A new formulation of state estimation in distribution systems including demand and generation states

Patricia Rousseaux
Dept. Elec. Eng. & Comp. Sc.
University of Liège, Belgium
p.rousseaux@ulg.ac.be

Jean-François Toubeau
Zacharie De Grève
François Vallée
Dept. Elec. Eng.
University of Mons, Belgium
francois.vallee@umons.ac.be

Mevludin Glavic
Dept. Elec. Eng. & Comp. Sc.
University of Liège, Belgium
mevludin.glavic@ulg.ac.be

Thierry Van Cutsem
Fund for Scientific Research
Dept. Elec. Eng. & Comp. Sc.
University of Liège, Belgium
t.vancutsem@ulg.ac.be

Abstract—**Distribution system state estimation faces a major difficulty: the lack of real-time measurements. This imposes to add information, usually pseudo-measurements from historical data. This paper proposes a different, novel formulation of state estimation relying on the classification of loads into components (e.g. residential, commercial, etc.) and accounting for dispersed generation. The approach "by-passes" the use of pseudo-measurements by expressing the medium-voltage bus injections as functions of a small number of active power components at low-voltage level, treated as additional state variables. The injections at medium-voltage buses become equality constraints. A procedure to identify the above functions is detailed, which takes advantage of data collected by smart meters.**

Index Terms—**distribution systems, state estimation, load components, dispersed generation, smart meters, least squares**

I. MOTIVATION AND BACKGROUND

State estimation has become a standard monitoring function in transmission systems [1]. Nowadays, it is envisaged in Medium Voltage (MV) distribution systems in support of active network management [2]. However, Distribution System State Estimation (DSSE) faces a major difficulty: the relative scarceness of real-time measurements, leading to low redundancy. It is thus required to add information in the form of pseudo-measurements. Most of the time, the latter relate to *MV bus injections*, i.e. the powers drawn from the distribution network.

Pseudo-measurements are usually obtained from historical data such as standardized load profiles, weather and time-of-day data, customer billing information, etc. [2], [3]. Three methods to compute the associated variances are: assumed variance, normal distribution fitting, and correlation model [3]. Gaussian mixture model [2], [4], artificial neural networks [4], and machine learning algorithms based on automated metering infrastructure [5] have also been proposed. Once the pseudo-measurements have been retrieved from the available data, they are usually treated as statistically independent data.

In fact few approaches acknowledge that a strong correlation exists among loads within the same area. The early reference [6] considered aggregation of loads in demand areas defined as connected sub-networks without real-time measurements,

connected to other demand areas through branches with real-time measurements. The central idea was to first estimate aggregated loads and then share them among individual loads according to pre-specified scaling factors. The correlation model in [3] relies on a regression analysis to determine the degree of dependency of non-measured loads with real-time measurements.

This paper proposes a novel formulation of DSSE based on the assumption of a strong correlation among loads of the same type (e.g. residential, commercial, etc.) within the same area. It also accounts for distribution generation, which is important in view of the rapid growth of photo-voltaic installations connected to Low Voltage (LV) networks. The approach "by-passes" the use of pseudo-measurements by expressing the MV bus injections as functions of a small number of active power components at LV level, treated as additional state variables. The injections at MV buses become equality constraints.

The paper is organized as follows. The new state estimation formulation is presented in Section II, followed by an illustrative example in Section III. A procedure to identify the relations between MV bus injections and LV active power components is detailed in Section IV, while simulation results using data from existing LV networks are given in Section V. The main points are summarized in Section VI.

II. NEW STATE ESTIMATION FORMULATION

A. Demand and dispersed generation components

For a single connected distribution grid, the whole active power consumed at LV level is split into c_D *demand components*, one for each type of load, e.g. residential, tertiary sector, industrial, etc.

Similarly, the whole active power generated at LV level is split into c_G *dispersed generation components*. In today's LV networks, the dominant Dispersed Deneration (DG) units are photo-voltaic installations, but other types can be included.

Let us define $c = c_D + c_G$ and let L be the number of MV bus injections corresponding to connections to LV networks.

