Electrical Power and Energy Systems 32 (2010) 1165–1174

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Bad data pre-filter for state estimation

D. Singh [a,*], R.K. Misra [a], V.K. Singh [b], R.K. Pandey [b]

[a] Department of Electrical Engineering, Institute of Technology, Banaras Hindu University, Varanasi 221 005, India
[b] Department of Applied Mathematics, Institute of Technology, Banaras Hindu University, Varanasi 221 005, India

ARTICLE INFO

Article history:
Received 19 September 2008
Received in revised form 12 November 2009
Accepted 7 June 2010

Keywords:
Bad data detection
Discrete wavelet transform
State estimation

ABSTRACT

Systematic approach for pre-filter for state estimation based on wavelet analysis to detect and eliminate bad data, is developed in this work. Unbiased (random/Gaussian) bad data such as, transient meter failures, transient meter malfunction, and measurements captured during system transients, are inherently in the form of large abrupt change of short duration in a measurement-sequence. These should be detected in pre-filtering stage because their presence poses an extra burden on post-SE bad data analysis. The test results of the proposed pre-filter on two test systems establish that there is a significant reduction in the number of iterations required for bad data detection and elimination.

1. Introduction

State estimation (SE) algorithm fits the measurements on the mathematical model of the power system in order to provide a reliable data-base for monitoring, security assessment, and control functions. The presence of bad data among the measurements processed by a least-squares estimator is, as a rule, detrimental to the estimator's performance, usually resulting in poor state estimates. The methods reported so far, for bad data detection, involve pre-filtering and post-filtering of measurement data using the concept of inter-data consistency of the measurements taken as a snapshot [1–6]. Explicitly erroneous data (bad data) can be rejected by pre-filtering the measurements. However, the pre-filtering tests can be used to detect bad data whose deviations are greater than 30 times the meter's standard deviations [7]. In the context of power system state estimation (PSSE), bad data are measurements much more inaccurate than those which are assumed during modeling [8]. In practice, bad data are caused by a variety of reasons, such as failures in telemetry, communication links, defective meters, system transients and errors in modeling pseudo-measurements.

The post-filtering of bad data is almost mandatory and is more accurate in detection, identification and elimination of bad data compared to the pre-filtering approach. Effective elimination of bad data using post-filtering requires multiple state estimations which are time consuming. In view of this, an efficient pre-filtering approach is desired to reduce the time and computations.

The pre-filtering approaches, in general, use limit checks and/or solve simplified system equations for detecting and identifying the gross errors. An efficient pre-filtering algorithm can in fact significantly improve the state estimator performance both in terms of time and accuracy. Data pre-filtering has deserved relatively scarce attention in the literature and available methods seem to be mainly of heuristic nature. A lack of systematic procedures makes room for application of new techniques to perform data pre-filtering.

The advantage of a pre-filtering is explained with help of flowchart (Fig. 1), depicting the standard power system state estimation and bad data elimination process. The purpose of pre-filtering algorithm proposed in this work is to reduce the bad data at the pre-filtering stage and thereby reducing the number of state estimation and bad data detection and identification (BDDI) loop iterations.

If the measurement characteristic and consistency of individual measurement (samples of same measurement over a time interval, i.e. measurement-sequence) are considered, the gross errors can be detected more easily and effectively. The drawback of the conventional approaches can be explained with the help of Fig. 2. Fig. 2 shows the measurements acquired at the control center. The data x_i^j indicates the measurement j taken at the time instant i. The column wise data, '*measurement-snapshots*,' are a set of system measurements taken as a snapshot at a particular instant of time whereas, the row wise data, '*measurement-sequences*,' are set of individual measurements evolving over a period of time. The conventional methods use the measurement-snapshot, say at kth instant only and try to find the gross error through identification of those measurements which misfit in the system model. But while estimating the state variables of the model, the measurements

* Corresponding author. Tel.: +91 9935343016, +91 542 2310593; fax: +91 542 2368727, +91 542 2368428.
E-mail addresses: dev_singh_vns@yahoo.com, dsitbhu@yahoo.com (D. Singh).

0142-0615/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijepes.2010.06.016

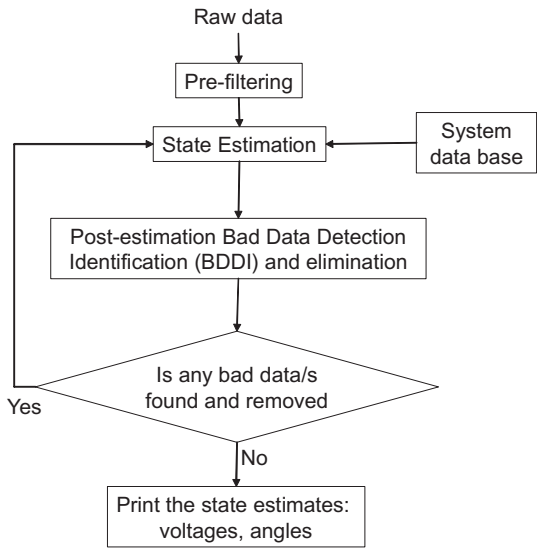

Fig. 1. Flowchart depicting state estimation and bad data detection.

with gross error are also used as independent variables. This assumption leads to problems such as bad data smearing, multiple bad data, and multiple state re-estimations, well known in state estimation literature.

2. Literature review

Salefar and Zhao [9] in their work proposed a neural network based pre-filter for on-line filtering of bad data due to meter malfunction and fast transients. Once trained off-line, the ANN based pre-filter may be used on-line to identify bad data before estimation and thus improves the efficiency of the conventional state estimators. The authors in their work concluded that designing a complete set of training examples, choosing the right size of the hidden layer, and actually training a network could be very tedious and expensive. Thus, there is a possibility of an incorrect response when a network is exposed to a new situation for which it has not been trained. Do Coutto Filho et al. [10] used forecasting aided bad data elimination algorithm to process bad data and data captured during unpredictable sudden changes. The authors used the ANN based forecaster for the purpose. Souza et al. [11] proposed pattern analysis based approach using Group Method of Data Handling (GMDH) and ANN approach for pre-filtering of bad data. However, although GMDH reduces the training time by appropriately reducing the input dimension, it is difficult to generate all type of possible system errors for effective training. A similar method was proposed by Do Coutto Filho et al. [12] which uses the difference in measurements of forecasted state by dynamic state estimator and the current measurement to suspect bad data. However, the method for filtering of bad data uses the traditional state estimator and residual tests. Richard Andrew Wiltshire et al. [13] developed short term alarming procedures based on statistics of power spectral density of Kalman filter innovations, to detect the modal changes in a large interconnected power system. There have been recent literature on ANN based state estimator [14–18] and fuzzy based state estimators [19,20] based on pattern recognition concepts which can run faster than conventional state estimators. However, the conventional WLSE based state estimators are still the main stay due to their track-back capability during failures. It is always easier to find the modeling defects as well as provide system upgrade adaptability in conventional state estimators.

