State Estimation for Future Energy Grids

Shervin Mehryar, Member, IEEE, Moe Z. Win, Fellow, IEEE

Laboratory for Information and Decision Systems (LIDS)
Massachusetts Institute of Technology (MIT)
77 Massachusetts Avenue, Room 32-D674A
Cambridge, MA 02139 USA
Tel.: (617) 253-9341
e-mail: moewin@mit.edu

Abstract—Today’s power generation and distribution networks are quickly moving toward automated control and integration of renewable resources - a complex, integrated system termed the Smart Grid. A key component in planning and managing of Smart Grids is State Estimation (SE). The state-of-the-art SE technologies today operate on the basis of slow varying dynamics of the current network and make simplifying linearity assumptions. However, the integration of smart readers and green resources will result in significant non-linearity and unpredictability in the network. Therefore in future Smart Grids, there is need for ever more accurate and real-time algorithms. In this work, we propose and examine new SE methods that aim to achieve these measures by approximating the true distribution of the state variables, rather than a linearized version as done for instance in Kalman filtering. Through simulations we show that in the presence of non-linearities and non-Gaussian noise, our general SE framework improves accuracy where linear and Kalman-like filters exhibit impaired performance.

I. INTRODUCTION

Operation, monitoring, and control of power grids rely on the state estimation (SE) problem, defined as the computation of line voltages and phase angles under steady-state conditions [1]. Estimated line parameters are utilized by supervisory entities to ensure safe-mode operation, efficient resource allocation, and low-cost power generation, among others. While traditionally energy networks are assumed to have slow and modest behavior, in future grids real-time dynamic SE is essential. Above all, for an expanding system that will heavily rely on estimations at every level, accuracy becomes paramount with minimal tolerance for error. This work focuses on state estimation methods that provide reliable and efficient performance with high scalability and low complexity cost for future grids.

The power grid of today has a 3-layer structure: 1) power is generated at Generation Plants, 2) power is transferred through transmission line (TL), 3) power is delivered to end users at Distribution Level. Figure 1 shows a diagram of the power flow through such a network. The Smart Grid in this context refers to all aspects of the power network that allow for smooth and seamless generation and distribution of power. It is, according to [2], [5], an automated, widely distributed energy delivery network, that will be characterized by a two-way flow of electricity and information and will be capable of monitoring everything from power plants to customer preferences and individual appliances. It incorporates into the grid the benefits of distributed computing and communications to deliver real-time information and enable the near-instantaneous balance of supply and demand at the device level. As opposed to a top-down flow, the Smart Grid will be highly integrated and dynamic.

Traditionally, SE only exists in static form [2], [4]–[7]. This is mainly due to the fact that current power generation is facilitated primarily by hydro (e.g. water powered) and steam (e.g. coal or gas fueled) turbines that follow slow-varying and predictable patterns. However, with the advent of renewable sources such as wind and solar panels that are more irregular, the exigence in dynamic estimation becomes apparent.

In addition, traditional SE focuses on TL level estimation, with SE at Distribution level happening very infrequently [2]. At the time, power meters that provide power and voltage magnitude readings on high-voltage lines, are the sole data acquisition source. A limiting factor here is the difficulty in deploying measurement units given the size of the network. Today however, affordable smart meters and phase measurement units (PMU) are available that allow for more distributed (at households) and content-rich data collection. Particularly with the emergence of green sources as key contributors to future grids, power generation no longer will follow a vertical flow from plant generation down to household consumption. This shift in generation trend puts distributed level SE on par with TL level SE.

Table 1 summarizes the current state of the grid versus the changes that are going to be seen in future energy networks.

1Image courtesy of: www.smartgridlegalnews.com/smart-grid-basics/the-electric-grid-101.
TABLE I

| Traditional Grids | Future Grids |
|-------------------|--------------|
| Vertical Structure | Integrated Structure |
| Static Estimation | Dynamic Estimation |
| Transmission Level SE | Distributed SE |

In view of these transformations, SE algorithms must be able to deal with highly dynamic and heterogeneous networks and operate reliably at Distributed Levels as well.

Optimal filtering methods, as they are often called, exist for finite state, linear, and Gaussian problems, as well as a constrained sub-class of non-linear problems. However, in general, continuous problems with non-linear/non-Gaussian models are intractable and approximate solutions that are only locally optimal are in application.

A. 4 Key Design Considerations for Future Energy Networks

Based on the discussion building up to here, there are four major considerations that are key to ensuring flawless and reliable design of the future smart grids. Any SE algorithm must have these 4 issues into account and addressed them in order to meet future operational and practical requirements of the network. They are enumerated in what follows.

1) Large Network Scalability: One of the key considerations in designing algorithms for the Energy Networks of tomorrow is the scale of such networks. Considering the U.S.A. power grid alone, it consists of more than 9,200 electric generating units with more than 1,000,000 megawatts of generating capacity connected through more than 300,000 miles of TL. The network itself is subdivided into three main interconnections: Eastern, Western, and Texas. Each of which is built from countless number of electricity transmission components. For what is orders of magnitude larger than the Internet backbone in the number of building blocks, while at the same time lacking an inherent monitoring layer like the Internet network, the architecture and build of operation services becomes crucial. Given such a scale, highly distributed and precise estimation algorithms are of utmost necessity where taking measurements form every node in the network is unrealistic and a minor error can cascade through the Energy Network with catastrophic consequences.

2) Integrated and Dynamic Nature: One of the key goals in building the future Smart Grids according to the report by the U.S. Department of Energy is for the grid to be fully accommodating to both traditional and renewable sources of all sorts, where the power flow is no longer unidirectional from Generating Plants down to the Distribution Level. The Energy Network of tomorrow will be highly integrated and power injection can happen realistically at any point in the network, with smart PMU’s collecting and sending feedback to the supervisory control and data acquisition (SCADA) centers at rates about 30 samples/second. As a result the future network will be highly dynamic and their state can change in unpredictable manners. This makes the need for adaptive and fast-response estimation algorithms indisputable.

3) Non-linear System Immunity: A technical challenge in Energy Networks that is often overshadowed, is the presence of non-linearities at the low-level system. The network is built of active and reactive components that have complex form and beset network parameters with non-linear relations. For instance, the relation between voltage magnitudes and angles and the corresponding real and imaginary powers for a line are given by an equation of the form:

\[ P_{ij} = |V_i|^2 (g_{ij} + g_{ij}) \]
\[ Q_{ij} = |V_i|^2 (b_{ij} + b_{ij}) \]

Such non-linearities have irreprehensible precautions for estimation algorithms and can degrade performance if not properly dealt with. Most state-of-the-art methods unfortunately are designed only to work well disregarding such conditions.

4) Uncategorized System Noise: Lastly, a major hurdle to computing the exact line states is the uncertainty in the network. There is noise arising at all levels of the system, from measurement units and meters to data fusion at the control centers. Given the scale of the network, this predicament can escalate very rapidly. To add salt to the injury, the exact form of the system noise is not known. Therefore, estimation algorithms must have built-in resilience to uncategorized noise models or at the very least mitigate the effect of uncertainty through other intelligent ways.

