Application of SMA and EMA algorithms for smoothing of data corrupted by random noise and comparison of their performances

Padma Lochan Bora and Ashish Kumar Nanda
Defence Research and Development Laboratory
Kanchanbagh Post, Hyderabad, India
e-mail: padmabora@gmail.com, aknanda80@gmail.com

Abstract—**During the development phases of any propulsion system, it is essential to perform different tests/experiments. During such tests, measurements of different physical parameters like pressure, temperature, thrust, displacement etc. are vital for evaluating the performance of the system. These parameters generally start from zero level. But in many cases, they may start from a nonzero constant value. When such a parameter is measured, the measured data should represent the true value. But when there is random noise present in the measurement channel, the measured data will not represent the true value; rather there will be fluctuations. So, it is inevitable to eliminate/reduce the effect of noise and deduce the true data.**

When the physical phenomenon/test begins, the same random noise may be predominantly present throughout the measured data. The data processing method applied in such situations should faithfully eliminate/reduce the random noise and provide the data that represent the true value. Simple Moving Average (SMA) algorithm is generally applied to reduce the noise and obtain the true value. Similarly, Exponential Moving Average (EMA) algorithms – Single EMA (SEMA) and Double EMA (DEMA) can also be applied to process the noisy data and deduce the near true values.

In this paper, we have presented the results of a comparative study carried out to reduce the effect of random noise using SMA, SEMA and DEMA and their performances with respect to ideal rectangular pulse data.

Keywords- SMA, SEMA, DEMA, MSE, time series.

INTRODUCTION

We carry out the measurement of different physical parameters like pressure, temperature, displacement, thrust etc. to get experimental data pertaining to certain systems. Based on the data, certain decisions are taken regarding design validation, process consolidation etc. We also evaluate performance of systems and subsystems based on the measurement of certain parameters. These measurements should conform to certain level of accuracy as mandated by the system to be validated.

In many situations it is observed that, the measured data are corrupted by random noise and do not conform to the required accuracy specifications. However, when the mean is taken or the effect of random noise is reduced by applying some methodology, the resultant data conform to the required accuracy specifications. It is essential to find out appropriate algorithms to process the noisy data, reduce the effect of noise and derive the best possible true value without introducing distortion, delay and deviation. The resultant data should retain the sharp rise/fall, if any, of the original signal.

Simple Moving Average (SMA): In general, Simple Moving Average (SMA) algorithm is used to process the raw data to get the smoothed data. This algorithm is simple to implement and does the job faithfully in case of slow varying signals/data. But it has the associated disadvantages of introducing delay, deviating from the mean position and reducing the peaks considerably in case of signals having abrupt rise/fall.

In this algorithm, the smoothing is performed by taking the average of the required number of available past [1] (or both past and future) raw samples. In our case of performance evaluation of propulsion systems, we use only the available past samples to deduce the processed/smoothed present sample.

Exponential Moving Average (EMA): Exponential Moving Average is a class of recursive algorithm used to reduce the effect of random noise. There are two commonly used algorithms in this class - Single Exponential Moving Average (SEMA) and Double Exponential Moving Average (DEMA).

In EMA, the value of a time series data at some point of time is determined by past values of the same series. This method is similar to other types of moving average algorithms with the exception that here exponentially decreasing weights are assigned to the observations as they get older and hence all past values contribute in some way or the other in forecasting/estimating the present value. In other words, recent observations are given relatively more weights than the older observations in forecasting/estimating the present value/sample [2]. It is a procedure for continually revising a forecast/estimate in the light of more recent experience.

Single EMA (SEMA): It is also known as Simple Exponential Smoothing. It is used for *short-range* forecasting/estimation, just one time interval into the future. This is also used to estimate the present value/sample based on the past estimated value and the present raw value. In case of the first sample, no past history/estimated value is available and hence it is taken as it is. This algorithm is best suited for data that fluctuate around a reasonably stable mean (no trend or consistent pattern of growth) [3].

Double EMA (DEMA): It is also known as Double Exponential Smoothing. It is used when the data pattern shows a trend. This method of smoothing of raw data with a trend works much like simple smoothing except that two components - level and trend, must be updated in each period. The level is a smoothed estimate of the value of the raw data at the end of each period and the trend is a smoothed estimate of average growth of the raw data at the end of each period [3].

