

# Cluster statistics based normalization for online fault diagnosis of self-powered neutron detectors

Soumallya Chatterjee, Sushil K. Surwase, Mani  
Bhushan, Madhu N. Belur  
Indian Institute of Technology Bombay  
Mumbai, India

Akhilanand P. Tiwari\*, Mahendra G. Kelkar\*\*,  
Mahitosh Pramanik\*\*  
\*Bhabha Atomic Research Center  
\*\*Nuclear Power Corporation of India Ltd.  
Mumbai, India

**Abstract**—Self Powered Neutron Detector (SPND) is a widely used sensor for measuring neutron flux in a nuclear reactor. In this work we propose a novel cluster statistics based normalization scheme to normalize SPND measurements. These normalized measurements are subsequently used in a recursive Principal Component Analysis (PCA) based approach for detecting faults and identifying faulty SPNDs in an online manner. The motivation behind cluster statistics based normalization is that faults effect only individual sensors, while simultaneous variations in multiple sensors are usually caused by dynamic variations in the reactor operation. The proposed normalization approach is applied on SPND data obtained from an operating nuclear reactor and results compared with existing sensor statistics based normalization approach. The results demonstrate the utility of the proposed normalization approach.

**Keywords**—SPND, recursive PCA, false alarms, detection rate, cluster normalization

## I. INTRODUCTION

Self-powered neutron detectors (SPNDs) are widely used sensors to measure the amount of neutron flux present in a nuclear reactor. SPNDs provide real time measurement of neutron flux without requiring any external power [8]. As neutron flux is associated with the reactor power output, measurement of the neutron flux provides a direct measure of the reactor output power. Hence, to ensure optimal and safe operation of a nuclear reactor, it is necessary to have reliable and accurate measurements of neutron flux from SPNDs placed at various locations in the nuclear reactor [1].

During the course of its operation, an SPND can develop a fault due to a variety of reasons. One prominent reason is abrupt change of insulation resistance and capacitance of the cable connecting the SPND sensor with the external circuit, which may occur due to moisture capture, temperature changes etc. [3]. When such faults occur the readings from that particular detector become erroneous and correspondingly the inference based on those readings can degrade reactor performance. Thus, it is necessary to identify the faulty SPND in real time so that appropriate corrective action can be taken. To perform this task in real-time, it is desirable to have an automated real-time diagnostic system which is applicable for

reactor operation over significantly varying power levels and is scalable to work with any number of SPNDs. The aim of the current work is to develop such a real-time diagnostic system.

## II. LITERATURE REVIEW AND OBJECTIVE

Various model based methods are available for fault detection and diagnosis in literature [7]. The models can be derived from first principles (white box models) or from past historical data (data-driven models or black box models) [7]. Such data-driven models involve identification of multivariate relationships amongst the variables of interest given the training data obtained during normal operation, and subsequent deployment of the identified models to detect and estimate faults. For the nuclear reactor operation considered in the current work, white box models are not available. However, operational data is routinely archived and hence data-driven models based on past historical data can be obtained. In [4] authors have proposed a Principal Component Analysis (PCA) based method to obtain data driven models and perform real time fault detection and diagnosis. Towards this end, as a first step they grouped the SPNDs in smaller clusters such that in each cluster SPNDs are highly correlated. After that for each cluster they derived set of linear relationships (models) between the SPND measurements belonging to that cluster using PCA based on historical data. Statistical analysis of the residuals obtained from those linear data driven models enabled them to detect the presence of a fault and subsequent use of generalized likelihood ratio based test enabled identification of faulty SPND in real-time without much computational overhead. However, the obtained data-driven models were implicitly assumed to be static in time.

As the operation of the nuclear reactor varies significantly with time, the models based on past data may not be applicable in the future. A recursive PCA based method known as Recursive PCA–Exact Calculation (RPCA-EC) has been proposed in [1], which uses the data at current time instant to recursively update the models if no fault is detected at the current time instant. This method was based on identifying models on normalized SPND values. Towards this end, each SPND was normalized by subtracting its mean and then dividing this deviation from the mean by the SPND standard deviation. The mean and variance of each SPND

**Paper Id:**

were also recursively updated at each time instant. Each SPND was thus normalized based on its individual mean and standard deviation. However, in presence of significant fast variations in the neutron flux, the mean and the variance will not be updated fast enough to reflect the changed dynamics. This in turn can cause model residuals to be high even in the absence of any fault, thereby leading to high false alarm rates.

