Methodology for Performing Synchrophasor Data Conditioning and Validation

Kevin D. Jones, Member, IEEE, Anamitra Pal, Student Member, IEEE, and James S. Thorp, Life Fellow, IEEE

Abstract—A PMU only state estimator is intrinsically superior to its SCADA analogue with respect to performance and reliability. However, ensuring the quality of the data stream which leaves the linear estimator is crucial before establishing it as the front end of an EMS. One way to do it is by pre-processing the phasor data before it arrives at the linear estimator. This paper presents an algorithm for synchrophasor data conditioning and validation that fits neatly into the existing linear state estimation formulation. The methodology has been tested using field data obtained from PMUs installed in Dominion Virginia Power’s (DVP’s) EHV network. The results indicate that the proposed technique provides a computationally simple, elegant solution to the synchrophasor data quality problem.

Index Terms—Data Quality, Kalman Filter, Phasor Measurement Unit (PMU), State Estimation.

I. INTRODUCTION

The cornerstone of all energy management systems (EMS) are the algorithms that process raw power system data for computing the states of the system. The traditional non-linear state estimation techniques rely on supervisory control and data acquisition (SCADA) measurements for providing this raw data. The state estimator results in turn become the basis for other network applications such as real-time contingency analysis, reliability studies, and system alarming. With the continued proliferation of phasor measurement unit (PMU) technology in the electric utility industry, the transition of state estimation from a traditional non-linear formulation to one which is purely phasor based and linear is imminent.

A purely PMU based state estimator has considerable advantages over a purely SCADA based or a mixed (SCADA-and-PMU based) state estimator [1-2]. Firstly, being linear, the PMU only state estimator does not require any iteration. Secondly, it is free from the data scan that is required in conventional estimators. Thirdly, despite its formulation as a state estimation problem, the time-tagged data produces an estimate at such a fast enough rate that it can be considered to be truly dynamic. However, similar to the conventional state estimators, a PMU only state estimator also depends on a consistent, reliable stream of input data. Due to the streaming nature of the phasor data, downstream applications which use this data are vulnerable to network congestion, configuration errors, equipment failures, etc. Reference [3] highlights some of the data quality issues associated with PMU data, but does not provide any algorithm for its conditioning/validation. It is the objective of this paper to address concerns regarding quality of synchrophasor data by establishing a computationally simple and efficient model for cleaning them.

Kalman filtering [4] is one of the most popular techniques in modern controls and it has been extensively used in power system applications [5-7]. In [5], it was used in the development and testing of wide frequency range adaptive phasor and frequency algorithms. In [6], it was used for integrating phasor measurements in a wide area power system stabilizer (PSS) design. A two-stage Kalman filter approach for robust and real-time power system state estimation was developed in [7]. In this paper, a data conditioning algorithm is developed that is based on the Kalman filter and which is used for cleaning real field synchrophasor data. Two methods for pre-screening the PMU data before it reaches the linear estimator are introduced first. This is followed by a combination of filtering and smoothing techniques based on a “quadratic prediction model” developed in [8]. The main contribution of this paper is to show how this quadratic prediction model based Kalman filter is able to condition actual synchrophasor data obtained from the field.

As part of a DOE Demonstration Project, Dominion Virginia Power (DVP) has developed a PMU only state estimator. Accordingly, the techniques proposed in this paper are demonstrated using real synchrophasor data obtained from DVP. In compliance with FERC: CEII regulations, all identifying marks from the data/models have been removed. It is observed that the proposed approach mitigates drop-outs, outliers, and other data quality issues that decrease the value of the phasor data at a downstream location, at the source itself.

The rest of the paper is structured as follows. Section II describes two techniques for pre-screening the synchrophasor data before it is received by the PMU only state estimator. The filtering and smoothing process for data conditioning is developed in Section III. Section IV shows how the quadratic prediction based Kalman filter is able to overcome the
challenges faced by a traditional dynamic/tracking estimator. The application of the data conditioning algorithm on real synchrophasor data is described in Section V. The conclusion is provided in Section VI.

