Tracking Power System State Evolution with Maximum-correntropy-based Extended Kalman Filter

Julio A. D. Massignan, *Student Member*, *IEEE*, João B. A. London Jr., *Member*, *IEEE*, and Vladimiro Miranda, *Fellow*, *IEEE*

***Abstract*—This paper develops a novel approach to track power system state evolution based on the maximum correntropy criterion, due to its robustness against non-Gaussian errors. It includes the temporal aspects on the estimation process within a maximum-correntropy-based extended Kalman filter (MCEKF), which is able to deal with both nonlinear supervisory control and data acquisition (SCADA) and phasor measurement unit (PMU) measurement models. By representing the behavior of the state variables with a nonparametric model within the kernel density estimation, it is possible to include abrupt state transitions as part of the process noise with non-Gaussian characteristics. Also, a novel strategy to update the size of Parzen windows in the kernel estimation is proposed to suppress the effects of suspect samples. By properly adjusting the kernel bandwidth, the proposed MCEKF keeps its accuracy during sudden load changes and contingencies, or in the presence of bad data. Simulations with IEEE test systems and the Brazilian interconnected system are carried out. The results show that the method deals with non-Gaussian noises in both the process and measurement, and provides accurate estimates of the system state under normal and abnormal conditions.**

***Index Terms*—Tracking state estimation, Kalman filter, maximum correntropy, power system, Parzen window.**

I. Introduction

UTILITIES rely on power system state estimation as a crucial computation tool in their control centers, which is capable of increasing the situational awareness about the power system condition in real time [1]. The increased adoption of new sensors in power systems, such as the phasor measurement units (PMUs), is currently driving changes in the operation of power systems towards a more data-driven and interconnected way. The fast sampling of these new sensors enables an important extension of state estimation formulation towards time-variant models [2]. The goal is to evaluate the operation condition under both normal and abnormal conditions, enabling a deeper understanding of power system behaviors.

The representation of system evolution in the time domain, instead of a static state model of the power system, has motivated the development of the so-called dynamic state estimators and the quasi-steady state estimators [2]. Even though the idea of including time domain aspects has been considered since the propositions of the first state estimators [3]-[5], it is only recently that this idea has gained new thrust with the increasing complexity of power systems and the advances in computation technologies [2]. Both dynamic and quasi-steady state approaches represent advances in terms of detailing in the estimation: the first aims at estimating the network state and the internal states of controllers, loads and machines, while the second focuses on the complex nodal voltages of the network and its operation constraints.

Despite their different model resolution, these estimators mostly rely on the Kalman filter theory or its variations such as the conventional linear [6], nonlinear extended [7], unscented [8] and cubature-based [9] versions to solve the estimation problem. These methods include a state space model in the form of a Markov chain process under multivariate Gaussian measurement and process noises. Despite the accurate results shown in the theory, the formulation of classical Kalman filter is prone to practical drawbacks related to the Gaussian hypothesis, such as the presence of bad data (gross errors or outliers) or non-Gaussian measurement noise, as demonstrated with practical data from PMUs in [10].

This limitation has motivated the development of robust approaches as the *H*-infinite [11], generalized maximum likelihood [12], polynomial-chaos loss function [13], robust hybrid [14], the least absolute value criterion [15], and the cubature robust [16] approaches. All these estimators are widely applied in the literature, and show good performance when dealing with bad data and non-Gaussian noise. Nonetheless, they all rely on the basic assumption that the Markov chain is a general probabilistic representation of the system temporal relation, which is a severe limitation since the structural changes on the system may occur [17].

Manuscript received: February 29, 2020; accepted: May 19, 2020. Date of CrossCheck: May 19, 2020. Date of online publication: July 9, 2020.

This work was supported by CNPq (No. 308297/2018-0), CAPES and FAPESP (No. 2016/19646-6). V. Miranda acknowledges ERDF (COMPETE 2020 Programme) and FCT (POCI-01-0145-FEDER-016731 INFUSE).

J. A. D. Massignan (corresponding author) and J. B. A. London Jr. are with the Department of Electrical and Computing Engineering, School of Engineering of São Carlos, University of São Paulo, São Carlos, Brazil (e-mail: julio.massignan@usp.br; jbalj@sc.usp.br).

V. Miranda is with the Faculty of Engineering, University of Porto and INESC TEC, Porto, Portugal (e-mail: vmiranda@inesctec.pt).

DOI: 10.35833/MPCE.2020.000122

Such assumption is translated into a belief that the system state evolution is a smooth process, and that the correct transition matrices (or state transition function or dynamic models of machines and controllers) are available to build the correct state space model. However, this is questionable, because the switching operations, sudden load or generation changes or failure events may abruptly change the condition of the network and introduce discrete discontinuities. To deal with such transitions, some adaptive techniques have been explored in [18] that basically adopt a different set of rules to interchange among different estimators or transition matrices. Nonetheless, such different decision rules limit their generalization in unpredicted scenarios and hamper their adoption by utilities.

This work is an effort towards a new interpretation of the state estimation problem that can encompass both the statistical robustness and flexibility to overcome system transitions in a single framework. By exploiting information theory concepts and modelling the noise characteristics through a kernel density estimation approach, this paper proposes a maximum-correntropy-based extended Kalman filter (MCEKF) to track power system state evolution. As shown in the initial work in [19], by incorporating information theory concepts within the traditional Kalman filter, it is possible to increase the accuracy without relying on the assumption of a Gaussian distribution for the measurement and process noises.

The first application of information theory in the power system static state estimation problem has been reported in the seminal work in [20], which proposes the adoption of the correntropy concept [21] in state estimation. It has been followed by many efforts that show an increase in the robustness against bad data in challenging scenarios such as the case of leverage points and critical measurements [22], [23], and recently, for detecting topology errors [24]. Such methods still focus on static steady state measurement models.

New approaches have come to light, addressing the temporal aspects of the estimation problem based on the correntropy concept [25]-[27]. Reference [25] includes a quasi-steady state estimator only for supervisory control and data acquisition (SCADA) measurements, while [26] includes a dynamic state estimator only for PMUs. Reference [27] uses the generalized correntropy concept within an unscented Kalman filter to improve the robustness against outliers. This paper extends these efforts with a new MCEKF version, enabling the incorporation of both SCADA and PMU measurements in the estimation process, a more practical approach to represent realistic metering systems. It also presents a novel way to deal with suspect samples and system transitions by performing a new special adjustment on the kernel size in the estimation process during the occurrence of such events. The main contributions of this paper can be summarized as follows.

1) A new concept is applied for the state variables, namely state evolution, which is formulated as a nonparametric probabilistic model for the state that encompasses the possibility of different state values and transitions.

2) The MCEKF is applied for tracking state estimation of power systems. This is a type of quasi-steady state estimator based on a modified-Newton algorithm, which is able to deal with nonlinear models from both SCADA and PMUs.

3) A novel method is proposed to suppress the effect of suspect samples and abrupt system transitions through a special adjustment of the kernel Parzen window sizes.

Simulation results on the IEEE test systems and the Brazilian interconnected system show that the MCEKF model provides an accurate state of the network even during severe and abrupt changes such as sudden load changes and contingencies.

II. Theoretical Background

A. Power System Static State Estimation

A nonlinear measurement model formulates the static state estimation problem for a power system with m measurements and n state variables [1], [2]:

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

where $\boldsymbol{z}$ is the measurement vector ($m \times 1$); $\boldsymbol{x}$ is the state vector ($n \times 1$); $\boldsymbol{h}(\boldsymbol{x})$ is the set of nonlinear equations that relates the measured electrical quantities with the state; and $\boldsymbol{e}$ is the random error vector ($m \times 1$) that represents the measurement noise. Traditional state estimation assumes the noise vector as multivariate Gaussian independent random variables, with zero mean and known covariance matrix $\boldsymbol{R}$ ($m \times m$). The state vector is typically composed of the complex nodal voltage of the system, represented in rectangular coordinates in this work.

The maximization of the likelihood function results in the weighted least squares (WLS) criterion to find the estimated state, i.e., WLS snapshot state estimator. Under an independent multivariate Gaussian noise assumption, the optimization problem is written as:

$$\max l(\boldsymbol{e}|\boldsymbol{x}) = \prod_{i=1}^{m} f(e_i|\boldsymbol{x}) = \prod_{i=1}^{m} \frac{1}{2\pi R_{ii}^2} \mathrm{e}^{-\frac{(z_i - h_i(x))^2}{2R_{ii}^2}} \tag{2}$$

where $l(\cdot)$ and $f(\cdot)$ are the likelihood function and probability distribution function of the conditional probability of the measurement error given the state, respectively; and R_{ii} is the diagonal elements of $\boldsymbol{R}$. Working with the above equation, the estimated state $\hat{\boldsymbol{x}}$ can then be obtained by the unconstrained minimization problem:

$$\hat{\boldsymbol{x}} = \min (\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x}))^{\mathrm{T}} \boldsymbol{R}^{-1} (\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x})) \tag{3}$$

B. Tracking State Estimation

The fast sampling of PMUs (from 1 to 120 samples per second) and synchronization via global positioning system (GPS) time stamps are strong motivations for formulating the state estimation as a time-dependent problem. Besides the PMU measurements, there are also measurements gathered from the SCADA system, which, thereby, still requires a nonlinear measurement model in order to include all the information available in the estimation.

