] \rightarrow R$$

$$S_{TgDisc\ast}(v) = \left[\frac{1}{a_l} \operatorname{tg} \frac{v + d_l}{c_l} + b_l \right] \cdot \frac{1 - \operatorname{sgn}(v - v_b)}{2} + \left[\frac{1}{a_r} \operatorname{tg} \frac{v + d_r}{c_r} + b_r \right] \cdot \frac{1 + \operatorname{sgn}(v - v_b)}{2} \quad (1)$$

In equation (1) $\operatorname{sgn}(x)$ has the following form:

$$\operatorname{sgn}(x) = \begin{cases} -1, & x < 0 \\ 0, & x = 0 \\ 1, & x > 0 \end{cases}$$

The support function (1) has 10 parameters for each branch: $a_p, b_p, c_p, d_p, a_r, b_r, c_r, d_r, v_b$ and ε . The index l stands for „left side”, while the index r stands for „right side”.

Out of the 10 parameters, the last 2 are initialized. Thus, the parameter v_b that marks the point of discontinuity of the function (1) approximates the voltage value for which the shape of the point cloud has a discontinuity. As well, the parameter ε is adopted so that the argument of the trigonometric tangent should not reach the values $+\pi/2$. This last condition introduces two relationships between the parameters. If the

domain of variation for the already measured voltage v is the interval $[V_{\min}, V_{\max}]$, where $V_{\min} = -V_{\max}$, and if it is identical to the definition domain of the signature, then the two relationships are:

$$d_l = -c_l \cdot \frac{\pi}{2} + c_l \cdot \varepsilon - V_{\min}, d_r = -c_r \cdot \frac{\pi}{2} - c_r \cdot \varepsilon - V_{\max}$$

Both equations are written in the form of substitution formulas, the two parameters d_p, d_r being dependent on c_l and c_r . Consequently, the sets of parameters of the discontinuous tangent class are:

$$\Pi_a = \{a_l^a, b_l^a, c_l^a, a_r^a, b_r^a, c_l^a\}, \quad (2.1)$$

$$\Pi_d = \{a_l^d, b_l^d, c_l^d, a_r^d, b_r^d, c_l^d\}$$

These sets include 6 parameters for the ascending branch, Π_a , and 6 for the descending branch, Π_d . The intervals in which the parameters receive values are:

$$\Pi_{\Pi_a} = \left\{ \left[a_l^a \min, a_l^a \max \right], \left[b_l^a \min, b_l^a \max \right], \left[c_l^a \min, c_l^a \max \right], \left[a_r^a \min, a_r^a \max \right], \left[b_r^a \min, b_r^a \max \right], \left[c_r^a \min, c_r^a \max \right] \right\}$$

$$\Pi_{\Pi_d} = \left\{ \left[a_l^d \min, a_l^d \max \right], \left[b_l^d \min, b_l^d \max \right], \left[c_l^d \min, c_l^d \max \right], \left[a_r^d \min, a_r^d \max \right], \left[b_r^d \min, b_r^d \max \right], \left[c_r^d \min, c_r^d \max \right] \right\} \quad (2.2)$$

2.2 Continuous Tangent Class

In Figure 2, the point cloud $\{(v, i)\}$ has resulted from measurements conducted on an individual consumer that belongs to the continuous tangent class. This class includes consumers like laptops.

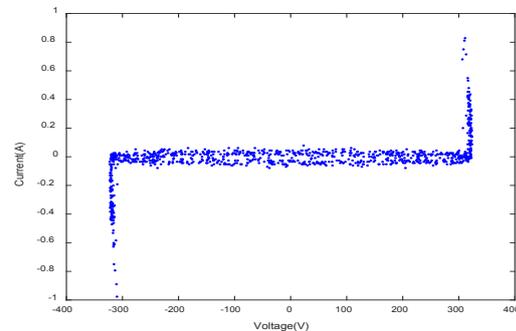


Figure 2. Point cloud for continuous tangent class

The support functions $i_a = S_{Tg_a}(v), i_d = S_{Tg_d}(v)$, associated with the continuous tangent class are deduced from equation (1).

The point cloud suggests a regression function that is continuous in $v = 0$ and presents a symmetry in relation to the value $v = 0$. As a result, in (1) $v_b = 0$ is hereby adopted and the following relationship is imposed:

$$\begin{aligned} \frac{1}{a_l} \operatorname{tg}\left(\frac{d_l}{c_l}\right) + b_l &= \frac{1}{a_r} \operatorname{tg}\left(\frac{d_r}{c_r}\right) + b_r \\ b_r &= \frac{1}{a_l} \operatorname{tg}\left(\frac{d_l}{c_l}\right) + b_l - \frac{1}{a_r} \operatorname{tg}\left(\frac{d_r}{c_r}\right). \end{aligned}$$

Then, the formula of substitution results

$$b_r = \frac{1}{a_l} \operatorname{tg}\left(\frac{d_l}{c_l}\right) + b_l - \frac{1}{a_r} \operatorname{tg}\left(\frac{d_r}{c_r}\right) \quad (3)$$

The other aspects discussed in the case of the discontinuous tangent class remain the same.