B. Review of State of the Art

In view of the network considerations discussed above, many recent studies and developments have been conducted to realize the possibility of green, adaptive, integrated Smart Grids. At the state estimation level, there are a few key components in the system that have been subject to extensive research. In this section, a review of systematic shortcomings related to these components is presented and the progress in addressing them is discussed.

The problem of SE can be cast as a Hidden Markov Model (HMM) whereby first the state of the system is predicted based on previous states (Prediction Step) and next the predictions are corrected by taking into account the network measurements (Update Step). Mathematically, the problem can be described by the computation of a state vector \( x_t \) at time \( t \) from the previous state through a function \( g(\cdot) \) and given a set of measurements \( y_t \) through a function \( h(\cdot) \). For instance, the state vector can be line voltages and measurements can be a set of power, voltage, and current readings taken by a batch of PMU’s. The equations that govern these relations are:

\[ x_{t+1} = g(x_t) + q_t \]
\[ x_t = h(y_t) + n_t \]
Three points of active research in this field are:
i) There exists a lack of a physically and mathematically sound dynamic model. Current models assume smooth and slow state progression \([14] - [16]\). For instance, Holt’s Smoothing method is a widely used linear Dynamic Model given by:
\[
x_{t+1} = F_t x_t + g_t + q_t
\]
In the future, the addition of renewable energy inlets throughout the system and automated network controllers will make the system significantly more unpredictable and fluctuating. There is therefore a need for a model that meets the requirements of future Smart Grids in that it can support the Integrated and Dynamic nature of the system.

ii) The measurement equation \( h(\cdot) \) in Power Systems are highly non-linear and complex. Current methods, such as Kalman-like filters \([14] - [16]\), assume linearization which impairs the accuracy. For instance, line voltage and power relation discussed previously are often simplified as:
\[
P_n = \sum_{i \neq n} b_{ni}(\theta_n - \theta_i)
\]
There is a need for filtering methods that can promise non-linear system immunity.

iii) With increasing deployment of wireless smart readers and PMU’s, more complicated uncertainty modeling is needed. Typically measurement noise is assumed to be additive Gaussian in all current literature, in the form:
\[
n_t \sim N(0, \Sigma_n)
\]
Since the true nature of the noise is not known yet, there is a need for better models that can capture more complicated, uncharted system uncertainty.

The goal here is to develop estimation algorithms customized to operate in accordance with the requirements of future power networks. In view of the discussions above, this work is after algorithms that are distributed, resilient to different noise formations, and adapt to different system dynamics. This work proposes methods that particularly focus on reducing estimation error in the face of non-linearities and non-gaussian noise. Much recent research has gone into addressing these aspects and a summary of most popular and up-to-date methods is provided in Table II. All these related studies make simplifying assumptions for the Observation Model that rely on linearization and Gaussian representation.

This work presents two well-known filtering methods, namely the belief condensation filter (BCF) and the particle filtering (PF) which have been used in other domains such as Network Localization and Navigation \([26]\), and applies them to the problem of SE in Energy Networks. These are recent filtering techniques that perform provably well under non-linear, non-gaussian conditions. The effect of dynamic modeling is left as a subject of future research in this work. To this effect, a forecasting method that is commonly used in the literature is also employed here. In Section III-A a review of the Power Systems Theory and the complex equations relating observed and state variables are presented. In Section III-A the HMM used for modeling and forecasting in power networks is explained. Detailed descriptions along with preliminary analysis of the proposed filtering methods (the BCF and PF) are provided in Sections III-B2 and III-B3. In Section III-C in-depth numerical experimentation is carried out to analyze the performance of the BCF and the PF. The simulation set-up is described in Section III-D. The characterization of non-linear versus linear filtering in power networks is provided in Section III-D. Section III-D details out the presence and effect of non-Gaussian uncertainty in the network. A full-system simulation that measures and compares the performance of pertinent algorithms in a real-time scenario is given in Section III-E. Final conclusions and remarks are drawn in Section IV.

II. SYSTEM MODEL AND METHODOLOGY

In this section, a review of the relevant parts of the Power Systems theory is presented with elaboration for their utilization within a HMM framework for SE. Based on this set-up, the development of the BCF and PF for the task of filtering over HMM’s is detailed out as well.

A. Power Systems

The network theory and equations for power systems have long been established \([27], [28]\). Figure 8 shows a simple diagram of a power microgrid. Any such system consists of 4 primary system components: 1) Bus lines - which transfer power; 2) Branches - as connections between bus lines; 3) Generators - as sources of power; 4) Loads - as sinks for power. As far as network control goes, the state of generators are fully known to the operators. We are only concerned with state estimation for bus, branch, and load elements at different points, known as nodes in the network. The parameters of interest at each node to compute are complex voltages, powers, and currents (also called phasors), as summarized in Table II for an \( n \)-node network.

For the purposes of SE in energy networks, it suffices to compute line voltage magnitudes and phases only. Given physical values (e.g. line resistances, impedances, and admittances), currents and powers can be calculated thereby. Put in Estimation Theory terminology, the complex voltage values form sufficient statistics, the knowledge of which tells us everything needed to know about the network. As such, the interesting part is the parameter relations in terms of voltage magnitudes and angles. The following sections are aimed at explaining these relation, which form the basis for SE’s Observation Model in this work’s analysis III:

1) AC Power Flow Model: A single-phase AC model for injected power at bus \( n \) is given by:
\[
P_n = \sum_{i \neq n} |V_i||V_n|(g_{in}\cos(\theta_i - \theta_n) + b_{in}\sin(\theta_i - \theta_n))
\]
Filtering Technique | Families of Distributions | Filtering Approach
--- | --- | ---
Least Squares Minimization (LSQ) | N/A | Minimize objective function by Newtons Method
Kalman Filtering (KF), Extended Kalman Filtering (EKF) | Gaussian Distribution | Linearize the models by Taylor Expansion
Unscented Kalman Filtering (UKF) | Gaussian Distribution | Approximate KF recursions by numerical integration with different quadrature rules
Belief Condensation Filtering (BCF) | Mixture of Exponential Families (e.g. Gaussian) | Approximate complex distributions by right sided Kullback-Leibler (KL) divergence
Particle Filtering (PF) | Discrete Distributions | Approximate complex distributions by random sampling with different proposal densities

### TABLE III

**Power system parameters for network of size n.**

| Parameters                          | Values |
|-------------------------------------|--------|
| Bus Voltage Magnitudes | $|V_1|, |V_2|, \ldots, |V_n|$ |
| Bus Voltage Phase Angles | $\theta_1, \theta_2, \ldots, \theta_n$ |
| Bus Real Powers                   | $P_1, P_2, \ldots, P_n$ |
| Bus Reactive Powers               | $Q_1, Q_2, \ldots, Q_n$ |
| Bus Current Magnitudes            | $|I_1|, |I_2|, \ldots, |I_n|$ |
| Bus Current Phase Angles          | $\angle I_1, \angle I_2, \ldots, \angle I_n$ |

\[
Q_n = \sum_{i \neq n} |V_i||V_n||g_{in} \sin(\theta_i - \theta_n) - b_{in} \cos(\theta_i - \theta_n)\]  