Reference [2] provides a comprehensive definition of different methods and objectives to include temporal aspects in the estimation problem. Basically, they can be separated in:

1) Dynamic estimators, which encompass differential equations, generator models, speed and excitation controllers of

synchronous machines. These estimators aim at estimating the dynamic state of the network and the internal variables of loads and generators.

2) Forecasting-aided state estimators, which use a forecasting model to forecast the states in a subsequent instant through smoothing functions for the state transition. These estimators aim at the algebraic state and the complex nodal voltages of the network.

3) Tracking state estimators, a particular case of the previous one where variations are assumed small enough, so that the forecasting corresponds to the last estimated state.

In this paper, we explore the tracking state estimation problem, since it is one of the most affected by abrupt state changes [2]. This approach may lead to fewer modifications in static estimators (and network database), while obtaining significant gains in terms of accuracy, which is a practical advantage. Tracking state estimation consists in the following discrete time-variant nonlinear model:

$$\begin{cases} \boldsymbol{x}_t = \boldsymbol{x}_{t-1} + \boldsymbol{\omega}_t \\ \boldsymbol{z}_t = \boldsymbol{h}(\boldsymbol{x}_t) + \boldsymbol{e}_t \end{cases} \tag{4}$$

where $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t-1}$ are the state variable vectors ($n \times 1$) at instant t and the previous one $t-1$, respectively; $\boldsymbol{z}_t$ is the measurement vector ($m \times 1$) at instant t; $\boldsymbol{e}_t$ is the random error vector ($m \times 1$) that represents the measurement noise at instant t, with zero mean and known covariance matrix $\boldsymbol{R}$ ($m \times m$); and $\boldsymbol{\omega}_t$ is the random error that represents process noise with zero mean and known process noise covariance matrix $\boldsymbol{Q}$ ($n \times n$). Traditional tracking state estimation also assumes the noise vectors as multivariate Gaussian independent random variables, and the state space corresponds to a hidden Markov model, as described in (2) and illustrated in Fig. 1.

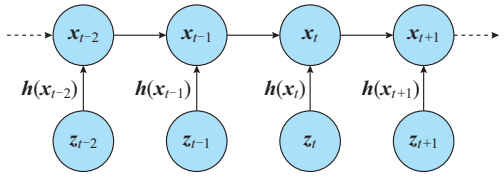

Fig. 1. Representation of hidden Markov model for tracking state estimation problem.

A notable conceptual trait in this formulation is that the state variables are now modeled with a random component $\boldsymbol{\omega}_t$. Despite the fact that static state estimation in (1) carries the notion of the estimator variance, it does not encompass the hypothesis of different possible values for the state within the inference problem. The solution of the estimation problem in such state space model corresponds to the well-known Kalman filter algorithm based on the WLS criterion [1]-[12]. It consists of a two-step recursive algorithm, with the following formulation for the tracking state estimation model:

1) Prediction step: the prior mean and covariance matrix are given by:

$$\hat{\boldsymbol{x}}_{t|t-1} = \hat{\boldsymbol{x}}_{t-1} \tag{5}$$

$$\boldsymbol{P}_{t|t-1} = \boldsymbol{P}_{t-1} + \boldsymbol{Q} \tag{6}$$

where $t|t-1$ stands for the conditional relation between two time-steps; $\boldsymbol{P}$ is the calculated state covariance matrix ($n \times n$).

2) Updating step: the Kalman filter gain matrix $\boldsymbol{K}_t$ ($n \times m$), the estimated posterior state $\hat{\boldsymbol{x}}_t$ and posterior covariance $\boldsymbol{P}_t$ are then updated:

$$\boldsymbol{K}_t = \boldsymbol{P}_{t|t-1}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{P}_{t|t-1}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R})^{-1} \tag{7}$$

$$\hat{\boldsymbol{x}}_t = \hat{\boldsymbol{x}}_{t|t-1} + \boldsymbol{K}_t(\boldsymbol{z}_t - \boldsymbol{H}\hat{\boldsymbol{x}}_{t|t-1}) \tag{8}$$

$$\boldsymbol{P}_t = (\boldsymbol{I} - \boldsymbol{K}_t\boldsymbol{H})\boldsymbol{P}_{t|t-1}(\boldsymbol{I} - \boldsymbol{K}_t\boldsymbol{H})^{\mathrm{T}} + \boldsymbol{K}_t\boldsymbol{R}\boldsymbol{K}_t^{\mathrm{T}} \tag{9}$$

where $\boldsymbol{H}$ is the Jacobian matrix ($m \times n$) of the measurement model; and $\boldsymbol{I}$ is the identity matrix ($n \times n$). The nonlinear version of the above model corresponds to the WLS extended Kalman filter (EKF) that performs linearization of the measurement equations, and the iterated EKF (IEKF) that updates such linearization iteratively, which presents more accurate results.

C. Information Theory Concepts

Based on the information theory, the concept of correntropy arises as a measure of similarity between two probability distribution functions (PDFs) [19]-[24]. It has a strong relation to the concept of entropy, which measures the information content of a PDF. The correntropy between two random variables A and B is defined in (10) for a finite number of samples.

$$V(A,B) = E(k_\sigma(A-B)) = \frac{1}{N}\sum_{i=1}^{N} k_\sigma(A_i - B_i) \tag{10}$$

where $E(\cdot)$ is the function to calculate the expectation; $k_\sigma(\cdot)$ is a kernel for the random variable in the sampled value; σ is the kernel bandwidth (or size of the Parzen window in the Parzen-Rosenblatt estimator [28]); and the subscript i represents the limited number N of samples of random variables. The concept of correntropy induces a distance function to measure the similarity of two PDFs, the correntropy induced metric (CIM). Depending upon the choice of kernel and size of the bandwidth, this metric may vary from the L_∞ to L_0 norms. Due to such property, it can be tuned to become insensitive to outliers, pushing it towards the L_0 norm (indifference in distance) [22].

Finally, the CIM distance motivates the creation of the maximum correntropy criterion (MCC) to perform inference. For instance, in a regression analysis, it aims at finding parameter values that maximize the similarity among a set of observations ($r_i = A_i - B_i$).

$$\hat{W} = \max_{W \in \Omega} \frac{1}{N}\sum_{i=1}^{N} k_\sigma(r_i) \tag{11}$$

where $\hat{W}$ is the optimal parameter value; and Ω is the feasible set for the parameters.

III. Tracking Power System State Evolution with MCEKF

A. State Evolution Concept in Power Systems

The temporal aspect of the state estimation problem and the probabilistic interpretation of the state variables provide a conceptual framework to evaluate the network condition under both measurement and system uncertainties. However, it has a limitation related to the parametric model typically

assumed for the process and measurement noises under conventional Kalman filter formulations. Many efforts have been made to explore non-Gaussian noise situations [2], [10]-[16] such as in the case of bad data, cyber-attacks or even particular noise characteristic of PMUs.

This paper proposes a new interpretation regarding the process noise towards a generalist non-Gaussian model based on the kernel density estimation principle. During a sequence of observations from the measurements, the state variables can present different possible values, which induces the proposed concept of state evolution. Such different values can be related to system transitions, which can be:

1) Systemic: due to sudden load variations, generation dispatching or controller actions.

2) Structural: due to contingencies, switching operations or changes in the network.

3) Random: due to intermittency or failures.

4) Induced: due to cyber-attacks or unsupervised switching.

An illustration of such transitions in a power system observed by a measurement set, monitored by a state estimation process is illustrated in Fig. 2. Due to the contingencies, changes in the controllers, natural load variation, and generation intermittency, each algebraic state of the network may present different values over time. The state evolution aims at capturing such different possible values within a single non-Gaussian PDF for the state.

