CHAPTER 2

RELATED WORK AND BACKGROUND STUDY

In this chapter, the literatures of past researches about EWMA controllers are reviewed. It is known that an integral (I) control action in a discrete proportional-integral-derivative (PID) controller can eliminate offsets or shifts and provide robustness. There is a close relation between I controllers and the exponentially weighted moving average (EWMA) statistic. In this chapter, some relative knowledge about feedback adjustment methods based on the EWMA statistic are presented. Beside this, a neural technique will also be introduced in this chapter.

2.1 Related Work

The EWMA statistic was first suggested by Roberts (1959) for process monitoring, but he referred to it as a geometric moving average (GMA). The use of the EWMA statistic has two distinct purposes (Fatin *et al.* 1990): as control charts (Box and Kramer 1992, Montgomery 1996, Box and Luceño 1997, Chen and Elsayed 2002) and as forecasts (Box and Jenkins 1976, Box *et al.* 1994, Brockwell and Davis, 1996). Recently, the statistic has been used widely for process adjustment purposes (Lucas and Saccucci 1992, Ingolfsson and Sachs 1993, Del Castillo and Hurwitz 1997, Del Castillo 2001, Pan and Del Castillo 2001, Del Castillo 2002, O'Shaughnessy and Haugh 2002, Fan *et al.* 2002).

In semiconductor manufacturing, EWMA controllers are sometimes called bias tuning controllers (Butler and Stefani 1994). The purpose of EWMA-based controllers is for compensating against disturbances which affect the run-to-run (batch-to-batch) variability in quality characteristics (Del Castillo 2002). Sachs *et al.* (1995) describe a controller that recommends process adjustments at each run of silicon wafers based on the EWMA.

The use of single EWMA controller has a relation to the pure-integral (I) control action, which was a part theme of the well-known PID controller. Box and Jenkins (1976) shown that a controller based on the single-EWMA statistic is a minimum mean square error (MMSE) controller when the underlying process disturbance model follows the IMA(1,1) (first-order integrated moving average) time-series process. Other than IMA(1,1) process, the single-EWMA controller had been shown to possess effective performance in some disturbances, for instance: step and ramp with slow drift rate disturbance models.

The performance of the EWMA controlled process output strongly depends on choosing the control parameters. As mentioned by Smith and Boning (1997), the controlled process output under a higher EWMA weight would return to the target much faster than a lower weight, but it would create more oscillations. Therefore, it is important to select the EWMA parameter carefully.

Due to a process environment is usually dynamic in a real manufacturing world, developing adaptive algorithm for self-tuning the single-EWMA controller is necessary. Satri (1988) used the theory of least squares estimation (LSE) for sequential parameter-detection and revision of the moving average parameter in the IMA(1,1) time series model. Luceño (1995) presented a computer program to choose the EWMA controller parameter in the EPC. His algorithm was based on the maximum likelihood estimate (MLE) theory. These above mentioned algorithms all have a common constraint in that the probability distribution must be known beforehand. Therefore, Smith and Boning (1997) utilized a neural network as an approximation function to map from the disturbance state (magnitude of linear drift and random noise) of a given process to the corresponding optimal EWMA weights.

However, it was a problem to estimate the slope of the controlled process. Del Castillo and Yeh (1998) presented an adaptive run-to-run multiple-input-multiple-output controller for linear and nonlinear semiconductor processes. Recently, an adaptive algorithm to estimate the EWMA gain was suggested by Patel and Jenkins (2000). Their objective was to design an automated scheme for optimizing the numerical parameter of the EWMA controller.

For a process that drifts considerably, the single-EWMA controller will tend to be significantly off-target. For this reason, Bulter and Stefani (1994) extended the single-EWMA controller with another EWMA equation in order to compensate for the ramp disturbance model. There are two mechanisms for the double EWMA controller (termed a predictor-corrector controller, PCC): one for estimating the drift rate and the other for estimating the step change deviation. The double-EWMA controller is a MMSE controller when the IMA(2,2) disturbance model exists in the process. Chen and Guo (2001) shown that the double-EWMA controller is not a discrete PID form, but an integral-double-integral (Ⅰ - Ⅱ) form.

Del Castillo (1999) presented an optimization form for solving the double-EWMA gains. The objective of his algorithm is to trade-off the initial transient effect and long-run variance. Del Castillo and Rajagopal (2002) extended the algorithm to a multiple-input multiple-output (MIMO) process.

Although Del Castillo (1999) suggested keeping the trade-off solution weights to control the process will provide adequate performance than a variety case. However, a time-varying weight can produce a superior performance than a fix one (Del Castillo and Hurwitz, 1997). Therefore, one objective of this research aims to develop a time-varying weights tuning strategy for the double-EWMA controller.

