Transactions of the Korean Nuclear Society Autumn Meeting
Gyeongju Korea, Oct. 26-27, 2017

Development of Cross Moving Median Filter for Denoising and Data Reconstruction in Nuclear Reactor

Moath Kassim, Gyunyoung Heo*
Kyung Hee Univ., Yongin-si, Gyeonggi-do, 17104, Republic of Korea
**Corresponding author:* gheo@khu.ac.kr

1. Introduction

For the sake of developing any reliable data processing model, signals should be presented in a clear waveform by preprocessing cleanup algorithm that can reduce noise and outliers effect, and reconstruct any missing data that would present in the entire sample range. Bad and missing data should be removed and replaced by the most probable data value [1].

In nuclear reactors, monitoring system can generate huge data that can explain the operability and safety of the reactor, this data always accompanied with random noise that can create a signal perturbation and, as a consequence, it would show unusual condition of the reactor parameters behavior.

Random noise in sensors consist of; process noise which is a natural parameter's fluctuation around its true value, measurement noise which is the inherent noise in the sensor, and electronic noise which is developed during data transfer by cables from the sensor to the central processor. If noise level is high in a sensor reading, the developed model performance would be low reliable, and if this noise is unavoidable, a denoising technique is important to be implemented before processing with the proposed averaging method [2]. Many empirical methods are available for freeing signals from noise and outlier such as; Discrete Wavelet transform (DTW) which includes both frequency and time domain analysis using consecutive filters and decimators, Adaptive filtering which minimize the error using iterative computation to model the relationship between two signals, methods based on least-squares polynomial approximation like the Savity-Golay filtering, and other statistical data driven models like the Moving average and Moving Median filters.

Data-driven techniques can widely simulate system behavior while being implemented quickly and cheaply. These models have ability to decrease the noise dimensions and transform it into lower dimensional information to be suitable for decision making [3].

In this paper, a new filter called Cross Moving Median (CMM) is proposed as to tackle three main problems attached with sensors data; noise, outliers, and missing data. This filter is based on the running median that was improved to account for prior estimate as to update the moving median result.

The CMM filter was validated using HANARO Cold Neutron Source data that was generated from a hydrogen transmitter.

2. Methodology

2.1 Moving Median (MM) model:

Moving Average (MA) filter is well-known denoising technique where the average of certain moving window can replace the bad measurement value. Mean (or average) is susceptible to the influence of outliers [4]. The median as a measure of central tendency can be defined as the center element (if the array has odd elements number) or the average of the two elements in the center (if the array has even elements number) of a given array after ascending its elements [5], so for dealing with noise and outliers together at once, Median or as so called Median filter is considered in this study.

The Moving Median (MM) filter chooses the median over a sliding window instead of the mean as to suppress the noise and outliers, and at the same time preserve the edges that would be a characteristic behavior of the signal. Figure 1 shows the difference of applying MA & MM denoising techniques on a hydrogen pressure transmitter sensor's signal.

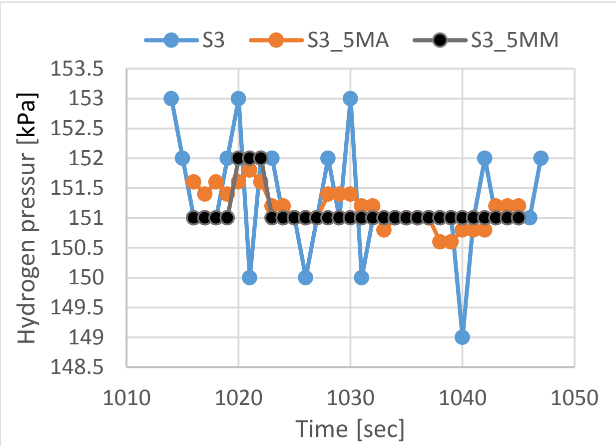

Figure 1. Difference between MA and MM filters applied on one hydrogen transmitter data (S3).

As proposed in this study, MM will be used for estimating the initial value of each sensor's signal and for involving missing data recovery process, MM will be updated for each movement for estimating the rest of sensors' data.

