Development of Black-Box Equivalent Models from Ambient Data

Theofilos Papadopoulos[1], Georgios Barzegkar-Ntovom[1], Eleftherios Kontis[2], and Nikolaos Batzis[1]

[1]Department of Electrical & Computer Engineering, Democritus University of Thrace, Xanthi, Greece
thpapad@ee.duth.gr, gbarzegk@ee.duth.gr, nikobatz@ee.duth.gr
[2]School of Electrical & Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece
kontislefteris@gmail.com

Abstract

In this paper, aggregated black-box equivalent models are developed from ambient data. The accuracy of the identified model parameters is evaluated using error indices. The applicability of the derived models for the analysis and simulation of large voltage disturbances is also demonstrated.

1. Introduction

The accuracy of power system stability studies relies heavily on the accuracy of the adopted models [1]. Therefore, precise models are required for all power system assets. Unlike all other components, power system loads and distribution networks (DNs) are integrated in stability studies using aggregated models [2]. The development of such models is very challenging and demanding.

Aggregated models for load and DN analysis can be developed using either the component- or the measurement-based approach [3]. In the former, aggregated models are built using a bottom to top approach by exploiting billing results, by analysing the behaviour of groups of consumers, and by applying statistical analysis [4]. Nevertheless, power system loads, and especially DNs, are becoming nowadays more and more complex and time-varying [5]. Therefore, the measurement-based approach has recently started gaining ground due to the advent of new measurement technologies, such as the smart meters and the phasor measurement units (PMUs). In this approach, aggregated models are developed by analyzing field measurements (system responses).

Usually aggregated models, developed using the measurement-based approach, are derived by analysing post large disturbance responses (PLDRs), [2], [4]. However, PLDRs do not occur frequently. Therefore, it is challenging to maintain up-to-date aggregated models using only PLDRs. Thus, recently researchers are focusing on the development of aggregated models from ambient data, [2], [4], [5]. Ambient data refer to the small disturbances, caused from the continuous random variations of electrical loads and renewable energy sources [5], [6].

In this paper, several aggregated black-box models are developed using ambient data. Black-box models can be divided into two main categories, namely the static and dynamic models [7]. Static models describe the relation between the real/reactive (*P*/*Q*) power at any time instant with the bus voltage; dynamic models express *P*/*Q* as a function of voltage and time. In particular, a static load model, namely, the polynomial (ZIP) model, and three dynamic models, i.e., the exponential recovery model (ERM), the second order recovery model (SORM) and the autoregressive moving average exogenous (ARMAX), are considered and examined. The accuracy of the identified model parameters is assessed by appropriate error indices; The applicability of the derived models for the analysis of large disturbances is also evaluated and quantified.

The rest of the paper is organized as follows: In Section 2 the examined black-box models are presented. The parameter estimation procedure is analysed in Section 3. Simulation results are provided in Section 4. Finally, Section 5 summarizes the main findings of the research and concludes the paper.

2. Black-box Model Structures

In this Section, the structure of the examined models is briefly explained.

2.1. Polynomial Model

The second order polynomial also known as ZIP model, since it consists of constant impedance (*Z*), constant current (*I*) and constant power (*P*) load components is a widely used static load model [7], [8]. The mathematical representation of the polynomial load model for the *P*/*Q* response is given in the following generalized form:

$$y(V) = y_0\left[y_1\left(\frac{V}{V_0}\right)^2 + y_2\left(\frac{V}{V_0}\right) + y_3\right] \qquad (1)$$

where, $y(V)$ is the calculated *P*/*Q* and V is the bus voltage; V_0 and y_0 are the voltage and *P*/*Q* immediately prior to the disturbance, respectively. The model parameters, y_i, express the participation of specific load types with respect to the total load demand. In particular, for i=1, y_1 denotes the fraction of the constant impedance, for i=2, y_2 represents the fraction of the constant current and for i=3, y_3 the fraction of the constant power load [9]. Note that, $\sum_{i=1}^{3} y_i = 1$ must apply.