(9)

for the real power and reactive power. Here, the summation is over all other connected buses to bus $n$, $g_{ij}$ denotes conductance between bus $i$ and $j$, and $b_{ij}$ denotes susceptance between bus $i$ and $j$. A subscript 0 denotes the reference (e.g. ground). The real and reactive power flow from bus $i$ to bus $j$ takes the form:

\[
P_{ij} = |V_i|^2(g_{ij} + g_{ij}) - |V_i||V_j|(g_{ij} \cos(\theta_i - \theta_j) - b_{ij} \sin(\theta_i - \theta_j))
\]

(10)

\[
Q_{ij} = -|V_i|^2(b_{ij} + b_{ij})
\]

(11)

2) **Line Current and Voltage Model:** The relation between voltage-phaser measurements and voltage estimations can be expressed by a linear model. That is, a measured voltage value is a noisy version of the actual voltage state value. Unlike voltage measurements, voltage estimations form a non-linear relation with current-phasers. This relation using a $\pi$-model is given by:

\[
\Re\{I_{ij}\} = (g_{ij} + g_{ij0})|V_i|\cos(\theta_i) - g_{ij}|V_j|\cos(\theta_j) - (b_{ij} + b_{ij0})|V_i|\sin(\theta_i) - b_{ij}|V_j|\sin(\theta_j)
\]

(12)

\[
\Im\{I_{ij}\} = (g_{ij} + g_{ij0})|V_i|\sin(\theta_i)
\]

Solving these non-linear equations is known to be a hard problem and often results in infeasible solutions. In general, three simplifying assumptions are made to reduce the AC model to a system of linear equations which can be efficiently solved - called the DC Power Model. The assumptions are:

1) $b_{ij} \gg g_{ij}$; 2) Phase angle differences are negligible, i.e. $\sin(\theta_i - \theta_j) \approx \theta_i - \theta_j$; 3) Voltage magnitudes are close to one, i.e. $|V_i| \approx 1$. Under the DC Model, then non-linearities collapse as:

\[
P_n = \sum_{i \neq n} b_{in}(\theta_n - \theta_i)
\]

(14)

The conjecture here is that, such simplifications can cause significant performance degradation for estimation algorithms. While the state-of-the-art algorithms as described in Section II-B resort to linearizations of this sort, the main idea in this work is to numerically verify that indeed non-linearities in the system are non-neglectable and propose filtering methods that overcome this deficiency.

### B. State Estimation

The problem of finding line voltage magnitudes and angles in the face of uncertainty can be cast as a HMM (also referred to as Sequential Inference) problem. In this framework, based on the knowledge of past states, a new state is first predicted. This prediction is further corrected by taking into account the observables (i.e. measurements). This section details out this formulation and describes the methodology used in computing the different components.

First let’s introduce some notation. Let

\[
x_t = [|V_1|, |V_2|, \ldots, |V_n|, \theta_1, \theta_2, \ldots, \theta_n]
\]

(15)

denote the state vector consisting of voltage magnitudes and angles that we wish to estimate. As stated in Section II-A computing the state vector allows us to compute any other line values. That is, knowing the state vector tells us everything we need to know about the system. Simultaneously, let $y_t$ be the vector of observations. To maintain generality, in this section
let us avoid specifying what each element of the observation vector is. In general the entries of \( y_t \) can be any independent subset of the bus parameters at a given time \( t \). In Section III for analysis this is explicitly set as:

\[
y_t = [V_1, V_2, \ldots, V_l, P_1, P_2, \ldots, P_m].
\] (16)

The relation between state parameters and observations according to the discussion in Section II-A can ideally be expressed through the following relation:

\[
x_t = h(y_t)
\] (17)

To be specific, \( h(\cdot) \) captures the observation equations (10) pertinent to power measurements, and (12-13) pertinent to voltage and current measurements. This at first glance insinuates an algebraic-based solution via setting up a system of equations. However, there are three major considerations that call for a closer examination of the problem:

i) The original relations between such parameters are highly non-linear.

ii) Even with linearization assumptions in (14), due to sparsity of measurements in large networks at any given time, not all state parameters can be obtained using this method.

iii) Observations themselves are inflicted by measurement noise.

In view of the issues identified above, rather than a deterministic approach, it is then logical to take a probabilistic view - a common approach in the literature through Bayesian networks. In this framework, in lieu of direct computation of states, the joint probability density of the state vectors and observations \( f(x_{1:T}, y_{1:T}) \) from start to time \( T \) is sought after. The joint probability density is of interest since other distributions of interest can be deduced via marginalization for instance, which in turn lend themselves to inference about single state parameters. The following sections explain the treatment of this density function.

1) A Hidden Markov Model (HMM): Following the above discussion, the goal here is to make estimations about the state vector parameters based on noisy measurements of the observables. As an additional relation, the evolution of the states through time is also driven by system dynamics. For instance, the state of voltage on a given line is a function of powers drawn from connected branches. The specific system dynamics can be described by a HMM in which the hidden parameters are the voltages to be estimated and observables are any subset of voltage, power, and current measurements (29-31).

The HMM framework leads to two immediate assumptions. Graphically, these conditions are expressed in Figure 2

1) The states \( x_t \) form a Markov chain, i.e. voltage magnitudes and angles at a time \( t \) only depend on those of the previous and/or next time step.

2) Observables \( y_t \) are independent conditioned on the states \( x_t \).

\[
\begin{align*}
\text{Fig. 2. A HMM for the sequential inference in Smart Grids.}
\end{align*}
\]
Equation (22) is referred to as the Prediction Step and (23) is referred to as the Update Step. Our development in this work focuses on the latter. While the Prediction Step captures the dynamics of the system as it evolves through time, the implications of the assumptions in equations (6) are far too superior to neglect in achieving good performance. As such, we seek filtering methods that are immune to system non-linearities and complex noise models. In what follows, two such filters are presented - the BCF and the PF. Both methods have been applied in the context of Localization and Navigation Systems [8], [26] - an application with similar problem setup where positional state of agents and measurements are related through highly non-linear, non-Gaussian relations. We present the details of the two filters below and later carry out analysis as to how much performance gain they achieve for the application of Smart Grids SE in Section III. The details of algorithmic implementation for the BCF can also be found in [24].

2) Belief Condensation: BCF is a filtering framework where the true posterior of the state vector is approximated by a mixture of probability density functions. It has been shown that under certain optimality conditions, BCF can provide accuracies approaching the theoretical bounds and outperforming existing techniques, particularly for non-linear/non-Gaussian problems. One of the main advantages of the BCF filtering method is its treatment of the observation function as an inverse problem. Whereas Kalman-like filters tend to linearize the observation method where needed, which in specific cases of Gaussian distributions for instance perform remarkably well, the BCF filter makes no assumptions as to what form $h(\cdot)$ must attain. This level of abstraction equips the BCF filter with interesting performance advantages while keeping the computational complexity low. In this section, the details of the BCF filter are outlined, together with analysis that explores such advantages.

