Measurement-Based Transmission Line Parameter Estimation with Adaptive Data Selection Scheme

Li, Changgang; Zhang, Yaping; Zhang, Hengxu; Wu, Qiuwei; Terzija, Vladimir

Published in:
IEEE Transactions on Smart Grid

Link to article, DOI:
10.1109/TSG.2017.2696619

Publication date:
2017

Document Version
Peer reviewed version

Measurement-Based Transmission Line Parameter Estimation with Adaptive Data Selection Scheme

Changgang Li, *Member*, *IEEE*, Yaping Zhang, Hengxu Zhang, *Member*, *IEEE*, Qiuwei Wu, *Senior Member, IEEE*, Vladimir Terzija, *Fellow*, *IEEE*

***Abstract*—Accurate parameters of transmission lines are critical for power system operation and control decision making. Transmission line parameter estimation based on measured data is an effective way to enhance the validity of the parameters. This paper proposes a multi-point transmission line parameter estimation model with an adaptive data selection scheme based on measured data. Data selection scheme, defined with time window and number of data points, is introduced in the estimation model as additional variables to optimize. The data selection scheme is adaptively adjusted to minimize the relative standard deviation (RSD) of estimated parameters. An iterative technique derived from the Newton method is adopted to solve the proposed model by fitting the relationship between the RSD and data selection scheme with exponential functions. Simulated data are applied to illustrate the performance of the proposed model. Some 500kV transmission lines from a provincial power system of China are estimated to demonstrate the applicability of the presented model. The superiority of the proposed model over fixed data selection schemes is also verified.**

***Index Terms*—Adaptive data selection, parameters estimation, power systems, supervisory control and data acquisition system (SCADA), transmission line, wide area measurement system (WAMS)**

I. Introduction

NUMERICAL simulation is essential for power system analysis, operation and control decision making. Accuracy of advanced applications, such as state estimation, dynamic simulation, and emergency control, is highly dependent on the accuracy of equipment models and parameters. Reports of some blackouts show that the detailed dynamics can hardly be reproduced by numerical simulation due to the inaccuracy of equipment parameters [1]-[3]. It is crucial to improve the quality of equipment parameters to ensure stable and secure operation of modern power systems [4]-[7].

Transmission lines are key components of power systems, and parameters of transmission lines can be calculated with designed layout theoretically or measured with field tests, assuming that the operating and geographical conditions remain unchanged [8]. However, it will lead to great error if the operating conditions are changed, e.g. the geographical condition along a transmission line can hardly be uniform and the ambient temperature is always varying. For the field test method, the transmission line to be tested needs to be switched out of service, and predefined voltages are applied to the transmission line with dedicated equipment [9]. The line parameters can be deduced with recorded voltage and current. The drawback of offline tests lies in its inability to take into account of actual operating conditions, especially for parallel transmission lines, mutual inductance of which cannot be neglected.

The deployment of supervisory control and data acquisition system (SCADA) and wide area measurement system (WAMS) provides online measured data of substations under actual operating conditions. With the online measured data, transmission line parameters can be estimated to overcome the drawbacks of the offline test method. In most literatures, phasor data are used for estimation. Various parameters were estimated in [10] with phasor data corrected. Transmission lines with distributed parameters were estimated in [11] for protection applications. Phasor data were also used in [12] to estimate a two-port model with ABDC parameters. In [13], a dynamic discrete filtering algorithm was applied to perform transmission line parameters estimation by minimizing the residual least absolute value. In [14], a single point estimation model was proposed to estimate the parameters of long transmission lines. In [15], with field SCADA data, a multi-point parameter estimation model was developed based on the Levenberg-Marquardt algorithm. In [16] and [17], WAMS and SCADA data were incorporated for estimating multi-terminal transmission lines with a nonlinear weighted least square algorithm.

For multi-point estimation models, estimation results will be more accurate theoretically if more data points are used. However, due to measurement error, more data points tend to increase the chance of nonconvergence of estimation [15]. Therefore, it is necessary to select appropriate measured data for estimation. In [16] and [17], a scheme of selecting 3 measurement points was proposed for estimation. In [10], 8 sets of simulated data within 16 minutes were chosen for estimation with even time interval of 2 minutes. In [12], 200 measurement

This work was jointly supported by National Natural Science Foundation of China (No: 51627811).

Changgang Li, Yaping Zhang, and Hengxu Zhang are all with the Key Laboratory of Power System Intelligent Dispatch and Control of the Ministry of Education (Shandong University), Jinan, 250061 China (e-mail: zhanghx@sdu.edu.cn).

Qiuwei Wu is with the Center for Electric Power and Energy (CEE), Department of Electrical Engineering, Technical University of Denmark (DTU), 2800 Kgs, Lyngby, Denmark.

Vladimir Terzija is with the School of Electrical and Electronic Engineering, the University of Manchester, Manchester, UK M60 1QD.

points with different loading conditions were chosen from measured data in 1 day with sampling intervals of 5 minutes. Through extensive tests, a data selection scheme was developed in [15] to show how to select 6 data points from a time window of 30 minutes.

The aforementioned data selection schemes in literatures are different which indicates that there is no uniform data selection scheme suitable for all lines. Those data selection schemes select a fixed number of data points from fixed length of time window regardless of measurement accuracy. The measurement accuracy, however, varies from line to line due to the measurement noise of remote terminal units (RTU) and phasor measurement units (PMU). Therefore, when estimating transmission line parameters with measured data of different accuracy, fixed data selection scheme is not suitable for all transmission lines. It is necessary to choose appropriate data selection scheme for different lines according to the unique measurement accuracy of each transmission line.

This paper proposes a transmission line parameter estimation model with adaptive data selection scheme to estimate parameters considering different measurement error. The major contribution of this paper is twofold. First, adaptive data selection scheme is introduced to the estimation model as additional variables to optimize. The solved data selection scheme is adaptive to lines to improve the accuracy of estimated parameters. Second, relative standard deviation (RSD) instead of absolute error of estimated parameters is chosen to minimize in the proposed model. It improves the practicability of the proposed model since the absolute error cannot be obtained due to the unavailability of accurate transmission line parameters.

The rest of this paper is organized as follows. The basic transmission line parameter estimation models are introduced in section II. The parameter estimation model with adaptive data selection scheme is proposed and solved in section III. The detailed explanation of why RSD can be chosen as objective to minimize is also presented in section III. To demonstrate the feasibility of the proposed method, seventy 500kV transmission lines from a provincial power system of China are estimated based on measured data in section IV, followed by conclusions in section V.

II. Basic Parameter Estimation Models

A. Single-Point Parameter Estimation Model

The lumped symmetry π equivalent model of transmission lines illustrated in Fig. 1 is widely used for power system analysis. The resistance R, reactance X, and shunt susceptance B are parameters to be estimated. U, I, P, and Q are voltage, current, active power, and reactive power of the sending side (with subscript s) or the receiving side (with subscript r).

