

A Statistical Selection Approach of Dynamic Load Model Parameters

Yulong Huang^{1, a *}, Xun Chen^{2, b}

¹Electrical and Information College, Jinan University, Zhuhai, 519070, China

²Electric Power Research Institute of Guangdong Power Grid Corporation, Guangzhou, 510600, China

^athuangyulong@jnu.edu.cn, ^bsteadyman@126.com

Keywords: Load model, Time-variation, Multiple linear regression

Abstract. Since real load characteristics are time-varying all along, load model parameters built on some historical data are only valid within limited scenarios. A real-time dynamic load model parameter selection method based on multiple linear regression (MLR) is proposed to find the load model parameters that are best matched with real time system operation condition from load model parameters history database. And along with the database size growing larger, the load model parameter matching accuracy will become higher and higher. The effectiveness and accuracy of the proposed method are verified on field measurement data collected from a substation in a metropolitan area of China.

Introduction

Real-time load dynamic characteristics are varying along with time, month, season, holiday, temperature, etc., since the aggregate load always consists of various and time-varying components with different characteristics [1]. The randomness of power loads and the variety of their composition make accurate load modeling a most challenging task [2-3].

Classification and synthesis for load dynamic characteristics has been proposed as a good way to settle the time-variation problem in load modeling in [4-8]. Yet limited classifications can't express constantly time-varying load characteristics well enough, and different classes may also intersect together. Based on assumptions of the error function, a multicurve identification technique has been proposed in [1] to derive load model parameters, and the generalization capability of the built model also investigated to meet most random scenarios in power system. However, it conceals the stochastic parameter fluctuation in the real-time load model, and also sacrifices the load model accuracy.

The thesis [9] suggested that more sophisticated measurements and signal processing such as stochastic parameter identification could help for the problem of load drift and change in composition. [1] also concluded that the load model that captures the inner law of the randomly varied load would be the best one. In order to disclose real-time dynamic load model parameters' statistically changing law and hence enhance load model accuracy, a load model parameter selection method based on MLR is proposed to find the load model parameters from load model parameter history database that are best matched with real time system operation condition. By this way, the load model parameter matching accuracy will become higher and higher along with the database size growing larger, which improves the precision of power system simulations under large disturbances so as to determine more precise settings for real-time voltage stability or transient stability control and protection devices, etc..

Real-Time Dynamic Load Model Parameter Selection

Load Model Parameters Database Foundation. The load model we built is mainly for power system stability simulation and control. Thus, it should be easily integrated into the power system analysis tool. Furthermore system operators prefer a load model with clearly physical interpretations. So ZIP augmented with induction motor is chosen as aggregate load model structure in this paper. The model consists of a static part modeled as ZIP and an equivalent circuit of the third-order induction motor, whose detailed equivalent circuit, algebraic and differential equations can be found in paper [1].

In this paper, differential evolution (DE) algorithm[10] is applied in load model parameter identification for all the historical recorded system disturbances data because of its strong global optimization searching capability, forming load model parameters history database. With accumulation of field measurements and their corresponding load model parameters history database, it becomes easier to find the statistical law behind the seemingly irrelevant load data, and the model that captures the inner law of the randomly varied load would be the best one.

Load Model Parameter Selection Based on MLR. Regression analysis approach can investigate the relationship between two or more variables related in a nondeterministic fashion, and are often used to perform prediction. Real-time load dynamic characteristics are varying indeed along with time, month, season, holiday, temperature, bus voltage and its fluctuation magnitude, active power, reactive power, apparent power, power oscillation amplitude, etc.. A load model parameter selection method based on MLR[11-12] is proposed to search the load model parameters from load model parameter history database that are best matched with real time system operation condition in this section.

In order to evaluate load model output errors between different disturbances, a revised load model output error E^k is defined in Eq. 1.

$$E^k = \sqrt{\frac{1}{n_k} \left[\frac{\sum_{j=1}^{n_k} (P_j^k - P_j'^k)^2 + (Q_j^k - Q_j'^k)^2}{(P'^{kmax} - P'^{kmin})^2 + (Q'^{kmax} - Q'^{kmin})^2} \right]} \quad (1)$$

where P_j^k / Q_j^k denote the field measured active, reactive power respectively for the specific system disturbance k with data size n_k ; P_j^k and Q_j^k are the corresponding simulated active, reactive power; P'^{kmax} / P'^{kmin} and Q'^{kmax} / Q'^{kmin} denote the field measured active, reactive power oscillation upper and lower limits respectively.

