Application of Data Mining Techniques for Sensor Drift Analysis to Optimize Nuclear Power Plant performance

S.Narasimhan, Rajendran

Abstract: ***The Power Plants are engineered and instrumented to ensure safety in all modes of operation. Hence they should be continuously monitored and maintained with necessary Instrumentation to identify performance degradation and the root causes to avoid calling for frequent maintenance. The degraded performance of Instrumentation & Control systems may also lead to plant outages.***

Different studies have suggested that a well maintained instrumentation with errors and response times within the permissible limits may increase the availability minimizing outages.

The I&C systems are designed for monitoring, control and safety actions in case of an event in a power plant. The sensors used are single, redundant, triplicated or diverse based on the type of application. Where safety is of prime concern, triplicated and 2/3 voting logic is employed for initiating safety actions. Diverse instruments are provided for protecting the plant from any single abnormal event. Redundant sensors are used to improve plant availability.

Wherever 2/3 logics are used, the sensors shall uniformly behave and the drifts across the sensor may lead to crossing the threshold, initiating a protective action. Instead of waiting for the regular preventive maintenance schedule for recalibrating the sensors, the drift in the sensors are analyzed by developing a combined overall online monitoring parameter which will give an early warning to the operator the need for recalibration of the redundant sensors.

This paper deals with development of one such parameter through data mining techniques for a representative process in a nuclear power plant.

Keywords: Calibration/Drift Measurement, Data-driven models, Data Collection, Data Mining, Data Pre-processing, Knowledge Discovery in Database, Mahalanobis Distance, Summary Statistics ,Z-score normalization, ,covariance matrix.

I. INTRODUCTION

All safety class instrumentation & control (I&C) systems, for nuclear industry have been effectively categorized by the International Electro-technical Commission (IEC), standard IEC-61226, as Class-IA (Safety), Class-IB (Safety Related) and Class-IC (Non-Safety). Specifically for both safety & safety-related instrumentation, it is essential to follow all the good engineering practices, right from the design stage to manufacturing, testing, calibration and subsequent operation & maintenance. The IEC-61226 standard [1], establishes the criteria and methods to be used to assign the I&C functions to categorize them based on the importance of the function for safety. Further, it also defines the mandatory characteristics for each of the category including requirements such as functional, performance, stability, stress, interfaces, QA, reliability, testing & maintenance. This standard also introduces the additional set of associated IEC standards, which together with IEC-61226, provide the essential guide-lines for design of safety and safety-related instrumentation.

On similar lines, the IEC-61508 standard [2], defines the functional safety of electrical, electronic and programmable electronic equipment. This standard focuses attention on risk based safety-related system design which should result in far more cost-effective implementation. It may also be noted that various other associated IEC standards, have evolved, for design & development of safety class instrumentation and computer based systems, essentially targeting the requirements of nuclear industry, but can very well be used for aerospace, chemical, petro-chemical, pharmaceutical, etc.

In order to effectively address plant safety, primarily, the total I&C (Instrumentation & Control) system configuration considers various safety criteria, which is built in to the plant layout, viz. single-failure, fail-safe, redundancy, de-energize to actuate, etc. are listed & explained in IAEA documents [3]. In addition, every sensor, transmitter, signal conditioning modules, alarm-trip unit modules, etc. undergo, necessary life-cycle testing in order to adhere to the safety requirements, as spelt out in the various standards. Further, it is also important that the instrument set-point & calibration uncertainties are also factored in, so as to accurately fit the requirements of the process LCO (limiting conditions of operation), LSS (limiting safety settings) and the actual safety limits. Considering all these, it becomes essential to understand in quantitative terms, the uncertainties in different sensors & transducers and the calibration & ageing management practices, during the operation & maintenance (O&M) phase of the installed instrumentation. This is all the more important in safety instrumentation and plant protection systems, since the true reading determines the limiting condition of operation (LCO) which must lie below the limiting safety settings (LSS), which in turn should be below the actual safety limits of the process being monitored/controlled.

Revised Manuscript Received on October 15, 2019

S.Narasimhan, IT&Instrumentation, Bharatiya nabhikiya Vidyut Nigam Limited, BHAVINI, Kalpakkam, India. E- Mail: snsimhan@igcar.gov.in.

Dr.Rajendran, Department of ECE , VELS Institute of Science, Technology and Advanced Studies(VISTAS) Chennai, India E-Mail: director.ece@velsuniv.ac.in.

