Robust Data-Driven State Estimation for Smart Grid

Yang Weng, *Student Member, IEEE*, Rohit Negi, *Member, IEEE*, Christos Faloutsos, *Member, IEEE*, and Marija D. Ilić, *Fellow, IEEE*

***Abstract*—AC power system state estimation process aims to produce a real-time "snapshot" model for the network. Therefore, a grand challenge to the newly built smart grid is how to "optimally" estimate the state with increasing uncertainties, such as intermittent wind power generation or inconsecutive vehicle charging. Mathematically, such estimation problems are usually formulated as Weighted Least Square (WLS) problems in literature. As the problems are nonconvex, current solvers, for instance the ones implementing Newton's method, for these problems often achieve local optimum, rather than the much desired global optimum. Due to this local optimum issue, current estimators may lead to incorrect user power cut-offs or even costly blackouts in the volatile smart grid. Frequent topology changes, poor measurement accuracy, and malicious attack can further deteriorate the state estimate. To solve the problem, in this paper, we propose utilizing historical data of Energy Management System to efficiently obtain a good state estimate. Specifically, kernel ridge regression is proposed in a Bayesian framework based on robust Nearest Neighbors search. To enable online data-driven SE, techniques such as dimension reduction and k-dimensional tree indexing are employed with 1000 times speed up in simulations. Further numerical results show that the new method produces an state estimate excelling current industrial approach.**

***Index Terms*—Smart grid, state estimation, historical data, robustness, k-nearest neighbors, kernel ridge regression, speed up.**

I. Introduction

Initiated by the U.S. government, the rapid-expanding smart grid aims to evolve into an efficient, reliable and sustainable modern grid by adopting, integrating, and advancing the communication and computing technologies already exist. To achieve such an ambitious goal, namely the "smartness" of the power grid, a highly accurate State Estimation (SE) process [1], [2] is necessary in providing bases for many key functionalities in the operation and control of smart grid.

However, the nonlinear measurement model of AC power system renders the SE problem a highly nonconvex character, which can not be optimally solved without great computational expense. To deal with nonlinearity, one can approximate AC power system with a DC model [3], [4], based on which robust state estimation can be further applied to deal with bad data [5]–[7]. One can also try to convexify the nonconvex SE problem by convex relaxation [8]–[10]. However, those technics usually come with an approximation cost, resulting in a relative poor estimate. Therefore, many successful and widely used SE algorithms for the power grid are to directly work with nonlinear measurement model, which is formulated in the Weighted Least Square (WLS) [11]–[15] form, with Newton's method (i.e. [3]) to be the solver. By successively finding a better approximation, Newton's method can reach a local optimum of the non-convex problem. However, obtaining global optimum is not guaranteed. If the initial guess used in Newton's method is (by pure chance) close to global optimum, then it is likely to find global optimum. Otherwise, it may merely reach a local optimum and stop.

In traditional transmission network, it is possible to use previous state estimate as a heuristic initial guess for SE, based on the belief that no significant change appears in a short time. However, such a belief will no longer hold in smart grid, where intermittent generation (wind and solar farms) and consumption (plug-in hybrid electric vehicles), and frequent topological changes can lead to significant state shift in power system operations. In such case, a previous state estimate computed around 2 minutes ago [16] may not truly reflect the operating point of the current power system and generates suboptimal results accordingly. Although not suitable for smart grids, utilizing the previous state estimate as a prior knowledge for the current SE does reflect an important idea in power systems analysis: using historical data in a smart way can enhance real time analysis against uncertainties.

On the other side, recent advances in communications, sensing, computing and control, as well as the targeted investments toward deploying advanced meter infrastructures (AMIs) and synchrophasors have become drivers and sources of data previously unavailable in the electric power industry. Such a database is expected to exhibit exponential growth. With vast amounts of data being generated in the power grids, researchers and engineers need to address questions, such as what patterns and trends to extract and how to use them to improve power systems reliability, security, sustainability, efficiency and flexibility.

This paper aims to utilize such valuable historical data resources to improve SE accuracy against the ever-changing hard-to-predict uncertainties in smart grids. Instead of using a single data point (last state estimate), a key idea in this paper is to use more historical data (i.e. state, topology, and measurement) for robustness. Our preliminary work [17] proposes a Bayesian approach based on historical data search, where a group of measurement sets and the corresponding state estimates are used in combination with the current measurement in a kernel ridge regression [18] to pursue a good estimate of current states. The proposed method is based on the idea that two similar system measurement sets usually indicate two similar operation conditions (system states). After collecting

The manuscript has been submitted to the Special Issue on Neural Networks and Learning Systems Applications in Smart Grid.

Yang Weng, Rohit Negi, and Marija D. Ilić are with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213 USA (e-mail: yangweng@andrew.cmu.edu,{negi, milic,}@ece.cmu.edu).

Christos Faloutsos is with the Department of Computer Science at Carnegie Mellon University, Pittsburgh, PA, 15213 USA (e-mail: christos@cs.cmu.edu).

a group of similar measurements in the past, a supervised learning framework is employed to map the historical data to the current state estimate.

Although showing promising result in smart grids, [17] implicitly places strong assumptions over historical data. For instance, it assumes that the historical data are without 1) topology changes [3], [19], [20], 2) bad data [7], [21], and 3) malicious attack [22]. Unfortunately, such assumptions is frequently violated in practice [3], [7].

To enable robustness for the algorithm, in this paper, we generalize system learning process. Instead of using only one step, we propose three steps to systematically locate good and robust nearest neighbor points. In particular, we firstly utilize historical state and system topology information to deal with topology changes. Secondly, historical measurements are used to refine the data set against bad data. Subsequently, we conduct a maximum agreement algorithm for collected states to identify malicious attack. Then the resulting information is used in a supervised learning process with Kernel Ridge Regression, leading to a robust data-driven state estimate. Such an estimate can also serve as an improved initial guess in iterative algorithms for potential improvement.

While resulting in a highly accurate state estimate as one will see in the simulation result, the new method faces a large computational overhand preventing its online application. For instance, the similarity evaluation over high dimensional power system measurement vectors is time consuming, especially in a large electric power grid. Besides, the time required to exhaust all historical data points is formidable, preventing any streaming estimation process with real-time guarantees for sustainable grid services [23].

To reduce the computational burden, the structure of power system measurement data is examined via singular value decomposition (SVD). We observe an important phenomenon. Thanks to the roughly periodic pattern of power system, the measurement data are highly clustered, creating the potential for speed up. A direct and natural idea is to reduce the measurement complexity (dimensionality) over measurement space. Random mapping is proposed in this paper [24], [25] to achieve dimension reduction quickly, and be adaptive to the new incoming data. To further reduce the computation over the space of time, we propose to organize/index the clustered data into a tree structure. In such a structure, the time for similarity check can be dramatically reduced due to data grouping. For instance, once a sub-tree is chosen, all the other sub-trees can be ignored, resulting in a log-reduction over time. The proposed method is advantageous in practical scenarios because it reorganizes the historical information in a form that can provide information in a compact way.

Whereafter, the performance of the data-driven SE approach is verified by simulations on the standard IEEE 300-bus test case [26], [27]. Provided with enough historical data, the new method can improve the performance of the traditional approach in a short time. 1000 times speed up is achieved, making our method feasible for online application.

The rest of the paper is organized as follows: Section II reviews the WLS state estimation and defines the problem of Data-Driven State Estimation; Section III describes the Nearest Neighbors approach; Section IV shows how to reduce computation time for online applications; Section V illustrates the simulation results and section VI concludes the paper.

II. Power System State Estimation

In this section we briefly review the state estimation (SE) problem in power systems. In general, the state estimation problem is a nonlinear problem that needs to be solved by implementing iterative algorithms.

