ACADEMIA

Accelerating the world's research.

Estimation of Transmission Line Parameters from Historical Data

J. Gronquist

2013 46th Hawaii International Conference on System Sciences

Cite this paper

Downloaded from Academia.edu 🗹

Get the citation in MLA, APA, or Chicago styles

Related papers

Download a PDF Pack of the best related papers 🗹

55965	(entropence
	(TOP) perce
Sharefully in	
ABOSED SATURA	
VERONAL PLAN	
The second second	
A contract over 1995	

Temperature-Dependent Power Flow Salman Mohagheghi

Allen J. Wood, Bruce F. Wollenberg, Gerald B. Sheblé Power Generation, Operation and Control Wiley Inte... giorgi jijishvili

Reactive Power Injection Strategies For Wind Energy Regarding Its Statistical Nature Joaquin Mur

Estimation of Transmission Line Parameters from Historical Data

Katherine R. Davis PowerWorld Corporation <u>kate@powerworld.com</u> Sudipta Dutta University of Illinois at Urbana-Champaign <u>dutta4@illinois.edu</u>

Abstract

Power system analysis results are based on a model of the system. The model is often assumed known, yet it may contain errors. Model errors can impact the accuracy of the results. This paper utilizes historical data over time to extract the transmission line model parameters. The focus is on application of line parameter estimation techniques for real data from a North American utility. The approach in this paper focuses on the challenges of dealing with uncertainty and errors encountered in the real-world data. Future applications are also discussed.

1. Introduction

Model development and validation is an area of growing importance for power systems. The reason for placing high emphasis on model accuracy is that in any discipline, results are only as good as the models upon which the analysis is based. Thus, identifying model errors is an important problem.

This work incorporates large volumes of historical data collected over time to estimate the values of transmission line model parameters. Real data often contain significant errors which do not have known probability distributions. Meters used for measuring data are a common source error. A key aspect of this work involves being able to perform analysis such that any errors in the available data do not corrupt the estimate.

With respect to data collection, phasor measurement units (PMUs) [1] have received high visibility. Increasing PMU deployment leads to an increase in the amount of data available and a desire to effectively manage, send, and process the data. It remains a challenge to intelligently analyze the data and obtain useful information to advance the state-ofthe-art for power system operations and control. Data from PMUs is often the focus of data mining Thomas J. Overbye University of Illinois at Urbana-Champaign overbye@illinois.edu Jim Gronquist Bonneville Power Administration jfgronquist@bpa.gov

efforts, yet other existing sources tend to be overlooked. One resource which has not been utilized to its full potential is the enormous volume of historical SCADA data which is readily available to many utilities.

This work addresses the research question of whether the sheer volume of available SCADA data facilitates acceptable parameter estimates in the presence of data errors. The idea exploited is to use "data over time," and the approach in this paper is advantageous over methods which consider a single snapshot of the system. The parameters of a pimodel of a medium length transmission line are determined from multiple measurements of complex power and voltage magnitude at both ends of the lines.

The organization of the paper is as follows. A brief background of the estimation problem is provided in Section 2. The transmission line model used in the work and an overview of the parameter estimation method is presented in Section 3. The proposed methodology is described in Section 4. Results obtained by applying the proposed methodology on a simulated test case are presented in Section 5 and results obtained on real SCADA data from a North American utility are presented in Section 6. Section 7 presents discussions on the results. Section 8 discusses future applications for transformer model parameter estimation. Finally, conclusions are made in Section 9.

2. Estimation Background

Many applications rely on correct underlying model information. In power systems, the area where estimation tends to receive the most attention has been in power system state estimation [2], [3]. Parameter identification has often been studied in the context of state estimation literature as state estimators also serve as data filters. Uncertainty in data with respect to state estimation is analyzed in [4]. Transmission system parameter inaccuracies can have adverse effects on state estimation. Thus, existing work is aimed at identifying and correcting transmission model parameter errors using telemetry data [3],[5],[6],[7],[8]. While a few methods provide a means for updating the estimates online, most existing approaches consider only a single snapshot in time. Also, typically, results for small simulated systems are presented as opposed to real systems.

