*Journal of Industrial Engineering International* *Islamic Azad University, South Tehran Branch*
*January 2010, Vol. 6, No. 10, 16-30*

# Identifying the time of a step change with MEWMA control charts by Artificial Neural Network

**F. Ahmadzadeh***

*Assistant Professor, Dep. of Industrial Engineering, Islamic Azad University, Karaj Branch, Karaj, Iran*

**R. Noorossana**

*Professor, Dep. of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran*

**A. Saghai**

*Assistant Professor, Dep. of Industrial Engineering, Islamic Azad University, Science and Research Branch, Tehran, Iran*

**Abstract**

Quality control charts have proven to be very effective in detecting out of control signals. It is very important to practitioners to determine at what point in the past the signal was initiated. If a control chart signals a change in the process parameter, identifying the time of the change will substantially help the signal diagnostics procedure since it simplifies the search for special causes. In this paper the researchers have proposed the observations following multivariate normal distribution. They have used Multivariate Exponentially Weighted Moving Average (MEWMA) control chart to detect signals. This research provides two ways to detect the change point, first MLE, and then neural network is used to identify the time of the change in the parameters (mean) in the past. The researchers intended to assess the performance of two approaches and compare them through computer simulation experiments. The results show that neural network performs effectively and equally well for the whole process dimensions while shift magnitudes are considered. Thus, the neural network provides process engineers with an accurate and useful estimate of the actual time of the change in the process mean.

**Keywords:** Statistical process control; Multivariate Exponentially Weighted Moving Average (MEWMA); Change point estimation; Monte Carlo Simulation; Neural Network; Maximum Likelihood Estimator ( MLE)

## 1. Introduction

Statistical Process Control (SPC) charts are the tools to monitor the state of a process by distinguishing between common causes and special causes of variability. When a control chart signals that a special cause is present, process engineers initiate a search for the special cause. Process engineers' expertise and knowledge of the process, help them identify which combination of many process variables caused the signal. This identification enables the engineers to improve product quality by preventing or avoiding changes in those variables which cause poor quality, and optimizing those variables which can lead to a higher quality. Knowing when a process has changed would simplify the search for the special cause by shrinking the time window within which to look for the special cause. Consequently, the special cause can be identified more efficiently, and corrective measures can be implemented sooner. When several characteristics of a manufactured component are to be monitored simultaneously MEWMA control charts may be used (Montgomery).The MEWMA chart gives an out of control signal as soon as $T_i^2 = \mathbf{z}_i' \boldsymbol{\Sigma}_{\mathbf{z}_i}^{-1} \mathbf{z}_i > L$ (where L (>O) is chosen to achieve a specified in control ARL and $\Sigma_{zi}$ is the covariance matrix of $Z_i$ ) and in this case the process is deemed out of control due to one or more special cause(s) and the investigation is carried out to detect the special causes other wise the process is assumed to operate under a stable system of common causes. When one or

* Corresponding author. E-mail: syahrood@gmail.com

more point(s) exceed(s) the L, it is well known that MEWMA control charts can signal a change in the multivariate process mean a substantial amount of time after the change has actually occurred. This is especially true for small shifts in the process mean since the average run length of the control chart can be large. Hence, examining the process for special causes only at the time of the signal may be ineffective knowing when the change had actually occurred would substantially help the signal diagnostic procedure.

In this paper the researchers have first identified the change point using MLE, and then described how neural network may be used to identify the change point when MEWMA signals a change in the process mean. Then they have assessed the performance of two approaches through computer simulation.

Next section describes MEWMA procedure and MLE approach to detect the change point, Section 3 illustrates artificial neural network architecture to detect the change point, section 4 studies the performance of two approaches using Monte Carlo simulation and comparison of them finally provides the conclusions.

## 2. The MEWMA procedure

Suppose that we observe $X_1, X_2 \ldots$ in the univariate case i.e. when p = 1. The univariate EWMA chart is based on these values:

$$Z_i = rX_i + (1 - r)Z_{i-1} \qquad i = 1, 2, \ldots, \tag{1}$$

where $Z_o = \mu_o = o$ and $o < r \le 1$. Roberts (1959) showed that, if $X_1, X_2 \ldots$ are iid N (o, σ2) random variables, then the mean of Zi is o and the variance is:

$$\sigma^2_{z_i} = \frac{r}{2-r}[1-(1-r)^{2i}]\sigma^2 \tag{2}$$

when the control value of the mean is 0, the control limits of the EWMA chart are often set at $\pm L\sigma_{Zi}$, here L and r are the parameters of the chart. Lucas and Saccucci (1990) have discussed the choice of r and L for the univariate EWMA chart in detail, although their control limits were based on the asymptotic form of $\sigma_{Zi}$ that is:

$$\sigma^2_{z_i} = \frac{r}{2-r}\sigma^2$$

In the multivariate case, a natural extension is to define the vectors of EWMA's,

$$Z_i = RX_i + (1 - R)Z_{i-1} \qquad i = 1, 2, \ldots, \tag{3}$$

where $Z_o = 0$ and R = diag(r1, r2, . . . , r p ), $O < r \le 1$, j = 1 , 2 , . . . ,p. The MEWMA chart gives an out of control signal as soon as:

$$T_i^2 = \mathbf{z}_i' \boldsymbol{\Sigma}_{\mathbf{z}_i}^{-1} \mathbf{z}_i > L \tag{4}$$

where L (>O) is chosen to achieve a specified in control ARL and $\Sigma_{zi}$ , is the covariance matrix of $Z_i$ If there is no a priori reason to weight past observations differently for the p quality characteristics being monitored, then $r_1 = r_2 = \ldots = r_p = r$. If the variables being monitored are not of equal importance and the desired ARL performance is such that the ARL should not be a function of λ, then the method of Hawkins (1991) is recommended. Another possibility, proposed by Tsui and Woodall (1991), is to use a different matrix in Equation (3) to calculate the quadratic form of the MEWMA chart.

If $r_1 = r_2 = \ldots = r_p = r$, then the MEWMA vectors can be written as:

$$Z_i = rX_i + (1 - r)Z_{i-1}, \qquad i = 1, 2, \ldots$$

$$\Sigma^2_{z_i} = \frac{r}{2-r}[1-(1-r)^{2i}]\Sigma^2 \tag{5}$$

Analogous to the situation in the univariate case, the MEWMA chart is equivalent to Hotelling's χ2 chart if r =1. As MacGregor and Harris (1990) pointed out for the univariate case, using the exact variance of the EWMA statistic leads to a natural fast initial response for the EWMA chart. Thus initial out of control conditions are detected more quickly.

This is also true for the MEWMA chart. Because, however, it may be more likely that the process will stay in control for a while and then shifts out of control, we will assume for a chart design and in our ARL comparisons that the asymptotic (as i → ∞) covariance matrix, that is:

$$\Sigma^2_{z_i} = \frac{r}{2-r}\Sigma^2$$

used to calculate the MEWMA statistic in (5) unless otherwise indicated.