A. Current Model

The following Static AC power system model is usually assumed in static SE:

$$\boldsymbol{z} = \boldsymbol{h}(\boldsymbol{x}) + \boldsymbol{w}, \tag{1}$$

where the vector $\boldsymbol{x}$

$$\boldsymbol{x} = (|v_1|e^{j\delta_1}, |v_2|e^{j\delta_2}, \cdots, |v_n|e^{j\delta_n})^T \tag{2}$$

represents the power system states, $\boldsymbol{w}$ is an $m \times 1$ vector denoting the additive measurement noises, presumably independent Gaussian random variables with zero means, i.e., $\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$, where Σ is a diagonal matrix, with the i^{th} diagonal element σ_i^2. $\boldsymbol{z}$ is an $m\times 1$ vector denoting the set of telemetered measurements, such as power flows and voltage magnitudes. $\boldsymbol{h}(\cdot)$ is a vector of nonlinear functions relating the states $\boldsymbol{x}$ to the measurements $\boldsymbol{z}$. In practice, the measurement set $\boldsymbol{z}$ is usually made redundant to guarantee the observability of the whole system.

The goal of power system SE is to find an estimate ($\hat{\boldsymbol{x}}$) of the true states ($\boldsymbol{x}$) that best fits the measurement set $\boldsymbol{z}$ according to the measurement model in (1). This is usually achieved by minimizing the following criterion:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} L_p(\boldsymbol{x}) = \sum_{i=1}^{m} \left(\frac{z_i - h_i(\boldsymbol{x})}{\sigma_i} \right)^p, \tag{3}$$

where the parameter p ($p \geq 0$) is used to achieve desired performance. For example, for $p = 2$, the above problem corresponds to the conventional Weighted Least Square (WLS) SE. For $p = 1$, the above problem reduces to the Weighted Least Absolute Value (WLAV) SE [3], which is well-known as robust to bad data.

1) WLS State Estimation: As we already discussed, the optimization problem in (3) is highly nonconvex, due to the non-convexity of the cost function in (3). Thus, it is very difficult to solve the problem optimally. In practice, the state estimation problem is usually solved by using Newton's method, which is essentially a local search algorithm. For instance, if we set $p = 2$, then the problem becomes equivalent to:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} J_2(\boldsymbol{x}) = (\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x}))^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x})). \tag{4}$$

After obtaining an initial guess $\boldsymbol{x}^{(0)}$, Newton's method updates the estimate according to the following rule:

$$\boldsymbol{x}^{(i+1)} = \boldsymbol{x}^{(i)} - \frac{J_2'(\boldsymbol{x}^{(i)})}{J_2''(\boldsymbol{x}^{(i)})} \Delta \boldsymbol{x}, \qquad \forall i \in \mathcal{N}, \tag{5}$$

where i is the iteration index, $\Delta \boldsymbol{x}$ is the step size, and $\mathcal{N}$ is the set of natural numbers.

Notice that, such a local search method is highly sensitive to the initial guess. As a result, for smart grid SE, the simple industrial initial guess approach (use the last SE result) may not be able to provide a good initial guess for the Newton's method to converge to the global optimum. In the next section, we are going to discuss a new systematic approach to obtain a robust data driven SE.

B. Problem Setup

Now we consider a new SE method to lessen the error in a static SE. We assume the availability of a database recording the historical measurements, topologies, and state estimates.

- Problem: Obtain a data-driven state estimator
- Given:
  - a sequence of historical measurement column vectors: $\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_k, \cdots \boldsymbol{z}_Q$, where k is the time index, and Q is the total number of data points in the database;
  - a sequence of historical state estimates column vectors: $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_k, \cdots, \boldsymbol{x}_Q$;
  - a sequence of historical measurement function sets: $\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_k, \cdots, \boldsymbol{h}_m$;
  - the current measurement column vector: $\boldsymbol{z}_{current}$;
  - the current measurement function set: $\boldsymbol{h}_{current}$.
- Find: Robust State Estimate without solving (3).

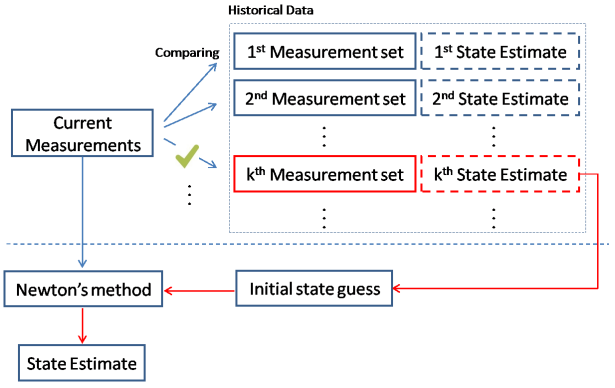


Fig. 1: Flow chart

III. Robust Data-Driven Approach

Mathematically, the proposed Robust Data-Driven SE algorithm can be decomposed into two parts:

- A minimization problem to obtain a group of likely historical data;
- A kernel ridge regression problem to obtain an "optimal" state estimate from the group.

We will detail the idea behind such a method in the next two subsections.

A. A Nearest Neighbors Approach

Intuitively, with the same topology, close-by states usually produce similar measurements. Therefore, a smaller distance between the current measurement set $\boldsymbol{z}_{\text{current}}$ and a historical measurement set $\boldsymbol{z}_k$ at time k implies that the associated historical state vector $\boldsymbol{x}_k$ stays closer to the current true state vector $\boldsymbol{x}_{\text{true}}$ with high probability. Provided with the aforesaid, we proposed a historical data search method. To make the estimation unbiased to a single data point, a group of nearest neighbors is obtained instead of a single historical data point.

Such a method, as shown in Fig.1, is called K-Nearest Neighbors (K-NNs) [1] approach in Statistics, which is a nonparametric method requiring no model to fit. Specifically, given a query point $\boldsymbol{z}_{current}$, we find K training points ($\boldsymbol{z}_k$) closest in distance to the query point. Despite mild structural assumptions and algorithmic simplicity, K-NNs's predictions are often accurate, leading to its successes in a large number of classification problems, including handwritten digits, satellite image scenes and EKG patterns.

Unfortunately, such a naive comparison in the measurement space did not count the scenarios of topology changes, bad data, and malicious attack:

1) The method above is non-robust to topology changes. If it happens, small distance between measurements may result in undesired large distance between states.
2) Data points with relatively large errors in the historical measurement may deteriorate the quality of the collected data points. Different level of bad data may harm the overall quality of the estimation result.
3) If unobserved malicious bad data injection appears in static SE, it should not be collected in the robust K-NNs due to its faking information.

To ensure robustness, we propose to use not only the measurements in the past and now [17], but also historical state, historical topology data and current topology. As more information is used, better performance is expected. The key is how to smartly map those information pieces into the current states via system learning. In the following, we illustrate how to adjust the K-NNs approach [17] to make it robust to topology changes, bad data, and occasional malicious data attack.

1) Topology Changes: To capture the topology changes in the smart grids, instead of using historical measurements for comparison, historical psudo-measurements are used.

$$\hat{\boldsymbol{z}}_k = \boldsymbol{h}_{current}(\boldsymbol{x}_k) \tag{6}$$

where $\boldsymbol{h}_{current}$ is the measurement function associated with the current topology. Mathematically, the K-NNs results that is robust to the topology changes can be expressed as the following optimization.

$$\hat{\boldsymbol{s}} = \arg \min_{|\boldsymbol{s}|=K} d(\boldsymbol{s}) = \sum_{k \in \boldsymbol{s}} ||\boldsymbol{z}_{\text{current}} - \hat{\boldsymbol{z}}_k||_2^2, \ k \leq Q, k \in \mathcal{N} \tag{7}$$

i.e., minimizing sum distance function $d(\boldsymbol{s})$. Here Q is the number of total data points in the database. $\mathcal{N}$ represents the

[1]The tuning parameter K can be chosen by cross-validation [18].