II. UNCERTAINTIES IN SENSOR/TRANSDUCER

The primary sensors used for various field measurements, such as bourdon/capsule/diaphragm, (for gauge pressure), orifice/venturi (for DP), thermocouple/RTD (for temperature), etc. which are mounted on the process equipment or inserted in the piping, are designed with the codal conformance matching with the mechanical/process standards and undergo similar manufacturing & in-stage testing procedures. However, the uncertainties and the response times need to be theoretically computed, so as to match the steady-state accuracy & transient response & requirements of the total instrumentation & control loop performance. Thus, in nutshell, one needs to factor in the following uncertainties/errors, during design and installation:

- Computation of Uncertainty in Design of Primary Sensors
- Inherent Errors
- Installed Errors
- Total Error Computation
- Time Constants

The manufacture of all precision sensors, for measurement of pressure, temperature, mechanical displacement, etc. necessarily undergoes the required study of the material properties, property changes due to the applied stressors along with the necessary compatibility with the process fluids (for all wetted parts), conformance to linearity in the selected operating range, response time to step changes in the process, etc. Details of such manufacture & fabrication process are well presented in [2, 3] and also in a variety of manufacturer's catalogues, wherein the modern-day 3D-simulation techniques, used to understand mechanical, thermal behavior of the sensors, are adequately explained. The limits of operation pertaining to stated accuracy and the inherent uncertainty in the design are also available and requires to be factored in, for all safety instrumentation.

Considering a temperature instrumentation system as a case study, it is important to understand the variables in a temperature measurement process, which requires a computation of the model of heat transfer from the process to the thermometer. Since the temperature measurement involves the transfer of the heat energy from the process through the thermo-well (as applicable) and the sheath material, it is also necessary to model the thermal conductivity and the transfer functions of these, in order to account for the uncertainty and the delay. Further it will also be necessary to account for the heat transfer coefficient of the various interfaces, which requires knowledge of the interface geometry and the corresponding thermal conductivity.

III. MEASUREMENT DRIFTS AND CALIBRATION

During the Operation & Maintenance (O&M) phase of the safety-class instrument, it is important to periodically assess the health of the sensor and also drifts found in monitoring and measurement. Calibration of the various field sensors, with laboratory standard instruments, traceable to primary standards, are in practice.

- Traceability in Calibration
- Calibration/Adjustment of Field Instruments
- Accounting for Errors in Lab & Field Instruments

As brought out by Kadilec.P et al in [4] the drifts observed in monitoring and measurements can be attributed to either process or sensor. The process may drift due to changes in the process conditions or due to some external factors. There can be a many mechanical equipment which are under continuous operation and may undergo a steady wear and tear during the operation of the plant. This may affect the process itself, e.g. the flow of coolant in a process may decrease due to the wear of mechanical pumps resulting in changes in process temperature. Drifts can also happen due to external influences like changing environmental conditions. Hence these drifts should be recognised, reported and appropriate corrective actions have to be taken to remove their cause. Whereas sensor drifts are caused by unintended variations in the functioning of sensors and cannot be attributed to the process itself. However this type of drifts, while still observed in the measured data, does not indicate any process related changes. Therefore when drifts in the sensors are observed, it should be taken for re-calibration without performing any corrective actions to the process.

M/s Beemex Corporation has brought out in their white paper on "Optimal Calibration Parameters on Process Instrumentation" [5], the following issues:

Determining a proper calibration interval is an educated guess based on several factors. A best practice is to set a conservative interval based on what the impact of a failure would be in terms of operating in a safe manner while producing product at the highest efficiency and quality. It is also important to review calibration methods and determine the best practices where there will be a minimal impact on plant operations. By focusing on the most critical instruments first, an optimum schedule can be determined and would allow for less critical testing if personnel have availability.

Since all instrumentation drifts no matter the make/model/technology, suppliers end up creating vastly different specifications making it difficult to compare performance. Many times, there are several complicating footnotes written in less than coherent terminology. Instrument performance is not always driven by price. The only true way to determine an optimum interval is to collect data and evaluate drift for a specific make/model instrument over time.

When collecting data on calibration, it is a good practice to not make unnecessary adjustments. For example, if the tolerance is +/-1% of span and the instrument is only out by -0.25% of span, an adjustment should not be made. How can drift be analysed (minimum of 3 points) with constant adjustment? For certain "personalities," not adjusting can be a challenge (people strive for perfection), but note that every time an adjustment is made, drift analysis gets more difficult.

In general, a best practice is to avoid adjusting until the error is significant. With a consistent schedule, a trim most likely will be needed on the next calibration cycle and not cause an As Found "Fail" condition. Of course, this may not be possible due to criticality, drift history, erratic scheduling or other factors, but when possible, do not automatically make calibration adjustments.

What if the drift is inconsistent, both increasing, then decreasing over time? More analysis is required; for instance, are the ambient conditions extreme or constantly changing? Depending on the process application, instrument performance may be affected by media, installation, throughput, turbulence or other variables. This situation indicates there is a level of "noise" associated with drift. When this is the case, analysis should show there is a combination random error and systematic error. Random error consists of uncontrollable issues (ambient conditions and process application) vs. systematic error that consists of identifiable issues (instrument drift). By focusing on systematic error and/ or clear patterns of drift, a proper calibration interval can be set to maximize operation efficiencies in the safest manner possible.

A. Modelling for Calibration/Drift Measurement

Estimation and monitoring of changes in the statistical parameters of the sensor/instrumentation outputs, form an important approach towards detection of drifts in sensors/instrumentation, both by a model-based estimation approach and/or a data processing approach. For effective implementation of on-line monitoring scheme there is a need to develop dynamic models of the instrumentation system so as to simultaneously obtain estimates of process variables and compute calibration uncertainties.

