2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Sensor Fault Diagnosis Using Spectral Principal Component Analysis and CNN Deep Learning

MOU Jianqiang[1], CUI Shan [1]

*[1]National Metrology Centre, Agency for Science, Technology and Research (A*STAR), Singapore*
E-mail address: mou_jianqiang@a-star.edu.sg, cui_shan@a-star.edu.sg

***Abstract* –A data driven methodology for sensor fault diagnosis in sensor network using principal component analysis (PCA) of coherence spectrum and convolutional neural network (CNN) deep learning is proposed. The methodology was evaluated with the measurement data of a sensor network for ambient relative humidity (RH) monitoring of a chemical laboratory. The results demonstrated accuracy up to 99% for sensor fault diagnosis in the sensor network functioning across a large spectrum of frequencies for environmental monitoring.**

Keywords: Sensor fault diagnosis, sensor network, coherence spectrum, principal component analysis (PCA), convolutional neural network (CNN), deep learning.

I. INTRODUCTION

Sensor fault diagnosis plays an important role for measurement data quality assurance (DQA) in environmental monitoring [1]. In recent years, data-driven algorithms for sensor fault diagnosis and identification have been extensively investigated [2].

In this the paper, a data driven methodology for sensor fault diagnosis in sensor network using principal component analysis (PCA) of coherence spectrum and convolutional neural network (CNN) deep learning traceable to the reference standard was proposed.

The proposed methodology was evaluated with the measurement data of a sensor network for ambient relative humidity (RH) monitoring of a chemical laboratory. The accuracy of the methodology for sensor fault diagnosis in the sensor network was analyzed, and the focus of future work was discussed.

II. METHODOLOGY AND PROCEDURE

The proposed methodology for sensor fault diagnosis uses PCA of coherence spectrum and CNN deep learning and has measurement traceability to the reference standard. Specifically, the methodology is applied to sensor fault diagnosis in sensor network functioning across a large spectrum of frequencies for environmental monitoring.

The methodology is developed by improving our previous work on sensor fault diagnosis in a sensor network for structure health monitoring excited by a single frequency force [3]. It consists of the steps as follows.

2.1 Sensor network configuration

Configure the sensor network as consisting of one pilot sensor x_p and other m sensors x_j $(j = 1, 2, \ldots, m)$. The pilot sensor x_p is defined for disseminating traceability in the sensor network, which is accessible for physical calibration by the reference standard.

2.2 Sensor baseline coherence calculation

At initial stage of the environmental monitoring process, collect time series measurement baseline data under the healthy condition of the sensors. Calculate the coherence spectrum between the pilot sensor and all the other sensors in the sensor network with the baseline time series measurement data. Identify the dominant frequencies ω_k $(k = 1,2,\ldots,l)$ of the coherence spectrum.

For time series measurement data $x_p(t)$ and $x_j(t)$ $(j = 1, 2, \ldots, m)$ for the pilot x_p and the sensor x_j $(j = 1,2, \ldots, m)$ respectively, the coherence spectrum $ch_j(\omega_k)$ between the $x_p(t)$ and each $x_j(t)$ is calculated as:

$$ch_j(\omega_k) = \frac{|S_{pj}(\omega_k)|^2}{S_{pp}(\omega_k)S_{jj}(\omega_k)} \quad (1)$$

Where, the $ch_j(\omega_k)$ is the coherence (magnitude-squared coherence) between the $x_p(t)$ and $x_j(t)$ at the dominant frequencies ω_k $(k = 1,2,\ldots,l)$, the $S_{pj}(\omega_k)$ is the cross-spectral density (CSD), and the $S_{pp}(\omega_k)$ and $S_{jj}(\omega_k)$ are the power spectral density (PSD) respectively [3].

