

A Pre-Estimation Filtering Process of Bad Data for Linear Power Systems State Estimators using PMUs

M.Pignati, L.Zanni, S.Sarri, R.Cherkaoui, J-Y.Le Boudec, M.Paolone

École Polytechnique Fédérale de Lausanne
EPFL

Lausanne, Switzerland

Email: marco.pignati@epfl.ch

Abstract—The paper proposes a specific algorithm for the pre-estimation filtering of bad data (BD) in PMU-based power systems linear State Estimators (SEs). The approach is framed in the context of the so-called real-time SEs that take advantage of the high measurement frame rate made available by PMUs (i.e., 50 – 60 frames per second). In particular, the proposed algorithm examines PMU measurement innovations for each new received set of data in order to locate anomalies and apply countermeasures. The detection and identification scheme is based on: (i) the forecasted state of the network obtained by means of a linear Kalman filter, (ii) the current network topology, (iii) the accuracy of the measurement devices and (iv) their location. The incoming measurement from each PMU is considered reliable, or not, according to a dynamic threshold defined as a function of the confidence of the predicted state estimated by using an AutoRegressive Integrated Moving Average (ARIMA) process. The performances of the proposed algorithm are validated with respect to single and multiple bad data of different nature and magnitudes. Furthermore, the algorithm is also tested against faults occurring in the power system to show its robustness during these unexpected operating conditions.

I. INTRODUCTION

In Active Distribution Networks (ADNs), the energy resources (i.e., distributed generation, storage, loads, etc.) can be actively controlled by a suitable Energy Management System (EMS) in order to achieve specific operation objectives. Typical ones refer to optimal voltage control, management of line-congestion, fault detection and location, post-fault management, loss minimization, etc. (e.g., [1], [2]). These functionalities are significantly improved if the knowledge of the network state is available. The time frames for these functionalities vary between few hundreds of ms (fault detection and location) to few tens/hundreds of seconds (fast and slow voltage control and line congestions), so they might require the knowledge of the network state with relatively high refresh rates. This is driving the development of the so-called “Real-Time State Estimators” (RTSEs) for which a drastic reduction of the computational burden can be achieved using Phasor Measurement Units (PMUs) (e.g., [3], [4]).

In this context, state estimation results are consistent as long as the input data provided to the estimation algorithm are correct. In the real field, electromagnetic interferences or telecommunication system failures might lead to measurements whose error is larger than an acceptable bound compatible with the accuracy of both metering and communication systems. Such measurements are commonly called Bad Data (BD). Some of them are easy to be identified and eliminated

by using simple plausibility checks (e.g., negative voltage magnitude, measurements with several orders of magnitude of difference from the expected value, etc.) Obviously, not all types of BD are immediately detectable and they can feed the State Estimator which is not designed to cope with such additional errors. For the above mentioned reasons, a SE deployed in the real field, has to be coupled with a suitable feature able to detect and identify any type of BD [5].

The existing schemes of BD detection and identification can vary from geometrical (e.g. [6]) to statistical approaches (e.g. [7]). Many different techniques are nowadays available as summarized in [8]. They can be further classified in pre-estimation and post-estimation filtering processes according to the position in the state estimation chain, where the detection of the BD takes place. Pre-estimation procedures are usually based on the analysis of the measurement innovations, defined as the difference between the most recent set of measurements and a predicted one inferred by using previous estimations. Therefore, in pre-estimation BD approaches, the set of measurements is analyzed and filtered before proceeding with the state estimation.

Post-estimation BD approaches, instead, usually deal with the estimation residuals so that the identification of the BD is part of an iterative process that filters-down the potentially corrupted measurements and re-estimate the state until the whole set has been checked. A large number of papers have been published using post-estimation analysis (e.g., [9], [10]) and this method has been recognized as providing good performances. On the other hand, when dealing with real-time monitoring of power systems, where speed and time determinism are key aspects (e.g., [3], [4]), one can take advantage of a predicted system state in order to obtain a consistent set of measurements to be compared with the real ones.

Several pre-estimation algorithms have been described in the literature (e.g., [11]–[15]) but, to the best of the authors’ knowledge, none of these contributions have discussed the possibility of detecting BD using measurements coming from PMUs. Indeed, the large number of measurements streamed by these devices (i.e. up to 50 frame-per-second (fps) [16]) enables the possibility of having SEs with a high refresh rate and calls for the need of reliable and fast, thus non-iterative, BD detection algorithms. Moreover, pre-estimation filtering of BD requires to be robust concerning the presence of a sudden change in the system’s operating point due to faults (i.e. short circuits) or components inrush, providing crucial help to the SE to keep track of the state of the system in

these particular transient conditions. Again, to the best of the authors' knowledge, this aspect has not been largely treated by the literature on the subject.

Within the context of PMU-based linear state estimators, this paper proposes a pre-estimation filtering of BD. It examines the measurement innovations for each new received set of measurements in order to locate anomalies and apply countermeasures. The incoming measurement from a PMU is marked as reliable, or not, according to a dynamic threshold defined as a function of the confidence of the predicted state estimated by using an AutoRegressive Integrated Moving Average (ARIMA) (0,1,0) process model. A fast and reliable boolean logic routine is proposed to detect the presence of BD and discriminate it from measurements recorded during fault conditions. The detection and identification scheme is based on the following inputs: (i) the forecasted state of the network obtained by means of a linear Kalman filter, (ii) the current network topology, (iii) the accuracy of the measurement devices and (iv) their location.