Physics-based models are developed based on the physical properties of the system and can be demonstrated by established theories. These models are general in nature and can precisely describe the process and its conditions. Further these models do not require experimental data and thus not limited to certain environmental or operational conditions. However building such a physics based models are quite difficult and complex. A model with better accuracy will be more complex, costlier and difficult to build. Sometimes it will be practically impossible to build such a model with a desired accuracy.

Whereas a data driven model are not based on the physics of the system. These models are developed from observed measurements of the processes. The models are built and trained using the historical data of measurements. Building a data driven model do not require any prior knowledge of the physics of the processes. Hence such models can be built practically for any kind of system complexity compared to the physical models. Data-driven models are more reality related and describe the true conditions of the process in a better way.

The model thus developed for the sensor can be used as a goodness function test against the recorded system data. Subsequently, the difference between the measured value (as obtained in the data acquisition system) and the estimated value (as computed from process models) is monitored for change, thereby prompting a human operator on instrumentation performance and ageing management.

Thus instead of waiting for the regular preventive maintenance schedule for recalibrating the sensors, the drift in the sensors are analyzed by developing a combined overall online monitoring parameter which will give an early warning to the operator the need for recalibration of the redundant sensors.

This paper deals with development of one such parameter through data mining techniques for a representative process in a nuclear power plant.

IV. KNOWLEDGE DISCOVERY IN DATABASE

In order to implement such calibration strategies during the O&M phase, it is necessary to have a digitized infrastructure, wherein most of the important plant parameters are available on computer based data acquisition systems. Subsequently, various, data processing algorithms including data correlation, parity among multi-variate data, etc. can be implemented in the back-end of the plant computer display screens or in plant historians.

Knowledge Discovery in Databases (KDD) refers to the process of extraction of hidden, previously unknown and potentially useful information from data stored in databases. The KDD process helps us to extract a valid, novel, and ultimately understandable patterns in data. The KDD process is an iterative process with many inter process interactions involving several steps such as

- Data Selection
- Data Pre-processing
- Data Transformation
- Data Mining
- Interpretation & evaluation

A typical steps of KDD are indicated in figure-1.

A. Data Mining

Data mining process extracts the information and patterns from large sets of data that can be presented in a form useful for the analyst. It involves the use of statistical and mathematical processes to derive patterns and trends that can be extracted from the data. These patterns and trends can be collectively called as a data mining model.

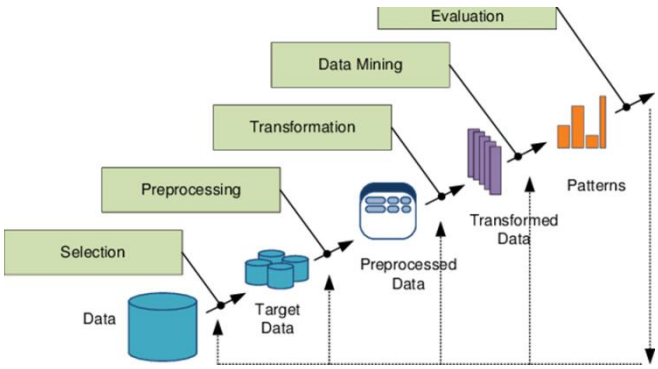

Figure 1 Steps of KDD

A data mining model is built by a process of generating recursive queries about the data that is collected and trying to find answers to such queries using statistical and mathematical means. The various tasks that can be deployed in a data mining process are as follows:

1. association
2. sequence or path analysis
3. clustering
4. classification
5. regression
6. visualization

The type of the task required to be performed in a data mining process depends on the kind of analysis that is required to be done.

A Regression analysis is done to predict new values based on the past inferences. The new values are computed for a dependent variable based on the values of one or more measured attributes. Classification process divides samples into classes. It uses a trained sets of previously labelled data for this purpose. Clustering is a process of grouping the data set into subsets based on some common characteristics. Classification is in some way similar to the clustering, but requires that the analyst know ahead of time how classes are defined.

Based on the data mining approach, a common monitoring parameter for drifts in the process variables of temperatures are discussed in this paper for a representative process in a Nuclear Power plant.

V. DESCRIPTION OF THE PROCESS UNDER STUDY

A typical coolant circuit of a nuclear power plant consists of a vertical centrifugal single stage pump housed in a fixed shell called a pump tank having the free liquid level. The space above the liquid is filled with inert cover gas. The pump shaft is guided by a hydrostatic bearing at the bottom and a thrust bearing at the top. Three mechanical seals, cooled by oil are provided at the top to prevent the leakage of cover gas into atmosphere. The oil acts as a buffer fluid. Two independent oil circuits are provided to circulate the oil in the three mechanical seals; one for the bottom mechanical seal

and the other one for the top bearing and mechanical seal assemblies. There are two oil catch pots provided below the bottom mechanical seals. The oil circuits are designed such that the maximum temperature rise in the oil is less than certain threshold under normal operating conditions. Alarms and trip are provided in case the oil temperature exceeds the set points. The coolant pump is designed to trip on high seal oil temperature and low oil flow/low oil pressure.