II. PRE-SCREENING OF SYNCHROPHASOR DATA

Similar to other algorithms that use SCADA data, purely synchrophasor data-based algorithms are also susceptible to data quality issues. Loss of GPS synchronization, incorrect PMU configuration, and communication network congestion are few of the problems commonly encountered in relation to phasor data. Two simple techniques for validating the quality of the incoming data even before it is received by the linear state estimator are described below.

A. Plausibility Checks

The first step for data conditioning involves passing the data through a series of online plausibility filters or sanity checks. It is common practice at DVP as a part of the commissioning process to authenticate the PMU data after an installation before the stream is connected to the operations center. By doing so, issues resulting from an incorrect configuration of the PMU, problems with the GPS clock or an incorrectly connected signal wire can be captured and immediately resolved. In addition to preventing unworthy data from being consumed by the application, the online algorithm for plausibility checks also makes engineers aware of data quality problems that will require manual intervention. Different types of plausibility checks that would be cause for alarm are:

- In-service buses having zero, near zero or negative voltage magnitude measurement readings
- In-service lines having zero or near zero current magnitude measurement readings
- Improper phase relationships in three phase systems
- Frames with the C37.118 DataValid bit asserted
- Frames with the C37.118 PMUSync bit asserted
- Frames with the C37.118 PMUError bit asserted
- Frames which have other problems communicated via the C37.118 status word

B. Evaluating Signal Quality Using Signal-to-Noise Ratio

The next step following the simple plausibility check is the evaluation of the Signal-to-Noise Ratio (SNR) of the input signal. Since reconstructing the original sinusoid would be difficult, the assumption made here is that the components of the phasor (magnitude and unwrapped, referenced phase angle) are DC signals. In such a scenario SNR is the mean of the signal divided by the standard deviation of the signal taken over a moving window [9] as shown in (1).

\[ \text{SNR}_{DC}(\text{in dB}) = 10 \log \left( \frac{\text{mean}}{\text{std. dev.}} \right) \]  

SNR evaluation not only indicates loose connections or potential hardware problems, but also helps diagnose equipment issues much more accurately than the raw voltage measurement. An example is shown in Fig. 1a and 1b, where the SNR magnitude and angle plot for a 500kV bus captured the faulty C-phase data (identified by the wider spread of the C-phase data in comparison to the other phases) days before the potential transformer (PT) actually failed. In this sense, SNR becomes a suitable candidate for setting alarm limits.

III. TECHNIQUES DEVELOPED FOR DATA CONDITIONING

The idea of a three phase linear state estimator using only synchrophasor data was originally proposed in [10]. Its high speed allowed it to appear dynamic, but being a static formulation, it considered each new frame as a separate problem. Therefore, it was impossible to detect/identify bad data using that model. An improvement to this model was made in [8] which identified the “quadratic” relationship between the past, present and future states given by:

\[ \dddot{x}(k + 1) + 3\dddot{x}(k) - 3\dddot{x}(k - 1) + \dddot{x}(k - 2) = f(k) \]
In (2), \( x \) refers to the individual state, \( \hat{x} \) (read as x-hat) denotes the estimated value of \( x \), while the symbol | is the “given” operator. Therefore, \( \hat{x}(k|j) \) reads as x-hat of \( k \) given \( j \). Equation (2) is based on the logic that for a linear increase in load at constant power factor (true for a power system at 30 times a second); the complex voltages (and currents, since currents are linear functions of voltage) will follow a quadratic trajectory with the next estimate depending on three previous estimates [8]. Using (2), two techniques for conditioning of synchrophasor data via Kalman filtering are developed here.