2.2. Exponential Recovery Model

The ERM describes the *P*/*Q* as a non-linear function of voltage and time, t, [10], as shown in (2).

$$y(V,t) = \left(y_0\left(\frac{V}{V_0}\right)^{a_s} - y_0\left(\frac{V}{V_0}\right)^{a_t}\right)\left(1 - e^{-\frac{t-t_0}{T_p}}\right) + y_0\left(\frac{V}{V_0}\right)^{a_t}. \qquad (2)$$

Here t_0 is the disturbance time instant, T_p the recovery time constant, a_s is the steady-state power exponent and a_t the corresponding transient parameter.

2.2. Second Order Recovery Model

The SORM is a higher order representation of ERM, that can simulate more accurately complex dynamic responses [11]. The formulation of SORM is given in generic form by:

$$y(t,V)=\left(y_0+\frac{q_0}{p_0}\left(V(t)-V_0\right)\right)+\left(e^{-\frac{p_1}{2}(t-t_0)}\right)\cdot$$
$$\left[\frac{2p_0q_1-p_1q_0}{2p_0\sqrt{p_0-\frac{p_1^2}{4}}}\sin\left(\sqrt{p_0-\frac{p_1^2}{4}}\left(t-t_0\right)\right)\right. \quad . \qquad (3)$$
$$\left.-\frac{q_0}{p_0}\cos\left(\sqrt{p_0-\frac{q_1^2}{4}}\left(t-t_0\right)\right)\right]\left(V(t)-V_0\right).$$

The development of the SORM requires the identification of the model parameters p_0, p_1, q_0 and q_1.

2.4 ARMAX model

An ARMAX model provides a description of the system with output, $y(t)$ and input, $u(t)$, distorted by white noise, $e(t)$, as shown in (4) in discrete form:

$$y[k]+\sum_{\ell=1}^{p_a}a_\ell y[k-\ell]=\sum_{\ell=1}^{p_b}b_\ell u[k-\ell]+\sum_{\ell=0}^{p_c}c_\ell e[k-\ell]. \qquad (4)$$

where, a_ℓ , b_ℓ and c_ℓ are the parameters of the autoregressive (AR) model, the exogenous inputs and the moving average (MA) model, of order p_a, p_b and p_c, respectively [12].

In z-domain the ARMAX formulation is given by:

$$Y(z)=\frac{B(z)}{A(z)}U_m(z)+\frac{C(z)}{A(z)}e(z). \qquad (5)$$

where, $B(z)/A(z)$ is the deterministic part of the model and $C(z)/A(z)$ is the stochastic one. Polynomials $A(z)$, $B(z)$, and $C(z)$ are defined as:

$$A(z)=1+\sum_{\ell=1}^{p_a}a_\ell z^{-\ell}$$
$$B(z)=\sum_{\ell=1}^{p_b}b_\ell z^{-\ell} \qquad (6)$$
$$C(z)=\sum_{\ell=0}^{p_c}c_\ell z^{-\ell}$$

3. Black-box Model Derivation

In this Section, the proposed black-box model parameter identification method is introduced. The parameters of the examined models are determined by using ambient data. Subsequently, the derived black-box models are used to simulate transient responses. More specifically, the following step-by-step procedure is adopted for the identification of the model parameters and their validation.

- Data recording: data of V and P/Q under normal operating conditions (ambient data) are recorded.
- Black-box models development: the P/Q model parameters of ZIP, ERM, SORM or ARMAX are identified by using ambient data (sub-Section 3.1). To access the performance of the developed models, evaluation indexes are used (sub-Section 3.2).
- Model validation: the effectiveness and accuracy of the developed black-box models is validated by using transient responses.

3.1. Parameter Estimation

The P/Q model parameters (θ) are determined by using the measured V and P/Q ambient data as input and output, respectively, to a curve fitting problem. The unknown model parameters of the ZIP, ERM and SORM are estimated via nonlinear least square optimization [11]. Specifically, following successive iterations the objective function of (7) is minimized:

$$J=\sum_{n=1}^{N}\left\|\left(y[n]-\hat{y}[n\,|\,\theta]\right)^2\right\| \qquad (7)$$

where N is the total number of samples, $y[n]$ and $\hat{y}[n]$ the original and the estimated P/Q at the n-th sample, respectively. The inputs to this optimization problem are the formulas of the examined model, the voltage and P/Q data and an initial random estimate of the model parameters. The trust-region-reflective algorithm is used to solve the problem.