The concept proposed in the present work is a step further in pre-filtering analysis in the sense that uses systematic analysis of the measurement-sequences. It uses row wise measurements (measurement-sequence) as well as column wise measurements (measurement-snapshots) as data in pre-filtering stage for detection, identification, and removal of the bad data as well as detecting major system changes and meter malfunctions. It uses well established wavelet analysis method for detection of bad data and sudden changes avoiding the problems of ANN based methods such as insufficient and inappropriate training. An ideal bad data pre-filter for state estimator should:

(i) Detect, identify and eliminate the bad data points in the pre-filtering stages of state estimation.
(ii) Detect and identify permanent meter failure and meter malfunctions in the systems for unbiased bad data.
(iii) Detect and monitor the major system changes due to line, load and generator switching in/out using analog measurements.

3. Wavelet analysis

Wavelet transform analysis has been applied in various on-line and off-line decision and analysis functions such as fault location and estimation [21–24], short-term disturbance modeling and analysis [25,26], transient analysis [27], and relaying [28–30].

It is shown in later sections that abruptness of the measurement is one of the characterizing feature of a measurement with bad data. The wavelet analysis of the data basically decomposes the data in two parts namely a smooth or slowly varying low frequency component and the abrupt or fast changing high frequency

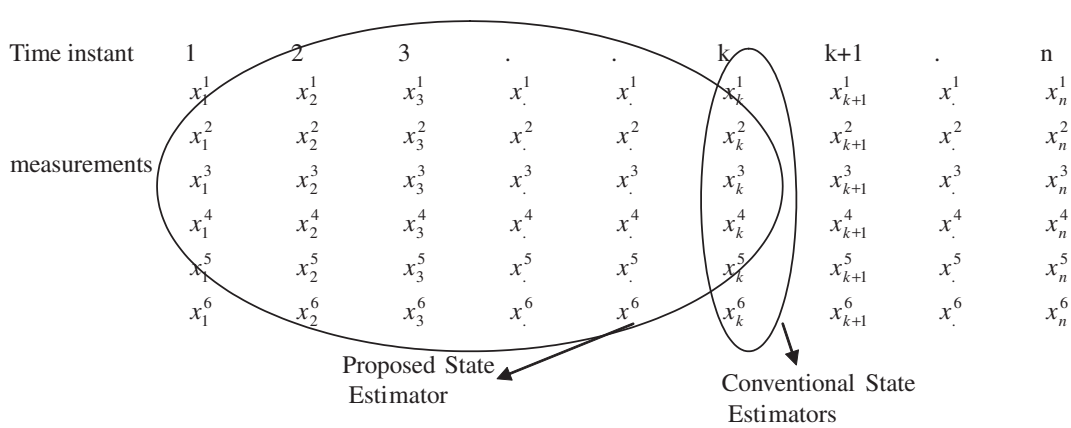

Fig. 2. Data considered by proposed method for gross error detection/elimination.

component. Also, the more important feature is that, while decomposing the data in high and low frequency components it retains the time identity of the instant in terms of when those frequencies appeared in the said data. The accuracy of time instant information is high for high frequency whereas the same is less accurate for low frequency; the phenomenon well known as uncertainty principle. This is because of lower scales (i.e. high frequencies) the time information is more precise whereas for higher scales (low frequencies) the frequency information is more precise than the time information [31]. With wavelet transform, it is possible to extract frequency information in time domain by decomposing the signal with short scale of window for high frequency band while with long window scale for low frequency band using scale and shift technique. This is an important feature which is used as an identifying feature of a bad data and therefore the reconstructed high frequency component are used to identify the abruptness in the measurement.

A discrete wavelet transform (DWT) of a signal $x[k], k = 1, 2, 3, \ldots, n$ is obtained by passing it simultaneously through a high-pass, $h(n)$, and a low-pass, $g(n)$, filters.

The filter outputs, down-sampled by 2, give the following relation for high-pass and low-pass filters, respectively.

$$y_{\text{high}}[n] = \sum_{k=-\infty}^{\infty} x[k] \cdot h[2n - k] \tag{1}$$

$$y_{\text{low}}[n] = \sum_{k=-\infty}^{\infty} x[k] \cdot g[2n - k] \tag{2}$$

The wavelet coefficients obtained through the high-pass and low-pass filters are called detail and approximation coefficients, respectively.

The detail and the approximation coefficients can be used to reconstruct the original signal when used together. Also they can be used separately to reconstruct the high frequency and low frequency components of the original signal. The detailed coefficients reconstruct the high frequency component of the signal thereby showing the time instants at which the measurement-sequence has high frequencies (changed abruptly). Thus, reconstructing the detail coefficients gives the information of time instants at which the measurement-sequence has changed abruptly. To detect abruptness, detection of instants at which high frequencies present are only required, therefore, in the later sections only detailed coefficients are used and discussed.

In the present work 'db10' has been taken as mother wavelet. The other type of wavelets were also tried and were found to have similar performance. The 'db10' is used in the present work for illustrative purposes only; other could also be used instead.

4. Problem formulation and proposed method

The proposed method removes those measurement-snapshots which contain unbiased bad data; thus passing the snapshots which contain biased bad data only. The significant reduction in number of bad data points in the measurement-snapshots will avoid additional computations required for re-estimations during bad data identification and elimination process. The proposed method does not eliminate or replace the existing post-estimation bad data analysis which is a comprehensive simultaneous study of the network model and measurements; rather it reduces the use of post-filtering analysis routine thereby enhancing the performance of the traditional state estimations through pre-filtering of measurement data.

Given the measurement data, the static state estimator calculates the system states in steady state. The static state estimators normally run in interval of every 2–3 min, however, the measurement data scan cycles obtained through telemetry are of order of 0.1 s to few seconds. Thus, there is sufficient time available for the pre-filter to process the measurement-sequences. In measurement acquisition usually some signal processing techniques are necessary to assure the quality of the measurement itself. Proposed method could be placed along with the other SCADA functions.

The static state estimators are used to derive those control and monitoring functions that are associated with power system operating under normal, slowly varying conditions i.e. the slow changes in the state of the system, or large changes on the long time scale, but not with abrupt large changes on very short time scale [32]. In view of the above practical situation, the characteristics of the measurements and their detection, based on reconstruction from detailed wavelet coefficients, can be made in the following manner.

4.1. Noise (error of 2–5% of meter range)

This may occur due to following reasons:

(1) Meter noise.
(2) Telemetry noise.

The reconstructed-detailed-coefficient will not show any large abrupt change for such a measurement-sequence. Thus, the reconstructed-detailed-coefficients will remain below pre-defined threshold which is tuned to the value of noise level (*THGROSS* ≈ *noise*). The proposed pre-filter is not intended to correct or de-noise such a measurement-sequence.

4.2. Gross errors (error above 15% of meter range)

The gross errors normally occur due to following reasons.