A. Filtering

The classical model of a Kalman filter is given in (3).

\[
x(k+1) = \Phi(k+1,k)x(k) + \Gamma(k+1,k)w(k) \\
z(k+1) = H(k+1)x(k+1) + v(k+1)
\]  

In (3), \( k \) is the discrete time step, \( x(k) \) and \( x(k+1) \) are the system states at time \( k \) and time \( k+1 \), respectively, \( z(k+1) \) is the actual measurement at time \( k+1 \), \( w(k) \) is the zero-mean Gaussian process noise at time \( k \), \( v(k+1) \) is the zero-mean Gaussian measurement noise at time \( k+1 \), \( \Phi(k+1,k) \) is the state transition matrix relating the transition of the state from time \( k \) to time \( k+1 \), \( \Gamma(k+1,k) \) is the disturbance transition matrix relating the transition of the disturbance from time \( k \) to time \( k+1 \), and \( H(k+1) \) is the measurement matrix at time \( k+1 \). Thus, \( k+1 \) is the current time stamp where the measurement arrives and the estimation is made. For optimal filtering we take the estimate of (3) and express it as a recursive relation in the Kalman filter notation as seen in (4a) and (4b).

\[
\hat{x}(k+1|k+1) = \Phi(k+1,k)\hat{x}(k|k) + K(k+1)(z(k+1) - \hat{z}(k+1|k)) \tag{4a}
\]

\[
\hat{z}(k+1|k) = H(k+1)\hat{x}(k+1|k) \tag{4b}
\]

In (4) \( K(k+1) \) is the Kalman gain. Equation (4) is solved using the standard Kalman filtering technique [11]. However, when applied to synchrophasor data, it can be further simplified as shown below.

Due to the nature of the quadratic prediction model, adjacent state vectors share two of the three state variables in common yielding an augmented state vector. This can be thought of as a moving window containing three snapshots of the system which moves forward only one snapshot at a time. Therefore, for predicting the next state based on (2), \( \hat{x}(k|k) \) and \( \hat{x}(k+1|k) \) can be expressed as,

\[
\hat{x}(k|k) = \begin{bmatrix} \hat{x}(k-1|k-1) \\
\hat{x}(k-2|k-2) \end{bmatrix} \tag{5a}
\]

\[
\hat{x}(k+1|k) = \begin{bmatrix} \hat{x}(k+1|k) \\
\hat{z}(k|k) \\
\hat{z}(k-1|k-1) \end{bmatrix} \tag{5b}
\]

Since the estimate of the future state depends on the three previous state estimates, for filtering purposes, it makes sense to depict \( \hat{x}(k|k) \) and \( \hat{x}(k+1|k) \) as \( 3 \times 1 \) matrices. Now, we know that \( \Phi(k+1,k) \) relates the \( k+1 \) state to the \( k \) state, i.e.

\[
\hat{x}(k+1|k) = \Phi(k+1,k)\hat{x}(k|k) \tag{5c}
\]

Therefore, based on (2), (5a), (5b) and (5c), \( \Phi(k+1,k) \) can be formulated as a constant as shown in (5d).

\[
\Phi(k+1,k) = \begin{bmatrix} 3 & -3 & 1 \\
1 & 0 & 0 \\
0 & 1 & 0 \end{bmatrix} \tag{5d}
\]

It is to be noted here that in (5a) and (5b), the states are the complex voltage and current measurements. Therefore, without any loss of generality, we can write (5e) and (5f).

\[
z(k+1) = x(k+1) + v(k+1) \tag{5e}
\]

\[
z(k+1) = \hat{x}(k+1|k) \tag{5f}
\]

Using (5b) and (5f) in (4b), we get \( H(k+1) \) as,

\[
H(k+1) = [1 \ 0 \ 0] \tag{5g}
\]