For the indoor ambient environment regulated by an air handling unit (AHU) operated as a linear or piecewise linear fluid dynamic system, the coherence $ch_j(\omega_k)$ between the pilot sensor x_p and the sensor x_j $(j = 1,2,\ldots,m)$ at each dominant frequency ω_k $(k = 1,2,\ldots,l)$ will be calculated as close to 1 [3].

When sensor faults occurred in the sensor network, the linearity of the system will be jeopardized and the coherence $ch_j(\omega_k)$ will be degraded.

2.3 Coherence baseline matrix formation

The coherence baseline matrix $C \in R^{n\times(m\times l)}$ is formed by n samples of the coherences for m sensors x_j $(j =$

1,2,…,m) in the sensor network at the *l* dominant frequencies, including the measurement uncertainties.

The type A uncertainty of coherence $ch_j(\omega_k)$ *(j = 1,2,…,m, k =1,2,…,l)* is reflected by the n samples of the coherence spectrum $ch_j(\omega_k)$. The synthetic type B uncertainties added to the matrix $C \in R^{n\times(m\times l)}$ is derived by the Monte Carlo method, following a procedure published in our previous work [4] developed according to the guideline specified in the BIPM JCGM 101:2008 [5].

Normalise the coherence baseline matrix $C \in R^{n\times(m\times l)}$ as the matrix $C_x \in R^{n\times(m\times l)}$, and rescale the coherence for each sensor x_j *(j = 1,2,…,m)* to have zero mean and unity variance.

2.4 Covariance matrix formation and singular value decomposition (SVD)

The covariance matrix C_{cx} of the matrix C_x is formed as:

$$C_{cx} = \frac{1}{n-1} C_x^T C_x \qquad (2)$$

Perform the singular value decomposition to decompose the covariance matrix as [6]:

$$C_{cx} = P \Lambda P^T \qquad (3)$$

and

$$C_{cx} P = P \Lambda \qquad (4)$$

Where, $P = [p_1\, p_2 \ldots p_{(m\times l)}] \in R^{(m\times l)\times(m\times l)}$, $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_{(m\times l)})$, P^T is the transpose of the P.

The $(\lambda_1, \lambda_2, \ldots, \lambda_{(m\times l)})$ are eigenvalues of the C_{cx}. The columns of matrix P are defined as the principal components (PCs) of the baseline dataset C_x, corresponding to the eigenvalues sorted in descending order.

2.5 Principal components transformation

The relative cumulative contribution rate (*CCR*) represents the selected first *r* principal component (PC) modes contributed to the cumulative variance of the dataset C_x, which is defined as [6]:

$$CCR = \sum_{i=1}^{r} \lambda_i \,/ \sum_{i=1}^{(m\times l)} \lambda_i \qquad (5)$$

Taking the *r* value corresponding to $CCR \geq 95\%$, the first *r* principal components $P_r = (p_1, p_2, \ldots, p_r)$ is selected for principal components transformation of the C_x

Project the dataset C_x into the principal components subspace $P_r = (p_1, p_2, \ldots, p_r)$ to calculate the principal components score matrix T:

$$T = C_x P_r \qquad (6)$$

The matrix C_x is then decomposed as:

$$C_x = T P_r^T + E \qquad (7)$$

$$E = C_x (I - P_r P_r^T) \qquad (8)$$

Where, the TP_r^T is the part of the coherence dataset C_x transformed into the principal components subspace P_r, and E is the PCA residual.

2.6 Principal components projection

Collect time series measurement training data at later stage during the environmental monitoring processes, including the sensor fault introduced.

Following the operations specified at step 2.2 and step 2.3, construct the coherence training matrixes and normalize as $C_{xs} \in R^{s\times(m\times l)}$ for the sensor network. Implement principal components projection using the equation (6), to calculate principal components score matrixes $T_s \in R^{s\times r}$ of the coherence training datasets [3].

2.7 CNN classification by deep learning

Train the convolutional neural network (CNN) classification model by deep learning using the images which are converted from the principal components score matrixes $T_s \in R^{s\times r}$ of the training datasets and categorially labelled with different fault modes.