4.2.1. Transient meter and/or telemetry malfunctions

Such a gross error will occur at few instants in only one or very few of the measurement-sequences for a chosen window. The reconstructed-detail-coefficients will show large abrupt change at such instants in those measurement-sequences. These large abrupt changes will cause the reconstructed-detailed-coefficients to show the values above *THGROSS* at those instants.

4.2.2. Permanent meter and/or telemetry failure/malfunctions

The malfunction in meter and/or telemetry starts producing gross error permanently and starts showing large errors in almost all the samples of the meter or the telemetry link in question. The reconstructed-detailed-coefficients signal will show the noise level above *THGROSS* in most of the samples. Therefore, if the reconstructed-detailed-coefficients start showing large number of data points above the threshold it can be concluded that it is a permanent meter failure of unbiased nature.

4.2.3. Transient measurements (measurements captured during power swings or other transients)

Such gross errors occur in a number of measurement-sequences at a particular instant (of transient). The reconstructed-detailed-coefficients will show abrupt change in those measurement-sequences at the corresponding instant. Since, the measurement-snapshot of this type is not used for state estimation [32]; it is desirable to eliminate such measurement-snapshots from the measurement-sequences.

4.3. Measurement transitions

4.3.1. Fast change in data due to switching actions (load rejection, line outage, etc.)

In such cases, the change in the set point will be reflected in the detailed coefficients as in case of gross errors. However, such

transitions are reflected in large number of measurements. Consider a case of switching in/out of a line, the flows in almost all the lines (or most of the lines in a large system) will change abruptly. The threshold crossing will be reflected at the same time instant in almost all the related measurements. Thus, fast changes in data due to switching actions can be detected when reconstructed-detailed-coefficients show values higher than *THGROSS* in large number of measurements at the same time instants.

4.3.2. Slow transition of measurements (in small time range) in normal operating conditions

Slow transitions are encountered in normal operating situations. The reconstructed-detail-coefficients will not show any abrupt change for such a data window indicating normal operation.

4.3.3. Biased gross errors

These errors are of special kind and occur due to wrong assumption or wrongly entered data-base item. Such examples are; incorrect tap-setting value saved in a data-base, incorrectly entered status of un-metered switch/ circuit breakers, assumed incorrect polarity of a measured item, and permanently failed meter or telemetry link stuck-up at a particular value [33]. These types of errors normally occur in newly installed measurement system or part of a system. Such errors can only be detected through rigorous post-estimation bad data analysis along with engineering judgment and intuition.

From the above discussion, it is obvious that all the erroneous measurement samples containing gross errors and transients can be detected and eliminated keeping only the normal signals i.e. the noise and slow transition signal samples. The algorithm, depicted in Fig. 3, shows the methodology along with typical values used in the work to eliminate the gross error samples. In the said algorithm, steps from 2 to 6 are treated as one pass through the wavelet analysis.

Ledwich and Palmer [34] introduced the experimentally motivated hypothesis that disturbance inputs to the power systems (corresponding to such aspects as load changes) can be well modeled as integrated white noise. The assumption of 2% as noise and more than 15% as gross error, in the present work, is only indicative of the error ranges normally encountered in practice. There is no clear indication of exact values for these parameters in literature. In fact, the estimated noise range is the only choice in actual practice for state estimation, because of the large number of measurement and absence of viable data. The ranges of such data are 30σ [1], $5\sigma - 20\sigma$ [14] and 20% or higher of the meter range [20] for bad data, and 2σ [14] and 8% (±4%)of the meter range [20] for normal noise.

The effectiveness of the proposed method is first demonstrated on measurements obtained from sample power system shown in Fig. 4. The actual values and corresponding measurements (X's) acquired are also shown in Fig. 4. The studies are also carried out for IEEE 14-bus test system and the summary of results for both the systems is presented in Table 1. Following matrices and matrix equations are defined.

$$X = \Gamma + \Sigma \tag{3}$$

and

$$X^k = \Gamma^k + \Sigma^k \quad \text{(at } k\text{th instant)} \tag{4}$$

```
Algorithm
1. Capture the predefined length of all the measurement-sequences. Set
THGROSS=2% of meter range, SET NLARGE=30. Set TOTAL_ITT=2. ITT=1.

2. Obtain the discrete wavelet decomposition at level 1 for all the
measurement-sequences and reconstruct the detailed wavelet coefficients
for all the measurement-sequences.

3. Find the time instants (k's), for each of the measurement-sequences
where the reconstructed-detailed-coefficients values cross the THGROSS.
Assign error marking of -1% to samples at these instants.

4. IF for a particular measurement- sequence, number of k's is greater
than NLARGE, THEN reject all samples of that measurement-sequence. FLAG
that particular measurement as PERMANENT METER FAILURE (REJECTED
MEASUREMENT).

5. Eliminate the measurement-snapshots (and nearby snapshots) at those
time instants where reconstructed-detailed-coefficients>=THGROSS. FLAG
the corresponding instants as large abrupt change (Major system
changes).

6. Shift the remaining measurement-snapshots to construct the new set
of measurement-sequences.

IF ITT=TOTAL_ITT, THEN OUTPUT measurement-snapshots. FLAG Major system
changes.       STOP
ELSE
Set ITT=ITT+1 Go to step 2.
```

Fig. 3. Algorithm for gross error detection.

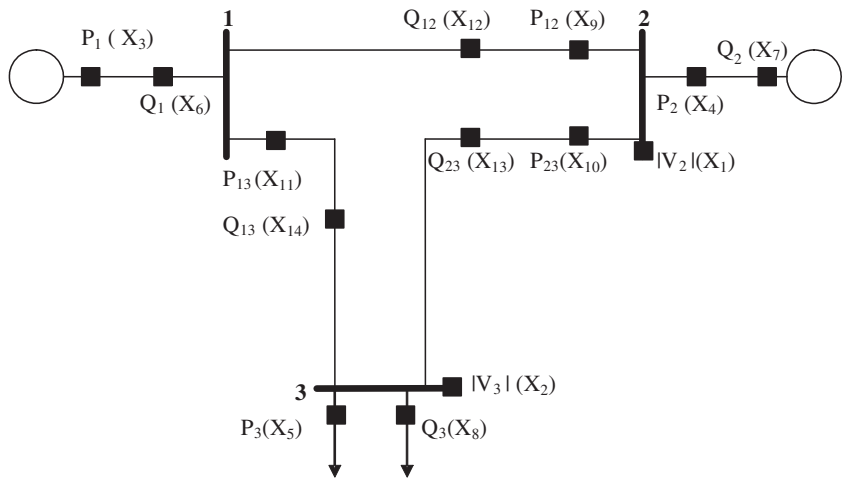

Fig. 4. Sample 3-bus system.

Table 1
Overall gross error elimination for sample system and IEEE 14-bus system.