Thus, on substituting (5d) and (5g) in RHS of (4a) and (4b), respectively, a simplified model of the filtering technique will be developed as shown in (6a) and (6b).
\[
\hat{x}(k+1|k+1) = \begin{bmatrix} 3 & -3 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \hat{x}(k|k) + K(k+1)(z(k+1) - \hat{x}(k+1|k)) \quad (6a)
\]
\[
\hat{z}(k+1|k) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \hat{x}(k+1|k) \quad (6b)
\]

The second term in RHS of (6a) corresponds to the steady state observation residual. The Kalman gain \(K(k+1)\) is specified by the following set of equations:
\[
K(k+1) = P(k+1|k)H^T(k + 1)[H(k+1)P(k+1|k)H^T(k + 1) + R(k + 1)]^{-1} \quad (7a)
\]
\[
P(k+1|k) = \Phi(k+1,k)P(k|k)\Phi^T(k+1,k) + \Gamma(k+1,k)Q(k)\Gamma^T(k+1,k) \quad (7b)
\]
\[
P(k+1|k) = [I - K(k+1)H(k+1)]P(k+1|k) \quad (7c)
\]

As the individual states are uncorrelated\[12\], \(R(k+1) = R\) and \(Q(k) = Q\), i.e. both of them are scalar. Typical values of these two quantities for field data were found to be in the order of \(10^{-2}\). Also, the \(3 \times 3\) initial state vector \(\hat{x}(0|0)\) and the \(3 \times 3\) initial estimation-error covariance \(P(0|0)\) are null matrices. The reason for this is that the values of the initial state vector and the initial estimation-error covariance will not affect the simulations as long as it is acknowledged that the values for these will not be correct until after the first window has been fully populated with data.

Fig. 2 shows the application of this filter to a voltage phasor oscillation that was measured at 30 times a second. Blue stars are real synchrophasor data while red circles are optimal filtered estimates. The oscillation starts from top left and moves to bottom right. From the figure, it can be seen that the estimate catches up with the measured value within a few cycles. It also becomes clear that it would be very difficult to design a model that matches such observations. The fact that the optimal filtered estimate is able to track real-time data so accurately is proof of the power and efficacy of the quadratic prediction model on actual synchrophasor data.

![Fig. 2. Kalman filtering estimate of an actual voltage phasor oscillation](image)

**B. Smoothing**

The smoothing algorithm estimates the previous states of the system using current measurements. Mathematically, this means solving for \(\hat{x}(k|j)\) where \(j > k\)\[11\]. The advantage of smoothing in comparison to filtering is that by smoothing, quality of the estimates can be considerably improved\[13\]. The quadratic prediction model associated with synchrophasor data that had been used for filtering purposes also applies to the smoothing process. The model of the fixed-lag smoother that has been used in this paper was developed in\[14\]. It has a discrete time state equation in the form of a recursive Kalman filter with an augmented state vector, an associated augmented dynamical system, and an augmented measurement equation as seen in (8). More details about it can be found in\[11, 14\].

\[
\begin{bmatrix}
\hat{x}(k+1|k+1) \\
\hat{x}(k|k+1) \\
\hat{x}(k-1|k+1) \\
\vdots \\
\hat{x}(k+1-N|k+1)
\end{bmatrix} =
\begin{bmatrix}
\Phi(k+1,k) & 0 & 0 & \ldots & 0 \\
1 & 0 & 0 & \ldots & 0 \\
0 & 1 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 1 & 0
\end{bmatrix}
\begin{bmatrix}
\hat{x}(k|k) \\
\hat{x}(k-1|k) \\
\hat{x}(k-2|k) \\
\vdots \\
\hat{x}(k-N|k)
\end{bmatrix} +
\begin{bmatrix}
K(k+1) \\
K_1(k+1) \\
K_2(k+1) \\
\vdots \\
K_N(k+1)
\end{bmatrix}
\begin{bmatrix}
z(k+1) - H(k+1)\hat{x}(k+1|k)
\end{bmatrix} \quad (8)
\]