Consider the mixture family $\mathcal{F}_{\Xi^n}$ with an instance member $g(x; \xi)$ given as:

$$
g(x; \xi) = \sum_{i=1}^{m} \alpha_i g_i(x; \theta_i) \tag{24}
$$

where $\{\alpha_1, \alpha_2, \cdots, \alpha_m\} \in \mathbb{R}_+$, $\sum_{i=1}^{m} \alpha_i = 1$, and $g_i(x; \theta_i)$, for each $i \in \{1, 2, \cdots, m\}$, belongs to an exponential family $\mathcal{F}_{\Theta_m}$, given by:

$$
g_i(x; \theta_i) = q_i(x) \exp \{\theta_i^T t_i(x) - A_i(\theta_i)\} \tag{25}
$$

Here $\theta_i \in \Theta_i$, $t_i(x)$, and $A_i(\theta_i)$ are the natural parameters, sufficient statistics, and log-partition function of $\mathcal{F}_{\Theta_m}$. The parameter set for $g(x; \xi)$ consists of $\xi = (\alpha_1, \theta_1, \cdots, \alpha_m, \theta_m) \in \Xi^m$.

Let $f \in \mathcal{P}$ from the distribution family $\mathcal{P}$ denote the posterior distribution that we wish to approximate by $g(x; \xi) \in \mathcal{F}_{\Xi^n}$ within the Sequential Inference framework. For instance, in our particular analysis $f(x) = f(x|y_{1:t})$. The Kullback-Leibler (KL) divergence $D_{KL}$ between the probability distributions $f(x)$ and $g(x; \xi)$ is defined by:

$$
D_{KL}(f, g_{\xi}) = \mathbb{E}_f \left\{ \log \frac{f(x)}{g_{\xi}(x)} \right\} \tag{26}
$$

It can be shown that, under the following regularity conditions, BCF recursions (see Theorem I), condense the probability distribution $f(x)$ into a mixture of exponential families:

1. **Theorem I.** If a continuous probability distribution $f(x)$ satisfies the regularity conditions A1-A4, then the sequence $\{D_{KL}(f, g_{\xi})\}_{\xi \in \Xi}$ is monotonically decreasing, where the sequence $\{\xi^{[0]}\}_{\xi \in \Xi}$ is recursively determined by:

$$
\alpha_{i}^{[l+1]} = \alpha_{i}^{[l]} \mathbb{E}_{g_i}(x|\theta_{i}^{[l]}) \left\{ \frac{f(x)}{g(x;\xi^{[l]})} \right\} \tag{27}
$$

for $i \in \{1, 2, \cdots, m\}$ and $\theta_{i}^{[l+1]}$ satisfying

$$
\mathbb{E}_{g_i}(x|\theta_{i}^{[l+1]}) \{t_i(x)\} = \frac{\mathbb{E}_{g_i}(x|\theta_{i}^{[l]}) \{t_i(x)\}}{\mathbb{E}_{g_i}(x|\theta_{i}^{[l]}) \{t_i(x)\}} \tag{28}
$$

for each $i \in \{1, 2, \cdots, m\}$ and any initial parameter $\xi^{[0]} = (\alpha_1^{[0]}, \theta_1^{[0]}, \alpha_2^{[0]}, \theta_2^{[0]}, \cdots, \alpha_m^{[0]}, \theta_m^{[0]}) \in U_f$.

In the case where the exponential families are Gaussian, $\theta_{i}^{[l+1]}$ in (28) can be obtained in a closed form as shown in the following Corollary.

**Corollary 1.** Let $\mathcal{F}_{\Xi^n}$ be the mixture family of $m$ Gaussian distributions, with each mixture component parameterized by $\theta_i = \{\mu_i, \Sigma_i\}$. That is,

$$
g_i(x; \theta_i) \sim N(\mu_i, \Sigma_i)
$$

If $f(x)$ is a continuous probability distribution satisfying the regularity conditions A1-A4, then the natural parameter $\theta_{i}^{[l+1]}$ at update step $l + 1$ can be obtained by:

$$
\mu_i^{[l+1]} = \frac{\mathbb{E}_{g_i}(x|\theta_{i}^{[l]}) \{f(x)\}}{\mathbb{E}_{g_i}(x|\theta_{i}^{[l]}) \{g(x;\xi^{[l]})\} x} \tag{29}
$$

$$
\Sigma_i^{[l+1]} = \frac{\mathbb{E}_{g_i}(x|\theta_{i}^{[l]}) \{f(x)\}}{\mathbb{E}_{g_i}(x|\theta_{i}^{[l]}) \{g(x;\xi^{[l]})\} x^T} - \mu_i^{[l+1]} (\mu_i^{[l+1]})^T \tag{30}
$$

Equations (29) and (30) provide a recursive method for calculating and updating the state variables in each step. The main complexity in this computation comes from carrying out the computation for the expectation integrals. The fact that these expectations are taken with respect to a member of an exponential family (namely a Gaussian distribution) can be exploited, for which efficient quadrature rules exist [32], [33]. In this case, these integrals can be efficiently carried out with polynomial time in $m$ the number of components, in $q$ the
number of quadrature points used, and in $d$ the dimension of the state vector.

**Example 1 (A Linear Model).** Consider a simple, linear observation model where we try to estimate parameter $x \in \mathbb{R}$ given $N$ i.i.d. noisy observations $y_i \in \mathbb{R}$, described by:

$$y_i = x + n_i, \quad i = 1, 2, \cdots, N$$

where $n_i$ are independent, additive noise terms distributed according to a normal Gaussian distribution with variance $\sigma_n^2 = 0.1$. We run the BCF filtering algorithm described above for different values of $m$, analyze, and compare the results to the extended Kalman filter (EKF), in order to provide insight about different aspects of the BCF algorithm.

To characterize the performance of the BCF algorithm, a Monte Carlo simulation is run and the mean squared error (MSE) over 100 trials is computed. In each trial, we generate $N = 5$ noisy observations from the true value of $x = 2.2$ and make point estimation using both BCF and EKF methods. We use the mean of the posterior as the point estimate. Figure 3 shows the MSE plot versus the size of gaussian mixture $m = \{1, 2, \cdots, 9\}$. Since the model is linear, the EKF and BCF show comparable performance (difference is within one standard deviation). As the number of gaussian mixtures used in the BCF algorithm is increased, the accuracy also increases. After mixture size of $m = 6$ however, the performance is degraded due to the fact that the model is becoming over-complex for the simple scenario under study here, hence impairing the accuracy. Figure 4 shows the convergence behavior of the coefficients $\alpha_i$ which in general settle after $l = 5$ to $l = 10$ steps.

**Example 2 (A Bi-modal Distribution).** In order to further investigate the power of the BCF filter in the face of more complex models, we consider a scenario where instead true values of $x$ are generated from a bi-modal distribution - a mixture of 2 gaussian distributions with means $\mu_1 = -1.1$ and $\mu_2 = 1.1$, and variance $\sigma_n^2 = 0.1$. Here we set the mixing value to be $p = 0.5$. Similar to the previous example, we run both the BCF and EKF algorithms and compare the behavior. In this case, we use $m = 2$ gaussians in the mixture model. Figure 5 shows that the BCF filter represents this model correctly in the probability density sense, while the EKF falls short in capturing the bi-modal nature of the underlying model and is concentrated around the mean with small standard deviation.