### 2.1. Multivariate process model

Suppose that a multivariate process is monitored by means of a MEWMA control chart on p important quality characteristics. Let $\mathbf{X}_{ij}$=( $X_{ij1}$ , $X_{ij2}$ , ..., $X_{ijp}$)′ be a p 1 vector which represents the p characteristics on the jth observation (j = 1, 2, ..., n) in the ith subgroup of size n.

Suppose further that when the process is in control, the $\mathbf{X}_{ij}$ 's are independent and identically distributed and follow a p variate normal distribution with mean vector $\boldsymbol{\mu}_o$ and covariance matrix $\Sigma_o$. that $\mathbf{X}$ij 's are iid $\mathbf{N\ p}$ ($\boldsymbol{\mu}_o$, $\Sigma_o$) , when the process is in control. We let n denote the subgroup size and $\mathbf{X}$ i denote the average vector of the ith subgroup. That is:

$$\bar{X}_{i_i} = \frac{1}{n}\sum_{j=1}^{n} X_{ij}$$

When the ith subgroup is observed, the statistic:

$$T_i^2 = \mathbf{z}_i' \boldsymbol{\Sigma}_{\mathbf{z}_i}^{-1} \mathbf{z}_i > \mathrm{L}$$

where L (>O) is chosen to achieve a specified in control ARL and $\Sigma_{zi}$ , is the covariance matrix of $Z_i$ .

We will assume that, when the multivariate process means changes, there has been a step change from its in control value of $\mu = \boldsymbol{\mu}_o$ to an unknown value $\boldsymbol{\mu} = \boldsymbol{\mu}_1$ where $\mu_o \neq \boldsymbol{\mu}_1$ . If the statistic of MEWMA exceeds the L (as mention), we may conclude that the step-change in the process mean has occurred after some unknown time $\tau$ where $o \le \tau \le T-1$ .

Hence, we assume that the subgroup averages $\mathbf{X}_1$ , $\mathbf{X}_2$ , ..., $\mathbf{X}_\tau$ came from the in control process and the subgroup averages $\mathbf{X}_{\tau^1}, \mathbf{X}_{\tau^2}, \ldots, \mathbf{X}_T$ from the out of control process. We further assume that there is no change in the covariance structure and that the process mean remains at the new level $\boldsymbol{\mu}_1$ until the special cause has been identified and removed. The maximum likelihood estimator of $\tau$ can be shown to be the value of t for which the statistic Mt attains its maximum. That is:

$$\hat{\tau} = argmaxM_t \quad t = 0,1,\ldots,T\text{-}1 \tag{7}$$

$$M_t = (T-t)(\bar{\bar{X}}_{t,T_i} - \mu_0)' \textstyle\sum_0^{-1} (\bar{\bar{X}}_{t,T_i} - \mu_0)$$

is the average of the T-t most recent subgroup averages.( See appendix for the derivation of this estimator).

**Table 1**. Bivariate normal observation and the MEWMA statistics.

| Row | Obs1 | Obs2 | z1 | z2 | $T^2_i$ |
|---|---|---|---|---|---|
| 1 | -0.70 | -0.48 | -0.07 | -0.05 | 0.51 |
| 2 | 0.29 | 1.19 | -0.03 | 0.08 | 0.70 |
| 3 | 0.25 | 0.06 | -0.01 | 0.07 | 0.33 |
| 4 | 1.45 | -0.56 | 0.14 | 0.01 | 0.81 |
| 5 | 1.43 | -0.13 | 0.27 | 0.00 | 2.83 |
| 6 | -0.99 | 0.70 | 0.14 | 0.07 | 0.54 |
| 7 | 1.04 | 1.69 | 0.23 | 0.23 | 1.75 |
| 8 | -0.30 | 0.50 | 0.18 | 0.26 | 1.61 |
| 9 | -0.51 | -2.33 | 0.11 | 0.00 | 0.37 |
| 10 | 1.40 | 1.72 | 0.24 | 0.17 | 1.31 |
| 11 | -0.51 | 0.31 | 0.16 | 0.18 | 0.86 |
| 12 | 0.68 | -0.17 | 0.22 | 0.15 | 1.01 |
| 13 | 1.16 | 1.54 | 0.31 | 0.29 | 2.43 |
| 14 | -0.69 | 0.28 | 0.21 | 0.29 | 1.77 |
| 15 | -2.54 | -0.85 | -0.07 | 0.17 | 1.20 |
| 16 | -0.98 | 1.24 | -0.16 | 0.28 | 3.85 |
| 17 | -0.24 | -1.75 | -0.16 | 0.08 | 1.19 |
| 18 | 0.70 | 1.73 | -0.08 | 0.24 | 2.17 |
| 19 | -0.75 | -1.18 | -0.15 | 0.10 | 1.18 |
| 20 | -1.14 | 1.31 | -0.25 | 0.22 | 4.19 |
| 21 | 0.57 | -0.67 | -0.16 | 0.13 | 1.68 |
| 22 | -0.12 | -1.47 | -0.16 | -0.03 | 0.55 |
| 23 | -0.55 | 0.35 | -0.20 | 0.01 | 1.05 |
| 24 | -0.32 | -0.90 | -0.21 | -0.08 | 0.86 |
| 25 | -1.66 | -1.38 | -0.36 | -0.21 | 2.44 |
| 26 | 0.03 | -1.92 | -0.32 | -0.38 | 3.18 |
| 27 | 0.83 | -0.30 | -0.20 | -0.37 | 2.67 |
| 28 | -0.88 | -1.77 | -0.27 | -0.51 | 5.03 |
| 29 | 1.00 | 0.40 | -0.14 | -0.42 | 3.53 |
| 30 | -1.58 | -0.91 | -0.29 | -0.47 | 4.31 |
| 31 | 0.03 | -0.16 | -0.26 | -0.44 | 3.73 |
| 32 | -0.68 | 0.30 | -0.30 | -0.37 | 2.89 |
| 33 | -0.20 | -1.75 | -0.29 | -0.50 | 4.88 |
| 34 | -1.02 | -2.41 | -0.36 | -0.70 | 9.20 |