	Sample system	IEEE 14-bus
Total number of samples	75,600 (5400 × 14)	432,000 (5400 × 80)
Number of gross error samples	2807	3007
Number of samples rejected after ITT = 1	29,049 (38.90%) **(2074)**	30,112 (7%) **(2157)**
Number of samples rejected after ITT = 2	29,452 (38.96%) **(2103)**	31,002 (7%) **(2239)**
Correct detection after ITT = 1	99%	99%
Correct detection after ITT = 2	99.8%	99.6%

where

$$X = \begin{bmatrix} x_1^1 & x_1^2 & \dots & x_1^k & \dots & x_1^m \\ x_2^1 & x_2^2 & \dots & x_2^k & \dots & x_2^m \\ \vdots & & & \vdots & & \vdots \\ x_i^1 & x_i^2 & \dots & x_i^k & \dots & x_i^m \\ \vdots & & & & & \vdots \\ x_n^1 & x_n^2 & \dots & x_n^k & \dots & x_{n1}^m \end{bmatrix},$$

$$\Gamma = \begin{bmatrix} \rho_1^1 & \rho_1^2 & \dots & \rho_1^k & \dots & \rho_1^m \\ \rho_2^1 & \rho_2^2 & & \rho_2^k & & \rho_2^m \\ \vdots & & & \vdots & & \vdots \\ \rho_i^1 & \rho_i^2 & \dots & \rho_i^k & \dots & \rho_i^m \\ \vdots & & & & & \vdots \\ \rho_n^1 & \rho_n^2 & \dots & \rho_n^k & \dots & \rho_{n1}^m \end{bmatrix},$$

$$\Sigma = \begin{bmatrix} \varepsilon_1^1 & \varepsilon_1^2 & \dots & \varepsilon_1^k & \dots & \varepsilon_1^m \\ \varepsilon_2^1 & \varepsilon_2^2 & & \varepsilon_2^k & & \varepsilon_2^m \\ \vdots & & & \vdots & & \vdots \\ \varepsilon_i^1 & \varepsilon_i^2 & \dots & \varepsilon_i^k & \dots & \varepsilon_i^m \\ \vdots & & & \vdots & & \vdots \\ \varepsilon_n^1 & \varepsilon_n^2 & \dots & \varepsilon_n^k & \dots & \varepsilon_{n1}^m \end{bmatrix},$$

$$X^k = \begin{bmatrix} x_1^k \\ x_2^k \\ \vdots \\ x_i^k \\ \vdots \\ x_n^k \end{bmatrix}, \quad \Gamma^k = \begin{bmatrix} \rho_1^k \\ \rho_2^k \\ \vdots \\ \rho_i^k \\ \vdots \\ \rho_n^k \end{bmatrix}, \quad \text{and} \quad \Sigma^k = \begin{bmatrix} \varepsilon_1^k \\ \varepsilon_2^k \\ \vdots \\ \varepsilon_i^k \\ \vdots \\ \varepsilon_n^k \end{bmatrix}.$$

$$X_i = \begin{bmatrix} x_i^1 & x_i^2 & \dots & x_i^k & \dots & x_i^m \end{bmatrix},$$

$$\Gamma_i = \begin{bmatrix} \rho_i^1 & \rho_i^2 & \dots & \rho_i^k & \dots & \rho_i^m \end{bmatrix},$$

and

$$\Sigma_i = \begin{bmatrix} \varepsilon_i^1 & \varepsilon_i^2 & \dots & \varepsilon_i^k & \dots & \varepsilon_i^m \end{bmatrix}.$$

In the above set of equations ρ_i^k is the *actual value* of the measured quantity, i, at the kth instant; ε_i^k is the random noise (or gross error) added to the *actual value* i, at the kth instant; and x_i^k is the corresponding measurement. The n and m are respectively the number of measurement-sequences and the number of samples of each measurement taken for wavelet analysis.

For generating test cases, the methodology of Refs. [36–38] was adopted. The white noise of ±2%, of the meter range was considered as normal noise and noise greater than ±15% of the meter range was considered as gross error. The gross errors are assumed to be present in 3–5 randomly selected samples of any of the measurement-sequence. For meter malfunction the random noise of ±15%, of the meter range is assumed to be continuously present in large number of samples in a particular measurement-sequence.

5. Simulation results

In the system of Fig. 4, data samples for 5400 instants were generated. For the sample system the total number of measurement–sequences, n, is 12. A window length (m) of 128 samples was selected for wavelet analysis. In a data window, set of five measurement-sequences were chosen such that one measurement-sequence has no gross errors or meter malfunction. Three measurement-sequences from the above are such that each has gross errors at 3–5 instants, and one measurement–sequence has a meter malfunction. This typical set of five measurement-sequences is chosen for demonstration of three cases, i.e. *cases 1–3*. All the samples were time stamped to retain their time identity, so that the samples obtained after wavelet analysis and the original samples

could be compared appropriately. The threshold (*THGROSS*) employed was 2% of the meter range. The development of different test cases and results for these test cases using proposed method of wavelet analysis are described in the following sections.

5.1. Case 1: measurements without gross errors

The wavelet analysis of measurement-sequence, X_2, containing no gross error i.e. having normal noise only, is presented in this case. This case is considered to demonstrate the fact that measurement-sequence without gross errors, do not show values larger than the *THGROSS* in the corresponding reconstructed-detailed-coefficients.

The measurement, X_2, corresponding to variable, $|V_3|$, is indicated in Fig. 5a. The large abrupt drop in voltage at the 100th sample is due to outage of lines 2–3 at that instant. The percentage absolute difference between the actual variable, $|V_3|$, and the corresponding measurement, X_2, denoted by % $|\Sigma_2|$, is indicated in Fig. 5b showing the noise of 2% as expected in the actual value of the measurement, $|V_3|$. Fig. 5c shows the percent absolute value of reconstructed-detailed-coefficients, $\Sigma_{2,d1}$, denoted by % $|\Sigma_{2,d1}|$, obtained after the first pass. Fig. 5d shows the percent absolute-reconstructed-detailed coefficients, $\Sigma_{2,d2}$, denoted by % $|\Sigma_{2,d2}|$, obtained after the second pass. In Fig. 5c and d, it can be observed that there are numerous instants having error markings of -1%. These error marking of -1% corresponds to that time instant (k), where the reconstructed-detailed-coefficient value crossed the *THGROSS* for the given measurement-snapshot in any of the measurement-sequences. The error marking of -1% is due to those x_i^k in which gross error was found in any of the measurement-sequence. This is in accordance with step 3 of *Algorithm 1*. If x_i^k is having gross error then the corresponding measurement-snapshot, $X^k = [x_1^k, x_2^k, \ldots, x_n^k]^T$, is to be eliminated from the matrix, X. All such snapshots are given an error marking of -1% in Fig. 7c and d and other such figures to follow. Thus, samples which are removed (having error marking of -1%) actually correspond to gross errors and signal transitions (abrupt switching) present in *other* measurement-sequences and not related to measurement-sequence X_2.

Thus, as desired, no gross errors are indicated in this case.