In (8), \(N\) is the window length, while \(K(k+1), K_1(k+1) \ldots K_N(k+1)\) are the gain matrices. Equation (8) can be solved recursively as shown in\[14\]. The matrices \(\Phi(k+1,k)\) and \(H(k+1)\) in (8) have same dimensions and values that were computed in the previous sub-section. A judicious choice of window length is important because although a higher value of \(N\) would introduce more delay between the received measurement and the smoothed state estimate, it will also significantly improve the quality of the estimate. The reason for this being that an estimate obtained using \(\hat{x}(k+1-N|k+1)\) will be intuitively better than one obtained using \(\hat{x}(k+1-N|k+1-N)\)\[11\]. For the simulations done in this paper \(N = 3\). In the next section, the performance of this quadratic prediction model based Kalman filter for conditioning real synchrophasor data obtained from the field is illustrated.

**IV. PERFORMANCE OF DATA CONDITIONING ALGORITHM**

The ultimate objective of this paper is to use the knowledge of the quadratic prediction model and the Kalman filter based filtering and smoothing techniques to improve the quality of the synchrophasor data stream provided to the linear state estimator. This means mitigating data problems like dropouts, randomly occurring outliers, repeated data, as well as other intermittent data problems. The ideal solution would be to replace “bad” measurements with estimates generated from previous measurements that have been deemed “good” quality. In accordance with this logic, the proposed algorithm takes action for each measurement at every time step based on...
the quality of preceding measurements. The following four scenarios encompass all data condition possibilities:

**Scenario 1:** If the algorithm receives good quality measurements at each time step for every signal in the stream, then the data conditioning algorithm needs to only compute for the optimal smoothed estimate using (8). This is the best case scenario and no further processing of the data is required.

**Scenario 2:** In the presence of manageable amount of bad data, the data conditioning algorithm replaces the bad data by the optimal predicted estimate calculated by the Kalman filter. This measure of good data progresses naturally through the smoothing window.

**Scenario 3:** In the case where there is not enough quality information for estimating value of the measurement, the algorithm reports the raw value (if available) with an “unreliable quality” flag. The resetting procedure developed in sub-section B performs this function.

**Scenario 4:** In the case of a discrete change in the network, the algorithm reports the data points as measured, with no conditioning until the smoothing window is completely filled with post-contingency raw data. This action is also performed by the resetting function described in sub-section B.

### A. Simulation Results – Data Drop-outs and Repeated Values

The difference between the current measurement and the optimal predicted estimate creates an observation residual. By comparing this residual with a pre-defined threshold, the data conditioning algorithm determines whether the data is suitable or not. Two occasions when the algorithm will receive unexpected data are: (1) the data is bad, or (2) a discrete change (contingency) occurs. In this sub-section, handling of bad data is analyzed.

If the observation residual is higher than the pre-defined threshold, then the algorithm will replace the measurement with the optimal predicted estimate calculated by the Kalman filter. This measurement will then be smoothed as it progresses naturally through the smoothing window. Thus, the algorithm will be able to provide reliable estimates in the presence of bad or missing data as long as it receives good data intermittently. Under such circumstances, the percentage of good data points received becomes a critical quantity. This is determined by the size of the smoothing window (N). If enough subsequent data points are not reported or are “bad”, then the algorithm cannot make accurate predictions.

In the simulations, the algorithm is subjected to varying levels of data drop-outs or unreported data (effectively 100% total vector error (TVE)). Uniformly distributed data drop-out rates of 1%, 2%, 5%, 10%, 20%, and 40% were analyzed in [15]. The drop-outs were simulated by removing data points from the stream probabilistically, i.e., for a 10% drop-out rate, the likelihood that a data point will be reported is 90%. On doing the analysis, it was realized that the algorithm is able to successfully provide quality performance up to drop-out rates of 20% (Fig. 3). As the drop-out rate increased further, the likelihood that adjacent frames were dropped increased substantially. This resulted in a corresponding decrease in the algorithm’s ability to make an accurate prediction as seen in the case with 40% data drop-outs (Fig. 4) where on numerous occasions the optimal smoothed estimate did not closely follow the raw data. However, since better data quality is demanded by other agencies (like regional transmission operators); data drop-outs of more than 10% are unrealistic. In Fig. 5, the data conditioning algorithm is applied to a portion of the nuclear oscillations example of Fig. 2, but with 30% drop-outs. In Fig. 5, red circles denote raw data points while green pluses denote optimal smoothed estimates. This figure clearly shows the effectiveness of this algorithm to condition actual phasor data even in the presence of poor data quality.