For new measurement data series, calculate the principal components score matrixes $T_s \in R^{s\times r}$ of the coherence spectrum datasets and convert them into images. Apply the trained CNN model to the images for diagnosis of the sensor fault in the sensor network.

III. MEASUREMENTS AND SPECTRAL PRINCIPAL COMPONENTS ANALYSIS

A sensor network for ambient relative humidity (RH) monitoring of a chemical laboratory is used for evaluation of the proposed methodology and procedure. The ambient condition in the chemical laboratory is regulated by an air handling unit (AHU). The sensor network consists of 7 high accuracy GREYSTONE humidity sensors, marked as *sRHi (i = 1,2…,7)*. Part of the chemical laboratory and the humidity sensor network is displayed in Figure 1. The humidity sensor *sRH1* is configured as the pilot sensor in the sensor network.

Fig. 1. Part of the chemical laboratory and the humidity sensor network.

The time domain measurement data acquisition for the humidity sensor network was carried out from 7.30 PM to 7.30 AM next day over one month under the normal operation of the AHU, with a sampling frequency 0.1Hz, as displayed in Figure 2.

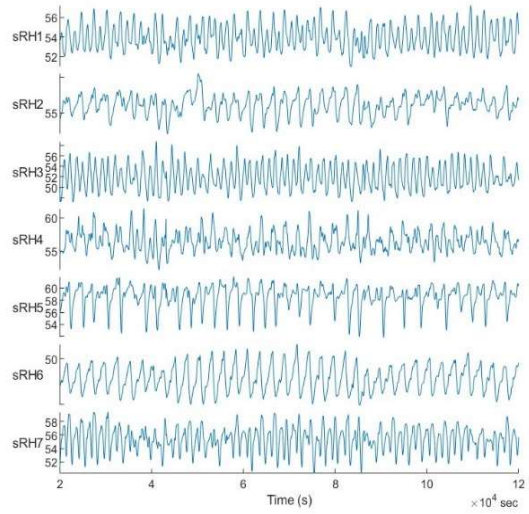


Fig. 2. Time domain measurement data of the RH sensor network.

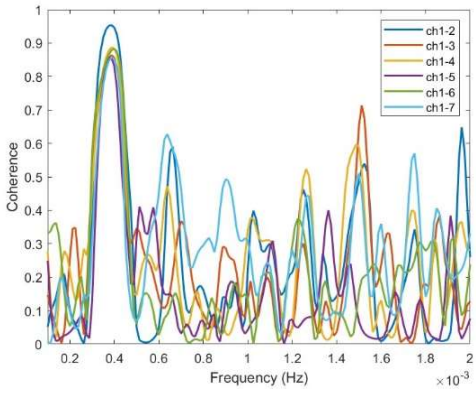


Fig. 3. Coherence spectrum w.r.t the pilot sensor of the baseline measurement data.

The coherence spectrum with respect to (w.r.t) the pilot sensor of the baseline time series measurement data was calculated, as displayed in Figure 3. The dominant frequencies of the coherence spectrum were identified as 13 frequencies in the frequency range from 0.2e-3 Hz to 0.5e-3 Hz, corresponding to the operational frequencies of the air handling unit (AHU). The coherences at the dominant frequencies were calculated and the coherence baseline dataset was constructed, at the initial stage of the testing process without sensor fault in the sensor network. Then the principal components analysis (PCA) on the normalized coherence baseline matrix $C_x \in R^{6000\times(6\times13)}$ for 6000 repeated measurements of the coherence spectrum at the dominant frequencies of the sensor network was performed. As displayed in Figure 4, the first *36* principal components were selected for the PCA model, corresponding to the *CCR ≥ 95%.*