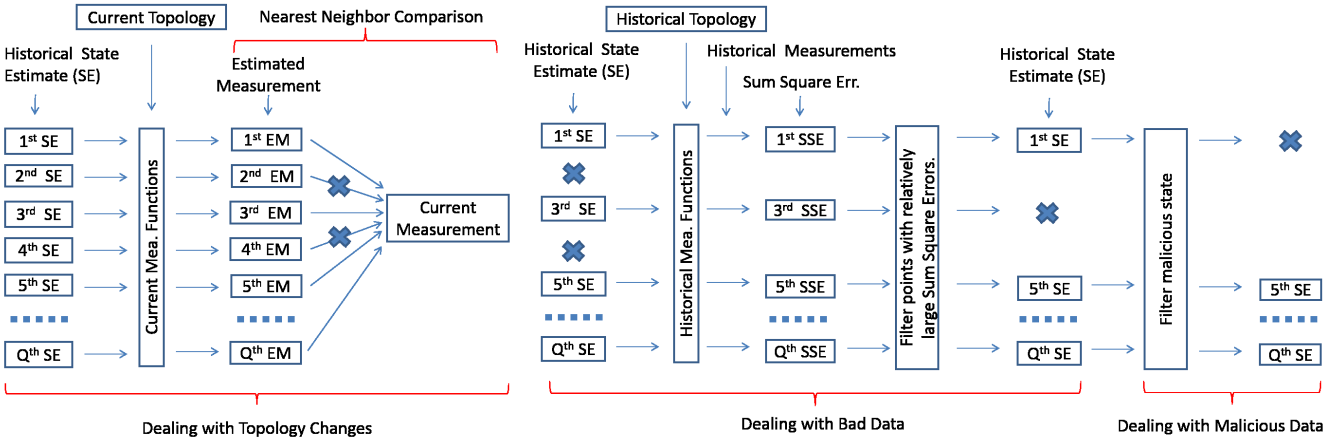


Fig. 2: Diagram for robust nearest neighbors search

set of natural number. k represents a particular index for a data point. As a result, $\boldsymbol{z}_k$ indicates the measurement set within the time slot of index k. Finally, K indicates the cardinality of the set $\boldsymbol{s}$. Essentially, during the searching step, the algorithm simply looks for an index set $\boldsymbol{s}$ with K elements which represents a group of psudo-measurement vector sets that have nearer distance to the current measurement $\boldsymbol{z}_{\text{current}}$. Such a process is illustrated on the left side of Fig.2. Notice that, such a searching process is time consuming, due to the high dimensionality of measurement vectors and a large volume of data points over a long time. Section IV is devoted to this problem for online practice.

2) Dealing with unfiltered bad data: Bad data occurs due to equipment failure, finite accuracy, infrequency of instrument calibration and measurement scaling procedure at the control center. Besides, telecommunication errors and incorrect topology information may also cause bad data. Traditional static bad data detection and filtering are based on Chi-square test [3]. After choosing a threshold over the probability of error, i.e. 5%, weighted sum square error are evaluated in a Chi-square distribution under the rules of Chi-square test. Unfortunately, the threshold is usually subjective and can not guarantee the absent of relative large bad data. Luckily, when many data points are used in a data-driven framework, one obtain the opportunity for the first time to further improve the SE result by selecting relative better data points. In this paper, instead of using the hard decision process, we propose to compare the sum square errors in different time slots by utilizing the extra degree of freedom over time. In this paper, as illustrated in the center of Fig. 2, we proposed to remove data points with relatively large sum square errors, i.e. 10% data points will be deleted.

3) Dealing with malicious bad data: [22] presents a new class of bad data, namely the false data injection, against state estimation in electric power grids. By exploiting the measurement function (topology information) of a power system, an attacker can successfully introduce arbitrary errors into certain state variables while bypassing existing techniques for bad measurement detection.

Such an attack is hard to detect for a static SE in power systems, due to its unobservability. However, with a long historical data, one can compare different data points across the time domain. For this reason, we can compare all the collected states and filter out the outliers. In other words, instead of looking into false data injection among different measurements in a single time slot, we propose to examine the state vectors in multiple time slots to filter out data points that are inconsistent with others according to some metric, such as the definition of relative outliers [28]–[30]. Fig.3 illustrates the idea over a two bus system via synthetic data. The x coordinate represents the voltage magnitude of the first bus. The y coordinate represents the voltage magnitude of the second bus. As power system data are highly clustered, which is shown in Sec.IV, the point o_1 and o_2 are regarded as suspected data points to be filtered out.

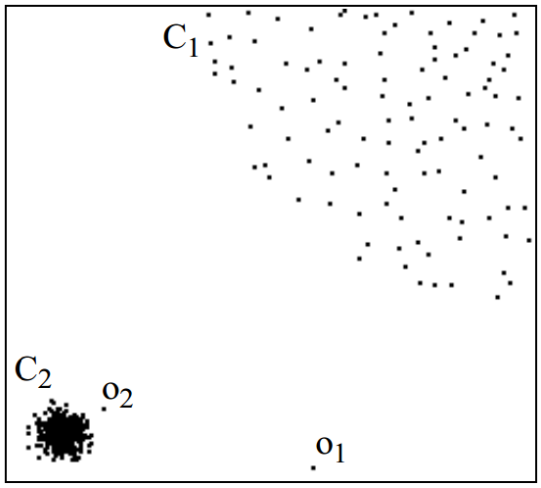


Fig. 3: Visualization of voltage magnitude pairs in a two bus system

Overall, the robust K-NNs method discussed in this section (Fig.2), aims at filtering out unnecessary data information to shrink the size of data. In the next section, we will "average" these historical similar information by conducing a regression problem. Together, such a two-stage process is called K-

Nearest Neighbors Regression in Statistics [31].

B. Bayesian Inference

To combine prior beliefs with data in a principled way, we proposed to conduct Bayesian inference over the collected K-Nearest Neighbor points to obtain a data-driven SE. For such an inference, one can use Generative model or Discriminative model. Although a Generative model is more informative, and can perhaps be obtained from physical principles, it needs to specify the probability distribution of the hidden parameters (power system states) and the conditional probability of the measurement given the hidden parameters. Unfortunately, the prior distribution of the hidden parameters needs to be constructed heuristically, which is unreliable. As the major goal is to conduct robust inference for the hidden parameters (states) based on the labeled data (historical measurements-state pairs), one can inverted the causality relation for a Discriminative model. Such a model is proposed purely for inference tasks, rather than to model some underlying reality. This is similar in spirit to curve fitting methods, such as polynomial regression. This type of reasoning, where we combine inductive (model choice) and deductive (Bayes inference) reasoning into one step has been called Transductive reasoning.

1) Kernel ridge regression:

- Ridge regression

We first consider a Normal model below, which is a popular discriminative model with unknown hyper-parameters $\boldsymbol{q}$ and Σ_d:

$$\boldsymbol{x}|\boldsymbol{z} : \boldsymbol{N}(\boldsymbol{q}^T\boldsymbol{z}, \boldsymbol{\Sigma}_d). \tag{8}$$

To identify such a discriminative model for our inference, a regularized (ridge regression) estimator is commonly used:

$$\hat{\boldsymbol{q}} = \arg\min_{\boldsymbol{q}} \sum_{i=1}^{K} (\boldsymbol{x}_i - \boldsymbol{q}^T\boldsymbol{z}_i)^2 + 2\gamma||\boldsymbol{q}||^2, \tag{9}$$

where $\sum_{i=1}^{K}(\boldsymbol{x}_i - \boldsymbol{q}^T\boldsymbol{z}_i)^2$ is used to minimizing the sum square error, and the regularization term $2\gamma||\boldsymbol{q}||^2$ is used to improve the estimator performance for ill-conditioned problems.