**Table 2**. Reverse cumulative average vectors and the $M_t$ statistics.

| **Row** | $\bar{\bar{X}}_{t,T}{}'$ | $\bar{\bar{X}}_{t,T}{}'$ | $\mathbf{M_t}$ |
|---|---|---|---|
| 1 | -0.16 | -0.23 | 1.19 |
| 2 | -0.15 | -0.22 | 0.94 |
| 3 | -0.16 | -0.27 | 1.08 |
| 4 | -0.17 | -0.28 | 1.23 |
| 5 | -0.23 | -0.27 | 2.05 |
| 6 | -0.28 | -0.27 | 3.11 |
| 7 | -0.26 | -0.31 | 2.49 |
| 8 | -0.31 | -0.38 | 3.38 |
| 9 | -0.31 | -0.41 | 3.26 |
| 10 | -0.30 | -0.34 | 2.98 |
| 11 | -0.37 | -0.42 | 4.37 |
| 12 | -0.36 | -0.45 | 4.05 |
| 13 | -0.41 | -0.47 | 4.95 |
| 14 | -0.49 | -0.56 | 6.60 |
| 15 | -0.48 | -0.60 | 6.02 |
| 16 | -0.37 | -0.59 | 3.40 |
| 17 | -0.33 | -0.69 | 2.65 |
| 18 | -0.34 | -0.63 | 2.59 |
| 19 | -0.40 | -0.78 | 3.45 |
| 20 | -0.38 | -0.75 | 2.88 |
| 21 | -0.33 | -0.90 | 1.97 |
| 22 | -0.39 | -0.92 | 2.69 |
| 23 | -0.42 | -0.87 | 2.78 |
| 24 | -0.41 | -0.98 | 2.41 |
| 25 | -0.41 | -0.99 | 2.28 |
| 26 | -0.27 | -0.95 | 0.91 |
| 27 | -0.31 | -0.83 | 1.05 |
| 28 | -0.48 | -0.90 | 2.12 |
| 29 | -0.41 | -0.76 | 1.34 |
| 30 | -0.69 | -0.99 | 3.18 |
| 31 | -0.47 | -1.01 | 1.17 |
| 32 | -0.63 | -1.29 | 1.60 |
| 33 | -0.61 | -2.08 | 0.99 |
| 34 | -1.02 | -2.41 | 1.38 |



## 2.2. Illustrative example

The bivariate normal distribution is considered with mean of (0, 0) unit variances and a correlation coefficient of 0.5. In Table 1, the values of ($X_1$, $X_2$) are the observations, the values ($Z_1$, $Z_2$) correspond to the MEWMA vector in (3) with r =0.1, and the values of $T^2$ were obtained using Equation (4). The use of Equation (5) provides the natural head start feature for the MEWMA chart. The values of L=8.79 were obtained using simulation to provide in control ARL's of 200. The MEWMA chart based on (4) signals out of control after the 34th observation. Note that when the MEWMA chart signals, the MEWMA vector elements, $Z_1$ and $Z_2$, give some indication of the direction of the shift.

The sample averages of 34 subgroups and the corresponding MEWMA statistics are shown in Table 1. The control chart has issued an alarm for the 34th subgroup since MEWMA, L=8.79 thus T = 34. Our proposed estimator can now be applied to estimate the change point. To do so, we need to find the reverse cumulative averages $\bar{\bar{X}}_{t,T}$ for t = 0, 1, 2... T1. In our example, the signal was issued at T = 34.

$$\bar{\bar{X}}_{t,34} = \frac{1}{34-t}\sum_{i=t+1}^{34} \bar{x}_i \qquad \text{t=0,1,…,T-1}$$

$$\bar{\bar{X}}_{33,34} = \frac{1}{34-33}\sum_{i=34}^{34} \bar{x}_{34} = [-1.02,-2.41]'$$

$$\bar{\bar{X}}_{32,34} = \frac{1}{34-32}\sum_{i=33}^{34} \bar{x}_{34} + \bar{x}_{33} = [\,-0.61,-2.08\,]'$$

and so on. All 34 of these reverse cumulative averages are shown in Table 2. The Mt statistics are then calculated for t = 0, 1, 2... 33 where:

$$M_t = (34-t)(\bar{\bar{X}}_{t,T} - \mu_0)' \sum_0^{-1} (\bar{\bar{X}}_{t,T} - \mu_0)$$

$$M_{34} = (34-33)(\bar{\bar{X}}_{33,34} - \mu_0)' \sum_0^{-1} (\bar{\bar{X}}_{33,34} - \mu_0) = 1.38$$

$$M_{33} = (34-32)(\bar{\bar{X}}_{34,32} - \mu_0)' \sum_0^{-1} (\bar{\bar{X}}_{34,32} - \mu_0) = 0.99$$

All 34 of these $M_t$ statistics are shown in Table 2. Process engineers could use this information in their investigations for the special cause responsible for the shift in the process mean. For example, they could review the process log books and study the information recorded during and around the time when subgroup 14 was machined to identify a special cause. Upon identifying this as the special cause, the process engineers could correct the misalignment in the tool. If the process engineers had examined their records corresponding to subgroup T = 34 only, i.e., the subgroup for which the MEWMA control chart issued the signal, they might have failed to identify the special cause or they might have incorrectly identified a special cause. Alternatively, the process engineers could have started examining their records at the time of signal and searched backwards until a special cause was found. But, using the researchers' proposed estimator is a more efficient way to search for the special cause. Process engineers could initiate their search at the time recommended by the proposed estimator and they could expand their search window by examining the records corresponding to subgroups machined before and after the estimated change point.

### 2.3. Performance evaluations of MLE

In this section, we will study the performance of the researchers' proposed estimator using Monte Carlo simulation. Two performance measures, namely, the average change point estimate and the empirical distribution of the estimated change point around the actual change point are considered.

The observations were assumed to come from a N p ($\mu_o$ , $\sigma_o$) 0 distribution when the process is in control. Two process dimensions, namely p = 2, 5 were considered. One hundred subgroups of size n = 1 were generated randomly from the in control distribution. If the MEWMA statistic for any of these subgroups exceeded the L, all data from that subgroup were discarded and replaced with new data. The new MEWMA statistic was then recomputed and compared with the L. This procedure was repeated, as needed, until 100 subgroups from the in control process had MEWMA statistics that did not exceed the L. Thus, the MEWMA control chart did not issue any false alarms. Starting with subgroup 101, the simulated process mean was changed from $\mu_0$ to $\mu_1$ by introducing a shift of magnitude $\lambda$ in the in control means.

Subgroups were then generated from the out-of-control process until a subgroup's MEWMA statistic exceeded its L, that is, until the control chart issued a genuine alarm signal. The change point estimate was then calculated following that genuine alarm signal using the method described earlier. This procedure was then replicated 10,000 times, and the average of those 10,000 change point estimates its standard error and the empirical distribution of the estimated change point around the actual change point were obtained. Seven shifts of magnitude $\lambda$ = 0.25, 0.5, 1, 1.5, 2, 2.5 and 3.0 were considered. The results of this simulation study are presented in the following tables. In Tables 3 and 4 we show E(T), the time at which the MEWMA control chart is expected to issue signal of a change in the process mean. Since the change had actually occurred following subgroup t =100, E(T) = 100 +ARL, where ARL is the average run length of the control chart for the out of control process. The average run length of a control chart is the expected number of subgroups required to detect a change in the process parameter. We also show the average of the change point estimates ($\bar{\bar{\tau}}$) and its standard error based on 10,000 replications. Since the actual change point was at subgroup t = 100, the average of the change point estimates ($\bar{\bar{\tau}}$) should be near 100. The results in Tables 3 and 4 show that the ($\bar{\bar{\tau}}$) averages are in fact close to the actual change point of t = 100 for approximately all shift magnitudes and for all dimensions ( p=2, p=5) considered.

## 3. Artificial neural network

The neural network is an approach to information processing that does not require algorithm or rule development. The three essential features of a neural computing network are the computing units, the connections between the computing units, and the training algorithm used to find values of the network parameters. Neural networks are trained by two main types of learning algorithms supervised and unsupervised learning. In general, supervised learning can be used in predicting or mapping problems and the clustering problems usually make use of unsupervised learning. Neural networks can be classified into two different categories, feed-forward and feed-back networks. In this study, the researchers have utilized the feed-forward network because it has been found to be an effective system for learning the distinguishing patterns from a body of examples.