Obviously E^k is a random dependent variable changing with predictor variables, e.g. time of day X_1 , day of week X_2 , month of year X_3 , temperature in the load area X_4 , initial voltage of the load bus $X_5 = U_0^k$, voltage temporarily descending magnitude X_6 (seen in Eq. 2), voltage temporarily ascending magnitude X_7 (seen in Eq. 3), actual initial active power $X_8 = P_0^k$, actual initial reactive power $X_9 = Q_0^k$, power oscillation magnitudes $X_{10} = \Delta S^k$ (seen in Eq. 4)) etc.. MLR model for load model output error is built as (5) accordingly.

$$X_6 = \Delta U^{kmin} = (U_0^k - U^{kmin}) / U_0^k \quad (2)$$

$$X_7 = \Delta U^{kmax} = (U^{kmax} - U_0^k) / U_0^k \quad (3)$$

$$\Delta S^k = \sqrt{\frac{(P_j'^{kmax} - P_j'^{kmin})^2 + (Q_j'^{kmax} - Q_j'^{kmin})^2}{(P_0^k)^2 + (Q_0^k)^2}} \quad (4)$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon \quad (5)$$

where U^{kmax} / U^{kmin} denote the field measured bus voltage upper and lower limits; $\beta_0, \beta_1, \dots, \beta_p$ are the regression coefficients, relating the revised load model output error $Y = E^k$ to the explanatory variables X_1, X_2, \dots, X_p ; ε is the residual error accordingly, assumed to be normally distributed with its mean value $E(\varepsilon) = 0$ and variance value $V(\varepsilon) = \sigma^2$.

After load model parameters history database has been built for all of the N_s history disturbances, simulate the other $N_s - 1$ disturbances and compute the corresponding revised load model output error E^k with one identified load model parameters in the database. Now N_s dimensional revised load model output error vector Y is obtained in the form $Y = [E^1, E^2, \dots, E^{N_s}]^T$, with N_s dimensional vector X_1, X_2, \dots, X_p for each independent variable, i.e. $X_1 = [x_{11}, x_{12}, \dots, x_{1N_s}]^T$ etc..

Since all the independent variables can be obtained online or determined by a preset disturbance k , real-time E^k can be predicted by the MLR model introduced in detail in this section as a match quality indicator measuring whether the chosen load model parameters are suitable to be selected as real-time load parameters or not. After all load model parameters history data are chosen to perform multiple linear regressions in order to predict E^k for the online conditions, the load model parameters with the smallest E^k prediction value are chosen as online load parameters.

Standardizing Variables. Let \bar{x}_i and s_i be the sample average and sample standard deviation of the x_{ij} ($j=1, \dots, N_s$), seen in Eq. 6. When values of variables are larger in magnitude, it is advantageous to code each variable x_i by $x'_i = (x_i - \bar{x}_i) / s_i$.

$$\bar{x}_i = \sum_{j=1}^{N_s} x_{ij} / N_s, \quad s_i = \sqrt{\sum_{j=1}^{N_s} (x_{ij} - \bar{x}_i)^2 / (N_s - 1)}. \quad (6)$$

Multiple Linear Regression. Substitute all the remained predictor values into Eq. 5, get a matrix form equation as:

$$Y = X\beta + \varepsilon. \quad (7)$$

where $X = [e, X_1, X_2, \dots, X_p]$, $e = [1, 1, \dots, 1]_{N_s \times 1}^T$, $\beta = [\beta_0, \beta_1, \dots, \beta_p]^T$, $\varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{N_s}]^T$.