The proposed contribution analyses the impact of single and multiple BD of different nature and magnitudes. Furthermore, the algorithm is also tested against faults to show its robustness and the help it provides to the state estimator during these peculiar operating conditions. The speed of the process itself is also taken into account since the BD processing is supposed to introduce a negligible time latency (in the order of few ms).

The structure of the paper is the following: Section II gives the mathematical background of the adopted Discrete Kalman Filter State Estimator based on measurements coming from PMUs only. Section III provides a detailed explanation of the developed BD algorithm. Section IV shows the validation of the algorithm. It provides the assessment of the algorithm's performances for different BD and during faults. A latency assessment is also provided in this section. Section V provides the final remarks and potential deployment of the proposed pre-estimation filtering of BD in existing SEs.

II. THE ADOPTED DISCRETE KALMAN FILTER

The proposed BD process is based on the pre-estimation filtering. It is based on a comparison between every new set of measurements and the corresponding set of predicted ones. In this paper it is supposed that measurements consist of phasors of bus voltages and nodal-injected current sent by a certain number of PMUs deployed in the network to guarantee its observability (e.g. [3]). The availability of bus voltage and nodal-injected current phasors allow the use of linear SEs. In this paper we have used a linear Kalman filter algorithm, that results in the so-called Discrete Kalman Filter (DKF) state estimation process.

A. Mathematical background

As known, the DKF is a dynamic estimator. This means that the state \mathbf{x}_k , namely the array composed by the voltage phasors in all the system's buses at a given time-step k , is obtained by taking into account information available from the previous time-step $k - 1$.

Assuming that measurements come from PMUs only, the measurement array can be defined as:

$$\mathbf{z} = [\mathbf{z}_V, \mathbf{z}_I]^T \quad (1)$$

where \mathbf{z}_V and \mathbf{z}_I are sub-arrays both composed by the real and imaginary part of the phasors of bus voltages and nodal-injected currents.

In general, the KF addresses the problem of estimating the state $\mathbf{x} \in \mathbb{R}^s$ (where \mathbb{R}^s is the s -dimensional set of real numbers, being $s = 3l$ where l is the number of network buses; we here consider a generic unbalanced 3 phase network) of a discrete-time system process governed by the following set of linear stochastic equations (e.g. [17]). We assume an ARIMA process model of order (0,1,0). In this case \mathbf{A} is a static matrix and $\mathbf{A} = [\mathbf{I}]$

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{w}_{k-1} \quad (2)$$

$$\mathbf{z}_k = \mathbf{H}\mathbf{x}_k + \mathbf{v}_k \quad (3)$$

where:

- $\mathbf{x}_k \in \mathbb{R}^s$ and $\mathbf{x}_{k-1} \in \mathbb{R}^s$ represent the system state at time-step k and $k - 1$, respectively. The state of a three-phase (3-ph) system is here defined as the nodal voltage phasors for each phase at each network bus.
- \mathbf{A} is a $s \times s$ matrix known as the state transition matrix. It links the system state at time-step $k - 1$ with the one at the current time-step k .
- $\mathbf{w}_{k-1} \in \mathbb{R}^s$ represents the process noise at time-step $k - 1$, it is assumed to be white and with a normal probability distribution $p(\mathbf{w}_{k-1}) \sim N(0, \mathbf{Q})$ where \mathbf{Q} is the so-called noise covariance matrix.
- $\mathbf{v}_k \in \mathbb{R}^m$ represents the measurement noise at time-step k , assumed to be white and with a normal probability distribution $p(\mathbf{v}_k) \sim N(0, \mathbf{R})$ where \mathbf{R} is the measurement noise covariance matrix. \mathbf{v}_k is also assumed to be independent from the process noise \mathbf{w} .
- \mathbf{H} is a $m \times s$ matrix that relates the measurements set and the system state, both referred to the current time-step k . In case \mathbf{z}_k is represented by nodal injected powers, the function that links the measurements with the system state is nonlinear. Using only nodal voltage and injected current phasors measurements, (i.e. as expressed in (1)), the above-mentioned function \mathbf{H} becomes linear if (2) and (3) are expressed in rectangular coordinates. In this case, \mathbf{H} , which is composed by the partial derivatives of the measurements as a function of the system state, consists of constant elements, namely: zeros, ones, and elements of the 3-ph compound admittance matrix [18].

As mentioned above, the estimator has to face and manage two main sources of error formally included in (i) the process noise covariance matrix \mathbf{Q}_k and (ii) the measurement noise covariance matrix \mathbf{R}_k . The former takes into account every approximation introduced in the process model, therefore it gives an indication of how much the filter trusts the model (2). The \mathbf{R}_k matrix, instead, is related to the noise \mathbf{v}_k added

by the measurement devices. The larger the coefficients of \mathbf{R}_k , the lower the filter trusts the measurements (3).

In this context, the DKF algorithm is composed by *prediction* and *estimation* equations. The prediction equations obtain an a-priori estimate \mathbf{x}_k^- of the true state \mathbf{x}_k , using the observations available up to and including time-step $k-1$. The estimation equations incorporate the new measurements obtained at time-step k into the a-priori estimate and obtain an improved a-posteriori estimate $\hat{\mathbf{x}}_k$ of the true state \mathbf{x}_k . In the following the consequent DKF prediction and estimation equations are given.