This paper circumvents both previous limitations. Historical SCADA data, recorded every five minutes over several months, is available and is presented in Section 6 of this paper. A close analogy to the key concept of this work is introduced in [9], where PMU data collected over a number of different operating points can be applied to construct a reduced equivalent model at the PMU buses. The statistical properties of the data are important to the quality of our estimate. Background on the statistical analysis of data [10] provides insight into the quality of the obtained estimate.

3. Transmission Line Model and Available Data

In this section, the transmission line model is described and analyzed in the context of the measured data which is used in this work. This provides a fundamental expectation of what can (and cannot) be estimated about the model based on the available data.

3.1. Transmission Line Model

The line model considered is the *pi model* for transmission lines, shown in Figure 1. This is a standard model which is often used to model lines of medium length [11], of approximately 50 to 150 miles.



Let V_1 , V_2 , I_1 , and I_2 denote the magnitudes of the phasor quantities. The assumption is that the measurements for this application are P_1 , Q_1 , V_1 , P_2 ,

 Q_2 , and V_2 . This is the 'raw' SCADA data, recorded every five minutes. Note that what is physically measured by potential transformers (PTs) and current transformers (CTs) is only voltage and current, respectively; however, these *true* raw measurements may be unavailable. The goal is to use available SCADA measurements to estimate values of r, x, and b in Figure 1. The assumption is that the angle difference across the line and the line parameters are unknown. If it is possible to know the angle difference across the line (as with PMUs), the problem is considerably simplified.

3.2. Basic SCADA Data Properties

The pi model equations are now considering the assumed *known* or measured quantities. Several quantities can always be derived directly from the SCADA data. Since P_1 , Q_1 , P_2 , and Q_2 are measurements, the complex powers S_1 , S_2 at the line ends are directly known. Thus, the apparent powers $|S_1|$, $|S_2|$ are also known. From the apparent power and voltage magnitudes, the current magnitudes, I_1 , I_2 , can be directly computed.

$$I_i = V_i / \left| S_i \right| \tag{1}$$

From the current magnitudes, power factor angles are also known at both line ends, $\theta_{1,pf}$, $\theta_{2,pf}$. Notably lacking in these quantities, which can be exactly computed from the SCADA data, are the line parameters and the angle difference across the line.

3.3. Pi Line Model

In the pi model, let b/2 denote the value of the shunt modeled at each bus. The admittance matrix of the transmission line model is Y=G+jB:

$$\mathbf{Y} = \begin{bmatrix} \frac{1}{r \Box jx} \Box jb/2 & \Box \frac{1}{r \Box jx} \\ \Box \frac{1}{r \Box jx} & \frac{1}{r \Box jx} \Box jb/2 \end{bmatrix}$$
(2)

The real and reactive components of **Y** gives matrices **G** and **B** with elements G_{ij} and B_{ij} . The equations for real and reactive power flow at both ends of the line are given by the following:

$$P_{1} = V_{1}^{2}G_{11} \Box V_{1}V_{2} \begin{bmatrix} G_{12} \cos \Box \theta_{12} \Box B_{12} \sin \Box \theta_{12} \Box \end{bmatrix}$$
(3)
$$P_{2} = V_{2}^{2}G_{22} \Box V_{2}V_{1} \begin{bmatrix} G_{21} \cos \Box \theta_{21} \Box B_{21} \sin \Box \theta_{21} \Box \end{bmatrix}$$
(4)

$$Q_{1} = V_{1}^{2} B_{11} \Box V_{1} V_{2} \Big[G_{12} \sin \Box \theta_{12} \Box B_{12} \cos \Box \theta_{12} \Box \Big]$$
(5)
$$Q_{2} = V_{2}^{2} B_{22} \Box V_{2} V_{1} \Big[G_{21} \sin \Box \theta_{21} \Box B_{21} \cos \Box \theta_{21} \Box \Big]$$
(6)