Another type of data quality issue that shares similar characteristics with drop-outs due to network congestion is the case of repeated values. Similar to data drop-outs, repeated values can occur intermittently. However, contrary to a missing data point which has a TVE of 100%, the TVE of a repeated data point is very small. Although, it may appear that this data quality problem is not necessarily a problem because the values do not change as much, in reality, the phase angle is changing with each new frame (based on the deviation from the nominal frequency). Therefore, repeated values create problems with the quality of the phase angles. Repeated values of 10% and 20% (probabilistically) were analyzed in [15]. The plot obtained for repeated values of 20% is shown in Fig. 6. Similar to the case of data drop-outs, from Fig. 6 it is realized that repeated values of up to 20% can be successfully addressed by the proposed algorithm.
Fig. 3. Performance of data conditioning algorithm with 20% drop-outs

Fig. 4. Performance of data conditioning algorithm with 40% drop-outs
B. Simulation Results – Divergence and Network Changes

As the probability of adjacent data point drops increases, the data conditioning algorithm loses its ability to make accurate predictions. Fig. 7 shows the phasor magnitude plot for a scenario where there is 50% likelihood (probabilistically) that each data point will be lost. While the algorithm is able to perform for some time, it does not take long until so many adjacent data points are lost that the prediction diverges from the raw measurements. This is because if the observation residual remains high, the measurement will keep getting replaced by the optimal predicted estimate. Therefore, even when the good data returns, the error in the optimal prediction would have compounded so many times that it would not be able to track the raw measurements anymore. In order to prevent the algorithm from becoming numerically unstable and diverging, a reset function is built-in to it that activates when the smoothing window is completely filled with estimated data. The pseudo-code for the algorithm’s reset function is provided in Fig. 8.
Fig. 7. Performance of data conditioning algorithm with 50% drop-outs

\[
\begin{align*}
\text{If} \left( \frac{\text{Measurement fails SNR/Plausibility Checks}}{\text{ObservationResidual}_{\text{Measurement}}} \geq \text{Predefined Threshold} \right) \\
\quad \text{Flag}_{\text{UnreliableQuality}} = 1 \\
\text{else} \\
\quad \text{Flag}_{\text{UnreliableQuality}} = 0 \\
\text{endif} \\
\text{If} (\text{Flag}_{\text{UnreliableQuality}} = 1) \\
\quad \text{Replace Measurement by Optimal Predicted Estimate} \\
\quad N_{\text{SubOptimalDataPoints}} = N_{\text{SubOptimalDataPoints}} + 1 \\
\text{else} \\
\quad N_{\text{SubOptimalDataPoints}} = 0 \\
\text{endif} \\
\text{If} \left( N_{\text{SubOptimalDataPoints}} \geq \text{SmoothingWindowLength} \right) \\
\quad \text{Reset Algorithm} \\
\quad N_{\text{SubOptimalDataPoints}} = 0 \\
\text{endif}
\end{align*}
\]