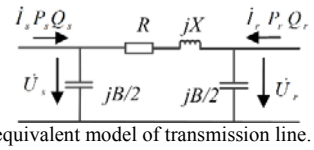


Fig. 1. Symmetry π equivalent model of transmission line.

Transmission line parameter can be estimated with SCADA or WAMS data. Since SCADA is a basic infrastructure of modern power systems and covers wider areas than WAMS [15], the SCADA data are used in this paper to build the following parameter estimation model.

With an independent set of variables $\{R, X, B, \dot{U}_r, \dot{I}_r\}$, $\dot{U}_s$, $\dot{I}_s$, P_s, Q_s, P_r, and Q_r can be expressed as follows,

$$\begin{cases} \dot{U}_s = \dot{U}_r + \left[j\dot{U}_r \dfrac{B}{2} - \dot{I}_r \right](R + jX) \\ \dot{I}_s = jB\left(\dot{U}_s + \dot{U}_r\right)/2 - \dot{I}_r \\ P_s = \mathrm{Re}\left(\dot{U}_s \dot{I}_s^*\right) \\ Q_s = \mathrm{Im}\left(\dot{U}_s \dot{I}_s^*\right) \\ P_r = \mathrm{Re}(\dot{U}_r \dot{I}_r^*) \\ Q_r = \mathrm{Im}(\dot{U}_r \dot{I}_r^*) \end{cases} \tag{1}$$

If the angle of $\dot{U}_r$ is treated as reference, i.e., the angle of $\dot{U}_r$ is 0, the independent set can be rewritten as $\{R, X, B, U_r, I_r, \theta_r\}$ where θ_r is the power factor angel of the receiving side. With the new independent set, $\dot{U}_s$ and $\dot{I}_s$ can be expressed as,

$$\begin{cases} \dot{U}_s = U_r + \left[jU_r \dfrac{B}{2} - I_r\left(\cos\theta_r + j\sin\theta_r\right) \right](R + jX) \\ \dot{I}_s = jB\left(\dot{U}_s + U_r\right)/2 - I_r\left(\cos\theta_r + j\sin\theta_r\right) \end{cases} \tag{2}$$

With measured data, line parameters are usually estimated in literatures with such method as least square (LS) algorithm to minimize the residual of estimated quantities. However, due to the unavailability of phase angles in SCADA, for $\dot{U}_s$ and $\dot{I}_s$, only their magnitudes can be compared with their measured counterparts. Therefore, the residual to minimize can be defined as,

$$\sqrt{\frac{1}{8}\sum_{i=s,r}\left(\left(\frac{\varepsilon_i^U}{U_i^m}\right)^2 + \left(\frac{\varepsilon_i^I}{I_i^m}\right)^2 + \left(\frac{\varepsilon_i^P}{P_i^m}\right)^2 + \left(\frac{\varepsilon_i^Q}{Q_i^m}\right)^2\right)} \tag{3}$$

where U^m_i, I^m_i, P^m_i, and Q^m_i are measured value of U, I, P, and Q of side i, and ε^U_i, ε^I_i, ε^P_i, and ε^Q_i are the difference between estimated and measured U, I, P, and Q of side i. The difference between estimated and measured quantities at side i is defined as,

$$\varepsilon_i^V = V_i - V_i^m$$

where V represents U, I, P, or Q.

With the U_s I_s, P_s, Q_s, P_r, and Q_r equations obtained in (1) and (2), the 6 independent variables $\{R, X, B, U_r, I_r, \theta_r\}$ can be estimated with one measured data point, i.e. a group data of U, I, P, and Q sampled at the same time. However, since measurement errors are inevitable for field SCADA data, more data points are preferred to get better results by increasing redundancy of estimation equations. It leads to the following multi-point estimation model.

B. Multi-point Parameter Estimation Model

With N data points selected from time window $[t, t+T]$, (3) can be applied to measured quantities of the N data points, and the objective function of the multi-point parameter estimation model can be expressed as minimization of the following residual,

$$Res=\sqrt{\frac{1}{8N}\sum_{k=1}^{N}\sum_{i=s,r}\left(\left(\frac{\varepsilon_{ik}^{U}}{U_{ik}^{m}}\right)^2+\left(\frac{\varepsilon_{ik}^{I}}{I_{ik}^{m}}\right)^2+\left(\frac{\varepsilon_{ik}^{P}}{P_{ik}^{m}}\right)^2+\left(\frac{\varepsilon_{ik}^{Q}}{Q_{ik}^{m}}\right)^2\right)} \quad (4)$$

where t is the starting time of the time window, T is the window size, and k is the index of selected data points.

For a series of measured data, they can be divided into W successive windows: $[t_0, t_1]$, $[t_1, t_2]$, ⋯, and $[t_{W-1}, t_W]$ with uniform T, i.e. t_w-t_{w-1}=T for w=1, 2…W. For window $[t_{w-1}, t_w]$, N data points can be selected and the multi-point estimation method can be applied to get the estimated parameter $x_{ew,}$ where x could be R, X, or B. With the W windows, the average value and standard deviation of estimated parameters can be calculated as,

$$\begin{aligned} x_{av}&=\frac{1}{W}\sum_{w=1}^{W}x_{ew}\\ S_x&=\sqrt{\frac{1}{W}\sum_{w=1}^{W}\left(x_{ew}-x_{av}\right)^2}\end{aligned} \quad (5)$$

and the RSD of estimated parameters can be defined as,

$$\sigma_x=\frac{S_x}{x_{av}} \quad (6)$$

The average value x_{av} can be treated as the final estimated result, and its credibility can be evaluated with σ_x which measures the dispersibility of estimated results in W windows.

When the accurate values of the transmission line parameters are available, the absolute error of the estimated parameter can be obtained as,

$$\Delta_x=\left|\frac{x_{av}-x_0}{x_0}\right|\times 100\% \quad (7)$$

where x_0 is the accurate value of parameter x.

C. Effects of T and N on Estimation

For modern measurement systems, a large amount of measured data are available. For the SCADA data with resolution of 1 minute, 1440 frames of data can be obtained per day. For PMU data with rate of one frame per cycle, about 4~5 million frames of data can be obtained per day. For multi-point parameter estimation models, due to the computation burden and numerical convergence, it is not necessary to use all measured data. Appropriate data should be selected to feed the estimation algorithm.

The window size T affects how many windows can be evenly divided. Greater T will lead to less W and less computation burden. However, if T is too great which would result in very few W, the RSD will give little information of the distribution of estimated results. Moreover, to get reasonable estimated results, the data used for estimation should be representative, i.e., the diversity of loading levels in a time window should be as much as possible. Due to the repetitive feature of load variation, a similar repetitive mode can be observed for transmission lines in different days. A typical window size of 1 to 2 days is suitable to cover the diversity of operation modes.

The number of data points for estimation in a time window, N, affects the estimation process in two ways. Firstly, greater N will lead to more equations of (1), and the computation burden will be increased exponentially. Secondly, for a given time window $[t_{w-1}, t_w]$, the diversity of line loading levels is fixed, i.e. the difference between the peak and valley loading levels is fixed. Greater N will lead to less difference between selected data points, and numerical convergence will deteriorate.