Replacing in (1) b_r with the expression from (3) and changing b_l in b , the following support function is obtained:

$$\begin{aligned} S_{Tg*} &: \left[-c_l \frac{\pi}{2} - d_l + \varepsilon, -c_r \frac{\pi}{2} - d_r + \varepsilon\right] \rightarrow R \\ S_{Tg*} &= \left[\frac{1}{a_l} \operatorname{tg} \frac{v+d_l}{c_l} - \frac{1}{a_l} \operatorname{tg} \frac{d_l}{c_l} + \frac{1}{a_r} \operatorname{tg} \frac{d_r}{c_r} \right] \cdot \frac{1 - \operatorname{sgn}(v)}{2} \\ &+ b + \left[\frac{1}{a_r} \operatorname{tg} \frac{v+d_r}{c_r} + b_r \right] \cdot \frac{1 + \operatorname{sgn}(v)}{2} \end{aligned} \quad (4)$$

In this case each set contains 5 parameters:

$$\Pi_a = \{a_l^a, b^a, c_l^a, a_r^a, c_l^a\}, \Pi_d = \{a_l^d, b^d, c_l^d, a_r^d, c_l^d\} \quad (5.1)$$

The related intervals are:

$$\begin{aligned} I_{\Pi_a} &= \left\{ \left[a_{l \min}^a, a_{l \max}^a \right], \dots, \left[c_{r \min}^a, c_{r \max}^a \right] \right\} \\ I_{\Pi_d} &= \left\{ \left[a_{l \min}^d, a_{l \max}^d \right], \dots, \left[c_{r \min}^d, c_{r \max}^d \right] \right\} \end{aligned} \quad (5.2)$$

2.3 Ellipse Class

In Figure 3 a point cloud belonging to a consumer from the ellipse class is represented. By analysing the figure, it can be noticed that the shape of the cloud resembles a peanut shape ellipse rotated with the angle β . The fridge is one of the consumers from this class.

The support functions that correspond to the ellipse class, $i_a = S_{ell_a}(v), i_d = S_{ell_d}(v)$, are obtained from the equation (6) of a non-rotated ellipse with the semi-axis of lengths a and b .

$$\begin{aligned} S_{ell_*} &: [V_{min}, V_{max}] \rightarrow R \\ S_{ell_*}(v) &= \frac{\left(1 - \frac{b^2}{a^2}\right) \cdot \sin^2(\beta) \cdot v \cdot \operatorname{ctg}(\beta) m \sqrt{\Delta(v)}}{1 - \left(1 - \frac{b^2}{a^2}\right) \cdot \sin^2(\beta)}, \end{aligned} \quad (6)$$

$$\Delta(v) = \left(1 - \frac{v^2}{a^2}\right) - \left(1 - \frac{b^2}{a^2}\right) \cdot \sin^2(\beta)$$

The necessary conditions for closing the two branches of the ellipse in $v = V_{min}$ and $v = V_{max}$ are:

$$\Delta(V_{max}) = 0, \Delta(V_{min}) = 0, \Delta(v) \geq 0, v \in [V_{min}, V_{max}] \quad (7)$$

Since $V_{min} = -V_{max}$ the first two relations from (7) are reduced to

$$\sin^2(\beta) = \frac{a^2 - V_{max}^2}{a^2 - b^2} \quad (8)$$

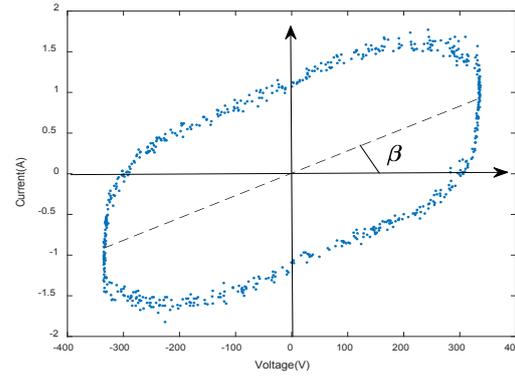


Figure 3. Point cloud of an ellipse class consumer

In order to make equation (8) explicit, the following conditions should be accomplished:

$$I > \frac{a^2 - V_{max}^2}{a^2 - b^2} > 0 \quad (9)$$

If this double inequality is satisfied, the last condition in (7) is accomplished.

