

Identification of Topology Changes in Power Grids using Phasor Measurements

Scott Vander Wiel Russell Bent Emily Casleton
Earl Lawrence

September 24, 2014

Abstract

Phasor measurement units (PMUs) are increasingly important for monitoring the state of an electrical power grid and quickly detecting topology changes caused by events such as lines going down or large loads being dropped. Phasors are complex valued measurements of voltage and current at various points of generation and consumption. If a line goes down or a load is removed, power flows change throughout the grid according to known physical laws and the probability distribution of phasor measurements changes accordingly. This paper develops a method to estimate the current topology of a power grid from phasor measurements and considers the design goal of placing PMUs at strategic points in a distribution system to achieve good sensitivity to single-line outages. From a vector of phasor measurements probabilities are computed corresponding to the scenario that all power lines are operational and to alternate scenarios in which each line goes down individually. These probabilities are functions of the joint distributions of phasor measurements under each possible scenario, obtained through Monte Carlo simulations with random load profiles. We use log-spline densities to estimate marginal distributions of phasor measurements and fold these into a multivariate Gaussian copula to capture important correlations. Sensitivity to outages varies according to which line goes down and where PMUs are placed on the grid. A greedy search algorithm is demonstrated for placing PMUs at locations that provide good sensitivity to single-line outages.

1 Introduction

A power grid operator must know the grid topology and demand profile in order to generate and dispatch power in the right locations and to run the grid stably and securely. Transmission grid operators have a long history of using state estimation for supervisory control and planning but in the *distribution* grid real-time measurements are sparsely available and much must be inferred. State estimation is a major research topic in the power engineering community [20, 28, 22, 21, 4, 27, 1, 23]. The problem of estimating the state of the grid is often divided into two interrelated phases. One phase is state estimation proper in which analog quantities are estimated—voltages at all bus locations and power flowing over all lines. The other phase is topology processing and topology error detection in which breaker status information is used to track the current topology of the grid and errors in the calculated topology are detected and corrected. These two stages iterate and the combined process is known as generalized state estimation. Future smart grids will be more highly instrumented with measurement devices, increasing the importance of bringing state and topology estimation techniques into the distribution portion of the electric grid. Reference [9] provides a good overview of state estimation challenges that arise in the development of smart grids.

1.1 Contributions

Phasor measurement units (PMUs) are sensors placed at select nodes in a power grid to collect synchronized measurements of electrical waveforms in real time. Synchronized real-time measurements are a contrast to traditional slowly sampled inject measurements with time accuracy that is too coarse to resolve phase differences between measured nodes. PMUs are revolutionizing methods for monitoring and control of electric power grids.

This paper makes two contributions in the area of topology error detection (e.g., detection of downed lines) in distribution grids. First, we simplify and extend the model bank classification methods of [16] to work with data from PMUs. Second, we use the new technique as part of a simple procedure for determining good PMU locations.

Many topology error detectors use relatively simple statistical methods that start with an assumed topology and modify it based on an examination of the state estimate residuals (e.g. [4, 28]). Recent work

detects topology errors using a predefined bank of possible topologies [23, 16]. Bayesian methods estimate the probability distribution over the model bank given the observed data. This paper also uses a model bank and extends the idea of [16] in which the search for topology errors proceeds with a sparse set of measurements insufficient to determine the full analog state of the grid. We also demonstrate the method as part of a technique for determining good PMU locations.

Power flow equations (discussed later) are important for estimating the analog portion of the power system state. Solving these nonlinear equations can be computationally demanding, especially for large networks, and real-time solutions are generally not available. Furthermore, measurements are too sparse in the distribution grid for full state estimation that is routine in the transmission grid. The method presented here meets these challenges by running the solver in a pre-computation step to build a statistical approximation that is used in real-time to detect unknown downed lines or other topology errors with many fewer measurements than needed for full analog state estimation.

The remainder of Section 1 gives an overview of work on the problems of topology error detection and optimal placement of PMUs. Section 2 discusses some of the background on solving power flow equations for electrical parameters, with loads generated randomly. Section 3 describes Monte Carlo-based detection of topology faults (i.e., downed lines) using log-spline copula models to represent the joint distribution of phasor data under a known grid topology. Section 4 applies downed line detection to a 37 bus distribution grid. Section 5 provides a greedy method to select a set of buses to instrument with PMUs for good sensitivity to single line outages. Section 6 applies PMU placement to the 37 bus network. Section 7 discusses ideas for follow-on work to expand the usefulness of these methods.

1.2 Overview of Topology Error Detection

Topology errors, in the power systems literature, are unexpected changes in network interconnections that are not (or not *yet*) captured by the topology processing system that receives breaker status data to track which buses (nodes) are electrically connected to one another. Downed lines, for example, are usually detected using breaker status data at the control center. However, sometimes breaker status is in error and a topology change is only detected in the analog measurements that

are available on the grid. Also, more highly instrumented smart distribution grids will need to identify abrupt changes more quickly than current topology processing systems are capable of doing.

A traditional method [5] for identifying topology errors in a power system compares residuals from estimation of the analog state (e.g. voltages and flows) to the residuals expected from each of many possible faults in the network. Another popular method [6] detects topology errors using normalized Lagrange multipliers from a constrained non-linear least squares formulation of the state estimation problem. The Lagrange approach has been extended [18] to a more efficient method that uses Bayesian hypothesis tests and avoids having to perform state estimation for each of a large number of alternative topologies,

Recent advances in topology error detection are based on the availability of PMUs that provide voltage measurements with GPS time stamps accurate enough to compare phases across multiple bus locations. Tracking algorithms, such as the unscented Kalman filter, are being applied to PMU data streams [26] to dynamically estimate the power system state and quickly detect anomalies such as sudden changes caused by downed lines. A two level state estimator has been proposed [29] in which PMU data are processed at substations to estimate the local state of the grid and remove local topology errors. The grid control center combines substation estimates into a higher level estimate and also detects remaining topology errors.

1.3 Strategies for Placing PMUs

PMUs are expensive and are being installed incrementally. Recent work on PMU placement [17] assumes known topology and uses a greedy algorithm to achieve near-optimal placement of PMUs as measured by a mutual information criterion. This technique sequentially places PMUs in locations that reduce uncertainty in estimates of the analog system state. Section 5 solves a similar problem but with an objective of better detecting topology changes such as downed lines. Integer programming methods have been employed [3] to place an optimal combination of traditional power injection sensors and PMUs at locations that are strategic for detecting topology errors at low cost.

For reliable operation of a modern transmission grid it is becoming more important to detect and localize a fault within a fraction of a second. New techniques [15] combine synchronized voltage measurements at widely dispersed locations and use a traveling wave the-

oretical framework to triangulate the location of a fault. Followup work [14] places a minimum number of sensors at strategic locations so that all possible faults become observable using the traveling-wave technique. Many algorithms exist to localize faults in future smart grids that will have vastly expanded measurement capability. In fact, one new method [10] uses the current topology of the grid and availability of measurements to choose the best algorithm among several choices for localizing a fault.