- 1) Prediction equations:

$$\mathbf{x}_k^- = \hat{\mathbf{x}}_{k-1} \quad (4)$$

$$\mathbf{P}_k^- = \mathbf{P}_{k-1} + \mathbf{Q}_k \quad (5)$$

- 2) Estimation equations:

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}^T (\mathbf{H} \mathbf{P}_k^- \mathbf{H}^T + \mathbf{R}_k)^{-1} \quad (6)$$

$$\hat{\mathbf{x}}_k = \mathbf{x}_k^- + \mathbf{K}_k (\mathbf{z}_k - \mathbf{H} \mathbf{x}_k^-) \quad (7)$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}) \mathbf{P}_k^- \quad (8)$$

where:

- \mathbf{P}_k^- is a $s \times s$ matrix that represents the a-priori estimate error covariance.
- \mathbf{P}_{k-1} and \mathbf{P}_k represent the a-posteriori estimate error covariance matrices at time-step $k-1$ and k , respectively.
- \mathbf{K}_k is a $s \times m$ matrix that minimizes the a-posteriori estimate error covariance. It is the so-called ‘‘Kalman gain’’.

III. THE BAD DATA PROCESS

A set of measurements might differ from the expected ones for two main reasons: a sudden change in the system operating point or BD. Generally speaking, there might be errors in the meter-communication or in the network configurations (e.g. faulty switch breaker status information). The algorithm here proposed only considers meter-communication errors and it is able to discriminate a BD from data recorded during a fault, a fast dynamic in the system or other unexpected conditions. The probability of simultaneous occurrence of a BD and a fault is neglected.

The proposed BD processing method might be seen as a process inserted in between the prediction and estimation parts of the DKF. It is divided in four basic steps as shown in Fig. 1. In the first step, called *detection of anomalies*, the difference between predicted and real measurements (innovation vector) is checked against a specific threshold that is analytically defined in what follows. If one or more anomalies are detected, the algorithm goes to the second step, otherwise the estimator goes ahead to the estimation part. In the second step, *fast dynamic vs. BD*, the algorithm infers whether the previously identified anomaly is due to a sudden change in the network

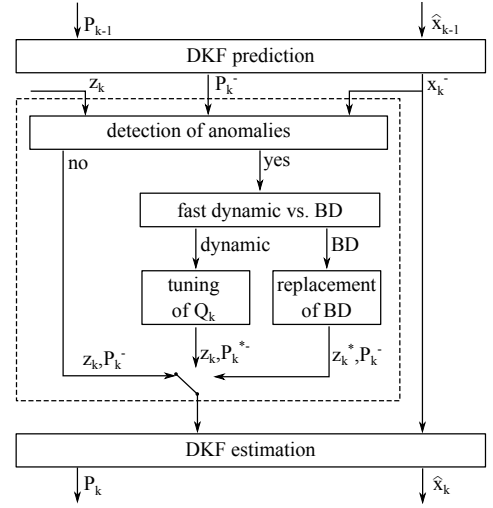


Fig. 1. Steps of the bad data algorithm process.

conditions (e.g., fault, inrush of a load, etc.) or to a set of incoming measurements affected by BD. In case a BD is identified, the algorithm proceeds to the final step, *replacement of BD*, where the wrong measurement is substituted with its predicted value (pseudo-measurement), obtained by using the ARIMA (0,1,0) process model combined with the outcome of the DKF. When the anomaly is found to be due to a sudden change in the network state or a fault in a bus, the algorithm enters in the *tuning of \mathbf{Q}_k* step in order to help the estimator to keep tracking the quickly changing state of the network.

A. Step #1: detection of anomalies

Let suppose the availability, at time-step $k-1$, of the a-posteriori estimated state $\hat{\mathbf{x}}_{k-1}$ and its associated error covariance matrix \mathbf{P}_{k-1} . A BD is now inserted between time-step $k-1$ and k so that the set of measurement \mathbf{z}_k is affected. By means of (4) and (5), \mathbf{x}_k^- and \mathbf{P}_k^- are obtained. The predicted set of measurements \mathbf{z}_k^- is then calculated as:

$$\mathbf{z}_k^- = \mathbf{H} \mathbf{x}_k^- \quad (9)$$

It is now possible to define the innovation vector \mathbf{v}_k :

$$\mathbf{v}_k = \mathbf{z}_k - \mathbf{z}_k^- \quad (10)$$

Knowing the measurement noise covariance matrix \mathbf{R}_k and the \mathbf{P}_k^- , it is possible to calculate the covariance matrix of the innovation terms that gives an indication on the confidence in the predicted measurements at time-step k :

$$\mathbf{S}_k = \mathbf{H} \mathbf{P}_k^- \mathbf{H}^T + \mathbf{R}_k \quad (11)$$

It is now possible to define a threshold for each element of the innovation vector that should not be exceeded during normal operating conditions. Therefore, each i^{th} element of \mathbf{v}_k should satisfy the following equation:

$$|v_k(i)| \leq \gamma \sigma_{S,i} \quad (12)$$

where:

- $|v_k(i)|$ is the absolute value of the i^{th} element of the innovations v_k ;
- $\sigma_{S,i} = \sqrt{\mathbf{S}(i,i)}$ is the standard deviation of the i^{th} predicted measurement;
- γ defines the confidence interval, usually taken equal to 3 for Gaussian noises;

It is important to note that the use of bus voltage and nodal-injected current phasors establishes a linear relation between measurements \mathbf{z}_k and state \mathbf{x}_k . Therefore, in the calculation of matrix \mathbf{S}_k , no approximations are introduced. This eliminates the need of including in γ the approximations introduced by the linearization process as done, for instance, in [13].

The first step of the detection of anomalies is therefore to check each $v_k(i)$ using (12) and raise a flag for those elements who do not satisfy it. In this way, the v_k is associated to a Boolean vector α_k that is filled with zeros or ones according to the absence or presence of anomalies, respectively.