Taking into consideration that the maximal numeric values for normal users regarding current and voltage, I_{max} and V_{max} , expressed in Ampere and Volt, differ with at least one order of magnitude, it results that for the ellipses $a > b$. In conclusion, from (9) it results: $a > V_{max}$.

Replacing $\sin^2(\beta)$ and $\operatorname{ctg}(\beta)$ in (6) based on equation (8) and applying a double correction (to

approximate the peanut shape of the cloud), which includes a sinusoidal function and an exponential function the expression of the signature becomes:

$$S_{ell_*} : [V_{min}, V_{max}] \rightarrow R$$

$$S_{ell_*}(v) = \left\{ \left[\frac{a^2}{V_{max}^2} - 1 \right] \cdot v \cdot \sqrt{\frac{V_{max}^2 - b^2}{a^2 - V_{max}^2}} m \frac{b \cdot a}{V_{max}} \cdot \sqrt{1 - \frac{v^2}{V_{max}^2}} \right\}$$

$$\left(1 + A \cdot \sin \left(3\pi \cdot \frac{v - V_{min}}{v - V_{max}} \right) \right) \cdot e^{\frac{|v|}{k \cdot V_{max}}}$$

$$a > V_{max}, b < V_{max}$$
(10)

The set of parameters contains 4 elements:

$$\Pi = \{a, b, A, k\} \quad (11.1)$$

The necessary intervals are:

$$I\Pi = \{[a_{min}, a_{max}], \dots, [k_{min}, k_{max}]\} \quad (11.2)$$

2.4 Hybrid Class

In Figure 4 the point cloud of a hybrid class consumer is shown. One type of consumer that belongs to this class is the vacuum cleaner.

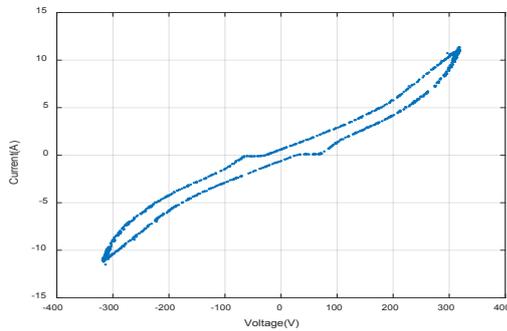


Figure 4. Point cloud of a hybrid class consumer

The support functions corresponding to the hybrid class, $i_a = S_{h_a}(v), i_d = S_{h_d}(v)$, have the following expressions:

$$S_{h_*} : [V_{min}, V_{max}] \rightarrow R$$

$$S_{h_*}(v) = \left\{ \sum_{j=1}^5 \left[I_j + a_j \cdot |v - V_j|^{n_j} \cdot (\text{sgn}(v - V_j))^{\alpha_j} \right] \cdot f_{c_1}(V_j) \right\}$$

$$+ k \cdot f_{c_2}(V_x),$$

$$f_{c_1}(V_j) = e^{b_{1j} \cdot (v - V_j) \frac{1 - \text{sgn}(v - V_j)}{2} + b_{2j} \cdot (v - V_j) \frac{1 + \text{sgn}(v - V_j)}{2}}$$

$$f_{c_2}(V_x) = e^{-[k_1(1 - \text{sgn}(v - V_x)) + k_2(1 + \text{sgn}(v - V_x))] |v - V_x|}$$
(12)

Function (12) contains 6 terms. Out of these, 5 terms are achieved in correlation with the expression $|v - V_j|^{n_j}$, where n_j is a rational number. Note that in any calculation on a computer the representations of the numbers correspond to rational expressions. All the terms contain correction factors (attenuation/ amplification) with asymmetric exponential variation related to differences $v - V_j$. The sixth term allows the rendering of angular points. Since (12) does not contain standard functions, the function has been named “hybrid”.

Two observations led to the writing of the formula (12): i) the branches of the signature have an odd function shape with respect to the vertical $v = 50$ V (ascending branch) and $v = -50$ V (descending branch) ii) the branches of the signature are made of arches with parabolic aspect near the voltage level of $v = V_j$. The number of elements from (12) in $|v - V_j|$ can be an odd number that is higher or equal to 3 (in (12) 5 terms have been considered).

In (12) the parameter α_j can take values of 1 or 2, thus resulting following cases:

$$f(x) = \begin{cases} a \cdot \text{sgn}(x) \cdot |x|^n, & \alpha = 1 \\ a \cdot |x|^n, & \alpha = 2 \end{cases}, a \in R$$

The term with $\alpha_j = 1$ allows the reproduction of the parts with inflexion points, while the one with $\alpha_j = 2$ shows the areas of local extreme points.