Fig. 8. Pseudo-code to depict data conditioning algorithm’s reset functionality

In Fig. 8, \( N_{\text{SubOptimalDataPoints}} \) denotes the number of estimates. If the number of successive bad measurements received by the algorithm equals or exceeds the smoothing window length, then the algorithm will reset itself. It will start operating normally (afresh) once the smoothing window gets filled with raw data (and not estimates). Thus, for the quadratic prediction model used here, there will be a delay of at least three frames. However, proper selection of initial conditions can help the algorithm track the synchrophasor stream faster. For example, the steady state error covariance matrix and Kalman-filter gains can be saved and used to reinitialize the algorithm when required. The effect of the reset function on the same phasor magnitude plot that was described in Fig. 7 is shown in Fig. 9. The plot shows that although the output is mostly raw data (since there are not enough adjacent data points to support a quality prediction), there is no divergence even after the simulation has been run for more than twice the length of time that the first simulation (Fig. 7) was run for.

Another advantage of the reset functionality is that by using it, contingencies or discrete changes in the system can be properly conditioned. Since the quadratic prediction model developed in [8] cannot account for step changes, it takes several samples until the window moves past the step change before it can properly track the stream again. However, by resetting the algorithm at the right time, a discrete network change can be immediately acknowledged. Fig. 10 demonstrates the effectiveness of the algorithm’s reset functionality during a loss of generation event simulated in the IEEE 118-bus system. From the figure it becomes clear that by resetting the algorithm at the correct instant, the optimal smoothed estimate (green star) is able to track the actual measurement (red circle) perfectly.

V. DISCUSSION: RELEVANCE OF THE QUADRATIC PREDICTION MODEL BASED KALMAN FILTER

Before the introduction of phasor measurements, the idea of tracking the state of the power system with a Kalman filter like process was suggested in [16]. The difficulty in the use of such a filter for the current application is that the number of measurements made using PMUs is inadequate to produce a successful estimate based on that approach. To elaborate, DVP has approximately 4,000 buses. Therefore, the traditional Kalman filter will require a state equation for the states of all the 4,000 buses. However, DVP has placed PMUs on only their high voltage network (500kV buses) which number about 30. In such a scenario, a traditional dynamic/tracking estimator will not work because all the states of the system will not be “observable”. The proposed approach considers each state individually. Therefore, the predicted value of a state is based on the previous predicted values of the same state. Hence, the proposed methodology is independent of the network model/size.
The issue of bad data was also not taken into account in [16]. Reference [17] was one of the first papers to propose a tracking state estimator with bad-data processing abilities under normal conditions using a Kalman-filter like model. It was later extended to handle abnormal conditions in [18]. But, since PMUs were not in existence at that time, these methodologies involved estimating the states from the flow measurements using the Jacobian matrix. Hence they were neither very fast, nor very accurate. A robust algorithm for dynamic state estimation that was proposed to be immune to polluted measurements was developed in [19]. But the reliability of their approach for an actual implementation was a major concern. The use of unscented transform and its combination with the Kalman filter (Unscented Kalman Filter) and wide area measurements for dynamic state estimation and bad data detection was proposed in [20], [21]. However, the approach developed in this paper shows that by using the quadratic prediction algorithm, even a simple Kalman filter is able to not only detect, but also clean synchrophasor data in the context of a PMU-only state estimator.

Another significance of the proposed approach is that it can detect bad data in the individual measurement using the history of that same measurement. The three-phase linear state estimator developed in [10] considered every measurement as a “fresh” measurement. In the proposed approach, the quadratic prediction model developed in [8] is integrated with the linear state estimation formulation. By doing so, an observation residual is created (Eq. (6)) that detects anomalies in individual measurements based on previous estimates of the same measurement, and thereby validates its quality.

It is also important to point out here that the quadratic prediction model developed in [8] was meant for detecting bad data or exogenous events. It had not been used to then subsequently mitigate the data quality issues while simultaneously being robust to network changes which may appear like bad data. This extension is the focus of this paper.