In the current work, we propose a new normalization approach known as cluster based normalization which uses instantaneous cluster statistics to normalize the measurements of each SPND in that cluster. The PCA based model is then updated using these normalized values. The advantage of using instantaneous cluster statistics is that it will lead to low false alarm rates even if the individual sensor data changes at a fast rate.

The rest of the paper is organized as follows: in section III we present the background material. In section IV we describe the methodology of the proposed cluster based normalization scheme and the relevant modelling aspects. In section V we present comparative results of our approach and the approach proposed in the earlier literature on SPND data obtained from a nuclear power plant. The paper is concluded in section VI.

### III. PRELIMINARIES

#### A. Brief description of Self Powered Neutron Detector (SPND)

Self-powered neutron detector is a sensor which requires no external power to operate. When neutrons bombard the sensor material (called emitter), it produces current which is proportional to the amount of neutron flux incident on the emitter. The cause for generation of current is the production of electrons mainly via two mechanisms: (i) when neutrons bombard the emitter material, gamma rays are produced which produce electrons by photoelectric/Compton effect, (ii) neutron bombardment on emitter atoms produce radioactive isotopes which produce electrons by radioactive beta decay. The sensors which produce electrons mainly by (i) respond quickly to the changing neutron flux and are called prompt SPNDs (e.g. Inconel and Cobalt). But the sensors which generate electrons by (ii) have a delayed response as the beta decay happens with a certain half-life. Such SPNDs are called delayed SPNDs (e.g. Vanadium). While the prompt sensors have the advantage of promptly responding to flux changes, they tend to be relatively less accurate, as they are sensitive to background noise [1]. In our work we have used the Vanadium sensor as delayed sensor, and Cobalt and Inconel sensors as prompt sensors. In the rest of the paper, Cobalt or Inconel sensors will be referred to as prompt sensors.

#### B. Slowing down of prompt sensor data

The prompt and delayed SPND values may not be directly correlated even when exposed to similar neutron flux [1]. To develop accurate multivariate models irrespective of the nature of the sensors, the measurements from prompt and delayed sensors needs to be made compatible to each other. Besides this requirement, the prompt sensors suffer from background noise which needs to be eliminated. To meet these two tasks,

the authors in [5] have proposed a method which slows down the prompt SPND response by passing the prompt SPND measurement through the Vanadium SPND transfer function. This makes the measurements of both type sensors compatible with each other.

#### C. Clustering of SPNDs

A nuclear reactor may contain hundreds of SPNDs located at different positions inside the reactor. But the neutron flux varies considerably in different regions of the reactor. Thus, to derive high fidelity models it is desirable to identify correlated or similar SPNDs and group them. Clustering using standard k-means algorithm is performed to find such groups of highly correlated SPNDs. Given the number of clusters to be identified, k-means clustering is a widely used unsupervised clustering approach that assigns points to different clusters such that the sum of distances of each point from its cluster centroid is minimized [2]. In the current work, the distance function to perform the k-means clustering is taken to be one minus the absolute value of the correlation coefficient between a pair of sensors [5].

#### D. PCA based modelling

After clusters are obtained, for each cluster initial models for the training data are obtained using principal component analysis (PCA). The available training data set corresponds to measurements of  $n$  SPNDs belonging to a cluster for  $N$  time instants. At each time instant  $t$  we consider the set of readings of  $n$  sensors to be a row vector, labelled  $\tilde{x}(t)$ . Before building a model, the raw values are normalized as follows:

$$x_i(t) = \frac{\tilde{x}_i(t) - \mu_{s,i}}{\sigma_{s,i}} \quad (1)$$

where  $\tilde{x}_i(t)$  is the raw measurement reading for  $i^{\text{th}}$  sensor at time instant  $t$ ,  $\mu_{s,i}$  and  $\sigma_{s,i}$  are the mean and standard deviation of  $i^{\text{th}}$  sensor for the training data set. The normalized row vectors  $x(1)$ ,  $x(2)$  up to  $x(N)$  are computed and stacked together to form the data matrix  $X \in \mathbb{R}^{N \times n}$  as  $X = [x(1); x(2); \dots; x(N)]$ . Models are obtained based on this normalized training data matrix. The detailed discussion of PCA based modelling is given in [1],[4]. Here we present only a brief description of that approach. For PCA based modelling we calculate the covariance matrix of  $x(t)$  from  $X$  as:

$$C = \left(\frac{1}{N}\right) X^T X \quad (2)$$

Let the eigenvectors of  $C$  be stacked as columns of a matrix  $P$  as  $P = [Q|T]$ , where columns of  $Q$  are the  $m$  eigenvectors corresponding to  $m$  smallest eigenvalues of  $C$  and columns of  $T$  are the eigenvectors corresponding to the remaining  $(n-m)$  eigenvalues.

