Electric Power Systems Research 163 (2018) 66–72

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

Statistical testing for load models using measured data

Jiaqing Lv[a,*], Mirosław Pawlak[a], U.D. Annakkage[a], Bagen Bagen[b]

[a] Department of Electrical & Computer Engineering, University of Manitoba, Canada
[b] Manitoba Hydro, Canada

ARTICLE INFO

Keywords:
Load modeling
Statistical testing
F-test
Regression nested-models
Parametric system identification

ABSTRACT

This paper applies the statistical testing theory to examine the validity of different load parametric models. Traditionally measurement-based static load modeling has been performed based on a single parametric model. Commonly utilized models include: ZIP (constant-impedance–constant-current–constant-power) model, exponential model, and frequency component adjusted ZIP/exponential models. It has been conjectured that the models making use of the frequency feature should be better compared to the ones purely based on the voltage component. However, there has not been any theoretical-based justification to confirm this claim. It is a goal of this paper to provide a formal method for verifying this claim by employing the theory of statistical testing for correct parametric model specification. In particular, a class of *F*-tests for checking the correctness of the specific load model is employed. Our methodology is verified on the real phasor measurement unit (PMU) data describing a radial load in the Manitoba Hydro power system. The obtained results confirm the usefulness of the frequency component based models.

1. Introduction

Load modeling plays an important role in stable operation of power systems. Load models are used in power flow and transient stability studies for planning and operation of power systems. It is well known [1] that the results of a transient stability analysis depend on the load models used. Therefore, it is important to model the loads as close as possible to the actual load. Modeling of loads is complicated because loads connected to a typical bus are composed of a large number of devices often with ever-changing characteristics. The composition of loads often has a complicated dependence on many factors including time, season, weather condition, and energy prices, just to name a few. Moreover, one should take into account the time-varying nature of the load systems as the load signals such as voltage, frequency, current are non-stationary processes.

Typical approaches for load modeling can fall into two categories: component-based methods [1–3] and measurement-based methods. Furthermore, the measurement-based approach can be based on static modeling, dynamic modeling, or the combined static–dynamic modeling. The following is a popular polynomial static model describing the dependence between the voltage (V) signal and the powers: the real power – P and the reactive power – Q:

$$P = a_1 V^2 + a_2 V + a_3 + \varepsilon, \tag{1a}$$

$$Q = a_4 V^2 + a_5 V + a_6 + \xi, \tag{1b}$$

where ε, ξ are modelling errors regarded to be random variables with zero mean and finite variance.

Another commonly used load model is the exponential model of the following form:

$$P = a_1 V^{a_2} + \varepsilon, \tag{2a}$$

$$Q = a_3 V^{a_4} + \xi. \tag{2b}$$

The above models assume that the output variables, i.e., P, Q are only influenced by the voltage signal V. In more refined models P, Q may depend both on the voltage V and the frequency f. This leads to the following extended models corresponding to (1) and (2)

$$P = (a_1 V^2 + a_2 V + a_3)(1 + a_4(f - f_0)) + \varepsilon, \tag{3a}$$

$$Q = (a_5 V^2 + a_6 V + a_7)(1 + a_8(f - f_0)) + \xi, \tag{3b}$$

and

$$P = a_1 V^{a_2}(1 + a_3(f - f_0)) + \varepsilon, \tag{4a}$$

$$Q = a_4 V^{a_5}(1 + a_6(f - f_0)) + \xi, \tag{4b}$$

where f_0 is the reference frequency being 60 Hz in the North American power system. The aforementioned models are defined up to the unknown coefficients $\{a_i\}$ that need to be determined from the available

* Corresponding author.
E-mail address: lvjiaqing@gmail.com (J. Lv).

https://doi.org/10.1016/j.epsr.2018.05.026
Received 24 October 2017; Received in revised form 6 April 2018; Accepted 28 May 2018
0378-7796/

data.