For the Normal model, with the historical data stored in $\boldsymbol{Z}_{\rm mat}$ and $\boldsymbol{X}_{\rm mat}$ as follows,

$$\begin{aligned} \boldsymbol{Z}_{\rm mat} &= (\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_K), \\ \boldsymbol{X}_{\rm mat}^T &= (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_K), \end{aligned} \tag{10}$$

we can obtain a closed-form solution:

$$\hat{\boldsymbol{q}} = (\boldsymbol{Z}_{\rm mat}\boldsymbol{Z}_{\rm mat}^T + 2\gamma\boldsymbol{I})^{-1}\boldsymbol{Z}_{mat}\boldsymbol{X}_{\rm mat}. \tag{11}$$

where the unknown hyper-parameter Σ_d has been absorbed into the penalty constant γ. Notice that due to the ridge regularization (since $\gamma > 0$), $\boldsymbol{Z}_{\rm mat}\boldsymbol{Z}_{\rm mat}^T + 2\gamma\boldsymbol{I} \succeq 2\gamma\boldsymbol{I} \succ 0$. Therefore, the matrix $\boldsymbol{Z}_{\rm mat}\boldsymbol{Z}_{\rm mat}^T + 2\gamma\boldsymbol{I}$ is always invertible. Thus, the regularized estimator in (11) always exists. In other words, with the choice of quadratic penalty, the ridge regression solution is again a linear function of the labels (states) in $\boldsymbol{X}_{\rm mat}$. The solution adds a positive constant to the diagonal of $\boldsymbol{Z}_{\rm mat}\boldsymbol{Z}_{\rm mat}^T$ before inversion. This makes the problem nonsingular, even if $\boldsymbol{Z}_{\rm mat}\boldsymbol{Z}_{\rm mat}^T$ is not of full rank, and was the main motivation for ridge regression when it was first introduced in statistics.

Once the hyper-parameter $\hat{\boldsymbol{q}}$ is estimated in (11), it can be used for Bayesian inference to generate the current state estimate $\hat{\boldsymbol{x}}^B$, as follows:

$$\begin{aligned} \hat{\boldsymbol{x}}_{\rm current}^B &= \hat{\boldsymbol{q}}^T\boldsymbol{z}_{\rm current} && (12)\\ &= \boldsymbol{X}_{\rm mat}^T\boldsymbol{Z}_{\rm mat}^T(\boldsymbol{Z}_{\rm mat}\boldsymbol{Z}_{\rm mat}^T + 2\gamma\boldsymbol{I})^{-1}\boldsymbol{z}_{\rm current} && (13)\\ &= \boldsymbol{X}_{\rm mat}^T\boldsymbol{T} && (14) \end{aligned}$$

Notice that, such a form is based on $\boldsymbol{Z}_{\rm mat}\boldsymbol{Z}_{\rm mat}^T$. One can also conduct the following derivation to obtain an alternative form based on $\boldsymbol{Z}_{\rm mat}^T\boldsymbol{Z}_{\rm mat}$. By employing the Matrix Inversion Lemma $(A+BDC)^{-1} = A^{-1} - A^{-1}B(D^{-1}+CA^{-1}B)CA^{-1}$ to expand the inversion above, the alternative form of $\boldsymbol{T}$ can be obtained, which may further simplify the computational need:

$$\begin{aligned} \boldsymbol{T} =& \boldsymbol{Z}_{\rm mat}^T(\boldsymbol{Z}_{\rm mat}\boldsymbol{Z}_{\rm mat}^T + 2\gamma\boldsymbol{I})^{-1}\boldsymbol{z}_{\rm current} && (15)\\ =& \frac{1}{2\gamma}(\boldsymbol{Z}_{\rm mat}^T\boldsymbol{z}_{\rm current} - \boldsymbol{Z}_{\rm mat}^T\boldsymbol{Z}_{\rm mat}(\boldsymbol{Z}_{\rm mat}^T\boldsymbol{Z}_{\rm mat} \\ & + 2\gamma\boldsymbol{I})^{-1}\boldsymbol{Z}_{\rm mat}^T\boldsymbol{z}_{\rm current}) && (16)\\ =& \frac{1}{2\gamma}(\boldsymbol{Z}_{\rm mat}\boldsymbol{z}_{\rm current}^T - (\boldsymbol{Z}_{\rm mat}\boldsymbol{Z}_{\rm mat}^T + 2\gamma\boldsymbol{I} - 2\gamma\boldsymbol{I}) \\ & \cdot(\boldsymbol{Z}_{\rm mat}\boldsymbol{Z}_{\rm mat}^T + 2\gamma\boldsymbol{I})^{-1}\boldsymbol{Z}_{\rm mat}^T\boldsymbol{z}_{\rm current}) && (17)\\ =& (\boldsymbol{Z}_{\rm mat}^T\boldsymbol{Z}_{\rm mat} + 2\gamma\boldsymbol{I})^{-1}\boldsymbol{Z}_{mat}^T\boldsymbol{z}_{\rm current}, && (18) \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{Z}_{\rm mat}^T\boldsymbol{Z}_{\rm mat} &= (\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_n)^T(\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_n) \\ &= \begin{pmatrix} \boldsymbol{z}_1^T\boldsymbol{z}_1 & \cdots & \boldsymbol{z}_1^T\boldsymbol{z}_n \\ \vdots & \ddots & \vdots \\ \boldsymbol{z}_n^T\boldsymbol{z}_1 & \cdots & \boldsymbol{z}_n^T\boldsymbol{z}_n \end{pmatrix}, \end{aligned} \tag{19}$$

$$\begin{aligned} \boldsymbol{Z}_{\rm mat}^T\boldsymbol{z}_{\rm current} &= (\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_n)^T\boldsymbol{z}_{current} \\ &= \begin{pmatrix} \boldsymbol{z}_1^T\boldsymbol{z}_{current} \\ \vdots \\ \boldsymbol{z}_n^T\boldsymbol{z}_{current} \end{pmatrix}. \end{aligned} \tag{20}$$

Because the matrix $\boldsymbol{Z}_{\rm mat}^T\boldsymbol{Z}_{\rm mat}$ appears in the calculation (18), as opposed to the original calculation (15) involving $\boldsymbol{Z}_{\rm mat}\boldsymbol{Z}_{\rm mat}^T$, the pairwise inner product in $\boldsymbol{Z}_{\rm mat}^T\boldsymbol{Z}_{\rm mat}$ creates the potential to improve estimation performance in the nonlinear kernel space. This is because power system's measurement functions are usually nonlinear. Direct use of (8) and (12) implicitly assumes linear model, leading to a relatively poor state estimate.

- The Kernel trick for Normal Discriminative model

Kernels are important building blocks for high-dimensional learning techniques. There is a trick called kernelization for improving a computationally simple classifier/regressor [18]. The idea is to map the covariate $\boldsymbol{z}_i^T\boldsymbol{z}_j$ into a higher dimensional space and apply the regression in the bigger space. This can yield a more flexible estimator while retaining computational simplicity. The Point is that to get a richer set of regression models we do not need to give up the convenience of linear

regression model. We simply map the covariates to a higher-dimensional space. This is akin to making linear regression more flexible by using polynomials.