2 The Power Flow Equations and Phasor Measurements

Figure 1 shows a *oneline* diagram of a power distribution grid. Labeled segments are buses that attach to transmission lines, loads, and generators. Generators are shown as double arrows on eight of the buses (e.g., two at bus 28 in the lower right of the diagram). Transformers separate regions that operate at different voltages and are indicated by pairs of squiggly lines. Loads are shown with arrows indicating real and reactive power consumption, capacitor symbols represent shunt capacitor banks that supply reactive power at large load buses to help stabilize voltage (e.g., at node 44 in the lower left). Each of these elements has associated electrical parameters such as the power output of a generator and the resistance of a transmission line. A given electrical parameter can affect power flow through every line on the network.

Flow of power on an alternating current (AC) electrical grid is described by complex-valued voltage and power. Figure 2 illustrates the most important parameters related to power flow. Power injected (i.e., net consumption or generation) at bus i is denoted by P_i (real) and Q_i (imaginary, *reactive*). Similarly, power flowing from bus i to bus j is denoted by P_{ij} (real) and Q_{ij} (reactive). Complex-valued voltage at bus i is given in polar form by a voltage magnitude $|V_i|$ and angle θ_i . Flow across the grid is affected by two admittance parameters of power lines, conductance, denoted $G_{i,j}$ and susceptance, denoted $B_{i,j}$ (which are derived from the resistance and reactance properties of the line); both are assumed known. Obviously, power grids have a wide variety of elements and many other electrical characteristics, but this simple description is sufficient to explore detection of line outages.

Power engineers make heavy use of the AC power flow equations

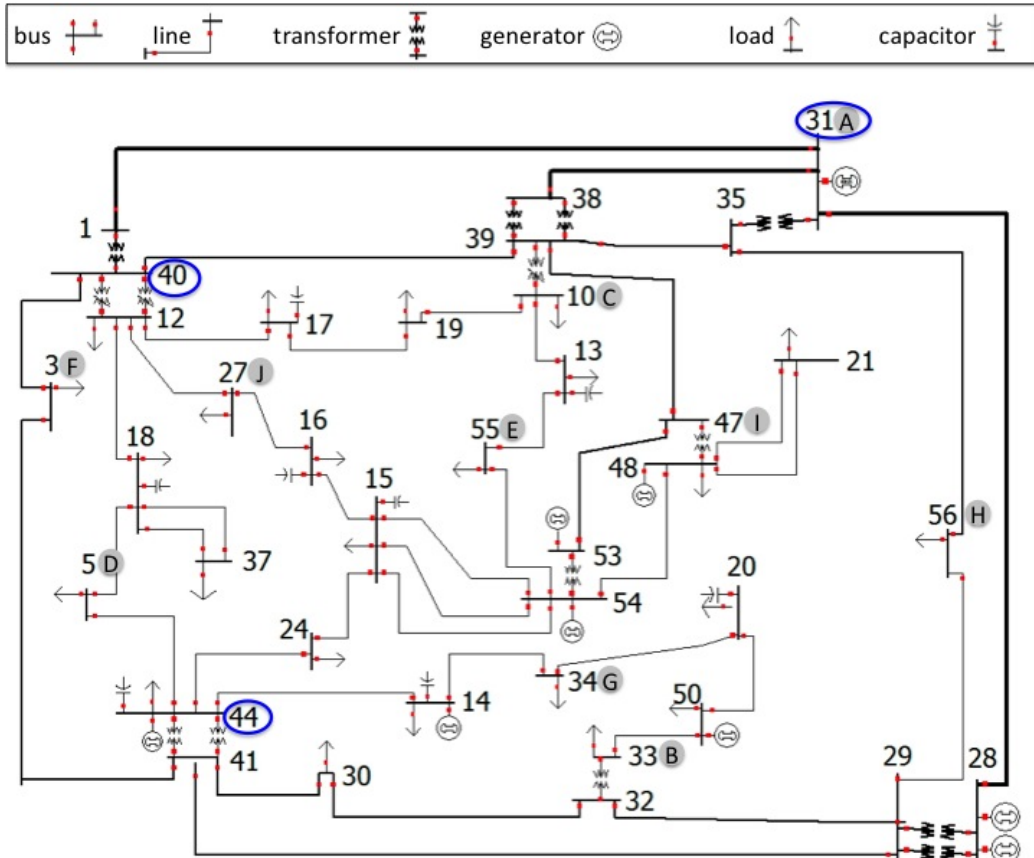


Figure 1: Oneline diagram of a 37-bus distribution system studied in [24] and [25]. Buses 31, 40 and 44 (encircled) were selected for PMUs in [24]. PMUs placed at the ten buses labeled A through J provide good sensitivity for detecting a downed line throughout most of the network (Section 6).

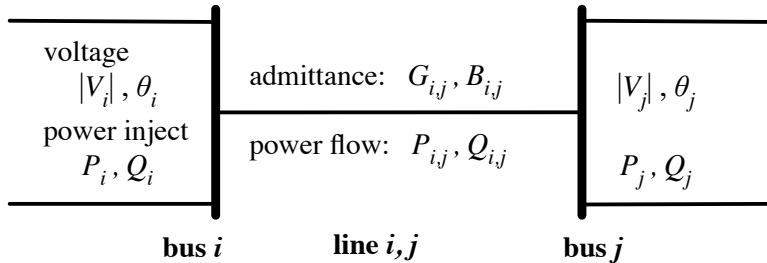


Figure 2: Basic parameters that determine flow of power over the lines of an electric grid.

that relate the real and reactive injections at each bus ($i = 1, \dots, B$) to voltage magnitudes and phases at all buses in the grid.

$$\begin{aligned}
 P_i &= \sum_{j=1}^B |V_i| |V_j| [G_{i,j} \cos(\theta_i - \theta_j) + B_{i,j} \sin(\theta_i - \theta_j)] \\
 Q_i &= \sum_{j=1}^B |V_i| |V_j| [G_{i,j} \sin(\theta_i - \theta_j) - B_{i,j} \cos(\theta_i - \theta_j)].
 \end{aligned} \tag{1}$$

Individual terms in these sums represent power flowing into bus i on line i,j . Thus, the real injection P_i is the sum real flows $P_{i,j}$ into bus i and the reactive injection Q_i is the sum of reactive flows $Q_{i,j}$ into bus i . Individual nodes can either be generators (net power going out), loads (net power coming in), or neutral¹.