The eigenvectors of the covariance matrix give the orthonormal directions of variability of the normalized data. In particular, small eigenvalue signifies that the variability in the direction of the corresponding eigenvector is small. In our work we want to find a model  $A$  such that  $AX^T = 0$ , and hence we consider  $A = Q^T$ . For each time instant  $t$ , we now assume that  $x(t)$  will satisfy:

Paper Id:

$$Ax(t)^T = 0 \quad (3)$$

The number of lowest eigenvalues  $m$ , which is equal to number of rows of model matrix  $A$ , is chosen by performing a ratio-test on the eigenvalues of  $C$  [1].

We have assumed that  $x(t)$  for  $t = 1, 2, \dots, N$  will follow (3). This would happen if the  $m$  least eigenvalues were identically 0. But this is not the case in general. The  $m$  least eigenvalues are small but non-zero. So (3) will not give 0 residual but it leads to a  $m$ -dimensional residual vector, labelled  $r(t)$ , at each time instant.  $r(t)$  is obtained as follows:

$$Ax(t)^T = r(t) \quad (4)$$

The covariance matrix of the residuals, labelled  $E$  is a diagonal matrix with diagonal entries being the chosen  $m$  least eigenvalues of data covariance matrix  $C$ .

*E. Gross error analysis at current time instant*

At current time instant  $t$  we have considered the raw measurements vector of  $n$  sensors as  $\tilde{x}(t)$ . Using (1) we normalize  $\tilde{x}(t)$  to obtain normalized data vector  $x(t)$ . The residual vector  $r(t)$  is calculated at time  $t$  using (4). After calculating  $r(t)$  the scaled length of the residual vector is calculated as:  $\gamma(t) = r(t)^T E^{-1} r(t)$ .

In absence of any sensor fault,  $\gamma(t)$  will follow a  $\chi^2$  distribution with  $m$  degrees of freedom [6]. Thus a sensor fault is detected at significance level  $\alpha$  if  $\gamma(t) \geq \chi^2_{1-\alpha, m}$  where  $\chi^2_{1-\alpha, m}$  is the value from a  $\chi^2$  distribution with  $m$  degrees of freedom such that the area to the left is  $(1-\alpha)[1]$ . This test is labeled as the  $\chi^2$  test. Violation of the  $\chi^2$  threshold implies that there is a fault in at least one sensor. The faulty sensor in that cluster is then identified using generalized likelihood ratio test [6].

*F. Model update*

To accommodate the change of data nature the model is updated using the Recursive PCA-Exact Calculation (RPCA-EC) approach as implemented in [1]. In this approach, the following update steps are used to update various quantities using the current measurement vector  $x(t)$  if no fault is detected (i.e.  $\gamma(t) < \chi^2_{1-\alpha, m}$ ) at time  $t$ :

- 1) Mean of each sensor is updated as :

$$\mu_{s,i}(t) = (1-\theta)\mu_{s,i}(t-1) + \theta\tilde{x}_i(t) \quad (5)$$

- 2) Variance (square of standard deviation of each sensor) is updated as follows:

$$\sigma_{s,i}^2(t) = (1-\theta)\sigma_{s,i}^2(t-1) + \theta(\tilde{x}_i(t) - \mu_{s,i}(t))^2 \quad (6)$$

- 3) Covariance matrix of the cluster is updated as :

$$C(t) = (1-\theta)C(t-1) + \theta x(t)x(t)^T \quad (7)$$

where  $\mu_{s,i}(t-1)$  and  $\sigma_{s,i}(t-1)$  are the mean and standard deviation for  $i^{\text{th}}$  sensor based on data up to time instant  $(t-1)$ ,  $C(t-1)$  is the corresponding covariance matrix of the cluster based on data up to time instant  $(t-1)$ ,  $\theta$  is a tuning parameter

known as forgetting factor and  $x(t)$  is the vector of sensor values at time instant  $t$  which has been normalized using the corresponding updated means  $\mu_{s,i}(t)$  and standard deviations  $\sigma_{s,i}(t)$ . After computing  $C(t)$ ,  $m$  least eigenvalues and corresponding eigenvectors are chosen by performing the ratio test as in [1] and the new model matrix  $A$  is calculated by stacking those  $m$  eigenvectors as rows. The covariance matrix of residuals (matrix  $E$ ) is also updated. Note that  $m$  may vary as a function of time. The above update steps are performed only if the current measurement vector is found to be fault-free. In case a fault is detected at current time instant  $t$  then the above updates are not performed.