For an observation array S (m, n), The initial estimated value of an estimated array $\hat{S}$ (m, n-w-1) is given in equation 1:

$\hat{S}_k(1) = meadian[S_k(1) \cdots S_k(w)]$ (1)

Where; $k = 1,2,3 \ldots, m$
m is the redundant signal number,
n is the range of observation sample,
and w is the proposed sliding window.

2.2 Cross Moving Median (CMM) filter:

Missing data is a real problem that can really affect any prediction model's performance, especially when they corrupt long range of the processed signal, in this case, removing them from database even for one signal is not a reliable method [6]. Therefore, missing data can be either reconciled based on energy balance calculation, replaced by statistical values, or reconstructed using some available methods such as; Auto Associative Artificial Neural Networks (ANN), Principle Component Analysis (PCA), and Auto-associative Kernel Regression (AAKR).

For achieving a good and fast estimate for each sensor readings, a missing data recovery process is performed in this study based on the statistical replacing approach proposing new idea to update the posterior estimated value using the immediate past estimate; the immediate past estimate will be involved along with the sliding MM window for estimating the median. This Cross Moving Median (CMM) of a predetermined sliding window can efficiently and cheaply recover all missing data even if they present in a wide range.

Figure 2 shows how the cross median window is moving on both the actual observation vector and the estimated vector.

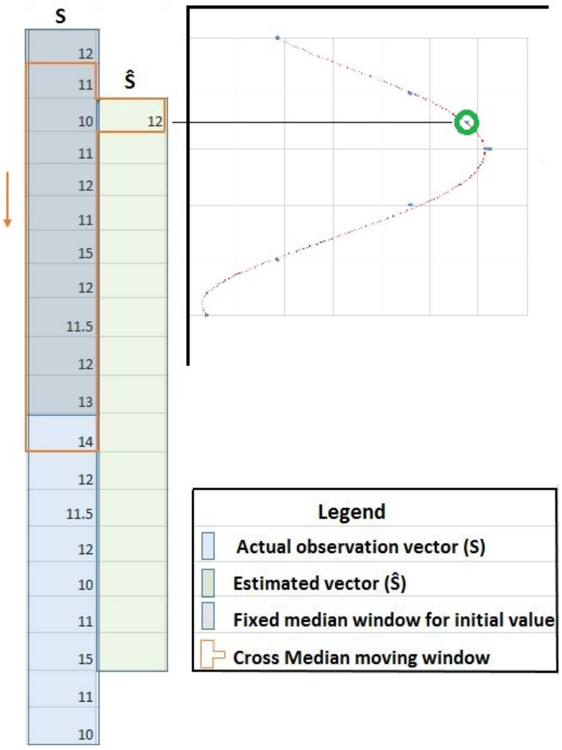

Figure 2. Cross Moving Median window

To implement CMM for calculating the rest of signal estimated values, the immediate past estimate will be added to the proposed sliding window as shown in equation 2:

$\hat{S}_k(j) = median[S_k(j) \cdots S_k(j + w - 1),\ \hat{S}_k(j - 1)]$
(2)

Where; $j = 2, \ldots \ldots \ldots, (n - w) + 1$

2.3 SNR & Correlation Analysis:

The reduction of noise and outliers' effects is not only important for smoothing the signals but also to minimize the uncertainty estimate that would exceed the drift allowance even when no drift is present.

As this study mentioned nuclear reactors where data is highly steady states and the independent random noise is a dominant factor, the signal to noise ratio should be calculated for choosing the optimal sliding window which its results show the highest correlation between redundant sensors' estimated signals. Low correlation should be avoided in order to keep good performance of the model and as a consequence a reliable averaging estimate can be achieved.

Signal to Noise Ratio (SNR) can be calculated [7] for the raw signals and the filtered signals respectively as in the equation bellow:

$SNR_{S_k} = \frac{\mu_{S_k}}{\sigma_{S_k}}$ and $SNR_{\hat{S}_k} = \frac{\mu_{\hat{S}_k}}{\sigma_{\hat{S}_k}}$ (3)

Where; μ_{S_k} & $\mu_{\hat{S}_k}$ are the means of actual signal (observation) and estimated signal respectively,
σ_{S_k} & $\sigma_{\hat{S}_k}$ are the standard deviation of actual signal (observation) and estimated signal respectively.