B. Step #2: discrimination between fast dynamic and BD

The second step discriminates whether the previously identified anomalies are in fact BD or they are due to an unpredictable fast dynamic in the system. A reliable logic routine is implemented to distinguish between the two possibilities. The idea is simple: assuming that a sudden change in the system is occurring in a certain region of the network, an unexpected variation in correlated variables has to be detected by different PMUs, especially from those who are deployed in a region nearby the anomaly. Obviously, it is assumed that network topology and substations where the PMUs are installed, are known. A simple function takes as input the adjacency matrix¹ and the PMU positions and provides, for each PMU, the indices, in the innovation vector v_k , of the neighbor measurement devices. In this case, the term neighbor devices indicates the PMUs that are closer to the one whose measurements are under analysis. Each anomaly detected for bus voltage or nodal-injected current phasors in subsection III-A, is then checked against the corresponding quantity of the neighbor PMUs. If all the devices are recording the same anomaly, it is a symptom of the presence of a fast network dynamic at time-step k , otherwise the anomaly is marked as definitive BD. A simplified pseudo-algorithm that summarizes this process is provided in (13).

input: PMU index i and time slot k such that
 $\alpha_k(i) \neq 0$
 $J = \text{set of neighbors of } i$
if $(\forall j \in J, \alpha_k(j) \neq 0)$
 $z_k(i) = \text{dynamic};$
else $z_k(i) = \text{bad data};$

(13)

When the anomaly is detected on an injected current, the comparison described in (13) is performed with the corresponding voltage phasors of the neighbor PMUs. This is due to the relatively low influence that a fast dynamic has on the injected currents whereas the voltage profile is more affected by sudden changes in the power flows.

¹In power system, the adjacency matrix provides an indication of which are the neighboring nodes of each bus.

C. Step #3: replacement of BD

The previous step has enabled the discrimination between BD and sudden changes in the operating point of the network. The measurements that have been identified as wrong, are simply replaced with their prediction and the new set of measurements \mathbf{z}_k^* (see Fig. 1) is fed to the estimation part of the DKF. The variance associated to the pseudo-measurements is indeed larger than the one of the incoming set, because it includes the uncertainty added in the prediction step. Therefore, the outgoing \mathbf{P}_k , resulting from the state estimation at time-step k , will have larger values for those elements obtained by using the pseudo-measurements.

If the BD is not temporary, i.e. it is not cleared immediately, the uncertainty in the innovation terms, especially for those nodes whose state is strongly based on the pseudo-measurements, will eventually be so large that a BD could fall again within the confidence interval that defines the acceptable measurements and, therefore, it could be fed again to the DKF. The discussion of this specific point, however, is beyond the scope of this paper.

D. Step #4: tuning of process noise covariance matrix (\mathbf{Q}_k)

As anticipated, even when the anomalies turn out to be triggered by a fault, the algorithm has to take further actions before proceeding with the estimation part. As explained above, the \mathbf{Q}_k gives an indication of how much the DKF trusts its process model and thus its previous estimations. Provided that the process model is correctly defined, a diagonal \mathbf{Q}_k with elements in the order of $10^{-4} - 10^{-10}$ is usually selected (e.g., [19], [20]) meaning that high belief is given to the previous estimations to infer the next ones. A fault in the network, instead, eliminates completely the correlation between the previous estimation and the current state. In order to take into account the above considerations, it is necessary that, once a fault is detected, the elements of the \mathbf{Q}_k are increased to values large enough² to strongly rely on the measurements only until the estimator has tracked the new faulty state. At that point, the \mathbf{Q}_{k+n} can be decreased again to the pre-fault values trying to re-enhance the accuracy performance of the estimation. An appropriate value of n for a BD process coupled with PMUs and a DKF working at 50 fps is $n \geq 3$.

The BD algorithm, after having detected a fault and increased the \mathbf{Q}_k , has also to compute an updated version of the a-priori estimated covariance \mathbf{P}_k^{*-} (see Fig. 1) that incorporates the changes in the \mathbf{Q}_k matrix. The set of measurements can then be forwarded to the estimation part of the DKF.

IV. SIMULATIONS

The performances of the pre-estimation filtering of BD are tested on a fully unbalanced distribution network based on a modified version of the IEEE 13-bus test feeder shown in Fig. 2 [21]. The choice of a distribution system rather than a transmission network is driven by the large interest the PMUs are raising when installed in the challenging active distribution networks context (e.g., [22], [23]). From the point of view of the algorithm definition, there are no differences when dealing with transmission or distribution networks.

²Note that the assessment of the most suited value of \mathbf{Q}_k during fault conditions is beyond the scope of this paper.

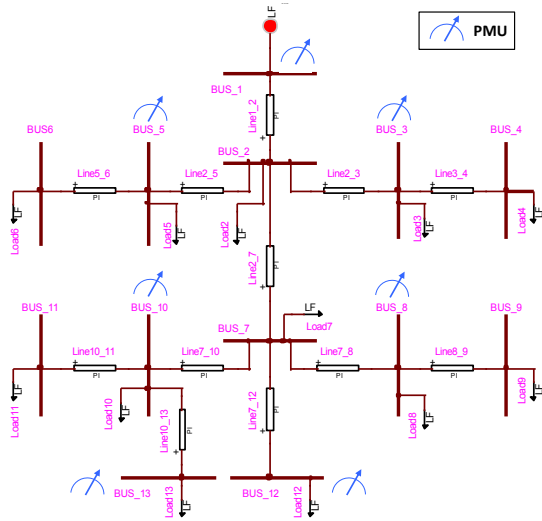


Fig. 2. The simulated IEEE 13-bus distribution test feeder.