The parameters I_l and I_s (with the physical dimension expressed in amperes) result from imposing the following conditions on the interval extremities $[V_{min}, V_{max}]$: $i_{min} = S_{h_*}(V_{min})$ and $i_{max} = S_{h_*}(V_{max})$, where i_{min} and i_{max} are approximation values for the electrical current i at the end of each interval:

$$i_{min} = i_{min_a} + \beta_{min} i_{min_d}, \quad i_{max} = \beta_{max} i_{max_a} + i_{max_d} \quad (13)$$

In these equations i_{min_a} and i_{min_d} are the values of the current corresponding to the points from the left extremities of sets M_a and M_d , whereas i_{max_a} and i_{max_d} are the values of the current that corresponds to the points from the right extremities of the same sets. The factors $\beta_{min} \in (0, 1)$ and $\beta_{max} \in (0, 1)$ have been previously adopted.

In this case 74 parameters are defining the whole signature, out of which 37 for the ascending branch and 37 for the descending one:

$$\begin{aligned} \Pi_a = & \{V_1^a, \dots, V_5^a, n_1^a, \dots, n_5^a, \alpha_1^a, \dots, \alpha_5^a, b_{11}^a, \dots, b_{15}^a, \\ & b_{r1}^a, \dots, b_{r5}^a, V_x^a, k_1^a, k_2^a, k^a, I_2^a, \dots, I_4^a, a_1^a, \dots, a_5^a\} \\ \Pi_d = & \{V_1^d, \dots, V_5^d, n_1^d, \dots, n_5^d, \alpha_1^d, \dots, \alpha_5^d, b_{11}^d, \dots, b_{15}^d, \\ & b_{r1}^d, \dots, b_{r5}^d, V_x^d, k_1^d, k_2^d, k^d, I_2^d, \dots, I_4^d, a_1^d, \dots, a_5^d\} \end{aligned} \quad (14.1)$$

The parameters must lie in intervals as follows:

$$\begin{aligned} I_{\Pi_a} = & \{[V_{1min}^a, V_{1max}^a], \dots, [a_{5min}^a, a_{5max}^a]\} \\ I_{\Pi_d} = & \{[V_{1min}^d, V_{1max}^d], \dots, [a_{5min}^d, a_{5max}^d]\} \end{aligned} \quad (14.2)$$

3. The Approach for Determining the Individual Consumer Signature

The applied approach contains 2 steps:

- i) application of an algorithm of point's allocation, which results in generating the subsets M_a and M_d ;
- ii) application of a genetic algorithm for each subset in order to obtain the values of the specific parameter set for the selected class, and thus the signature.

Note that in this context, the class is selected in advance through inspection of the point cloud $M = \{(v, i)\}$.

3.1 The Allocation Algorithm

The algorithm of point's allocation performs the separation of the subsets M_a and M_d from the sets of the measured points M , after M was ordered according to voltage values $v \in [V_{min}, V_{max}]$. The flowchart of the allocation algorithm is displayed in Figure 5 and described as follows.

The allocation is based on how the points $\{(v_k, i_k)\}$ are placed in different parts of the cloud. Two types of areas can be distinguished: vertical and horizontal. In the horizontal areas the variation of the values of the absorbed current is very low, while in the vertical areas the variation is high, the points being almost vertically arranged.

In principle, the distribution of the zones is based on voltage reference values that delineate

the vertical areas from the horizontal ones. From figures 1-5 one can notice that there is one horizontal area, two vertical areas and two different reference voltages $-V_z$ and V_z . They are determined through monitoring the number of measured points (v, i) on intervals of 10 V range. When the number of points from the interval increases significantly the absolute minimum value of the interval limits becomes V_z or $-V_z$.

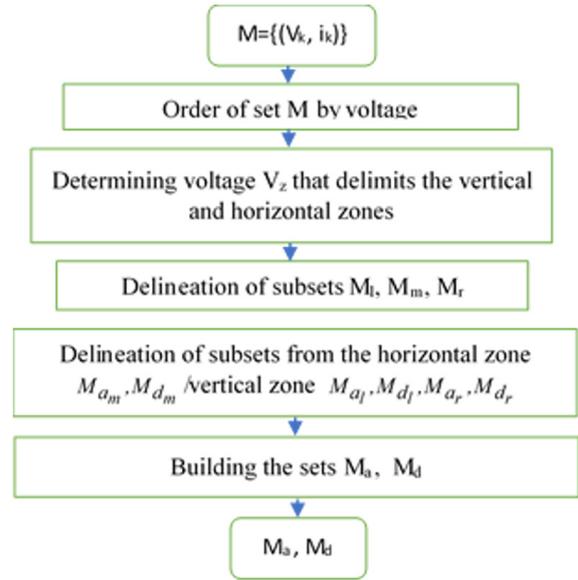


Figure 5. Flowchart of the allocation algorithm

Once the reference values are determined, the subsets M_m (horizontal area), M_l (left vertical area), and M_r (right vertical area), are delineated according to the following relations:

$$\begin{aligned} M_m = & \{(v_k, i_k) | (v_k, i_k) \in M, -V_z < v_k < V_z\} \\ M_l = & \{(v_k, i_k) | (v_k, i_k) \in M, V_{min} < v_k < -V_z, i_k < -I_{ref}\} \\ M_r = & \{(v_k, i_k) | (v_k, i_k) \in M, V_z < v_k < V_{max}, i_k > I_{ref}\} \\ M \setminus (M_l \cup M_m \cup M_r) = & \emptyset \end{aligned}$$

The conditions $i_k > I_{ref}$ and $i_k < -I_{ref}$ are used for eliminating the points from the curbed areas, i.e. from the areas in which it is hard to establish whether the dots belong to the ascending or the descending branch. For I_{ref} an acceptable value seems to be 0.1A.

After the delineation of these three areas, their sets of points are split in two subsets: one subset for the ascending branch M_{am} , and one subset for the descending branch M_{dm} . The division takes place as follows:

- For the horizontal area the average current I_{avg} is determined on 10 V intervals. Let N be

the number of the intervals, and n the serial number of an interval. The points for which $i < I_{avg}$ are allocated to the ascending branch, while the ones with values $i > I_{avg}$ belong to the descending branch. As a result the following sets are obtained:

$$M_{a_m}^n = \left\{ (v_k^n, i_k^n) \mid (v_k^n, i_k^n) \in M_m^n, i_k^n \leq I_{avg}^n \right\},$$

$$M_{a_m} = \bigcup_n M_{a_m}^n, n = \overline{1 \dots N}$$

$$M_{d_m}^n = \left\{ (v_k^n, i_k^n) \mid (v_k^n, i_k^n) \in M_m^n, i_k^n > I_{avg}^n \right\},$$

$$M_{d_m} = \bigcup_n M_{d_m}^n, n = \overline{1 \dots N}$$

- Depending on the range of variation of the consumed current two types of allocation are considered for the vertical area.
- The first allocation variant corresponds to the situation in which for the vertical area the current variation is reduced. In this case the average voltage V_{avg^*} is calculated for the points of the entire area. For the right vertical area, the dots with $v < V_{avg^*}$ are allocated to the ascending branch while those with $v > V_{avg^*}$ to the descending branch. This gives:

$$M_{a_r} = \left\{ (v_k, i_k) \mid (v_k, i_k) \in M_l, v_k \geq V_{avg_l} \right\},$$

$$M_{d_r} = \left\{ (v_k, i_k) \mid (v_k, i_k) \in M_l, v_k < V_{avg_l} \right\}$$

Similarly, for the left vertical area the sets M_{al} and M_{dl} are obtained.

- The second allocation variant corresponds to the situation when in the vertical area there is a significant electrical current variation. In this case the sets M_r and M_l are built without using the conditions $i_k > I_{ref}$ and $i_k < -I_{ref}$ and the intervals $[V_z, V_{max}]$ and $[-V_{max}, -V_z]$ are divided into subintervals of $0.5 \text{ V} - 3 \text{ V}$, in descending order towards the extremities. All the subintervals are regarded as horizontal areas. The sets M_{ar} and M_{dr} , respectively M_{al} and M_{dl} are built through allocation relations like the ones in which the sets M_{am} and M_{dm} are obtained.

In the end, the sets M_a and M_d are built as a reunion of the above-mentioned subsets:

$$M_a = M_{a_l} \cup M_{a_m} \cup M_{a_r},$$

$$M_d = M_{d_l} \cup M_{d_m} \cup M_{d_r}.$$

3.2 The Genetic Algorithm

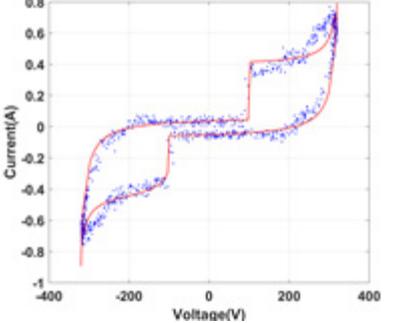
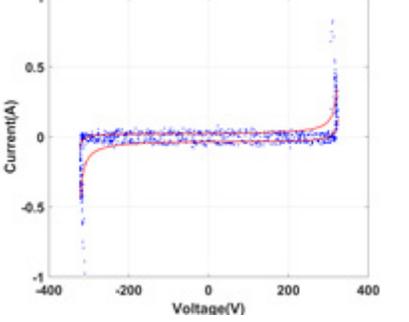
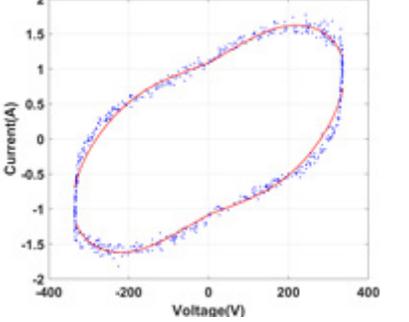
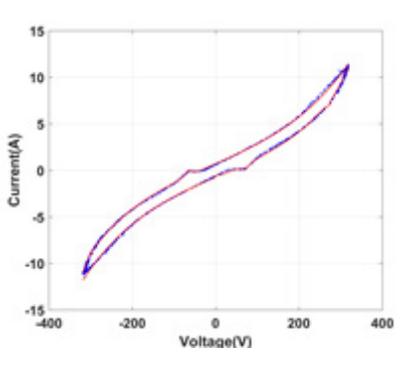
Once the two subsets M_a and M_d have been determined and the pair of support function for the class $i_a = S_{C_a}(v), i_d = S_{C_d}(v)$, and the intervals I_{Π_a}, I_{Π_d} for the parameters were chosen, the genetic algorithm is applied to determine the values of the parameters for each branch independently. An individual is represented by the set of parameters Π_a , or parameters Π_d , of the corresponding support function. As fitness function, function (16) is used, in which $i(v)$ is the regression function:

$$F_{fitness} = \frac{1}{N_p} \sum_k |i_k - i(v_k)|^{n_i} \cdot \left(1 + 0.2 \cdot e^{-\frac{v_k}{V}} \right) \quad (16)$$

In (16) i_k is the measured value of the current which corresponds to the measured voltage value v_k , $i(v_k)$ is the estimated value with the chosen support function, N_p the number of points from set M_a or M_d . The exponential term in equation (16) increases the contribution of the horizontal area in fitness. It was added because the values of the current in horizontal areas are lower than in the vertical ones. Since the household consumer is supplied at an actual voltage of 230 volts, the parameter was set on the 300- volt level. The exponent n_i contributes to balancing the involvement of the values $|i_k - i(v_k)|$ in building the fitness. In the following $n_i=1$ was considered.

The (Matlab, 2019) function $[x, fval]=ga(_)$, where x is the solution and $fval$ is the value of the fitness has been used to implement the genetic algorithm. For its running, the implicit assigned parameterization and operators of the genetic algorithm have been used. That means that the population scale = 50, when the number of the parameters is lower than 5, and 200 when this number is higher than 5; number of generations = 100 multiplied by number of parameters; elite number = 0.05; cross fraction = 0.8; migration fraction among subpopulations = 0.02 (for each 20th generation), respectively the generating of the initial population with help of random generation function, with consistent probability distribution; the scaling

Table 1. Experimental Results

Consumer/class / signature/ parameters/ restrictions/ n_{Ma}/n_{Md}	Measured points, signature's graph	Parameter values (Π_a, Π_d), / fitness_a / fitness_d	Parameter constrains (I_{Π_a}, I_{Π_d})
LCD television / C_{TgDisc} / (1) / (2.1) / (2.2) / 518 / 506		$\Pi_a = \{19.059, -0.444, 80.773, 90.157, -0.0781, 832.624\};$ $\Pi_d = \{81.715, 0.0636, 778.379, 81.895, 0.394, 433.848\} / 0.05572 / 0.05966$	$I_{\Pi_a} = I_{\Pi_d} = \{[-50, 700], [-0.6, 0.6], [50, 850], [50, 400], [-06, 0.6], [50, 900]\}$
Laptop / C_{Tg} / (4) / (5.1) / (5.2) / 440 / 466		$\Pi_a = \{223.629, 0.0372, 839.255, 355.598, 378.021\}$ $\Pi_d = \{699.906, 0.0147, 500.477, 277.691, 801.051\} / 0.05042 / 0.04817$	$I_{\Pi_a} = I_{\Pi_d} = \{[-50, 700], [-0.2, 0.2], [50, 850], [104, 400], [50, 850]\}$
Fridge / C_{ell} / (11) / (12.1) / (12.2) / 475 / 462		$\Pi = \{340.001, 1.1378, 0.0472, 5.5977\} / 0.18291 / 0.17857$	$I_{\Pi} = \{[340, 700], [1, 10], [0, 8], [1, 50], \}$
Vacuum cleaner / C_h / (13) / (15.1) / (15.2) / 521 / 502/ in (14): $\beta_{min}=\beta_{max}=0.1$		$\Pi_a = \{271.929, 0.553, 342.173, 461.323, -401.940, -0.815, 1.157, 11.868, 1.651, -48.742, 0.00672, 37.935, 14.240, 16.954, 1.246, 0.0081, 0.560, 1.874, 7.021, 0.00563, 1.999, 1.275, 1.863, 1.069, 1.548, 4.963, 6.319, 6.637, 71.842, 0.0363, 0.0424, -0.648, 0.861, 4.897, 0.407, 23.197, 10.217\}$	$I_{\Pi_a} = \{[-600, 600], *, *, *, *, [-100, 100], *, *, *, *, [-50, 50], *, *, *, *, [1, 2], *, *, *, *, [-25, 25], *, *, *, *, [50, 120], [-50, 50], *, *, *, *, [...], *, *, *, *, *, \}$ For I_{Π_d} : $[-120, -50]$ instead $[50, 120]$, =