From (2)-(6), the equations for the real and reactive power losses in the line are thus given by P_1+P_2 and Q_1+Q_2 respectively:

$$P_{losses} = \boxed{V_1^2 \Box V_2^2} \boxed{G_{11} \Box 2V_1V_2} \begin{bmatrix} G_{12} \cos \Theta_{12} \end{bmatrix}$$
(7)

$$Q_{losses} = \left[V_1^2 \Box V_2^2 \right] B_{11} \Box 2V_1 V_2 \left[B_{12} \cos \Box \theta_{12} \Box \right]$$
(8)

Thus, four out of the six equations in (3)-(8) are independent. The loss equations are obviously not independent of the first four since they are each the sum of two other equations. In equations (3)-(6), the unknowns are G_{12} , B_{12} , b/2, and $\theta_{12} = (\theta_1 - \theta_2)$. Thus, there are four equations and four unknowns.

This looks similar to a traditional state estimation problem except with different unknown variables. The difficulty is that the unknowns appear in the equations in a nonlinear manner. Thus, it follows that use of a linear technique to estimate these quantities will not yield complete success. To see this, observe that what appears in the equations is always a product term of G_{ij} and B_{ij} multiplied with either the term $sin(\theta_{12})$ or $cos(\theta_{12})$. Solution requires a nonlinear iterative approach such as Newton's method to find b, x, and r to minimize the difference between the measurements and the calculated quantities. Essentially, this is state estimation. The key difficulty is that from one snapshot, there are more unknowns than equations so traditional state estimation is not sufficient. The goal of this work is to improve the ability to directly estimate b, x, and rusing data collected over multiple points in time.

The unknowns G_{ij} , B_{ij} , and b/2 are not dependent on time. This is true at least over some window of time, whereas the angle difference θ_{12} is a function of the operating point and changes over time. Thus, the terms in (3)-(6) which depend on time are the following:

$$G_{12} \cos \left[\theta_{12} \right], \quad B_{12} \sin \left[\theta_{12} \right] \\G_{12} \sin \left[\theta_{12} \right], \quad B_{12} \cos \left[\theta_{12} \right]$$
(9)

The proposed approach of considering multiple snapshots in time to estimate the transmission line parameters is only successful if there are fewer unknowns which depend on time than the number of linearly independent equations. In this case, there are four linearly independent equations. Each additional time point will produce four more equations but will also produce four more unknowns. Thus, in the SCADA estimation problem, some approximation is necessary. The choice of the approximation is therefore important.

3.4. Estimation Equations

The equations used to perform the SCADA estimation are presented in this section. The following two equations for real power losses and reactive power losses contain the three unknowns (r, x, and b) which are time independent.

$$P_{losses} = I^2 r \tag{10}$$

$$Q_{losses} = \Box V_1^2 \Box b / 2 \Box D_2^2 \Box b / 2 \Box I^2 x \qquad (11)$$

The need for approximation arises since the current magnitude I in (10) and (11) is not known exactly from the measurable quantities. Due to current injections from the shunt elements, I differs from I_1 and I_2 which are known from (1). As an approximation, the average of I_1 and I_2 is taken for I. This avoids the problem of unknowns which depend on time.