**Example 3 (A Non-linear Model).** As a final example, let’s consider an observation function that has a general form similar to those that appear in Power Grids. In particular, imagine the relation between the observation $y$ and state parameter $x$ to be expressed by the non-linear function $h(x) = x^2 \sin(x)$, which gives rise to the following observation model:

$$y_i = x^2 \sin(x) + n_i, \quad i = 1, 2, \cdots N$$

As before, we run 100 Monte Carlo trials and compute the MSE from both the BCF and EKF algorithms. The results are shown in Figure 6. From the figure, it can be seen that in the presence of non-linearity in the model, the BCF outperforms the Kalman filter due to linearization assumptions such methods make. It can also be seen that, with increasing $m$, the accuracy in estimation for BCF also improves.
3) Particle Filtering: A density representation method that has recently obtained much popularity in a variety of applications is the PF. These filters are a powerful tool that theoretically can approximate any density function in the limit of sampling size. In essence, this method discretizes the probably space via an ensemble of samples called particles, each of which is weighted to represent the density at that sample. Therefore, as the number of samples \( n \) approaches infinity, the gap between the true and approximated distribution closes ever so tightly. In this section, the mathematical formulation of the PF method is explained, along with examples to provide insightful perspective.

The PF representation of a posterior density of interest \( f(x) \) consists of a set \( \{ \hat{x}_i, w_i \}_{i=1}^n \) of \( n \) particles (or samples) \( \hat{x}_i \) and their respective weights \( w_i \), for each \( i \in \{1, 2, \ldots, n\} \). Consequently, we can write an approximate representation for \( f(x) \) with:

\[
f(x) \approx \sum_{i=1}^{n} w_i \delta(x, \hat{x}_i)
\]

(31)

where \( \delta(\cdot) \) is the Dirac delta function, and weights sum to one, \( \sum_{i=1}^{n} w_i = 1 \). As \( n \to \infty \), the approximation would be exact in theory, i.e. \( f(x) = \sum_{i=1}^{\infty} w_i \delta(x, \hat{x}_i) \).

There exist a number of different implementations of the PF, a survey of which is provided in [34–36]. In this work, given the particle representation for a density function \( f(x) \), the recursive update equation for sample weights is given by [25]:

\[
w_{i}^{[t+1]} = f(x^{[t]}|\hat{x}_{i}^{[t]}) w_{i}^{[t]}
\]

(32)

and the task of inference on the state vector \( x \) can be then carried out via the minimum-mean square estimate (MMSE) given by:

\[
\hat{x}_{\text{MMSE}} = \sum_{i=1}^{n} w_i \hat{x}_i
\]

(33)

or the Maximum A-Posteriori (MAP) estimate given by:

\[
\hat{x}_{\text{MAP}} = \arg \max_{\hat{x}_1, \ldots, \hat{x}_n} \{ w_i \}
\]

(34)

III. NUMERICAL EXPERIMENTATION AND ANALYSIS

In this section, we apply the BCF and PF methodologies to the problem of SE in power networks. The main goal of this experimentation is to investigate and numerically support the following claims:

i) In the presence of non-linearities in the system, Kalman-like filters that resort to linear approximations, trade-off accuracy for efficiency. In such cases, continuous mixture models (e.g. BCF) and discretized alternatives (e.g. PF) provide an ideal substitute, maintaining both accuracy and efficiency.

ii) In the presence of more complex and/or unknown noise (e.g. non-Gaussian) in the system, the said methods (i.e. BCF and PF) significantly reduce the estimation error by abstracting out the noise model.

In order to quantify the outcomes under each claim, first a measure of accuracy and a measure of efficiency are established. Characterizing the accuracy of the algorithms is done via the MSE per unit [3]. Given estimates \( \hat{x}_t \) at time \( t \), the MSE is defined by:

\[
\text{MSE} = \frac{1}{T} \sum_{t=1}^{T} ||x_{\text{true}} - \hat{x}_t||^2
\]

(35)

\(^2\) All variables are normalized to avoid unit mismatches in MSE computation. This is due to the fact that the state vector contains both voltage and angle values which have different units.
where \( x_{\text{true}} \) denotes the true values of the state vector, and the mixture mean and the MMSE are used as the final point estimate in BCF and PF, respectively. The efficiency is analyzed by looking at either the computation complexity using the \( \mathcal{O}(\cdot) \) notation or the computation time per operation, depending on the type of algorithm under study.

Two popular methods from the literature are used to benchmark the results: 1) least squares estimation (LSQ) - a generic method that is widely used in the current power networks; 2) uncented Kalman filter (UKF) - the most state-of-the-art Kalman-like filter in the community.

Using the measures defined here, the claims made above are verified through 3 sets of numerical experiments. In the first set of simulations described in Section III-B the effect of linear (Kalman) versus non-linear (BCF and PF) filtering is investigated by computing the MSE of each algorithm. By holding the dynamic changes and noise model constant in the system, in this case the performance behaviors of the BCF and PF are singled out and measured against the UKF.

In the second set of simulations, in Section III-C a mixture noise model is introduced that is additive, identically distributed, and is controlled for Gaussian-ness by a parameter \( p \) - the noise mixing coefficient. In a similar fashion, the resiliency of each algorithm to noise is measured in terms of the MSE as a function of varying noise coefficient. Lastly, in Section III-D the filtering methods are simulated in a real-time setting where the state values are dynamically changed over time. In this important step, the goal is to address issues related to practical and operational aspects of the proposed methods in a real life power grid scenario. In particular, we are interested in knowing the performance gain in employing non-linear filters as an alternative in future power grids.

A. Simulation Set-up

In order to verify the performance improvements by the proposed methods in this work, the IEEE 14-bus test system is utilized for numerical simulation. This test system is available for free access at [37] as a part of the MATPOWER package which contains full line state variables and parameters. Figure 8 shows a diagram of the IEEE 14-bus system. Bus 1 is assigned as the slack/reference bus (\( V_1 \geq 1 \) and \( \theta_1 = 0 \)). The task is to estimate \( \{ |V_2|, |V_3|, \ldots, |V_{14}|, \theta_2, \theta_3, \ldots, \theta_{14} \} \) corresponding to Bus 2 through to 14. In all the experiments that follow, a single iteration of the BCF and PF (\( l = 1 \)) is performed in each trial.

Measurements used in the Observation Model are taken at random from power injections as well as voltage amplitude, for a total of 28 measurements in each time step. The measurements are generated by running the real state values through the observation function and superimposing an additive noise. The Gaussian portion of all noise models hereon have a zero mean with standard deviation 2% and 0.1% for power and voltage measurements, respectively. The mixture noise models are described as necessary in the following sections.

In order to simulate the slow dynamics of the systems, a subset of loads is selected initially and changed over a period of 50 time sample intervals. The loads are varied following a linear trend of increase or decrease by 10% in each time step. In addition to these linear changes, a fluctuation of 3% is also superimposed. The dynamic model used in the simulations is the Holt’s Smoothing method, with elements of diagonal matrix \( Q \) equal to \( 10^{-6} \) and \( P_0 \) initially equal to \( 10^{-6} \). Since Holt’s method is 2-step smoothing filter, it’s assumed that the state of the lines are known at first, and the simulation starts at \( t = 2 \).

B. Linear and Non-Linear Filtering

As the first case in point, let us investigate the performance characteristics in the face of non-linearities in the system model for the BCF and PF filtering methods and their treatment thereof. As argued in the previous sections, non-linearities arise in the system due to the complex nature of power equations captured by the Observation Model. In order to isolate this effect, let us fix the time-varying dynamics and consider the network in a static mode. This reduces the network equations down to a single Observation Model given by:

\[
x_t = h(y_t) + n_t
\]  

(36)

Each method is tested by estimating \( h(y_t) \) through Monte Carlo sampling from 1000 independent measurements \( y_t \), from which an aggregate MSE is computed. This is done for different settings of BCF with \( m = 1, 2, \ldots, 9 \) components and PF with \( n = 500 \) to \( n = 10000 \) particles. The noise here is assumed to be i.i.d. Gaussian. The fundamental question here is, under equal dynamic and noise models, do non-linear filters achieve better accuracy (MSE) and efficiency than linear filters?