T and N jointly determine the data selection scheme. The whole data with bad data eliminated are firstly divided into W windows of length T. Then N data points are selected from each window for estimation. To cover as many loading levels as possible, the measured data in each time window are sorted with ascending current, and N data points with equal current interval are picked in this paper.

III. Multi-Point Parameter Estimation Model with Adaptive Data Selection Scheme

A. Basic Adaptive Parameter Estimation Model

The target of parameter estimation is to get the parameters of transmission lines as accurate as possible. Therefore, the objective of parameter estimation model can be intuitively defined as the minimization of absolute error of estimated parameters,

$$\begin{aligned} &\min\ \Delta_X,\Delta_B\\ &s.t.\\ &\begin{cases}\Delta_X=F(T,N)\\ \Delta_B=G(T,N)\\ N\ge 1\\ T\ge Nh\end{cases}\end{aligned} \quad (8)$$

where h is the sampling interval of measured data, and $F(T, N)$ and $G(T, N)$ are functions to represent the relationship between the absolute errors and the data selection scheme. With T and N adjusted, (8) can be used to adaptively select measured data for estimation of each transmission line. As X and B have great influence on the characteristics of transmission line, X and B are the two key parameters focused on in this paper.

Two difficulties are encountered when solving (8). First, since the accurate values of estimated parameters are unknown under the circumstance of estimation, the absolute error of estimated parameters is unavailable. Second, $F(T,N)$ and $G(T,N)$ are implicit functions of T and N. It is hard to get their analytical expression and solve the model. To overcome the two difficulties, (8) should be transformed into a solvable form which is discussed in the next subsection.

B. Improved Adaptive Parameter Estimation Model

The absolute error (7) and the RSD (6) are the two indices of estimated parameters. To find the relationship between the absolute error and the RSD, some simulated data are generated in the following way to perform parameter estimation with (4).

(1) Get U, P, and Q of receiving side from SCADA. General operation modes of the transmission line are kept in the data.

(2) Assuming that U, P, and Q at the receiving side are accurate, the corresponding I at the receiving side and U, I, P, Q at the sending side are calculated with given R_0, X_0, and B_0. Now the simulated data with no measurement error at both

sides are generated.

(3) Noise following normal distribution (Gaussian distribution) is added to the simulated data to construct measured data with noise. It should be noted that the noise generated in computers are pseudo-random numbers. To draw general conclusions, 100 tests were performed to generate different noise. Average results of the 100 tests are examined to reveal the influence of T and N on the estimated parameters.

A 500 kV transmission line (line 26# in section IV) with given R_0=0.7232Ω X_0=10.1248Ω and B_0=73.766μS is modeled. The sampling time interval is 1 minute, and generated data of 66 days without noise are shown in Fig. 2. Noise with normal distribution is then added in the following subsections for further analysis.

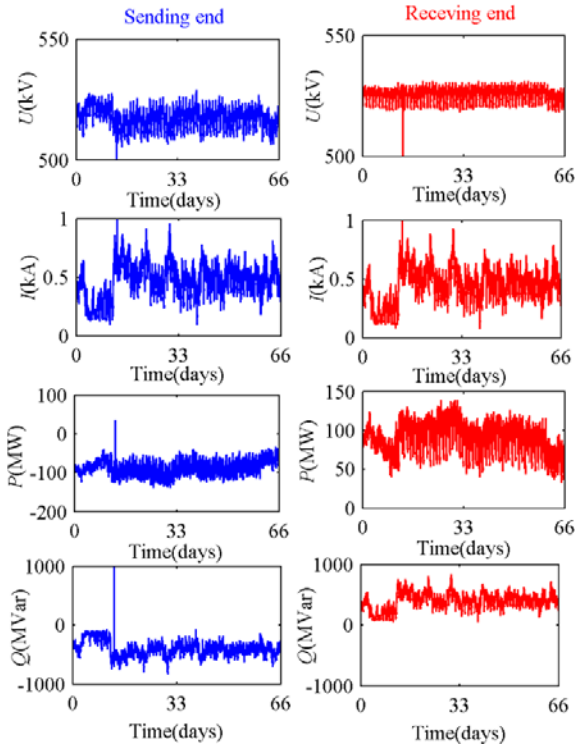


Fig. 2. Simulated data for line 26# without noise.

With N=200, the estimated results with measurement error of 3% and 6% are illustrated in Fig. 3 as right-pointing triangle and circle markers with variable T.

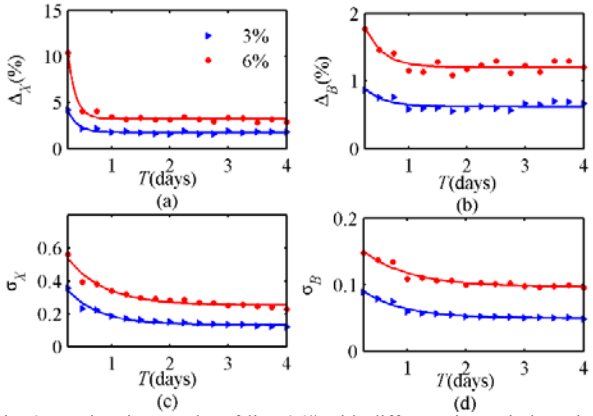


Fig. 3. Estimation results of line 26# with different time window sizes and different noise levels (N=200).

It can be seen from Fig. 3 that with the same noise level, time window greatly affects the accuracy of estimation results. Both absolute error and RSD decrease with increasing window size. The trend of RSD is similar to that of absolute error. With the same time window, both absolute error and RSD increases with increasing noise level.

Similar conclusions can be drawn with fixed T and variable N. With T=1 day, the estimated results are shown in Fig. 4 with variable N. Both absolute error and RSD decrease with increasing number of selected data points. With the same N, greater measurement error would lead to less accurate estimated results.

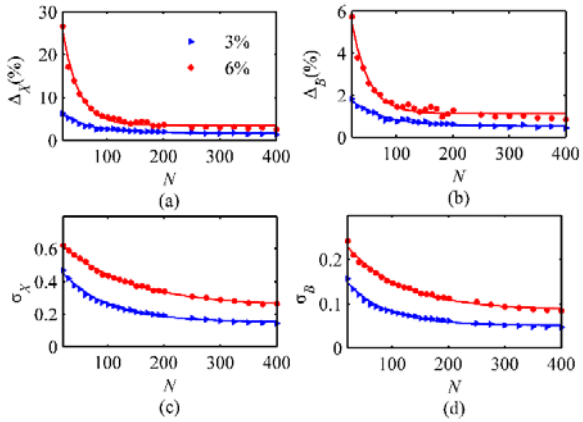


Fig. 4. Estimation results of line 26# with different data points number and different noise levels (T=1day) .