There are some additional constraints on these equations. First, in a lossless model, the sum of the power injects over all of the buses is zero, $\sum_{i=1}^N P_i = 0$ and $\sum_{i=1}^N Q_i = 0$. In other words, all of the power generated in the system is used by the system. More generally, line losses constrain these sums to non-zero values. Second, the voltage angles only matter up to a rotation because they enter the equations only through their differences. Thus, solutions are invariant to adding a constant to all angles.

Each bus provides two of the four power and voltage parameters $P_i, Q_i, |V_i|, \theta_i$ and the remaining two parameters are found by solving the nonlinear power flow equations (1). Unfortunately, the nonlinear equations can be difficult to solve. We assume the grid operator desig-

¹Physically, a node can have both load and generation, e.g. Figure 1. However these equations aggregate the load and generation at a node.

Bus Includes	Input	Computed	PMU Measurements
Slack	$\theta_i = 0$	P_i	$A_i \theta_i$
Generator	$ V_i $	Q_i	
Regular	P_i	θ_i	$A_i \theta_i$
Generator	$ V_i $	Q_i	
No	P_i	θ_i	$A_i \theta_i$
Generator	Q_i	$ V_i $	$M_i V_i$

Table 1: Power flow equations (1) are set up with different input variables, depending on the elements at each bus. The slack generator runs at a known voltage magnitude and its voltage angle is take to be the reference, $\theta_i = 0$. A regular generator also runs with a known voltage and produces a known amount of real power. For buses without a generator, the real and reactive power are provided as inputs. In each case, the other two quantities are computed. All PMUs provide angle measurements, whereas only PMUs at non-generator buses provide useful measurements of voltage magnitude.

nates a single generator to pick up the slack between nominal levels of consumption and generation and the actual levels at any given time. The generator at the slack bus operates at a specified voltage magnitude and its voltage angle is take to be the reference phase for the whole grid.

Table 1 lists which parameters are given as inputs and which are computed for three different types of bus. The bus with the slack generator specifies the generator’s voltage magnitude $|V_i|$ and sets the reference voltage angle $\theta_i = 0$. Any other bus with a generator specifies the generator’s voltage magnitude $|V_i|$ and real power output P_i as inputs ². A bus with no generator may have a load or capacitor bank and these will determine non-zero values of P_i and Q_i . If a bus has no elements, its input quantities are $P_i = Q_i = 0$. In each case, the unspecified quantities are computed by solving the AC power flow equations (1).

Power injections (P_i, Q_i) are generally sums of

$$P_i = P_i^{\text{load}} + P_i^{\text{gen}} \quad \text{and} \quad Q_i = Q_i^{\text{load}} + Q_i^{\text{gen}} + Q_i^{\text{cap}} \quad (2)$$

where “load” indicates power consumed, “gen” is power generated and “cap” is reactive power injected by a capacitor bank. Bus 44 in

²If a generator is not voltage controlled, it is specified by Q_i and P_i

Figure 1 has all three components of injection whereas bus 41 has none of them. The load terms are generally not known with certainty, so we take them to be *a priori* independent Gaussian quantities with 20% standard deviations (as in [23]):

$$\begin{aligned} P_i^{\text{load}} &\sim N[\mu_i, (0.2\mu_i)^2] \\ Q_i^{\text{load}} &\sim N[\nu_i, (0.2\nu_i)^2] \end{aligned} \quad (3)$$

where μ_i and ν_i are specified nominal loads. By contrast, P_i^{gen} and Q_i^{cap} are specified deterministic quantities. Finally, each generator's reactive power output Q_i^{gen} is determined by subtraction from the computed injection Q_i :

$$Q_i^{\text{gen}} = Q_i - Q_i^{\text{load}} - Q_i^{\text{cap}}.$$

Phasor Measurement Units (PMUs) are devices with GPS clocks that can measure complex-valued voltages and currents with precise timestamps. The third column of Table 1 indicates voltage measurements that a PMU would provide if installed. PMUs can measure current and other quantities as well; we discuss voltage measurements to illustrate their utility for topology estimation. A PMU provides voltage angle measurements, A_i , on any bus and voltage magnitude measurements, M_i , on a non-generator bus. For given parameters θ_i and V_i the two types of voltage measurements are unbiased and independent:

$$\begin{aligned} A_i | \theta_i &\sim N[\theta_i, 0.01^2] \\ M_i | |V_i| &\sim N[|V_i|, (0.001V_{0,i})^2] \end{aligned} \quad (4)$$

where $V_{0,i}$ is the nominal voltage at bus i . Thus, voltage angles are measured with standard deviation equal to 0.01° and voltage magnitudes are measured with standard deviation of 0.1% of nominal. These are small but realistic measurements errors; see [7].

Although PMUs could, in theory, be placed at every bus on the grid, this is not currently practical in distribution grids because of PMUs are costly and communication lines are inadequate at some bus locations. Full deployment of PMUs is a goal that lies many years in the future. Measurements from PMUs placed on several buses in a grid can be collected at a central Phasor Data Concentrator where they are synchronized in time so that relative phases between buses will provide valuable information to infer power flow over the grid.

3 Topology Identification with Log-Spline Copula Models

Denote by \mathbf{y} the vector of PMU measurements from a given set of buses on a distribution grid. Our goal is to use \mathbf{y} to determine the probability of each topological scenario in a predefined list. Given \mathbf{y} , the scenario probabilities are determined by combining their prior probabilities with the probability density values for \mathbf{y} under each scenario and this motivates an effort to approximate the distributions of \mathbf{y} for each scenario of interest. Specifically, we illustrate the method using a bank of topologies corresponding to all single-line down scenarios and the nominal case of no lines down. Other types of network faults can be handled identically.

The distribution of \mathbf{y} is not available analytically, but is driven by propagating the prior distributions for the loads in (3) through the power flow equations in (1) and adding the measurement noise in (4). Following [16], we will obtain an offline Monte Carlo simulation of the observed data and build a statistical approximation to the distribution of \mathbf{y} . That paper approximated solutions to the power flow equations and used importance sampling to connect to data observed with measurement error. Here, we simulate the data generation process all the way through to the measurements and build an approximate model for observed data. This eliminates the need for importance sampling and makes the real-time part of the procedure even faster.

Assume that the slack generator is instrumented with a PMU and that voltage angles are measured relative to that of the slack bus, designated as bus i_0 . Thus, \mathbf{y} contains measured angle differences $A_i - A_{i_0}$ for each bus i that is instrumented with a PMU, except i_0 . In addition \mathbf{y} has voltage magnitude measurements M_i for each non-generator bus with a PMU, as indicated in the final column of Table 1. A specific example is given for three PMUs in Section 4.

3.1 Log-spline Copula Models for PMU Measurements

To classify the topology with too few measurements for full state estimation, a flexible statistical model is needed to represent data generated under each scenario. A general purpose class of distributions with dependent components can be constructed using log-spline densities to

represent the univariate marginals of \mathbf{y} and these can be folded into a multivariate Gaussian copula to capture dependence amongst the variables. Pairing copulas with flexible log-spline densities is a powerful means of representing dependence among PMU measurements, leading to a method for estimating the current topology of the grid.