Figure 9 shows the MSE as a function of mixture components \( m \) for BCF. Inset the figure also is UKF’s MSE - a constant line in this case. It can be seen that in general the BCF curve lies below the UKF line, suggesting performance improvements by non-linear filtering.
From the figure, at \( m = 1 \) it can be seen that BCF achieves a slight improvement in accuracy. This gain in performance is due to BCF’s treatment of non-linear equations \( h(\cdot) \) as an inverse problem. Based on the discussion in Section II-B2 while Kalman-like filtering methods resort to linearization, the BCF estimates the density function for the state variables from the measurements using the true observation model. With more components, i.e. increasing \( m \), the complexity of the BCF model increases, resulting in better estimation power. It can be seen that at \( m = 9 \), near an order of magnitude increase in accuracy can be obtained by non-linear filtering (the BCF) versus linear filtering (the UKF). Ultimately, with a minimum of \( m = 6 \) gaussian components the BCF performance is optimized.

To give this discussion context, let us revisit the IEEE 14-bus test system and re-interpret the results. For a sub-network that contains 14 bus lines at nominal voltage levels of 110V per line (typical in U.S. households), the UKF and the baseline BCF \((m = 1)\) have net MSE’s of approximately 1.5V and 1.3V, respectively \[\text{Fig. 9}\] BCF with \( m = 2 \) components almost doubles the performance with the net MSE of 0.75V. With \( m = 6 \) components, the net MSE is reduced even further to 0.31V.

Figure \[\text{Fig. 10}\] shows the MSE curve for the PF, versus UKF, as a function of the number of particles. For the number of particles \( n \) under 1000, we can see an oscillatory behavior in unstable estimation. From the figure, at about \( n = 4000 \) particles the best accuracy (i.e. lowest MSE) achieved by \( m = 9 \) components from BCF is surpassed. At \( n = 10000 \) particles, the PF obtains an order of magnitude improved performance over the UKF. For the IEEE 14-bus test system, this corresponds to a net MSE of 18mV.

3The net MSE is computed as the value of MSE per unit per line \( \times \) the nominal voltage value \( \times \) the number of bus lines .

C. Non-gaussian Noise Model

In the previous section, it became evident that in the presence of non-linearities, filters that abstract out the role of the observation function are needed in order to achieve better accuracy. Other elements of the model were held constant to verify this conjecture. Specifically, the observation noise was treated as additive, independent, and gaussian with:

\[
\mathbf{n}_t \sim \mathcal{N}(\mu_{t}, \Sigma_{t})
\]  

(37)

While this noise model suffices in cases where only device-related measurement errors are of interest, in more realistic networks the true noise model is not known. For instance, with the advent of wireless PMU and meters, communication errors due to packet drops among others may contribute to measurement errors. In such scenarios, more sophisticated filtering methods are required that can alleviate uncertainties in the system.

In this section, let’s investigate the effect of non-gaussian error in power grids and expose the filtering methods to more general error models to understand their resilience. In particular, let us consider a general mixture noise model given by:

\[
\mathbf{n}_t = p_e \mathbf{n}_t^g + (1-p_e) \mathbf{n}_t^u
\]  

(38)

which is the sum of a gaussian random variable \( \mathbf{n}_t^g \) and a uniform random variable \( \mathbf{n}_t^u \), regulated by the mixture coefficient \( p_e \in [0,1] \):

\[
\mathbf{n}_t^g \sim \mathcal{N}(\mu_{t}, \Sigma_{t})
\]  

(39)

\[
\mathbf{n}_t^u \sim \mathcal{U}(a, b)
\]  

(40)

The uniform component can be thought of as an event where the measurement is fully corrupted due to packet collision for instance, with probability of occurrence \( 1 - p_e \). Let’s repeat the simulation process as above, and look at the MSE as a function of the mixing coefficient \( p_e \). Notice that at \( p_e = 1 \), the numerical results collapse to that seen in Section III-B while at the other extreme when \( p_e = 0 \), we only see uniform noise.
Figure [11] shows the effect of non-gaussian noise on the BCF algorithm with \( m = 1, 2, 3, 5, 9 \) components in terms of the MSE. The UKF curve is shown in red. In general, there is an increasing pattern as the noise mixture approaches uniform distribution (to the left). This is expected since uniform distribution implies the most uncertainty in the system, contributing most to the degradation of the estimates. The main point to note from this figure is the relative degradation in the BCF as opposed to the UKF. While due to increasing uncertainty the UKF’s MSE experiences a 4.4-fold increase, the BCF exhibits more noise resilience, keeping this degradation gradual and limited to under 2 times. As expected, the BCF’s performance improves with the number of components included (i.e. \( m = 2, 3, 5, 9 \)) due to better accuracy in the update step through the observation model.

Figure [12] shows these results for the PF with \( n = 200, 1000, 5000, 8000, 10000 \) particles. At very low number of particles, the PF exhibits oscillatory behavior, seen from the curve for \( n = 200 \). This is due to insufficient number of particles to rightfully represent the true distribution (under 1000 particles according to discussion in Figure [10] Section III-B). With larger number of particles, the performance also improves. As was the case with BCF, with increasing uncertainty (from right to left on the figure) the performance is impaired. In contrast to the BCF, at smaller number of particles, the PF has a steeper increasing pattern with the mixing noise coefficient. It can be seen that with the number of particles \( n \) large enough, the steadiness is regained. At \( n = 10000 \), the change in MSE is still 16-fold - a considerable jump. This is attributed to the fact that the MMSE estimate is a weighted sum of all the particles that are overfitting to the complex noise model in this case.

**D. Full System Simulation**

As a final test, in this section we put the pieces together and a full simulation is carried to compare algorithms as the state changes through time. The evolution of states through time is described by the Dynamic Model given in (19). To this end, a subset of the state variables are selected pre-run at random. During the simulations, the selected subset’s true values are changed according to a linear model (increased or decreased) as per descriptions in Section III-A. Measurement noise is considered in the same fashion as before.

In each time step, first a prediction is made based on the previously computed state values and measured observations. This step corresponds to Prediction Step in (22) in the HMM model. Here, Holt’s smoothing is used that provides a linearized way (See Appendix B for details). In Holt’s method, innate parameters are set as \( \alpha_t = 0.81 \) and \( \beta_t = 0.56 \), empirically. Once a new state is obtained, predictions are updated with measurements using each of the algorithms under study, in a set-up similar to previous sections.

Figure [13] shows the MSE per unit for 4 different filtering methods - LSQ, UKF, baseline BCF with \( m = 1 \), and the PF with \( n = 10000 \). The LSQ filter as a fourth method is considered here due to wide acceptance currently in the community as an estimation filter. In each time step, a prediction \( x_t \) is made for the state vector according to the dynamic model (same for all filtering methods), followed by an update step by each filter to obtain \( x_t \).