Other variations of the estimation equations are possible depending upon which approximations are desirable to make. Future work may investigate the impact of the model approximation on the parameter estimation. For example, a bias term u can be added to the estimation equation for r,

$$P_{losses} = u \,\Box \, I^2 r \tag{12}$$

$$Q_{losses} = \Box \left(\frac{V_1^2 \Box V_2^2}{2} \right) b \Box I^2 x \tag{13}$$

which represents a P_{loss} static error in the measurements. Based on (12) and (13), a system of equations is obtained by appending the contributions of repeated measurements to the matrix **A**. This may be expressed as (14) or (15):

$$\begin{bmatrix} 1 & I^{2} & 0 & 0 \\ 0 & 0 & \Box I^{2} & \boxed{V_{1}^{2} \Box V_{2}^{2}} \boxed{V} 2 \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{Ae}_{y} = \mathbf{f}_{p,q} \end{bmatrix} \begin{bmatrix} u \\ r \\ x \\ b \end{bmatrix} = \begin{bmatrix} P_{losses} \\ Q_{gen} \\ \vdots \end{bmatrix} (14)$$

As additional rows are added corresponding to additional time points, this becomes an overdetermined problem. Least squares estimation minimizes the residual,

$$\mathbf{r} = \mathbf{A}\mathbf{e}_{\mathbf{y}} \Box \mathbf{f}_{\mathbf{p},\mathbf{q}} \tag{16}$$

and results in the following estimate (14),

$$\mathbf{e}_{\mathbf{y}} = [\mathbf{A}^{\mathrm{T}}\mathbf{A}]^{\Box \mathbf{I}}\mathbf{A}^{\mathrm{T}}\mathbf{f}_{\mathbf{p},\mathbf{q}}$$
(17)

where $\mathbf{e}_{\mathbf{y}} = [u, r, x, b]$ denotes a vector of the estimated quantities. The vector $\mathbf{f}_{\mathbf{p},\mathbf{q}}$ contains real power losses P_{losses} and the negative of reactive power losses Q_{gen} . It is possible to estimate P_{losses} and Q_{gen} independently since the equations are completely decoupled, as evident in (14).

4. Proposed Estimation Approach

An overview of the proposed methodology is now presented. Initially, a data selection module returns the dataset and also accomplishes data 'screening' functions such as value filtering and initial bad data rejection. Then, for a given set of data, regression is applied to detect and reject possibly wrong data based on analysis of a fitted model.

4.1. Dataset Selection

The procedure outlined in Algorithm 1 is responsible for selecting and returning a set of data on which the analysis is to be performed. For a specified line, the sample dataset consists of measurements of P, Q, and V at both line ends. This function also provides initial screening and filtering of the raw SCADA data. Currently, all data is screened simply by removing data points which have all values of zero. Additional screening may also be performed. A *data point* or a *row* refers to one set of SCADA measurements from one point in time. Value filters, when applied, allow Algorithm 1 to return a dataset for particular conditions, such as a < x < b, where x is any of the measured or computed variables.

Function DataSelection

```
Input: RawData, whichLine, filter, chunkSize,

Output: DataSet consisting of P, Q, V at line ends

keepIndex = (rand) to (rand+chunkSize)

DataSet = RawData(keepIndex)

Foreach row of DataSet do begin

If row = 0 then begin

Remove row

End

End

clear keepIndex

Foreach row of DataSet do begin

If row.MeetsFilter then begin

keepIndex.add(row.Index);

End

End
```

DataSet = DataSet (keepIndex) Algorithm 1. Procedure for Dataset Selection

In summary, Algorithm 1 returns a random consecutive data chunk of size *chunkSize* which has undergone initial screening and value filtering.

4.2. Estimation and Outlier Rejection

Once an initial data set is selected, regression is applied, as outlined in Algorithm 2. An initial fit to the dataset is obtained from least squares. However, in this initial fit, all of the data are used Thus, outliers can have a indiscriminately. significant impact on the estimate. By removing outliers or problematic data points, the estimate improves. At each iteration, the residuals or deviations are computed using (16) with the current values of the estimated parameters. High residuals indicate a poor match of the particular data point to the model. The strategy is to identify data points with residuals which exceed a predetermined threshold and reject them. The process continues and iteratively fits the data to the refined data set until all data points have residuals within the specified threshold. The procedure stops when the number of rows removed from the dataset during any iteration, n_R , is equal to zero.