Solve the following equation in order to find least squares estimates $\hat{\beta}$ of regression coefficients β .

$$\hat{\beta} = (X^T X)^{-1} X^T Y. \quad (8)$$

and the regression function is:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_p X_p. \quad (9)$$

The estimate of variance σ^2 is:

$$\hat{\sigma}^2 = \frac{(Y - \hat{\beta}X)^T (Y - \hat{\beta}X)}{N_s - p - 1} = \frac{S_{SE}}{N_s - (p+1)}. \quad (10)$$

where S_{SE} is the sum of squared errors. However the estimated $\hat{\beta}$ and $\hat{\sigma}^2$ can not be used in the regression function to predict yet.

Outlier Check. When standard error absolute value $|(y_i - \hat{y}_i) / \hat{\sigma}| > 2$ or 3, this data is regarded as an outlier and should be deleted^[12], or check the original data again.

Variable Selection. If at least one predictor highly depends on other predictors, the data is said to exhibit multicollinearity. This can be found by correlation coefficient $\rho(X_{i_1}, X_{i_2})$ computation^[11-12] between any two predictors: X_{i_1} and X_{i_2} . If $|\rho(X_{i_1}, X_{i_2})| > 0.7$ and $|\rho(X_{i_1}, Y)| > |\rho(X_{i_2}, Y)|$, X_{i_1} has a stronger relationship with Y than X_{i_2} , and should be remained in the regression model while X_{i_2} should be treated properly depending on the value of $|\rho(X_{i_2}, Y)|$. In the test case, X_{i_2} is deleted.

The Coefficient of Determination. The coefficient of determination, denoted by R^2 , is given by $R^2 = S_{SR} / S_{ST} = S_{SR} / (S_{SR} + S_{SE}) \in [0, 1]$. The total sum of squares S_{ST} measures the total amount of variation in observed y values, defined in Eq. 11. Regression sum of squares S_{SR} is the amount of total variation that is explained by the regression model, seen in Eq. 12. The higher the value of R^2 , the more successful is the simple linear regression model in explaining y variation. As in polynomial regression, R^2 is often adjusted for the number of parameters in the model by the formula Eq. 13.

$$S_{st} = \sum_{i=1}^{N_s} (y_i - \bar{y})^2. \quad (11)$$

$$S_{SR} = \sum_{i=1}^{N_s} (\hat{y}_i - \bar{y})^2 = S_{ST} - S_{SE}. \quad (12)$$

$$R_a^2 = \frac{(N_s - 1)R^2 - p}{N_s - (p + 1)} = 1 - \frac{S_{SE} / (N_s - p - 1)}{(S_{SE} + S_{SR}) / (N_s - 1)}. \quad (13)$$

The Model Utility Test. The value of R^2 is sometimes deceptive because it can be greatly inflated by using a large number p of predictors relatively to the sample size N_s . The model utility test in linear regression involves the null hypothesis $H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$, which says that there is no useful relationship between Y and any of the predictors. The test is based on a statistic that has a particular F distribution by the formula

$$F = (S_{SR} / p) / [S_{SE} / (N_s - p - 1)] \sim F(p, N_s - p - 1). \quad (14)$$

If the proportion of explained variation is high relative to unexplained, e.g. in the rejection region for a level α test $F > F_{1-\alpha}(p, N_s - p - 1)$, we would naturally want to reject H_0 and confirm the utility of the model.

Model Parameters Validation. A test for the null hypothesis $H_{0j}: \beta_j = 0$, uses the t statistic value $t_{\hat{\beta}_j}$, which means that there is no useful relationship between Y and the predictor X_j .

$$t_{\hat{\beta}_j} = \hat{\beta}_j / S_{\hat{\beta}_j} \sim t(N_s - p - 1).$$

$$S_{\hat{\beta}_j} = \sqrt{S_{SE} / (N_s - p - 1) \times \psi(j, j)}.$$

$$\psi = (X^T X)^{-1}. \quad (15)$$

If $t_{\hat{\beta}_j}$ is in the rejection region for a level α test $|t_{\hat{\beta}_j}| > t_{1-\alpha/2}(N_s - p - 1)$, we would naturally want to reject H_0 and confirm the model parameter validation. Otherwise, delete the variable X_j and make regression again.