2.2 Single EWMA Controller

2.2.1 Brief Review

Assume that the relation between the input and the output of a manufacturing process can be expressed as follows:

$$y_t = \alpha + \beta x_{t-1} + \varepsilon_t \tag{2.1}$$

where y_t denotes the observed output deviation from target, ε_t is a white noise stochastic process, and x_t is the manipulated variable. The parameter of α represents the process offset, β is the process gain, and both parameters need to be estimated. Equation (2.1) implies that all the effects of a change in the compensating variable will be realized at the output, in one time interval. Such a system is called a responsive system (Box and Luceño 1997) and is commonly seen in the discrete part manufacturing. Let b represents an estimate of the gain (β) that can be estimated off-line by fitting the regression model. The single EWMA scheme can be expressed as follows:

$$x_t = -\frac{a_t}{b} \tag{2.2}$$

where

$$\begin{aligned} a_t &= \lambda(y_t - bx_{t-1}) + (1-\lambda)a_{t-1} \\ &= \lambda[y_t - bx_{t-1} + (1-\lambda)(y_{t-1} - bx_{t-2}) + (1-\lambda)^2(y_{t-2} - bx_{t-3}) + \ldots] \end{aligned} \tag{2.3}$$

is an estimate of the offset, and is computed recursively based on the EWMA statistic with the last measurement data. The previous estimate a_{t-1}, and λ are the controller parameters which can be adjusted to achieve a desired output. As λ approaches 1, more weight is given to the most recent observations. As λ decreases, more weight is given to older observations, and in the limit when $\lambda = 0$, all the $a_t's$ equal to a_0.

Substituting Equation (2.3) into Equation (2.2) lead to

$$x_t = -\frac{\lambda}{b}\sum_{j=-\infty}^{t} y_j \tag{2.4}$$

Therefore, the single EWMA controller is a pure integral (I) controller with integral constant $K_I = -\lambda/b$, which is a particular case of the well-known PID controller. A discrete PID form can be expressed as follows

$$x_t = -K_P y_t - K_I \sum_{j=0}^{\infty} y_{t-j} - K_D (y_t - y_{t-1}) \tag{2.5}$$

where K_P, K_I and K_D are the tuning parameters of the controller. From the PID controller, the integral action can eliminate offsets or shifts, and provides robustness in the controlled process.

Box and Jenkins (1976) showed that the single EWMA statistic is a minimum mean square error (MMSE) controller when the integrated moving average (IMA(1,1)) disturbance model exists in the process. The IMA(1,1) is an important non-stationary time series model in the study of process regulation and adjustment (Box and Luceño, 1997). Other than the IMA(1,1), the single EWMA controller has been shown to perform effectively in some disturbance models such as the step and ramp with slow drift rate disturbance model.

2.2.2 Effects of Incorrectly Setting the EWMA Parameter

As mentioned by Smith and Boning (1997), the controlled process output under a higher EWMA weight would return to the target much faster than a lower weight, but it would create more oscillations. Therefore, it is important to select the EWMA parameter carefully. In this section, the effect of incorrectly choosing the EWMA gain will be described in more detail.

Consider a process that can be modeled by

$$y_t = \alpha + \beta x_{t-1} + N_t \tag{2.6}$$

where N_t denotes the disturbance model we want to compensate for. Assume it follows an IMA(1,1) stochastic process as follows:

$$(1-B)N_t = (1-\theta B)\varepsilon_t \tag{2.7}$$

where θ is the moving average parameter. An IMA(1,1) model is widely used for modeling the drift in the discrete manufacturing (Box *et al.* 1994, Lucas and Saccucci 1992). If we use the single EWMA controller to compensate for the disturbance, then the controlled process is as follows:

$$(1-(1-\lambda\xi)B)y_t = (1-\theta B)\varepsilon_t \tag{2.8}$$

We can see that the controlled process exhibits an ARMA(1,1) process, and that the stable condition is $|1-\lambda\xi| \le 1$. Therefore, the inflation factor of the controlled process will be:

$$\sigma_y^2 / \sigma_\varepsilon^2 = 1 + \frac{(1-\lambda\xi-\theta)^2}{1-(1-\lambda\xi)^2} \tag{2.9}$$

Figure 2.1 shows the inflation factor ($\sigma_y^2 / \sigma_\varepsilon^2$) versus λ and θ given the process gain is known ($\xi = 1$). Consider that the disturbance model follows a white noise process, which implies $\theta = 1$ in Equation (2.8) and we use a full adjustment to the process ($\lambda = 1$), then the controlled output variance will be inflated twice as much than if there was no adjustment. This is what Dr. Deming (1986) meant by "tampering with the process". In order to achieve the minimum mean square error (MMSE) controlled process output, we should set the controller parameter to be as follows:

$$\lambda^* = 1-\theta \tag{2.10}$$

Although we already knew the actual optimal controller parameter in the above equation, it was only an optimal value in the static sense. In practice, the parameter of the disturbance model changes with time, thus it is necessary to develop an approach to adjust the controller parameter dynamically in order to obtain a better performance of the controlled process output. The following section will introduce an adaptive algorithm that was recently suggested by Patel and Jenkins (2000).