$\Pi_d = \{395.829, 396.853, 184.563, -418.157, -298.311, -56.611, 1.261, 2.691, 94.695, -0.537, 0.00537, 8.033, 2.565, 6.522, -0.0112, -34.104, 4.865, -27.178, -8.999, -0.00626, 1.998, 1.704, 1.325, 1.733, 1.947, -0.171, -1.568, 2.115, -63.739, 0.0425, 0.0612, 0.765, 5.183, 5.442, 0.062, 2.553, 0.159\} / 0.17 / 0.14$

of the fitness function (for suitability) through rank scaling; the selection using a stochastic standard algorithm; the cross-breeding through the dispersal function; and the mutation through a feasible adaptive function.

The parameters of the support function Π_a, Π_d have been restricted to the variation intervals I_{Π_a}, I_{Π_d} stated in previous section. For the instance in which the parameters have voltage or current physical dimensions, these intervals have been established according to the variation area of the voltage and of the current. All the others intervals had been determined according to the arrangement of the points in the subsets M_a, M_d and according to the associated support functions. Consequently, after running the genetic algorithm, the values in these intervals have resulted for the parameters Π_a, Π_d .

4. Experimental results

In this section the signatures obtained for each of the 4 individual household consumers using the measurements made as presented in (Căiman & Dragomir 2018a) are described. The acquisition of the points $M = \{(v,i)\}$ was done using a sample period of approximately 2.5 supply voltage periods (≈ 20 msec).

The results are summarized in Table 1. For each case are specified: the consumer, the class, the formula of signature from section 2, the measured point cloud, the resulted signature graph,

The numbers of points n_{M_a} and n_{M_d} allocated to the subsets M_a and M_d , the value of the fitness function (*fitness_a* for the ascending branch and *fitness_d* for the descending one), the parameters of the signatures and the domains imposed on parameters for determining their value with genetic algorithms. The value of the fitness function should be also considered as a measure of the quality of the performed approximation using the support function associated with the class of the individual consumer.

Each signature was obtained in 2 stages: i) the allocation of the subsets M_a and M_d ; ii) the computing of the set of parameters that

accompany the signature by applying the genetic algorithm.

For the hybrid class consumers, the second stage was achieved in 2 steps. In the first step formulae (12) were applied for $k = 0$. In the second step the parameters were initialized with the values obtained in the first step and $k \neq 0$ was taken. This working method addresses both the correction nature of the term affected by k , and the fact that the correction is done only on a certain area of the domain of the voltages (in the area of the angular point illustrated in Figure 4).

Matlab has been used as working environment. The settings for the genetic algorithm were the ones mentioned in section 3.

In the case of a television device (class C_{TgDisc}) the values of the fitness function from the two branches are very similar, both ranging below 0.06. For the laptop (class C_{Tg}) the fitness values are close to 0.05. For the fridge (class C_{ell}) the fitness values are approximately three times higher than those of the other 2 cases. These values are comparable to the fitness values for the vacuum cleaner (class C_h).

For this latter case, the significance of the notations in the table is the following: - the underlined parameters correspond to the correction $k \cdot f_{C2}(V_x)$, - for the parameters corresponding to the blank intervals no variation limits have been required, - symbol “*” states for the intervals that are identical to the last clearly completed interval. The parameters Π_d have been written under the table. The parameters resulted from applying the genetic algorithm as mentioned in section 3.

5. Conclusion

For the individual household consumers, voltage-current signatures can be determined using the parameterizable functions set up through regression analysis. Due to the fact that the parameterizable functions are non-linear with respect to the parameters, the regression is non-linear.