Differently from what it is defined in [21], here the network rated voltage is assumed to be 15kV, the lines are unbalanced and correspond to the #602 line conductor configuration of [21]. With reference to Fig. 2, bus 1 represents the connection to the sub-transmission network. It is modeled as an ideal generator imposing the nominal voltage and frequency in series with its internal short-circuit impedance ($S_{sc} = 300MVA$, balanced $R_{sc}/X_{sc} = 0.1$).

The network is implemented inside a phasor-domain simulator to reproduce different test conditions and save the corresponding *true* phasors of voltage and currents. These sets of measurements are saved every 20 ms, so that it is possible to simulate incoming synchrophasors streamed at 50 fps from a certain number of PMUs deployed in the field (the PMU's characteristics are those described in [24]). A Gaussian noise with zero mean is added to the measurements in order to simulate the use of class 0.1 CT and VT as described in [25] and [26]. The noise added by the PMUs is neglected since it is a few orders of magnitude smaller than the one added by the sensors [24]. For this reason the PMU's accuracy is not inserted in the model of the system. As for the measurements, also the true state of the network is saved every 20 ms in order to allow an offline accuracy assessment of the SE performances when pseudo-measurements are inserted in the DKF process. The power profiles of loads and distributed generation (DG) are unbalanced and they come from measurements taken in a real distribution network located in the South-East region of Switzerland. Data refers to residential and commercial buildings for the loads and photovoltaic (PV) and mini-hydro power plant for the DGs.

Tests are run with the network in normal operating conditions and also when affected by fast dynamics like the ones characterizing single or multi-phase faults. The algorithm proposed in this paper has to identify the eventual presence of single or multiple BD during the normal operating conditions, and detect the sudden change in case of faults without replacing the unexpected, but correct, received measurements.

Both DKF and BD algorithms described in section II and III are fully implemented in LabVIEW. A library of

potential BD is developed in order to simulate the meter devices behavior when affected from several types of BD. The library includes: offset in magnitude and phase for each phasor of every PMU, single or multiple packet losses from one or more PMUs due to communication failures, drift in the phase estimation that typically affects PMUs when the time synchronization with the GPS signal is lost, etc.

A total of 7 PMUs are assumed to be installed in bus 1, 3, 5, 8, 10, 12, 13 and every PMU estimates 6 synchrophasors (three nodal voltages phasors and three injected/absorbed nodal current phasors). This number of PMUs guarantees the observability of the system. For simulation purposes, this number of PMUs, together with the absence of other meters, creates a system with very low redundancy: removing only one PMU from the set, would make the system not fully observable. We have decided to use this critical measurement configuration since no PMU data can be dropped without losing the observability of a certain area of the network. The importance of an effective replacement in case of a telecommunication system failure is therefore highlighted. The matrix R_k is diagonal and it is assumed to be constant. Each element is equal to $0.1 \cdot 10^{-6}$ that corresponds to the variance of real and imaginary part of the current and voltage sensor measurements [25], [26]. The elements of matrix Q_k are set to 10^{-9} during normal operating conditions, following a probabilistic assessment of the process-noise covariance matrix described in [20].

A. Tests and results

The algorithm is tested against several scenarios of BD present in everyday life of power systems. Some characteristic results are reported in what follows. All the plots are shown in per unit with reference to the following base voltage and current values: $V_b = 15kV$, $I_b = 666.6A$.

1) *Gross errors*: When a packet containing the synchrophasors estimated in a node is lost, the missing information in the z vector has to be replaced with *NaN* according to what specified in [16] and [27]. This test simulates a temporary communication failure from the PMU placed in node 1. At time $t = 0.5s$, all the information coming from bus 1 is lost and the communication is supposed to be re-established after 1s. The gross BD is identified with no delay and replaced with pseudo-measurements. Fig. 3 shows the replacement for real and imaginary part of the voltage phasors of phase *c*. A similar behavior is recorded for all the missing synchrophasors. As it can be seen, the replacement with pseudo-measurements, increases the confidence interval for which an incoming measurement is defined as acceptable or not. This is due to the growing uncertainty associated to the pseudo-measurements when they are constantly replacing the missing ones.

Fig. 4 shows the absolute value of the error in the estimation of the real and imaginary components of the state for the node affected by the specific BD and its neighbors. The error is plotted as a function of the node position and the simulation time. The packets are supposed to be missing between the two red dotted lines. As it can be noticed, the replacement does not worsen the accuracy in the estimated state for all the surrounding nodes since the errors of the estimated state after the BD event are comparable with those before the event. In case of damage of one of the PMU's components, the device can stream synchrophasors largely different from the real ones.

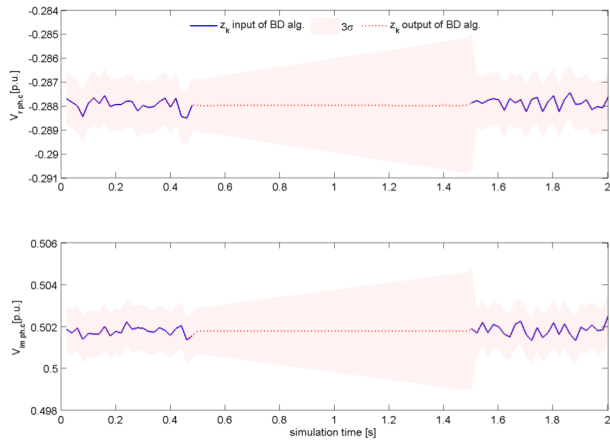


Fig. 3. Replacement of synchphasors due to temporary communication failure. The pink area represents the 3σ confidence interval.