5.2. Case 2: meter malfunction

The measurement-sequence, X_1, corresponding to variable, $|V_2|$, is indicated in Fig. 6a. The measurement-sequence, X_1, has large number of gross errors indicating meter malfunction. The percent absolute difference between the actual variable, $|V_2|$, and the corresponding measurement, X_1, denoted by % $|\Sigma_1|$, is indicated in Fig. 6b. Fig. 6b shows gross errors in all the samples after 43rd instant. Thus, there are large number of samples (>NLARGE) crossing the *THGROSS*. Fig. 6c and d shows that *Algorithm 1* assigns an error marking of -1 to all the samples of the measurement-sequence indicating that all the samples are to be rejected and there is a meter malfunction corresponding to this measurement- sequence.

For such a case, whole of the measurement-sequence, X_1, is rejected and is not included for state estimation. The existence of such a situation is flagged for correction or repair of meter or the communication link. However, the meter stuck-up at a particular value cannot be detected by the proposed method but can easily be detected by post-estimation bad data analysis.

5.3. Case 3: transient meter malfunctions/transient telemetry malfunctions/system transients

The measurement-sequences, X_{13}, X_{11}, and X_{10} corresponding to actual values, Q_{23}, P_{13} and P_{23}, respectively are indicated in Figs. 7a, 8a and 9a. These measurement- sequences contain gross errors due to transient meter malfunctions/ transient telemetry malfunctions/ system transients. The percent absolute difference between the actual variables and the corresponding measurements, denoted by $\%|\Sigma_{13}|$, $\%|\Sigma_{11}|$, and $\%|\Sigma_{10}|$, are indicated in Figs. 7b, 8b, and 9b, respectively, showing the large percent absolute differences at the instants of gross errors. Thus, large number of samples cross *THGROSS*. Figs. 7c, 8c, and 9c show the percent absolute reconstructed-detailed-coefficients denoted by $\%|\Sigma_{13,d1}|$, $\%|\Sigma_{11,d1}|$, and $\%|\Sigma_{10,d1}|$, respectively obtained after the first pass. The samples which are to be removed have error marking of -1% on the same figures. Samples which are removed actually correspond to gross errors and signal transitions (abrupt switching) present in *other* measurement-sequences. These are the samples removed from

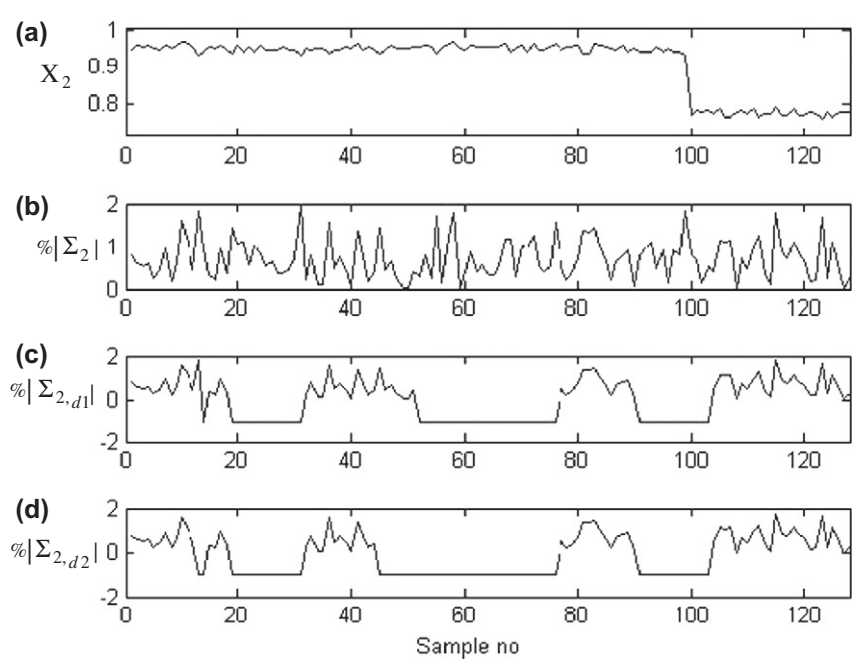

Fig. 5. (a–d) Wavelet analysis of measurement without gross errors.

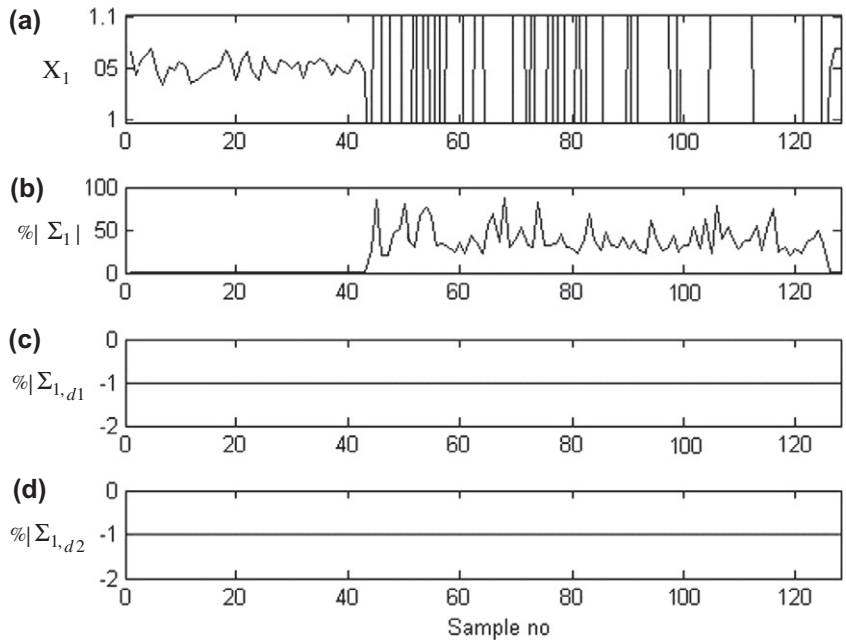

Fig. 6. (a–d) Wavelet analysis of measurement with permanent meter malfunction.

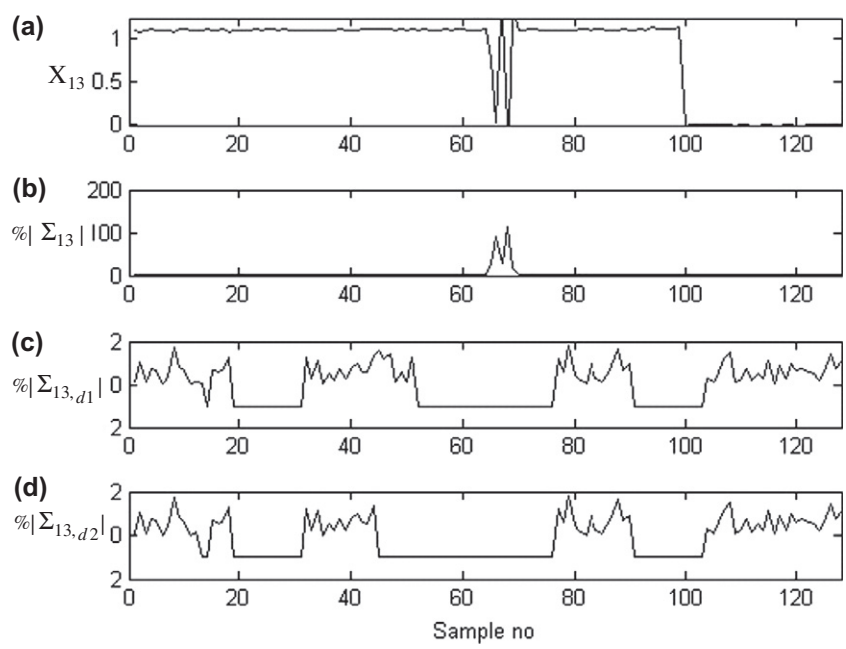

Fig. 7. (a–d) Wavelet analysis of measurement with gross errors.

all the measurement-sequences as reported in the earlier case, *case A*.