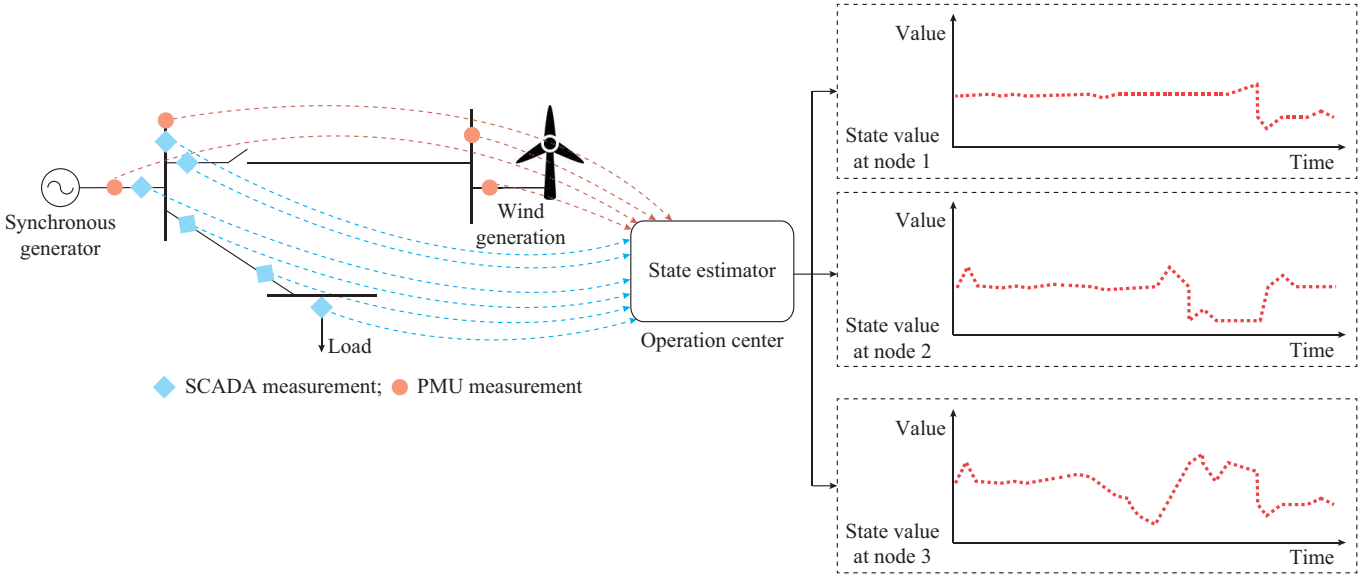

Fig. 2. Example of a power system and respective estimated state during different system transitions monitored by SCADA and PMU measurements.

The state evolution concept is defined by a set of state vectors, and each element is related to a respective set of observations at the instant t during an observation window Δt.

$$\boldsymbol{x}=\left\{\boldsymbol{x}_0,\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{t-1},\boldsymbol{x}_t,\boldsymbol{x}_{t+1},\ldots,\boldsymbol{x}_{\Delta t}\right\}=\boldsymbol{x}_t|_{t\in\Delta t} \tag{12}$$

Associated with this set of state vectors, the underlying PDF of the state can be obtained in a nonparametric manner with the introduction of a kernel density estimator. The Parzen-Rosenblatt method [28] for kernel density estimation then obtains the empirical PDF for the state through (13). Figure 3 illustrates this concept for a single state variable during different conditions of the network and the respective empirical probability distribution function.

$$\boldsymbol{x}\sim\frac{1}{\Delta t}\sum_{t\in\Delta t}k_\sigma(\boldsymbol{x}_t) \tag{13}$$

Since the goal is to capture the state transitions, an observation kernel is associated with the process noise through the state space model in (4).

$$\boldsymbol{\omega}_t\sim k_\sigma(\boldsymbol{x}_t-\boldsymbol{x}_{t-1}) \tag{14}$$

Along with the state vector modeled as a nonparametric random variable, this work also revises the hypothesis that the measurement noise follows a Gaussian distribution. In the case of PMUs, [10] has shown with experimental results that such a hypothesis does not hold and that more robust models are required for PMU-based state estimators.

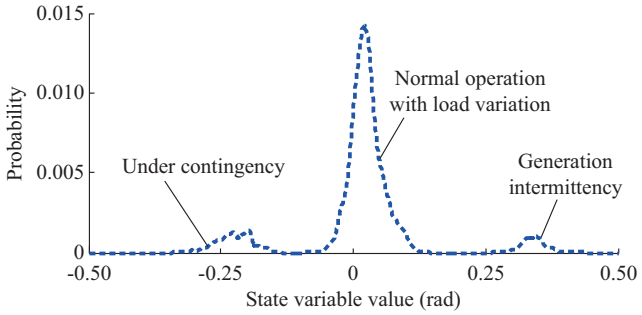

Fig. 3. State evolution concept for a single state variable (e. g., voltage phase angle) under different conditions obtained by kernel density estimation.

In a different perspective, [20] has shown that the presence of gross errors (bad data) consists of a non-Gaussian noise situation. To encompass such general probabilistic model for the measurement model, in the same way as for the state vector, the measurement noise is modeled under a kernel assumption related to each observation.

$$\boldsymbol{e}_t \sim k_\sigma(\boldsymbol{z}_t - \boldsymbol{h}(\boldsymbol{x}_t)) \tag{15}$$

B. EKF Based on MCC

The MCC Kalman filter was first introduced in [19], which aims at performing inference under a state space model without the assumption of Gaussian noise. The only assumption on the noise vectors is that the second-order moment is known and given by:

$$E\left(\begin{bmatrix}\boldsymbol{\omega}_t\\ \boldsymbol{e}_t\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\boldsymbol{\omega}_t\\ \boldsymbol{e}_t\end{bmatrix}\right)=\begin{bmatrix}\boldsymbol{P}_{t-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{R}\end{bmatrix}=\begin{bmatrix}\boldsymbol{B}_P^{\mathrm{T}}\boldsymbol{B}_P & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{B}_R^{\mathrm{T}}\boldsymbol{B}_R\end{bmatrix} \tag{16}$$

where $\boldsymbol{B}_P$ and $\boldsymbol{B}_R$ are the Cholesky decomposition of the process and measurement covariance matrices, respectively. The state space in (4) is rescaled by the respective decomposition.

$$\boldsymbol{B}_P^{-1}\boldsymbol{x}_t = \boldsymbol{B}_P^{-1}(\boldsymbol{x}_{t-1} + \boldsymbol{\omega}_t) \tag{17}$$

$$\boldsymbol{B}_R^{-1}\boldsymbol{z}_t = \boldsymbol{B}_R^{-1}(\boldsymbol{h}(\boldsymbol{x}_t) + \boldsymbol{e}_t) \tag{18}$$

The MCEKF is then devised in [25] for nonlinear models in the context of power systems. This paper applies the maximum correntropy principle for tracking power system state evolution, which yields the following optimization problem for the i^{th} measurement and j^{th} state variable:

$$\max J_{MCC} = \sum_{i=1}^{m} k_\sigma(z_{i,t} - h_i(\boldsymbol{x}_t)) + \sum_{j=1}^{n} k_\sigma(x_{j,t} - x_{j,t-1}) \tag{19}$$

A Gaussian kernel shown in (13) is assumed to compose the process and measurement noise. It is worth mentioning that the assumption of the kernel function does not mean the same assumption for the PDF of the random variable.

$$k_\sigma(r_i) = \mathrm{e}^{-\frac{r_i^2}{2\sigma^2}} \tag{20}$$

Thereby, the estimation problem consists in the following optimization problem:

$$\max J_{MCC} = \sum_{i=1}^{m} \mathrm{e}^{-\frac{(z_{i,t} - h_i(\boldsymbol{x}_t))^2}{2\sigma^2}} + \sum_{j=1}^{n} \mathrm{e}^{-\frac{(x_{j,t} - x_{j,t-1})^2}{2\sigma^2}} \tag{21}$$