Log-spline densities. A log-spline density is the exponential of a natural cubic spline with k knots and constrained to integrate to 1. It can be represented as

$$f(y; \mathbf{b}, \mathbf{c}) \propto \exp \left[b_0 y + \sum_{i=1}^k b_i (y - c_i)_+^3 \right]$$

where $(\cdot)_+ \equiv \max(0, \cdot)$, $\mathbf{c} = (c_1, \dots, c_k)$ are the knots, and the coefficients $\mathbf{b} = (b_0, \dots, b_k)$ have k degrees of freedom because $\log f$ is constrained to be linear outside the range of knots. Log-spline densities are flexible for fitting a wide range of distributional shapes, while also having a convenient parametric form that is easy to store, reuse and evaluate.

Stone and Kooperberg [12, 13] develop a method for fitting log-spline densities, including choice of k . We interfaced Kooperberg's [11] C++ library to work within Matlab [19] in order to evaluate log-spline densities within an existing computing environment that simulates electric power loads and solves power flow equations.

Gaussian Copula Each element of \mathbf{y} is modeled with a log-spline density $y_i \sim f_i \equiv f(y; \mathbf{b}_i, \mathbf{c}_i)$ with F_i as its corresponding CDF. Taking the probability integral transform of each coordinate produces a random vector $\mathbf{u}(\mathbf{y}) = (F_1(y_1), \dots, F_p(y_p))$ that has uniform $U(0,1)$ marginals.

Copulas are multivariate distributions with $U(0,1)$ marginals. The Gaussian copula has joint cumulative distribution function equal to

$$\mathcal{C}_{\text{Gauss}}(\mathbf{u}; \mathbf{0}, \mathbf{\Sigma}) = \Phi_p \left[(\Phi_1^{-1}(u_1), \dots, \Phi_1^{-1}(u_p)); \mathbf{0}, \mathbf{\Sigma} \right] \quad (5)$$

where $\mathbf{\Sigma}$ is a correlation matrix, $\Phi_p(\cdot; \boldsymbol{\mu}, \mathbf{\Sigma})$ is the multivariate Gaussian CDF with mean $\boldsymbol{\mu}$ and variance $\mathbf{\Sigma}$, and Φ_1 is the standard univariate Gaussian CDF. Our models actually use the more general distribution $\mathcal{C}_{\text{Gauss}}(\mathbf{u}; \boldsymbol{\mu}, \mathbf{\Sigma})$ where non-zero $\boldsymbol{\mu}$ and general positive-semidefinite covariance $\mathbf{\Sigma}$ are permitted. The more general form is not necessary but the extra degrees of freedom over (5) will not harm the multivariate fit when an adequate Monte Carlo sample is used.

Modeling $\mathbf{u}(\mathbf{y}) \sim \mathcal{C}(\mathbf{u}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ implies that the density of \mathbf{y} is

$$g(\mathbf{y}; \boldsymbol{\theta}) = \phi_p(\mathbf{z}(\mathbf{y}); \boldsymbol{\mu}, \boldsymbol{\Sigma}) \prod_{i=1}^p \left| \frac{f(y_i; \mathbf{b}_i, \mathbf{c}_i)}{\phi_1(z_i(y_i))} \right| \quad (6)$$

with parameters $\boldsymbol{\theta} \equiv (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{b}_1, \mathbf{c}_1, \dots, \mathbf{b}_p, \mathbf{c}_p)$ and where

$$\mathbf{z}(\mathbf{y}) \equiv [(\Phi_1^{-1}(F_1(y_1))), \dots, \Phi_1^{-1}(F_p(y_p))] \quad (7)$$

and ϕ_p is the density of Φ_p . Notice that this model implies $\mathbf{z}(\mathbf{y}) \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In particular, the construction implies that each component of \mathbf{z} is approximately $\mathcal{N}_1(0, 1)$, which provides a basis for checking goodness of fit for the marginal distributions of \mathbf{y} ; see the example of Section 4.

Numerical evaluation of $g(\mathbf{y}; \boldsymbol{\theta})$ requires evaluation of p log-spline densities and standard Gaussian densities along with a multivariate Gaussian density in which $\boldsymbol{\Sigma}^{-1}$ can be pre-computed. Additionally, evaluation of the quantities $\mathbf{z}(\mathbf{y})$ requires evaluation of p log-spline CDFs and standard Gaussian inverse CDFs. These computations are straightforward, although the log-spline CDFs rely on one-dimensional quadrature for numerical integration. All told the combination of log-spline densities and a Gaussian copula is an efficient and flexible tool for modeling the distribution of a general measurement vector \mathbf{y} and computing its density.

Algorithm 1 steps through the process of estimating scenario-specific distributions of the measurement vector \mathbf{y} . For each scenario s the algorithm produces a set of simulated measurement vectors $\mathbf{Y}_s = \{\mathbf{y}_1, \dots, \mathbf{y}_J\}$, corresponding to PMUs on a set of buses \mathcal{B} . \mathbf{Y}_s is obtained by repeatedly drawing independent random loads, solving the power flow equations (1), and collecting random PMU measurements of voltage angles and magnitudes. Log-spline copulas are fit to each scenario-specific measurement set, producing a corresponding parameters vector $\boldsymbol{\theta}_s$.

3.2 Topology Estimation

Estimated distributions of \mathbf{y} for each possible scenario provide a direct means of estimating the current topology of the grid based on a new set of measurements \mathbf{y} . Let $g_s(\mathbf{y}; \boldsymbol{\theta}_s)$ denote the density of \mathbf{y} under the Gaussian log-spline copula \mathcal{C}_s for scenario s . Let $\pi(s)$ represent a

Algorithm 1 Estimate Scenario-Specific Distributions of \mathbf{y}

for all $s \in \{\text{nominal and single line down scenarios}\}$ **do**
 for $j = 1, \dots, J$ **do**
 draw random loads $P_i^{\text{load}}, Q_i^{\text{load}}$ as in Equation (3)
 compute P_i, Q_i for non-generator buses from Equation (2)
 set two of $P_i, Q_i, \theta_i, |V_i|$ as inputs according to Table 1
 solve power flow Equations (1) under scenario s
 draw PMU measurements $A_i | \theta_i, M_i | |V_i|$ as in Equation (4)
 form \mathbf{y} from measurements $A_i - A_{i_0}$ and M_i
 end for
 fit $\mathcal{C}_s = \text{Gaussian log-spline copula to } \mathbf{Y}_s = \{\mathbf{y}_1, \dots, \mathbf{y}_J\}$
end for

prior probability that the topology of the grid is given by scenario s . Then the posterior probability of scenario s is

$$\Pr(s | \mathbf{y}) = \frac{\pi(s)g_s(\mathbf{y}; \boldsymbol{\theta}_s)}{\sum_{s'} \pi(s')g_{s'}(\mathbf{y}; \boldsymbol{\theta}_{s'})}. \quad (8)$$

Equation (8) gives the probability of each topological scenario amongst the set of possibilities. Let S represent the number of scenarios (i.e. topologies) under consideration. The ability of PMU measurements to identify the current topological scenario is summarized by an $S \times S$ confusion matrix, \mathbf{C} , containing probabilities of classifying into one scenario when observations are generated under another scenario. The s, s' element of \mathbf{C} is

$$C_{s,s'} = E_s \Pr(s' | \mathbf{y}) \quad (9)$$

where E_s denotes expectation for \mathbf{y} generated according to scenario s . Diagonal elements of the confusion matrix, called *sensitivities*, are probabilities of correct classification under each given scenario.