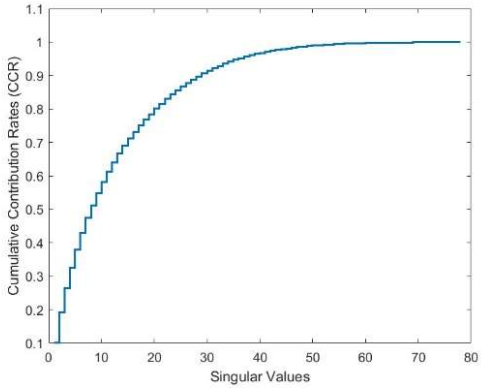
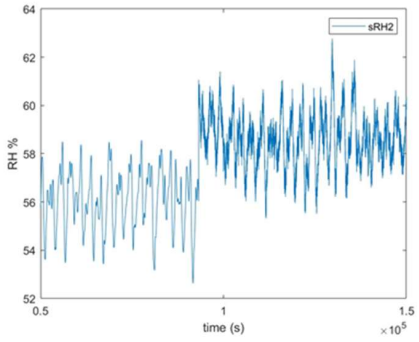
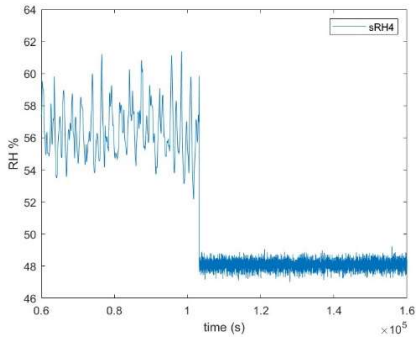


Fig. 4. The CCR contributed to the cumulative variance of the coherence baseline dataset.

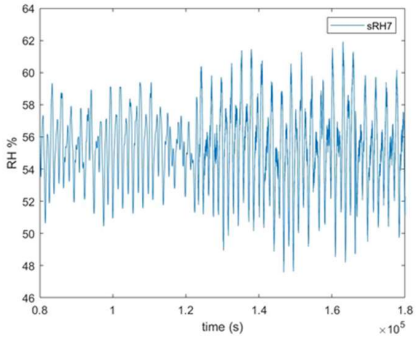
Sensor fault was introduced randomly at the later stage of measurement data acquisition in the evaluation process. Specifically, the sensor fault modes [2, 3] such as senor bias, sensor complete failure and sensor gain were introduced in the time domain measurement data acquired for the sensors sRH2, sRH4 and sRH7 respectively, as illustrated in Figure 5.



a) Sensor sRH2 bias fault.



b) Sensor sRH4 complete failure.



c) Sensor sRH7 gain fault.

Fig. 5. Fault modes of the relative humidity sensors.

Figure 6 displays the principal components scores of the coherence training datasets at the principal component PC1, PC2 and PC3 subspace. It is found that the PCA scores of coherence dataset of the healthy sensors were partially mixed with those for the faulty sensors. To diagnose the sensor fault in the sensor network using the PCA scores of the coherence spectrum at the dominant frequencies, a CNN classification model developed by deep learning is deployed.

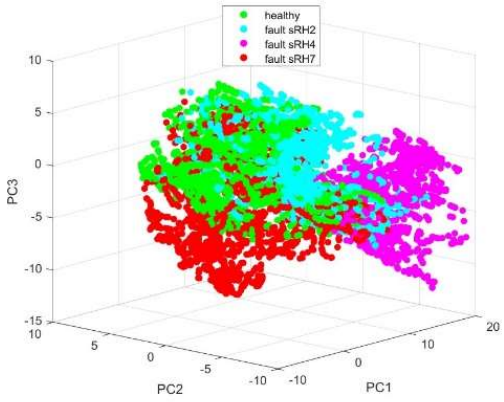


Fig. 6. The PCA scores of coherence training dataset at PC1, PC2 and PC3 subspace.

IV. CNN DEEP LEARNING

A convolutional neural network (CNN) classification model [3] is constructed, as displayed in Figure 7, for sensor fault diagnosis and classification.