From Fig. 3 and Fig. 4, it can be concluded that when RSD is reduced, the estimated results are improved. Therefore, for field power systems where accurate transmission line parameters are unavailable, the accuracy of the estimation results can be reflected by the RSD. The minimum of the RSD can be selected as the objective to obtain the estimated parameters. Thus, the difficulty of getting absolute errors when accurate values of parameters are unknown is overcome, and an improved parameter estimation model with adaptive data selection scheme is formulated as,

$$\begin{aligned} &\min \ \sigma_X, \sigma_B \\ &s.t. \\ &\begin{cases} \sigma_X = f(T,N) \\ \sigma_B = g(T,N) \\ N \ge 1 \\ T \ge Nh \end{cases} \end{aligned} \tag{9}$$

where $f(T, N)$ and $g(T, N)$ are nonlinear functions of the relationship between the RSD and the data selection scheme.

The estimation model (9) is a multi-objective optimization problem. To simplify the solution to model (9), the estimation model is transformed into a single-objective optimization problem in this paper by minimizing the average of σ_X and σ_B,

$$\min \ 0.5(\sigma_X + \sigma_B) \tag{10}$$

C. Iterative Solution to the Proposed Model

To solve the nonlinear model (9), an iterative technique derived from the Newton method is adopted in this paper. Suppose in the n-th iteration, the window size is T_n and the

number of selected data points is N_n, σ_{Xn} and σ_{Bn} can be determined with the parameter estimation model. The correction equation of T and N, i.e. ΔT_n and ΔN_n, can be described as,

$$\begin{pmatrix} \sigma_{Xn}-\sigma_{X\min n} \\ \sigma_{Bn}-\sigma_{B\min n} \end{pmatrix} = J \begin{pmatrix} \Delta T_n \\ \Delta N_n \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial T} & \frac{\partial f}{\partial N} \\ \frac{\partial g}{\partial T} & \frac{\partial g}{\partial N} \end{pmatrix} \begin{pmatrix} \Delta T_n \\ \Delta N_n \end{pmatrix} \quad (11)$$

where $\sigma_{X\min n}$ and $\sigma_{B\min n}$ are the minimum value of σ_X and σ_B in the n-th iteration. The elements of the Jacobin matrix J are partial derivatives of $f(T, N)$ and $g(T, N)$ over T and N.

Similar to $F(T, N)$ and $G(T, N)$, $f(T, N)$ and $g(T, N)$ are implicit functions. The partial derivatives in (11) cannot be explicitly expressed. Moreover, without knowing $f(T, N)$ or $g(T, N)$, $\sigma_{X\min n}$ or $\sigma_{B\min n}$ cannot be obtained. To overcome the difficulties, an approximation technique is introduced as follows to solve the estimation model for better practicability.

According to Fig. 3 and Fig. 4, the relationship between the RSD and T and N generally follows exponential functions. Thus, $f(T, N)$ and $g(T, N)$ can be decoupled into $f_N(T)$ and $g_N(T)$ with fixed N, and $f_T(N)$ and $g_T(N)$ with fixed T. The decoupled functions can be approximated with exponential functions as,

$$f_N(T) = A_{XT} + C_{XT}\, e^{-T/T_X} \quad (12)$$

$$g_N(T) = A_{BT} + C_{BT}\, e^{-T/T_B} \quad (13)$$

$$f_T(N) = A_{XN} + C_{XN}\, e^{-N/N_X} \quad (14)$$

$$g_T(N) = A_{BN} + C_{BN}\, e^{-N/N_B} \quad (15)$$

where A_{XT}, C_{XT}, T_X, A_{BT}, C_{BT}, T_B, A_{XN}, C_{XN}, N_X, A_{BN}, C_{BN}, and N_B are variables for fitting the 4 decoupled functions. The fitted $f_N(T)$, $g_N(T)$, $f_T(N)$, and $g_T(N)$ of the line 26# with n=1 are shown in Fig. 3 and Fig. 4 as solid lines.

With the approximated decoupled functions, the Jacobin matrix, J, can be approximated by,

$$J \approx \begin{pmatrix} \frac{\partial f_N}{\partial T} & \frac{\partial f_T}{\partial N} \\ \frac{\partial g_N}{\partial T} & \frac{\partial g_T}{\partial N} \end{pmatrix} \quad (16)$$

The minimum value of $f_N(T)$, $g_N(T)$, $f_T(N)$, and $g_T(N)$ are A_{XT}, A_{BT}, A_{XN}, and A_{BN}. Thus, $\sigma_{X\min n}$ and $\sigma_{B\min n}$ can be approximated as,

$$\begin{aligned} \sigma_{X\min n} &= \min(A_{XT}, A_{XN}) \\ \sigma_{B\min n} &= \min(A_{BT}, A_{BN}) \end{aligned} \quad (17)$$

Once J, $\sigma_{X\min n}$, and $\sigma_{B\min n}$ are obtained, the correction of T and N can be calculated with (11). The new window size T_{n+1} and the number of selected data points N_{n+1} can be updated as,

$$\begin{aligned} T_{n+1} &= T_n + \alpha \Delta T_n \\ N_{n+1} &= N_n + \alpha \Delta N_n \\ \alpha &= 0.5^l \end{aligned} \quad (18)$$

where α is the optimal factor and l is the iteration for searching the optimal factor α. If the objective (10) calculated with new T_{n+1} and N_{n+1} are greater than that calculated with T_n and N_n, α will be halved till $(\sigma_{Xn+1}+\sigma_{Bn+1})$ is less than $(\sigma_{Xn}+\sigma_{Bn})$. The maximum iteration for updating α is denoted as l_{max}.

To improve the convergence, ΔT_n and ΔN_n should be checked before updating T_{n+1} and N_{n+1}. According to extensive tests on simulated data, T has great impact on σ_{Xn} and σ_{Bn}, and ΔT_n is limited to the range of [-0.1, 0.1]. If ΔT_n is out of the range, ΔT_n is set as 0.1 or -0.1 depending on the sign of original ΔT_n. ΔN_n will be scaled accordingly.

Fig. 5 shows the flowchart of solving the proposed estimation model. After loading the measured data and eliminating bad data in step ①, initial line parameters and some arguments of the estimation process are set in step ② where n_{max} is the maximum iteration. The kernel estimation process in section II.B is shown in step ③. Then $f_N(T)$ and $g_N(T)$ are fitted with fixed N, and $f_T(N)$ and $g_T(N)$ are fitted with fixed T in step ④ for getting Jacobian, $\sigma_{X\min n}$, $\sigma_{B\min n}$, ΔT_n, and ΔN_n in step ⑤. With step ⑥, T and N are updated with (18). The RSD with the new T and N are calculated in step ⑦ for checking convergence in step ⑧. The convergence criteria are discussed in the next subsection.