IV. METHODOLOGY OF OUR PROPOSED WORK

We now present the proposed cluster based normalization approach and its use in online fault diagnosis.

*A. Motivation*

As discussed in Section III, in the approach available in literature ([1], [4]) each sensor reading is normalized using its individual sensor mean and standard deviation. As individual sensor statistics are used so we name this scheme as sensor statistics based normalization (SSBN). In a nuclear reactor, the flux magnitude may undergo sharp variations due to change in reactor power level. The normalization approach presented in [1] may lead to high false alarms in presence of such sharp variations of the flux magnitude. In the current work, we propose a novel cluster based normalization scheme that is not sensitive to sharp variations in the flux profiles as long as all detectors within a cluster register such sharp variations in response to changes in reactor power level.

To motivate our proposed normalization approach, consider the following example:

**Example 1:** Consider the dynamic behaviour of four variables as shown in Fig. 1 where variables  $x_1, x_2$  were generated as:

$$x_1(t) = \begin{cases} 0.01t; & \text{for } 0 \leq t \leq (T/4) \\ 0.015t; & \text{for } (T/4)+1 \leq t \leq (T/2) \\ 600; & \text{for } (T/2)+1 \leq t \leq (T) \end{cases} \quad (8)$$

$$x_2(t) = \begin{cases} 2-0.02t; & \text{for } 0 \leq t \leq (T/4) \\ 2-0.03t; & \text{for } (T/4)+1 \leq t \leq (T/2) \\ -1198; & \text{for } (T/2)+1 \leq t \leq (T) \end{cases} \quad (9)$$

with  $T=80000$ . Variables  $x_3, x_4$  were generated as linear combination of variables  $x_1, x_2$  as

$$\begin{bmatrix} x_3(t) \\ x_4(t) \end{bmatrix} = \begin{bmatrix} 0.2653 & 0.6356 \\ 1.5633 & 0.0883 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \quad (10)$$

Note that there is no fault in any sensor (variable) at any time instant and the sharp changes observed in Fig. 1 represent dynamic variation in the operating values. The corresponding sensor statistics based normalized profiles are plotted in Fig 2. It is seen that normalization is not able to remove sharp variations present in the original data. A diagnostic method working on this normalized data will likely label the sharp changes as faults even though no fault has been introduced in any sensor. We now propose an alternate approach which uses cluster statistics for normalization.

**Paper Id:**
**B. Cluster statistics based normalization (CSBN)**

In cluster statistics based normalization (CSBN), for a cluster, the cluster-mean and cluster-standard deviation of the sensor readings belonging to that cluster at each time instant are computed. As before we consider that the raw measurements of  $n$  sensors for  $N$  time instants are available as the training data set in a given cluster. Then the sensor readings are normalized using the cluster-mean and cluster-standard deviation at that time instant. Let  $\tilde{x}(t)=[\tilde{x}_1(t),\tilde{x}_2(t),\dots,\tilde{x}_n(t)]$  be set of readings of  $n$  sensors belonging to the cluster at time instant  $t$  ( $t = 1,2,\dots,N$ ). Then the cluster mean  $m(t)$  and cluster standard deviation  $\sigma(t)$  at time instant  $t$  are computed as:

$$m(t)=\left(\frac{1}{n}\right)\sum_{i=1}^n \tilde{x}_i(t); \quad \sigma(t)=\sqrt{\left(\frac{1}{n}\right)\sum_{i=1}^n (\tilde{x}_i(t)-m(t))^2} \quad (11)$$

Note that  $m(t)$  and  $\sigma(t)$  are functions of time and can be thought of as instantaneous cluster mean and cluster standard deviation. These values depend only on the sensor measurements at the current instant and are not affected by the past behaviour. After the instantaneous cluster mean and cluster standard deviation are obtained, the cluster statistics based normalized sensor readings are obtained as:

$$y_i(t)=\frac{\tilde{x}_i(t)-m(t)}{\sigma(t)}, \quad i=1,2,\dots,n \quad (12)$$

To see the utility of the above CSBN scheme, consider Example 1. Fig. 3 shows the dynamic profile of the four sensors normalized using CSBN. It can be seen that despite sharp changes in the raw data, the normalized data shows constant behaviour after brief initial transients. A diagnostic approach working with such normalized data will not identify a fault at the time instants when all the sensors underwent sharp changes which were part of normal operation and not caused by any fault.