Now the predicted point estimate of y_f when $\mathbf{x}_f = [1, x_{f(N_s+1)}, \dots, x_{f(N_s+1)}]$ is obtained by Eq. 16.

$$\hat{y}_f = \mathbf{x}_f \hat{\boldsymbol{\beta}}. \quad (16)$$

A $1-\alpha$ confidence interval for y_f is:

$$\hat{y}_f \pm t_{1-\alpha/2}(N_s - p - 1) \times \hat{\sigma} \times \sqrt{1 + 1/N_s + (\mathbf{x}_f - \bar{\mathbf{X}})(X^T X)^{-1}(\mathbf{x}_f - \bar{\mathbf{X}})^T}. \quad (17)$$

Real-time Load Model Parameter Selection Procedure. Firstly the load model parameters identification based on DE is performed for all the latest history disturbances for a period of time so as to build load model parameters database, and then history disturbances are classified according to voltage oscillation magnitudes, i.e., descending and ascending magnitudes.

Load model structure is related to system disturbance magnitude. ZIP augmented with induction motor model is valid within voltage variation scope between 0.8 p.u. and 1.2 p.u., which can be applied to short-term voltage stability^[13-14] and transient stability analysis. When voltage and frequency fluctuate larger, broad spectrum load modeling^[15] in medium-term and long-term dynamic progress needs to consider the effect of low-frequency and low-voltage load shedding^[16], low-voltage relay action^[17], etc., which is beyond the study topic in this paper and will be studied further in our future works. So we classify disturbances into small disturbance set Z_s , large disturbance set Z_L and common disturbance set Z_c as following.

$$Z_s = \left\{ k \mid (\Delta U^{kmax} < 0.01) \cap (\Delta U^{kmin} < 0.01) \right\}. \quad (18)$$

$$Z_c = \left\{ k \mid \left\{ \left(\Delta U^{k_{\max}} < 0.2 \right) \cap \left(0.01 \leq \Delta U^{k_{\min}} < 0.2 \right) \right\} \cup \left\{ \left(\Delta U^{k_{\max}} \geq 0.01 \right) \cap \left(\Delta U^{k_{\min}} < 0.01 \right) \right\} \right\}. \quad (19)$$

$$Z_L = \{ k \mid 0.2 \leq \Delta U^{k_{\min}} \}. \quad (20)$$

Secondly using the identified load model parameters for each disturbance in the same class, MLR model is established to predict the real-time revised load model output errors. At last choose the load model parameters whose point estimate is the smallest in the class as the real-time load model parameters after all the regression models have been accomplished.

The selection procedure is performed online every 5 minutes or so to find out the best load model parameters from history database since the MLR CPU time is short enough for online analysis, which serves for online voltage stability, transient stability analysis, etc..

Test Case

In this section, the results of applying the proposed real-time load modeling parameter selection method to a 220 kV substation in a metropolitan city in China, are presented to show its effectiveness. All computations were performed in a 2.4GHz Intel CPU, 4GB RAM machine. 49 useful disturbances were measured in the substation from February 2010 to May 2012 by means of voltage oscillation detection. The base value of original capacity and voltage are chosen as $S_B=100\text{MVA}$, $U_B=220\text{kV}$.

The 49 disturbances are classified into 3 sets according to disturbance classification criterion in Eq. 18~ Eq. 20, leading to 31 disturbances left in set Z_c . The load model parameters are firstly identified by means of the DE algorithm proposed in [10] with the same load model parameters limits.

Since the prediction capability of the model can only be found in new, unseen measured data, the whole data set can be divided into the training data set and the test data set testifying real-time load model parameters matching ability of the proposed method. The last 3 disturbances are used as 3 online data and the others are utilized to establish MLR models. In this paper we only present results of common disturbance set Z_c , and do not discuss large disturbance set Z_L . So $N_s=28$.

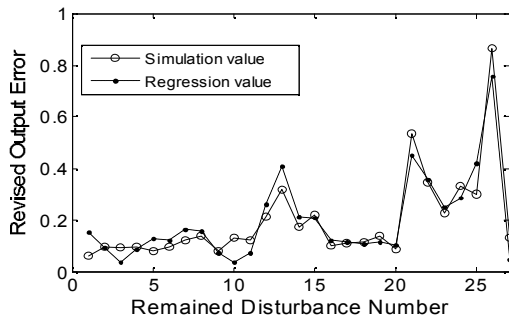


Fig.1. y_i versus \hat{y}_i for the 1st disturbance.