A feed-forward network is composed of several layers, an input layer, one or more hidden layer(s), and an output layer. Neurons in the feed-forward network receive inputs only from the previous layer and feeds outputs only to the next layer. Multilayer feed-forward neural networks are used for modeling of many manufacturing processes which are typically trained through a back propagation algorithm.

The back propagation algorithm involves a forward and a backward pass. The purpose of the forward pass is to obtain the activation value, and the purpose of the backward pass is to adjust weights according to the differences between the desired and actual network outputs.

**Table 3**. Expected time of a signal with MEWMA, Average of change point estimates ( $\bar{\bar{\tau}}$ ) standard error for P=2 , τ=100,N=10000.

| **λ** | 0.25 | 0.50 | 0.75 | 1.00 | 1.50 | 2.00 | 2.50 | 3.00 |
|---|---|---|---|---|---|---|---|---|
| **E(T)** | 163.89 | 121.87 | 111.91 | 108.10 | 105.03 | 103.65 | 102.95 | 102.49 |
| **τ** | 124.85 | 101.81 | 98.72 | 97.95 | 98.98 | 99.55 | 100.00 | 100.36 |
| **Std(τ)** | 0.51 | 0.22 | 0.16 | 0.14 | 0.11 | 0.08 | 0.06 | 0.05 |

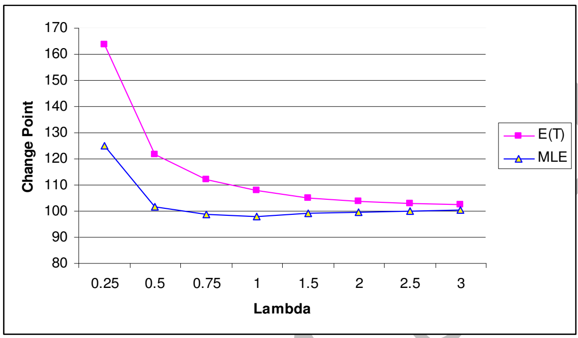


**Figure 1.** Expected time of a signal with MEWMA chart, average of change point estimates ( $\bar{\bar{\tau}}$ ) by MLE, 10000 time simulation for τ=100 and p=2.

**Table 4**. Expected time of a signal with MEWMA, average of change point estimates ( $\bar{\bar{\tau}}$ ) standard error for P=5 , τ=100, N=10000.

| **λ** | 0.25 | 0.50 | 0.75 | 1.00 | 1.50 | 2.00 | 2.50 | 3.00 |
|---|---|---|---|---|---|---|---|---|
| **E(T)** | 173.22 | 123.92 | 112.72 | 108.56 | 105.21 | 103.82 | 103.06 | 102.58 |
| **τ** | 117.97 | 99.17 | 97.26 | 97.38 | 98.67 | 99.49 | 100.09 | 100.37 |
| **Std(τ)** | 0.51 | 0.24 | 0.18 | 0.15 | 0.11 | 0.08 | 0.06 | 0.05 |

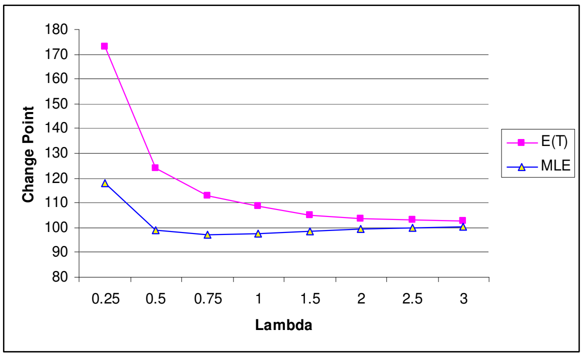


**Figure 2.** Expected time of a signal with MEWMA chart, average of change point estimates ( $\bar{\bar{\tau}}$ ) by MLE, 10000 time simulation for τ=100 and p=5.

**Table 5.** Expected time of a signal with MEWMA, Average of change point estimates ( $\bar{\bar{\tau}}$ ) by NN, standard error for P=2 , τ=100,N=10000.

| **λ** | 0.25 | 0.50 | 0.75 | 1.00 | 1.50 | 2.00 | 2.50 | 3.00 |
|---|---|---|---|---|---|---|---|---|
| **E(T)** | 163.86 | 121.78 | 112.07 | 108.20 | 105.02 | 103.66 | 102.96 | 102.51 |
| **τ** | 102.69 | 102.16 | 102.02 | 102.62 | 101.19 | 101.04 | 100.89 | 100.81 |
| **Std(τ)** | 0.40 | 0.30 | 0.17 | 0.10 | 0.05 | 0.02 | 0.02 | 0.02 |

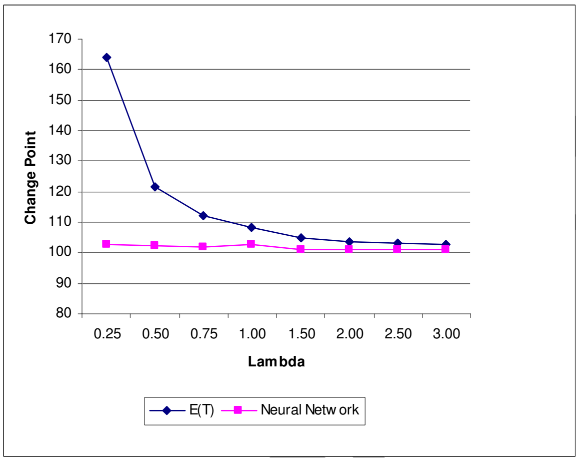


**Figure 3.** Expected time of a signal with MEWMA chart, Average of change point estimates ( $\bar{\bar{\tau}}$ ) by NN, 10000 time simulation for τ = 100 and p=2.

**Table 6.** Expected time of a signal with MEWMA, Average of change point estimates ( $\bar{\bar{\tau}}$ ) by NN, standard error for P=5, τ=100, N=10000.

| **λ** | 0.25 | 0.50 | 0.75 | 1.00 | 1.50 | 2.00 | 2.50 | 3.00 |
|---|---|---|---|---|---|---|---|---|
| **E(T)** | 172.63 | 124.25 | 112.80 | 108.68 | 105.30 | 103.86 | 103.08 | 102.59 |
| **τ** | 101.91 | 101.92 | 101.91 | 101.88 | 101.18 | 100.79 | 100.76 | 100.83 |
| **Std(τ)** | 0.38 | 0.30 | 0.18 | 0.11 | 0.04 | 0.03 | 0.02 | 0.02 |

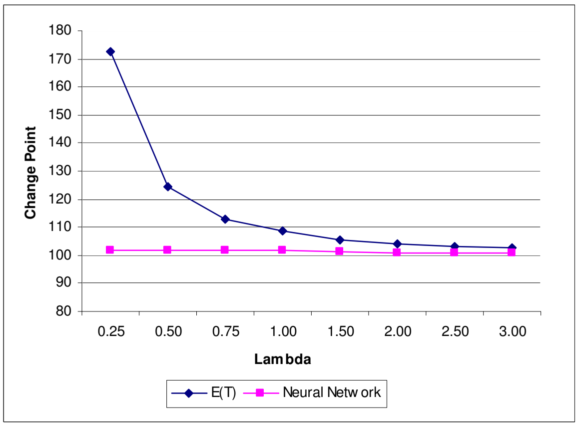


**Figure 4.** Expected time of a signal with MEWMA chart, Average of change point estimates ( $\bar{\bar{\tau}}$ ) by NN, 10000 time simulation for τ = 100 and p=5.