The ARMAX model parameters are calculated by using a prediction error method [12], [13].

3.2. Model Evaluation

The performance of the examined models is evaluated on the basis of three indexes. The parameter percentage error (*PPE*), defined in (8) is used to evaluate the accuracy of the estimated model parameters:

$$\%PPE=\frac{|\theta_S-\theta_E|}{|\theta_S|}\cdot 100 \qquad (8)$$

where, θ_S and θ_E are the actual (Table I) and the simulated (in terms of black-box modelling) parameters. Further, to access the accuracy of the simulated responses, the coefficient of determination (R^2) and the relative error (*RE*), defined in (9) and (10), respectively are calculated.

$$R^2=\left(1-\frac{\sum_{n=1}^{N}\left(y[n]-\hat{y}[n]\right)^2}{\sum_{n=1}^{N}\left(y[n]-\bar{y}\right)^2}\right)\cdot 100\% \qquad (9)$$

$$RE=\sqrt{\frac{\left(\sum_{n=1}^{N}\left(y[n]-\hat{y}[n]\right)^2\right)}{\sum_{n=1}^{N}y[n]^2}}\cdot 100\% \qquad (10)$$

where $\bar{y}$ is the mean value of the response. Note that, a 100% R^2 and 0% *PE* values, indicate perfect matching.

4. Simulation Studies

The feasibility of the proposed method is demonstrated on the basis of synthetic signals. Real power synthetic signals are generated by using the simple simulation model of Fig. 1. The black-box block corresponds to the DN that can be represented by the models described in Section 2. In particular, the following two network configurations are examined:

- grids dominated by residential-commercial motors, represented by either the ZIP or the ERM,
- grids dominated by small induction machines, represented by the SORM.

For the examined models typical real power parameters have been used as summarized in Table 1 [13].

Initially ambient data are generated by distorting the V signal used to excite the model, with Gaussian noise to imitate random load demand fluctuations and other relevant small perturbations under normal operating conditions. The simulations are conducted, considering a rate of 1000 samples per second (sps) and a signal-to-noise ratio (SNR) 30 dB. Fig. 2 demonstrates an example of ambient data generated by using the SORM.

The transient responses used to evaluate the effectiveness and accuracy of the developed black-box models are excited by applying the following voltage step responses (SR):

- SR#1: voltage step-down from 1.0 per-unit (p.u.) to 0.8 p.u.
- SR#2: voltage step-down from 1.0 p.u. to 0.9 p.u.

In addition, to replicate measurement errors, the transient responses are intentionally distorted by additive white Gaussian noise (AWGN), assuming SNR 30 dB.

Table 1. Real power response parameters

Model	Set of parameters
ZIP	$\theta_{P,ZIP}$=[0.5, 0.4, 0.1]
ERM	$\theta_{P,ERM}$=[0.1070, 1.0411, 0.2047]
SORM	$\theta_{P,SORM}$=[2100, 55, 2.5, 150]

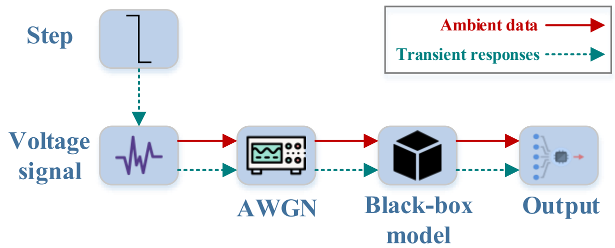

Fig. 1. Block diagram of the simple simulation model to generate synthetic signals.