Function PerformRegression

Input: DataSet Output: $e_y = [u, r, x, b]$, DataSet $e_y = LeastSquaresEstimate(DataSet)$ $r = Ae_y-b$ While $n_R > 0$ then begin $n_R = 0$ [DataSet, n_R] = OutlierFilter (DataSet, r, n_R) $e_y = LeastSquaresEstimate(DataSet)$ End

Algorithm 2. Regression

The regression algorithm thus progressively identifies and removes bad data and computes a new estimate from the refined data set. However, any errors which exist in the initial dataset can still have an impact on the quality of the initial fit.

Outlier filtering is performed as indicated in Algorithm 3, which is applied as a component of Algorithm 2. While Algorithm 3 is the currently implemented, other outlier filtering algorithms can easily be substituted in its place.

Function OutlierFilter Input: DataSet, r, n_R Output: DataSet, n_R Foreach row of DataSet do begin [F, x] = ksdensity(r) $[\mu, \sigma, \mu_{ci}, \sigma_{ci}]$ = normfit(r)

```
threshold = f(r, \sigma)
      If residual(row) > threshold then begin
         Remove row from DataSet
         Increment(n_{\mathbb{R}})
      End
   End
End
```

Algorithm 3. Outlier Filtering

In Algorithm 3, Matlab functions ksdensity and normfit have been used from the Statistics toolbox. The function ksdensity (r) is used to compute a probability distribution estimate for the sample vector r. This function returns a vector of density values F evaluated at the points in vector x_i . The function normfit(r) fits r to a normal distribution and returns the estimates of mean μ and standard deviation σ of the data with a 95% confidence interval. Estimates on the bounds of the confidence intervals are given by μ_{ci} and σ_{ci} . The objective of Algorithm 3 is to detect and remove 'bad' data characterized by deviation of residual from the zero mean. If the error residual for a data point exceeds the threshold, the point is removed.

5. **Application to Simulated Data**

In this section, the procedures are applied to simulated data from a test system to serve as a reference. To perform these simulations, PowerWorld Simulator is used via SimAuto and script commands. A program allows SimAuto to retrieve SCADA data from the simulated system. Then, the system is perturbed to a new operating point. By repeating this procedure, data for a number of operating points is obtained. Simulations provide a reference case since both the model and the data are exactly known. Application to real data is presented in the next section. The system used for the simulations is the seven bus system shown in Figure 2.



Figure 2. Seven bus study system

The estimated line is between Bus 3 and Bus 4, with model parameters: r=0.01 pu, x=0.03 pu, b=0.02 pu.

5.1. Simulated Error-Free Data

A simulated SCADA dataset for line (3,4) is obtained from five different operating points. The units of P, Q, and V measurements are MW, MVAr and per unit, respectively. Using only these points with no added error, the estimate matches the model, as expected. The values of the estimates are given in Table 2. Real and reactive power plots of the actual data, the estimate, and the model are shown in Figure 3 and Figure 4. The per unit base is 100 MVA.

Table 1. Study system noise-free estimates

Line	u (pu)	r (pu)	x (pu)	b (pu)
(3, 4)	-1.99e-5	0.010055	0.030161	0.020058



Figure 3. Seven-bus error-free data Ploss plot



Figure 4. Seven-bus error-free data Q_{gen} plot

5.2. Simulated Data with Error

From the data for the five 'error-free' operating conditions above, 200 perturbations are made of each point. This reflects the fact that during each quasisteady-state operating point, the SCADA values do not change much but may change slightly. Thus,

from the five exact operating points, we create 200*5 = 1000 hypothetical measurement data points. Each measurement is per-unitized and randomized by a normal probability distribution of mean zero and standard deviation 2 MW, 2 MVAr, and 0.02 per unit voltage. These 1000 points are used to compute the estimate. The results from the new dataset are shown in Figure 5 and Figure 6, and the corresponding estimates are shown in Table 2.