### 3.1. Training Neural Networks (ANNs)

ANNs are extensively parallel computing mechanisms modeled on the operation of a human brain. These networks consist of numerous interconnected processing elements called neurons which are typically organized into layers linked via weights. The networks are trained in advance to exhibit a desired behavior, by iterating through several input, output vectors and adjusting the weights. The fundamentals of ANNs can be found in Zurada (1992).

The application of neural networks involves selecting feature vectors, establishing the network architecture, choosing the activation function and training as follows:

### 3.2. Selecting the feature vector

The selection of feature vector in a crucial training affects the neural network. The feature vector must be able to help classifying shifts. Herein, each subgroup has 10 samples. Samples of subgroups correspond with the mean of 5 after simple observations.

### 3.3. Establishing the network architecture

The neural network adopted in this study is a three-layer, fully connected, feed-forward network with a back propagation training algorithm and is successfully applied to various classification problems. The neural network architecture including an input layer with n nodes, a hidden layer with 40 nodes, and an output layer with n nodes. (n = size of observation until MEWMA gives out the control signal). Many theoretical and simulative investigations of engineering applications have demonstrated that the number of hidden layers need not exceed two (Kecman, 2001). Whether one hidden layer or two hidden layers are required is determined on a case by case basis (Chester, 1990; Hayashi, Sakata, & Gallant, 1990; Hush & Horne, 1993; Kurkova, 1992). One hidden layer suffices for the back propagation network herein to approximate any continuous mapping from the input to output patterns, to an arbitrary degree of accuracy. Selecting the suitable number of neural units in the hidden layer is seldom as straightforward as determining the number of input and output neurons. No general guidelines exist for specifying the optimal number of nodes required in the hidden layer, although some hints have been proposed by Hirose, Yamashita, and Hijiya (1991) and Sietsma and Dow (1991). The optimal number of nodes depends on the problem. If a network has too few hidden neurons, it cannot learn the training set and generalize it well. On the other hand, a network with too many neurons may tend to memorize the training set and cannot generalize it well. In this study, the number of hidden neurons was selected based on trials and errors, having 40 nodes in one hidden layer.

### 3.4. Choosing the activation function

The activation function is important in the training stage. The one most extensively applied in back propagation algorithm is the log sigmoid transfer function $f(y)=1/(1+e^{-y})$, with output values that range [0,1] . Another activation function is the tangent-sigmoid transfer function with output values in the interval [-1 , 1]. It is defined as:

$$f(y) = \frac{e^{y}-e^{-y}}{e^{y}+e^{-y}}$$

Cheng (1995) claimed that the tangent sigmoid transfer function effectively detects process changes in every direction. The tangent sigmoid function is used herein as the activation function in the hidden layer while linear function is for the output layer.

### 3.5. Training

The input vectors are presented to the network and propagated forward to yield the output. While the weights and biases are iteratively modified according to the difference between the target and generated outputs. The Mean Square Error (MSE) associated with the output layer is propagated backward through the network, by modifying the weights. The popularity of the gradient descent search method is based more on its simplicity than on its search power. This study implements the Levenberg Merquardt Quasi-Network approach to replace the gradient descent method. Principe, Euliano, and Lefebvre (2000) stated that the Levenberg Merquardt Quasi Network method is particularly appropriate for training ANNs with MSE. Furthermore, it interfaces well with the back-propagation formalism. The training is terminated when the MSE of the difference of two successive iterations is within a predetermined tolerance 0.001 or when 100 epochs had been performed. When the MEW-

MA control chart gives signals for a sample which is out of control, the sample data are collected as input data to the trained network. (through this stage it is supposed that generating data has bivariate normal distribution with mean (0, 0) unit variances and a correlation coefficient of 0.5 .The values of ($X_1$, $X_2$) are the observations, the values ($Z_1$, $Z_2$) correspond to the MEWMA vector in (3) with r =0.1, and the values of $T^2$ were obtained using Equation (4). The values of L=8.79 were obtained using simulation to provide in control ARL's of 200, and we train and test the network with $T^2$ estimator of MEWMA).

In the training stage Mont Carlo simulation is used, i.e. generation of 100 bivariate normal with mean $\mu_0$ for first class, and creating shift magnitude of $\lambda$ at observation 100th and generating again 100 bivariate normal data with mean $\mu_1$ as the second class. To start training stage some transformation has been done on the raw data. First we calculate the mean of five consequent data to smooth data and consider it as Xi (i=1…200) then we put 10 consequent Xi as input vector. Output data for the first class are 0 and for the second class are 1. This work has been done hundred times for each shift magnitude all over the training phase. Seven shifts of magnitude $\lambda$ = 0.25, 0.5, 1, 1.5, 2, 2.5 and 3.0 were considered. As soon as the input vector is passed through the trained network, the output activation at each output is examined against a pre specified decision interval to yield a transference output 0 or l.

### 3.6. Test and performance evaluations of neural network

The researchers study the performance of their proposed network using Monte Carlo simulation. The observations were assumed to come from a N p ($\mu_o$, $\sigma_o$) 0 distribution when the process is in control. Two process dimensions, namely p = 2, 5 were considered. One hundred subgroups of size n = 1 were generated randomly from the in-control distribution.

If the MEWMA statistic for any of these subgroups exceeded the L, all data from that subgroup were discarded and replaced with new ones. The new MEWMA statistic was then recomputed and compared with L. This procedure was repeated, as required, until 100 subgroups from the in control process had MEWMA statistics that did not exceed L. Thus, the MEWMA control chart did not issue any false alarms. Starting with subgroup 101, the simulated process mean was changed from μ0 to μ1 by introducing a shift of magnitude $\lambda$ within the in control mean where subgroups were then generated from the out of control process until a subgroup's MEWMA statistic exceeded its L, that is, until the control chart issued a genuine alarm signal. The change point estimation was then calculated following that genuine alarm signal, using the aforementioned method. This set of data is considered as input vector to the network to test the network. The output of the network concludes a vector of 0 and 1; if our network has a good training it must include a series of 0 and 1, that is shown where 0 is changed to 1 is, where the change or shift in mean occurs. This procedure was then replicated 10,000 times, and the average of those 10,000 change point estimates its standard error.

Seven shifts of magnitude $\lambda$ = 0.25, 0.5, 1, 1.5, 2, 2.5 and 3.0 are considered. The results of this simulation study are presented in the following tables. In tables 5-6 we show E(T), the time at which the MEWMA control chart is expected to issue a genuine signal of a change in the process mean. Also results show that the ($\bar{\bar{\tau}}$) averages are in fact close to the actual change point of t = 100 for all shift magnitudes and for all dimensions considered.