C. Numerical Solution Method

Since it is a nonlinear optimization problem, an iterative numerical procedure must be employed to find the solution, i.e., the state vector values at instant t. In this work, a modified-Newton algorithm [29] is implemented to find the solution. It is worth mentioning that it consists in a different solution method compared with the first MCEKF proposed in [25], which applies a fixed-point iterative method. This work employs a nonlinear numerical optimization method. The main advantage is that the problem can be solved under a fully optimization framework with a globally convergent method [29]. An initial condition for the state $\boldsymbol{x}_t^k = \boldsymbol{x}_t^0$ starts the method at iteration k. The gradient function is calculated as:

$$\frac{\partial}{\partial x_j} J_{MCC} = \sum_{i=1}^{m} \frac{1}{\sigma^2} \mathrm{e}^{-\frac{(z_{i,t} - h_i(\boldsymbol{x}_t^k))^2}{2\sigma^2}} (z_{i,t} - h_i(\boldsymbol{x}_t^k)) H_{ij} - \frac{1}{\sigma^2} \mathrm{e}^{-\frac{(x_{j,t}^k - x_{j,t-1})^2}{2\sigma^2}} (x_{j,t}^k - x_{j,t-1}) \tag{22}$$

where $H_{ij} = \partial h_i(\boldsymbol{x}_t^k)/\partial x_j$ is the element of the Jacobian matrix of the nonlinear measurement model calculated at a specific point $\boldsymbol{x}_t^k$. This can be rewritten in a matrix representation as:

$$\nabla J_{MCC}(\boldsymbol{x}_t^k) = \begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{H}\end{bmatrix}^{\mathrm{T}} \boldsymbol{D} \begin{bmatrix} x_{t-1} - x_t^k \\ z_t - h(x_t^k)\end{bmatrix} \tag{23}$$

where $\boldsymbol{I}$ is an identity matrix ($n \times n$); and $\boldsymbol{D}$ is a diagonal matrix ($(m+n)\times(m+n)$) with elements $D_{ii}(\boldsymbol{x}_t)$ given by:

$$D_{ii}(\boldsymbol{x}_t) = \begin{cases} \dfrac{1}{\sigma^2} \mathrm{e}^{-\frac{(x_{i,t}^k - x_{i,t-1})^2}{2\sigma^2}} & i < n \\ \dfrac{1}{\sigma^2} \mathrm{e}^{-\frac{(z_{i,t} - h_i(x_t^k))^2}{2\sigma^2}} & i \geq n \end{cases} \tag{24}$$

The Hessian matrix is calculated by:

$$\nabla^2 \boldsymbol{J}_{MCC}(\boldsymbol{x}_t^k) = -\begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{H}\end{bmatrix}^{\mathrm{T}} \boldsymbol{D}\left[\boldsymbol{I} - \boldsymbol{R}_p\right]\begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{H}\end{bmatrix} + \sum_{i=1}^{m} D_{ii}(\boldsymbol{x}_t) \frac{(z_{i,t} - h_i(\boldsymbol{x}_t^k))}{2\sigma^2} \frac{\partial^2 h_i(\boldsymbol{x}_t^k)}{\partial \boldsymbol{x}^2} \tag{25}$$

where $\boldsymbol{R}_p$ is a diagonal matrix ($(m+n)\times(m+n)$) with elements given by:

$$R_{pii} = \begin{cases} \dfrac{1}{\sigma^2}(x_{i,t}^k - x_{i,t-1})^2 & i < n \\ \dfrac{1}{\sigma^2}(z_i - h_i(\boldsymbol{x}_t^k))^2 & i \geq n \end{cases} \tag{26}$$

In order to create ascending directions (remind that the solution is the maximum correntropy), the optimization method requires an approximation of the Hessian matrix $\boldsymbol{M}_k$ that keeps negative definite during the entire optimization process. In this way, a good approximation candidate is given by (27), which neglects the second-order derivatives terms in (25).

$$\nabla^2 \boldsymbol{J}_{MCC}(\boldsymbol{x}_t^k) \approx \boldsymbol{M}_k = -\begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{H}\end{bmatrix}^{\mathrm{T}} \boldsymbol{D}\left[\boldsymbol{I} - \boldsymbol{R}_p\right]\begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{H}\end{bmatrix} < \boldsymbol{0} \tag{27}$$

With the gradient vector and the Hessian matrix approximation, it is possible to calculate the directions $\boldsymbol{p}_k$ followed by the maximization algorithm at each iteration k.

$$\boldsymbol{M}_k \boldsymbol{p}_k = \nabla \boldsymbol{J}_{MCC}(\boldsymbol{x}_t^k) \tag{28}$$

A backtracking algorithm is used to find the step-size α_k and satisfy the Armijo condition [29], which is the equivalent version for the maximization problems. This will result in a sequence of iterations expressed in (29) to find the optimal value $\hat{\boldsymbol{x}}_t$.

$$\boldsymbol{x}_t^{k+1} = \boldsymbol{x}_t^k + \boldsymbol{\alpha}_k \boldsymbol{p}_k \tag{29}$$

The convergence is given by the difference of the state vector values in two successive iterations and a pre-specified tolerance (in this work, 1.0×10^{-6}). This iterative solution corresponds to the Prediction Step of the MCEKF. With the estimated state, the process covariance matrix is also updated in the Updating Step of the MCEKF. It can be deduced as follows:

$$\boldsymbol{P}_t = (\boldsymbol{P}_{t-1}^{-1} + \boldsymbol{H}^{\mathrm{T}} \boldsymbol{R}^{-1} \boldsymbol{H})^{-1} \tag{30}$$

D. Suppression of Suspect Samples and System Transitions Through Parzen Window Adjustment

As shown in [19]-[24], the accuracy strongly depends on the strategy to choose the size of the Parzen windows or kernel bandwidth. The current practice on correntropy-based estimation is a reduction of the kernel bandwidth towards an indifference of the suspect samples (for instance, outliers). This approach of successively reducing the size of the Parzen windows to achieve convergence when dealing with correntropy is a widely-used method, which was firstly proposed for training mappers under correntropy and entropy cost criteria [30]. This process, sometimes referred to as "kernel annealing", has also been adopted for power system state estimation [26], [31]. The goal was to suppress the effect of gross errors in the estimation, encompassed as a particular case of non-Gaussian errors.

The effect of reduced bandwidth kernel is well-known and conceptually proved in [28]. In this situation, correntropy is reduced to the conditional expected value of the residue equal to zero for that sample. However, from the perspective of algorithmic optimization, this approach has some difficulties. The success of the approach depends on the starting point of iteration, and the flat starting voltage is no longer a universal good point. Also, in order to create ascending directions (remind that the solution is the maximum correntropy), the optimization method requires the Hessian (or its approximation) matrix to be negative definite during the entire optimization process and at the optimal value of $\boldsymbol{x}_t$ in (27). Hence, the matrix $\boldsymbol{I}-\boldsymbol{R}_p$ must be a positive definite matrix, which implies the following lower bound for the Parzen window size:

$$\begin{cases}(x_{i,t}^k - x_{i,t-1})^2 < \sigma^2 \\ (z_i - h_i(\boldsymbol{x}_t^k))^2 < \sigma^2\end{cases} \tag{31}$$

In practice, the reduction in the bandwidth cannot be severe, and is limited to a minimum size $\sigma_{\min}$. Thereby, such annealing strategy can create a bias on the suspect samples, even if it is small in the final estimate.

This work introduces a new strategy to update the Parzen window sizes of suspect measurements, by enlarging the bandwidth of the suspect samples while maintaining the size of the non-suspect samples. Shrinking the kernel means placing an outlier in a region with metric similar to L_0 (indifference). But a Gaussian kernel with σ large enough can reach the same effect of indifference, because it has a smooth slope close to zero with an almost constant value across the domain, and does not introduce any spikes in the optimization landscape. We take advantage of this property.

Note that under the MCEKF, the suspect samples can be either caused by gross errors and uncalibrated measurements or by large transitions of the system (encompassed in the state evolution concept). The effect is shown directly on the objective function of the MCEKF, given a suspect set $k \in S$:

$$\lim_{\sigma_k \to \infty} J_{MCC} = \sum_{i=1}^{m} \mathrm{e}^{-\frac{(z_{i,t}-h_i(\boldsymbol{x}_t))^2}{2\sigma_i^2}} + \sum_{j=1}^{n} \mathrm{e}^{-\frac{(x_{j,t}-x_{j,t-1})^2}{2\sigma_j^2}} \tag{32}$$

Expanding the summation for the suspect samples, we obtain:

$$\lim_{\sigma_k \to \infty} J_{MCC} = \sum_{k \notin S} \mathrm{e}^{-\frac{(z_{k,t}-h_i(\boldsymbol{x}_t))^2}{2\sigma_k^2}} + \sum_{k \in S} \lim_{\sigma_k \to \infty} \mathrm{e}^{-\frac{(z_{k,t}-h_i(\boldsymbol{x}_t))^2}{2\sigma_k^2}} + \sum_{j=1}^{n} \mathrm{e}^{-\frac{(x_{j,t}-x_{j,t-1})^2}{2\sigma_j^2}} = \sum_{k \notin S} \mathrm{e}^{-\frac{(z_{k,t}-h_i(\boldsymbol{x}_t))^2}{2\sigma_k^2}} + \sum_{k \in S} \mathrm{e}^{0} + \sum_{j=1}^{n} \mathrm{e}^{-\frac{(x_{j,t}-x_{j,t-1})^2}{2\sigma_j^2}} \tag{33}$$