**C. Use of CSBN for SPND fault diagnosis**

**Training phase or offline phase:** Given the training data  $\tilde{X}=[\tilde{x}(1);\tilde{x}(2);\dots;\tilde{x}(N)]$  for a particular cluster, we consider the cluster statistics based normalized row vector  $y(t)$  at time instant  $t$  as  $y(t)=[y_1(t),y_2(t),\dots,y_n(t)]$ . After calculating  $y(1), y(2), \dots, y(N)$  in this manner we form the matrix  $Y \in \mathbb{R}^{N \times n}$  as  $Y=[y(1); y(2); \dots; y(N)]$  and do the following:

- 1) The covariance matrix of  $y(t)$  is computed from  $Y$  in a similar manner as in (2).
- 2) The model matrix  $A$  is computed where rows of  $A$  are the eigenvectors corresponding to  $m$  least eigenvalues of covariance matrix as stated in Section III-D and  $m$  is found by the ratio test implemented in [1].
- 3) The covariance matrix  $E$  of the residual is obtained by stacking the  $m$  smallest eigenvalues on the diagonal.

**Fault detection phase or online phase:** At current time instant  $t$ , given the readings  $\tilde{x}(t)$ , the cluster mean  $m(t)$  and cluster standard deviation  $\sigma(t)$  are computed using (11), and cluster based normalized readings  $y(t)$  are computed. After that the

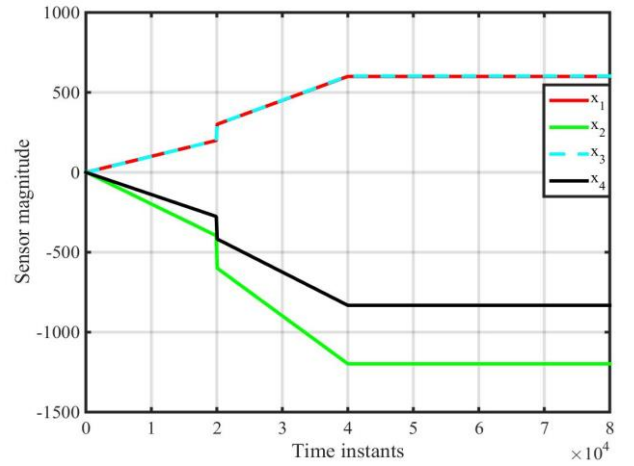


Fig. 1. Dynamic variations in variables

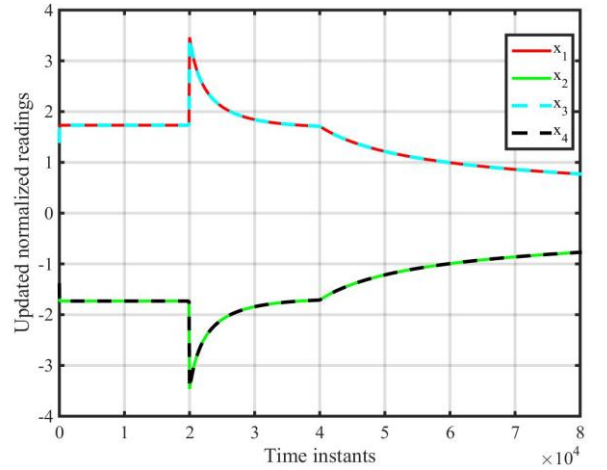


Fig. 2. Normalized values with SSBN scheme

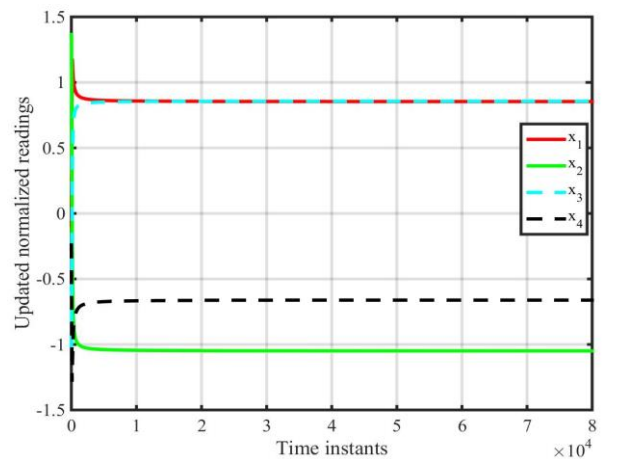


Fig. 3. Normalized values with CSBN scheme

**Paper Id:**

residual vector  $r(t)$  for  $y(t)$  is computed as shown in (4) and correspondingly  $\gamma(t)$  is computed. If  $\gamma(t)$  is less than the  $\chi^2$  threshold then we conclude that the cluster is not faulty and model is updated as done in (7) by using  $y(t)$  instead of  $x(t)$ . If  $\gamma(t)$  is greater than or equal to the  $\chi^2$  threshold, then it is concluded that there is a faulty sensor in the cluster which is then identified using generalized likelihood ratio test. The flowchart in Fig. 4 shows the steps involved in our proposed approach.