Moreover, the data used in [8] was simulated data (in MATLAB), whereas the data used in this paper is real
synchrophasor data obtained from the field. The reason why this is relevant is because in [8] the prediction was made for load changes occurring at constant power factor. Loads on the transmission system of a real utility are composite loads. In order to make the quadratic prediction work for actual loads, it had to be formulated in the form of a Kalman-filter whose goal is to minimize the observation residual. This was done in this paper and is therefore its main “technical” contribution.

The second important contribution is the use of a smoothing algorithm for improving the quality of the estimates. This is based on the fact that a fixed-lag smoother significantly lowers the error covariances [13].

VI. CONCLUSION

A data conditioning algorithm for synchrophasor data is developed in this paper. Plausibility checks and signal-to-noise ratios are presented as viable validation methods for preventing poor quality data from propagating downstream as well as for alerting engineers of problems which will require manual intervention. Then, by combining a quadratic prediction model with Kalman filter based filtering and smoothing techniques, conditioning of real synchrophasor data is demonstrated. Some key features of the proposed approach are summarized below –

- Provides optimal smoothed estimate under ideal conditions
- Uses optimal predicted estimate to replace bad/missing measurement
- Automatic reset functionality prevents divergence in presence of high volumes of bad/missing data
- Smoothed estimate tracks measurement accurately when discrete changes (contingencies) occur

The simulations indicate that these algorithms can compensate for bad data values of up to 20% with relative ease. Since the threshold for data quality set by most RTOs is much less, the proposed algorithm is guaranteed to provide a clean data stream to the linear estimator as well as to the downstream consumers and/or network applications that depend on it.

VII. ACKNOWLEDGEMENTS

This work was supported by the Lawrence Berkeley National Lab (LBL) subcontract 7048974 of prime contract DE-AC03-05CH11231 between LBL and Department of Energy (DOE). The authors gratefully acknowledge the support provided by Lawrence Berkeley National Lab (LBL), Department of Energy (DOE), and Dominion Virginia Power (DVP). This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, or The Regents of the University of California.

VIII. REFERENCES

IX. BIOGRAPHIES

Kevin D. Jones (M'11) received the B.S., M.S. and Ph.D. degrees in electrical engineering from the Bradley Department of Electrical & Computer Engineering at Virginia Polytechnic Institute & State University, Blacksburg, VA in 2009, 2011, and 2013, respectively. He has had work experience with Blue Ridge Electric, Lenoir, NC, ExxonMobil Research & Engineering, Fairfax, VA, and ExxonMobil Baton Rouge Refinery, Baton Rouge, LA. He is currently employed by Dominion Virginia Power as an engineer in the Electric Transmission System Operations Center. His research interests are in phasor based state estimation and wide area measurement.

Anamitra Pal (S'10) is currently pursuing his PhD degree in the Bradley and Computer Polytechnic Institute Blacksburg. He Engineering (B.E.) Institute of (India) in 2008 and Degree from Virginia State University, Blacksburg in 2012. From 2008-2010, he worked as a Manager in Tata Steel Ltd. (TSL), Jamshedpur, India. His present research interests include Power System Modeling, Small Signal Stability and Control and Wide Area Measurements.

James S. Thorp (LF 2002) was the Hugh P. and Ethel C. Kelley Professor of Electrical and Computer Engineering and Department Head of the Bradley Department of Electrical and Computer Engineering at Virginia Tech from 2004 to 2009. He was the Charles N. Mellowes Professor in Engineering at Cornell University from 1994-2004. He was the Director of the Cornell School of Electrical and Computer Engineering from 1994 to 2001, a Faculty Intern, American Electric Power Service Corporation in 1976-77 and an Overseas Fellow, Churchill College, Cambridge University in 1988. He was an Alfred P. Sloan Foundation National Scholar and was elected a Fellow of the IEEE in 1989 and a Member of the National Academy of Engineering in 1996. He received the 2001 Power Engineering Society Career Service award, the 2006 IEEE Outstanding Power Engineering Educator Award, and shared the 2007 Benjamin Franklin Medal with A.G. Phadke.