It is also worth mentioning that the model in (1) is called the ZIP model, since the different terms of the formula in (1) correspond to constant impedance (Z), constant current (I) and constant power (P). Model (2) is commonly referred to as the exponential model. Models (3), (4) are the extended ZIP and exponential models, respectively. They have correction terms depending on the additional input variable being the frequency component. Note that if $a_4 = a_8 = 0$ then the model in (3) reduces to the simpler model in (1). An analogous reduction applies to the model in (4).

The exponential model in (2) has been examined in [4,5], whereas in [6] the frequency-dependent models in (3), (4) have been evaluated. Besides the aforementioned models there have been other, less commonly used in practice, modelling approaches that can be regarded as some modifications of the above introduced popular models. For instance, the static model in [7] can be viewed as an intermediate model between models (1) and (3), while the EPRI LoadSyn program [8] uses a model with the real power part being a linear combination of (2a) and (4a), and the reactive part being a linear combination of the two models in (4b) with different parameters [8]. The latter approach has been also applied in [2,8]. Composite load models are the extension of the static models by allowing the dynamic part to be involved. Their static parts are as in (1)–(4), whereas the dynamic part is connected in a certain block-oriented structure. For instance, in [9], the models in (1), (2) are combined in the parallel structure with a dynamic part represented by an induction motor. These overall composite models are referred to as the ZIP-induction motor (GZIP-IM) load model and the exponential-induction motor (Exp-IM) load model. Similarly, in [10–14] a combination of (1) and the motor dynamic part has been taken into consideration. On the other hand, in [11] a composite model with the combination of (3) and the motor dynamic part has been utilized. Besides, a dynamic model can be extended from the basic form of static model through other approaches. For instance, the so-called “exponential recovery model” in [15] provide such a means.

Recently, artificial intelligence methods [16,17] as well as Hidden Markov models [18] have also been applied in load modeling.

In this paper, the basic static load models defined in (1), (2), (3), and (4) are examined. The problem of fitting, i.e., estimating unknown parameters of these models has been addressed in the aforementioned references. In practice, however, one needs to decide which model gives the most significant fit to the observed data. In fact, every loading modeling problem should be equipped with the proper model check. Hence, the principal goal of this paper is to consider testing procedures for such hypotheses. A formal statistical test is designed that is able to choose either the simple models in (1), (2) or the more complex models in (3), (4). These are nested models and the version of the F-test to verify the posed hypotheses is applied. The F-test is the general strategy for verifying the hypothesis that the proposed restrictive model fits the data well compared to the unrestricted larger model [19,20]. The test is based on the relative distance in residual sum of squares between the restrictive model and the full model. The developed methodology is then applied to the real PMU data representing the observations of a radial load in the Manitoba Hydro power system. Our findings are that for this particular data the frequency component is essential for the load modeling problem, i.e., the complex models in (3) and (4) are more preferable than the simple ones. Due to the random nature of the observed data this conclusion should be interpreted in the statistical framework, i.e., with the high probability of acceptance one can conclude that the models with the frequency component better fit the given data set. It is also clear that the developed methodology can be used by power engineers to select the appropriate load model that is suitable for a particular application.

The rest of the paper is organized as follows. Section 2 gives the description of the problem of parametric nested models specification and the corresponding test statistics. The problem of load models fitting under the hypotheses is examined in Section 3. Section 4 implements the introduced methodology to the real PMU data derived from the Manitoba Hydro power system. Concluding remarks are presented in Section 5.

2. Model specification and *F* statistics

In the regression analysis one obtains the data set $D_n = \{(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)\}$, where the component Y_i is the response variable that we try to explain from the values of the input variables $\{\mathbf{X}_i\}$. The postulated model between $\mathbf{X}_i$ and Y_i is of the parametric form

$$Y_i = m(\mathbf{X}_i; \theta^\star) + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $\{\varepsilon_i\}$ is the random non-observed noise process that is assumed to be a sequence of independent and identically distributed ($i.\ i.\ d.$) random variables with variance σ^2. The nonlinearity $m(\mathbf{x};\theta)$ is the known function except for the p-dimensional parameter θ, where $\theta^\star$ is the true value of θ. If the model $m(\mathbf{x};\theta)$ is correctly specified then one can estimate $\theta^\star$ by minimizing the least-squares criterion, i.e.,