## V. RESULTS

In this section we present the comparative results between our proposed CSBN method and the SSBN method available in literature [1] which was summarized in Section III and IV. The data used in our work is the set of measurements of 42 prompt and 102 Vanadium SPNDs of a 540 MWe operating nuclear reactor. Measurements are taken at 1 second sampling time interval for a period of 2 hours during April 2015. The entire data-set is assumed to be fault-free. The measurements of first 1000 time instants are used as training data and the measurements from 1001<sup>th</sup> time instant to 3000<sup>th</sup> time instant are used for comparing the fault diagnostic performances of the two approaches. We present results for a pure Vanadium cluster comprising 10 sensors, obtained using k-means clustering approach applied to the training data. The dynamic profiles of the sensors in this cluster are presented in Fig 5. To compare the diagnostic performances, we introduce different types of faults in sensor V53. The significance level  $\alpha$  (required for computing  $\chi^2$  threshold) is chosen to be 0.05 and the forgetting factor  $\theta$  in (5)-(7) is taken as (1/300) which is the inverse of the time constant of the Vanadium SPND transfer function.

The SSBN and CSBN approaches are compared using the following metrics:

- False alarm rate (FAR) =

$$\frac{\text{No. of time instants chosen cluster detected as faulty} \times 100}{\text{No. of time instants at which no sensor is faulty}}$$

- Detection rate (DR) =

$$\frac{\text{No. of time instants chosen cluster detected as faulty} \times 100}{\text{No. of time instants at which fault is present}}$$

- Identification rate (IR) =

$$\frac{\text{No. of time instants faulty sensor correctly identified} \times 100}{\text{No. of time instants at which fault is present}}$$

### A. Comparison of false alarm rates

When no fault is added to any sensor, the false alarm rates for SSBN and CSBN based approaches are 7.8 % and 0.57%, respectively, which demonstrates the advantage of cluster statistics based normalization.

### B. Comparison in presence of constant bias fault

During this test the bias is added as follows:

$$\tilde{x}_{53}^f(t) = \begin{cases} \tilde{x}_{53}(t); & \text{for } 1 \leq t \leq 1100 \\ \tilde{x}_{53}(t) + 0.01b\mu_{53}; & \text{for } 1101 \leq t \leq 3000 \end{cases} \quad (13)$$