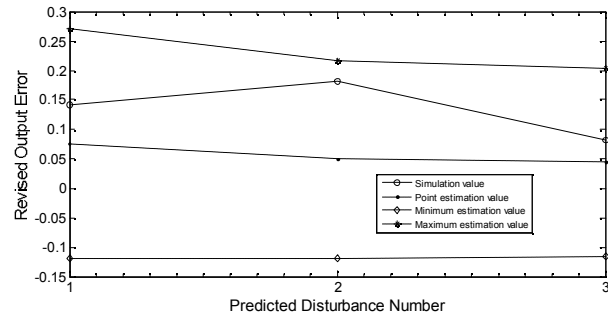


Fig.2. Prediction based on the 2st disturbance.

All the disturbances pass F tests. Fig. 1 shows y_i versus \hat{y}_i for the first disturbance data. It is clearly shown that revised load model output error fluctuates largely with different disturbances under different operation conditions along with time variation when load model parameters built on one disturbance field measurement data are adopted to simulate all the other disturbances, and the error does not definitely become bigger when disturbance occurs longer time later. Table 1 gives the regression coefficients and coefficient of determination for two regression models according to two disturbances respectively. In the first disturbance, $|\rho(X_6, X_{10})| = |0.895| > 0.7$ and $|\rho(X_6, Y)| = |-0.254| < |\rho(X_{10}, Y)| = |0.608|$, so predictor X_6 is deleted, and because of $|(y_{27} - \hat{y}_{27}) / \hat{\sigma}| > 2$, so the 27th disturbance data is outlier and deleted from the linear regression model.

Table 1 Regression parameters and Computational time

Disturbance	1	2	Disturbance	1	2
β_0	0.19744	0.19136	β_7	0.02271	0.01379
β_1	-0.02179	-0.01659	β_8	0.04273	0.04949
β_2	-0.00554	0.00851	β_9	0.06992	0.06768
β_3	0.14318	0.11203	β_{10}	0.03061	0.01622
β_4	-0.14826	-0.10913	R_a^2	0.83140	0.82064
β_5	0.02765	0.00245	Time (s)	0.327	0.312
β_6	-	-			

Point estimation and 1-0.05 confidence interval of y_f for the last 3 disturbances in set Z_c are obtained through the built MLR models; Fig. 2 shows the prediction effects based on the second disturbance. The estimations of y_f are surrounding the real load model output error by simulation using the same load model parameters.

After all regression models of the disturbances in each set are obtained, estimation values of y_f are compared with each other, and the load model parameter of the smallest one is chosen as real-time load model parameter. Table 2 shows some typical estimation values and real simulation values of y_f .

Table 2 Estimation values of y_f

Num.	\hat{y}_{f1}	\hat{y}_{f2}	\hat{y}_{f3}	y_{f1}	y_{f2}	y_{f3}	R_a^2
2	0.076	0.049	0.044	0.142	0.182	0.082	0.82
13	0.024	0.015	0.220	0.117	0.116	0.170	0.85
17	0.180	0.128	-0.055	0.150	0.197	0.100	0.77
21	0.692	0.698	0.404	0.605	0.801	0.536	0.88
23	0.715	0.637	0.127	0.502	0.684	0.392	0.73
26	0.145	0.103	0.590	0.185	0.200	0.272	0.80

Table 3 Part predictor values

Num.	Year	Date	Time	Te.[$^{\circ}$]	U_0 [pu]	P_0 [pu]	Q_0 [pu]
13	2010	12-21	07:19:42	24.8	1.064	0.486	0.078
17	2011	4-17	13:09:57	23.0	1.058	0.696	0.124
22	2011	10-1	05:22:34	23.4	1.067	0.467	0.075
29	2012	4-29	03:17:57	22.0	1.061	0.318	0.090
30	2012	4-29	03:20:07	22.0	1.061	0.315	0.085
31	2012	5-6	15:35:48	29.9	1.045	0.635	0.119