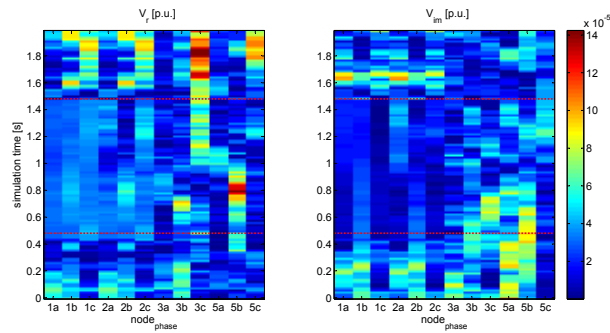


Fig. 4. Missing packets: Absolute value of the estimation error for real and imaginary part in node 1 and its neighbor buses.

This scenario was also tested and the algorithm performs well even in this case without affecting the quality of the estimations for the neighbor buses.

2) *Small deviations*: The algorithm is able to identify and detect BD even when the deviation from the true value is limited. Fig. 5 shows a temporary offset of 0.004 p.u. in the voltage magnitude of phase *b* streamed by the PMU in bus 10. The BD is properly detected and replaced.

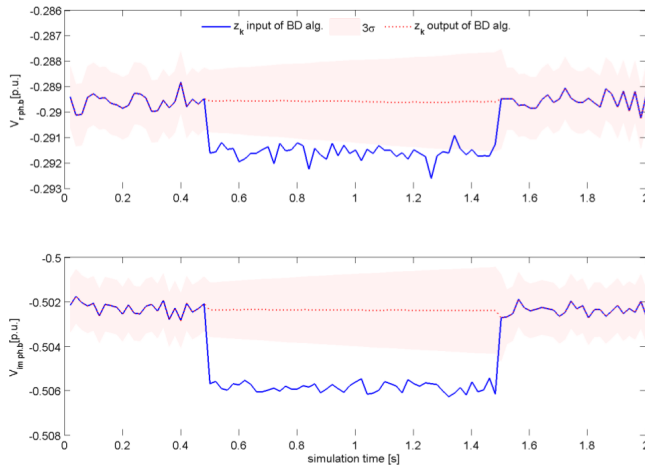


Fig. 5. Phasors replacement due to an offset in the voltage magnitude value.

3) *Phase Drift*: This is a BD specific for PMUs that could appear when the device loses its time synchronization with the GPS for a physical reason or an intentional jamming. It corresponds to a phase drift in all the synchphasors estimated by the PMU. The PMU deployed in node 10 is supposed to be affected by a deviation in the estimated phases of $1.5\text{mrad}/\text{sec}$. The synchronization is lost at time $t = 2\text{s}$ and the drift continues till $t = 18\text{s}$ when the PMU is re-synchronized with the GPS. As shown in Fig. 6, for the first two seconds, the BD does not detect the anomaly since the obtained residuals are within the bounds defined by (12). As soon as this condition is not satisfied, i.e. at $t = 4\text{s}$, the BD is correctly identified. Due to the slow deviation, the identification is not immediate and this results in a temporary worsening in the accuracy of the estimated state, especially for nodes 10 and 11 as shown in Fig. 7. The replacement of the wrong measurements brings then back the accuracy to its usual values even if the PMU is still not time-synchronized.

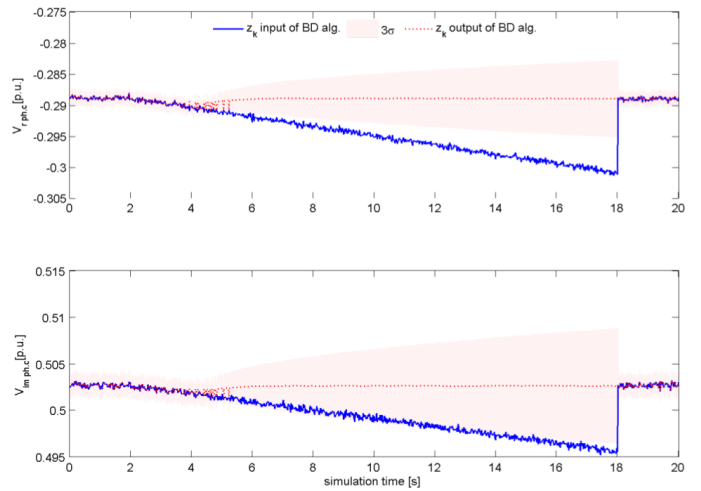


Fig. 6. Detection of a phase drift in node 10 due to GPS-synchronization issues.

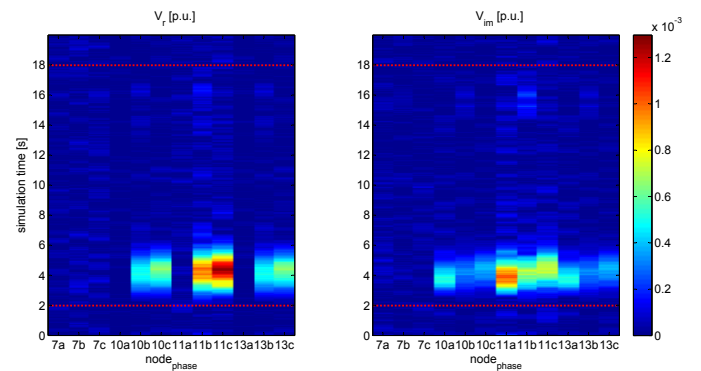


Fig. 7. Phase drift: Absolute value of the estimation error for real and imaginary part in node 10 and its neighbor buses.