By kernel trick, there exists a high-dimensional mapping $\boldsymbol{u}_i = w(\boldsymbol{z}_i)$, from which the inner product $\boldsymbol{u}_i^T \boldsymbol{u}_j = (w(\boldsymbol{z}_i))^T w(\boldsymbol{z}_j)$ can be calculated by a kernel $K(\cdot, \cdot)$, as below,

$$\boldsymbol{u}_i^T \boldsymbol{u}_j = K(\boldsymbol{z}_i, \boldsymbol{z}_j). \tag{21}$$

Therefore, the kernel calculation uses only the (low-dimensional) $\boldsymbol{z}$'s, rather than the high-dimensional $\boldsymbol{u}$'s. Therefore, the computational complexity of calculating the inner products in (19) and (20) is low, even though $\dim(\boldsymbol{u})$ itself may be very large. This idea of using a cost-effective kernel calculation to implement a high-dimensional Normal model is called 'the kernel trick'. In this paper, we employ the following kernel forms as candidates. This process is called kernel model assessment and selection.

- Homogeneous polynomial: $K(\boldsymbol{u}_i, \boldsymbol{u}_j) = (\boldsymbol{u}_i^T \boldsymbol{u}_j)^d$.
- Inhomogeneous polynomial: $K(\boldsymbol{u}_i, \boldsymbol{u}_j) = (1 + \boldsymbol{u}_i^T \boldsymbol{u}_j)^d$.
- Gaussian (Radial Basis function): $K(\boldsymbol{u}_i, \boldsymbol{u}_j) = \exp(-\mu ||\boldsymbol{u}_i^T \boldsymbol{u}_j||^2),\ \mu > 0$.

2) Model Selection: In order to choose the best model, we need to assess the performance of various models based on different γs in (9) and kernels above.

If we are in data-rich situation, the best approach for both problems is to randomly divide the data-set into three parts: a training set, a validation set, and a test set. The training set is used to fit the models; the validation set is used to estimate prediction error for model selection; the test set is used for assessment of the generalization error of the final chosen model. Ideally, the test set should be kept in a "vault", and be brought out only at the end of the data analysis. In the Sec.V, we use inconsecutive data between validation and testing phases for this purpose.

Therefore, we divide the data into three phases.

- In the training phase, one applies part of the historical data on different kernel function and γ pairs to calculate different Ts.
- In the validating phase, another part of historical data are used to choose the best kernel function and γ.
- Finally, the chosen $\hat{\boldsymbol{x}}^B$, computed from the validated T, is used in (3) for the testing phase for state estimation.

IV. Speed Up for Data Driven SE

In the proposed data-driven approach described in the last section, nearest neighbors search requires exhaustive exploration of all data points and is slow in large network with huge historical data. In the following we analyze the power system historical data structure and propose two steps to speed up the similar data search process but preserve the accuracy.

A. Dimension Reduction

In order to reduce NN search time, we start by exploring the data structure. As the electrical power systems exhibit periodicity, one would expect the measurement data to be highly clustered, creating the possibility for great dimension reduction for measurement data. As an illustration, singular value decomposition (SVD) is conducted over the historical data of 300 bus systems [32]. 1073 measurements per time slot over one year are used to form historical measurement matrix $Z = [\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_Q]$. Mathematically, the SVD decomposition is represented as

$$Z = U \times S \times V', \tag{22}$$

where the diagonal entries of S are known as the singular values. U and V are unitary matrices [32]. By plotting the magnitude of singular values in Z with a $\log - \log$ scale, only 8 significant singular values show up in Fig.4a. Besides, in Fig.4b, we show the result of mapping the historical data onto some two dimensional features (two left-singular vectors of U) associated with significant singular values. As the data in the figure are highly clustered and far away from each other, dimension reduction is possible for historical electric power system data to remove redundancy in NN search.

In this work, we propose to use random projection for this highly clustered historical measurement data set. This is because other dimension reduction techniques such as SVD tend to be very time consuming, making them unsuited for online state estimation process [33]. Besides, dimension reduction of these techniques is typically a one-time operation, which means that the entire process has to be done every time new power system data come up, making them non-adaptive to the new coming data, which is important for real time power system analysis.

In contrary, random projection is fast and adaptive. Mathematically, the original m-dimensional measurement data is simply projected onto a m'-dimensional ($m' \ll m$) subspace using a random $m' \times m$ matrix R,

$$\boldsymbol{y}_{m' \times 1} = R_{m' \times m} \boldsymbol{z}_{m \times 1}. \tag{23}$$

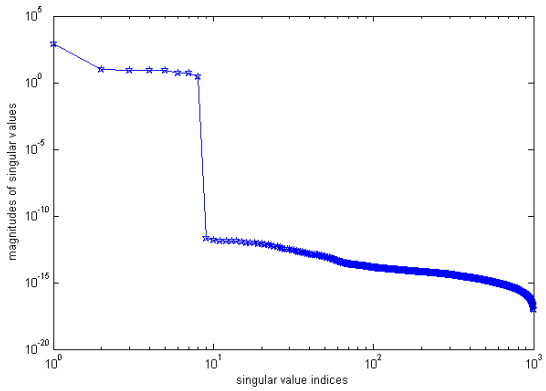
The elements r_{ij} of the random matrix R are often chosen to be normally independent and identically distributed with zero means, and the columns of R are normalized with unit lengths.

The key idea of such a random mapping arises from the Johnson-Lindenstrauss Lemma [34]: if points in a vector space are projected onto a randomly selected subspace of suitably high dimensions, then the distance between the points is approximately preserved.

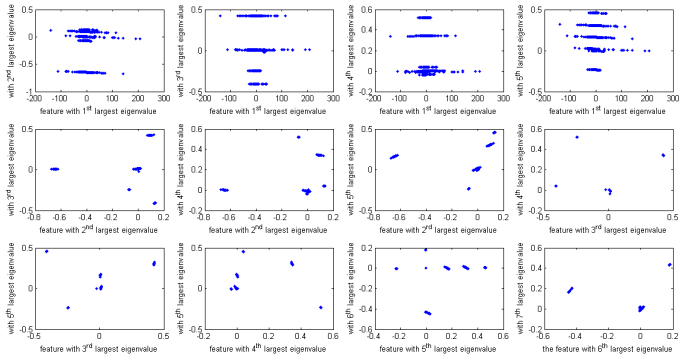
B. K-dimensional (k-d) Tree for Indexing

As the historical data is highly clustered, we propose to use a tree to index the data after dimension reduction. The basic idea of the proposed k-d tree approach comes from the binary-tree as illustrated on the left of Fig.5. In the Binary-tree, all nodes after the left pointer have smaller values than the current root value, and all nodes after the right pointer have bigger values. If one wants to search for a number in this seven-node tree, the maximum searching time is changing from 7 (by exhaustive search) to 3 by using the tree structure.

To extend the one dimensional data to high dimensional data such as the power system measurement vectors, we substitute a binary-tree with a k-d tree as illustrated in Fig.5. k-d tree



(a) Singular values



(b) Measurement data clustering

Fig. 4: Dimension reduction

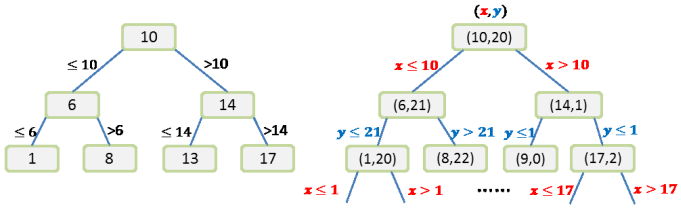


Fig. 5: Binary tree and k-d tree

alternates over different measurement as a discriminator on every level of the tree. Because of the efficient 'pruning' of the search space, the k-d tree has an average nearest neighbor search time of $O(\log(Q))$, where Q is the total number of historical data points.