The oil outlet from the coolant pump bottom mechanical seal has three triplicated temperature transmitters (TTRol20-804A, B and C) of range 0-1000C having the following logics:

When the temperature measured by three temperature sensors exceeds 670C, an alarm is given in the control room and if it exceeds 700C as measured by 2/3 sensors, it trips the bottom circuit oil pumps Pump-A/ Pump-B. This in turn may trip the main coolant Pump due to lack of lubrication and further may lead to reactor trip.

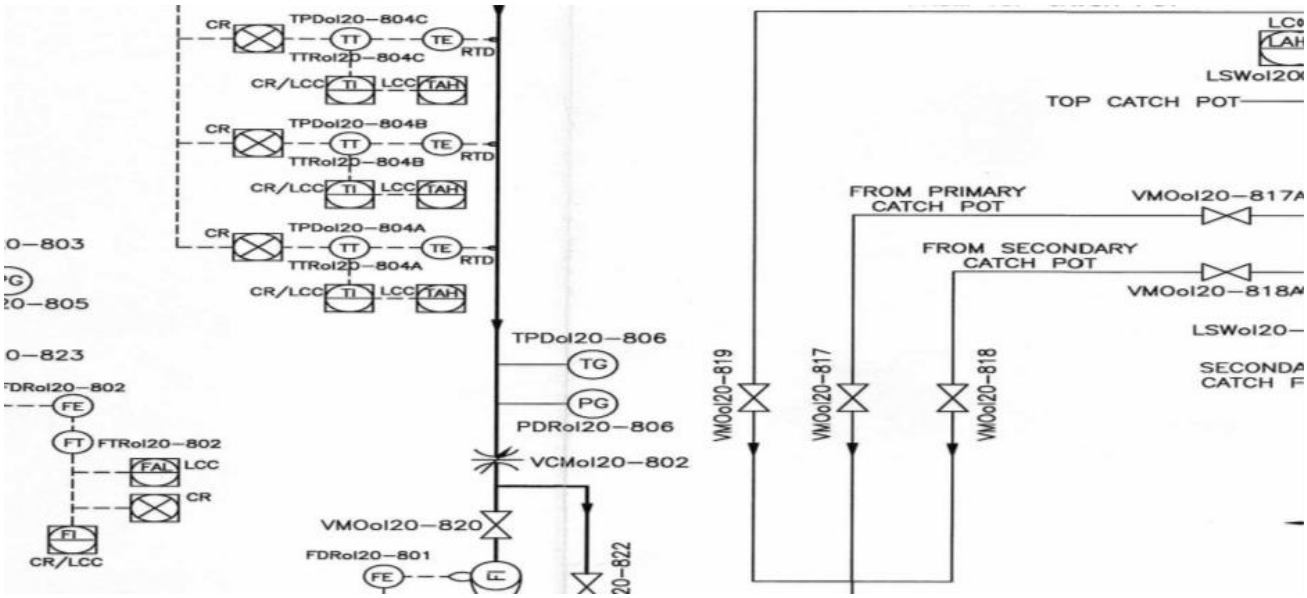

Fig.2 Lubricating Oil System Sensors and Interlock

A. Why This Sensors for Modelling

Triplicated sensors are generally used for protection of critical assets or a system. Ideally they are expected to read equal. If at least two out of the three sensors crosses the threshold, then protective action is initiated. This also indicates that the disagreement between the sensors due to drifts and not attributed to the process, also may result in spurious trips. On a long term, it may indirectly lead to reactor trip. Whereas there is a scheduled PM activity carried out to tackle the drifts by periodic calibration, it is possible to optimise the calibration interval by analysing the performance and possible drifts through data mining and analysis thereby avoiding costly outages to instrumentation drifts.

Hence these triplicated temperature monitoring sensors in lube oil circuit were taken for the study to identify a common drift indicator through data mining process.

VI. DATA MINING STEPS FOLLOWED IN THIS STUDY

A. Data Collection:

In order to obtain a large data sets for data modelling, it is necessary to have a digitized infrastructure, wherein most of the important plant parameters are available on computer based data acquisition systems. The subject Nuclear Power plant has a distributed digital Control System where all the plant parameters are scanned and logged in the historian. The scan time for each of the parameters is based on the criticality of the measurement. The data on the past measurements of the process variables under study for a period of one month was collected for the analysis. The collected data was of interval 5 seconds. Approximately 8 lacs of data tuples were taken up for data mining analysis.

B. Data Pre-processing:

The collected data was pre-processed by inspecting them for a possible missing values, noisy values etc.

Missing values in the context of a nuclear power plant may be due to various reasons. The general reasons for missing values are the unexpected failure of the field sensor or removal of the sensor for maintenance and calibration etc. As nuclear power plants have large number of instrumentation to cater to monitoring,control,protection and safety, the recorded process data may have a large number of diverse and redundant field variables. Hence it is always possible during the stable operation of the plant that a field instrument could have been taken out of service due to maintenance or failures. Missing data can also result due to loss of communication between the sensors and the data acquisition system, problems in assessing the server database, issues in historians, mismatch in connectivity protocols etc.