4 Topology Estimation in a 37 Bus Network

The system in Figure 1 consists of 37 buses, 57 lines, 24 loads, and 9 generators. It originally appeared in [8, Ex. 13.9] and was analyzed in [24] and [25]. This section demonstrates estimation of the system

topology using log-spline copulas built from Monte Carlo simulations of PMU measurements as in Algorithm 1. The prior distribution of topology is taken to be uniform over the nominal (no fault) topology and each single line down scenario. In actual usage, the nominal topology would probably be given a much higher prior probability.

To facilitate a number of investigations, Algorithm 1 was applied to Monte Carlo simulations with virtual PMUs at each of the 37 buses. Results are based on $J = 4600$ simulated measurement vectors \mathbf{y} for the nominal scenario and each of the 57 single line down scenarios. In constructing \mathbf{y} , all 36 non-slack buses contribute voltage angle differences, $A_i - A_{i_0}$ and the 30 non-generator buses also contribute measurements of voltage magnitude. Therefore, \mathbf{y} has 66 components for a grid that is fully instrumented with PMUs. The slack generator is at bus 31.

Figure 3 has panels with normal probability plots of $(A_3, A_{17}, A_{30}) - A_{31}$ (left) and of M_3, M_{17}, M_{30} (right). Each panel contains 58 curves, one for each scenario. The three buses were selected to illustrate the variety of distributions across all 37 buses in the grid. Voltage magnitudes are normalized to the nominal voltage of each bus. Voltage angle measurements are clearly non-normal, whereas magnitude measurements are essentially normal. Nevertheless, all measurements were fit with log-spline densities to illustrate that log-spline fits do no harm to Gaussian data.

Figure 4 shows normal probability plots of $\mathbf{z}(\mathbf{y})$, the measured angles and magnitudes after transforming to normality by way of log-spline density fits and Equation (7). All distributions are essentially standard normal, demonstrating that fitted log-spline densities are good marginal approximations to the Monte Carlo data sets. Fitting multivariate normal distributions to the simulated sets of $\mathbf{z}(\mathbf{y})$ for each scenario completes the log-spline copula modeling of PMU data.

As a simple example, consider the problem on the 37-bus system from [24]. In a manner similar to [24], the system will be monitored with PMUs located at buses $\mathcal{B} = \{31, 40, 44\}$ with bus 31 designated as the slack generator, 40 as a non-generator load bus, and 44 as a regular generator bus. Recall that voltage angle measurements in \mathbf{y} are differences from the measured slack bus voltage angle. Therefore, the full vector of measurements is

$$\mathbf{y}(\mathcal{B}) = (A_{40} - A_{31}, A_{44} - A_{31}, M_{40})^T.$$

Subtraction of A_{31} from the other angles is a source of correlation

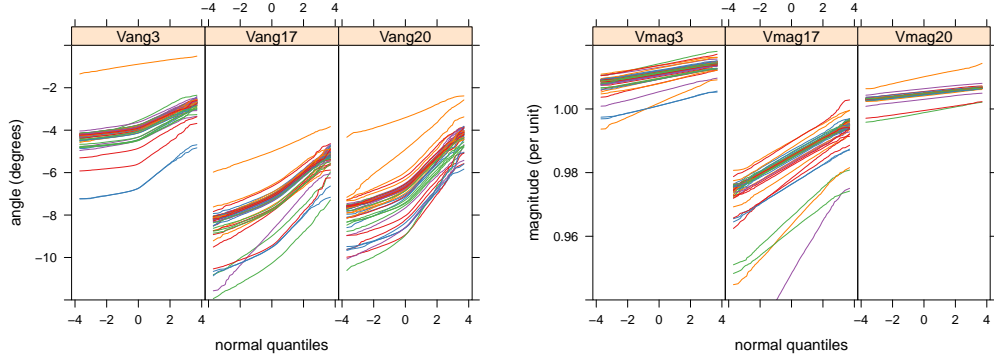


Figure 3: Normal probability plots of measured voltage angles (A_i , left) and magnitudes (M_i , right) at three representative buses in the 37-bus system of Figure 1. Fifty-eight scenarios are represented by the collection of curves in each panel, corresponding to the nominal topology and 57 single line scenarios. Voltage angles are decidedly non-normal. Marginal distributions for most of the scenarios are greatly overlapped with a few standouts. Correlations amongst measurements are important for estimating the current topology of the grid from such measurements.

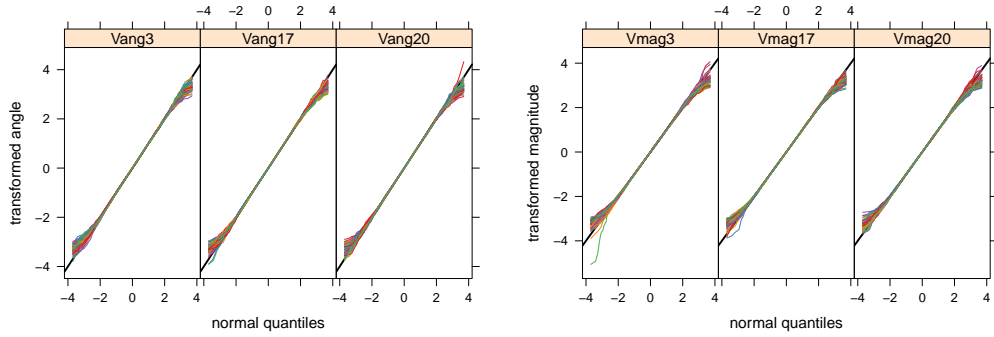


Figure 4: Normal probability plots of transformed voltage angles (left) and transformed magnitudes (right) at three representative buses in the 37-bus system of Figure 1. Each distribution from Figure 3 has been transformed to a standard normal by means of fitted log-spline densities and corresponding inverse probability transforms. Distributions on the diagonal indicate that the log-spline densities fit the data well.