DESCRIPTION OF ALGORITHMS, THEIR COMPARATIVE STUDY AND ERROR ANALYSIS

SMA is generally applied to get the smoothed output of a set of varying data samples. SEMA and DEMA are two popular algorithms of time series forecasting techniques. In our case, all these algorithms have been used for smoothing the noisy raw data. The algorithms for these three methods are mentioned below:

(a) Algorithm for SMA:

$$F(k) = \frac{1}{k}\sum_{i=0}^{k-1} R(k-i)\,, \quad for\ k < N$$

$$F(k) = \frac{1}{N}\sum_{i=0}^{N-1} R(k-i)\,, \quad for\ k \geq N$$

where 'N' is the no. of samples used for moving average

(b) Algorithm for SEMA [4]:

$$F(k) = R(k), \qquad for\ k = 1$$

$$F(k) = \alpha R(k) + (1-\alpha)F(k-1), \qquad for\ k > 1$$

(c) Algorithm for DEMA [2, 5]:

$$F1(k) = F2(k) = R(k), \qquad for\ k = 1$$

$$\left.\begin{aligned} F1(k) &= \alpha R(k) + (1-\alpha)F1(k-1) \\ F2(k) &= \alpha F1(k) + (1-\alpha)F2(k-1) \end{aligned}\right\} for\ k > 1$$

$$F(k) = \left(2 + \frac{\alpha}{1-\alpha}\right)F1(k) - \left(1 + \frac{\alpha}{1-\alpha}\right)F2(k)$$

In (b) and (c) above, R(k) is the raw signal/data, F(k) is filtered signal/data, 'α = 2/(N+1)' is the smoothing factor,
0 < α < 1 and 'N' is the no. of sampling intervals.

The analysis of processed data using SMA, SEMA and DEMA with different values of 'α' (smoothing factor) has been carried out. For the analysis, a rectangular pulse of width 4 seconds and an amplitude (maximum value) of 3.0 units superimposed with a Gaussian random noise of zero mean and variance of '0.05' was simulated & sampled at a rate of 200 samples per second. This sampled signal is shown in Fig.1 below:

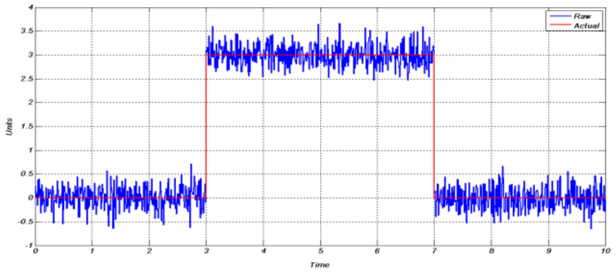

Fig.1. Raw data to be processed by using different algorithms

The above raw data have been processed by applying SMA with different no. of samples (N) and SEMA & DEMA with different 'α' values. The processed data have been analyzed to determine the suitable values of 'N' and 'α'.

In order to compare the performances of the algorithms, the Mean Square Errors (MSEs) of the filtered/processed data with respect to the ideal data (without noise) for different values of 'α' have been computed and compared. The algorithm, which gives the minimum MSE, is the preferred one as far as the overall error is concerned. Table-1 below shows the comparison of MSEs produced by application of SMA, SEMA and DEMA.

(d) Mean Square Error: Based on this measure, the appropriate value of 'α' is decided. The MSE [2] is computed as given below:

$$MSE = \frac{1}{n}\sum_{k=1}^{n}(F(k) - A(k))^2$$

where A(k) is the actual/ideal data & F(k) is the filtered data.

The mean square errors of filtered data with respect to the ideal data obtained by applying SMA, SEMA and DEMA for different 'α' values have been presented in Fig. 2 below. For SMA, the no. of samples has been mapped to corresponding 'α' values in SEMA and DEMA. These values have been used for plotting.

Apart from minimum MSEs, the following preferred characteristics of the filtered/smoothed data have been considered for the purpose of evaluation of the performance of each algorithm:

- Minimum delay in reaching the level of ideal data.
- Minimum overshoot or undershoot at the transition/ sharp rise/fall points.
- The profile of the filtered data should be as close as possible to the mean profile of the noisy data.