Depicted as bold character in Table 2, \hat{y}_{f1} and \hat{y}_{f2} of the 13th disturbance, \hat{y}_{f3} of the 17th disturbance are the smallest among all the disturbances, so the load model parameters identified by the 13th disturbance data are chosen as real-time load parameters at the time when the 29th disturbance occurred and also at the time when 30th disturbance occurred, and use these load parameters to simulate the 29th (seen in Fig. 3) and 30th disturbances with their revised output errors 0.117/0.116 respectively; the load model parameters identified by the 17th disturbance data are chosen as real-time load parameters at the time when the 31th disturbance occurred, and use these load parameters to simulate the 31th disturbance with its revised output errors 0.100. On the contrary, \hat{y}_{f1} of the 23th disturbance, \hat{y}_{f2} of the 21th disturbance and \hat{y}_{f3} of the 26th disturbance are the biggest each among all the disturbances; Simulate the 29th disturbance with the load model parameters identified by the 23th disturbance data, the revised output error is 0.502 (seen in Fig. 4). The fitting accuracy is greatly improved to use the load model parameters identified by the 13th disturbance data than by 23th disturbance data as real-time load

parameters at the time when the 29th disturbance occurred. In order to validate the prediction effect further, load model parameters are identified by aforementioned DE algorithm based on the the 29th disturbance field measurement data, with its revised output error 0.097 which is little smaller than 0.117.

It can be seen in Table 3 that the disturbance occurrence time of day, temperature, U_0 , P_0 and Q_0 among the 13th, 29th, and 30th disturbance are closely adjacent, which is a necessary but not sufficient condition of good fitting since these values of the 22th disturbance are also close to them of 29th and 30th disturbance. This also verifies better validation of statistic method than cluster and synthesis methods in load modeling.

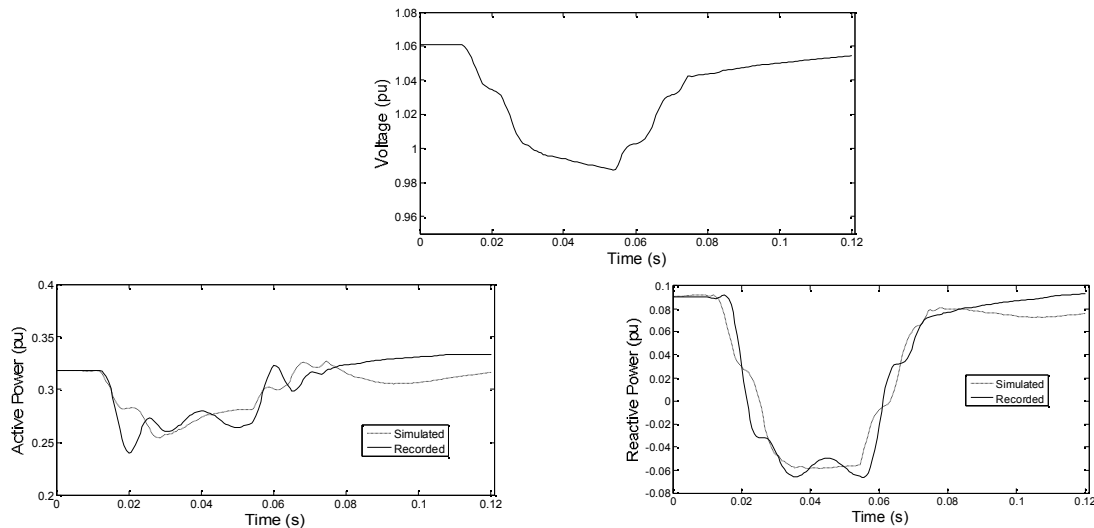


Fig.3. Voltage and power curves based on the 29th disturbance recorded data and simulation using the load model parameters identified by the 13th disturbance data.