For a lot of consumers such as laptops, fridges, televisions and vacuum cleaners, the class

signatures are paired: the continuous tangent, discontinuous tangent, ellipse and hybrid classes are presented in the paper. In order to determine the values of the parameters the proposed method can be applied based on regression calculations in which genetic algorithms are used.

The experimental results have led to signatures that approximate very well the measured

values of the current-voltage pairs. It should be mentioned that the signatures from the 4 classes can also be identified when the total consumption of several household appliances connected in parallel is measured. The result is to be presented in a future paper.

REFERENCES

1. Azaza, M. & Wallin, F. (2017). Evaluation of classification methodologies and Features selection from smart meter data, *Energy Procedia*, 142, 2250-2256.
2. Bouhouras, A. S. et al. (2019). A NILM Algorithm with Enhanced Disaggregation Scheme under Harmonic Current Vectors, *Energy and Buildings*, 183, 392-407.
3. Căiman, D. V. & Dragomir, T. L. (2018a). Non-Intrusive load monitoring: analytic expression as load signature. In *12th International Symposium on Applied Computational Intelligence and Informatics (SACI 2018)* (pp. 15-20).
4. Căiman, D. V. & Dragomir, T. L. (2018b). Two voltage-current load signature classes for residential consumers. In *22nd International Conference on System Theory, Control and Computing (ICSTCC)* (pp. 176-181).
5. De Baets, L. et al. (2018). VI-Based Appliance Classification Using Aggregated Power Consumption Data. In *IEEE International Conference on Smart Computing (SMARTCOMP)* (pp. 179-186).
6. EEA (2018). Final energy consumption by sector and fuel. Available at: <<https://www.eea.europa.eu/data-and-maps/indicators/final-energy-consumption-by-sector-9/assessment-4>>, created: 29 Nov 2018, accessed: 20 March 2019.
7. Ghorbanpour, S. & Mallipeddi, R. (2018). Significance of Classifier and Feature Selection in Automatic Identification of Electrical Appliances. In *IEEE International Conference on Systems, Man, and Cybernetics* (pp. 4184-4189).
8. Hart, G. W. (1992). Nonintrusive Appliance Load Monitoring. In *Proceedings of the IEEE*, 80(12) (pp. 1870-1891).
9. Hassan, T., Javed, F. & Arshad, N. (2014). An Empirical Investigation of V-I Trajectory Based Load Signatures for Non-Intrusive Load Monitoring, *IEEE Transactions on Smart Grid*, 5(2), 870-878.
10. Held, P. et al. (2016). Frequency invariant transformation of periodic signals (FIT-PS) for signal representation in NILM. In *42nd Annual Conference of the IEEE Industrial Electronics Society (IECON 2016)* (pp. 5149-5154).
11. Huang, B. et al. (2017). Non-intrusive load monitoring system performance over a range of low frequency sampling rates. In *IEEE 6th International Conference on Renewable Energy Research and Applications (ICRERA)* (pp. 505-509).
12. Kialashaki, A. & Reisel, J. R. (2013). Modeling of the energy demand of the residential sector in the United States using regression models and artificial neural networks, *Applied Energy*, 108, 271-280.
13. Lam, H. Y., Fung, G. S. K. & Lee, W. K. (2007). A novel method to construct taxonomy of electrical appliances based on load signature, *IEEE Transaction on Consumer Electronics*, 53(2), 653-660.
14. Longjun Wang, A. et al, (2018). Non-intrusive load monitoring algorithm based

-
- on features of V–I trajectory, *Electric Power Systems Research*, 157, 134-144.
15. Matlab (2019). ga - Function using genetic algorithm. Available at: <https://www.mathworks.com/help/gads/ga.html?s_tid=doc_ta>
16. Minghua, S. et al. (2017). Regression modeling based on improved genetic algorithm, *Tehnički vjesnik*, 24(1), 63-70.
17. Opreș, A. (2018). Balanța energetică și structura utilajului energetic, p. 8, INS. Available at: <http://www.insse.ro/cms/sites/default/files/field/publicatii/balanta_energetica_si_structura_utilajului_energetic_in_anul_2017.pdf>
18. Vlada, M. (2012). Modele neliniare. Teorie și aplicații. In *Conferința Națională de Învățământ Virtual, ediția a X-a* (pp. 57-65).
19. Zhuang, M., Shahidehpour, M. & Li, Z. (2018). An Overview of Non-Intrusive Load Monitoring: Approaches, Business Applications, and Challenges. In *International Conference on Power System Technology (POWERCON)* (pp. 4291-4299).