As data modelling techniques cannot handle a missing data situation efficiently, it is required to adopt suitable mechanisms to deal with such situations. There are different methods to deal with such a situation. A commonly applied practice is to provide the missing values with the average value of the observations so far. Another method is to delete the data tuple consisting of the missing values, i.e., case deletion.

In this study, as steady state analysis was performed, the data tuples that had missing values were deleted from the data set for analysis.

Out of the 8 lacs of data tuples, 1 lac samples were drawn by Random sampling without replacement. This type of sampling methodology had the following implications [17]:

(a) It gives each data tuple in the population an equal probability of getting into the sample; and all choices are independent of one another.

(b) It gives each possible sample combination an equal probability of being chosen.

Thus the data set so selected was unbiased and is a true representation of the process data under steady state.

As the intent of the study is to monitor the drift in the sensors, the individual deviations of each sensor are taken for analysis and study using the following calculated attributes.

1. Average outlet oil temperature

TTRol20_804Av= $\frac{A+B+C}{3}$ where

A is TTRol20_804A
B is TTRol20_804B
C is TTRol20_804C

2. Deviation of sensor A from average value
"Difference A"= TTRol20-804A- TTRol20-804Av
3. Deviation of sensor B from average value
"Difference B"= TTRol20-804B- TTRol20-804Av
4. Deviation of sensor C from average value
"Difference C"= TTRol20-804C- TTRol20-804Av

C. Descriptive Data Analysis

a) Summary Statistics of the Attributes:

Summary statistics summarize and provide information following information about the sample data:

- Measures of location (also called central tendency).
- Measures of spread.
- Graphs/charts.

The summary statistics of the variables (triplicated temperature sensor values) for both total population and samples were analysed for both the physical sensors and the calculated variables. The summary statistics were analysed for

1. Temperature Transmitter TTRol20-804A
2. Temperature Transmitter TTRol20-804B
3. Temperature Transmitter TTRol20-804C
4. Average outlet oil temperature TTRol20-804Av
5. Deviation of sensor A from average "Difference A"
6. Deviation of sensor B from average "Difference B"
7. Deviation of sensor C from average "Difference C"

The various summary statistics of the physical, calculated and the average attribute are brought out in the table below:

Attribute	Summary Statistics values									
	Mean	Median	Low	High	Std.Dev	skewness	Kurtosis	5% percentile	95% percentile	IQ range
TTRol20-804A	50.994	51.410	28.680	54.130	3.0020	-5.5414	33.242	49.860	52.950	1.7100
TTRol20-804B	50.979	51.370	29.070	54.090	2.9595	-5.5017	33.841	49.860	52.930	1.7100

TTRol20-804C	50.364	50.770	28.440	3.450	2.9881	-5.5696	33.368	49.270	52.270	1.6700
TTRol20-804Av	50.779	51.183	28.740	53.877	2.9827	-5.5392	34.162	49.667	52.717	1.7000

Table.1 Summary Statistics of the physical attributes

Attribute	Summary Statistics values									
	Mean	Median	Low	High	Std.Dev	skewness	Kurtosis	5% percentile	95% percentile	IQ range
Difference A	0.21529	0.21667	-1.2967	0.5566	0.0610	-1.8101	30.609	0.1200	0.3000	0.0800
Difference B	0.19974	0.19667	-0.13667	1.4500	0.05331	2.9720	34.155	0.13667	0.26333	0.05667
Difference C	-0.41503	-0.41667	-0.88667	0.0366	0.05515	-1.2601	11.611	-0.4833	-0.3367	0.0633

Table.2 Summary Statistics of the calculated attributes

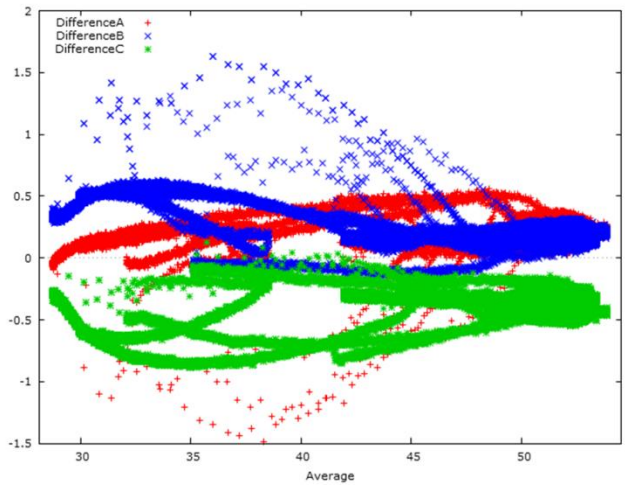

Figure 3 Scatter Plot of Deviations from the Average

The summary statistics reveals the following information:

1. The mean value of redundant sensor TTRol20-804C is comparatively lower than TTRol20-804A& TTRol20-804B.
2. The distribution of the values of both physical and calculated attributes are not normal.
3. All of them except Difference B attribute exhibit a strong negative skewness indicating that the left hand tail is longer. Difference B attribute exhibits a positive skew
4. The distribution also indicates leptokurtic behaviour with thick tails with Difference C attribute indicating a lower value compared to the other attributes.
5. There exists a significant negative bias in the deviations of sensor C from the average compared to the deviations of A&B.