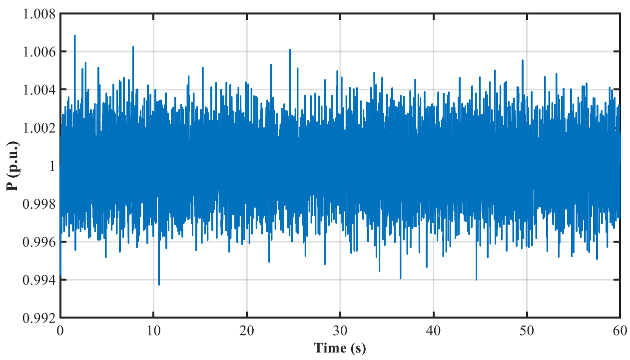

Fig. 2. Ambient response using the SORM.

4.1. ZIP results

By considering the ZIP as the simulation model of Fig. 1, ambient P data are generated. These are used to estimate the corresponding ZIP model parameters assuming that the identification window length (ambient data) varies from 30 s to 600 s. The identified model parameters are practically identical with the actual ones of Table 1 for all window lengths, as %PPE differences are very low, i.e., well below 0.1 %.

The derived models for each window length are used to simulate the two step responses; an exemplary case is presented in Fig. 3 for SR#1 and window length 600 s. The calculated R^2 and RE are summarized in Fig. 4. It is evident that, the simulated responses match very accurately the actual, as the R^2 is above 99.99 % and the corresponding RE below 0.0002%. Results also indicate that the window length has an insignificant impact on the estimation of the model parameters and consequently to the simulated responses.

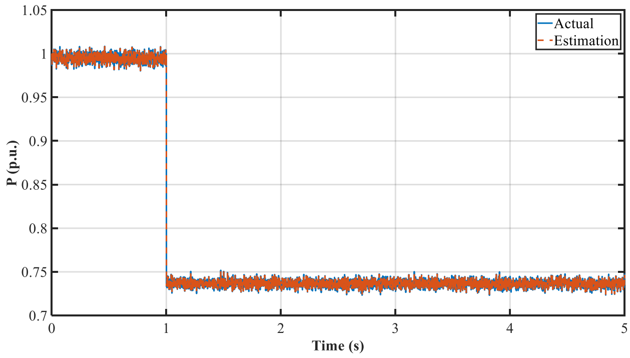

Fig. 3. SR#1 ZIP real power response for a window length of 600 s.

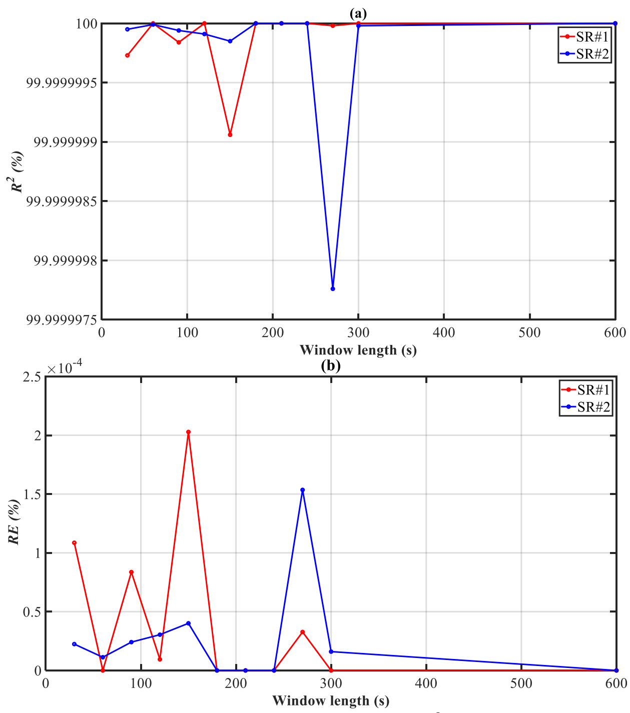

Fig. 4. ZIP real power response (a) R^2 and (b) RE.