An illustrative example is given in Fig. 1 for a simple, 8-bus MV system. $L = 4$ since there are connections to LV networks at buses 3, 4, 5 and 7. The LV grid fed by bus 4
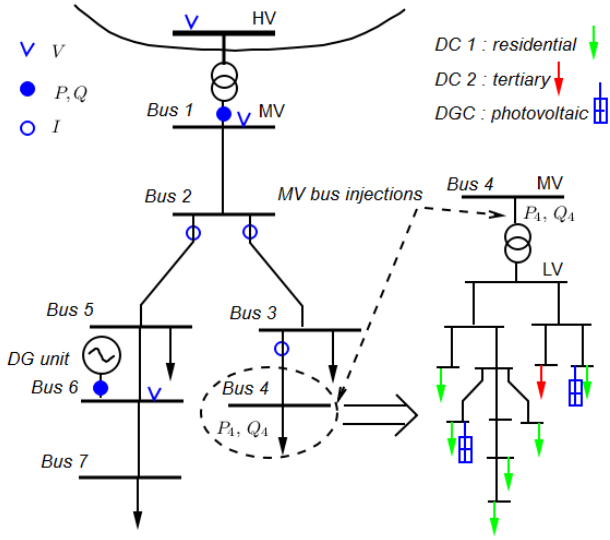
Fig. 1. Example of MV grid with real-time measurements, and detailed view of the LV grid connected to bus 4, corresponding to the injection at this bus

is also sketched. It supplies two types of loads, respectively residential and tertiary sector, and hosts photo-voltaic units. The same holds true in the LV networks connected to buses 3, 5 and 7. In this example, $c_D = 2$ and $c_G = 1$.

The demand and dispersed generation components contribute to the L MV bus injections of the distribution system. Let π_i ($i = 1, \ldots, c$) be the active power consumed by all loads, or produced by all generation units of the same component, with $\pi_i > 0$ for a demand component and $\pi_i < 0$ for a dispersed generation component. These variables are grouped into: $\quad \boldsymbol{\pi} = [\pi_1, \pi_2, \ldots, \pi_c]^T$.

It is assumed that the relationship between the active power injection P_ℓ at the ℓ-th MV bus and the π_i variables can be expressed through a general nonlinear model:

$$P_\ell = \varphi_\ell(\boldsymbol{\pi}), \quad \ell = 1, \ldots, L \tag{1}$$

and similarly for the reactive power injection:

$$Q_\ell = \psi_\ell(\boldsymbol{\pi}), \quad \ell = 1, \ldots, L \tag{2}$$

The simplest forms of the above relations would consist of linear models of the type:

$$P_\ell = \sum_{i=1}^{c} \alpha_{\ell i} \pi_i \qquad Q_\ell = \sum_{i=1}^{c} \beta_{\ell i} \pi_i \tag{3}$$

where each demand or dispersed generation component contributes as a fraction of the total injection. This linear model does not easily account for losses in LV networks and MV/LV transformers. A more accurate model can be obtained with the quadratic formulas:

$$P_\ell = \sum_{i=1}^{c} \alpha_{\ell i} \pi_i + \alpha_\ell \left(\sum_{i=1}^{c} \pi_i \right)^2 \tag{4}$$

$$Q_\ell = \sum_{i=1}^{c} \beta_{\ell i} \pi_i + \beta_\ell \left(\sum_{i=1}^{c} \pi_i \right)^2 \tag{5}$$

Clearly, two issues have to be addressed: (i) how to identify the $(\alpha_{\ell i}, \alpha_\ell, \beta_{\ell i}, \beta_\ell)$ parameters of the above model from LV network models and data, and (ii) how Eqs. (4,5) can be handled in the state estimation formulation. Section IV is devoted to the first aspect, while the second is addressed in the remaining of this section.

B. New formulation of DSSE

Real-time measurements involving voltage magnitudes, active and reactive power flows and injections, and current magnitudes, are described by the usual model:

$$\boldsymbol{z} = \boldsymbol{h}(\boldsymbol{x}) + \boldsymbol{e} \tag{6}$$

where, for an N-bus system, $\boldsymbol{x} = [V_{x1}, V_{y1}, \ldots, V_{xN}, V_{yN}]^T$ is the $2N$-dimensional state vector of rectangular components of bus voltages, $\boldsymbol{z}$ is the m-dimensonal measurement vector, $\boldsymbol{h}$ is the vector of nonlinear functions relating error-free measurements to state variables, and $\boldsymbol{e}$ is the vector of measurements errors, assumed to obey a Gaussian distribution with zero mean and known diagonal covariance matrix $\boldsymbol{R} = diag(\sigma_1^2 \ldots \sigma_m^2)$.

In the standard approach of DSSE, pseudo-measurements are added on the active and reactive powers to the L MV-bus injections (with larger σ_i values to reflect the lower accuracy). In the proposed approach, on the contrary, each of these injections is replaced by the model (1,2). This leads $2L$ "exact" equality constraints. At the same time, the state vector is extended to include $\boldsymbol{\pi}$ in addition to $\boldsymbol{x}$.