The CNN model is trained by deep learning [7] using the images converted from the principal components score matrixes of the coherence spectrum dataset.

Each coherences spectrum dataset is constructed with s = 40 measurements of the coherence spectrum with respect to the pilot sensor, by a time step 120 seconds. The principal components score matrixes $T_s \in R^{40\times36}$ was then calculated by projecting the normalized coherence spectrum dataset to the principal component subspace truncated at rank r = 36.

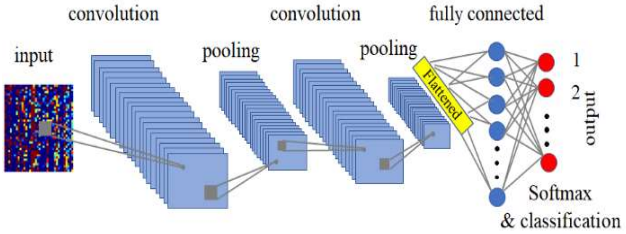
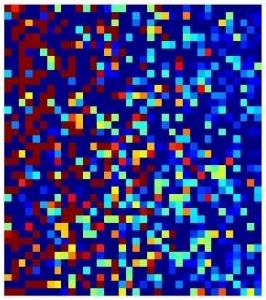
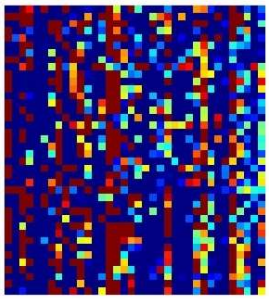


Fig.7. The architecture of the CNN model for sensor fault diagnosis and classification.

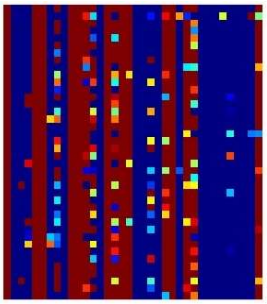
The principal component score matrix $T_s \in R^{40\times36}$ of the coherence spectrum was converted into a 40×36 pixels image and categorically labelled as input to train the CNN model by deep learning, as displayed in Figure 8.



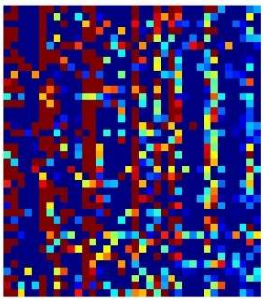
a) Healthy sensor.



b) sRH2 bias fault.



c) sRH4 complete failure.



d) sRH7 gain fault.

Fig. 8. The images converted from the principal components score matrixes of coherence spectrum.

The convolutional neural network (CNN) model was trained by deep learning using the images converted from the principal components score matrixes, which were divided into the training, validating and testing dataset proportionally by 50 %, 10 % and 40 % respectively.

Figure 9 displays the progress of the classification accuracy of the CNN model in the deep learning process using the images of the training dataset and validating dataset. The CNN model was finalized at the classification accuracy of 99.06% in the deep learning process.