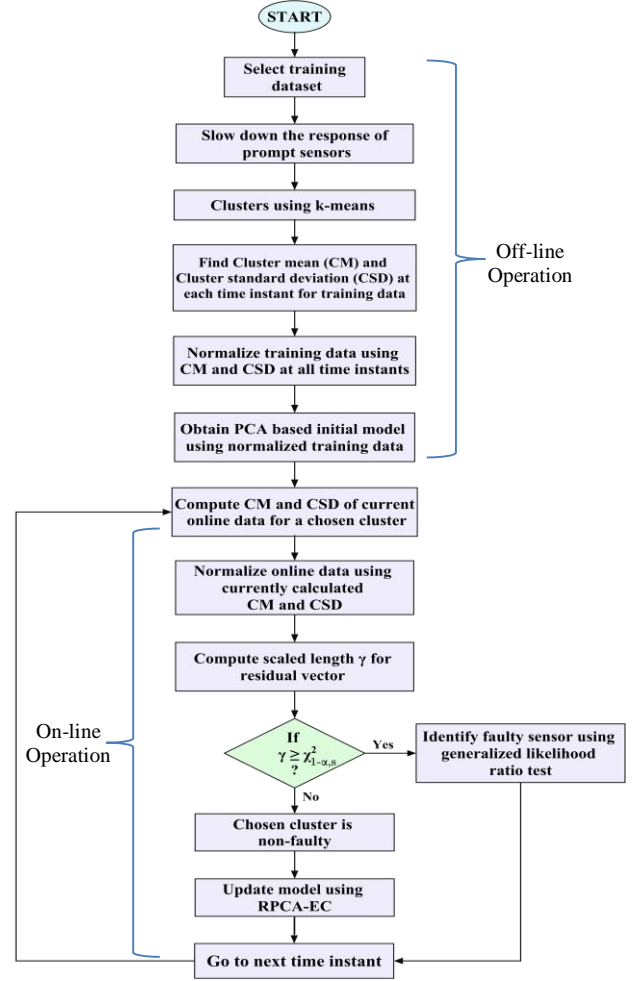


Fig. 4. Overall framework of the proposed work

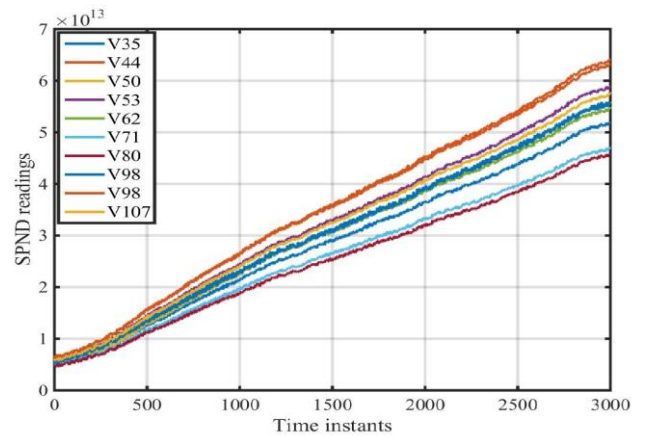


Fig. 5. Dynamic profiles of SPND measurements

where  $\tilde{x}_{53}(t)$  is the un-normalized fault free raw measurement of sensor V53,  $\mu_{53}$  is the sample mean of the V53 measurements in the interval  $1 \leq t \leq 1000$  and  $b$  is the bias percentage.  $\tilde{x}_{53}^f(t)$  is un-normalized measurement value of V53 after the bias addition. The diagnostic results are presented in

**Paper Id:**

Table I. We find that the sensor based normalization results in almost full detection rate when the bias percentage  $b$  is 6 (magnitude is  $8.67 \times 10^{11}$ ) while the cluster based normalization results in full detection rate when the bias percentage  $b$  is 14 (magnitude is  $2.02 \times 10^{12}$ ). This relatively low detection rate may be due to use of faulty data for computing cluster statistics before normalizing the raw sensor measurements in the CSBN approach. But as the sensor raw value is varying from  $1 \times 10^{13}$  to  $7 \times 10^{13}$ , this kind of performance difference may be tolerable.

*C. Comparison when the bias is proportional to the raw value*

$$\tilde{x}_{53}^f(t) = \begin{cases} \tilde{x}_{53}(t); & \text{for } 1 \leq t \leq 1100 \\ \tilde{x}_{53}(t) + 0.01b\tilde{x}_{53}(t); & \text{for } 1101 \leq t \leq 3000 \end{cases} \quad (14)$$

As shown in (14), the bias is time varying and is proportional to the raw value at each time instant. The results for the two approaches are presented in Table II. We observe that SSBN results in very high (95%) DR for  $b \geq 4$ , while CSBN requires the bias percentage  $b$  to be at least 6 to result in very high DR.

*D. Comparison in presence of exponential bias*

$$\tilde{x}_{53}^f(t) = \begin{cases} \tilde{x}_{53}(t); & \text{for } 1 \leq t \leq 1100 \\ \tilde{x}_{53}(t) - c(1 - e^{-0.1(t-1100)}); & \text{for } 1101 \leq t \leq 1220 \\ \tilde{x}_{53}(t) - c(1 - e^{-0.1(t-1100)}) - c(1 - e^{-0.1(t-1220)}); & \text{for } 1221 \leq t \leq 3000 \end{cases} \quad (15)$$

Coefficient  $c$  in the above expressions is chosen as a percentage of  $\tilde{x}_{53}(1100)$ . This kind of error profile can occur in practice due to leakage currents in the SPND cables. The comparative results for the two normalization schemes are given in Table III for different values of  $c$ . We observe that the SSBN approach results in high detection rate for  $c=1.06 \times 10^{12}$  onwards, while the CSBN approach results in high detection rate for  $c=1.59 \times 10^{12}$  onwards. The difference in the  $c$ -values at which these two approaches result in high DR is only  $0.53 \times 10^{12}$  which is much less compared to the raw value of the sensor measurements.