And also,

$$\begin{aligned}\lim_{\sigma_k \to \infty} \frac{\partial}{\partial x_j} J_{MCC} = &\sum_{k \notin S} \frac{1}{\sigma^2} \mathrm{e}^{-\frac{(z_{k,t}-h_i(\boldsymbol{x}_t))^2}{2\sigma^2}} (z_{k,t} - h_i(\boldsymbol{x}_t)) H_{kj} + \\ &\sum_{k \in S} \lim_{\sigma_k \to \infty} \frac{1}{\sigma^2} \mathrm{e}^{-\frac{(z_{k,t}-h_i(\boldsymbol{x}_t))^2}{2\sigma^2}} (z_{k,t} - h_i(\boldsymbol{x}_t)) H_{kj} + \\ &\frac{1}{\sigma^2} \mathrm{e}^{-\frac{(x_{j,t}^k - x_{j,t-1})^2}{2\sigma^2}} (x_{j,t}^k - x_{j,t-1}) = \\ &\sum_{k \notin S} \frac{1}{\sigma^2} \mathrm{e}^{-\frac{(z_{k,t}-h_i(\boldsymbol{x}_t))^2}{2\sigma^2}} (z_{k,t} - h_i(\boldsymbol{x}_t)) H_{kj} + \sum_{k \in S} 0 + \\ &\frac{1}{\sigma^2} \mathrm{e}^{-\frac{(x_{j,t}^k - x_{j,t-1})^2}{2\sigma^2}} (x_{j,t}^k - x_{j,t-1})\end{aligned} \tag{34}$$

The portion of the derivatives related to samples in the suspect set equal to zero. Hence, this does not affect at all the rest of the optimization process and the estimated values. In this situation, only the other measurements and the process equations are used to find the state estimates. Figure 4 shows an illustrative example of kernel density estimation and the effect of reducing and enlarging the Parzen window size for suspect samples. An outlier is associated with a large bandwidth and the other observations keep their initial window sizes. The effect of suspect samples with large window sizes spreads along the real line (from $-\infty$ to $+\infty$), and does not influence the estimated PDF.

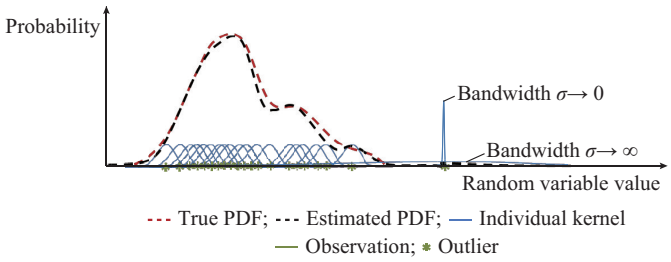

Fig. 4. Effect of different sizes of bandwidth (Parzen window sizes) for suspect samples on kernel density estimator.

If the detected event is a systemic transition, such as sudden load change, the suspect set becomes the process noise equations. In this scenario, only the Parzen windows related to the state space model are enlarged. The following equation in this situation is obtained by a similar demonstration:

$$\lim_{\sigma_j \to \infty} J_{MCC} = \sum_{i=1}^{m} \mathrm{e}^{-\frac{(z_{i,t}-h_i(\boldsymbol{x}_t))^2}{2\sigma_i^2}} + \sum_{j=1}^{n} \mathrm{e}^{0} \tag{35}$$

The portion of derivatives related to process equation in the state space then equals to zero. Hence, this does not affect the rest of the optimization process and the estimated values. It has an effect of momentarily neglecting the state transition equations, thus the estimates rely only on the current observed values, similar to the snapshot WLS state esti-

mator. In practice, this is equivalent to breaking the time dependency on the Markov chain, which has the same effect observed in [17] for treating non-stationary events.

Another situation is the simultaneous process and measurement noise due to structural transitions, caused by topological and parameter errors. In such a situation, both the process noise equations and the adjacent measurements of the affected part of the network become the suspect set, and their Parzen window sizes are enlarged.

A proper method for detecting and identifying suspect measurements is required. This paper will not treat these methods in detail, which can be obtained by a residual analysis, such as in [31] and [32] for bad data detection, or with other techniques based on artificial intelligence such as in [33] for system transitions. In this work, the standard normalized residual analysis [31], [32] triggers the identification of suspect samples for gross errors, while the system transitions are assumed as known. It is noteworthy that there are other and perhaps even more suitable ways to detect and identify suspect samples in the context of information theory, such as using the generalized correntropy along with the interior point method [22], [23].

The algorithm consists basically on the MCC Kalman filter recursive equations and the Parzen window update strategy. The window size updating consists in multiplying the bandwidth by a large value (1.0×10^4).

Algorithm 1: MCEKF with Parzen window update

Input: network data, network topology, measured data, $\boldsymbol{x}_0$, $\boldsymbol{P}_0$
Begin:
Initialize Parzen windows $\sigma_k = 10.0$
Set $\boldsymbol{x}_{t-1}$ as $\boldsymbol{x}_0$ and $\boldsymbol{P}_{t-1}$ as $\boldsymbol{P}_0$
For each t in Δt:
1) Rescale the state space according to (17) and (18)
2) Calculate $\boldsymbol{x}_t$ solving the MCC optimization in (21)
3) Update posterior state covariance according to (30)
4) Identify the suspect set S
5) Update the Parzen window sizes according to the type of event:
Bad data (gross error):
For each z_k with $k\in S$, $\sigma_k = 10000\sigma_k$
Return to 2)
Systemic transition:
For each x_j in $\boldsymbol{x}_t$, $\sigma_j^{t+1} = 10000\sigma_j$
Structural transition:
For each x_j in $\boldsymbol{x}_t$, $\sigma_j^{t+1} = 10000\sigma_j$
For each z_k with $k\in S$, $\sigma_k^{t+1} = 10000\sigma_k$
6) Update to next instant t and return to 1)
End

IV. Simulation Results

Monte Carlo simulations have been performed to evaluate and validate the proposed MCEKF. A sequence of load flow conditions creates the reference values for the state variables $x_{t,i}^{ref}$ and measurements $z_{i,t}^{ref}$. The random noise in the reference load flow values is considered in the simulation to obtain the measured values, and then state estimation is performed [34]. The noise characteristic is given based on the precision of the measurement pr_i (2% for SCADA and 0.1% for PMUs), and the i^{th} measured values are obtained using the following equation:

$$z_{i,t} = z_{i,t}^{ref} + u_i \frac{pr_i \left| z_{i,t}^{ref} \right|}{3} \tag{36}$$

where u_i is the underlying PDF of the noise. This equation provides the diagonal elements of the measurement noise covariance matrix $R_{ii} = pr_i \left| z_{i,t} \right| /3$. The process noise covariance matrix $\boldsymbol{Q}$ is a diagonal matrix with large values 1.0×10^3 to reduce the effect of the initial conditions on the estimation.

Different instants and the respective conditions are monitored by PMUs at a rate of 60 samples per second, while the SCADA measurements are updated at a rate of one sample per second. Each sample represents a different instant t and the respective measured values. Thus, the simulation consists of a sequence of quasi-stationary conditions with the estimation process triggered by the latest information available, sometimes monitored by SCADA and others by PMUs, with different time scales and random noise. The mean absolute error (MAE) evaluates the accuracy of the estimated state variables:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \left| \hat{x}_{t,i} - x_{t,i}^{ref} \right| \tag{37}$$

where $\hat{x}_{t,i}$ is the i^{th} state variable estimated. The MCEKF is then compared with two approaches based on the maximum likelihood principle: the WLS snapshot and the traditional WLS EKF.

A. Effect of Non-Gaussian Measurement Noise

The first test has been performed with the IEEE 14-bus test system to show aspects of non-Gaussian measurement noise under a stationary condition. The measurement set presented in Fig. 5 is used without process noise.

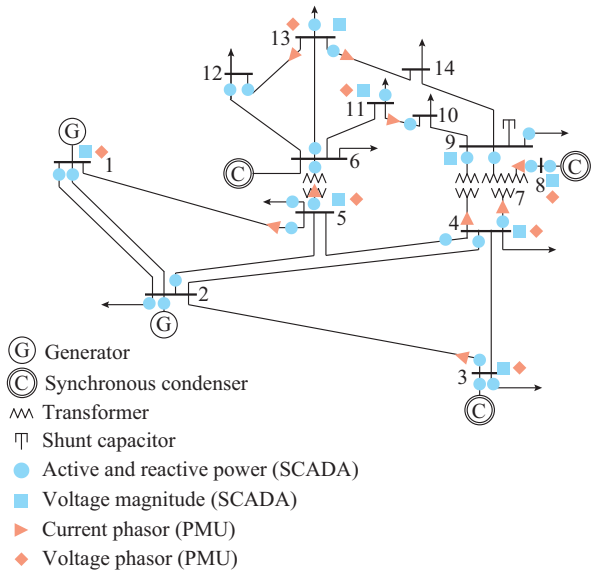

Fig. 5. IEEE 14-bus test system and respective SCADA and PMU measurement set for simulation.