4.2. ERM results

The same procedure is followed also for the ERM. The calculated *PPE* for all model parameters and window lengths is presented in Fig. 5. It can be deduced, that a_t and T_p are estimated with relative accuracy, as the corresponding average *PPE* is 0.5 % and 9.3 %, respectively. Conversely, a_s is not identified accurately as *PPE* exceeds 400 % for all cases. Therefore, it can be realized that the ERM cannot be used for the ambient analysis. This is also substantiated by Fig. 6 plots, where the actual and the simulated transient responses for SR#1 and SR#2 are compared.

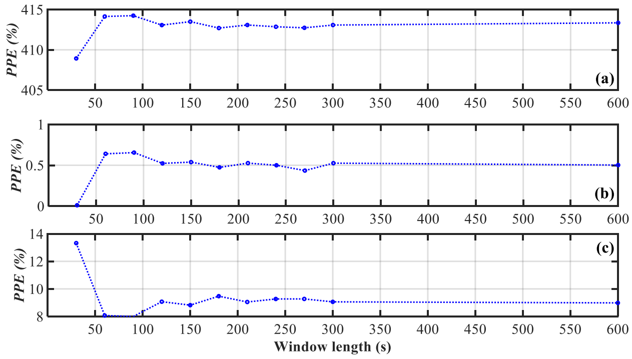

Fig. 5. *PPE* for (a) a_s, (b) a_t and (c) T_p.

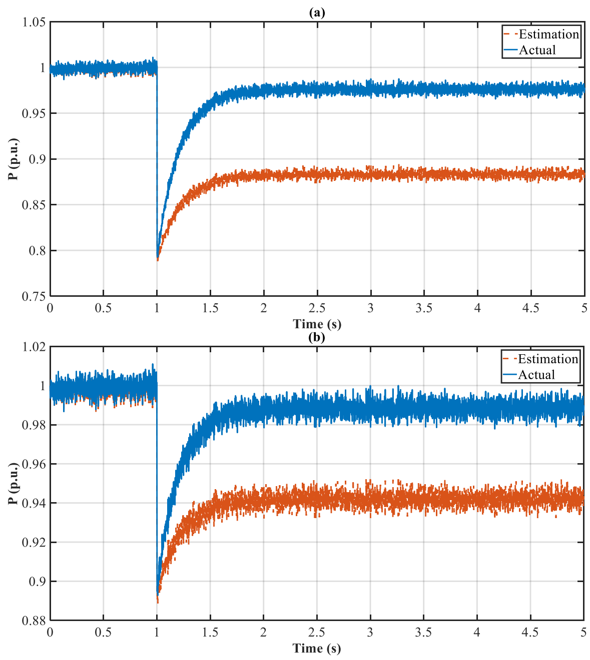

Fig. 6. ERM real power response for (a) SR#1 and (b) SR#2. The window length is 300s.

4.3. SORM results

SORM is also used to simulate transient responses by using the model parameters that have been identified from ambient data. In this case the proposed procedure fails to estimate accurately the model parameters. This entails that the simulated transient responses present significant discrepancies from the actual as shown in Fig. 7. In fact, the response prior to and after the disturbance is captured with relative accuracy. However, the developed model cannot represent the transient part at all.

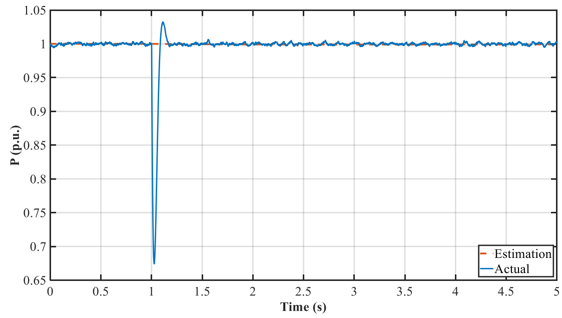

Fig. 7. SORM real power response.

4.3. ARMAX results

From the ERM and SORM results it can be deduced that the dynamic behaviour of the grid cannot be represented accurately by using these models. To investigate the complex dynamic grid performance under ambient conditions, ARMAX modelling is examined. In particular, the ambient data generated by incorporating the ERM and the SORM in the black-box block of the simulation model of Fig. 1, are used to identify the ARMAX model parameters. The model order of the ARMAX model is assumed to be varying from 1 to 3 and the window length from 60 s to 1800 s. Subsequently, the developed ARMAX models are used to simulate the step responses of the two test cases.