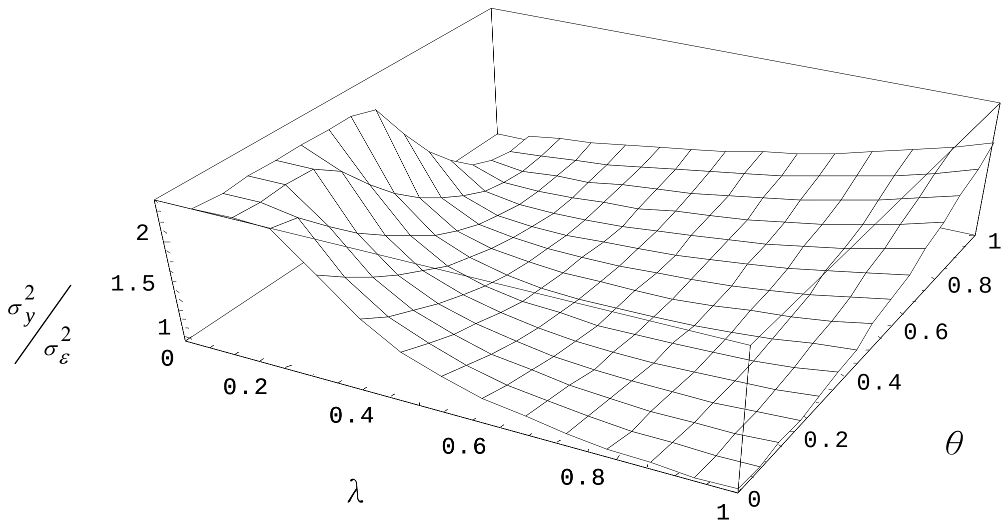

Figure 2.1 Inflation factor versus λ and θ

2.2.3 Patel-Jenkins Adaptive Algorithm

In the sense of an adaptive system, Satri (1988) used the theory of least squares estimation (LSE) for sequential parameter-detection and revision of the moving average parameter in the IMA(1,1) time series model. Luceño (1995) presented a computer program to choose the EWMA controller parameter in the EPC. His algorithm was based on the maximum likelihood estimate (MLE) theory. These above mentioned algorithms all have a common constraint in that the probability distribution must be known beforehand. Therefore, Smith and Boning (1997) utilized a neural network as an approximation function to map from the disturbance state (magnitude of linear drift and random noise) of a given process to the corresponding optimal EWMA weights. However, it was a problem to estimate the slope of the controlled process. Del Castillo and Yeh (1998) presented an adaptive run-to-run multiple-input-multiple-output controller for linear and nonlinear semiconductor processes. Recently, an adaptive algorithm to estimate the EWMA gain was suggested by Patel and Jenkins (2000). Their objective was to design an automated scheme for optimizing the numerical parameter of the EWMA controller. Figure 2.2 shows the adaptive EWMA controller block diagram they proposed.

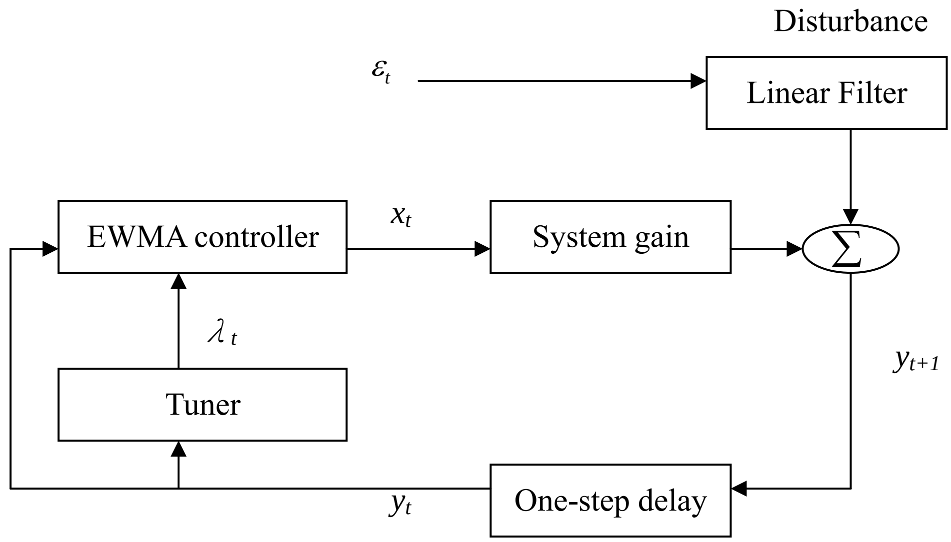

Figure 2.2 Structure of adaptive tuning controller

The Patel-Jenkins adaptive algorithm can simply be described as the following system:

$$\mu_{t+1} = \mu_t + \tau_t(e_{t+1} - \mu_t)$$

$$\zeta_{t+1} = \zeta_t + \tau_t(e_{t+1}^2 - \zeta_t)$$

$$\lambda_t = \frac{\delta^2 + 4\mu_t^2}{\delta + \mu_t^2 + \zeta_t} \quad (2.11)$$

where $\{\mu_t\}$ are the estimates of the mean of the output, and $\{\zeta_t\}$ are the estimates of the mean square value of the output. Initial conditions of (μ_0, ζ_0) satisfy $0 \le \mu_0^2 \le \zeta_0$. δ is a constant with a very small value which satisfies $0 < \delta < 1$, and $\{\tau_t\}$ is a sequence such that $0 \le \tau_t < 1$ and satisfies (1) $\lim_{t\to\infty} \tau_t = 0$, (2) $\sum_{t=0}^{\infty} \tau_t = \infty$, (3) $\sum_{t=0}^{\infty} \tau_t^2 < \infty$. The form of λ_t in Equation (11) intuitively provides a measure of the signal-to-noise (SN) ratio that satisfies $0 \le \lambda_t \le 2$. According to the above mentioned adaptive system, the EWMA control equation can be updated dynamically as follows:

$$x_t = -\frac{\lambda_t}{b}\sum_{j=-\infty}^{t} y_j \tag{2.12}$$

where the adaptive parameter λ_t follows the system in Equation (2.11).

2.3 Double EWMA Controller

2.3.1 Brief Review

Due to the single EWMA controller can not compensate for the wear-out process, that is a considerable offset will be produced. Butler and Stefani (1994) extended the single EWMA controller with another EWMA equation in order to compensate for the ramp disturbance model. In this section, we will briefly introduce the double EWMA controller that was proposed by Butler and Stefani (1994). Note that they do not call it a double EWMA controller, but refer to it as a predictor corrector control (PCC) scheme.

Consider a drifting process model as follows:

$$y_t = \alpha + \beta x_{t-1} + \delta t + \varepsilon_t \tag{2.13}$$

where δ denotes the drifting speed. A PCC control equation for x_t can be expressed as:

$$x_t = \frac{-c_t - D_t}{b} \tag{2.14}$$

where b is the estimate of β which can be obtained off-line by using designed experiments in a pre-control phase. c_t and D_t can be expressed as follows:

$$c_t = \lambda_1(y_t - bx_{t-1}) + (1-\lambda_1)c_{t-1}; \quad 0 < \lambda_1 \le 1 \tag{2.15}$$

$$D_t = \lambda_2(y_t - bx_{t-1} - c_{t-1}) + (1-\lambda_2)D_{t-1}; \quad 0 < \lambda_2 \le 1 \tag{2.16}$$

where λ_1 and λ_2 are the weights for the first and second EWMA equations. Note that if we set $\lambda_1 = 0$ or $\lambda_2 = 0$, then the double EWMA controller will reduce to a single-EWMA controller. From Equations (2.14) and (2.15), it is clear that the performance of a double EWMA controller depends on selecting both parameters of λ_1 and λ_2. To appropriately select both parameters, the stability conditions of controller parameters should be held.

By substituting Equation (2.14) into Equation (2.13), we obtain the closed-loop description of the output given be

$$y_t = \alpha - \xi c_{t-1} - \xi D_{t-1} + \delta t + \varepsilon_t \tag{2.17}$$

where $\xi = \beta / b$, is the bias in the gain estimate. Substituting this into Equations (2.15) and (2.16), we get, respectively,

$$c_t = \lambda_1(\alpha\delta t + \varepsilon_t) + (1 - \lambda_1\xi)c_{t-1} + \lambda_1(1-\xi)D_{t-1} \tag{2.18}$$

$$D_t = \lambda_2(\alpha + \delta t + \varepsilon_t) - \lambda_2\xi c_{t-1} + (1-\lambda_2\xi)D_{t-1} \tag{2.19}$$