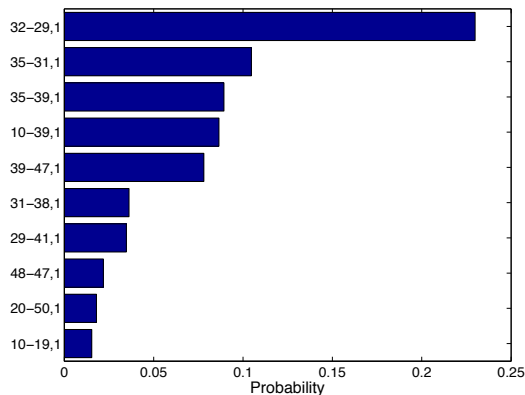


Figure 5: Top ten candidates for a single line outage when line 32-29 is down and PMU measurements are available from buses 31, 40 and 44. Probabilities are averaged over 4600 Monte Carlo samples of the measurements. These results are comparable to [24] who demonstrated that changes in voltage angles can finger the correct downed line, but with some ambiguity among other candidates.

among elements of \mathbf{y} and this is modeled by the covariance Σ in the log-spline copula distributions $g_s(\mathbf{y}; \theta_s)$ associated with \mathcal{B} . In [24], only the voltage angle measurements were used to demonstrate a method of detecting and locating a single line outage. They did not use the voltage magnitude measurement M_{40} , that we assume is available from the non-generator bus.

Figure 5 is a bar plot of the top ten candidates for a single line outage when line 32-39 is down. Each bar represents a confusion element $C_{s,s'}$ from Equation (9) with the true scenario $s = \text{line 32-39 down}$. The largest bar is the correct scenario, showing that three PMUs at buses 31, 40 and 44 have about 23% sensitivity for discovering this particular line outage.

These results are qualitatively comparable to those of [24, Table 2] in which a certain quality of fit score correctly identified the line outage with several other candidates obtaining substantially worse fits. The method of [24] monitors PMU time series data, responding to step changes in the level of the series by matching changes in voltage angles with what would be expected from line outages. On each bus instrumented with a PMU, the most recent phasor angle measurement is subtracted from the corresponding phasor angle measured at

a previous point in time. An outage is signaled if a lagged difference exceeds a predetermined threshold. Post-processing then determines time steps that are before and after the outage and computes voltage angle changes between those times. Voltage angle changes are compared to what is expected from each single line down scenario with the best fitting scenario (measured by a quality of fit score) flagged as the most likely outage. This method makes effective use of the unfolding time series of PMU measurements, whereas our results have focused on a vector of PMU measurements at a single point in time.

5 Selection of buses for PMUs

PMUs are becoming more common but are not yet widely installed in the U.S. [2], especially in the distribution portions of the power grid. Operators need to decide where to place PMUs, and sensitivity to line outages is an important consideration. This section discusses placing PMUs so that their measurements have good sensitivity to single line failures.

Let $\mathbf{C}(\mathcal{B})$ denote the confusion matrix associated with observations \mathbf{y} from PMUs at a subset \mathcal{B} of the buses. Average sensitivity

$$S(\mathcal{B}) = \frac{1}{S} \text{tr}(\mathbf{C}(\mathcal{B})) \quad (10)$$

is a sensible figure of merit for selecting a good subset of buses to instrument with PMUs.

Optimal selection of a subset of buses is combinatorially hard and the associated computations can only be performed for small numbers of PMUs. Greedy stepwise placement is a tractable alternative. Given a reference starting bus j_0 , Algorithm 2 sequentially adds a PMU to the next most advantageous bus for improving average sensitivity. This is known as greedy or forward-stepwise selection. The reference bus can be selected using a variety of factors such as a bus that already has a PMU installed or availability of communication lines to a central Phasor Data Concentrator. Alternatively, the algorithm could begin with the best small subset of buses, chosen by exhaustive search, or a subset of buses that are already fitted with PMUs.

Each forward step (i) of Algorithm 2 adds the most advantageous bus to the current set of buses \mathcal{B}_i . The best bus to add is determined from estimates of average sensitivity $\hat{S}(\mathcal{B}_{i,j})$ attained by sequentially

Algorithm 2 Stepwise Selection of PMU Locations

```
recall  $\mathbf{Y}_1, \dots, \mathbf{Y}_S$ , simulated measurement sets from Algorithm 1
 $\mathcal{B}_1 \leftarrow \{j_0 = \text{reference bus}\}$ 
for  $i \in \{2, \dots, B - 1\}$  do
  for  $j \in \{1, \dots, B\} \setminus \mathcal{B}_{i-1}$  do
     $\mathcal{B}_{ij} \leftarrow \mathcal{B}_i \cup \{j\}$ 
    for  $s \in \{1, \dots, S\}$  do
       $\mathbf{Y}_s(\mathcal{B}_{ij}) \leftarrow$  rows of  $\mathbf{Y}_s$  corresponding to buses in  $\mathcal{B}_{ij}$ 
       $\hat{C}_{s,s}(\mathcal{B}_{ij}) \leftarrow$  Equation (8) averaged over sample  $\mathbf{Y}_s(\mathcal{B}_{ij})$ 
    end for
     $\hat{S}(\mathcal{B}_{ij}) \leftarrow \sum_s \hat{C}_{s,s}(\mathcal{B}_{ij})/S$  ▷ sensitivity to add bus  $j$ 
  end for
   $\mathcal{B}_i \leftarrow \mathcal{B}_{i-1} \cup \{\arg \max_j \hat{S}(\mathcal{B}_{ij})\}$  ▷ add max sensitivity bus
end for
Stepwise PMU sets are  $\mathcal{B}_2, \dots, \mathcal{B}_B$ 
```

trying a PMU on each available bus j with the current set \mathcal{B}_i . Evaluation of $\hat{S}(\mathcal{B}_{ij})$ requires, for each scenario s , averaging $\Pr(s|\mathbf{y})$ (8) over all $\mathbf{y} \in \mathbf{Y}_s$. Recall that \mathbf{Y}_s is the Monte Carlo measurement set produced in Algorithm 1 to estimate the multivariate density of all-possible PMU measurements under scenario s .

Within the innermost loop of Algorithm 2 each realization of complete measurements $\mathbf{y} \in \mathbf{Y}_s$ is reduced to a vector containing measurements only on the subset of PMUs under consideration, \mathcal{B}_{ij} . Correspondingly, the log-spline copula model is reduced for evaluation of the confusion matrix associated with \mathcal{B}_{ij} . These subsetting operations are straightforward and do not require refitting of any univariate log-spline densities or Gaussian copulas. However, if inverse covariance matrices are stored with the copulas, these will need to be recomputed for each different subset of variables entertained in the stepwise loops over i and j . Some computational efficiency can be gained by forming inverse covariance matrices of the Gaussian copula for \mathcal{B}_{ij} as a one dimensional expansion of the inverse covariance matrix for the copula associated with \mathcal{B}_i .