Figure 5. Seven-bus noisy dataset Ploss plot



Figure 6. Seven-bus noisy dataset Q_{gen} plot

As indicated by the estimates for $(3, 4)^*$ in Table 2, the accuracy of the *r* estimate is considerably improved when the static error *u* is not estimated.

Cable 2.	Study	system	estimates	with nois	e
		~ , ~			-

Line	u (pu)	r (pu)	x (pu)	b (pu)
(3, 4)	0.0012165	0.006369	0.02932	0.017988
(3, 4)*	-	0.010153	0.02931	0.020321

It is interesting to note that the five operating points result in five distinct clusters of points on the graphs. In real data, the authors have observed the presence of these bands, especially when a relatively small set of points is examined. Thus, multiple noisy measurements of the same operating point seem to cause this feature. As may be evident when comparing the results, some sets of operating points facilitate better estimates than others. The issue is that when simulating data, it is challenging to reproduce data with a realistic distribution. The question arises of how one should perturb the system to obtain representative operating points since there are infinite possible ways to change the system and obtain a new operating point. Thus, rather than speculate on what representative operating points exist that should be studied, the most value is obtained from testing the procedure on real data.

6. Application to Real SCADA Data

The proposed procedures are applied to historical SCADA data from a real North American power system. Identifiers of buses and lines in the real system have been made anonymous. From a practical point of view, it is extremely valuable to be able to present these results for actual SCADA data. Estimation on artificially constructed data is a special/trivial case. However, parameter estimation using real world data can be extremely challenging as evident in the results and discussions presented in the following sections. Real world data rarely follows a Gaussian or known probability distribution.

6.1. Study System - North American Utility

From a North American utility, SCADA data collected over several months is available for the three transmission lines illustrated in Figure 7. The model values of the line parameters for these lines are also available, tabulated in Table 3. Results for these three lines are presented.



Figure 7. Three Lines with Available Data

Each line identifier $\alpha = (i,j)$ in Table 3 denotes a line α between bus *i* and bus *j*, from Figure 7.

Table 3. Existing Model Values for Lines

Tuble of Emisting forder values for Emes					
Line	r (pu)	x (pu)	b (pu)		
A=(1, 2)	0.00056	0.01054	0.9079		
B=(2, 3)	0.00052	0.00999	0.75194		
C=(3, 4)	0.00025	0.00445	0.32042		

6.2. Comparison of Model to Data

An initial dataset with a size of 30,000 data points is selected according to Algorithm 2. No value filtering is applied. In the next sub-sections, Figures 8, 10, 12, 14, 16, and 18 plot P_{loss} and Q_{gen} vs. *I* for each line. The red and blue lines on these plots show respectively the expected data points when using the estimated and model parameter values. Figures 9, 10, 13, 15, 17, and 19 show distributions of the residuals about the parameter estimate as well as Gaussian distributions of the same mean and standard deviation (red dotted lines). The available datasets for lines *A*, *B*, and *C* lead to reasonable and consistent parameter estimates.

Resistance r and static power offset u are estimated from the real power losses, while x and b are estimated from the reactive power losses. These estimates for all of the lines are given in Table 4. Each estimate is based on 30,000 data points.

6.2.1. Line *A* plots.



Figure 9. Q_{gen} residual distribution (Line A)



Figure 11. P_{loss} residual distribution (Line A)

6.2.2. Line *B* plots.





Figure 13. Q_{gen} residual distribution (Line *B*)





Figure 15. P_{loss} residual distribution (Line *B*)

6.2.3. Line C plots.





Figure 19. P_{loss} residual distribution (Line C)

6.2.4. Parameter Estimates for All Lines

The estimates of the line parameters based on the data above are given in Table 4 below.