The above modelling and formulation yields a higher ratio of the number of measurements to the number of unknowns, since $c \ll L$ in practice. At the same time, the estimate of $\boldsymbol{\pi}$ provides valuable real-time information to the Distribution System Operator (DSO).

The estimates of $\boldsymbol{x}$ and $\boldsymbol{\pi}$ are obtained by solving the following constrained Weighted Least Squares (WLS) problem:

$$\min_{\boldsymbol{x}, \boldsymbol{\pi}, \boldsymbol{r}} \quad \frac{1}{2} \boldsymbol{r}^T \boldsymbol{R}^{-1} \boldsymbol{r} \tag{7}$$

$$\text{subject to} \quad \boldsymbol{f}_0(\boldsymbol{x}) = \boldsymbol{0} \tag{8}$$

$$\boldsymbol{f}_\pi(\boldsymbol{x}, \boldsymbol{\pi}) = \boldsymbol{0} \tag{9}$$

$$\boldsymbol{r} - \boldsymbol{z} + \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0} \tag{10}$$

where $\boldsymbol{r}$ is the vector of measurement residuals.

The equality constraints (8) relate to zero current injections at transit buses with no load and no generation:

$$\text{at the } j\text{-th bus:} \quad I_{xj}(\boldsymbol{x}) = 0 \qquad I_{yj}(\boldsymbol{x}) = 0 \tag{11}$$

as well as the voltage phase angle at the reference bus r :

$$V_{yr} = 0 \tag{12}$$

The equality constraints (9) express the active and reactive power balance at the L buses connected to an LV network, i.e. Eqs. (1,2) rewritten as:

$$-\varphi_\ell(\boldsymbol{\pi}) + P_\ell(\boldsymbol{x}) = 0 \qquad \ell = 1, \ldots, L \tag{13}$$

$$-\psi_\ell(\boldsymbol{\pi}) + Q_\ell(\boldsymbol{x}) = 0 \qquad \ell = 1, \ldots, L \tag{14}$$

where $P_\ell(\boldsymbol{x})$ and $Q_\ell(\boldsymbol{x})$ are the power injections expressed as functions of the MV bus voltages.

The optimization problem (7 - 10) is solved using Hachtel's augmented matrix method, for its high numerical stability [7]. Details of implementation are outside the scope of this paper.

III. ILLUSTRATIVE EXAMPLE OF THE DSSE ALGORITHM

The proposed DSSE algorithm is illustrated on the 8-bus system detailed in Fig. 1. One DG unit is connected to bus 6. Buses 1 and 2 have zero current injection.

A reference system loading has been considered by choosing the π_i values of residential loads, tertiary-sector loads, and photo-voltaic production respectively: $\pi_1 = 12.4$, $\pi_2 = 3.9$, $\pi_3 = -4.7$ MW. The linear model (3) has been adopted, for simplicity. It has been used to determine the P_ℓ and Q_ℓ power injections at the $L = 4$ MV buses. The $\alpha_{\ell i}$, $\beta_{\ell i}$ parameters have been set to plausible values, reported in Table I. A unit power factor has been assumed for photo-voltaic production, leading to $\beta_{\ell 3} = 0$. It has also been assumed that no photo-voltaic production is present in the LV networks connected to buses 4 and 5, and no tertiary-sector load in the LV network connected to bus 5. Finally, the active and reactive productions of the DG unit have been chosen. All these data have been used in a power flow calculation on the MV grid, providing the "exact" system state.

TABLE I
PARAMETERS OF THE MV BUS INJECTION MODELS

| At bus $\ell =$	$\alpha_{\ell 1}$	$\alpha_{\ell 2}$	$\alpha_{\ell 3}$	$\beta_{\ell 1}$	$\beta_{\ell 2}$	$\beta_{\ell 3}$
3	0.42	0.37	0.63	0.11	0.21	0
4	0.14	0.54	0	0.16	0.20	0
5	0.23	0	0	0.19	0	0
7	0.26	0.14	0.42	0.08	0.20	0

The set of $m = 10$ real-time measurements is shown on the one-line diagram of Fig. 1. Note that this set of real-time data is not sufficient for a conventional DSSE : the number of state variables to be estimated, minus the number of constraints (8) is $16 - 5 = 11$, which exceeds the number of measurements. On the other hand, with the addition of the $2L = 8$ constraints (9), the number of real-time measurements is sufficient, and state estimation can be performed.

The measured values were simulated by adding to the exact values Gaussian noises with zero mean and standard deviations representative of the measurement accuracy. Two "noise levels" were considered, as detailed in Table II.