D. Cluster Analysis of the Samples

a) Scatter Plots of the Samples

Scatter plots of Deviation of sensor values from average value was plotted with against the average value in Figure-3 above.

The scatter plot reveals the following information on the attributes:

1. The pattern of their distribution against the average indicates similarity.
2. The range of deviations from their average is higher at lower temperatures.
3. There exists a fixed offset in the readings between channels A and B &C.
4. There are few out of bound values/outliers in all the sensor readings of A&B.

b) Data Transformation

The process instrumentation values as collected by the data acquisition system may have different ranges depending on the parameter measured, the units of measurement and the behavior of the process. This characteristic will make identification of the system behavior difficult in a data analysis process since variables with larger range of values may dominate over variables with smaller ones. Hence, to bring all the variables to a common standard for processing, we may perform data scaling.

The most common scaling methods are the min-max and the z-score normalizations defined by [6]

Min-max normalization:

$$x' = \frac{(X_i - X_{min})}{(X_{max} - X_{min})} \times (New_{max} - New_{min}) + New_{min}$$

Z-score normalization: $x' = (xi - \mu)/\delta$

Where x_i is the unscaled measurement

x' is the scaled version of the variable

x_{min} is the minimum value of the variable range of measurement.

x_{max} is the maximum value of the variable range of measurement.

μ is the mean value of measurement

δ is the standard deviation of the variable.

Whenever the data is observed to have outliers, the more robust z-score normalization approach is commonly preferred.

As identified in the scatter plot and the summary statistics, to normalize the bias and to bring all the readings to a common reference, the normalization of data tuples were done using z-score.

All variables are normalized with a mean of 0 and standard deviation of 1 with the z-score.

Now the root mean square deviation value (Drms) of the sensor readings from the average value for each of the tuple as a single deviation indicator was calculated as follows:

$$Drms = \sqrt{(A - Av)^2 + (B - Av)^2 + (C - Av)^2}$$

Where
A is TTRol20_804A
B is TTRol20_804B
C is TTRol20_804C
Av is TTRol20_804Av

The summary statistics of the obtained values are as listed in Table-3:

Attribute	Summary Statistics values									
	Mean	Median	Low	High	Std.Dev	skewness	Kurtosis	5% percentile	95% percentile	IQ range
Drms	0.76030	1.5828	0.0085	3.6004	0.82495	-0.12427	-1.1537	0.17944	2.6978	1.4419

Table.3 Summary Statistics of the deviation values

As can be seen from the table, the deviation value is not normally distributed. It has a negative skew indicating that all deviations are positive. (Since sum of square of deviations are measured). The distribution is platykurtic with light tails. As 95 percentile value is 2.69, the data tuples whose deviation thus calculated above as 2.7 are only considered for further analysis. Keeping only those tuples with that threshold, 3714 outliers were detected and removed.

VII. MAHALANOBIS DISTANCE

The Mahalanobis distance is a measure of the distance between a point P and a distribution D, introduced by P. C. Mahalanobis in 1936[7]. This value generalizes on multiple dimensions of the data to identify the distance of point P from the mean of the distribution D in terms of the standard deviations. If P is exactly at the mean value of D, then the distance from P to D will be 0. Further the distance increases as P moves away from the mean along each principal component axis. If the axes of the distribution is re-scaled to have unit variance, then the Mahalanobis distance will correspond to standard Euclidean distance in the transformed space.

The Mahalanobis distance is thus unit less and scale-invariant, and takes into account the correlations of the data set. The Mahalanobis distance measures distance relative to the centroid — a base or central point which can be thought of as an overall mean for multivariate data. The centroid is a point in multivariate space where all means from all variables intersect. The larger the mahalanobis distance, the further away from the centroid the data point is.

The most common use for the Mahalanobis distance is to find multivariate outliers, which indicates unusual combinations of two or more variables.

Thus in case of a highly correlated variables like the triplicated temperature sensors in this study, if there is no specific deviation of one or more sensors, the mean value of all the three sensors will intersect at a common point at which the calculated mahalanobis distance will be 0.

Whereas in case of variation of one or more variables from their mean, the calculated distance from the centroid will be more. The larger the mahalanobis distance, the further away from the centroid the data point is and thus the value can be sensed as a deviation from the expected reading.

Hence mahalanobis distance calculated for the triplicated sensor can be the single best indicator of drift in any of the sensor readings. Monitoring the distance in real time will indicate how the sensors are behaving and setting a threshold will help us to generate the required annunciation to take up corrective calibration action.

A. Calculation of Mahalanobis Distance

Mahalanobis distance is different from Euclidean distance due to its treatment of correlation and follows the following steps:

1. It converts the correlated variables into uncorrelated
2. Perform data transformation so that mean value is 0 and variance is 1
3. It calculates the Euclidean distance on the transformed variables.

The formula to compute Mahalanobis distance is as follows:

$$D^2 = (x - m)^T \bullet C^{-1} \bullet (x - m)$$

where,

- D^2 is the square of the Mahalanobis distance.
- x is the vector value of the observation (row in a dataset),
- m is the vector of mean values of independent variables (mean of each column),
- C^{-1} is the inverse covariance matrix of independent variables.