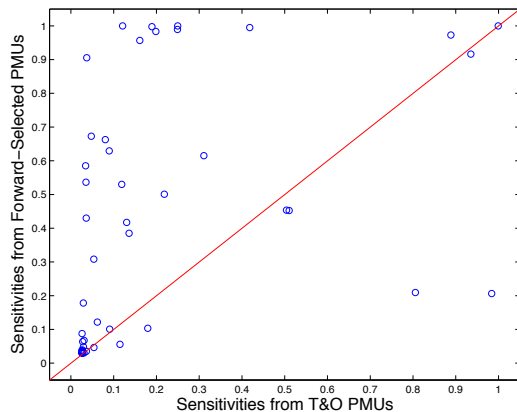


Figure 6: Sensitivities for each of 66 single-line down scenarios with measurements from two sets of three PMUs—forward stepwise selection (buses 31, 33, 10) versus the placement of [24] (buses 31, 40, 44). Stepwise PMU selection improves sensitivity for 43 of 58 outage scenarios.

6 PMU Placement in the 37 Bus Network

Algorithm 2 was used to place PMUs at buses in the 37 bus network studied by [24], beginning with a PMU at the slack bus $i_0 = 31$, shown with a label A in Figure 1. With stepwise selection, bus 33 was added first and bus 10 was added next; these are labeled B, and C in the figure.

Recall that [24] used PMUs at buses $\{31, 40, 44\}$, as did our demonstration of outage detection. Figure 6 plots sensitivities for each of the 58 scenarios for two placements of three PMUs—stepwise selection $\mathcal{B}_3 = \{31, 33, 10\}$ on the vertical axis versus buses $\{31, 40, 44\}$ on the horizontal axis. Average sensitivity is 0.250 for stepwise compare to 0.168 for the choice of [24]. Forty-three of the 58 single line outages are easier to detect with the three stepwise PMUs; two outages are substantially more difficult to detect. In either case substantial uncertainty remains with PMUs at only three buses.

Stepwise bus selection continued from three PMUs to all 37 buses. The lower-right plot in Figure 7 show the average sensitivity verses step number. After placing about 10 PMUs on this grid, little additional information is gained for detecting single line outages. In part this is due to parallel lines in the power system that have different

properties. It can be difficult to distinguish which line has failed. The other three plots in Figure 7 show sensitivity traces for each of the 58 scenarios, grouped into four qualitatively different behaviors: traces in the top-left plot have either high sensitivity or low sensitivity, that is largely unaffected by PMU placement. Traces in the top-right panel have low sensitivity with two PMUs but their sensitivities quickly climb to a high level with some number of PMUs no larger than 10. Traces in the lower-left have more gradual increase in sensitivity as PMUs are added.

The first 10 PMUs added by stepwise selection are shown in Figure 1 by labels A through J. These appear to be sequentially well-spread throughout the network. Recall that bus 31 was selected for PMU placement because it is the slack bus. Notice that none of the next nine buses has a generator attached. The reason is apparently related to PMUs at non-generator buses providing two measurements, A_i and M_i , whereas PMUs at generator provide only a single measurement, A_i .

7 Future Work

We have presented a method for identifying the probabilities of possible power grid topologies based on PMU data. We have also used this method to discover good PMU locations. There are a number of interesting additional research directions to pursue.

One possible direction is to consider the time series of PMU measurements as done by [24, 25]. Our model bank formation could be extended by building it into a hidden Markov model with transitions between the topologies in the model bank. This formulation would be a natural fit to PMU data and would likely help distinguish some of the more ambiguous observations. An important possible component of this model is that the prior for loads may exhibit diurnal patterns or include correlations based, for example, on weather.

A second direction for new research would be exploring other methods of PMU placement. Here, we use the method of forward selection from the variable selection literature. Following this further, two additional methods, backward elimination and forward-backward stepwise, should be investigated. We used the average sensitivity across the model bank, but there may be better choices. Further, for small numbers of PMUs, these heuristics need to be evaluated by compari-

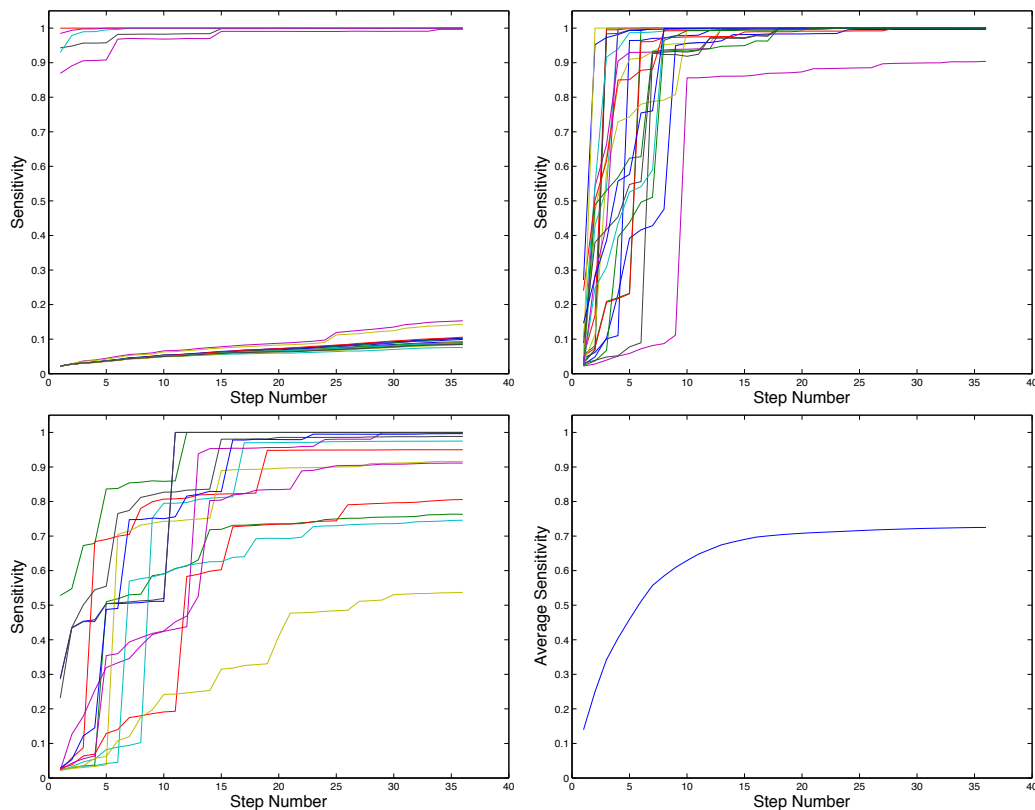


Figure 7: Sensitivity traces for the single-line down scenarios as a function of number of PMUs in the measurement set. Some outages (top left) are detected easily (5 lines) or poorly (14 lines), regardless of how many PMUs are used. Others (top right, 23 lines) have large jumps in sensitivity as the first 10 PMUs are placed. Still others (bottom left, 14 lines) improve more slowly with increasing numbers of PMUs. The lower right plot shows that the average sensitivity climbs readily at first but eventually additional PMUs improve outage identification only slightly.

son with an exhaustive search of the best locations.