**VI. CONCLUSIONS AND FUTURE WORK**

In this work, we proposed a cluster statistics based normalization approach which is used to develop PCA based online fault diagnostic method for SPNDs. The approach was applied to perform fault diagnosis on data obtained from a nuclear reactor and results were compared with sensor statistics based normalization approach available in literature. The proposed approach resulted in significantly low false alarm rates. However, the detection rates were also low for the proposed approach, especially for constant bias faults, but the detection rates were higher for the more practical scenario of time varying fault. Future work involves modifying the CSBN approach to result in high detection rate for constant bias faults as well.

TABLE I. COMPARATIVE PERFORMANCE WHEN BIAS IS CONSTANT

% of mean	Actual bias magnitude	DR (%) for SSBN	IR (%) for SSBN	DR (%) for CSBN	IR (%) for CSBN
2	$2.8 \times 10^{11}$	7.65	0.8	0.1	0.1
4	$5.7 \times 10^{11}$	11.0	4.05	0.1	0.1
6	$8.67 \times 10^{11}$	95.0	95.0	0.4	0.3
8	$1.15 \times 10^{12}$	95.0	95.0	2.95	2.85
10	$1.44 \times 10^{12}$	95.0	95.0	24.4	24.4
14	$2.02 \times 10^{12}$	95.0	95.0	95.0	95.0

TABLE II. COMPARATIVE PERFORMANCE WHEN THE BIAS IS PROPORTIONAL TO THE CURRENT RAW VALUE

Bias(% of raw value)	DR (%) for SSBN	IR (%) for SSBN	DR (%) for CSBN	IR (%) for CSBN
2	10.25	3.1	0.1	0.1
4	95.0	95.0	2.5	2.5
6	95.0	95.0	95.0	95.0
8	95.0	95.0	95.0	95.0

TABLE III. COMPARATIVE PERFORMANCE WHEN THE BIAS IS OF EXPONENTIAL NATURE

(%) of $\tilde{x}_{53}(1100)$	Value of $c$	DR (%) for SSBN	IR (%) for SSBN	DR (%) for CSBN	IR (%) for CSBN
2	$5.3 \times 10^{11}$	7.65	0.5	0.05	0.05
4	$1.06 \times 10^{12}$	94.7	94.65	0.1	0.1
6	$1.59 \times 10^{12}$	94.7	94.65	94.5	94.5
8	$2.12 \times 10^{12}$	94.7	94.65	94.65	94.65

**References**

- [1] A. Bhattacharyya, V. Yogi, S. Singla, M. Bhushan, M.G. Kelkar, A.P. Tiwari, M. Pramanik and M.N. Belur, "Adaptive, online models to detect and estimate gross error in SPNDs", 3rd IEEE Indian Control Conference, Indian Institute of Technology Guwahati, pp. 149-154, 4-6 January 2017.
- [2] A.K. Jain and R.C. Dubes, Algorithms for Clustering Data, Englewoods Cliffs, NJ, USA, Prentice-Hall, 1988.
- [3] P.K. Tamboli, A. Sharma, A.D. Prasad, N. Singh, J. Antony, M.G. Kelkar, Kaurav, Reetesh, and M. Pramanik, "Analysis of effect of cable degradation on SPND IR calculation", Proceedings of the National Symposium on Nuclear Instrumentation (NSNI), Mumbai, India, 18-21 November 2013.
- [4] R.A. Razak, M. Bhushan, M.N. Belur, A.P. Tiwari, M.G. Kelkar and M. Pramanik, "Data reconciliation and gross error analysis of self-powered neutron detectors: Comparison of PCA and IPCA based models", International Journal of Advances in Engineering Sciences and Applied Mathematics, 4(1-2), pp. 91-115, June 2012.
- [5] R.A. Razak, M. Bhushan, M.N. Belur, A.P. Tiwari, M.G. Kelkar and M. Pramanik, "Clustering of self powered neutron detectors: combining prompt and slow dynamics", IEEE Transactions on Nuclear Science, 61(6), pp. 3635-3643, December 2014.
- [6] S. Narasimhan and R.S.H. Mah, "Generalized likelihood ratios for gross error identification in dynamic processes", AIChE Journal, 34(8), pp. 1321-1331, August 1998.
- [7] V. Venkatasubramanian, R. Rengaswamy and S.N. Kavuri, "A review of process fault detection and diagnosis: Part I: Quantitative model-based methods", Computers & Chemical Engineering, 27(3), pp. 293-311, March 2003.
- [8] W.H. Todt, "Characteristics of self powered neutron detectors used in power reactors", Proceedings of the European Nuclear Congress (ENC), Nice, France, 1998.