 Table 4. Estimates from 50,000 Data 1 onits				
Line	u (pu)	r (pu)	x (pu)	b (pu)
A=(1, 2)	0.0067853	0.0008664	0.010794	0.90628
B = (2, 3)	-0.008649	0.0004789	0.01163	0.7144
C=(3, 4)	-0.010291	-0.000257	0.004415	0.35742

Table 4. Estimates from 30.000 Data Points

7. Discussion of Results

As can be seen from the above plots, the estimated parameters are close to the available model parameters. The estimated resistance for Line A is slightly higher than its model parameter; this may be due to the fact that the available model resistance value was computed using the dimensions of the line, material resistance, etc. for different conditions than actually seen in operation. The reason this does not occur for the other two lines may be because Line A is much longer in comparison to the Lines B and C.

The estimated resistance of Line C is negative when the algorithms are applied on the available raw SCADA data. However, the magnitude of the estimated resistance is close to the model resistance available. This may indicate the presence of measurement sign errors. The suspected sign errors are corrected, and a corrected dataset is returned after applying this filtering in

Algorithm 1. The new estimate of the resistance, r = 0.000305 pu, is considerably improved. Figure 20 shows the new P_{loss} vs. I plot.

Next, estimation is done after applying regression for all the lines. Since the actual parameters of the transmission lines are not known or rather cannot be exactly known, there is no way to validate the final results. However, the values provided in Table 5 are expected to be the best parameter estimates. It is clear that 'screening' the data is a critical stage of the estimation process, as the dataset which is chosen ultimately impacts the quality of the final estimate.

The data for reactance and shunt estimation for the three lines was 'better' than the data for resistance estimation, as evident from the plots for distribution of residuals. The latter data is noisy and the P_{loss} values are small compared to the value of the parameter being estimated. Measurements may also contain systemic biases which are not accounted for in the approach. A more thorough coverage of the detection of data problems is beyond the scope of this paper.



Figure 20. Ploss vs. I (Line C, corrected data)

Table 5. Estimates after Regression

Tuble et Estimates alter Regression				
Line	u (pu)	r (pu)	x (pu)	b (pu)
A = (1, 2)	0.0058326	0.0009012	0.010774	0.90586
B=(2, 3)	-0.010256	0.0004928	0.011749	0.71445
C=(3, 4)	0.0077929	0.00030975	0.004401	0.3589

An effect which has not been included is how actual line resistance varies with temperature. There is approximately a 0.4% increase in resistance per °C rise in temperature [12]. In seasonal planning models, the *r* values are not adjusted since *x* values typically dominate. When we use multiple months of SCADA data to obtain an estimate, any changes in the actual parameters are ignored.

8. GIC Model Validation Applications

A future use of historical data to estimate model parameters is in geomagnetic disturbance (GMD) modeling. GMDs induce geomagnetically induced currents (GICs) in transmission lines and transformers. Background on the problem and its effects in large power systems is presented in [13].

GIC-related transformer reactive power losses are observed to vary linearly with terminal voltage [14], and may be represented, for example, as

$$Q_{loss} = I_{line}^{2} \cdot x_m \,\Box V_{pu} K I_{GIC} \tag{18}$$

where K is the constant to be estimated and is specific to the transformer, V_{pu} is the terminal voltage in per unit, and I_{GIC} is the DC GIC. In the simple case of a wye-delta step-up transformer, I_{GIC} is the neutral current I_N . Units of K are MVAr/A. For an autotransformer, I_{GIC} is given by

$$I_{GIC} = \left| \frac{a_t I_H \Box I_L}{a_t} \right| \tag{19}$$

from [15] [16] where the two DC coil currents I_H and I_L may be measured, and a_t is the transformer turns ratio.

When the K values of transformers are not known, they are often approximated in the analysis by default values. Obviously, it is preferable to use actual values to improve the meaningfulness of the GIC results. The NERC report [13] identifies as an action item the development of tools for GIC flow and subsequent reactive power loss modeling. Parameter estimation can help validate transformer parameters which are needed for GIC analysis.