A Monte Carlo simulation with the addition of non-Gaussian noise, in all SCADA and PMU measurements, evaluates the performance of the MCEKF during a sequence of 600

samples (equivalent to 10 s). The noise characteristic added in the simulation follows the Gaussian mixture in (38).

$$u_i \sim 0.7N(0,1) + 0.2N(3,3) + 0.1N(0,20) \tag{38}$$

Figure 6 shows the performance index over time in order to illustrate the effect of including the temporal aspect in the estimation. Note that the snapshot approach does not improve the estimation over time since it performs an independent estimation each instant only with the respective sample values, either from SCADA or PMUs. The traditional WLS EKF does present an improvement by including the temporal relation of the state; however, the MCEKF approach increases even further the accuracy, since it is more suitable to treat non-Gaussian noise characteristics.

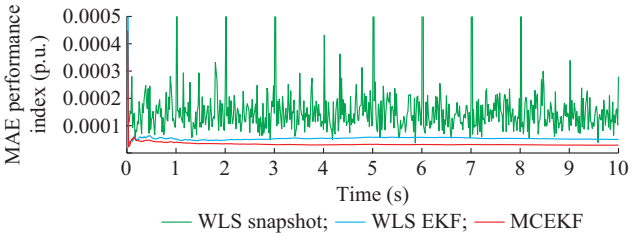

Fig. 6. MAE performance index over time.

Table I presents the comparison in terms of the overall MAE performance index. A different set of fixed values for the Parzen window sizes is also compared.

TABLE I
COMPARISON OF MAE PERFORMANCE INDEX FOR REAL AND IMAGINARY PARTS OF STATE VARIABLES

Method	Parzen window	MAE performance index (p.u.)	
		Real part of nodal voltage	Imaginary part of nodal voltage
WLS snapshot	-	2.5121×10^{-4}	0.5899×10^{-4}
WLS EKF	-	0.9387×10^{-4}	0.1941×10^{-4}
MCEKF	$\sigma_i = 0.1$ p.u.	14.6335×10^{-4}	23.0463×10^{-4}
	$\sigma_i = 1.0$ p.u.	1.6242×10^{-4}	0.2897×10^{-4}
	$\sigma_i = 10.0$ p.u.	0.9119×10^{-4}	0.1813×10^{-4}
	$\sigma_i = 100.0$ p.u.	0.9387×10^{-4}	0.1941×10^{-4}
	With Parzen update strategy	0.5545×10^{-4}	0.1024×10^{-4}

For the smaller window size, the lower bound is reached, and the estimation loses its performance. Another aspect is that by choosing larger size of the Parzen window, the MCEKF gets closer to the traditional WLS EKF. It is a known property in correntropy estimation that the CIM becomes the L_2 norm for high Parzen window size, which is similar to the WLS criterion. In this case, it is only with the Parzen window update strategy that more accurate estimation is achieved by suppressing the suspect set based on the largest residual.

B. Effect of Non-Gaussian Process Noise

The second test intends to evaluate the effect of process noise on the estimation. Process noise is added in the form of random load variation. In order to obtain a sensitivity of the effect of process noise, the percentage of load variation is increased gradually on this test, from 0% to 10.0%. Figure 7 presents the accuracy in terms of the estimation error empirical cumulative distribution function for the MCEKF. Table II summarizes the results for the three estimators.

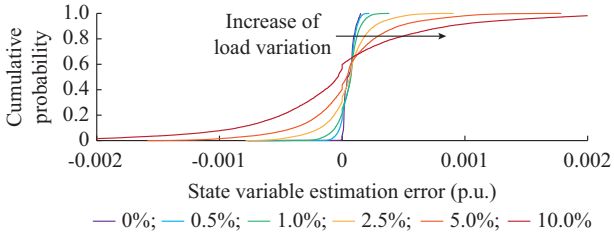

Fig. 7. Cumulative probability distribution of estimation error for MCEKF under different load variation (process noise) levels.

TABLE II
MAE PERFORMANCE INDEX FOR STATE VARIABLES WITH DIFFERENT LEVELS OF LOAD VARIATION

Load variation (%)	MAE performance index (p.u.)		
	WLS snapshot	WLS EKF	MCEKF
0.5	1.5510×10^{-4}	1.0368×10^{-4}	0.6748×10^{-4}
1.0	1.5510×10^{-4}	1.2205×10^{-4}	0.8533×10^{-4}
2.5	1.5509×10^{-4}	1.8705×10^{-4}	1.4698×10^{-4}
5.0	1.5507×10^{-4}	3.0779×10^{-4}	2.6412×10^{-4}
10.0	1.5503×10^{-4}	5.6942×10^{-4}	5.2730×10^{-4}

The effect of the process noise is a progressive increase in the estimator variance, up to the point that the most reliable estimator becomes the WLS snapshot instead of the WLS EKF or MCEKF approaches. It is noteworthy that large variations up to 5% of the load, in the short interval of time between two consecutive PMU samples (1 cycle), can be considered as a very abrupt and abnormal condition. Furthermore, such a scenario implies on a sensitive reduction of the premise that the last observation is a good estimation for the current one, the underlying hypothesis of the Markov chain in the tracking state space model in (4). In a typical transmission system, such scenario can be related to very large load variations or contingencies. In modern power systems, with the increase of intermittent distributed energy resources, such scenario can occur more frequently. However, despite the random variation that may occur, it is more likely that such large and abrupt changes happen less often in the time span of the tracking state estimation computing horizon.

This has motivated the pursuit for a window size tuning strategy that is able to suppress the effect of such large and abrupt variations. The proposition of this paper suppresses the negative effect of the systemic transition on the estimation accuracy by enlarging the Parzen window size of the process noise equation, as shown in Section III-C. The reasoning behind such strategy is that an abrupt change on the system state can be interpreted as a particular case of process noise with a heavy tailed PDF. Hence, the process equations become the suspect set and their Parzen window sizes

are enlarged. In order to evaluate this strategy, we consider 0.5% of random load variation and two sudden load changes in the simulation: ① an increase of all loads by 10% from $t=2.5$ s to $t=8.5$ s; ② a decrease of generation by 30% at node 2 from $t=5.8$ s to $t=7.2$ s. Figure 8 presents a performance comparison of the estimators regarding this scenario with a sudden load change. The effect is also compared for one of the estimated voltage magnitudes at nodes 2 and 14 in Fig. 9.

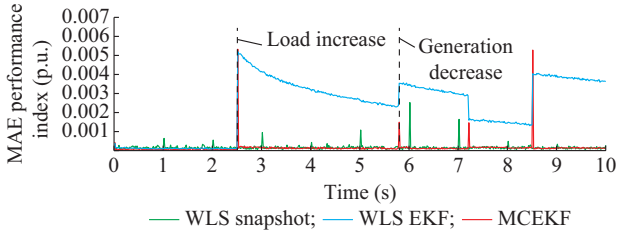

Fig. 8. MAE Performance index over time.

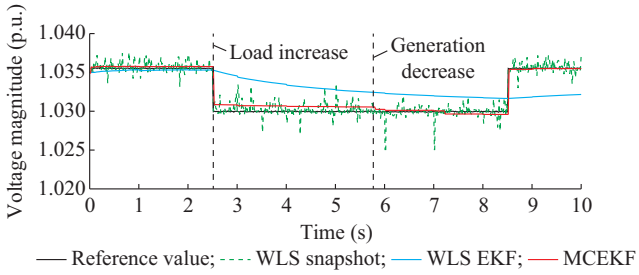

Fig. 9. Estimated voltage magnitude at node 14 during a load increase from $t=2.5$ s to $t=8.5$ s and generation decrease from $t=5.8$ s to $t=7.2$ s.

The effect of updating the Parzen windows related to the previous state consists in pushing the MCEKF momentarily towards the WLS snapshot. It counters the effect of the system transition by breaking the temporal relation of the Markov chain within the method. Otherwise, it would propagate a false assumption that the previous instant is a good approximation for the next one. Besides, it keeps a recursively improvement for estimating the steady state when the transition ends.

C. Suppression of Gross Errors

This third test on the IEEE 118-bus test system addresses the effect of gross errors and different measurement noise PDFs. In addition, the simulation considers the process noise with uniform load variation of 0.5%. The test assumes the non-Gaussian noise for the SCADA measurements in (38) and different characteristics for the PMUs according to the proportion and PDFs below: 80% of voltage phasor follows (39) and the remaining follows (40); and 50% of the current phasor follows (41) and the remaining follows (42).

$$u_i \sim 0.9N(0,1)+0.1N(0,20) \tag{39}$$

$$u_i \sim 0.6N(0,1)+0.3N(0.5,0.5)+0.1N(0,20) \tag{40}$$

$$u_i \sim Gamma(0.2,4) \tag{41}$$

$$u_i \sim Unif(-1,1) \tag{42}$$

The simulation consists also in a SCADA and PMU observable metering system with a 5 s horizon. It includes the following gross errors to illustrate the effect of bad data:

1) Addition of 30 standard deviations in PMU measurement V_{31} (voltage phasor measurement at node 31) at $t=0.6$ s.