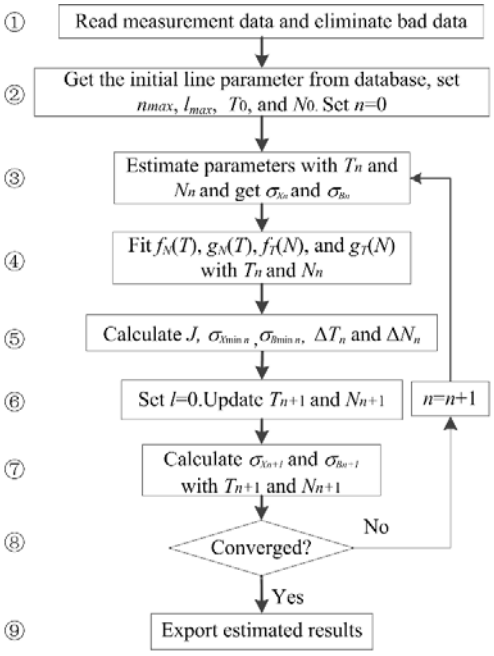


Fig. 5. Flowchart for solving the proposed estimation model.

D. Convergence Criteria

Due to the random measurement error, the inconsistency between the actual estimated results and fitted smooth exponential functions may result in failure to converge for the proposed method with typical convergence criteria. To improve convergence and practicability, three practical convergence criteria are adopted to check whether the estimation should be stopped.

(1) When objective of the (n+1)-th iteration doesn't change comparing with that of the n-th iteration, i.e., the average value of σ_{Xn+1} and σ_{Bn+1} is equal to the average value of σ_{Xn} and σ_{Bn}, the estimation will stop,

$$0.5(\sigma_{Xn+1}+\sigma_{Bn+1})=0.5(\sigma_{Xn}+\sigma_{Bn}) \quad (19)$$

(2) When the objective function of the (n+1)-th iteration is worse than the objective function of the n-th iteration and the

objective function of the n-th iteration is also worse than the objective function of the (n-1)-th iteration, the parameters estimation will stop, i.e.,

$$\begin{cases}(\sigma_{Xn+1}+\sigma_{Bn+1})/(\sigma_{Xn}+\sigma_{Bn})>=\beta \\ (\sigma_{Xn}+\sigma_{Bn})/(\sigma_{Xn-1}+\sigma_{Bn-1})>=\beta \\ \alpha<0.5^{l_{\max}}\end{cases} \quad (20)$$

where β is the threshold of deterioration (β>1.0) .

(3) When n is greater than n_{max}, the parameters estimation will stop. If estimation stops when this condition is met, it is desirable to adaptively change the initial values as discussed in section III.F.

E. Impact of Measurement Error on Estimation

To numerically check the impact of measurement error on line parameter estimation, measurement error of 3%, 6%, 9%, and 12% were added as in section III.B. Estimated results are illustrated in TABLE I. The *Res* is the residual defined in (4) with estimated parameters, and the Res_0 is the residual obtained with accurate line parameters.

TABLE I
ESTIMATED RESULTS OF LINE 26# WITH DIFFERENT LEVELS OF MEASUREMENT ERROR

Error/%	ΔX/%	ΔB/%	σ_X	σ_B	*Res*/%	Res_0/%
0	1.467×10^{-7}	1.221×10^{-7}	1.527×10^{-9}	4.492×10^{-10}	4.062×10^{-7}	2.206×10^{-14}
3	1.587	0.951	0.0561	0.0374	0.936	0.915
6	2.460	1.710	0.0983	0.0601	1.889	1.840
9	3.498	3.163	0.1244	0.0981	2.836	2.767
12	4.732	4.221	0.1693	0.1153	3.898	3.719

As shown in TABLE I, the estimated results are accurate when there is no measurement error. With increasing measurement error, the absolute error and RSD of estimated line parameters generally increase in a linear manner. Comparing *Res* with Res_0, it can be found that the residual is mainly contributed by the measurement error. With greater estimation error of line parameters, the residual contributed by inaccurate line parameters increases. For the case with measurement error of 12%, the inaccurate estimated line parameters contribute about 4.8% to the residual.

F. Adaptive Initialization of T and N

The initial values of T and N affect the convergence of the estimation model. For field measured data with error, $f(T,N)$ and $g(T,N)$ are not smooth. Inappropriate initial values of T and N may increase the number of iterations or make the estimation results trapped in local optimum. It is necessary to select appropriate initial values for different transmission lines.

In this paper, a N_0-based scheme is proposed to set up initial values of T and N. With a given number of data points used for estimation, i.e. N_0, (12) and (13) are fitted. To choose a reasonable T from (12) and (13), a small number τ ($0<\tau\leq1$) is defined to indicate how far the RSD is from its desired minimum value. Let,

$$\begin{aligned} e^{-T/T_X}&<\tau \\ e^{-T/T_B}&<\tau \end{aligned} \quad (21)$$

Thus, the initial value of T can be determined as,

$$T_1=\max\left\{-T_X\log(\tau),-T_B\log(\tau)\right\} \quad (22)$$

With T_1 selected, N_1 should be updated in a similar way as T_1 with fitted (14) and (15),

$$N_1=\max\left\{-N_X\log(\tau),-N_B\log(\tau)\right\} \quad (23)$$

Once T_1 and N_1 are determined, line parameters can be iteratively estimated with the method shown in Fig. 5.

If the estimation fails to converge within n_{max} iterations, greater τ should be adaptively set to get new initial values for estimation till converged results are obtained. For example, τ can be doubled to get reduced T_1 and N_1 to improve the convergence. An external loop can be added to Fig. 5 to control the update of τ, as shown in Fig. 6.

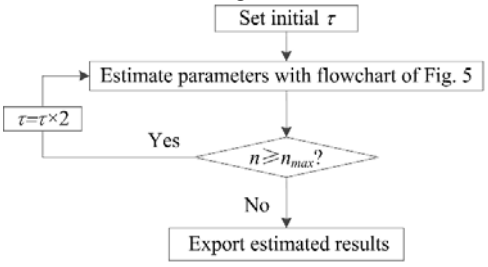


Fig. 6. Flowchart for estimation with adaptive update of τ

IV. CASE STUDIES

To verify the feasibility of the proposed model, a provincial power system in China with forty two 500kV substations and seventy one 500kV transmission lines are tested. Seventy lines are successfully estimated with three months of measured data with sampling interval of 5 minutes (288 data points per day), while the other one cannot be estimated due to loss of measured data. Through extensive tests, estimation configuration parameters in this paper are set as: β=1.05, n_{max}=50, l_{max}=10, N_0=50, and initial τ=0.02.

A. Demonstration of Estimation with Adaptive Data Selection Scheme

A transmission line (line 36#) with database parameters R=1.2497Ω, X=18.67Ω, B=298.6μS is used to demonstrate the estimation process. To show the process of selecting initial value of T and N for iteration, the fitted $f_N(T)$and $g_N(T)$ are illustrated in Fig. 7(a) as solid lines with fixed N_0=50.