Thus, on average, our proposed network change point estimate is close to the actual time of change regardless of the values of the shift magnitude and process dimension. For example Table 5 shows, when p = 2 and $\lambda$ = 1, the MEWMA control chart on average will signal the change in the process mean on subgroup 108.20 when the actual change occurred after the 100th subgroup. That is, the control chart on average signals the change 8.20 subgroups after the actual change. But Design NN on average will detect the change in the process mean on subgroup 102.62, it means 2.62 subgroup after the actual time of the change neural network can detect the change.

When p = 2 and $\lambda$ = 3, the MEWMA control chart on average will signal the change in the process mean on subgroup 102.51 when the actual change occurred after the 100th subgroup. But Design NN on average will detect the change in the process mean on subgroup 100, 81. When p = 2 and $\lambda$ = 0.25, the MEWMA control chart on average will signal the change in the process mean on subgroup 163.86 when the actual change occurred after the 100th subgroup. But Design NN on average will detect the change in the process mean on subgroup 102.69. Thus, seeking special causes at the time of the signal might be futile, if the process engineers only search their databases for special causes that may had occurred around the time used instead. The process engineers would appropriately conclude, on average, the shift had occurred after the 100th subgroup and consequently have a much better chance of identifying the special cause. Results show that the neural network change point estimates around the actual change point will not depend on the shift magnitude and for all shift magnitude is approximately up to two subgroups after the actual change point. The researchers have also designed the network and tested it by 5 variables.

Table 6 shows the same results, i.e. the design is good for all shift magnitude. For example: when p = 5 and $\lambda$ = 0.25, the MEWMA control chart on average will signal the change in the process mean on subgroup 172.63 when the actual change occurs after the 100th subgroup. But Design NN on average will detect the change in the process mean on subgroup 101.91 and when p = 5 and $\lambda$ = 3 , the MEWMA control chart on average will signal the change in the process mean on subgroup 102.59 when the actual change occurs after the 100th subgroup. But Design NN on average will detect the change in the process mean on subgroup 100.83.

Results show that MEWMA will signal soon for shift magnitudes greater than one, but Design NN will detect soon for all shift magnitudes. Also design network is examined when actual time of change points are 50.150; the results are included in the appendix. We might note that using a simulated change point different from $\tau$ =100, does not alter our conclusions (see the appendix).

## 4. Comparison of MLE and NN

Comparison performance of MLE and proposed NN using Monte Carlo simulation is provided in Table 7. As it is shown in Table 7 on average, the researchers' proposed network estimate change point is close to the actual time of change regardless of the values of the shift magnitude and process dimension, but MLE change point estimator for shift magnitudes greater than one is close to the actual time of change point but for shift magnitudes less than one is not close to the real change point. For example table shows, when p = 2 and $\lambda$ = 1, the MEWMA control chart on average will signal the change in the process mean on subgroup 108.15 when the actual change occurred after the 100th subgroup. That is, the control chart on average signals the change 8.15 subgroups after the actual change. But Design NN on average will detect the change in the process mean on subgroup 102.62 , i.e. 2.62 subgroup after the actual time of the change neural network can detect the change and MLE on average will detect the change in the process mean on subgroup 97.95, i.e. 2.05 subgroup before the actual time change. Also when p = 2 and $\lambda$ = 3, the MEWMA control chart on average will signal the change in the process mean on subgroup 102.50 when the actual change occurs after the 100th subgroup. But Design NN on average will detect the change in the process mean on subgroup 100.81 and MLE on average will detect the change in the process mean on subgroup 100.36, thus the difference between the two ways is 0.45. When p = 2 and $\lambda$ = 0.25, the MEWMA control chart on average will signal the change in the process mean on subgroup 163.87 when the actual change occurs after the 100th subgroup. But Design NN on average will detect the change on subgroup 102.69 and MLE on average will detect the change in the process mean on subgroup 124.85, thus the deference between the two ways is 22.16. The same result can be observed for P=5. Also the same results can be seen when actual time of change point is 50, 150; the results are included in the appendix. We note that using a simulated change point different from $\tau$ = 100, does not alter our conclusions (see the appendix).

## 5. Conclusions

In this paper the researchers have proposed two approaches for identifying the time of a step change in a multivariate process mean. The first approach is MLE and the next is NN. They described how the MEWMA control charts can be used in conjunction with MLE and NN. When a MEWMA control chart signals a change in the process mean, process engineers carry out a search for special causes responsible for the change. Confining the search only to the time of the signal is likely to be ineffective since the actual change may have taken place a substantial amount of time before the signal. This is especially true when there is only a small change in the process mean since the average run length can be quite large. Process engineers can improve the chances of identifying the specific cause by using the proposed estimator especially for NN. They illustrated the use of their proposed NN and MLE with an example involving a MEWMA control chart. The researchers presented the results of some simulation experiments carried out to evaluate the performance of the MLE and NN.

**Table 7**. Expected time of a signal with MEWMA, average of change point estimates ( $\bar{\bar{\tau}}$ ) , standard error for P=2,5, τ=100, N=1000, with MLE , NN.

| **p=2 , λ** | | 0.25 | 0.50 | 0.75 | 1.00 | 1.50 | 2.00 | 2.50 | 3.00 |
|---|---|---|---|---|---|---|---|---|---|
| **MLE** | **E(T)** | 163.89 | 121.87 | 111.91 | 108.10 | 105.03 | 103.65 | 102.95 | 102.49 |
| | **τ** | 124.85 | 101.81 | 98.72 | 97.95 | 98.98 | 99.55 | 100.00 | 100.36 |
| | **Std(τ)** | 0.51 | 0.22 | 0.16 | 0.14 | 0.10 | 0.08 | 0.06 | 0.04 |
| **NN** | **E(T)** | 163.86 | 121.78 | 112.07 | 108.20 | 105.02 | 103.66 | 102.96 | 102.51 |
| | **τ** | 102.69 | 102.16 | 102.02 | 102.62 | 101.19 | 101.04 | 100.89 | 100.81 |
| | **Std(τ)** | 0.40 | 0.30 | 0.17 | 0.10 | 0.05 | 0.02 | 0.02 | 0.02 |
| **p=5 , λ** | | 0.25 | 0.50 | 0.75 | 1.00 | 1.50 | 2.00 | 2.50 | 3.00 |
| **MLE** | **E(T)** | 173.22 | 123.92 | 112.72 | 108.56 | 105.21 | 103.82 | 103.06 | 102.58 |
| | **τ** | 117.97 | 99.17 | 97.26 | 97.38 | 98.67 | 99.49 | 100.09 | 100.37 |
| | **Std(τ)** | 0.51 | 0.24 | 0.18 | 0.15 | 0.11 | 0.08 | 0.06 | 0.05 |
| **NN** | **E(T)** | 172.63 | 124.25 | 112.80 | 108.68 | 105.30 | 103.86 | 103.08 | 102.59 |
| | **τ** | 101.91 | 101.92 | 101.91 | 101.88 | 101.18 | 100.79 | 100.76 | 100.83 |
| | **Std(τ)** | 0.38 | 0.30 | 0.18 | 0.11 | 0.04 | 0.03 | 0.02 | 0.02 |