(x – m) indicates the distance of the vector from the mean value. The distance so arrived is now divided by the matrix of covariance C.

If the variables are highly correlated, then, the covariance will be high. Dividing by a large covariance will effectively reduce the distance.

Likewise, if the variables are less correlated, then the covariance will be low and the distance will not be reduced so much.

Thus mahanalobis distance effectively addresses both the problems of scale as well as the correlation of the variables.

In our study of the process with the triplicated temperature sensors, after removal of 3714 outliers as in section 4.4.2, the mahanalobis distance is calculated for 96286 data tuples. The calculated variable (Mdist) is arrived as:

Mahalanobis Distance (MDist) for the samples under study

MDist = $(x-m)^T \bullet C^{-1} (x-m)$ where

X is the matrix (3 X 1) of vector (TTRol20-804A TTRol20-804B TTRol20-804C)

m is the matrix (3 X 1) of average of the vectors (Mean of A Mean of B Mean of C)

C is the covariance matrix (3 X 3) of the triplicated sensors TTRol20-804A, TTRol20-804B and TTRol20-804C.

Covariance matrix (3 x 3) as calculated for the samples under study are:

9.2612	9.1246	9.2144
9.1246	8.9982	9.0841
9.2144	9.0841	9.1782

Inverse of covariance matrix C-1 (3 x 3) is

134.80	-88.415	-47.824
-88.415	196.60	-105.83
-47.824	-105.83	152.86

The summary statistics of the obtained values are tabulated Table-4.

Attribute	Summary Statistics values									
	Mean	Median	Low	High	Std.Dev	skewness	Kurtosis	5 %Percentile	95 %percentile	IQ range
MDist	1.1693	1.1282	0.0176	7.9175	0.58467	1.1700	4.1238	0.33130	2.1076	0.72822

Table.4 Summary Statistics of the Mahanalobis Distance

As can be seen from the table, the MDist value is not normally distributed. It has a positive skew indicating that the right hand tail is longer. The distribution also indicates leptokurtic behaviour with thick tails.

Scatter plots of Mahalanobis distance variable Mdist with respect to the deviations of individual temperatures from their mean was plotted as below:

As seen, the Mdist value increases with increase in drift of the temperature values for each of the sensor. The variations in the MDist values are the result of combined variations in all the three sensors. The scatter plot also reveals a bidirectional behaviour considering the directions of deviations of the individual sensors. Hence this can qualify to be a composite variable for drift and can effectively monitor the overall behaviour of these triplicated sensors.

A suitable thresh hold can be fixed for this variable to provide an indication for drift in sensor values and alert for calibration requirement. This model can be deployed online to generate alert.

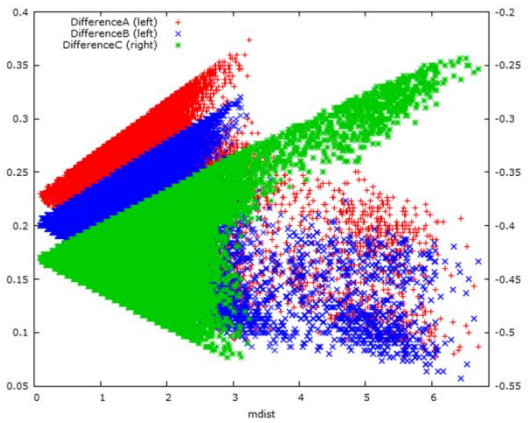

Figure 4 Scatter Plot of MDist with respect to deviations in the sensor

VIII. CONCLUSION AND FUTURE SCOPE OF STUDY

With the statistical observations, the drift in triplicated sensors of oil temperature of main coolant pumps can be monitored by a single modelled variable- MDist. This study concentrated on finding a common drift variable of redundant temperature sensors of the process monitoring the temperature at the same location. The study can also be applied to process sensors of spatial redundancy also by using suitable data transformations. For example, the study can extend to identifying composite drift variable for oil outlet temperature and inlet oil pressure which will govern the efficiency of the process.

Further subsequent to this observation, it is also possible to extend the study to estimate the values of MDist through regression analysis, artificial neural networks, Bayesian estimation etc., so that an annunciation can be generated for the maintainer to plan for calibration of these sensors if the MDist value so obtained by the model exceeds the threshold value decided based on the process. Further the model can also estimate the time required for the MDist value to reach

The threshold level given the present readings. This will be helpful in finalising the schedule for Preventive maintenance and periodic calibration

ACKNOWLEDGMENT

I am grateful to all who have supported me in every manner for this study and would especially like to thank the following:

Dr. Kallol Roy, CMD, BHAVINI and Shri V. Rajan Babu, Director (Technical) for their encouragement and support.

I am greatly indebted for the kind permission given to me to take up the study in the process system and to have an access to the data for the purpose of study.

Dr. V. Rajendran, Director-ECE,VELS Institute of Science and Advanced Studies (VISTAS) for giving me the opportunity to be a PhD student in Electronics & communication Engineering and for the encouragement and support provided by him.