TABLE II
STANDARD DEVIATIONS OF MEASUREMENT NOISES

	Voltage magn. (pu)	Current magn. (pu)	P/Q power flow (MW)	P/Q power injection (MW)
Low noise	0.001	0.05	0.01	0.01
High noise	0.005	0.1	0.05	0.05

First, it has been assessed that the DSSE algorithm correctly converges to the exact π_i values when fed with error-free measurements. The results are shown in Fig. 2(a). The exact π_i values are found, with a good convergence of Hachtel's
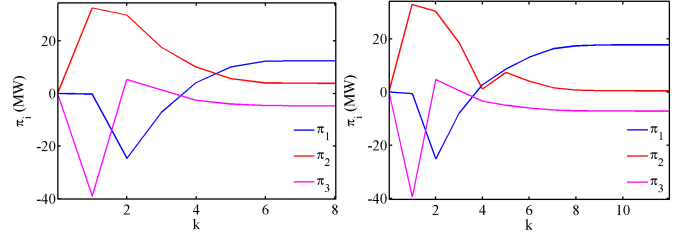
(a) with exact measurements   (b) with noisy measurements
Fig. 2. Convergence of π variables (k is the iteration counter)

augmented matrix method. As can be seen in Fig. 2(b), only a slightly slower convergence is observed in case measurements are affected by the high noise.

The estimation accuracy has been checked by computing the following indices for the voltage magnitudes V (in pu), voltage phase angles θ (in degree), active and reactive power injections P_ℓ, Q_ℓ (in MW and Mvar):

$$S_y = \sum_i |\hat{y}_i - y_{ti}| \quad \text{and} \quad |\Delta y|_{max} = \max_i |\hat{y}_i - y_{ti}| \quad (15)$$

where $\hat{y}_i$ is the estimated and y_{ti} the true value. Estimation errors have also been computed for the π_i variables, as well as for the net active load at LV level: $P_{LV} = \sum_{i=1}^{3} \pi_i$. The results are gathered in Table III. As expected, the estimation accuracy is impacted by the quality of the measurements. Although the π_i variables are the most affected, it is noteworthy that the accuracy of P_{LV} is very satisfactory.

TABLE III
ESTIMATION ACCURACY (IN PU, DEG, MW AND MVAR)

	S_V	S_θ	$|\Delta V|_{max}$	$|\Delta\theta|_{max}$
Low noise	$1.6\,10^{-3}$	0.23	$5\,10^{-4}$	0.07
High noise	$7.9\,10^{-3}$	0.91	$2\,10^{-3}$	0.24

	S_{P_ℓ}	S_{Q_ℓ}	$|\Delta P_\ell|_{max}$	$|\Delta Q_\ell|_{max}$
Low noise	0.86	0.49	0.37	0.31
High noise	1.96	1.60	1.09	1.03

	$|\Delta\pi_1|$	$|\Delta\pi_2|$	$|\Delta\pi_3|$	$|\Delta P_{LV}|$
Low noise	1.58	1.11	0.6	0.13
High noise	5.33	3.45	2.48	0.60

IV. IDENTIFICATION OF THE MV BUS INJECTION MODELS

The identification of the MV bus injection model (4,5) relies on a preliminary analysis and classification of LV loads and generations into the already mentioned c categories. Furthermore, advantage can be taken from the data collected by smart meters monitoring some of the loads or generations. The usual period of this reporting is one quarter of an hour.

The procedure can be decomposed into the four steps described hereafter. For simplicity, the description mainly refers to loads but, unless otherwise mentioned, the same procedure is applied to the dispersed generations.

A. Characterizing the load dispersion from smart meter data

Let $\mathcal{N}_i(i = 1, \ldots, c)$ be the set of all LV nodes associated to π_i, and $\mathcal{J}_i$ the subset of $\mathcal{N}_i$ corresponding to nodes equipped with smart meters; thus, $\mathcal{J}_i \subset \mathcal{N}_i$. Note that a bus can belong to more than one set $\mathcal{N}_i$ or $\mathcal{J}_i$, for instance when a domestic consumer is equipped with a photo-voltaic installation.

Let us consider a given quarter of an hour (though, the dependence on time is omitted in the notation, for clarity). For the i-th demand component ($i = 1, \ldots, c$), let us define:

N_i the number of loads in the component ($N_i = \#\mathcal{N}_i$)

N_i^{sm} the number of loads in the component equipped each with a smart meter ($N_i^{sm} = \#\mathcal{J}_i$)

$E_{j,i}$ the energy consumed during the quarter of an hour by the j-th load ($j = 1, \ldots, N_i$)

$E_{j,i}^{tot}$ the total energy consumed annually by the same load.

The objective of the first step is to characterize the statistical dispersion of the loads pertaining to the same demand component, based on the measurements collected by smart meters every quarter of an hour.