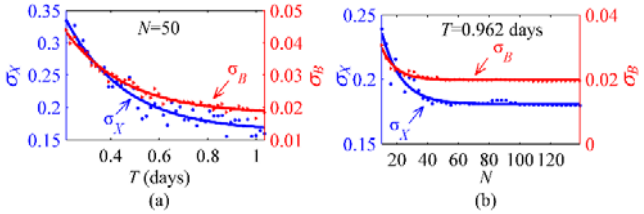


Fig. 7. The process of selecting initial T and N for line 36#.

It can be seen from Fig. 7(a) that the RSD of X and B decreases exponentially with increasing T, which verifies the conclusion made with simulated data. With the fitted $f_N(T)$ and $g_N(T)$, T_1 is determined as 0.9620 days with (22). With T_1=0.962 days, $f_T(N)$ and $g_T(N)$ are fitted in Fig. 7(b), and N_1=49 is selected with (23).

With T_1=0.962 days and N_1=49, parameters of line 36# is

estimated with the proposed estimation model iteratively. Estimation converges in 3 iterations. The fitted curves of each iteration are shown in Fig. 8, and Fig. 9 shows the iteration of σ_X, σ_B, N, and T.

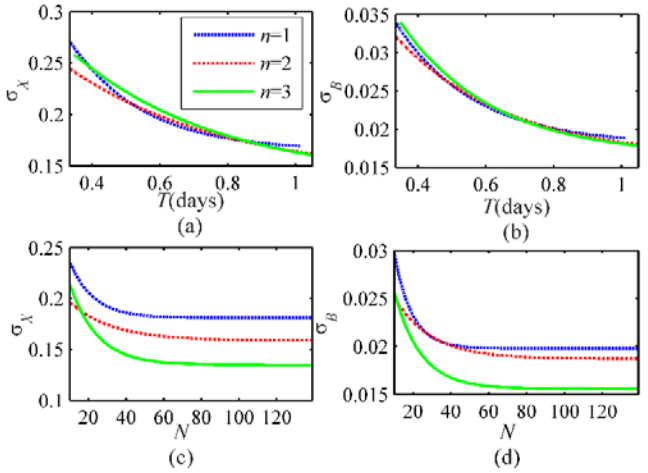


Fig. 8. Fitted $f_N(T)$, $g_N(T)$, $f_T(N)$ and $g_T(N)$ of each iteration of line 36#.

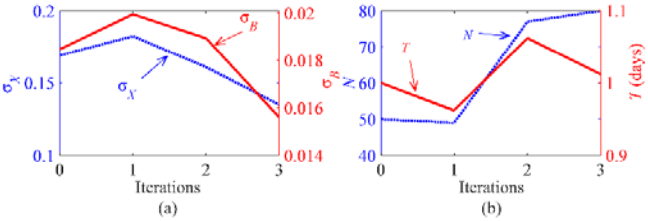


Fig. 9. Iteration of σ_X, σ_B, N, and T of line 36#.

It can be observed from Fig. 8 and 9 that the RSD decreases with iterations for better credibility and the data selection scheme is optimized. After 3 iterations, the estimation stops because of convergence criteria (1). T and N converge to 1.012 days and 80, respectively.

TABLE II lists the estimated results of line 36#. Though the final estimated results are almost the same as that of initial iteration, σ_X and σ_B reduce from 0.1690 and 0.0185 (iteration 0) to 0.1351 and 0.0156 (iteration 3). It indicates that the estimated results of iteration 3 are better than iteration 0.

TABLE II
ESTIMATED PARAMETERS OF LINE 36# OF EACH ITERATION

Iteration	0	1	2	3
X(Ω)	17.42	11.31	17.43	17.39
B(μS)	297.4	297.1	297.5	297.6

B. Adaptive Scheme v.s. Fixed Scheme

To further verify the performance of the solved scheme (scheme 1 with T=1.012 days and N=80), 6 other schemes are compared in TABLE III. Schemes 2～5 are derived from scheme 1. The data point number of scheme 2 is half of scheme 1. The time window of scheme 3 is half of scheme 1. For scheme 4, the data point number is the same as that of scheme 1, and the time window is reduced to N/288, i.e., all data points in the time window are selected. Scheme 5 has the same time window as scheme 1, and all data points in the time window are selected. Scheme 6 is the 6-point in 30 minutes scheme of [15] and scheme 7 is the 200-point in 1 day scheme of [12].

TABLE III
ESTIMATED RESULTS OF LINE 36# WITH DIFFERENT DATA SELECTION SCHEMES

Scheme	T(days)	N	X(Ω)	B (μS)	σ_X	σ_B
1	1.012	80	17.39	297.6	0.1351	0.0156
2	1.012	40	17.28	297.3	0.1468	0.0175
3	0.506	80	17.64	298.0	0.1901	0.0224
4	0.2778	80	17.58	297.9	0.3253	0.0386
5	1.012	291	17.39	297.6	0.1345	0.0155
6	1/48	6	23.32	311.28	0.7168	0.1527
7	1	200	17.41	297.4	0.1663	0.0181

It can be seen from TABLE III that the RSD of scheme 1 is much less than that of schemes 2~4. The RSD of scheme 5 is almost the same as scheme 1, but with greater computation burden. For scheme 6, the RSD is much greater than that of other schemes which indicates that the estimated results of scheme 6 are not trustable. It should be noted that there are only 6 measured data in 30 minutes for lines in the test system. No optimal selection can be performed for scheme 6 which was used in [15] to optimally select 6 points from 300 measured frames in 30 minutes (1 frame per 6 seconds). The performance of scheme 6 can be improved with measured data of higher sampling rates.

For scheme 7, its performance is similar to that of scheme 1 with slightly higher RSD. Therefore, scheme 7 is a moderate scheme for general applications, but won't always give the best estimation results. Detailed comparison between scheme 7 and the proposed adaptive data selection scheme is made in the next subsection.

C. Estimated Results of 70 Transmission Lines

With the proposed estimation model, the estimated RSD and absolute error of the 70 transmission lines sorted with ascending σ_X are shown in Fig. 10. It should be noted that in Fig. 10 (c) and (d), the absolute error is calculated with database parameters as reference since the accurate transmission line parameters are unavailable.

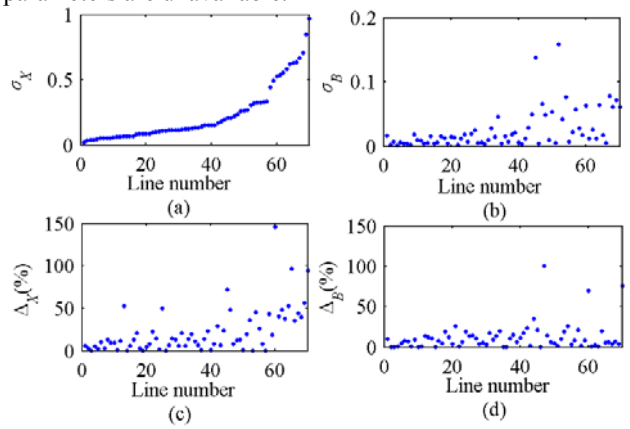


Fig. 10. Estimation results of 70 transmission lines with proposed model

As shown in Fig. 10, σ_B is much less than σ_X and σ_X should be paid more attention to. There are 51 lines with σ_X less than 0.3, and 11 lines with σ_X greater than 0.5. Those lines with greater σ_X are usually measured with worse accuracy and estimation results are not trustable. There are 45 lines with estimated X deviating from its database parameter by less than

20%, and 7 lines with estimated X deviating from its database parameter by greater than 50%.