The simulated ARMAX responses for SR#1 and SR#2 are compared to the corresponding actual ones for the dominated by residential-commercial motors grid and the dominated by small induction machines grid cases in Figs. 8 and 9, respectively; the window length is 300s. Results indicate that the 1st order ARMAX model cannot represent the transient part as well as the steady-state after the disturbance. Conversely, the 2nd and the 3rd order models present almost a perfect match (as indicated in the Fig. close-up) with the actual response as the average R^2 is 98.3 % and 99.66% respectively; the *RE* is 0.13 % and 0.12 %. It is also worthy of note, that the examined window lengths have an insignificant impact on the calculated parameters and in turn to the simulated responses.

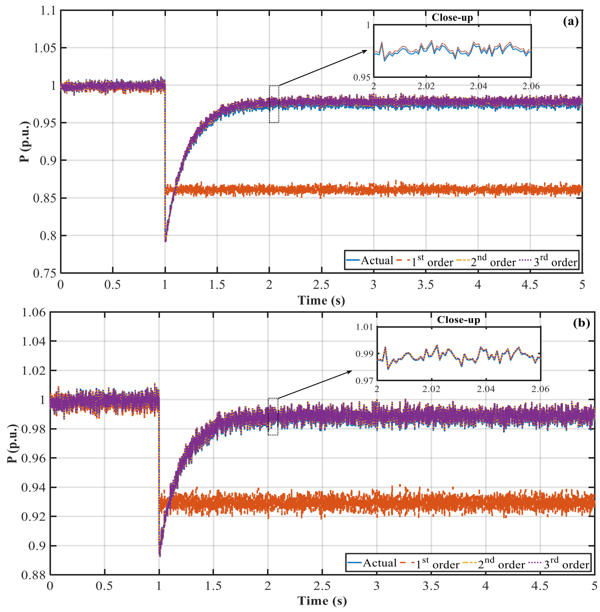

Fig. 8. Effect of the ARMAX model order on the real power response for (a) SR#1 and (b) SR#2.

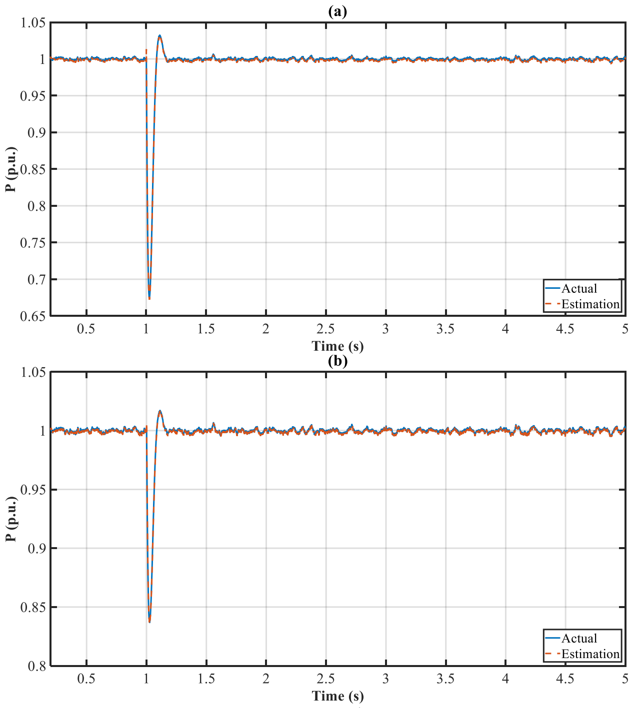

Fig. 9. Comparison of actual and 3rd order ARMAX model real power responses for (a) SR#1 and (b) SR#2 for the small induction machine dominated grid.