4) *Multiple BD*: The presence of multiple BD coming from a single or several PMUs is also simulated. The algorithm correctly identifies and replaces them as long as they are not highly correlated. In fact, the presence of bad data on the same phasor quantities in all the PMUs delimiting a certain region, might deceive the algorithm of the existence of a fast dynamic

in the region itself. It should be noticed that such events are not likely to occur.

5) *Tracking of faults:* Tests with single and multi-phase faults on different buses are performed to check whether the BD algorithm takes the proper countermeasures described in section III-D. The case illustrated in Fig. 8 consists in a 3-ph fault on bus 7 with a fault impedance of 1Ω . The fault begins at time $t = 1s$ and it is assumed to be cleared after $500ms$. As it can be seen, the fault is correctly distinguished from a BD since it is not classified as such. According to the algorithm presented in [20], $Q_{t=1s}$ is increased to an appropriate value for $n=3$ iterations and then it is set back to the original value of 10^{-9} . The same happens at $Q_{t=1.5s}$. Thanks to help of the BD algorithm, the SE keeps following the state, with only a slight worsening in the accuracy performance immediately after the begin and the end of the 3-ph fault. This test is repeated for different faults and fault impedances with the same positive results.

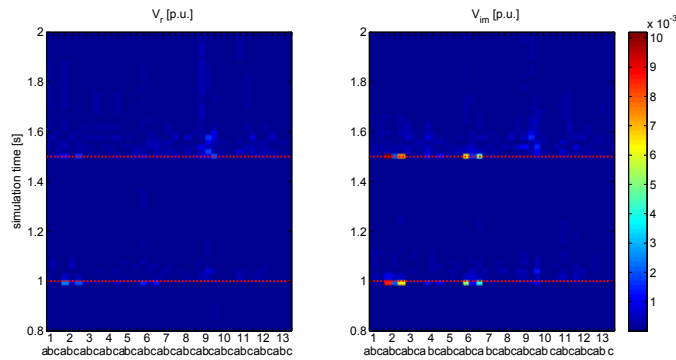


Fig. 8. Absolute value of the estimation error for real and imaginary part in every node, during a 3-ph fault at bus 7.

6) *Timing assessment:* Fig. 9 shows the cumulative distribution function (CDF) of the time added by the BD process described in this paper (red line) and also the CDF of the whole process (prediction, BD assessment and estimation). The test was run on a Intel core i7, 8gb of R.A.M. This result is only indicative and shows that BD detection obviously adds a latency to the SE. It can be seen that the latency is approximately one third of the total time that is taken to perform the complete estimation. The specific figure refers to the test where complete sets of incoming measurements are lost, and therefore replaced with pseudo-measurements, for one second. Obviously, the latency for both SE and BD algorithm is proportional to the number of measurements to be processed. In addition, the latency of the BD layer is also slightly dependent on the number of BD which are not likely to occur simultaneously.

V. CONCLUSION

Linear state estimators relying on PMU measurements might represent one of the main tools for the real-time monitoring of active distribution networks enabling the simplification of several control functionalities like: fault detection and location, post-fault management, optimal voltage control, management of line-congestion, etc.

In this context, the paper has proposed, discussed and validated a bad data pre-estimation filtering of PMU-based

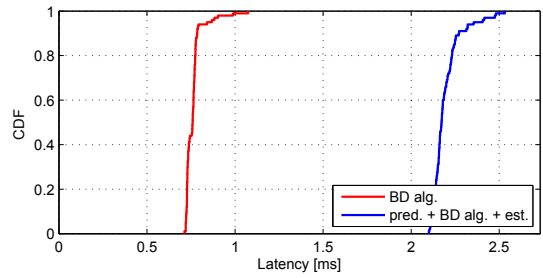


Fig. 9. CDF of the time latencies for the BD algorithm and the whole SE process.

linear state estimators. The proposed algorithm is coupled with a Discrete Kalman Filter (DKF) state estimator that uses only PMU measurements. The detection of potential anomalies relies on a set of dynamic thresholds inferred by predicted measurements using an ARIMA (0,1,0) process combined with the outcome of the adopted DKF state estimator. If a potential bad data is detected, it is compared with adjacent PMU data in order to verify whether it is a real bad data or a non-predictable fast dynamic appearing in the network (for instance, a fault). The paper has also proposed a way to replace the bad data by pseudo-measurements provided by a prediction process fed by the state estimation itself. The variance associated to these pseudo-measurements has been proposed to be computed using the a-priori error covariance of the DKF.

The performances of the proposed algorithm have been validated against single and multiple bad data of different nature and magnitudes, namely: gross errors, small synchrophasors-magnitude deviations and synchrophasors phase-drifts. Furthermore, the algorithm has been tested against faults to show its robustness during these unexpected operating conditions.

The performance assessment of the proposed algorithm allows to conclude that it appears robust and compatible with the real-time nature of the considered state estimation process enabling the use of this functionality also to observe fast network dynamics.

ACKNOWLEDGMENT

The research leading to these results has received funding from the NanoTera Swiss National Science Foundation project S^3 -Grids. The authors alone are responsible for the content of this paper.