It can be seen that the baseline BCF and PF reduce the MSE by an order of magnitude per unit over their counter parts. The LSQ has the poorest and most fluctuating performance. This is due to the fact that it minimizes the squared error between the measurements \( y_t \) and \( h(x_t) \) using a gradient descent approach. It disregards non-linear effects and lacks a smoothing mechanism over time. The remaining curves exhibit a smooth prediction and correction tracking overall. Due to their strength in mitigating the effect of non-linearities, the BCF and PF curves lie below the UKF curve. The UKF suffers from the highly non-linear nature of the observation model, resulting in impaired performance as expected.

To further investigate the behavior of BCF algorithm in a dynamic setting, the same experiment is repeated for \( m = 3 \) and \( m = 9 \). The results are shown in Figure [14] along the
$m = 1$ plot from Figure 13. As before, the three filters start and settle into a steady state tracking the linear changes in the network. The slight jump in the plots at step 10 is where the state of randomly selected bus lines start to change for all simulations, so to leave Holt’s method slack to settle in. As more BCF components are added to BCF, the update from the measurements is improved in each step. These incremental improvements carry over to next time steps through the dynamic model, making the overall performance better as confirmed by $m = 3$ and $m = 9$ curves. Furthermore, in Section III-B it was observed that with enough components, the BCF can achieve accuracies close to the PF. From Figure 14 with $m = 9$ the dynamic BCF achieves accuracy of around $0.002$ per unit or $0.29$ Volts. This is almost identical to $PF^*$’s performance at $n = 10000$ particles which confirms our observation.

Average MSE values over $500$ runs are shown in Table IV. To put these numbers in perspective, let us revisit the IEEE 14-bus example that is used as a test bench in this experimentation. Reading from the third column in Table IV in this test case (14 buses, with each bus at nominal voltage $110V$) the LSQ has an MSE of $14V$. For a relatively small sized network such as the IEEE 14-bus, this is a considerable error in estimation. The UKF has an MSE of $2.9V$, a much smaller deviation comparatively. The BCF further keeps the error margin under a volt, while the PF algorithm reduces MSE to a few hundred milli-volts. This is an improvement in accuracy by at least an order magnitude.

### E. Complexity Analysis

The results thus far have shown that non-linear filters such as the BCF and PF improve upon Kalman-like filters in estimating the state of power grids. In particular, the PF, with the right ensemble size, outperforms the rest of the methods (see Table IV). In the current literature, it is claimed that in the limit this filtering method indeed can achieve the theoretical bounds of estimation accuracy. On the other hand, it is known that PF suffers from the curse of dimensionality as it utilizes Monte Carlo methods to approximate multidimensional integrals [38]–[40]. To this date, there is no analytical formula that capture the complexity of PF as a function of the size and computation time of the problem.

In light of this obstacle and in order to trade off accuracy versus practical considerations, we pursue and compare here the computation time of the BCF and BF through numerical experimentation. We adopt a methodology presented in [41], whereby a dimension free metric for error is introduced. The dimension-less nature of this metric allows us to abstract out performance degeneration caused due the size of the problem and get a true comparison by fixing the accuracy.

Adopting the notation of [41], let $r$ denote the Mean Dimension-Free Error defined by:

$$r = \frac{\mathbb{E}\{(x - y)^* J (x - y)\}}{d}$$  \hspace{1cm} (41)$$

where $y$ is the estimate of $x$ from the BCF or PF, $J$ is the inverse of the estimation error covariance matrix, $x$ is the state vector to be estimated, and $(\cdot)^*$ denotes the transpose of $(\cdot)$. First the number of mixture components $m$ (in the case of BCF) or particles $n$ (in the case of BF) are selected that obtain a fixed value of $r$. The complexity of each filter is then defined as the time it takes the algorithms to reach $r$ with parameters selected in this fashion.

Figure 15 and Figure 16 show the Mean Dimension-Free Error plots for the BCF and PF respectively, one curve per dimension $d$. It can be seen that, for each curve, the error...
measure decreases with diminishing return as a function of increasing number of BCF components and number of particles. This is in agreement with the MSE analysis from Figure 9 and Figure 10 where accuracy improves with better parameters.

Here, we set the Dimension-Free Error to be $r = 2$ as shown by the black dotted horizontal lines. Where the curves cross the $r = 2$ line specify the choice of $m$ and $n$ for each dimension. For higher values of $d$, the intersecting points are moved further right, corresponding to larger component and ensemble sizes. This is expected, implying increased computation time as the dimensionality of the problem grows.

Figure 17 shows the computational time for each filtering method as a function of varying state dimension following the above process above. From the figure, the PF at very low dimensions shows a fast start, laying below the BCF curve with only a few hundred particles. As the size of the problem increases, the computational time for PF grows very rapidly, surpassing its counterpart after $d = 5$ from the numerical results. At dimension $d = 10$ alone, with only a few thousand particles it can be seen that PF’s computation time has doubled that of BCF’s.

Figure 18 shows the elapsed time for a PF update as a function of the number of particles up to $n = 10000$ for the IEEE 14-bus example. Here, the state is changed according to the dynamic behavior described above. Consider the case in which we are interested in tracking a household line voltage at SCADA rate of 30 samples/second. This requires a sample roughly every 33ms. Theoretically, as the number of particles approaches infinity, the PF can represent probability densities of any form no matter how complex they get. Unfortunately, it is evident that working with the PF with $n > 10000$ encounters major practical challenges.

IV. CONCLUSION

Power generation and distribution is quickly moving from a traditional top-down structure to smart grids, where power injection into and from the grid can occur at any point in the network. In order to ensure flawless operation through this paradigm shift, the need for SE across the network with higher accuracies than ever before is inarguable.

Classical approaches cast SE as a HMM whereby the state of the system is first predicted based on a priori knowledge (Prediction Step) and then the prediction is corrected using measurements (Update Step). The full system can be described by a forecasting equation capturing the time-varying dynamics of the system (Dynamic Model) and an observation function relating the state predictions to network measurements (Observation Model). There are four challenges that stand in the way of any algorithmic solution: 1) Scalability; 2) Integrated Dynamics; 3) Non-linear System Characteristics; 4) Uncharted System Noise. These factors are highly correlated and essential to the performance of SE algorithms.

The state-of-the-art filtering methods include LSQ (widely adopted) and different variations of the Kalman filter (UKF being the most up-to-date). The key assumptions that sit at the heart of deficiencies in current methodologies are the linear
Fig. 18. Operation time of PF update step as a function of the size of the state vector.

The treatment of measurement function \( h(\cdot) \) and the Gaussian simplification of the noise model \( n_t \). The fundamental questions raised in this work are: 1) Can the non-linearities in the system be more realistically represented? 2) Can more general noise models be utilized to capture the network uncertainty?

This work puts forth two filtering methods, namely the BCF and the PF, which have successfully been used in other domains, and applies them for the first time to the SE problem in power grids. We remove linearity constraints on the Observation Model and represent the complex power equations in their original form. Furthermore, we introduce a mixture uncertainty model that enables modeling of non-Gaussian noise in the system. Based on simulations of the IEEE 14-bus test-bench, numerical analysis in this work reveals that under system non-linearities and non-Gaussian noise, the BCF and PF achieve significant improvement in performance compared to LSQ and UKF.