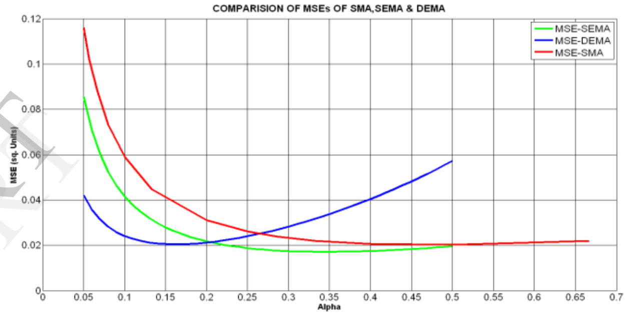

Fig.2. Comparison of MSEs for different 'α' values

Table-1 Salient points of the comparison of MSE values

Sl. No.	Type of Algorithm	MSE	% MSE of max. unit value of ideal signal
1	SEMA with α = 0.33	0.017113	0.5704%
2	SEMA with α = 0.34	0.017079	0.5693%
3	SEMA with α = 0.35	0.017075	0.5691%
4	SEMA with α = 0.36	0.017098	0.5699%
5	SEMA with α = 0.37	0.017146	0.5715%
6	DEMA with α = 0.14	0.020796	0.6932%
7	DEMA with α = 0.15	0.020532	0.6844%
8	DEMA with α = 0.16	0.020422	0.6807%
9	DEMA with α = 0.17	0.020445	0.6815%
10	DEMA with α = 0.18	0.020584	0.6861%
11	SMA with 2 samples (α=1.00)	0.027821	0.9273%
12	SMA with 3 samples (α=0.67)	0.021842	0.7280%
13	SMA with 4 samples (α=0.50)	0.020316	0.6772%
14	SMA with 5 samples (α=0.40)	0.020576	0.6858%
15	SMA with 6 samples (α=0.33)	0.021869	0.7289%

As seen from Table-1 above, minimum MSE for SMA has been achieved against 04 no. of samples (corresponding to α = 0.50). Similarly, for SEMA and DEMA, minimum MSEs have been obtained at ‘α = 0.35’ and ‘α = 0.16’ respectively.

The pictorial depiction of the three minimum MSEs in each case has been presented as bar chart in Fig. 3 below:

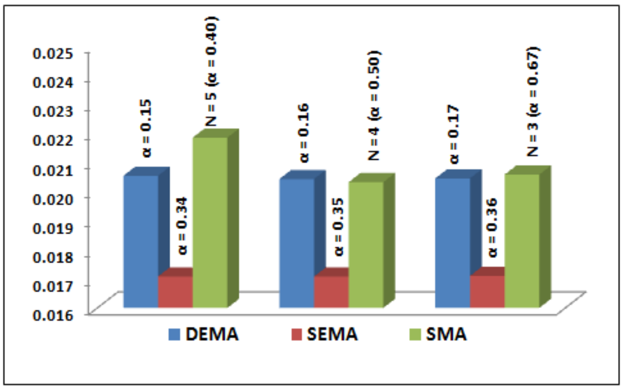

Fig.3. Comparison of MSEs for SMA, SEMA & DEMA

The plot for square of errors obtained against the three minimum ‘α’ values is given in Fig. 4 below:

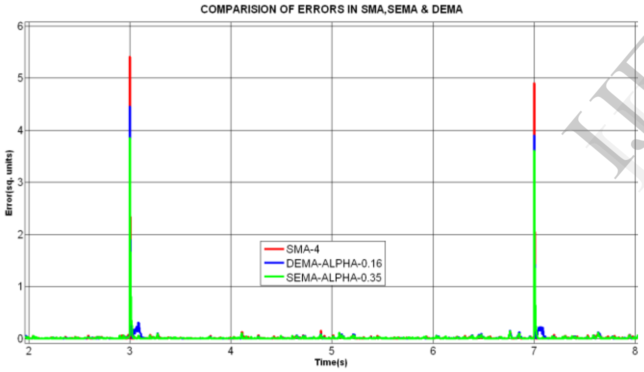

Fig.4. Comparison of Errors for SMA, SEMA & DEMA

RESULTS OF PROCESSING SIMULATED DATA

The results obtained by applying different algorithms for smoothing the simulated raw data have been presented in the following figures (Fig.5 to Fig.10):