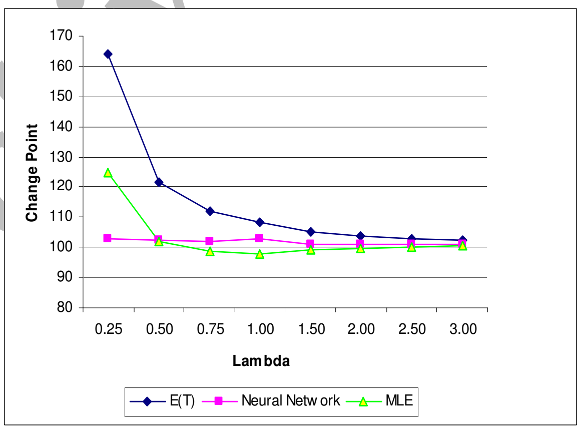


**Figure 5.** Expected time of a signal with MEWMA chart, average of change point estimates ( $\bar{\bar{\tau}}$ ) by MLE and NN, 10000 time simulation for τ=100 and p=2.

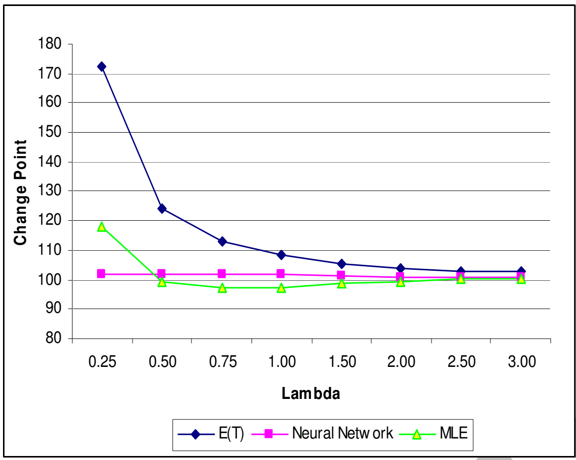


**Figure 6**. Expected time of a signal with MEWMA chart, average of change point estimates ( $\bar{\bar{\tau}}$ ) by MLE and NN 10000 time simulation for $\tau$=100 and p=5.

The simulation studies showed that given a change in the process mean, the network performed effectively and equally well in detecting the actual change point for all shift magnitudes and all dimensions considered but MLE perform well in detecting the actual change point for shifts greater than one; especially for shifts equal to three but for shifts less than one MLE estimator is not effectively well; especially for shifts equal to 0.25. Also results show design network will detect change point of the process mean for all shift magnitudes up to two subgroups after the actual change point but MLE for some shift magnitudes detect change before and for some shifts after the actual change point; thus the size of searching windows in MLE is larger than NN. Design network is also examined when the actual time of change point is 50, 150 and results show that using a simulated change point different from 100 does not alter the conclusions. Ultimately, the results show that the proposed network performs better than MLE in detecting the shifts.

## References


[1] Casella, G. and Berger, R. L., 1990, *Statistical Inference*. Wadsworth and Brooks/Cole Belmont, CA.

[2] Cheng, C. S, 1995, A multi-layered neural network model for detecting changes in the process mean. *Computers & Industrial Engineering*, 51-61.

[3] Cheng, C. S. , 1997 , A neural network approach for the analysis of control chart patterns. *International Journal of Production Research,* 35(3), 667-697.

[4] Guo, Y. and Dooley, K. J, 1992 , Identification of change structure in statistical process control. *International Journal of Production Research,* 30(7), 1655-1669.

[5] Hinkley, D. V, 1970, Inference about the Change-Point in a Sequence of Random Variables. *Biometrika*, 57, 1-17.

[6] Hwarng, H. B. and Hubele, N. F, 1993a , Back propagation pattern recognizers for x-bar control. *Computers & Industrial Engineering*, 219-235.

[7] Hwarng, H. B, 1992, *Pattern Recognition on Shewhart Control Charts Using a Neural Network Approach.* PhD. Thesis, Arizona State University.

[8] Hwarng, H. B. and Hubele, N. F., 1993b, X-bar control chart pattern identification through efficient offline neural network training. *IIIE Transactions*, 25(3), 27-40.

[9] Johnson, R. A. and Wichern, D. W, 1992, *Applied Multivariate Statistical Analysis.* Prentice Hall, Englewood Cliffs, NJ.

[10] Long-Hui C. and Tai-Yue W., 2004, Artificial neural networks to classify mean shifts from multivariate $X^2$ chart signals. *Computers and Industrial Engineering*, 195-205.

[11] Lowry, C. A. and Montgomery, D. C, 1995, A

review of multivariate control charts. *IIIE Transaction,* 27, 800-810.

[12]Lowry, C. A.and Woodall and W. H., Champ, C. W. and Rigdon, S. E, 1992, A multivariate exponentially weighted moving average control chart. *Technometrics*, 34, 46-53.

[13]Lucas, J. M. and Saccucci, M. S, 1990, Exponentially weighted moving average control schemes properties and enhancements. *Technometrics*, 32, 1-12.

[14]Montgomery, D. C, 1991, *Introduction to Statistical Quality Control.* John Wiley & Sons, New York.

[15]Nishina, K, 1986, *Estimation of the Change Point From Cumulative Sum Tests*. Reports on Statistical Applications and Research, JUSE 33, 1-14.

[16]Nishina, K, 1992, A comparison of Control charts from the viewpoint of change point estimation. *Quality and Reliability Engineering International*, 537-541.

[17]Perry, M. B., Spoerre, J. K., and Velasco, T, 2001, Control chart pattern recognition using artificial neural. International Journal of Production Research, 39(15), 3399-3418.

[18]Pham, D. T. and Oztemel, E, 1994, Control chart pattern recognition using learning vector quantization networks. *International Journal of Production Research,* 721-729.

[19]Pugh, G. A, 1989, Synthetic neural networks for process control. *Computers and Industrial Engineering,* 17(1-4), 24-26.

[20]Pugh, G. A, 1991, A comparison of neural networks to SPC charts. *Computers and Industrial Engineering,* 21(1-4), 253-255.

[21]Samuel, T. R. and Pignatiello, J. J and Calvin, J. A, 1998, Identifying the time of a step change with khido control charts. *Quality Engineering*, 10(3), 521-527.

## Appendix: The effect of τ on the average performances of the NN change point estimators

This appendix sorts out the average performances of NN following a signal from a MEWMA control chart. The researchers examined values of P=2, p=5 for $\tau = 50, 150$ and found that in all these cases the NN estimator has an overall superior average performance regardless of which point is considered as actual change point.