Estimated results of 5 example lines are listed in TABLE IV. It can be seen from TABLE IV that different transmission lines have different optimal data selection schemes for estimation. The varying T and N from line to line shows that the data selection schemes are indeed adaptively set by the proposed estimation model.

TABLE IV
ESTIMATED RESULTS OF 5 EXAMPLE LINES WITH MEASURED DATA

Line	T(days)	N	X(Ω)	B (μS)	σ_X	σ_B
1#	1.024	77	12.974	239.608	0.0212	0.0169
9#	0.823	104	15.581	228.491	0.0543	0.0179
15#	1.300	101	20.485	314.318	0.0684	0.0047
48#	2.307	69	6.789	87.146	0.2360	0.0492
67#	0.385	110	5.712	154.164	0.6640	0.0778

As a comprehensive demonstration of the performance of the proposed estimation model, the estimated RSD of the proposed adaptive scheme and the fixed scheme 7 discussed in previous section are compared in Fig. 11.

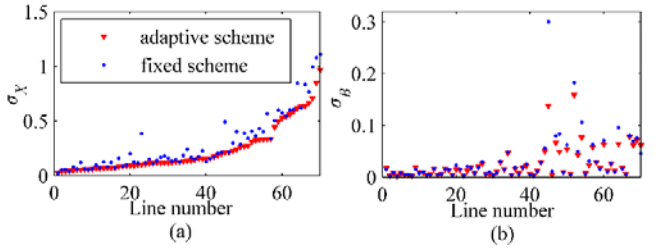


Fig. 11. Estimation results of 70 transmission lines with adaptive and fixed schemes.

As shown in Fig. 11, σ_X and σ_B of the fixed scheme are usually greater than those of the adaptive scheme. Among the 70 lines, 62 lines (88.6%) have lower σ_X with the adaptive scheme than the fixed scheme. Forty nine lines (70%) have lower σ_B with the adaptive scheme than the fixed scheme. The average improvement of σ_X of the adaptive scheme over the fixed scheme is 19.34%, while the best improvement is 74.33% (line 23#). The worst case of the 70 lines is line 1#, for which the σ_X with the adaptive scheme is 3.17% greater than the fixed scheme. The proposed estimation model with the adaptive data selection scheme generally provides better estimation results than the fixed data selection scheme.

D. Non-converged Cases

In field power systems, the estimation may fail to converge when iterations reach the maximum, e.g. line 9# with database parameter R=1.021Ω, X=14.2429Ω, and B=208.233μS. The estimated results with initial τ=0.02 are shown in Fig. 12. The estimation stops after 50 iterations. For demonstration, only the first 20 iterations are shown.

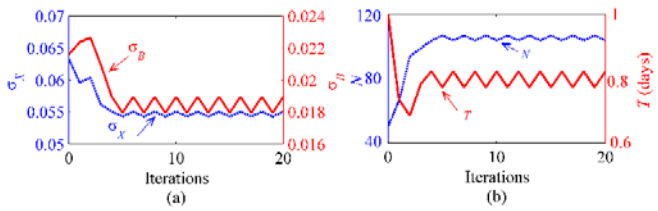


Fig. 12. Iteration of RSD and N and T for line 9# with τ=0.02.

It can be seen from Fig.12 (a) that the RSD decreases in the first few iterations. However, the estimation was trapped into oscillation after 3 iterations. When σ_X reaches minima, σ_X is 0.0543, and σ_B is 0.0179. With increased τ=0.04, the estimation of line 9# converges in 16 iteration as shown in Fig. 13. The estimated σ_X is 0.0513 and σ_B is 0.0178 which are slightly better than that in Fig. 12.

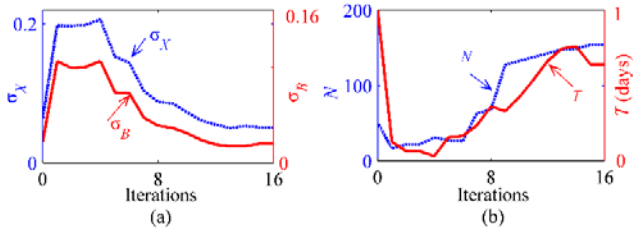


Fig. 13. Iteration of RSD and N and T for line 9# with τ=0.04.

E. Comments on Computational Efficiency

The proposed estimation model involves high computation burden. In this paper, the estimation was carried out with MATLAB® R2014a on a server with two 8-core CPUs (Intel® Xeon® E5-2640 v3@2.6GHz) with 3 months of measured data. A script-base estimation package was developed with parallel computing toolbox enabled to start 15 workers for improving estimation speed.

The average estimation time of the aforementioned 70 transmission lines is 1568s (26 minutes) per line and the average number of iteration is 25.5 per line. Twenty nine out of the 70 lines failed to converge in 50 iterations. For the other 41 lines, estimation averagely converges within 7.5 iterations per line, and the average estimation time is 515s (8.6 minutes) per line. Among the 41 lines, the longest estimation time is 3296s (55 minutes) for line 13# which converged in 28 iterations, while the shortest estimation time is 106s (1.8 minutes) for line 17# with 3 iteration. Since the data selection scheme is adaptively optimized for each line, the estimation time of lines with less iterations is not always less than that of lines with more iterations.

For such applications as parameter database calibration, the estimation will be started when new lines are commissioned, or state estimation fails to converge in some areas. Offline estimation would be appropriate for those applications, and time consumption is not the key restriction. Besides, the script-based estimation package can be further optimized with better solver, e.g., commercial solver in FORTRAN, to improve the estimation speed.

V. CONCLUSIONS

When estimating transmission line parameters with measured data, the data selection scheme has great influence on estimation results. This paper proposes a measurement- based transmission line parameter estimation model with an adaptive data selection scheme to make the estimation model suitable for transmission lines with different measurement accuracy. The RSD of estimated parameters is correlated with the absolute error of the estimated parameters. Therefore, the RSD of estimated parameters is minimized to improve the credibility of

estimated results. It is validated by tests with field measured data that the proposed estimation model is superior to models with fixed data selection schemes. It can be used for calibrating transmission line parameter database to improve power system operation and control, and is applicable for both SCADA and WAMS data.

De-noising of measured data and eliminating bad data can improve the estimation performance, which is one of the key issues that should be studied further. The empirical parameters used in the paper, such as β, n_{max}, l_{max}, N_0, and initial τ, are determined based on the cases of the test system, and should be validated with more test systems. The computation burden of the proposed model is much heavier than fixed data selection schemes. Though it is not a problem for offline applications, more efficient ways to solve the proposed model will help to improve its practicability.