In order to combine the measurements relative to loads of different "sizes", the $E_{j,i}$ values are normalized according to:

$$\bar{E}_{j,i} = \frac{E_{j,i}}{E_{j,i}^{tot}} \qquad j = 1, \ldots, N_i^{sm}; \; i = 1, \ldots, c \qquad (16)$$

The N_i^{sm} values $\bar{E}_{j,i}$ are considered to be realizations of a random variable $\bar{E}_i$ that characterizes the dispersion of the individual loads of the i-th demand component. The cumulative distribution function of $\bar{E}_i$, denoted $F(\bar{E}_i)$, is thus built from the collected values. It is sketched in Fig. 3, where:

$$\bar{E}_i^{min} = \min_{j=1,\ldots,N_i^{sm}} \bar{E}_{j,i} \text{ and } \bar{E}_i^{max} = \max_{j=1,\ldots,N_i^{sm}} \bar{E}_{j,i} \quad (17)$$

are respectively the lowest and the highest among the N_i^{sm} collected values of $\bar{E}_{j,i}$. The median of this distribution is denoted by $\bar{E}_i^{med}$. Note that there is one function $F(\bar{E}_i)$ per quarter of an hour, and per demand component ($i = 1, \ldots, c$).
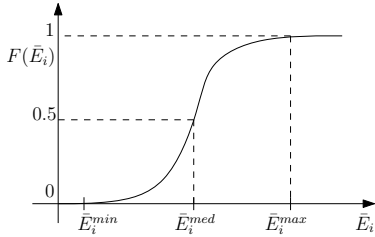

Fig. 3. Cumulative distribution function of normalized energy

The same procedure is applied to the data collected from photo-voltaic installations. The only difference lies with $E_{j,i}^{tot}$, which is the total energy produced over the whole year.

B. Building a large training set of operating points

The objective of the second step is to build a large number r of operating points, according to a non-chronological Monte-Carlo procedure, in which the powers at the various LV buses are varied in accordance with the information gathered from the smart meters. Each operating point is obtained as follows.

First, the values of $\pi_1, \ldots, \pi_c$ are drawn at random. Each variable π_i is assumed statistically independent of the other π_j's, and uniformly distributed in some range $[0 \; \pi_i^{max}]$. To keep the sample size reasonably small and still have satisfactory accuracy, a variance reduction sampling is needed. The Latin Hypercube sampling technique [8] has been chosen for its simplicity and ease of implementation.

Various choices are possible for π_i^{max}, for instance:

$$\pi_i^{max} = N_i \max_{j=1,\ldots,N_i^{sm}} \frac{E_{j,i}}{\Delta t} \qquad (18)$$

where Δt is the duration of a quarter of an hour. This formula assumes that, at the most loaded operating point, all loads have the maximum power observed from the smart meters, during the quarter of hour of concern.

Next, it is assumed that all loads pertaining to the same component obey the same statistical behavior. In other words, the normalized energy of the k-th load, denoted by $\bar{E}_{k,i}$ ($k = 1, \ldots, N_i$), is considered to be a random variable with the cumulative distribution function $F(\bar{E}_i)$ that has been extracted at Step A (see also Fig. 3) from the set $\mathcal{J}_i$ of loads equipped with smart meters.

The power $P_{k,i}$ consumed by the k-th load ($k = 1, \ldots, N_i$) is obtained from $\bar{E}_i$ through the linear mapping:

$$P_{k,i} = \frac{E_{k,i}^{tot}}{\sum_{j=1}^{N_i} E_{j,i}^{tot}} \pi_i \frac{\bar{E}_i}{\bar{E}_i^{med}} \qquad (19)$$

According to this formula, $P_{k,i}$ is a random variable with median equal to the fraction $E_{k,i}^{tot}/(\sum_{j=1}^{N_i} E_{j,i}^{tot})$ of the total power π_i consumed by the i-th demand component. This is coherent with the assumption that all loads of the same components behave similarly.

Thus, for each load, a number is drawn at random from a uniform distribution between 0 and 1. This number is assigned to $F(\bar{E}_i)$ and, from the curve in Fig. 3, the corresponding value of $\bar{E}_i$ is found. The latter is used in Eq. (19) to obtain the value of $P_{k,i}$.

If $\bar{E}_i$ was equal to $\bar{E}_i^{med}$, it is easily seen from Eq. (19) that the total power consumed in the i-th demand component:

$$\pi_i^\star = \sum_{k=1}^{N_i} P_{k,i} \qquad (20)$$

would be equal to the value initially assumed for π_i. However, since each power $P_{k,i}$ varies around its median, the power $\pi_i^\star$ differs from the initially assumed π_i, under the effect of the load power dispersion.