3. Results & Discussion

After implementing a sliding windows of different sizes (w=3, w=5, w=7, w=9, w=11, w=13, w=15, w=17 w=19, w=21, w=23, w=25, w=27, w=29, and w=31) for each hydrogen pressure transmitter signal and based on the correlation analysis and the calculation of Signal to Noise Ratio (SNR) performed, the window of 31 periods was found to be the optimal window. Figure 3 shows the difference between some representative CMM windows implemented on test data set.

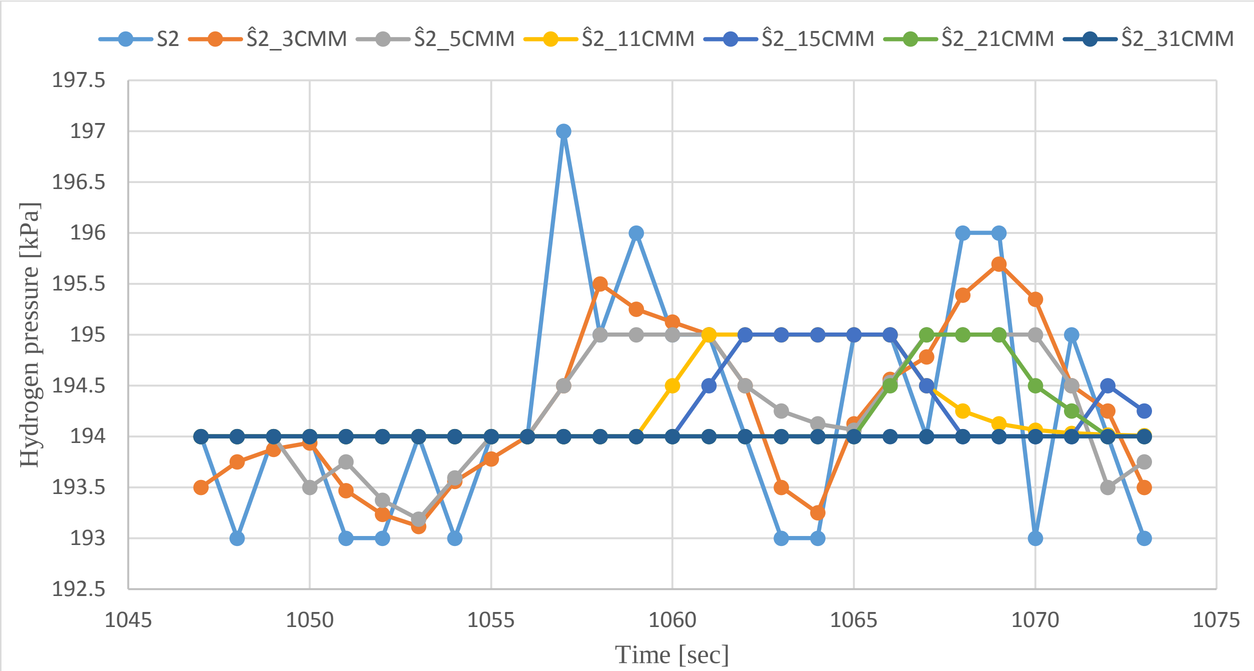

Figure 3. Signals' behaviors of 3, 5, 11, 15, 21 and 31 periods CMM with respect to the actual observation vector.

Due to the long range of missing data in the training and test datasets, it was not easy to calculate the correlation and SNRs of original signals, thus, it was determined to take the part of the original signal excluding missing data and compare it with the same range of its estimated vector in order to correctly identify the maximum SNR. Since we are dealing with steady state data, the bigger size the sliding window will be, the higher signal to noise ratio that can be obtained. Thus, CMM of 31period window was chosen as the optimal minimum window for our filtering model. Table 1 shows the enhancement of SNR value using two different sizes of sliding windows.