It is observed from Figs. 7d, 8d and 9d that there is elimination of extra samples (around instant 50) in the second pass of the algorithm even though the gross error are eliminated after the first pass. This elimination of extra samples is due to the abrupt change which appears after the relocation of two measurement-sequences obtained after removal of some of the snapshots during first pass.

5.4. Case 4: overall statistics

Fig. 10a–d shows the complete wavelet analysis for particular measurement-sequence X_8 for 5400 samples. A complete scenario of a said measurement for 5400 instants is depicted in Fig. 10a. It is observed that the smooth changes are not flagged, even though the values change considerably; however, the abrupt changes and gross errors are detected successfully by the proposed method. The overall statistics for total number of sample, number of gross errors, and number of data eliminated after first and the second pass for the two test systems is given in Table 1. The number of gross errors per window for the sample system and IEEE 14-bus system are same. For the sample system, the amount of data rejection is almost 40% of the total number of samples.

This is due to following reasons: (i) all the samples in a rejected measurement-snapshot due to gross error in single measurement are counted individually and (ii) in the cases of meter malfunction,

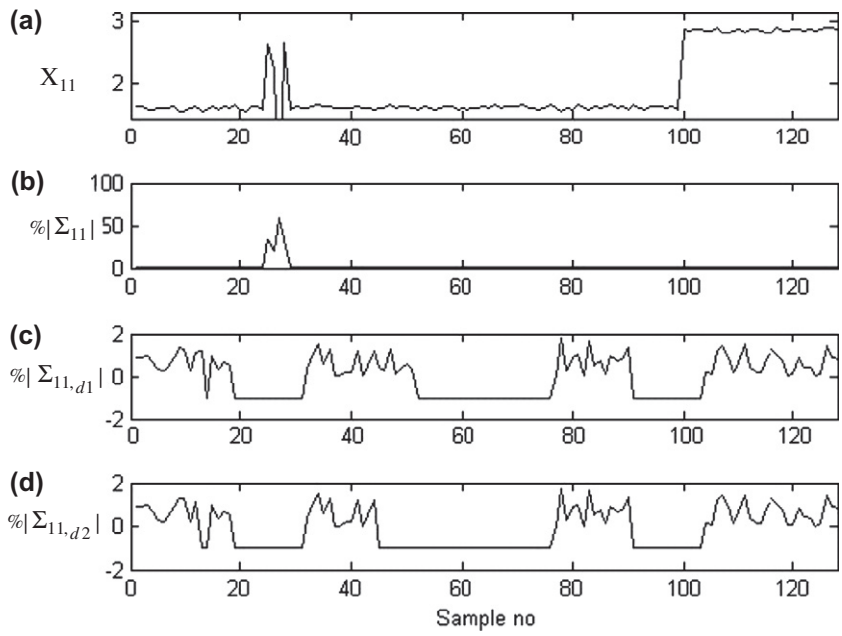

Fig. 8. (a–d) Wavelet analysis of measurement with gross errors.

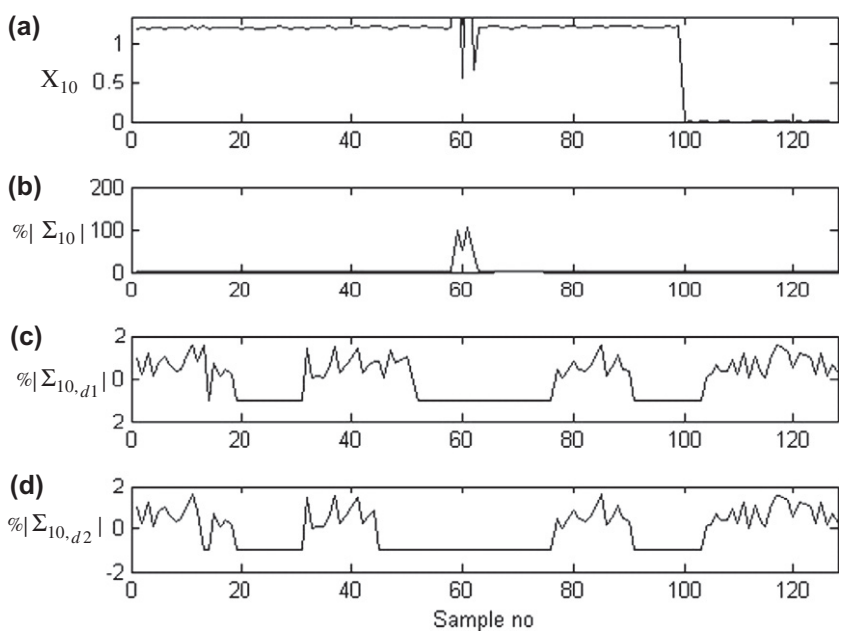

Fig. 9. (a–d) Wavelet analysis of measurement with gross errors.

even if only NLARGE data samples actually show gross errors; samples in the whole measurement-sequence are counted as rejected samples. In fact, the actual number of measurement-snapshots (instants) rejected is quite low and even less than the number of gross errors. This is because some of the gross errors are overlapping. The actual number of instants rejected is indicated in Table 1 in bold face.

Table 2 shows the advantage gained in terms of the number of iterations saved during state estimation when the proposed pre-filter is incorporated. The WLSE state estimator was employed as the state estimation routine with residual analysis for BDDI was used to test the efficacy of the proposed method. The general pre-filtering limit checks were incorporated in both the instances (with and without proposed pre-filter). Table 2 shows the number of iterations required in the state estimation and post-SE BDDI (SE-BDDI) loop. It can be observed that the average number of iterations in SE-BDDI loops were considerably reduced showing the effectiveness of the proposed filter. The average number of iteration is calculated as the ratio of total number of iterations required for the converged cases to that of total number of converged cases. The converged cases correspond to those, in which the system could remain observable after the elimination of bad data. When the proposed pre-filter is used, the number of iterations in SE-BDDI loops was almost negligible and only in few cases the errors could escaped to SE-BDDI loop.