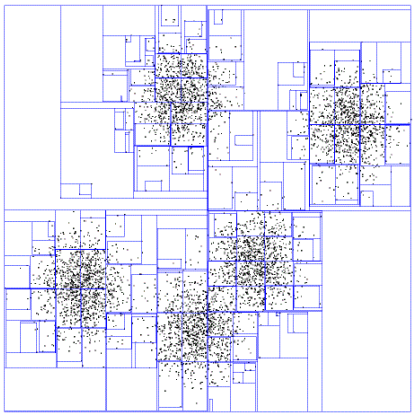


Fig. 6: K-d tree for clustered data set

As an illustration, Fig.6 shows a case where k-d tree is conducted over two-dimensional data set. By properly using the clustering properties of the spatial data points, the k-d tree achieves good performance as the search within it can omit large portion of the clustered points in the space.

C. Adaptation to the robust K-NNs Approach

The speed up method introduced in this section is based on the historical measurements. However, our robust K-NNs approach in Sec.III-A is based on the pseudo-measurements in (6). To adapt the algorithm, we will first chose $10 \times K$ Nearest Neighbors based on the historical measurements. Then we will reduce neighbor numbers via the robust K-NNs approach discussed in section III.

V. Numerical Result

In this section, we simulate and verify the performance of the proposed robust K-NNs regression approach, and compare it to the industrially used Newton's method in the standard IEEE 300 test case.

A. Simulation Method

1) Data Preparation: Such simulations are completed in MATLAB environment in accordance with MATLAB Power System Simulation Package (MATPOWER) [26], [27]. Further, to simulate the power system behavior in a more practical pattern, online load profile from New York ISO [35] is adopted in the subsequent simulation. Specifically, it has online load profiles in New York ISO area recorded every five minutes. The load data used is between February 2005 and December 2013 with a consistent data format. Therefore, we use load data between February 2005 and May 2013 in training and validation sessions. The load profiles between July 2013 and December 2013 are used in the testing session.

To generate data for SE, we first fit the normalized load data into the case file. Subsequently, an AC power flow is run to generate the true states of the power system, followed by creating true measurement sets with Gaussian noises (standard deviations in Table I). Hereby, we assume that the measurement set includes 1) power injection; 2) line power flows; 3) voltage magnitudes; 4) some phase angle measurements.

TABLE I: Standard Deviation of Measurement Noise

Measurement type	Standard deviation
Active (Reactive) power injection	0.015
Active (Reactive) power flow (from)(to)	0.02
Voltage magnitude	0.01
Phase angle	0.002

2) Data Adjustment: Since we propose a data-driven SE that is claimed to be robust to topology changes, bad data, and malicious attack, we will adjust the generated data above for simulation with respect to each of them.

- Topology Changes: In this case, before running the power flow to generate measurements, several randomly chosen topology connections are changed with probability 20% to imitate certain feature of the smart grid.
- Bad Data: Beside topology changes, we will randomly generate bad data and insert them into the measurements.
- Malicious Attack: Beside topology changes and bad data, we intentionally inject several malicious data in the historical database before May 2013, which is unobservable by Chi-square test [22].

3) Training, Validation, and Testing:

- **Training Phase**: By randomly selecting one measurement between July 2013 and December 2013 as a test case, a group of robust nearest neighbor measurements in (7) between Feb. 2005 and Dec. 2012 is selected via Fig.2 in Sec.III-A. Then the matrix T in (18) is computed for different choices of γ and kernel function.
- **Validation Phase**: The matrix T is validated on the data between January 2013 and May 2013 to obtain the best choice for γ and kernel pair.
- **Testing Phase**: We use the matrix T chosen in the validating phase to calculate Bayesian state estimate $\boldsymbol{x}^B$ via (14) in the data between July 2013 and December 2013. For comparison purpose, the industrial approach of Newton's method initialized via the last state estimate is also applied to the same testing data.

B. Improved Accuracy

In the testing phase, filtering out bad data and malicious data reduces measurement number. Therefore, instead of Sum Square Error, we will employ Mean Squared Error (MSE) defined as

$$\text{MSE} = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{z_i - h_i(\hat{x}_{\text{current}})}{\sigma_i}\right)^2. \tag{24}$$

1) Robustness to Topology Changes: In this part, we assume there is no bad data and malicious attack in the historical data to emphasis the robustness feature of the proposed method. To deal with topology changes, we employ the first building block only (left block) (Fig.2) to conduct R-KNNs search. Then, the gathered data is outputted to the kernel ridge regression method discussed in Sec.III-B1 for an estimate. Training, Validation and Testing procedure, discussed in Sec.III-B2, are subsequently used in this estimation process. For fairness, testing is conducted for 400 times. Each time, we obtain a MSE, voltage magnitude estimates, and voltage phase angle estimates. For comparison purpose, we also compute the corresponding results with Newton's method initialized by a previous state estimate.

To show the improved accuracy over 400 testing cases, we define relative error in the i^{th} testing case as

$$\gamma_i = \frac{MSE_{\text{RobustK-NearstNeighborsMethod}}}{MSE_{\text{Newton'sMethodwithPre.Est.Start}}}. \tag{25}$$

Subsequently, Fig.7a shows the histogram of 400 simulation results. By looking at the x coordinate, we observe that the Robust K-Nearest Neighbors (R-KNNs) approach has greatly reduced estimation errors with an average ratio of $10^{-2.5}$. From this fact, we can reasonably conclude that the proposed R-KNNs method is able to handle topology changes.

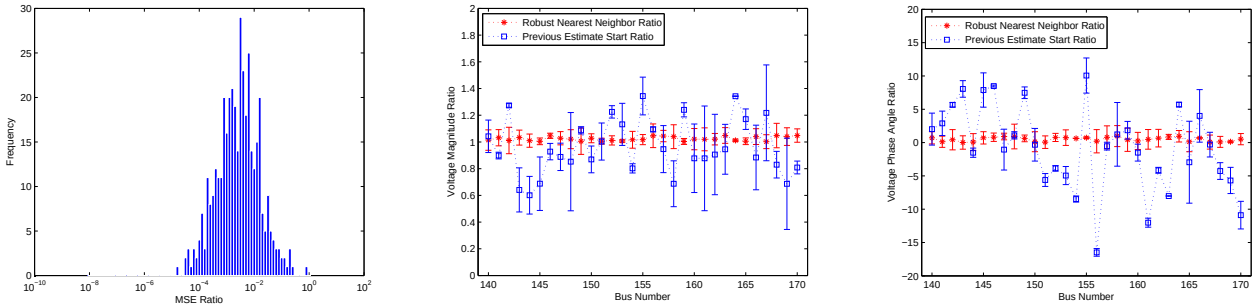
Further to the comparison in the MSE domain, Fig.7b and Fig.7c provide state domain plots. The x coordinate represents the bus number. The y coordinate represents the voltage magnitude ratio ($\frac{|V_{R-KNNs}|}{|V_{True}|}$ and $\frac{|V_{Pre.Est.}|}{|V_{True}|}$) in Fig.7b and the voltage phase angle ratio ($\frac{\angle V_{R-KNNs}}{\angle V_{True}}$ and $\frac{\angle V_{Pre.Est.}}{\angle V_{True}}$) in Fig.7c, respectively. It can be observed that the R-KNNs method has a voltage ratio in red close to 1, and its less variance (in red) indicates its ability to track the true system states while Newton's method with a previous estimate start has a ratio (in blue) far away to 1 with large variance. Such a poor result is caused by the local-search behavior of Newton's method, which is suboptimal with an inferior initial guess.