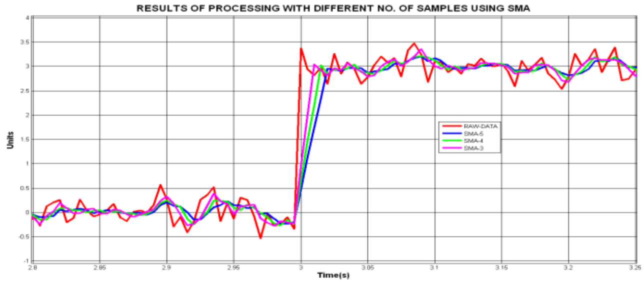

Fig.5. Comparison of filtered & raw data using SMA

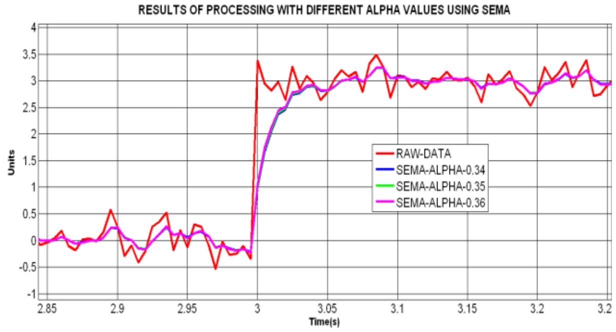

Fig.6. Comparison of filtered & raw data using SEMA

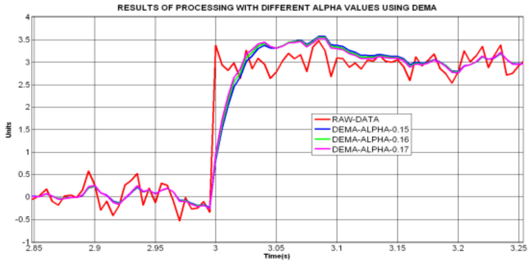

Fig.7. Comparison of filtered & raw data using

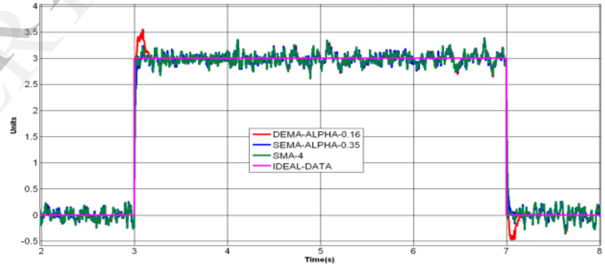

Fig.8. Comparison of filtered data using SMA, SEMA & DEMA with the ideal data

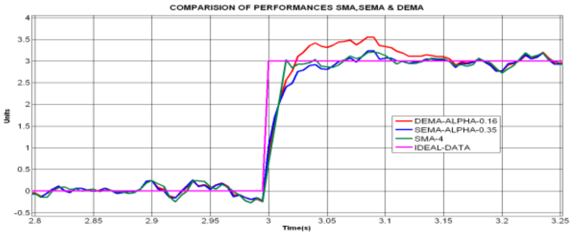

Fig.9. Zoomed view of Fig. 8 at the rising portion

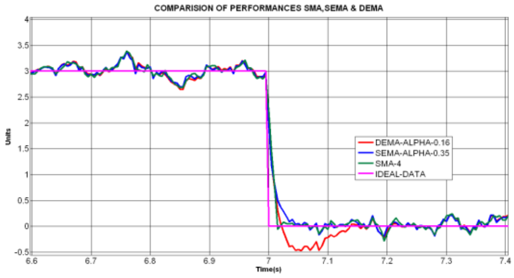

Fig.10. Zoomed view of Fig. 8 at the falling portion

RESULTS OF PROCESSING TEST DATA

From section III above, it has been observed that minimum MSEs were obtained for SMA, SEMA and DEMA at N = 4 (α = 0.50), α = 0.35 and α = 0.16 respectively. These values were used to process noisy test data. This test data was obtained during testing of a propulsion system.