REFERENCES

1. IEC-61226 : International Standard, Edition 3.0, 2009-07, Nuclear Power Plants --- Instrumentation & Control Important to Safety --- Classification of Instrumentation & Control Functions.
2. IAEA Safety Glossary, 2007.
3. IEC-61508 : Standard for Functional Safety of Electrical/ Electronic/ /Programmable Electronic Safety Related Systems.
4. Kadlec, P., Gabrys, B., & Strandt, S., Data-driven Soft Sensors in the Process Industry, Computers and Chemical Engineering (2008).
5. Optimal Calibration Parameters for Process Instrumentation, white paper on calibration, Beemex, www.beemex.com.
6. Ana Sofia Ramos Brásio,"Industrial process monitoring methodologies", University of Coimbra, PhD thesis in Chemical Engineering,2015.
7. Mahalanobis, Prasanta Chandra , "On the generalised distance in statistics". Proceedings of the National Institute of Sciences of India. 2 (1): 49–55,(2016).
8. Generic Qualification of the Rosemount 3051N Pressure Transmitter, EPRI, June 2001
9. Alexander Wiegand, WIKA Handbook : Pressure & Temperature Measurement, Wika Alexander Wiegand GmbH & Co., Caruna Druck, Mittenberg, 1995
10. Douglus M. Considine, Process/Industrial Instruments & Controls Handbook, 4th Edition, McGraw Hill Inc. 1993
11. Ernest O. Doebelin, Measurement Systems, Applications & Design, McGraw Hill, 1990
12. IAEA. (2008). On-Line Monitoring for Improving Performance of Nuclear Power Plants Part 1: Instrument Channel Monitoring, International Atomic Energy Agency, Vienna.
13. Hashemian, H.M. (2006). Maintenance of Process Instrumentation in Nuclear Powerplants, Chapter 6, Springer, Berlin.
14. Hashemian, H.M. (2011). Measurement of Dynamic Temperatures and Pressures in Nuclear Power Plants, the University of Western Ontario, Ontario, Canada.
15. Wikipedia article on mahalanobis distance: https://en.wikipedia.org/wiki/Mahalanobis_distance
16. Analytics India, Magazine https://www.analyticsindiamag.com/understanding-mahalanobis-distance-and-its-use-cases/, ABHISHEK SHARMA, JUL 10, 2018
17. Jiawei Han and Micheline Kamber,"Data Mining: Concepts and Techniques,"
18. Juha Juselius, Lappeenranta University of Technology , LUT School of Energy Systems,"Advanced condition monitoring methods in thermal power plants", master's thesis, 2018
19. D. Flynn, J. Ritchie and M. Cregan,"Data Mining Techniques Applied to Power Plant Performance Monitoring", The Queen's University of Belfast, N. Ireland, 16th Triennial World Congress, Prague, Czech Republic,2005
20. Li J., Niu C., Liu J., Zhang L. "Research and Application of Data Mining in Power Plant Process Control and Optimization." In: Yeung D.S., Liu ZQ., Wang XZ., Yan H. (eds) Advances in Machine Learning and Cybernetics. Lecture Notes in Computer Science, vol 3930. Springer, Berlin, Heidelberg ,2006.
21. Tony Ogilvie, B W Hogg, "Use of data mining techniques in the performance monitoring and optimization of a thermal power plant", IEEE Colloquium on Knowledge Discovery and Data Mining, 1998.

AUTHORS PROFILE

S. Narasimhan is a Fellow in Institution of Engineers (India) in Electronics & Communication Engineering and has a Master's degree in Business Administration on Operations Management from Indira Gandhi National Open University (IGNOU). He is presently working as Scientific Officer/H and Additional Chief Engineer in Bharatiya Nabhikiya Vidyut Nigam Limited (BHAVINI) under Department of Atomic Energy, India. He has the multiple responsibilities in the field of Instrumentation, IT, Communication Systems, Information Security, Electrical systems and cabling in Prototype Fast Breeder Reactor currently being commissioned in India. He has over 30 years of experience in the field of engineering, procurement, erection, commissioning and maintenance of Neutronics and Computer Based Instrumentation & Control Systems of Nuclear power Plants.

Currently he is pursuing PhD in Vels Institute of Science, Technology and Advanced Studies (VISTAS).

Velayutham Rajendran received the degree from Madurai Kamaraj University, and the M.Tech. degree from IISC Bangalore, and the Ph.D. degree from Chiba University, Japan, in 1993. He has been with different institutions like Indian Institute of Science, Bangalore, IIT Madras, and NIOT Chennai at different position. He is currently a Professor and Director with the Department of Electronics and Communication Engineering, VISTAS. He received MONBUSHO Fellowship Award of Japanese Government and Distinguished Scientist Award '07 from Jaya Engineering College. He has also been Elected Member twice as Vice Chairman - Asia of Executive Board of Data Buoy Co-operation Panel (DBCP) of Inter-Governmental

Oceanographic Commission (IOC) / World Meteorological Organization (WMO) of UNESCO, in October 2008 and September 2009. He has published more than 44journal papers and over 55 conference papers both International and national level.