To analyze the stability conditions of the system, define the state vector $E_t^{'} = (c_{t-1}, D_{t-1}, t)$, where the apostrophe means transpose. With this setting we have the state-space representation

$$\begin{aligned} E_{t+1} &= AE_t + G_t \\ y_t &= F'E_t + R_t \end{aligned} \tag{2.20}$$

where

$$A = \begin{pmatrix} 1-\lambda_1\xi\lambda_1 & 1-\xi & \lambda_1\delta \\ -\lambda_2\xi & 1-\lambda_2\xi & \lambda_2\delta \\ 0 & 0 & 1 \end{pmatrix} \qquad w_t = \begin{pmatrix} \lambda_1(\alpha+\varepsilon_{t+1}) \\ \lambda_2(\alpha+\varepsilon_{t+1}) \\ 1 \end{pmatrix}$$

$F' = (-\xi, -\xi, 0)$, and $R_t = \alpha + \delta t + \varepsilon_t$. To solve the state equation with

$$E_t = A^{t-k_0}E_{k_0} + \sum_{j=k_0}^{t-1} A^{t-j-1}G_j \tag{2.21}$$

where k_0 is the point in time where initial condition (E_{k_0}) are known. One way of computing A^t is to write $A^t = P\Gamma^t P^{-1}$, we get

$$\Gamma = \begin{pmatrix} 1-0.5\xi(\lambda_1+\lambda_2)+0.5e & 0 & 0 \\ 0 & 1-0.5\xi(\lambda_1+\lambda_2)-0.5e & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.22}$$

where $e=\sqrt{\xi(\xi(\lambda_1+\lambda_2)^2-4\lambda_1-\lambda_2}=e_1-e_2$, and e_1 and e_2 are the first two eigenvalues of A. Thus, the process will be asymptotically stable if and only if

$$\left|1-\frac{1}{2}\xi(\lambda_1+\lambda_2)+\frac{1}{2}\sqrt{\xi\left[\xi\left(\lambda_1+\lambda_2\right)^2-4\lambda_1\lambda_2\right]}\right|<1$$
$$\left|1-\frac{1}{2}\xi(\lambda_1+\lambda_2)-\frac{1}{2}\sqrt{\xi\left[\xi\left(\lambda_1+\lambda_2\right)^2-4\lambda_1\lambda_2\right]}\right|<1 \qquad (2.23)$$

Note that if $\lambda_2=0$, the stability conditions reduce to $|1-\lambda_1\xi|<1$, the condition for stability in a single EWMA controller.

From Equations (2.20) and (2.21), the expected value of the output is:

$$E[y_t]=E[F'\sum_{j=1}^{t-1}A^{t-j-1}G_j]+\alpha+\delta t \qquad (2.24)$$

assuming $E_1^{'}=(0,0,1)$, so we have $k_0=1$. It can be shown, after considerable algebraic manipulation, that

$$E[y_t]=T+\frac{\delta(e_1^{t-1}-e_2^{t-1})}{e_1-e_2}+\frac{\alpha}{e_1-e_2}(e_1^{t-1}(e_1-1)+e_2^{t-1}(1-e_2)) \qquad (2.25)$$

Given $\lambda_2=0$, Equation (2.25) gives $\lim_{t\to\infty}E[y_t]=\delta/\lambda_1\xi$, the bias incurred by the single EWMA controller applied to a system that drifts.

If $\xi=1$, it follows from Equation (2.17), (2.20) and (2.21) that as time approaches infinity, each c_t and D_t works as follows:

$$\lim_{t\to\infty}E[c_t]\to\alpha+\delta(t+1)-\delta/\lambda_1 \qquad (2.26)$$

$$\lim_{t\to\infty}E[D_t]\to\delta/\lambda_1 \qquad (2.27)$$

we can see that c_t is an asymptotical estimate of the ramp disturbance with a bias term (δ/λ_1), and D_t is the asymptotical estimate of the bias term. Thus, c_t+D_t becomes an asymptotically unbiased one-step-ahead estimate of the ramp disturbance model. The double EWMA formula can be further rewritten as

$$c_t = \lambda_1 \left[\frac{1-(1-\lambda_2)B}{1-2B+B^2} \right] y_t \tag{2.28}$$

$$D_t = \lambda_2 \left(\frac{1}{1-B} \right) y_t \tag{2.29}$$

where B is the backward shift operator ($By_t = y_{t-1}$). We can see that the c_t term is a second order filter ($\lambda_2 \neq 0$) and can filter out the trend. The term, D_t, is the first order filter that can filter out the process offsets. Therefore, the PCC control equations can be expressed as follows:

$$\begin{aligned} c_t + D_t &= \lambda_1 \left[\frac{1-(1-\lambda_2)B}{1-2B+B^2} \right] y_t + \lambda_2 \left(\frac{1}{1-B} \right) y_t \\ &= w_1 \sum_{i=1}^{t} y_i + w_2 \sum_{i=1}^{t} \sum_{j=i}^{t} y_j \end{aligned} \tag{2.30}$$

where, $w_1 = \lambda_1 + \lambda_2 - \lambda_1 \lambda_2$ and $w_2 = \lambda_1 \lambda_2$.