The programs for the proposed method and conventional WLSE were coded in MATLAB [35] software and run on a Compaq PIV machine. The sparsity feature was utilized for WLSE method and the

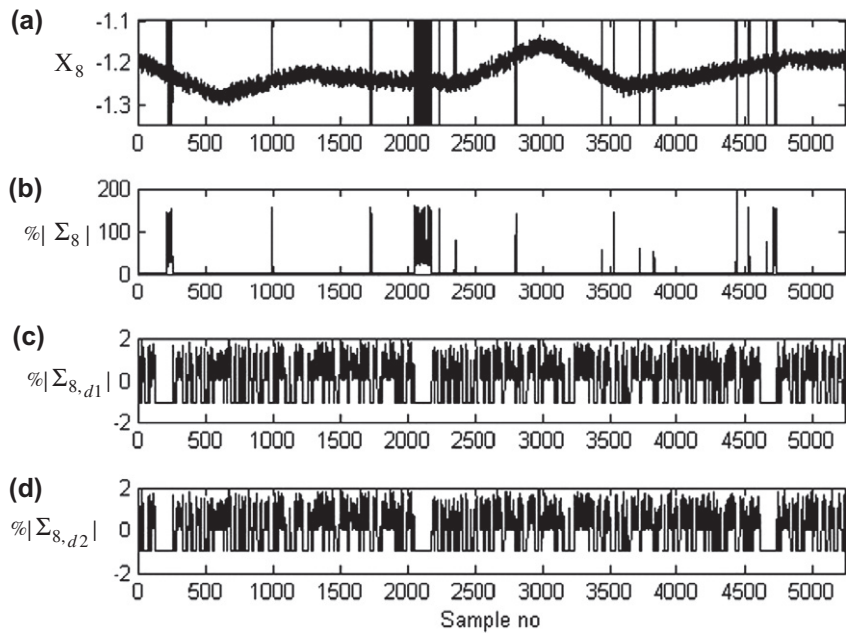

Fig. 10. (a–d) Wavelet analysis of measurement 5400 data samples showing gross errors and permanent meter malfunction.

Table 2
Iterations required in the state estimation and post-SE BDDI (SE-BDDI) loop.

System	3-Bus	IEEE 14-bus
Number of cases tested	5400	5400
Converged cases	4215	2942
Average number of iterations of SE-BDDI loop.	2.5	3.8
Average number of iterations of SE-BDDI loop with proposed pre-filter	0.01	0.01

tolerance of 0.00001 was taken as convergence criteria. The computation times reported in the following lines are mean time for 540 test cases since the time will be different due to different number of iterations for the same system in different loading and random errors. The time required for the proposed method, i.e. pre-filter with WLSE was found to be 0.12 s, as compared to 0.26 s for the WLSE without pre-filter in case of IEEE 14-bus system (80 measurements). The time required for IEEE 57-bus system (186 measurements) is 2.24 s for the proposed method as against 4.57 s for the conventional method. The time taken for IEEE 118-bus system (371 measurements) for the proposed method was 5.62 s, whereas the WLSE took 8.87 s. The wavelet decomposition can, in fact, be performed in parallel for all the measurements. Thus, the time for pre-filtering remains same irrespective of the number of measurements. The time taken for such a computation was of the order of few milliseconds (3 ms) which is considerably small compared to actual SE routine.

5.5. Case 5: on-line monitoring/detection of major system changes

It can be observed that after making the two passes through the wavelet analysis, all the gross errors are removed. The examination of the large noise values in the remaining data shows that they only correspond to system switching (major system changes) such as line outage, generation outage, and load rejection. In such cases, the large values of the reconstructed-detailed-coefficients exhibit a special characteristic that they appear in large number of measurements at almost same time instants. Therefore, values greater than the *THGROSS* at the same time instants in large number of measurements, indicate that a system switching has occurred. However, the line or load cannot be identified but since there would be only few such changes in a window (of say 128 s) it can be used as a mechanism to detect and verify the topology changes shown by the status measurements.

6. Discussion

If the threshold value is too low (less than the noise range of the meter) large number of samples will be removed and if it is too high then many bad data samples may pass through wavelet analysis undetected. It is found that not more than two passes (TOTAL_ITT = 2) are sufficient to get excellent results; at least 99% of bad data are detected and eliminated. One percent of bad data pass through the second iteration because of following two reasons.

(1) Samples at the end (sample number 128) or the beginning (sample number 1) are seen as a large discontinuity by discrete wavelet transform. However, this can be avoided by not considering the reconstructed-detailed-coefficients for last two samples; i.e. the two samples will never be used.

(2) The second reason is a specific type of bad data wherein all subsequent samples have noise of same sign and almost same magnitude, making the bad data to be almost constant for a reasonable amount of time (i.e. of sufficiently low frequency). The sample at beginning and the end are successfully detected. However, remaining part of bad data passes undetected for such a measurement-sequence.

The method proposed in the paper is for pre-filtering which is obviously very different from normally performed post-SE bad data analysis. The post-SE bad data analysis should be performed for elimination of biased bad data such as, time skew, meter malfunction of biased kind (meter stuck-up at steady value), and conforming bad data. These bad data are difficult to detect in any measuring system and up to an extent can be detected through post-SE bad data analysis only.

In fact, unbiased (random/Gaussian) bad data such as, transient meter failures, transient meter malfunction, and measurements

captured during system transients, are inherently in the form of large abrupt change of short duration in the measurement-sequence. These should be detected in pre-filtering stage because their presence poses an extra burden on post-SE bad data analysis.

The pre-filter in proposed method is developed to remove measurement-snapshots having "unbiased" bad data. In addition to the above, the proposed method also detects and flags the large changes such as, load rejection, pick up, and line outage, which provide important information about the system conditions to the system operator. In this sense the proposed method also performs on-line data monitoring functions.

The noise level is a system characteristic and is usually estimated. The threshold *(THGROSS)* is kept slightly higher than the noise level so that the measurements with noise pass through the filter but those with the bad data are detected. The threshold is a tunable parameter.

7. Conclusion

The application of wavelet analysis for systematic approach to pre-filtering bad data has been implemented and tested in this work. The studies show that wavelet analysis can be applied for bad data elimination with excellent bad data filtering capabilities in pre-filtering stage. A large reduction in number of SE-BDDI loop iterations is observed. Thus, the proposed pre-filter avoids multiple state re-estimations and bad data elimination cycles resulting in considerable saving in time and improvement in the efficiency of the state estimator. The on-line detection of major system changes are added advantage of the method.

References

[1] Vempati N, Shoults RR. Sequential bad data analysis in state estimation using orthogonal transformation. IEEE Trans on Power Syst 1991;6(1):157–66 [February].

[2] Garcia A, Monticeli A, Abreu P. Fast decoupled state estimation and bad data processing. IEEE Trans Power App Syst 1979;PAS-98(5):1645–52 [September].

[3] Kotiuga WW, Vidyasagar M. Bad data rejection properties of weighted least absolute value techniques applied to static state estimation. IEEE Trans Power App Syst 1982;PAS-101(4):844–53 [February].