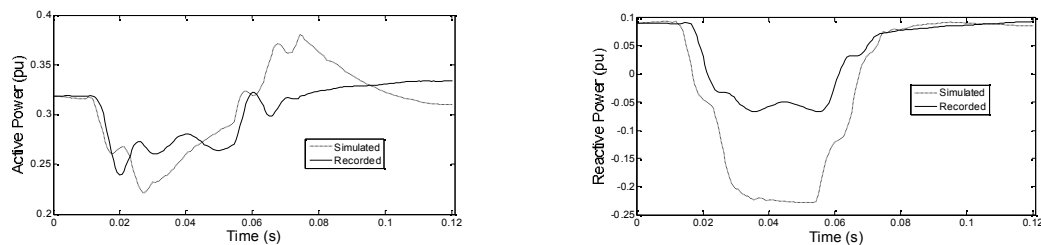


Fig.4. Power curves based on the 29th disturbance recorded data and simulation using the load model parameters identified by the 23th disturbance data.

Conclusions

Based on the MLR method, a real-time load model parameters selection method is proposed in this paper to deal with load model parameter variation problem. After disturbances classification based on voltage oscillation magnitude, the constructed MLR method can well delete the worst load model parameters and find the best load model parameters from load model parameters history database in most cases.

The proposed real-time load model parameters selection method breaks through traditional load model parameter classification mode. With the increasing number of identified load model parameters in the database, the load model dynamic characteristics can be better grasped and the chosen parameters will fit the online load with enough accuracy, while the traditional method can only classify load model parameters into several classes which cannot depict the varying load dynamic characteristics well enough.

The CPU times associated with the MLR method shown in Table 1 completely meet online speed requirements. Even after considering history load model parameters database increasing in the future, 5

minutes period is enough to exert real-time parameters selection computation. So all the load parameters are updated online periodically every 5 minutes, and serve for online voltage stability and transient stability analysis, etc..

Acknowledgements

This work was financially supported by the National Science Foundation of China under Grant No. 51377072 and Jinan University Scientific Research Starting Foundation.

References

- [1] R.M. He, J. Ma, and J.H. David: IEEE Trans. Power Syst., Vol. 21 (2006), p. 663.
- [2] H. Bai, P. Zhang, V. Ajjarapu: IEEE Trans. Power Syst., Vol. 24 (2009), p. 1145.
- [3] Q. Ai, D. Gu, C. Chen: IEEE Trans. Power Syst., Vol. 21 (2006), p. 1864.
- [4] R.M. He, W. Zhou: Automation of Electric Power Systems, Vol. 23 (1999), p. 12. (In Chinese).
- [5] X.R. Li, Y.X. Chen, and S. Su: Automation of Electric Power Systems, Vol. 25 (2001), p. 20. (In Chinese).
- [6] P. Ju, Y. Jin, F. Wu, Y. Cheng: Automation of Electric Power Systems, Vol. 28 (2004), p.64. (In Chinese).
- [7] J.H. Shi, R.M. He: Proceedings of the CSEE, , Vol. 24 (2004), p. 78. (In Chinese).
- [8] S.J. Lin, X.R. Li, H.H. Chen, et al: A New Classification and Synthesis Method for Load Dynamic Characteristics based on Field Measured Response, Proc. Int. Conf. Power System Technology, Chongqing, China, Oct. 22-26, 2006.
- [9] D. Karlsson, D.J. Hill: IEEE Trans. Power Syst., Vol. 9 (1994), p. 157.
- [10] Y.L. Huang, X. Chen, M.B. Liu, et al: Transactions of China Electrotechnical Society, Vol. 28 (2013), p. 270. (In Chinese).
- [11] Jay L Devore: Probability and statistics for engineering and the sciences, Higher Education Press, Beijing, 2004 (In Chinese).
- [12] N.M. Xia: New theory of probability and mathematical statistics, East University of Science and Technology press, Shanghai, 2011 (In Chinese).
- [13] Y.Q. Jing, X.Y. Li, X.M. Guo, et al: Automation of Electric Power Systems, Vol. 35 (2011), p. 10. (In Chinese).
- [14] S.J. Lin, X.R. Li, Y.H. Liu, et al: Proceedings of the CSEE, Vol. 29 (2009), p. 14. (In Chinese).
- [15] P. Ju: The theory and method of power system modeling, Science Press, Beijing, 2010 (In Chinese).
- [16] S.X. Zhou, L.Z. Zhu, X.J. Guo, et al: Power system voltage stability and its control, China Electric Power Press, Beijing, 2009 (In Chinese).
- [17] P. Kundur: Power System Stability and Control, McGraw-Hill, New York, 1993.