From Equation (2.30), we can see that the double EWMA controller is not a discrete PID form, but an integral-double-integral (Ⅰ-Ⅱ) form (Chen and Guo, 2001). It can be shown that the Ⅰ-Ⅱ controller is an MMSE controller when the IMA(2,2) disturbance model affects the process. Compared to the non-stationary IMA(1,1) process (i.e. drift in any direction with equal probability), the IMA(2,2) noise model can be interpreted as a process that experiences random changes in the slope coefficient in Equation (2.13) (Del Castillo, 1999). Figure 2.3 shows a block diagram of the double EWMA controller model when a ramp disturbance model exists in the process.

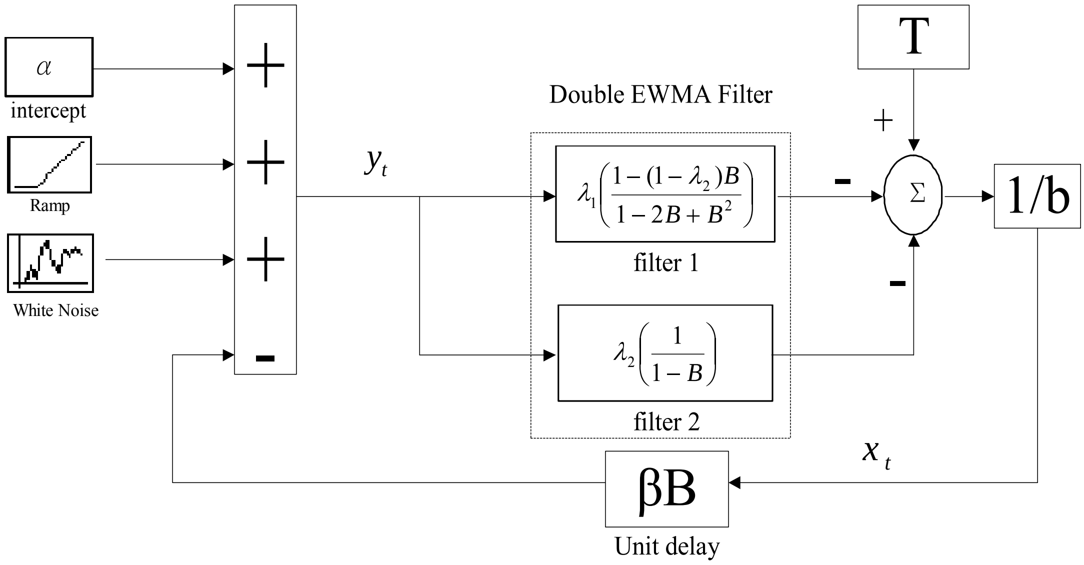

Figure 2.3 Double EWMA controller

2.3.2 Trade-Off Tuning Strategy

The control strategies of EWMA controllers can simply be divided into time-invariant and time-varying weights control schemes. The time-invariant control scheme means that the EWMA weights do not change with time, but that the weights are fixed to control the process. Del Castillo (2001) presented a solution of balancing the adjustment and output variances to control the single-EWMA controller. He also designed a trade-off solution of transient and steady-state performance to control the double-EWMA controller (1999). The time-varying control scheme is sometimes called a self-tuning or adaptive control, because the EWMA gains change with time. Smith and Boning (1997) used the neural technique to self-tune the EWMA controller. Del Castillo and Hurwitz (1997) used the recursive least squares (RLS) theory to continuously estimate the process parameters. Patel and Jenkins (2000) proposed an adaptive EWMA control algorithm by taking the signal-to-noise ratio (SNR) into consideration. The above mentioned adaptive algorithms all have one point in common, they self-tune the single-EWMA controller, but not the double-EWMA controller. Up till now, when it came to the topic of the double-EWMA control scheme, only Del Castillo (1999) has presented an optimization form for solving the double-EWMA gains. This algorithm is introduced below (assume $\xi = 1$).