The processed data obtained by application of these three algorithms along with noisy test data are presented in the following figures (Fig.11 & Fig.12):

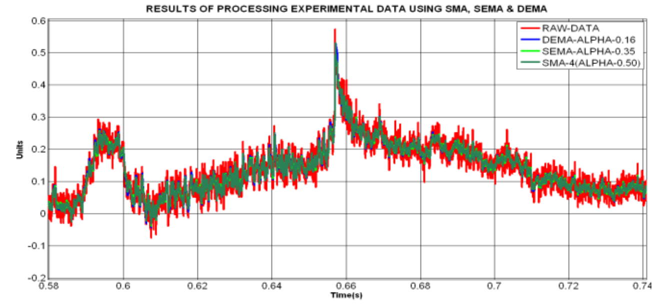

Fig.11. Comparison of noisy & filtered test data (using SMA, SEMA & DEMA)

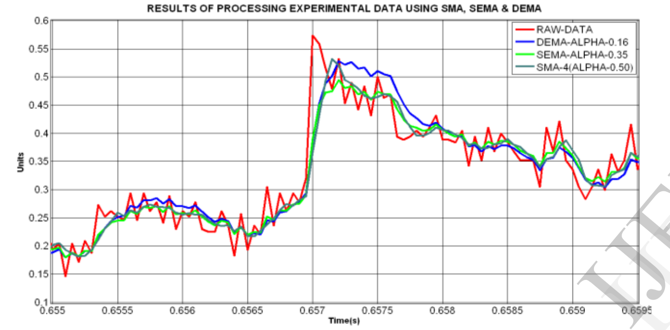

Fig.12. Zoomed view of the peak portion of Fig. 11

CONCLUSION

Both simulated and experimental data have been processed by applying SMA, SEMA and DEMA using the 'α' values corresponding to minimum MSEs. The results obtained have been compared. It has been observed that DEMA has introduced large overshoot or undershoot when there is a sharp transition in signal level. However, compared to DEMA, the overshoot or undershoot introduced by SMA and SEMA are reasonably low. As far as delay in reaching the high/low level (during transition) is concerned, SMA performs the best, i.e., with minimum delay. SEMA introduces the worst delay and DEMA performs in between. As far as mean square errors are concerned, SEMA with 'α = 0.35' results in minimum value.

In practical situations, both SMA and SEMA can be used to reduce the effect of random noise. DEMA may not be useful due to the overshoot/undershoot introduced by it. However, all these three algorithms may not be applicable at all if no delay is acceptable. This is because all these algorithms introduce considerable delays at the transition points when the raw data contains abrupt rise/fall.

REFERENCES

[32] Wenying Guo, Deren Chen, Timothy K. Shih, "Automatic Forecasting Agent for e-Commerce Applications," Proceedings of the 20th International Conference on Advanced Information Networking and Applications (AINA '06).

[33] Man Ho Kim, Suk Lee and Kyung Chang Lee, "Predictive Hybrid Redundancy using Exponential Smoothing Method for Safety Critical Systems," International Journal of Control, Automation, and Systems, vol. 6, no. 1, pp. 126-134, february 2008.

[34] Prajakta S. Kalekar, "Time Series Forecasting usign Holt-Winters Exponteial Smoothing," Kanwal Rekhi School of Information Technology.

[35] Everette S. Gardner, Jr., "Exponential smoothing : The state of the art – Part II," Bauer College of Business, University of Housten, Housten, Texas.

[36] Joseph J. LaViola Jr., "Double Exponential Smoothing: An Alternative to Kalman Filter-Based Predictive Tracking," Brown University Technology Center for Advanced Scientific Computing and Visualization.

[37] Nien Fan Zhang, "The batched moving averages of measurement data and their applications in data treatment," Measurement 39 (2006) 864-875, ELSEVIER.

[38] R. D. Snyder, A. B. Koehler, R. J. Hyndman, and J. K. Ord, "Exponential smoothing models: Means and variances for lead-time demand," European Journal of Operational Research, vol.158, no. 2, pp. 444-455, October 2004.