Simulations are run for different parameter settings of the proposed algorithms. In particular, the MSE is used as a measure of accuracy for the BCF with \( m = \{1, 2, \ldots, 9\} \) components and for the PF with up to \( n = 10,000 \) particles. Under i.i.d additive Gaussian noise, even the baseline BCF (\( m = 1 \)) and a small ensemble of particles (\( n = 1000 \)) reduce the MSE over the UKF, which is a second-order Kalman Filter. Increasing the number of components in the BCF or particles in the PF only improves the results. With \( m = 9 \) components and \( n = 10,000 \) particles, an MSE per unit per line of \( 1.93 \times 10^{-4} \) and \( 1.2 \times 10^{-4} \) are achieved by each algorithm, in contrast to the UKF's MSE per unit per line of \( 9.66 \times 10^{-5} \). For a network of size 14, this corresponds to net errors of \( 310mV \) (the BCF) and \( 180mV \) (the PF) versus \( 1.5V \) (the UKF) - an order of magnitude improvement by the BCF and PF. The non-Gaussianity in the system only makes these advantages more prominent. Under the most uncertainty (i.e. uniform noise), the MSE's increase to \( 1.4 \times 10^{-3} \) for the PF with \( n = 10,000 \), \( 1.6 \times 10^{-3} \) for the BCF with \( m = 1 \), and \( 4.5 \times 10^{-3} \) per unit per line for the UKF.

The treatment of observation model as an inverse problem by the BCF and the ensemble representation of the posterior density function by the PF, make them an ideal choice for the problem of state estimation in power grids. The numerical analysis in this work show enhanced performance of at least an order of magnitude for these methods over the current technology - a difference that becomes crucial at scales and dynamic complexities required by future grids. Although the PF itself can achieve better accuracies than the BCF, this gain comes at a price. As the dimension of the problem at hand increases, numerical results in this work have shown that the computational complexity of the PF increases in an almost exponential fashion, more than doubling in computation time at dimension \( d = 10 \). Therefore, where computational cost is an important factor, the BCF proves itself to be an ideal candidate - coming close to PF's accuracy while offering promising computing times to meet the needs of future Smart Grids.
where \( I \) is the identity matrix, and \( \alpha_t \) and \( \beta_t \) are constants between 0 and 1. Also, \( x_t^- \) here denotes the prediction for the state vector at time \( t \), before updating with the measurement information. Further, the vectors \( a \) and \( b \) are computed through the recursions:

\[
\begin{align*}
\mathbf{a}_t &= \alpha_t \mathbf{x}_t + (1 - \alpha_t) \mathbf{x}_t^- \\
\mathbf{b}_t &= \beta_t (\mathbf{a}_t + \mathbf{a}_{t-1}) + (1 - \beta_t) \mathbf{b}_{t-1}
\end{align*}
\]
[21] T. Bi, X. Qin, and Q. Yang, “A novel hybrid state estimator for including synchronized phasor measurements,” Electric Power Systems Research, vol. 78, no. 8, pp. 1343–1352, 2008.

[22] A. Monticelli and A. Garcia, “Reliable bad data processing for real-time state estimation,” IEEE Trans. Power App. Syst., vol. PAS-102, no. 5, pp. 1126–1139, May 1983.

[23] H. Wang and N. N. Schulz, “A revised branch current-based distribution system state estimation algorithm and meter placement impact,” IEEE Trans. Power Syst., vol. 19, no. 1, pp. 207–213, Feb 2004.

[24] S. Mehryar, P. Malekzadeh, S. Mazuelas, P. Spachos, K. N. Plataniotis, and A. Mohammadi, “Belief condensation filtering for rssi-based state estimation in indoor localization,” in ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019, pp. 8385–8389.

[25] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, “A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking,” IEEE Trans. Signal Process., vol. 50, no. 2, pp. 174–188, Feb 2002.

[26] H. Wymeresch, J. Lien, and M. A. Win, “Cooperative localization in wireless networks,” Proceedings of the IEEE, vol. 97, no. 2, pp. 427–450, Feb 2009.

[27] F. C. Schweppe and D. J. Wildes, “Power system static-state estimation - Part I: Exact model,” IEEE Trans. Power App. Syst., vol. PAS-89, no. 1, pp. 120–125, Jan 1970.

[28] F. C. Schweppe, “Power system static-state estimation - Part II: Approximate model,” IEEE Trans. Power App. Syst., vol. PAS-89, no. 1, pp. 125–130, Jan 1970.

[29] M. I. Jordan and T. J. Sejnowski, Graphical models: Foundations of neural computation. MIT press, 2001.

[30] C. M. Bishop, Pattern recognition and machine learning. Springer New York, 2006, vol. 1.

[31] B. C. Arnold, N. Balakrishnan, R. M.iguez, and J. M. Sarabia, Advances in Mathematical and Statistical Modeling. Springer New York, 2008.

[32] A. H. Stroud and D. Secrest, “Approximate integration formulas for certain spherically symmetric regions,” Mathematics of Computation, vol. 17, no. 82, pp. 105–135, 1963.

[33] L. Arasaratnam, S. Haykin, and T. R. Hurd, “Cubature kalman filtering for continuous-discrete systems: Theory and simulations,” IEEE Trans. Signal Process., vol. 58, no. 10, pp. 4977–4993, Oct 2010.

[34] “Special issue on monte carlo methods for statistical signal processing,” IEEE Trans. Signal Process., vol. 50, no. 2, Feb 2002.

[35] N. J. Gordon and D. J. Salmond, and A. F. M. Smith, “Novel approach to nonlinear/non-gaussian bayesian state estimation,” in IEE Proceedings F (Radar and Signal Processing), vol. 140, no. 2. IET, 1993, pp. 107–113.

[36] R. V. D. Merwe, A. Doucet, N. de Freitas, and E. A. Wan, “The unscented particle filter,” in Advances in Neural Information Processing Systems 13, T. Leen, T. Dietterich, and V. Tresp, Eds. MIT Press, 2001, pp. 584–590.

[37] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, “Matpower: Steady-state operations, planning and analysis tools for power systems research and education,” Power Systems, IEEE Transactions on, vol. 26, no. 1, pp. 12–19, Feb 2011.

[38] S. Arulampalam and B. Ristic, “Comparison of the particle filter with range-parameterized and modified polar ekfs for angle-only tracking,” in AeroSense 2000. International Society for Optics and Photonics, 2000, pp. 288–299.

[39] S. Musick, J. L. Greenewald, K. Kastella, and C. Kreucher, “Comparison of particle method and finite difference nonlinear filters for low snr target tracking,” in Proceedings of International Conference on Fusion, Aug 2001.

[40] A. Farina, B. Ristic, and D. Benvenuti, “Tracking a ballistic target: comparison of several nonlinear filters,” IEEE Aerosp. Electron. Syst. Mag., vol. 38, no. 3, pp. 854–867, Jul 2002.

[41] F. Daum and J. Huang, “Curse of dimensionality and particle filters,” in Aerospace Conference, 2003. Proceedings. 2003 IEEE, vol. 4, March 2003, pp. 1979–1993.

[42] S. G. Makridakis and S. C. Wheelwright, Forecasting methods and applications. John Wiley & Sons, 1978.