$$S(\theta) = \sum_{i=1}^{n} (Y_i - m(\mathbf{X}_i; \theta))^2. \tag{5}$$

The resulting least-squares estimate $\hat{\theta}$ can converge to $\theta^\star$ under very general regularity condition on the nonlinearity $m(\mathbf{x};\theta)$, see Chapter 12 in [20] for details. Moreover, if the noise $\{\varepsilon_i\}$ is normally distributed then the $\hat{\theta}$ is also the maximum-likelihood estimator that reveals some further efficiency properties [20]. In practice, one is confronting with several alternative regression models and would like to choose the model that explains the data the best. Hence, let

$$H_0\colon Y_i = m_0(\mathbf{X}_i; \theta_0^\star) + \varepsilon_i \quad i = 1, \ldots, n, \tag{6}$$

represents the restricted model, where $\theta_0^\star \in \mathbf{R}^{p_1}$. On the other hand

$$H_1\colon Y_i = m_1(\mathbf{X}_i; \theta_1^\star) + \varepsilon_i \quad i = 1, \ldots, n, \tag{7}$$

describes the alternative full model, where $\theta_1^\star \in \mathbf{R}^{p}$. These are nested models as $p = p_1 + p_2$, some $p_2 \geq 0$ and if $p_2 = 0$ then $\theta_1^\star = \theta_0^\star$ and consequently $m_0(\mathbf{x}; \theta_0^\star) = m_1(\mathbf{x}; \theta_1^\star)$. Hence, for the nested models, the model under H_0 is the special case of the model under H_1.

An important example is the following linear null hypothesis model

$$H_0\colon Y_i = a_0 + a_1 X_{1i} + \cdots + a_{p_1-1} X_{p_1-1,i} + \varepsilon_i, \quad i = 1, \ldots, n, \tag{8}$$

where $\mathbf{X}_i = (1, X_{1i}, \ldots, X_{p_1-1,i})^T$ is the p_1-dimensional vector of the input variables. The alternative full linear regression model is

$$H_1\colon Y_i = a_0 + a_1 X_{1i} + \cdots + a_{p_1-1} X_{p_1-1,i} + \cdots + a_{p-1} X_{p-1,i} + \varepsilon_i, \tag{9}$$

for $i = 1, \ldots, n$, where now $\mathbf{X}_i = (1, X_{1i}, \ldots, X_{p-1,i})^T$ is the p-dimensional input vector. In this case the choice between the two competitive models is equivalent to testing the null hypothesis H_0 that the last $p_2 = p - p_1$ of the a_i's are 0, i.e., $a_i = 0$ for $i = p_1, \ldots, p - 1$.

The general approach for testing the nested nonlinear regression hypotheses defined in (6) and (7) is based on the generalized likelihood ratio test [20,21] that requires the evaluation of the log-likelihood ratio

$$\log \frac{p(\mathbf{Y}|H_1)}{p(\mathbf{Y}|H_0)},$$

where $p(\mathbf{Y}|H_0)$ is the joint density of the observed data $\mathbf{Y} = (Y_1, \ldots, Y_n)$ when the null hypothesis model in (6) holds. This density should be conditioned on the input data $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ if they reveal the random nature. This is the case in our load model specification problem that is examined in this paper. The same interpretation applies to $p(\mathbf{Y}|H_1)$ for the alternative model in (7). The explicit evaluation of the likelihood ratio is generally difficult. If, however, one assumes that the noise process $\{\varepsilon_i\}$ is Gaussian then the direct calculation, see [20,21], yields the so-called F-statistic

$$\mathbf{F}_n = \frac{(\mathrm{SS}_0 - \mathrm{SS}_1)}{\mathrm{SS}_1} \frac{n - p}{p - p_1}, \tag{10}$$