C. Computing the MV bus injections for each operating point

Let $\mathcal{N}$ be the whole set of LV nodes with load and/or generation ($\mathcal{N} = \cup_i \mathcal{N}_i$). Let N_{LV} be the number of nodes in $\mathcal{N}$ ($N_{LV} = \#\mathcal{N}$). For each of the r operating points generated at Step C, the active powers at the N_{LV} nodes are known by adding the contributions of the various demand and dispersed generation components; at the n-th bus ($n = 1, \ldots, N_{LV}$):

$$P_n = \sum_{i=1}^{c} P_{n,i} \qquad (21)$$

The reactive power is obtained assuming known power factors $\cos\varphi_i$ for the various components:

$$Q_n = \sum_{i=1}^{c} P_{n,i} \tan\varphi_i \qquad (22)$$

Using these nodal powers, a power flow computation is performed on the LV networks together with their MV/LV transformers. This yields the active and reactive powers entering the MV/LV transformers, which are the MV bus power injections P_ℓ and Q_ℓ ($\ell = 1, \ldots, L$) considered in Section II.

Three-phase balanced operation is assumed in that power flow computation, for simplicity.

D. Identifying the load model coefficients

At the ℓ-th MV bus ($\ell = 1, \ldots, L$), it is assumed that the active and reactive power injections are related to the $\pi_i^\star$ values through the quadratic models (4,5):

$$P_\ell = \sum_{i=1}^{c} \alpha_{\ell i} \pi_i^\star + \alpha_\ell \left(\sum_{i=1}^{c} \pi_i^\star \right)^2 \quad (23)$$

$$Q_\ell = \sum_{i=1}^{c} \beta_{\ell i} \pi_i^\star + \beta_\ell \left(\sum_{i=1}^{c} \pi_i^\star \right)^2 \quad (24)$$

At each of the r operating points generated at Step C, Eqs. (23, 24) can be written with the $\pi_i^\star$ variables replaced by their corresponding values. This leaves linear relations between P_ℓ (resp. Q_ℓ) and the various α (resp. β) coefficients. The active power relations can be written in matrix form as:

$$
\begin{bmatrix} P_\ell^{(1)} \\ P_\ell^{(2)} \\ \vdots \\ P_\ell^{(r)} \end{bmatrix} = \begin{bmatrix} \pi_i^{\star(1)} & \cdots & \pi_c^{\star(1)} & \left(\sum_{i=1}^{c} \pi_i^{\star(1)}\right)^2 \\ \pi_i^{\star(2)} & \cdots & \pi_c^{\star(2)} & \left(\sum_{i=1}^{c} \pi_i^{\star(2)}\right)^2 \\ \vdots & \vdots & \vdots & \vdots \\ \pi_i^{\star(r)} & \cdots & \pi_c^{\star(r)} & \left(\sum_{i=1}^{c} \pi_i^{\star(r)}\right)^2 \end{bmatrix} \begin{bmatrix} \alpha_{\ell 1} \\ \vdots \\ \alpha_{\ell c} \\ \alpha_\ell \end{bmatrix}
$$
$$(25)$$

where the upper script in parentheses denotes the number of the operating point. This can be rewritten in compact form as:

$$\boldsymbol{p}_\ell = \boldsymbol{\Pi} \, \boldsymbol{\alpha}_\ell \quad (26)$$

where $\boldsymbol{p}_\ell$ is a vector of dimension r, $\boldsymbol{\Pi}$ a matrix of dimensions $r \times (c+1)$, and $\boldsymbol{\alpha}_\ell$ the vector of $c+1$ unknowns. The least-square solution of this overdetermined system is given by:

$$\hat{\boldsymbol{\alpha}}_\ell = \left(\boldsymbol{\Pi}^T \boldsymbol{\Pi} \right)^{-1} \boldsymbol{\Pi} \, \boldsymbol{p}_\ell \quad (27)$$

The corresponding reactive power relations are:

$$\boldsymbol{q}_\ell = \boldsymbol{\Pi} \, \boldsymbol{\beta}_\ell \quad (28)$$
$$\hat{\boldsymbol{\beta}}_\ell = \left(\boldsymbol{\Pi}^T \boldsymbol{\Pi} \right)^{-1} \boldsymbol{\Pi} \, \boldsymbol{q}_\ell \quad (29)$$

which involve the same matrix $\boldsymbol{\Pi}$ as Eqs. (27) and (29).

The procedure is repeated at each MV bus.