2) Robustness to Bad Data: In this section, we conduct similar simulations as the last subsection. However, besides using the first building block (left block of Fig.2), we will also use the second (middle) building block of Fig.2. This block targets at improving the result, by filtering out data points with relative large residuals, by creating a soft decision rule as discussed in Sec.III-A2.

For comparison purpose, we simulate this part in parallel to the last subsection, namely the robustness to topology changes. Therefore, we observe the same estimation mean and variance in blue for the Newton's method initialized by the previous state estimate. As one can observe, the only difference between Fig.8b and 7b are that the mean of R-KNNs ratio in Fig.8b is flatter and its variance is smaller than Fig.7b. This is caused by filtering out measurement with relative large residuals, thanks to richer data comparison across horizontal time line, instead of a single data point used in the traditional static SE. As another evidence to this observation, Fig.8a shows smaller MSE ratio than the one in Fig.7a.

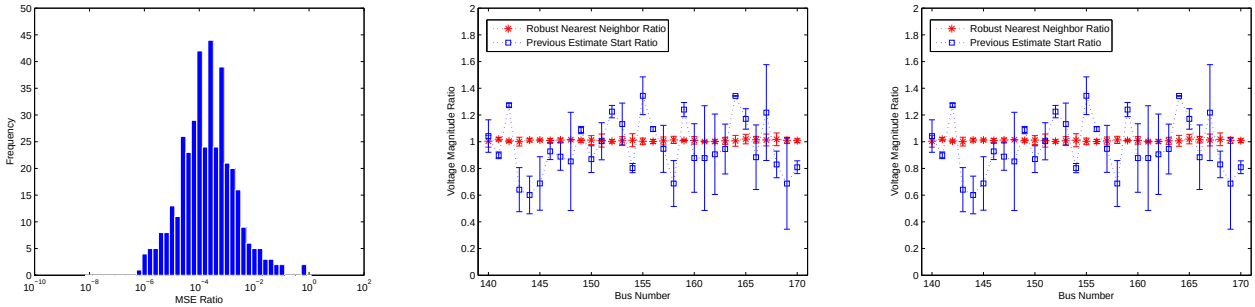
3) Robustness to Malicious Attack: The data used in this subsection are slightly different than the two simulations above. To mimic that at some time instance that human beings try to inject bad data that can pass the chi-square test, we injected malicious data intentionally into the measurement set only in the testing sets. From Fig.9a, we can see that the Mean Square Error before or after malicious data filtering out are not so different. However, we can observe large errors in the state space in Fig.9b and Fig.9c. For example, the rectangular plot in blue shows a long distance to 1 in Fig.9b. Similar observation is made in Fig.9c as well. This shows that the malicious data can change the state estimate dramatically without triggering the bad data alarm.

In contrast, the red star like curve is close to one in the voltage magnitudes plotting and close to zero in the voltage magnitudes plotting. Besides, the variance in red is much smaller than the variance in blue, illustrating the robustness of the proposed data drive approach against Malicious data.



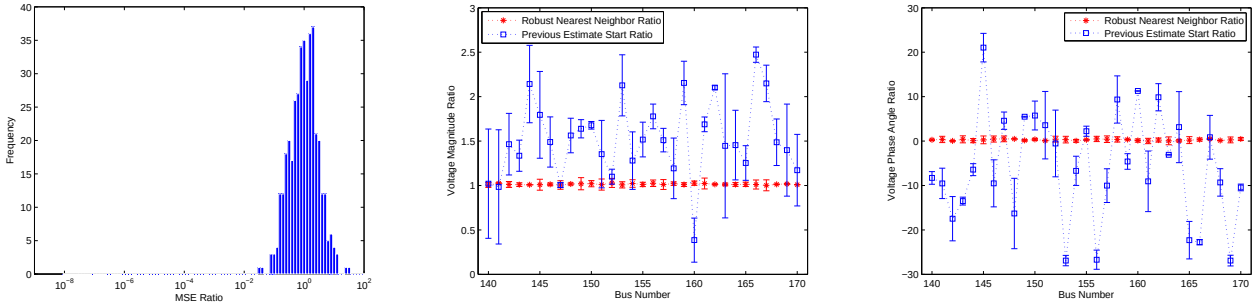
(a) Histogram of MSE ratio (b) Voltage magnitudes (c) Voltage phase angles

Fig. 7: Simulation results with only the first block of Fig.2.



(a) Histogram of MSE ratio (b) Voltage magnitudes (c) Voltage phase angles

Fig. 8: Simulation results with the first and the second blocks of Fig.2



(a) Histogram of MSE ratio (b) Voltage magnitudes (c) Voltage phase Angles

Fig. 9: Simulation results with all three blocks of Fig.2

C. Speed Up

In this section the speed up method proposed in Sec.IV is used. The comparison process is conducted for 400 times with and without speedup.

1) Time Reduction: To illustrate the speedup, we define the relative execution time ϕ for the i^{th} testing data during the R-KNNs search:

$$\phi_i = \frac{t_{wp,i}}{t_{ex,i}}, \tag{26}$$

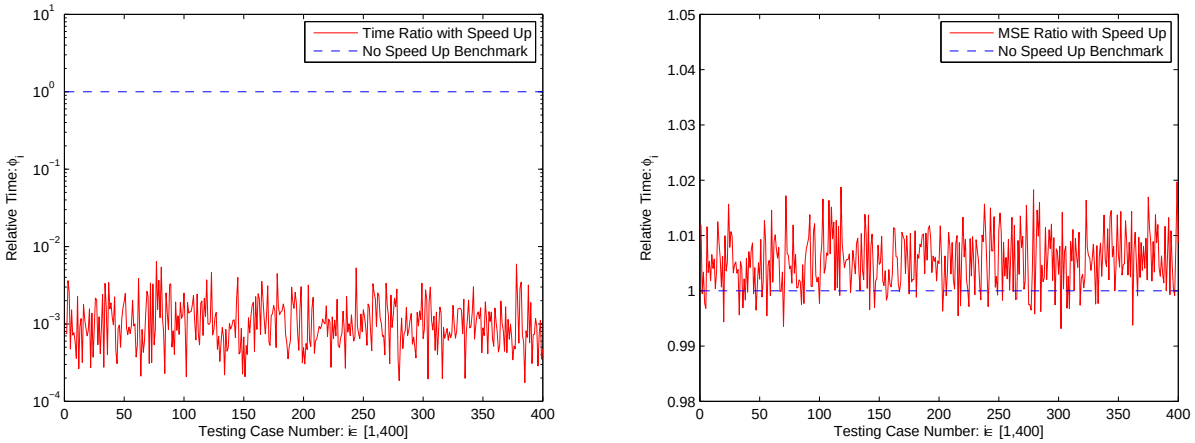
where t_{wp} represents the R-KNNs search time with preprocessing (dimension reduction and k-d tree indexing). t_{ex} represents the exhausted R-KNNs search time without preprocessing.

The plot of ϕ_i is drawn in Fig.10a. It shows that the proposed preprocessing approach has greatly reduced historical NN search time. For example, the average ratio (ϕ) for all testing cases is around 10^{-3}, creating 99.9% reduction in the search time. Such a result leads to a rational interpretation for the proposed procedure: since the historical data is organized in a compact way, the nearest neighbors search can be conducted much more efficiently.