9. Conclusion

The value of using data collected over time is apparent when estimating transmission line parameters, as shown in this paper. Results for both simulated and real data are discussed. It is clear that estimation based on real measurement data presents challenges which are not encountered in simulation.

The presentation of the proposed solution approach to real data is thus valuable. The approach is validated on actual SCADA data from a real North American utility. In summary, the proposed methodology is a valuable tool which can allow planners to find and correct model errors based on data which is already collected.

10. Acknowledgments

The authors would like to acknowledge support from the Stanford University Global Climate and Energy Project (GCEP), the Power Systems Engineering Research Center (PSERC), Trustworthy Cyber Infrastructure for the Power Grid (TCIPG), and Consortium for Electric Reliability Technology Solutions (CERTS).

11. References

- A. G. Phadke, "Synchronized phasor measurements in power systems," *IEEE Computer Applications in Power*, vol. 6, no. 2, pp. 10-15, Apr 1993.
- [2] A. Monticelli, *State estimation in electric power* systems: a generalized approach. Kluwer Academic Publishers, 1999.
- [3] A. S. Debs, "Estimation of steady-state power system model parameters," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-93, no. 5, pp. 1260-1268, Sept. 1974.

- [4] A. K. Al-Othman, M. R. Irving, "Uncertainty modelling in power system state estimation," *IEE Proceedings-Generation, Transmission and Distribution*, vol. 152, no. 2, pp. 233- 239, March 2005.
- [5] G. L. Kusic, D. L. Garrison, "Measurement of transmission line parameters from SCADA data," *IEEE PES Power Systems Conference and Exposition*, 2004, pp. 440- 445 vol. 1, Oct. 2004.
- [6] J. Zhu, A. Abur, "Identification of network parameter errors," *IEEE Transactions on Power Systems*, vol. 21, no. 2, pp. 586- 592, May 2006.
- [7] J. Zhu, A. Abur, "Improvements in network parameter error identification via synchronized phasors," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 44-50, Feb. 2010.
- [8] J. F. Dopazo, G. Irisarri, A. M. Sasson, "Real-time external equivalent for on-line contingency analysis," *IEEE Transactions on Power Systems*, vol. PAS-98, no. 6, pp. 2153-2171, Nov. 1979.
- [9] K. M. Rogers, R. D. Spadoni, T. J. Overbye, "Identification of power system topology from synchrophasor data," *IEEE Power Systems Conference and Exposition (PSCE)*, pp. 1-8, March 2011.
- [10] J. Mandel, *The Statistical Analysis of Experimental Data*, (Reprint) Dover, 1984, Mineola, N.Y.
- [11] R. Bergen, V. Vittal, *Power System Analysis*, 2nd ed., Prentice Hall, 2000.
- [12] S. Smolleck, J. Sims, "Guidelines for the selection and operation of bare ACSR conductors with regard to current carrying capacity," *Electric Power Systems Research*, vol. 5, 1982, pp. 179-190.
- [13] "2012 Special Reliability Assessment Interim Report: Effects of Geomagnetic Disturbances on the Bulk Power System," NERC, Feb. 2012.
- [14] X. Dong, Y. Liu, J. G. Kappenman, "Comparative analysis of exciting current harmonics and reactive power consumption from GIC saturated transformers," *IEEE Power Engineering Society Winter Meeting*, pp. 318-322, Feb 2001.
- [15] V.D. Albertson, J. G. Kappenman, N. Mohan, G. A. Skarbakka, "Load-flow studies in the presence of geomagnetically-induced currents," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-100, no.2, pp.594-607, Feb. 1981.
- [16] K. Zheng, et. al., "Influence of system characteristics on the amplitudes of geomagnetically induced currents," submitted to *IEEE Transactions on Power Delivery*, March 2012.