V. SIMULATION RESULTS: MV BUS INJECTION MODELS

A. Available data and scenarios

The procedure presented in the previous section has been applied to a simple but realistic LV system for which data from smart meters were available. The network is made up of four non connected LV sub-networks with respectively 35, 27, 21 and 19 nodes. Each of them is connected to an MV bus through an MV/LV transformer (thus $L = 4$). The corresponding area

TABLE IV
VALUES OF R^2 TO ASSESS THE GOODNESS OF FIT OF THE MODEL

quarter of an hour No	variant #1	variant #2	variant #3
MV-bus active power P_1			
12	0.8561	0.8703	0.6533
40	0.6165	0.7769	0.7256
60	0.7948	0.8291	0.7976
90	0.7617	0.7669	0.7430
mean	0.7852	0.8341	0.7674
MV-bus reactive power Q_1			
12	0.8514	0.8680	0.6446
40	0.5901	0.7604	0.7111
60	0.7805	0.8213	0.7924
90	0.7496	0.7545	0.7353
mean	0.7713	0.8253	0.7594

encompasses 97 residential consumers, among which 20 are equipped with rooftop photo-voltaic units, and five consumers are from the tertiary sector with an installed peak power much larger than the residential customers.

The procedure has been applied to three variants, which differ by the way of splitting the total LV active power.

In variant #1, two demand ($c_D = 2$) and no dispersed generation ($c_G = 0$) components are considered. The two associated variables, π_1 and π_2, represent the net consumptions of residential households and tertiary sector, respectively.

In variant #2, two demand ($c_D = 2$) and one dispersed generation ($c_G = 1$) components are considered. The corresponding variables are $\pi_1, \pi_2 > 0$ and $\pi_3 < 0$.

Finally, in variant #3, the residential consumption is divided into two components, based on their annual energy consumptions. The corresponding variables are $\pi_1 > 0$ (for small loads), $\pi_2 > 0$ (for large loads), $\pi_3 > 0$ (for tertiary sector), and $\pi_4 < 0$ (for photo-voltaic generation).

The simulations have been performed for a typical day of the month of July. That month was selected in order to have an important impact of photo-voltaic generation. A different model is computed for each of the 96 quarters of an hour.

B. Goodness of fit of models

The goodness of fit of the models (23, 24) has been assessed through the *coefficient of determination* [9]:

$$R^2 = 1 - \frac{SS_E}{SS_T} \quad (30)$$

where SS_E is the sum of squared errors between the data and the fitted function, and SS_T the sum of squared differences between the components of $\boldsymbol{p}_\ell$ (resp. $\boldsymbol{q}_\ell$) and their average. If there is a good fitting between the data and the model, SS_E is much smaller than SS_T and R^2 is close to one. Conversely, if the quality of the model worsens, the value of R^2 decreases towards zero.

The values of R^2 are shown in Table IV for respectively the active and reactive power injections at the first MV bus, whose MV/LV transformer serves 35 LV consumers. Note that a different model has been fitted to the data of each of the 96 quarters of an hour in the day; only a representative sample
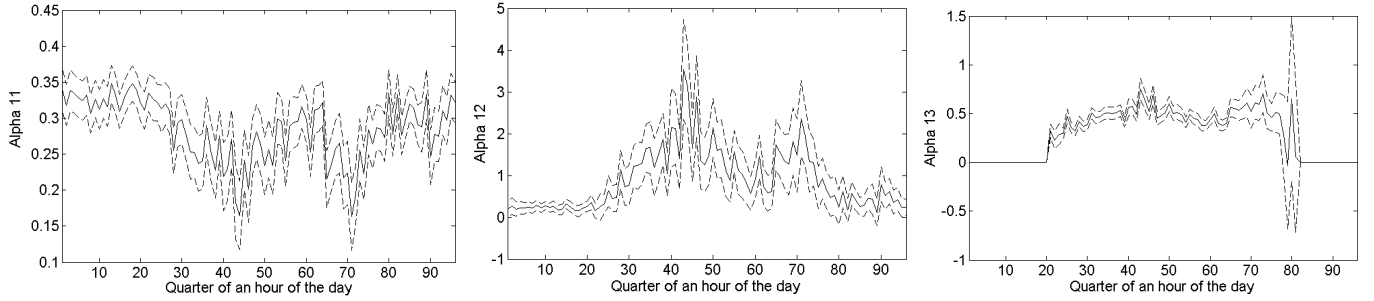
Fig. 4. Time evolution of coefficients α_{11}, α_{12} and α_{13} in Eq. (31) and their 95% confidence intervals
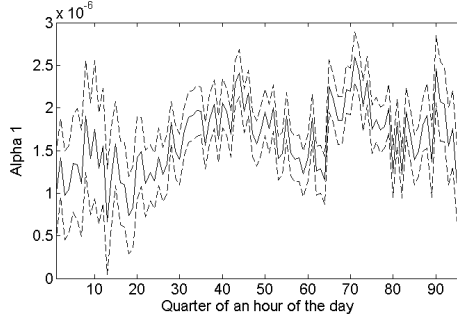

Fig. 5. Time evolution of α_1 in Eq. (31) and its 95% confidence interval

of four of them is reported in these tables, together with the mean value of R^2 over the whole day.