[4] Nian-de Xiang, Shi-Ying Wang, Er-Keng Yu. A new approach for detection and identification of multiple bad data in power system state estimation. IEEE Trans Power App Syst 1982;PAS-101(2):454–62 [February].

[5] Mili L, Van Cutsem Th, Ribbens-Pavella M. Hypothesis testing identification: a new method for bad data analysis in power system state estimation. IEEE Trans Power App Syst 1984;PAS-103(11):3239–52 [November].

[6] Felix F, Wu W, Liu HE, Holten L, Jelsvik A, Aam S. Observability analysis and bad data processing for state estimation using Hachtel's augmented matrix method. IEEE Trans Power Syst 1988;3(2):541–8 [May].

[7] Handschin E, Bongers C. Theoretical and practical considerations in the design of state estimators for electric power systems. In: Handschin E, editor. Real time control of electric power systems. Elsevier; 1972. p. 104–36.

[8] Quintana VH, Simoes-Costa A, Miert M. Bad data detection and identification techniques using estimation orthonormal methods. IEEE Trans Power App Syst 1982;PAS-101(9):3356–64 [September].

[9] Salehfar H, Zhao R. A neural network pre-estimation filter for bad-data detection and identification in power system state estimation. Electr Power Syst Res 1995;34:127–34.

[10] Do Coutto Filho MB, Souza JCS, Matos RSG, Schilling MTh. Revealing gross errors in critical measurements and sets via forecasting-aided state estimators. Electr Power Syst Res 2001;57:25–32.

[11] Souza JCS, Liete da Silva AM, Alves da Silva P. Data debugging for real time monitoring based on pattern analysis. IEEE Trans Power Syst 1996;11(3): 592–9 [August].

[12] Do Coutto Filho MB, Leite da Silva AM, Calvo Cantera JMC, da Silva RA. Information debugging for real-time power systems monitoring. In: Proc IEE, Pt-C, gener transm distri, vol. 136, No. 3; May 1989. p. 145–52.

[13] Wiltshire Richard Andrew, Ledwich Gerard, O'Shea Peter. A Kalman filtering approach to rapidly detecting modal changes in power systems. IEEE Trans Power Syst 2007;22(4):1698–706 [November].

[14] Vinod Kumar DM, Srivastava SC, Shah S, Mathur S. Topology processing and static state estimation using artificial neural network. In: Proc IEE, Pt-C, Gener Transm Distri, vol. 143. No. 1; January 1996. p. 99–105.

[15] Souza JCS, Leite da Silva AM, Alves da Silva AP. On-line topology determination and bad data suppression in power system operation using artificial neural networks. IEEE Trans Power Syst 1998;13(3):796–803 [August].

[16] Souza JCS, Leite da Silva AM, Alves da Silva AP. Data visualization and identification of anomalies in power system state estimation using artificial neural networks. In: Proc IEE, Pt-C, Gener Transm Distri, vol. 144. No. 5; September1997. p. 445–55.

[17] Alves da Silva AP, Quintana VH, Pang GKH. Solving data acquisition and processing problems in power systems using a pattern analysis approach. In: Proc IEE, Pt-C, Gener Transm Distri, vol. 138. No. 4; July 1991. p. 365–76.

[18] Alves da Silva AP, Quintana VH. Pattern analysis in power system state estimation. Int J Electr Power Energy Syst 1995;17:51–60.

[19] Shahidehpour M, Marwali M. Role of fuzzy sets in power system state estimation. Int J Emerg Electr Power Syst 2004;1 [article 1003].

[20] Singh D, Pandey JP, Chauhan DS, Identification Topology. Bad data processing and state estimation using fuzzy pattern matching. IEEE Trans Power Syst 2005;20(3):1570–9 [August].

[21] Jung Hosung, Park Young, Han Moonseob, Lee Changmu, Park Hyunjune, Shin Myongchul. Novel technique for fault location estimation on parallel transmission lines using wavelet. Int J Electr Power Energy Syst 2007;29: 76–82.

[22] Magnago FH, Abur A. Fault location using wavelets. IEEE Trans Power Deliv 1998;13(4):1475–80 [October].

[23] Evrenosoglu CY, Abur A. Travelling wave based fault location for teed circuits. IEEE Trans Power Deliv 2005;20(2):1115–21 [April].

[24] Bhowmik PS, Purkait P, Bhattacharya K. A novel wavelet transform aided neural network based transmission line fault analysis method. Int J Electr Power Energy Syst 2009;31(5):213–9 [June].

[25] Pillay P, Bhattacharjee A. Application of wavelets to model short term power system disturbances. IEEE Trans Power Syst 1996;11(4):2031–7 [November].

[26] Gaouda AM, Kanoun SH, Salama MMA, Chikhani AY. Wavelet-based signal processing for disturbance classification and measurement. In: Proc IEE, Pt-C, gener transm Distri, vol. 149. No. 3; May 2002. p. 310–18.

[27] Wikinson WA, Cox MD. Discrete wavelet analysis of power system transients. IEEE Trans Power Syst 1996;11(4):2038–44 [November].

[28] Youssef Omar AS. Online applications of wavelet transforms to power system relaying. IEEE Trans Power Syst 2003;18(4):1158–65 [October].

[29] Huang Shyh-Jier, Yang Tsai-Ming, Huang Jiann-Tseng. FPGA realization of wavelet transform for detection of electric power system disturbances. IEEE Trans Power Syst 2002;17(2):388–94 [April].

[30] Valsan Simi P, Swarup KS. Wavelet transform based digital protection for transmission lines. Int J Electr Power Energy Syst 2009;31(7-8):379–88 [September].

[31] Daubechies I. Ten lectures on wavelets. Philadelphia (PA): SIAM; 1992.

[32] Larson Robert E, Tinney William F, Hajdo Laszlo P, Piercy Dean S. State estimation in power systems part II: implementation and applications. IEEE Trans Power App Syst 1970;PAS-89(3):353–9 [March].

[33] Assadian Mehdi, Hong Dan French Goddard H. Field operational experiences with on line state estimator. IEEE Trans Power Syst 1994;9(1):50–8 [February].

[34] Ledwich G, Palmer E. Modal estimates from normal operation of power systems. In: 2000 IEEE power eng, soc winter meeting, conf proc, Singapore, vol. 2; 2000. p. 1527–31.

[35] Wavelet toolbox Matlab user's guide, Version 2.1, Mathworks Inc.; 2001.

[36] Schewepe FC, Wildes J. Power system state estimation part I: exact model. IEEE Trans Power App Syst 1970;1(1):120–5 [January].

[37] Schewepe FC, Rom DB. Power system state estimation part II: approximation model. IEEE Trans Power App Syst 1970;PAS-89(1):125–30 [January].

[38] Schewepe FC. Power system state estimation part III: implementation. IEEE Trans Power App Syst 1970;PAS-89(1):130–5 [January].