REFERENCES

[1] G. Andersson, P. Donalek, R. Farmer, et al, "Causes of the 2003 major grid blackouts in North America and Europe, and recommended means to improve system dynamic performance," *IEEE Trans. Power Syst.*, vol. 20, no. 4, pp. 1922-1928, Nov. 2005.

[2] UCTE. Final report of the investigation committee on the 28 September 2003 blackout in Italy. [Online]. Available: http://www.rae.gr/old/cases/C13/italy/UCTE_rept.pdf

[3] UCTE. Final report on system disturbance on 4 November 2006. [Online]. Available: http://ecolo.org/documents/documents_in_english/blackout-nov-06-UCTE-report.pdf

[4] C. Huang, F. Li, L. Zhan, et al, "Data quality issues for synchrophasor applications Part II: problem formulation and potential solutions", *J. Mod. Power Syst. Clean Energy*, vol. 4, no. 3, 2016, pp. 353-361, Aug. 2016.

[5] Z. Huang, P. Du, D. Kosterev, et al, "Generator dynamic model validation and parameter calibration using phasor measurements at the point of connection", *IEEE Trans. on Power Syst.*, vol. 28, no. 2, pp. 1939-1949, May 2013.

[6] A. A. Hajnoroozi, F. Aminifar, H. Ayoubzadeh, "Generating unit model validation and calibration through synchrophasor measurements", *IEEE Trans. on Smart Grid*, vol. 6 no. 1, pp. 441-449, Jan. 2015.

[7] J. E. Gomez, I. C. Decker, "A novel model validation methodology using synchrophasor measurements", *Elec. Power Syst. Research*, vol. 119, pp. 207-217, Feb. 2015.

[8] Y. Wang, X. Xu, H. Xue, "Method to measure the unbalance of the multiple-circuit transmission lines on the same tower and its applications", *IET Gen. Trans. & Dist.* Vol. 10, no. 9, pp. 2050-2057, June 2016.

[9] S. Kurokawa, J. Pissolato, M. C. Tavares, et al, "A new procedure to derive transmission-line parameters: applications and restrictions", *IEEE Trans. on Power. Del.*, vol. 21, no. 1, pp. 492-498, Jan. 2006.

[10] D. Ritzmann, P. S. Wright, W. Holderbaum, et al, "A method for accurate transmission line impedance parameter estimation", *IEEE Trans. on Instru. and Meas.*, vol. 65, no. 10, pp. 2204-2213, Oct. 2016.

[11] Y. Liao, M. Kezunovic, "Online optimal transmission line parameter estimation for relaying applications," *IEEE Trans. Power Del.*, vol. 24, no. 1, pp. 96–102, Jan. 2009.

[12] D. Shi, D. J. Tylavsky, K. M. Koellner, et al, "Transmission line parameter identification using PMU measurements," *Eur. Trans. Elect. Power*, vol. 21, no. 4, pp. 1574–1588, May 2011.

[13] Al-Othman, K. M. El-Naggar, M. E. Alsharidah, "On-line estimation of transmission line parameters using synchronized measurements," *Electr. Power Comp. Syst.*, vol. 44, no. 2, pp. 233-239, Jan. 2016.

[14] C. S. Indulkar, K. Ramalingam, "Estimation of transmission line parameters from measurements," *Int. J. Elect. Power Energy Syst.*, vol. 30, no. 5, pp. 337–342, June 2008.

[15] Y. Wang, W. Xu, J. Shen, "Online tracking of transmission-line parameters using SCADA data," *IEEE Trans. Power Del.*, vol. 31, no. 2, pp. 674-682, Apr. 2016.

[16] S. S. Mousavi-Seyedi, F. Aminifar, S. Afsharnia, "Parameter estimation of multiterminal transmission lines using joint PMU and SCADA data," *IEEE Trans. Power Del.*, vol. 30, no. 3, pp. 1077–1085, June 2015.

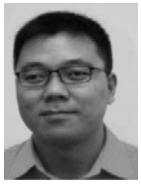
[17] S. S. Mousavi-Seyedi, F. Aminifar, S. Afsharnia, "Application of WAMS and SCADA Data to Online Modeling of Series-Compensated Transmission Lines," [Online]. Available: http://ieeexplore.ieee.org/document/7401091/

BIOGRAPHY

Changgang Li (M'15) received the B.E. and Ph.D. degrees in Electrical Engineering from Shandong University, Jinan, China, in 2006 and 2012, respectively. He was a Post-Doctoral Fellow with the Department of Electrical Engineering & Computer Science, the University of Tennessee, Knoxville, from 2012 to 2014. He is now an Associate Research Fellow with the School of Electrical Engineering, Shandong University, China. His research interests are power system operation and control.

Yaping Zhang received the B.E. degree in Electrical Engineering from Shandong University, Jinan, China, in 2014. She is now a M.E. candidate with the School of Electrical Engineering, Shandong University, China. Her main research interest is power system analysis and monitoring.

Hengxu Zhang (M'06) received the B.E. degree in Electrical Engineering from Shandong University of Technology, China, in 1998, and the M.S. and Ph.D. degrees in Electrical Engineering from Shandong University, China, in 2000 and 2003, respectively. He was a Post-doctoral Fellow in the Department of Electrical & Computer Engineering, Virginia Tech, from 2005 to 2006. He is now a Professor with the School of Electrical Engineering, Shandong University, China. His main research interests are power system security and stability assessment, power system monitoring, and numerical simulation.



Qiuwei Wu (M'08-SM'15**)** obtained the B. Eng. and M. Eng. from Nanjing University of Science and Technology, Nanjing, P. R. China, in 2000 and 2003, respectively, both in Power System and Its Automation. He obtained the PhD degree from Nanyang Technological University, Singapore, in 2009, in Power System Engineering. He was a senior R&D engineer with VESTAS Technology R&D Singapore Pte Ltd from Mar. 2008 to Oct. 2009. He has been working at Department of Electrical Engineering, Technical University of Denmark (DTU) since Nov. 2009. He was a visiting scholar at Department of Industrial Engineering & Operations Research (IEOR), University of California, Berkeley from Feb. 2012 to May 2012. He is also a visiting professor at Shandong University, China. His research interests are integration of wind power and electric vehicle into power systems, active distribution networks, and real time power system simulation and analysis.

Vladimir Terzija (M'95-SM'00-F'15) is the EPSRC Chair Professor in Power System Engineering in the School of Electrical and Electronic Engineering, the University of Manchester, U.K., where he has been since 2006. From 1997 to 1999, he was an Assistant Professor at the University of Belgrade, Serbia. From 2000 to 2006, he was with ABB AG, Germany, working as an expert for switchgear and distribution automation. His main research interests are Smart Grid, application of intelligent methods to power system monitoring, control, and protection, switchgear and fast transient processes, as well as DSP applications in power systems.