REFERENCES

- [1] "IEEE Guide for Design, Operation, and Integration of Distributed Resource Island Systems with Electric Power Systems," *IEEE Std 1547.4-2011*, pp. 1–54, 2011.
- [2] J. P. Lopes, N. Hatzigiorgiou, J. Mutale, P. Djapic, and N. Jenkins, "Integrating distributed generation into electric power systems: A review of drivers, challenges and opportunities," *Electric Power Systems Research*, vol. 77, no. 9, pp. 1189 – 1203, 2007.
- [3] K. D. Jones, J. S. Thorp, and R. Gardner, "Three-phase linear state estimation using Phasor Measurements," in *IEEE Power and Energy Society General Meeting (PES)*, 2013, 2013, pp. 1–5.
- [4] M. Paolone, M. Pignati, P. Romano, S. Sarri, L. Zanni, and R. Cherkaoui, "A Hardware-in-the-Loop Test Platform for the Real-Time State Estimation of Active Distribution Networks using Phasor Measurement Units," in *Proceedings of the 2013 Cigré SC6 Colloquium*.

- [5] A. Abur and A. Expósito, *Power System State Estimation: Theory and Implementation*, ser. Power Engineering (Willis). Taylor & Francis, 2004.
- [6] N. Bretas, S. Piereti, A. Bretas, and A. Martins, "A Geometrical View for Multiple Gross Errors Detection, Identification, and Correction in Power System State Estimation," *IEEE Transaction on Power Systems*, vol. 28, no. 3, pp. 2128–2135, 2013.
- [7] T. Van Cutsem, M. Ribbens-Pavella, and L. Mili, "Hypothesis Testing Identification: A New Method For Bad Data Analysis In Power System State Estimation," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-103, no. 11, pp. 3239–3252, 1984.
- [8] A. Monticelli, "Electric power system state estimation," *Proceedings of the IEEE*, vol. 88, no. 2, pp. 262–282, 2000.
- [9] E. Handschin, F. Schweppe, J. Kohlas, and A. Fiechter, "Bad data analysis for power system state estimation," *IEEE Transaction on Power Apparatus and Systems*, vol. 94, no. 2, pp. 329–337, 1975.
- [10] A. Monticelli, "Reliable Bad Data Processing for Real-Time State Estimation," *IEEE Power Engineering Review*, vol. PER-3, no. 5, pp. 31–32, 1983.
- [11] D. Falcao, P. A. Cooke, and A. Brameller, "Power System Tracking State Estimation and Bad Data Processing," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-101, no. 2, pp. 325–333, 1982.
- [12] K. Nishiya, J. Hasegawa, and T. Koike, "Dynamic state estimation including anomaly detection and identification for power systems," *IEE Proceedings on Generation, Transmission and Distribution*, vol. 129, no. 5, pp. 192–198, 1982.
- [13] A. Leite da Silva, M. Do Coutto Filho, and J. F. De Queiroz, "State forecasting in electric power systems," *IEE Proceedings on Generation, Transmission and Distribution*, vol. 130, no. 5, pp. 237–244, 1983.
- [14] A. Abur, A. Keyhani, and H. Bakhtiari, "Autoregressive Filters for the Identification and Replacement of Bad Data in Power System State Estimation," *IEEE Transaction on Power Systems*, vol. 2, no. 3, pp. 552–558, 1987.
- [15] A. Leite da Silva, M. B. D. C. Filho, and J. M. C. Cantera, "An Efficient Dynamic State Estimation Algorithm Including Bad Data Processing," *IEEE Power Engineering Review*, vol. PER-7, no. 11, pp. 49–49, 1987.
- [16] "IEEE Standard for Synchrophasor Data Transfer for Power Systems," *IEEE Std C37.118.2-2011 (Revision of IEEE Std C37.118-2005)*, pp. 1–53, 2011.
- [17] G. Welch and G. Bishop, "An Introduction to the Kalman Filter," Chapel Hill, NC, USA, Tech. Rep., 1995.
- [18] J. Arrillaga, D. Bradley, and P. Bodger, *Power system harmonics*, ser. A Wiley-interscience publication. Wiley, 1985.
- [19] J. Zhang, G. Welch, G. Bishop, and Z. Huang, "A Two-Stage Kalman Filter Approach for Robust and Real-Time Power System State Estimation," *IEEE Transaction on Sustainable Energy*, vol. PP, no. 99, pp. 1–1, 2013.
- [20] L. Zanni, S. Sarri, M. Pignati, R. Cherkaoui, and M. Paolone, "Probabilistic assessment of the process-noise covariance matrix of discrete Kalman filter state estimation of active distribution networks," in *International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, Jul 2014.
- [21] W. Kersting, "Radial distribution test feeders," in *IEEE Power Engineering Society Winter Meeting, 2001.*, vol. 2, 2001, pp. 908–912 vol.2.
- [22] J. Liu, F. Ponci, A. Monti, C. Muscas, P. Pegoraro, and S. Sulis, "Optimal placement for robust distributed measurement systems in active distribution grids," in *IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2013*, May 2013, pp. 206–211.
- [23] M. Shahidehpour and M. Khodayar, "Cutting Campus Energy Costs with Hierarchical Control: The Economical and Reliable Operation of a Microgrid," *IEEE Electrification Magazine*, vol. 1, no. 1, pp. 40–56, Sept 2013.
- [24] P. Romano and M. Paolone, "Enhanced Interpolated-DFT for Synchrophasor Estimation in FPGAs: Theory, Implementation, and Validation of a PMU Prototype," *IEEE Transaction on Instrumentation and Measurement*, 2014.
- [25] "International Electrotechnical Commission. Instrument transformers-Part 1: Current transformers," *IEC Standard EN.60044-1*, 2003.
- [26] "International Electrotechnical Commission. Instrument transformers-Part 2: Inductive voltage transformers," *IEC Standard EN.60044-2*, 1997.
- [27] "IEEE Guide for Phasor Data Concentrator Requirements for Power System Protection, Control, and Monitoring," *IEEE Std C37.244-2013*, pp. 1–65, 2013.