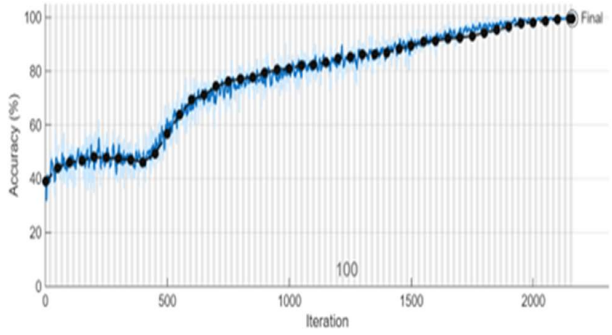


Fig. 9. The progress of the classification accuracy in CNN deep learning

The accuracy of the CNN model was further examined using the images of the testing dataset, by comparing the CNN predicated results with the true class of the sensor status in the sensor network. Table 1 displays the confusion chart of the CNN model trained by deep learning for sensor fault classification examined using the images of the testing dataset, where the classes from 1 to 4 represent the healthy sensor, sRH2 bias fault, sRH4 complete failure and sRH7 gain fault in the sensor network respectively. The results demonstrated that the accuracy up to 99% was achieved for sensor fault diagnosis in the sensor network. However, 2.2% of the healthy sensor (class 1) was misclassified as faulty sensor and confused with the bias fault (class 2) and gain fault (class 4), because of the measurement uncertainties of the coherence spectrum and fluctuation of the AHU operational process. The issue will be further addressed in our future work.

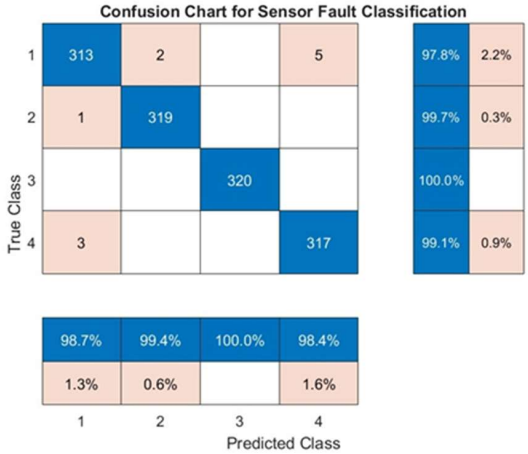


Table 1. Sensor fault classification confusion chart using the CNN model trained by deep learning.

V. CONCLUSIONS

A methodology for sensor fault diagnosis in sensor network using principal component analysis (PCA) of coherence spectrum and convolutional neural network (CNN) deep learning is proposed. The methodology was evaluated with the measurement data of a sensor network for ambient relative humidity (RH) monitoring of a chemical laboratory. The results demonstrated accuracy up to 99% for sensor fault diagnosis in the sensor network. The methodology is applicable for data quality assurance (DQA) in sensor network functioning across a large spectrum of frequencies for environmental monitoring beyond relative humidity, such as for hazardous gas monitoring. Future work will focus on enhancing the accuracy and robustness of the methodology.

ACKNOWLEDGMENTS

This research is supported by A*STAR under its RIE 2025 Urban Solutions and Sustainability domain's Low Carbon Energy Research Phase 2 Programme, Hydrogen and Emerging Technologies Funding Initiative, Directed Hydrogen Programme (Award: U2303D4001). Disclaimer: Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of the A*STAR.

REFERENCES

[1] Y. Liu, N. Tang, "Humidity sensor failure: a problem that should not be neglected", Atmos. Meas. Tech., 7, 3909–3916, 2014.

[2] S. Chauhan, G. Vashishtha and R. Zimroz,

"Analysing Recent Breakthroughs in Fault Diagnosis through Sensor: A Comprehensive Overview", CMES, Vol.141, No.3, pp. 1983-2020, 2024.
[3] J. Mou, S. Cui, C. Cheng, "Sensor fault diagnosis using principal component analysis and convolutional neural network for offshore structural health monitoring", Measurement: Sensors, 28 December 2024, 101465.
[4] R. Chen, S. Cui, D. W. Y. Khoo, B. C. Khoo, "Measurement uncertainty analysis of leak localization in a gas pipeline", Measurement: Sensors, Volume 18, December 2021, 100069.
[5] JCGM101:2008, Evaluation of measurement data — Supplement 1 to the "Guide to the expression of uncertainty in measurement — Propagation of distributions using a Monte Carlo method".
[6] Steven L. Brunton, J. Nathan Kutz, "Data Driven Science & Engineering: Machine Learning, Dynamical Systems, and Control", Cambridge University Press, 2019.
[7] I. Goodfellow, Y. Bengio, and A. Courville (2016), "Deep learning", The MIT press, 2016.