$$
\begin{aligned}
AVAR(y_t) &= \lim_{t\to\infty} Var(y_t) \\
&= \sigma_\varepsilon^2\left[1 + \frac{1}{(\lambda_1-\lambda_2)^2}\left(\frac{\lambda_1\lambda_2^2 + \lambda_1(\lambda_1-\lambda_2)^2}{2-\lambda_1} + \frac{\lambda_1^2\lambda_2 + \lambda_2(\lambda_1-\lambda_2)^2}{2-\lambda_2}\right)\right] \qquad (2.31)
\end{aligned}
$$

where σ_ε^2 represents the variance of the white noise term. The transient effect is measured by averaging the mean square deviation up to run *m*, and can be expressed as follows:

$$
\begin{aligned}
\overline{MSD} &= \frac{1}{m}\sum_{t=1}^{m} E(y_t)^2 \\
&= \frac{1}{m(\lambda_1-\lambda_2)^2}\left\{\frac{(\delta-\alpha\lambda_2)^2[1-(1-\lambda_2)^{2(m+1)}]}{1-(1-\lambda_2)^2}\right. \\
&\quad + \frac{2(\delta-\alpha\lambda_2)(\alpha\lambda_1-\delta)[1-(1-\lambda_2)^{m+1}(1-\lambda_1)^{m+1}]}{1-(1-\lambda_1)(1-\lambda_2)} \\
&\quad \left. + \frac{(\delta-\alpha\lambda_1)^2[1-(1-\lambda_1)^{2(m+1)}]}{1-(1-\lambda_1)^2}\right\} \qquad (2.32)
\end{aligned}
$$

Therefore, the optimization form can be modeled as follows:

$$
\begin{aligned}
&\min_{\lambda_1,\lambda_2} \quad k_1 AVAR(y_t) + k_2\overline{MSD} \\
&S.T \quad 0 < \lambda_1 \le 1 \qquad (2.33) \\
&\qquad\ \ 0 < \lambda_2 \le 1
\end{aligned}
$$

where the parameters (k_1, k_2) are determined by the engineers. For $(k_1, k_2) = (0,1)$, it is an all-bias solution. For $(k_1, k_2) = (1,0)$, it becomes an all-variance solution. If we set $k_1 = k_2 = 1$, then a trade-off solution will be obtained. Del Castillo (1999) suggested that keeping the trade-off solution weights to control the process would provide adequate performance than a variety case. For example: if we first use the all-bias solution to cancel out the transient effect, and then abruptly change it to the all-variance solution at a specific time, then the controlled process will incur a new transient at that specific time. Figure 2.4 shows the above condition in which, we simulate the process with a drift rate of 0.5. By solving Equation (2.33), we obtain the all-bias solution (0.1811, 0.7917) and the all-variance solution (0.0173, 0.1067). Suppose we change the control scheme at run 100. Clearly, a new transient occurs at run 100.

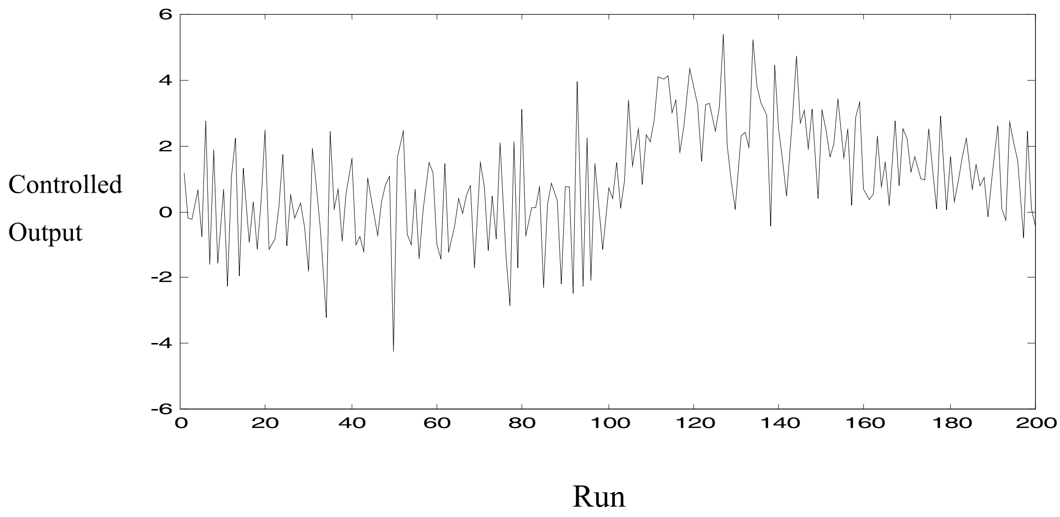

Figure 2.4 Change control scheme at run 100

2.4 Neural Network Techniques

Neural networks are of particular interest because they offer a means of efficiently modeling large and complex problems in which there may be hundreds of predictor variables that have many interactions. Neural nets have been used widely in pattern recognition (Su *et al.* 2002), function approximation (Smith and Boning 1997), optimization (Sjoberg and Agarwal 2002), and data clustering (Andrews and Geva 2002). In general, neural networks can be classified into two different categories: feed-forward and feedback networks (Cheng and Titterington 1994). In this study, we utilized the feed-forward network because it has been found to be an effective system for learning distinguishing patterns from a body of examples.