2) Approximately the Same Accuracy: Similar to the ratio metric defined in the last subsection, We define the i^{th} relative error π_i for testing data as

$$\pi_i = \frac{MSE_{\text{wp,i}}}{MSE_{\text{ex,i}}}. \tag{27}$$

The plot of π_i shows that the relative error is only slightly larger than 1 (less than 0.01 on average), so the proposed two-step approach returns a group of historical data highly similar to the ones in exhaustive search. The slightly larger



(a) Histogram of improvements.

(b) Histogram of improvements.

Fig. 10: Simulation results for speed up

error ($\approx 1\%$) may be caused by dimension reduction. But it is negligible when compared to the 1000 times speedup. Therefore the proposed approach achieves approximately the same accuracy as the slow but highly accurate data-driven SE as shown in Fig.7, Fig.8, and Fig.9.

VI. Conclusions

In this paper, we discuss how to systematically obtain a robust data-driven state estimation for AC power systems. Based on the intuition that similar measurements and topology reflect similar power system states, we formulate the finding of the initial SE as a minimum distance search problem. Further, a Bayesian estimate is obtained via kernel ridge regression. We further propose how to systematically reduce the computational cost for the proposed method. It is based on the observation that power system consumption features periodic pattern. In particular, dimension reduction and efficient indexing over trees are proposed. Numerical results show that the proposed method can achieve a highly accurate robust data-driven state estimate in a short time with 1000 times speedup over IEEE benchmark systems.

Acknowledgment

This work was supported in part by US NSF awards 0931978, 0831973 and 0347455. The authors would also like to thank ABB for supporting this work.

References

[1] A. Wood and B. Wollenberg, *Power generation, operation, and control*, 2nd ed., 1996.

[2] F. C. Schweppe, J. Wildes, and D. B. Rom, "Power system static state estimation, parts 1, 2, 3." *IEEE Transactions on Power Apparatus and Systems*, p. 120, Jan. 1970.

[3] A. Abur and A. G. Exposito, "Power system state estimation: Theory and implementation," *CRC Press*, Mar. 2004.

[4] E. Handschin, F. C. Schweppe, J. Kohlas, and A. Fiechter, "Bad data analysis for power system state estimation," *IEEE Transactions on Power Apparatus and Systems*, p. 329, Apr. 1975.

[5] L. Mili, M. G. Cheniae, N. S. Vichare, and P.J.Rousseeuw, "Algorithm for least median of squares estimation of power systems," *Proc. 35-th Midwest Symp. Circuit and Systems*, pp. 1276–1283, Aug. 1992.

[6] S. Gastoni, G. P. Granelli, and M. Montagna., "Robust state estimation procedure based on the maximum agreement between measurements," *IEEE Transactions on Power Systems*, p. 2038, Nov. 2004.

[7] Y. Weng, R. Negi, Q. Liu, and M. D. Ilić, "Robust state-estimation procedure using a least trimmed squares pre-processor," *IEEE Innovative Smart Grid Technologies*, pp. 1–6, Jan. 2011.

[8] H. Zhu and G. B. Giannakis, "Estimating the state of ac power systems using semidefinite programming," *Proceedings 43rd North America Power Symposium (NAPS)*, pp. 1–7, Aug. 2011.

[9] Y. Weng, Q. Li, R. Negi, and M. D. Ilić, "Semidefinite programming for power system state estimation," *IEEE Power and Energy Society General Meeting*, Jul. 2012.

[10] Y. Weng, Q. Li, M. Ilic, and R. Negi, "Distributed algorithm for sdp state estimation," *IEEE Innovative Smart Grid Technology Conference*, Aug. 2013.

[11] F. F. Wu, "Power system state estimation: A survey," *International Journal of Electrical and Power Engineering*, vol. 12, pp. 80–87, 1990.

[12] A. Monticelli, "The impact of modeling short circuit branches in state estimation," *IEEE Transactions on Power Systems*, vol. 8, no. 1, pp. 364–370, Feb. 1993.

[13] A. G. Exposito, A. Abur, A. V. Jaen, and C. G. Quiles, "A multilevel state estimation paradigm for smart grids," *Proceedings of the IEEE*, p. 952, Jun. 2011.

[14] B. V. Tuykom, J. C. Maun, and A. Abur, "Use of phasor measurements and tuned weights for unbalanced system state estimation," *North American Power Symposium (NAPS)*, p. 1, Sep. 2010.

[15] A. P. S. Meliopoulos, B. Fardanesh, and S. Zelingher, *Power system state estimation: modeling error effects and impact on system operation*, Jan. 2001.

[16] H. Wu and J. Giri, "Pmu impact on the state estimation reliability for improved grid security," *IEEE Power Energy Society Transmission and Distribution Conference and Exhibition*, p. 1349, Mar. 2005.

[17] Y. Weng, R. Negi, and M. Ilic, "A search method for obtaining initial guesses for smart grid state estimation," *IEEE SmartGridComm Symposium*, Nov. 2012.

[18] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, Feb. 2009.

[19] Y. Weng, M. D. Ilić, Q. Li, and R. Negi, "Convex relaxation-based state estimators in electric power systems c part i: Theoretical formulation," *IEEE Transaction on Power Systems*, 2014.

[20] ——, "Convex relaxation-based state estimators in electric power systems c part ii: Distributed implementation," *IEEE Transaction on Power Systems*, 2014.

[21] L. Mili, T. V. Cutsem, and M. R. Pavella, "Hypothesis testing identification: A new method for bad data analysis in power system state estimation," *IEEE Transactions on Power Apparatus and Systems*, p. 3239, Nov. 1984.

[22] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Conference on Computer and Communications Security*, pp. 21–32, Nov. 2009.

[23] M. Ilic, "Data-driven sustainable energy systems," *The 8th Annual Carnegie Mellon Conference on the Elctricity industry*, Mar. 2012.

[24] S. Kaski, "Dimensionality reduction by random mapping: fast similarity computation for clustering," *Proc. IEEE International Joint Conference on Neural Networks Proceedings*, vol. 1, pp. 413–418, 1998.

[25] E. Bingham and H. Mannila, "Random projection in dimensionality reduction: Applications to image and text data," *The 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 245–250, 2001.

[26] R. D. Zimmerman, C. E. Murillo-Sanchez, , and R. J. Thomas, "Matpower's extensible optimal power flow architecture," *IEEE Power and Energy Society General Meeting*, pp. 1–7, Jul. 2009.

[27] R. D. Zimmerman and C. E. Murillo-Sanchez, "Matpower, a matlab power system simulation package," *http://www.pserc.cornell.edu/ matpower/manual.pdf*, Jul. 2010.

[28] D. Hawkins, "Identification of outliers," *Chapman and Hall, London*, 1981.

[29] M. M. Breunig, H. P. Kriegel, R. T. Ng, and J. Sander, "Lof: Identifying density-based local outliers," *Proc. ACM SIGMOD Conf. 2000*, pp. 93–104, 2000.

[30] E. M. Knorr and R. T. Ng, "Algorithms for mining distance based outliers in large datasets," *Proceeding 24th Int. Conf. on Very Large Data Bases*, pp. 392–403, 1998.

[31] L. Wasserman, "All of nonparametric statistics," *Springer*, 2007.

[32] G. Strang, "Introduction to linear algebra (section 6.7)," *Wellesley-Cambridge Press*, 1998.

[33] M. Sahlgren, "An introduction to random indexing," *in Methods and Applications of Semantic Indexing Workshop at the 7th International Conference on Terminology and Knowledge Engineering, TKE*, 2005.

[34] W. B. Johnson and J. Lindenstrauss, "Extensions of lipshitz mapping into hilbert space," *American Mathematical Society Conference in Modern Analysis and Probability, Contemporary Mathematics*.

[35] NYISO, "Load data profile," *http://www.nyiso.com*, May. 2012.