2) Reduction of 30 standard deviations in PMU measurement $I_{63\text{-}59}$ (current phasor measurement at branch between nodes 63 and 59) from $t=1.2$ s to $t=1.5$ s.

3) Addition of 30 standard deviations in SCADA measurement $P_{76\text{-}118}$ and $Q_{76\text{-}118}$ (active and reactive power flow measurements at branch between nodes 76 and 118) at $t=2$ s.

4) Addition of 30 standard deviations in PMU measurement $I_{63\text{-}59}$ and $I_{49\text{-}66}$ from $t=2.5$ s to $t=2.8$ s.

5) Addition of 30 standard deviations in SCADA measurement V_{100} (voltage magnitude measurement at node 100) at $t=3$ s.

6) All the above errors simultaneously from $t=3.5$ s to $t=4.5$ s.

The MAE performance index keeps the same accuracy pattern as the previous tests, with an overall value of 3.5317×10^{-5}. In order to show the effect of the Parzen window update, Fig. 10 shows the largest normalized residue, initially before any size update, and after all updates are done. Note that in the final estimates, all normalized residues are less than the chosen threshold, which, therefore, has a reduced influence in the final estimates. It is noteworthy that there are instants when the measurement noise is enough to trigger the Parzen window update, since the noise PDFs have heavy tailed distributions.

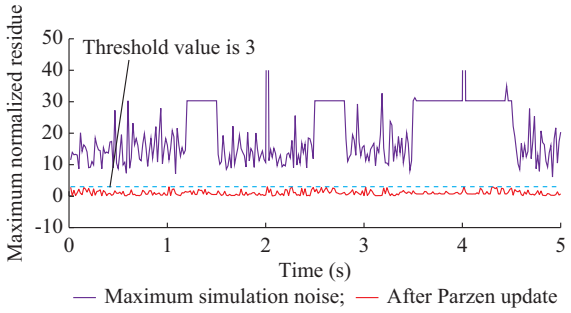

Fig. 10. Largest normalized residue after updating the Parzen window size for supression of effect of gross errros in MCEKF.

To illustrate the effect of the Parzen window update in the normalized residue, Fig. 11 shows the noise characteristics and respective normalized residue for measurement V_{31} with the MCEKF.

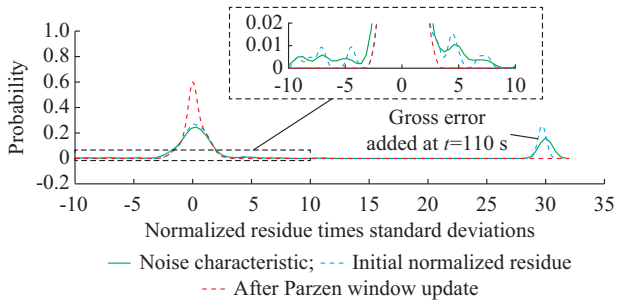

Fig. 11. Heavy tailed measurement noise with addition of gross error and respective normalized residue obtained by MCEKF.

The gross errors as well as the heavy tailed noise from the underlying measurement PDF are both properly suppressed by the Parzen window update strategy. This shows that the method can sustain accurate estimates in relatively large networks, even in the presence of a series of bad data contaminating the measurement vector.

D. Accuracy During Structural Transitions

This section evaluates the effect of structural transitions, i.e., abrupt changes of the topology of the system that has a large impact on both state and measurement models. The objective is to illustrate the state evolution concept and the state tracking through normal and abnormal network conditions in the Brazilian interconnected system (BR107). The network consists of 107 nodes along with a hypothetical set of SCADA and PMU observable sets. The simulation emulates a voltage instability situation. It starts with 1 s of steady state with random load variation of 0.5%, followed by an increase of load at a rate of 0.5% per second. Then, a sudden load increase of 5% in all loads occurs at $t = 2.5$ s. After another 0.5 s, a contingency occurs in one of the 500 kV transmission lines that connect the south to the southeast region. Following the contingency, the load ramp increases to 1.0% per second, and voltage instability occurs in less than 1 s.

Figure 12 illustrates the empirical PDF for the estimated voltage magnitude at two substations of the system. The MCEKF provides an accurate estimation of the state PDF that can be used in further analysis, for instance, to calculate probabilistic voltage stability margins in real time. Besides, the MCEKF presents a lot of flexibility by properly choosing the Parzen window sizes, which can suppress the negative effects of both systemic and structural transitions on the estimation accuracy. Such flexibility motivates the further exploration of improved tuning methods for the Parzen windows.

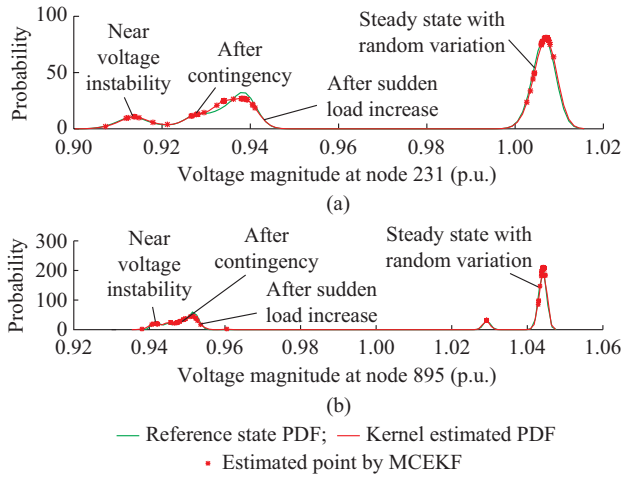

Fig. 12. Tracking state evolution PDFs for nodes 231 (minimum voltage) and 895 (terminal from transmission line that is switched off). (a) Node 231. (b) Node 895.

E. Computation Aspects

Finally, the computation aspects of the previous simulations are presented in Table III. The tests are performed using a microcomputer with a Core i7 3.60 GHz, 16 GB RAM with C programming language. The MCEKF shows good convergence characteristics in the power system state estimation problem. However, the computation burden must be taken into account for real-time applications, especially when PMUs has sampling rate up to 1 per cycle. The processing time shown in Table III is obtained without any concern on efficient programming and without using sparse matrix techniques or other computing efficiency resources. Therefore, they should be regarded only as an indication of the feasibility of the technique.

TABLE III
COMPUTATION ASPECTS FOR MCEKF

Test system	Average processing time (ms)	Iterations with SCADA measurement	Iterations with PMU measurement
IEEE 14-bus	31.0	4-5	2-3
IEEE 118-bus	360.0	5-7	2-3
BR107	313.0	5-7	2-3

V. CONCLUSION

This paper has proposed a tracking state estimator based on an EKF under MCEKF that deals with both SCADA and PMU measurements. A new update strategy for Parzen window suppresses the effect of suspect samples related to gross errors, process noise as well as system transitions. The importance of expanding the static state estimation problem to include the temporal aspects is emphasized in this work.

Even though many efforts have been made towards dynamic models, it is clear that currently implemented static state estimators can benefit a lot from simple modifications that leads to the tracking model. However, special attention must be taken for possible system transitions, such as systemic and structural transitions. This has been accomplished in this work by treating the measurement and process noises within a non-Gaussian kernel density estimation.

In this sense, the MCEKF outperforms the WLS snapshot, since it is based on a state space model. It also outperforms the conventional WLS EKF for treating non-Gaussian process and measurement noise. With the proposed concept of state evolution, the state variables also gain a non-parametric interpretation that can be further used in post-processing evaluation of the condition of the network under a probabilistic framework.

The method still depends on a proper suspect sample detection and identification process. In this sense, future work includes the development of feature extraction methods to automatically trigger the Parzen window update strategy based on a broader notion of the residual analysis with Bayesian inference concepts. An important direction is to increase the robustness of the method by adopting the interior point method as solver, instead of the Newton-Raphson iteration method. Finally, the model should be extended for dynamic estimation, to treat generators, loads and controllers, and evaluate detailed transient events.

REFERENCES

[1] A. Abur and A. Gomez-Exposito, *Power System State Estimation: Theory and Implementation*. New York: CRC Press, 2004.

[2] J. Zhao, A. Gomez-Exposito, M. Netto *et al*., "Power system dynamic state estimation: motivations, definitions, methodologies and future work," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 3188-3198, Jul. 2019.