The back-propagation learning algorithm is the most commonly used algorithm to train multilayer feed-forward networks by implementing a local gradient-search to minimize the square error between realized and desired outputs. A typical back-propagation neural network always has an input layer, an output layer and at least one hidden layer. There is no theoretical limit on the number of hidden layers, but typically there will be one or two. Figure 2.5 shows a three layers network. Each layer is fully connected to the succeeding layer. The back-propagation algorithm involves a forward pass 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 difference between the desired and actual network outputs. The above statement can be explained by the following mathematical equations:

Forward pass:

The net input to node i for pattern p is

$$net_{pi} = \sum_j w_{ij} g_{pj} + v_i \tag{2.34}$$

$$g_{pj} = \frac{1}{1 + e^{-net_{pj}}} \tag{2.35}$$

where w_{ij} is the weight from unit j to unit i, v_i is a basis associated with unit i, and g_{pj} is the activation value of unit j with sigmoid function for pattern p.

Backward pass

The sum of the squares error function is as follows:

$$\boldsymbol{E}_p = \frac{1}{2}\left\| \boldsymbol{t}_p - \boldsymbol{o}_p \right\|_2^2 \tag{2.36}$$

where $\boldsymbol{t}_p$ is the target output for the pth pattern and $\boldsymbol{o}_p$ is the actual output for the pth pattern. By minimizing the errors E_p using the gradient decent method, the weights can be updated using the following equation:

$$\Delta_p w_{ij} = \eta r_{pi} g_{pj} \tag{2.37}$$

where

$$r_{pi} = \begin{cases} (t_{pi} - o_{pi}) o_{pi} (1 - o_{pi}) & \text{if unit } i \text{ is an output unit} \\ o_{pi} (1 - o_{pi}) \left(\sum_k r_{pk} w_{ki} \right) & \text{if unit } i \text{ is a hidden unit} \end{cases} \tag{2.38}$$

and η is the learning rate. In general, a larger learning rate will increase the training speed, however it may oscillate widely. One way to increase the learning rate without oscillating is to modify Equation (2.37) to the following equation:

$$\Delta_p w_{ij} = \eta r_{pi} g_{pj} + u \Delta_{p-1} w_{ij} \tag{2.39}$$

where u is the momentum coefficient ($u \in [0,1]$) that determines the effect of past weight changes on the current direction of movement in weight space. There is no principle to determine the parameters of η and u; they are chosen by the neural network trainer via the trial and error approach. Concerning the model selection, one of the most useful methods in selecting problems is the cross-validation (CV) method. Breiman and Spector (1992) found 10-fold and 5-fold cross-validation to work better than leave-one-out method for choosing subsets of inputs in linear regression. Zhang (1993) showed that the delete-*d* multifold cross-validation (MVC) criterion is asymptotically equivalent to the well known FPE criterion under a regression model. Detail discussions about CV method can be found in Witten and Frank (2001).

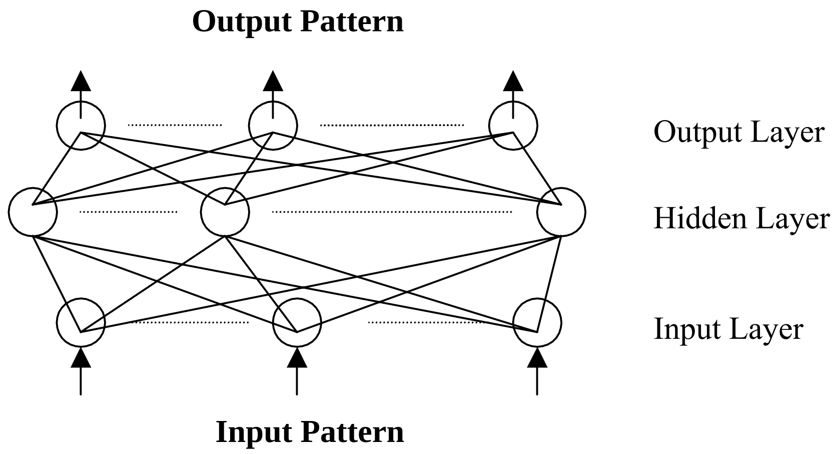

Figure 2.5 The backpropagation neural network structure