**Table 8.** Expected time of a signal with MEWMA, Average of change point estimates ( $\bar{\bar{\tau}}$ ) By NN standard error for P=2 , τ=50,N=10000.

| **λ** | 0.25 | 0.50 | 0.75 | 1.00 | 1.50 | 2.00 | 2.50 | 3.00 |
|---|---|---|---|---|---|---|---|---|
| **E(T)** | 113.86 | 72.41 | 62.06 | 58.24 | 55.04 | 53.70 | 52.95 | 52.51 |
| **τ** | 52.62 | 52.03 | 52.40 | 52.37 | 51.24 | 51.00 | 50.93 | 50.75 |
| **Std (τ)** | 0.38 | 0.32 | 0.16 | 0.10 | 0.04 | 0.02 | 0.01 | 0.02 |

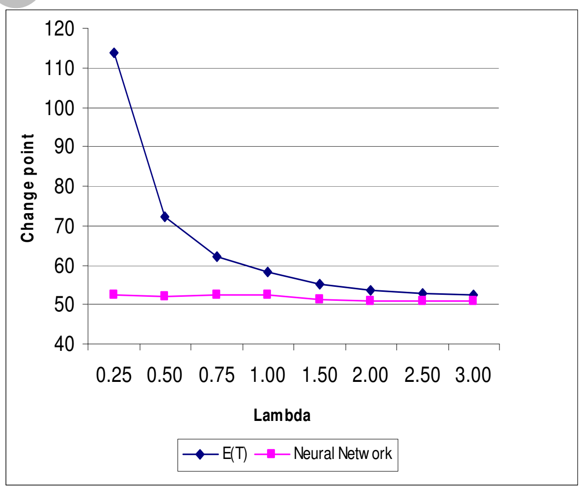


**Figure 7.** Expected time of a signal with MEWMA chart, average of change point estimates ( $\bar{\bar{\tau}}$ ) by NN 10000 time simulation for $\tau = 50$ and p = 2.

**Table 9.** Expected time of a signal with MEWMA, average of change point estimates ( $\bar{\bar{\tau}}$ ) by NN standard error for P=2, τ=150,N=10000.

| **λ** | 0.25 | 0.50 | 0.75 | 1.00 | 1.50 | 2.00 | 2.50 | 3.00 |
|---|---|---|---|---|---|---|---|---|
| **E(T)** | 214.53 | 172.15 | 162.11 | 158.35 | 155.03 | 153.70 | 152.96 | 152.50 |
| **τ** | 151.76 | 151.13 | 151.74 | 151.81 | 151.26 | 150.95 | 150.95 | 150.81 |
| **Std (τ)** | 0.38 | 0.28 | 0.18 | 0.10 | 0.05 | 0.02 | 0.02 | 0.02 |

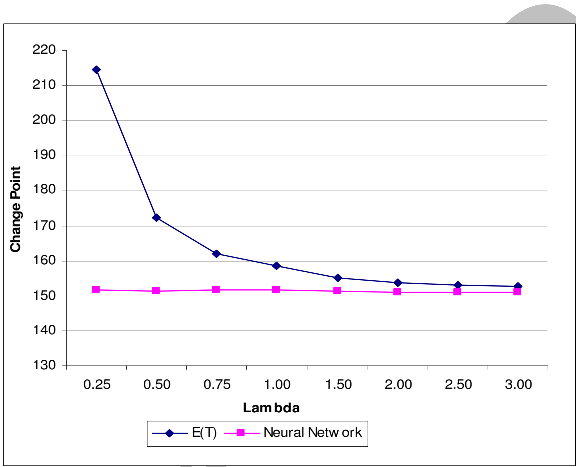


**Figure 8.** Expected time of a signal with MEWMA chart, average of change point estimates ( $\bar{\bar{\tau}}$ ) by NN 10000 time simulation for τ = 150 and p = 2.

**Table 10.** Expected time of a signal with MEWMA, average of change point estimates ( $\bar{\bar{\tau}}$ ) By NN standard error for P=5 , τ=50, N=10000.

| **λ** | 0.25 | 0.50 | 0.75 | 1.00 | 1.50 | 2.00 | 2.50 | 3.00 |
|---|---|---|---|---|---|---|---|---|
| **E(T)** | 122.85 | 74.29 | 62.92 | 58.69 | 55.30 | 53.88 | 53.09 | 52.59 |
| **τ** | 52.79 | 52.90 | 52.85 | 51.23 | 50.99 | 50.90 | 50.76 | 50.87 |
| **Std (τ)** | 0.44 | 0.31 | 0.20 | 0.12 | 0.06 | 0.02 | 0.03 | 0.02 |

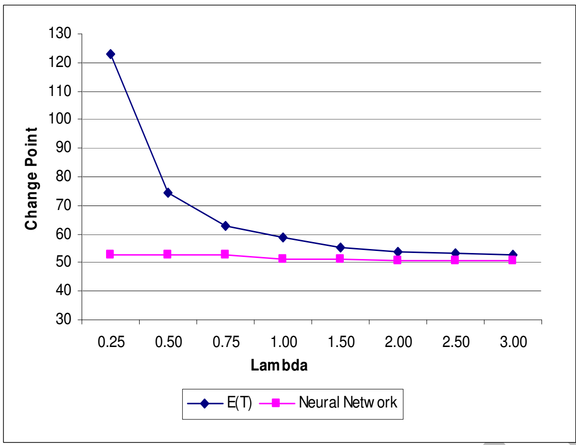


**Figure 9.** Expected time of a signal with MEWMA chart, average of change point estimates ( $\bar{\bar{\tau}}$ ) by NN 10000 time simulation for τ=50 and p=5.

**Table 11.** Expected time of a signal with MEWMA average of change point estimates ( $\bar{\bar{\tau}}$ ) by NN standard error for P=5 , τ=150,N=10000.

| **λ** | 0.25 | 0.50 | 0.75 | 1.00 | 1.50 | 2.00 | 2.50 | 3.00 |
|---|---|---|---|---|---|---|---|---|
| **E(T)** | 222.78 | 174.18 | 162.98 | 158.65 | 155.30 | 153.87 | 153.09 | 152.58 |
| **τ** | 152.83 | 152.19 | 151.41 | 151.66 | 150.75 | 150.84 | 150.85 | 150.80 |
| **Std (τ)** | 0.42 | 0.31 | 0.19 | 0.11 | 0.06 | 0.03 | 0.02 | 0.02 |

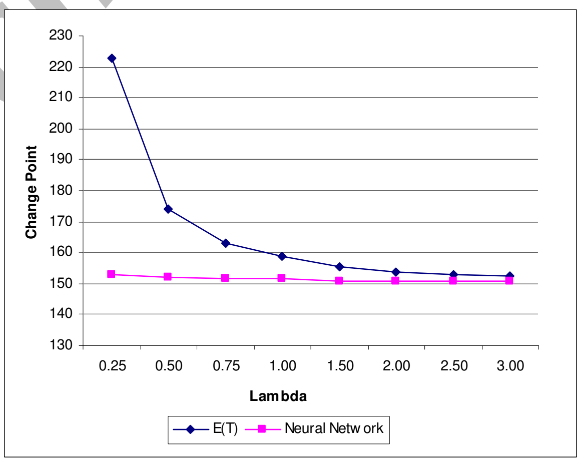


**Figure 10.** Expected time of a signal with MEWMA chart, average of change point estimates ( $\bar{\bar{\tau}}$ ) by NN10000 time simulation for τ=150 and p=5.