[3] A. S. Debs and R. E. Larson, "A dynamic estimator for tracking the state of power system," *IEEE Transactions on Power Apparatus and Systems*, vol. 93, no. 7, pp. 1581-1588, Sept. 1970.

[4] R. D. Masiello and F. C. Schweppe, "A tracking static state estimator," *IEEE Transactions on Power Apparatus and Systems*, vol. 90, no. 3, pp. 1025-1033, May 1971.

[5] D. M. Falcão, P. A. Cooke, and A. Brameller, "Power system tracking state estimation and bad data processing," *IEEE Transactions on Power Apparatus and Systems*, vol. 101, no. 2, pp. 325-333, Feb. 1982.

[6] S. Sarri, L. Zanni, M. Popovic *et al*., "Performance assessment of linear state estimators using synchrophasor measurements," *IEEE Transactions on Instrumentation & Measurement*, vol. 65, no. 3, pp. 535-548, Mar. 2016.

[7] L. Fan and Y. Wehbe, "Extended Kalman filtering based real-time dynamic state and parameter estimation using PMU data," *Electric Power Systems Research*, vol. 103, pp. 168-177, Oct. 2013.

[8] G. Valverde and V. Terzija, "Unscented kalman filter for power system dynamic state estimation," *IET Generation, Transmission & Distribution*, vol. 5, no. 1, pp. 29-37, Jan. 2011.

[9] A. Sharma, S. C. Srivastava, and S. Chakrabarti, "A cubature Kalman filter based power system dynamic state estimator," *IEEE Transactions on Instrumentation & Measurement*, vol. 66, no. 8, pp. 2036-2045, Aug. 2017

[10] S. Wang, J. Zhao, Z. Huang *et al.*, "Assessing Gaussian assumption of PMU measurement error using field data," *IEEE Transactions on Power Delivery*, vol. 33, no. 6, pp. 3233-3236, Dec. 2018.

[11] J. Zhao, "Dynamic state estimation with model uncertainties using $H\infty$ extended Kalman filter," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 1099-1100, Jan. 2018.

[12] J. Zhao, M. Netto, and L. Mili, "A robust iterated extended Kalman filter for power system dynamic state estimation," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3205-3216, Jul. 2017.

[13] Y. Xu, L. Mili, and J. Zhao, "A novel polynomial-chaos-based Kalman filter," *IEEE Signal Processing Letters*, vol. 26, no. 1, pp. 9-13, Jan. 2019.

[14] J. Zhao and L. Mili, "A framework for robust hybrid state estimation with unknown measurement noise statistics," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 5, pp. 1866-1875, May 2018.

[15] A. Rouhani and A. Abur, "Linear phasor estimator assisted dynamic state estimation," *IEEE Transactions on Smart Grid*, vol. 9, no. 1, pp. 211-219, Jan. 2018.

[16] Y. Wang, Y. Sun, V. Dinavahi *et al*., "Adaptive robust cubature Kalman filter for power system dynamic state estimation against outliers," *IEEE Access*, vol. 7, pp. 105872-105881, Jul. 2019.

[17] J. A. D. Massignan, J. B. A. London, C. D. Maciel *et al.*, "PMUs and SCADA measurements in power system state estimation through Bayesian inference," in *Proceedings of 2019 IEEE PowerTech Conference*, Milano, Italy, Jun. 2019, pp. 1-6.

[18] M. B. D. C. Filho and J. C. S. de Souza, "Forecasting-aided state estimation Part I: panorama," *IEEE Transactions on Power Systems*, vol. 24, no. 4, pp. 1667-1677, Nov. 2009.

[19] B. Chen, X. Liu, H. Zhao *et al.*, "Maximum correntropy Kalman filter," *Automatica*, vol. 76, pp. 70-77, Feb. 2017.

[20] V. Miranda, A. Santos, and J. Pereira, "State estimation based on correntropy: a proof of concept," *IEEE Transactions on Power Systems*, vol. 24, no. 4, pp. 1888-1889, Nov. 2009.

[21] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: a localized similarity measure," in *Proceedings of 2006 International Joint Conference on Neural Networks*, Vancouver, Canada, Jul. 2006, pp. 4919-4924.

[22] S. Pesteh, H. Moayyed, V. Miranda *et al*., "A new interior point solver with generalized correntropy for multiple gross error suppression in state estimation," *Electric Power Systems Research*, vol. 176, pp. 1-13, Nov. 2019.

[23] S. Pesteh, H. Moayyed, and V. Miranda, "Favorable properties of interior point method and generalized correntropy in power system state estimation," *Electric Power Systems Research*, vol. 178, pp. 1-10, Jan. 2020.

[24] R. Meneghetti, A. S. Costa, V. Miranda *et al*., "Information theoretic generalized state estimation in power systems," *Electric Power Systems Research*, vol. 182, pp. 1-13, May 2020.

[25] Z. Zhang, J. Qiu, and W. Ma, "Adaptive extended Kalman filter with correntropy loss for robust power system state estimation," *Entropy*, vol. 21, no. 3, pp. 1-18, Mar. 2019.

[26] S. M. Mohiuddin and J. Qi, "Maximum correntropy extended Kalman filtering for power system dynamic state estimation," in *Proceedings of 2019 IEEE PES General Meeting*, Atlanta, USA, Aug. 2019, pp. 1-5.

[27] W. Ma, J. Qiu, X. Liu *et al*., "Unscented Kalman filter with generalized correntropy loss for robust power system forecasting-aided state estimation," *IEEE Transactions on Industrial Applications*, vol. 15, no. 11, pp. 6091-6100, Nov. 2019.

[28] V. Freitas, A. S. Costa, and V. Miranda, "Robust state estimation based on orthogonal methods and maximum correntropy criterion," in *Proceedings of 2017 IEEE PowerTech Conference*, Manchester, UK, Jun. 2017, pp. 1-6.

[29] J. Nocedal and S. Wright, *Numerical Optimization*, 2nd ed., New York: Springer, 2006.

[30] D. Erdogmus and J. C. Principe, "An error-entropy minimization algorithm for supervised training of nonlinear adaptive systems," *IEEE Transactions on Signal Processing*, vol. 50, no. 7, pp. 1780-1786, Jul. 2000.

[31] W. Wu, Y. Guo, B. Zhang *et al*., "Robust state estimation method based on maximum exponential square," *IET Generation, Transmission & Distribution*, vol. 5, no. 11, pp. 1165-1172, Nov. 2011.

[32] M. B. D. C. Filho, J. C. S. de Souza, and M. A. R. Guimaraes, "Enhanced bad data processing by phasor-aided state estimation," *IEEE Transactions on Power Systems*, vol. 29, no. 5, pp. 2200-2209, Sept. 2014.

[33] V. Miranda, P. A. Cardoso, R. J. Bessa *et al*., "Through the looking glass: seeing events in power systems dynamics," *International Journal of Electrical Power & Energy Systems*, vol. 106, pp. 411-419, Mar. 2019.

[34] R. Singh, B. C. Pal, and R. B. Vinter, "Measurement placement in distribution system state estimation," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 668-675, May 2009.

Julio A. D. Massignan received the B.S. and M.S. degrees in electrical engineering, in 2011 and 2016, respectively, from the University of São Paulo, São Carlos, Brazil, where he is currently working toward the Ph.D. degree in power systems engineering. His research interests include state estimation for distribution and transmission systems, nonlinear optimization, Bayesian inference and probabilistic models, and applications related to smart grids.

João B. A. London Jr. received the B.S.E.E. degree from the Federal University of Mato Grosso, Cuiaba, Brazil, in 1994, the M.S.E.E. degree from São Carlos School of Engineering (EESC), University of São Paulo, São Paulo, Brazil, in 1997, and the Ph.D. degree from University of São Paulo, São Paulo, Brazil, in 2000, all in electrical engineering. He is currently an Associate Professor with the EESC, University of São Paulo. His research interests include power system state estimation, network reconfiguration, and application of sparsity techniques to power system analysis.

Vladimiro Miranda received the Ph. D. degree in electrical engineering from the Faculty of Engineering of the University of Porto (FEUP), Porto, Portugal, in 1982. In 1981, he joined FEUP and currently holds the position of Full Professor. He is President of INESC P&D Brasil, Brazil, and Associate Director of INESC TEC, Portugal, where he has been a Researcher since 1985. He holds scientific advisory positions in several institutions such as the IIT in Madrid, Spain, the IEE in San Juan, Argentina, IRESEN in Morocco, as well as in the Portuguese Army and Navy. He was the 2013 recipient of the IEEE PES Ramakumar Family Renewable Energy Excellence Award. His research interest includes application of computational intelligence to power systems.