PMUs also provide more data than we and others in the literature have used. Specifically, they can also provide measurements of the current flowing on the lines connected to the PMU's node. This provides indirect information about the voltage magnitudes and angles at neighboring nodes. We need to explore the utility of this seemingly valuable information.

Finally, we would like to move beyond the model bank formulation, which is necessarily limiting. This would mean developing methodology for searching over the more general space of possible topologies. Such a search would present computational challenges, but would provide more reliable estimation.

References

- [1] A. P. Alves da Silva, V. H. Quintana, and G. K. H. Pang. Solving data acquisition and processing problems in power systems using a pattern analysis approach. *Proceedings of the Institute of Electrical Engineering C*, 138(4):365–376, 1991.
- [2] Jim Y Cai, Zhenyu Huang, John Hauer, and Ken Martin. Current status and experience of wams implementation in north america. In *Transmission and Distribution Conference and Exhibition: Asia and Pacific, 2005 IEEE/PES*, pages 1–7. IEEE, 2005.
- [3] Jian Chen and Ali Abur. Enhanced topology error processing via optimal measurement design. *Power Systems, IEEE Transactions on*, 23(3):845–852, 2008.
- [4] K. A. Clements and P. W. Davis. Detection and identification of topology errors in electric power systems. *IEEE Transactions on Power Systems*, 3(4):1748–1753, 1988.
- [5] KA Clements and PW Davis. Detection and identification of topology errors in electric power systems. *Power Systems, IEEE Transactions on*, 3(4):1748–1753, 1988.
- [6] Kevin A Clements and A Simões Costa. Topology error identification using normalized lagrange multipliers. *Power Systems, IEEE Transactions on*, 13(2):347–353, 1998.
- [7] James W. Evans. Interface between automation and the substation. In John D. McDonald, editor, *Electric Power Substations Engineering*, chapter 6. CRC Press, 2012.

- [8] Antonio Gómez-Expósito, Ali Abur, Patricia Rousseaux, Antonio de la Villa Jaén, and Catalina Gómez-Quiles. On the use of PMUs in power system state estimation. In *Proc. 17th Power Systems Computation Conference*, pages 22–26, 2011.
- [9] Yih-Fang Huang, Stefan Werner, Jing Huang, Neelabh Kashyap, and Vijay Gupta. State estimation in electric power grids: Meeting new challenges presented by the requirements of the future grid. *Signal Processing Magazine, IEEE*, 29(5):33–43, 2012.
- [10] Mladen Kezunovic. Smart fault location for smart grids. *Smart Grid, IEEE Transactions on*, 2(1):11–22, 2011.
- [11] Charles Kooperberg. *logspline: Logspline density estimation routines*, 2009. R package version 2.1.3.
- [12] Charles Kooperberg and Charles J Stone. A study of logspline density estimation. *Computational Statistics & Data Analysis*, 12(3):327–347, 1991.
- [13] Charles Kooperberg and Charles J Stone. Logspline density estimation for censored data. *Journal of Computational and Graphical Statistics*, 1(4):301–328, 1992.
- [14] Mert Korkali and Ali Abur. Optimal deployment of wide-area synchronized measurements for fault-location observability. *Power Systems, IEEE Transactions on*, 28(1):482–489, 2013.
- [15] Mert Korkali, Hanoch Lev-Ari, and Ali Abur. Traveling-wave-based fault-location technique for transmission grids via wide-area synchronized voltage measurements. *Power Systems, IEEE Transactions on*, 27(2):1003–1011, 2012.
- [16] Earl Lawrence, Russell Bent, and Scott Vander Wiel. Model bank state estimation for power grids using importance sampling. *Technometrics*, 2013.
- [17] Qiao Li, Tao Cui, Yang Weng, Rohit Negi, Franz Franchetti, and Marija D Ilic. An information-theoretic approach to pmu placement in electric power systems. *Smart Grid, IEEE Transactions on*, 4(1):446–456, 2013.
- [18] Elizete Maria Lourenco, A Simões Costa, and Kevin A Clements. Bayesian-based hypothesis testing for topology error identification in generalized state estimation. *Power Systems, IEEE Transactions on*, 19(2):1206–1215, 2004.

- [19] MATLAB. *version 7.10.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts, 2010.
- [20] D Singh, J. P. Pandey, and D. S. Chauhan. Topology identification, bad data processing and state estimation using fuzzy pattern matching. *IEEE Transactions on Power Systems*, 20(3):1570–1579, 2005.
- [21] H. Singh and F. L. Alvarado. Network topology determination using least absolute value state estimation. *IEEE Transactions on Power Systems*, 10(3):1159–1165, 1995.
- [22] N. Singh and H. Glavitsch. Detection and identification of topological errors in online power system analysis. *IEEE Transactions on Power Systems*, 6(1):324–330, 1991.
- [23] Ravindra Singh, Efthymios Manitsas, Bikash C. Pal, and Goran Strbac. A recursive bayesian approach for identification of network configuration changes in distribution system state estimation. *IEEE Transactions on Power Systems*, 25(3):1329–1336, 2010.
- [24] Joseph Euzebe Tate and Thomas J Overbye. Line outage detection using phasor angle measurements. *Power Systems, IEEE Transactions on*, 23(4):1644–1652, 2008.
- [25] Joseph Euzebe Tate and Thomas J Overbye. Double line outage detection using phasor angle measurements. In *Power & Energy Society General Meeting, 2009. PES'09. IEEE*, pages 1–5. IEEE, 2009.
- [26] Gustavo Valverde and Vladimir Terzija. Unscented kalman filter for power system dynamic state estimation. *IET generation, transmission & distribution*, 5(1):29–37, 2011.
- [27] D. M. Vinod Kumar, S. C. Srivastava, S. Shah, and S. Mathur. Topology processing and static state estimation using artificial neural networks. *IEE Proceedings Generation, Transmission & Distribution*, 143(1):99–105, 1996.
- [28] Felix F. Wu and Wen-Hsiung E. Liu. Detection of topology errors by state estimation. *IEEE Transactions on Power Systems*, 4(1):176–183, 1989.
- [29] Tao Yang, Hongbin Sun, and Anjan Bose. Transition to a two-level linear state estimator part i: Architecture. *Power Systems, IEEE Transactions on*, 26(1):46–53, 2011.