5. Conclusions

In this paper, the development of black-box dynamic equivalent models from ambient data is investigated. Towards this objective, the effectiveness of the ZIP, ERM, SORM and ARMAX modelling is examined using synthetic signals generated from a simple simulation model. The results of the conducted analysis have shown that:

- The ZIP can be used to model residential grids and more specifically grids that practically behave as constant impedance under disturbances.
- The ERM and SORM cannot be used to develop black-box models from ambient data.
- The 2nd and the 3rd order ARMAX models can accurately represent the distinct characteristics of the transient response, i.e., overshoot, the power recovery, and the new steady state after the disturbance.

In summary, it can be deduced that high order models are required to analyse the complex dynamic behaviour of power DNs. This is more pronounced in modern grids rich in inverter-based loads and resources.

7. References

[1] Y. Zhu and J. V. Milanović, "Automatic Identification of Power System Load Models Based on Field Measurements," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3162-3171, May 2018.

[2] X. Zhang, D. J. Hill, C. Lu and Y. Song, "A Hierarchical Framework for Ambient Signals Based Load Modeling: Exploring the Hidden Quasi-Convexity," *IEEE Trans. Power Syst.*, vol. 36, no. 6, pp. 5780-5791, Nov. 2021.

[3] E. O. Kontis, T. A. Papadopoulos, A. I. Chrysochos and G. K. Papagiannis, "Measurement-Based Dynamic Load Modeling Using the Vector Fitting Technique," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 338-351, Jan. 2018.

[4] X. Zhang, D. J. Hill, and L. Zhu, "Enhanced ambient signals based load model parameter identification with ensemble learning initialisation," *IET Gener. Transm. Distrib.*, vol. 14, no. 24, pp. 5877-5887, Dec. 2020.

[5] X. Zhang, C. Lu, and Y. Wang, "A two-stage framework for ambient signal based load model parameter identification," *Int. J. Electr. Power Energy Syst.*, vol. 121, Oct. 2020.

[6] J. W. Pierre, D. J. Trudnowski and M. K. Donnelly, "Initial results in electromechanical mode identification from ambient data," *IEEE Trans. Power Syst.*,, vol. 12, no. 3, pp. 1245-1251, Aug. 1997.

[7] Cigre Working Group C4.605, "Modelling and Aggregation of Loads in Flexible Power Networks," 2014.

[8] G. A. Barzegkar-Ntovom, E. O. Kontis, T. A. Papadopoulos and P. N. Papadopoulos, "Methodology for Evaluating Equivalent Models for the Dynamic Analysis of Power Systems," *IEEE Trans. on Power Del.*, vol. 37, no. 6, pp. 5059-5070, Dec. 2022.

[9] T. A. Papadopoulos, E. N. Tzanidakis, P. N. Papadopoulos, P. Crolla, G. K. Papagiannis and G. M. Burt, "*Aggregate load modeling in microgrids using online measurements,*" MedPower 2014, Athens, 2014.

[10] D. Karlsson, D. J. Hill, "Modelling and identification of nonlinear dynamic loads in power systems," *IEEE Trans. Power Syst.*, vol.9, no.1, pp. 157-166, Feb 1994.

[11] G. A. Barzegkar-Ntovom, T. A. Papadopoulos and E. O. Kontis, "Robust Framework for Online Parameter Estimation of Dynamic Equivalent Models Using Measurements," *IEEE Trans. on Power Syst.*, vol. 36, no. 3, pp. 2380-2389, May 2021.

[12] T. A. Papadopoulos, E. O. Kontis, G. A. Barzegkar-Ntovom and P. N. Papadopoulos, "A Three-Level Distributed Architecture for the Real-Time Monitoring of Modern Power Systems," *IEEE Access*, vol. 10, pp. 29287-29306, 2022.

[13] A. Ubolli and B. Gustavsen, "Comparison of methods for rational approximation of simulated time-domain responses: ARMA, ZD-VF, and TD-VF," *IEEE Trans. on Power Del.*, vol. 26, no. 1, pp. 279– 288, 2011.

[14] J. V. Milanović and Y. Xu, "Methodology for Estimation of Dynamic Response of Demand Using Limited Data," in *IEEE Trans. on Power Syst.*, vol. 30, no. 3, pp. 1288-1297, May 2015.