Table 1. SNR results comparison for actual observation, 5 periods CMM window, and 31 periods CMM window

SNR	Actual observation	5CMM	31CMM
S#1	443.5671	436.8226	639.9879
S#2	148.8465	162.2184	168.4977
S#3	406.1051	422.0335	401.7409
S#4	411.6105	414.9048	639.8151

The correlation analysis was also performed on same data set for actual observation, the 5 period estimation window, and the 31 period estimation window. Results, as in tables 2, 3, and 4 showed that the highest correlation was obtained at 31 period window.

Table 2. Correlation results for actual observation

	S#1	S#2	S#3	S#4
S#1	1			
S#2	0.8243	1		
S#3	0.8921	0.8676	1	
S#4	0.8773	0.8289	0.8112	1

Table 3. Correlation results obtained at 5 CMM window

5 CMM	S#1	S#2	S#3	S#4
S#1	1			
S#2	0.7555	1		
S#3	0.8920	0.8154	1	
S#4	0.9026	0.7814	0.8284	1

Table 4. Correlation results obtained at 31 CMM window

31 CMM	S#1	S#2	S#3	S#4
S#1	1			
S#2	0.8792	1		
S#3	0.8718	0.8780	1	
S#4	0.9164	0.8406	0.8418	1

Another consideration for choosing the 31 period window as a suitable sliding window is that the minimum length for an observation sample to be treated as normally distributed shouldn't be less than 30 values.

As it can be inferred form table 4, the CMM filter recovered all missing data successfully, which is the most important feature that makes CMM better than normal MM, along with denoising each signal and decreasing the outlier effects. The only problem that would affect the data recovery is the present of Zero value as a very extreme value that can influence the model to make bad estimation of the missing value. This case would really affect the estimated average especially when the missing values present in a continues range contain at least one Zero value in a window of 31 periods, in this case the CMM filter will take the average of Zero value and the immediate past estimate, and that will be defiantly wrong recovery of the missing data. To resolve this issue, a conditional software command was made to replace any Zero value (or very extreme value) with a missing space before the CMM machine works.

It should be noticed here that the length of estimated vectors will be less than the original observation of raw vectors by 30 points as they were replaced by the estimated median of the sliding window.

Conclusions

As many models, which are developed for nuclear reactor diagnostics and prognostics, are based on data generated from nuclear reactors computer, the need for high quality data is very important for developing reliable diagnostics and prognostics methods. In this paper, we proposed CMM filter that can deal with noise, outliers, and missing data providing reliable estimate of sensors' readings. The so called Cross Moving Median (CMM) with an optimal sliding window has the capability to attenuate noise, prune outliers, and reconstruct missing data.

The HANARO Korean research reactor cold neutron source (CNS) data sets of hydrogen pressure redundant transmitters were studied for the purpose of validation, and results showed that the CMM window was chosen to be of 31 periods was the minimum optimal window to be considered which gave the highest correlation and highest Signal to Noise Raito (SNR), especially for the use in on-line applications of steady state nuclear reactors data.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIP) (Grant Number: NRF-2016M2B2A9A02945090)

REFERENCES

[1] Eiler, J., Glöckler, O., and International Atomic Energy Agency. (2008). *On-line monitoring for improving performance of nuclear power plants. Part 1, Instrument channel monitoring.*
[2] Hines, J. W., and Seibert, R. (2006). Technical Review of On-Line Monitoring Techniques for Performance Assessment. Volume 1. State-of-the-Art. *Nuclear Engineering and Technology*, *1*.
[3] Ã, E. Z., and Maio, F. Di. (2010). A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system. *Reliability Engineering and System Safety*, *95*(1), 49–57.
[4] Notation, S. (2003). Mean, Median and Mode, 2–3.
[5] Tukey, J. W. (John W. (1977). *Exploratory data analysis*. Addison-Wesley Pub. Co.
[6] Baraldi, P., Di Maio, F., Genini, D., and Zio, E. (2015). Reconstruction of missing data in multidimensional time series by fuzzy similarity. *Applied Soft Computing Journal*, *26*, 1–9.
[7] Holmes, D. S., and Mergen, A. E. (2007). SIGNAL TO NOISE RATIO - WHAT IS THE RIGHT SIZE? 1–6.