It can be seen that the best results are given by variant #2. One could have expected the higher number of components (i.e. of π_i variables) of variant #3 to yield a better model fitting. On the contrary, the results tend to indicate that the splitting of the residential consumption contemplated in variant #3 was not appropriate. Such results also point out the need for a systematic clustering procedure to define meaningful load and dispersed generation components.

C. Example of model parameters

For variant #2, and for the first MV bus, Figs. 4 and 5 show the 96 successive values of the coefficients of the model (23), which takes on the form:

$$P_1 = \alpha_{11}\pi_1 + \alpha_{12}\pi_2 + \alpha_{13}\pi_3 + \alpha_1(\pi_1 + \pi_2 + \pi_3)^2 \quad (31)$$

The plots also show the corresponding 95% confidence intervals.

A statistical test has been also performed to assess the significance of the various $\alpha_{\ell 1} \ldots \alpha_\ell$ coefficients [9]. The null hypothesis is that a coefficient has little or no influence on the fitted model. This test has been applied to the four coefficients in Eq. (31) at the 96 quarters of an hour. The probability to reject the null hypothesis has been found much below 0.1%, thus indicating a high probability that the coefficients are important indeed. This is somewhat confirmed by the rather narrow confidence intervals shown in Figs. 4 and 5.

VI. CONCLUSION

A new formulation of state estimation in distribution networks has been proposed, which avoids the use of classical pseudo-measurements. To that end, MV bus power injections are expressed as functions of a small number of LV components, which are added to the system state vector. This makes state estimation possible, while providing valuable real-time information on the LV network state. Hachtel's method is used to solve the constrained least-square problem.

A Monte-Carlo procedure which takes advantage of available LV network smart meter measurements has been described for identifying the MV bus injection models. Various numbers of LV load components have been considered. For the tested system, the best model was obtained when considering two types of loads (i.e. households and tertiary sector) in addition to the photo-voltaic generation.

Future investigations will deal with a systematic clustering of LV customers in order to improve accuracy. Attention will be also paid to reducing the high number of MV bus injection models (presently one for each quarter of an hour of each typical day).

REFERENCES

[1] A. Abur, A. Gomez Exposito, *Power System State Estimation: Theory and Implementation*, Marcel Dekker Inc., New York-Basel, 2004.

[2] R. Singh, B. C. Pal, R. A. Jabr, "Choice of estimator for distribution system state estimation," *IET Gener., Transm., Distrib.*, vol. 3, pp. 666-678, Jul. 2009.

[3] W. Kong, G. Ault, D. Wang, A. Michiorri, C. Foote, R. Currie, "Advanced Load Modelling Techniques for State Estimation on Distribution Networks with Multiple Distributed Generators,"in *Proc. of 17-th PSCC*, Stockholm, Sweden, Aug. 2011.

[4] E. Manitsas, R. Singh, B. C. Pal, G. Strbac, "Distribution System State Estimation Using an Artificial Neural Network Approach for Pseudo Measurements Modeling," *IEEE Trans. on Power Systems*, vol. 27, pp. 1888-1896, Nov. 2012.

[5] J. Wu, Y. He, N. Jenkins, "A Robust State Estimator for Medium Voltage Distribution Networks," *IEEE Trans. on Power Systems*, vol. 28 pp. 1008-1016, May 2013.

[6] M. Lehtonen, M. Jalonen, A. Matsinen, "A novel state estimation model for distribution systems,"in *Proc. of 14-th PSCC*, Seville, Spain, Jun. 2002.

[7] A. Gjelsvik, S. Aam, L. Holten, "Hachtel's augmented matrix method - A rapid method improving numerical stability in power system static estimation", *IEEE Trans. on PAS*, vol. PAS-104, no. 11, pp. 2987-2993, Nov. 1985.

[8] M. Stein, "Large Sample Properties of Simulations Using Latin Hypercube Sampling," *Technometrics*, Vol. 29, Issue 2, 1987.

[9] D. C. Montgomery, *Design and Analysis of Experiments*, 5th Ed., Wiley, Hoboken, New York, 2001.