The Run-to-Run Control Using Adjustment Limit Scheme for SISO Process Quality Control

Chien-Chih Wang , Cheng-Tung Chang

Department of Industrial Engineering and Management
Mingchi University of Technology
Taishan, Taiwan 243

Abstract: In the past few years, Run-to-Run (R2R) control techniques have been developed in process control. The R2R control methods integrate statistical process control (SPC) and engineering process control (EPC) techniques in semiconductor manufacturing processes. In traditional R2R control scheme, the process controller provides the feedback control for process output so that each run can be close to the target. However, each run needs to control or to adjust a process would result in some unnecessary control cost when the process is stable. In other words, it isn't the best control method for a process. In this paper, we combines the upper control limit (UCL) and lower control limit (LCL) lines in EWMA (exponentially weighted moving average) control chart as the control criterion for adjusting threshold to determine whether to start a R2R process controller. When the process output goes beyond the threshold, the controller starts and revises the next run of process to compensate the deviation from target. The result of this research showed that when the autoregressive parameter $\phi > 0.7$, the MSE using self-tuning control plus appropriate adjustment limit is less than that of EWMA control plus adjustment limit. When the autoregressive parameter $\phi \leq 0.7$, there are not significant differences of MSE between self-tuning control plus adjustment limit and EWMA controller plus adjustment limit, but EWMA control plus adjustment limit has the advantage of easier calculation, so it would be more effective. The developed control scheme differs from the traditional R2R is not to adjust for every run in process. Then it decreases the frequency of adjustment in process. When the quality characteristic lose slightly, making the stable process and decreases the control cost in producing processes by reducing the adjust frequency.

1. INTRODUCTION

The production process is very complicated for the semiconductor industry, because continuous production is difficult to control. The common control method used is run-to-run (R2R), However, the EWMA algorithm is often preferred to R2R, but is unable to achieve the goal of controlling production effectively when there is a high dependence between each process. In addition, the traditional R2R control method involves making adjustments to parameters on every run, but it's not necessary to do this. We plan to use self-tuning to be replace EWMA under the single-input-single-output (SISO) condition and to join the adjustment limits machine. We hope to use adjustment limits to judge when the algorithm needs to start and make current R2R control methods effective with extensive use in SISO conditions.

The quality control method most used in traditional manufacturing processes is statistical process control (SPC). It mostly uses off-line quality improvement, but in processes like continuous semiconductor production, it is unable to make on-line adjustments and cannot prevent critical situations. In 1980, some scholars proposed combining engineering process control (EPC) with SPC. The effective combination of these two methods relies on mature measuring and control technology. Though there are advantages in the merger of SPC and EPC, only assignable causes can be quickly developed and removed. But a lot of situations are not like this, and it is difficult to find the assignable causes and remove them, especially in manufacturing situations. Integrating SPC and EPC could not answer the demands of actual processes. Ingolfsson and Sachs (1993) extended the concept of combining SPC and EPC and proposed the R2R control method to the silicon wafers produces processes. The result shows that applying the R2R control method is beneficial to processing. Furthermore, some scholars suggest R2R in semiconductor manufacturing processes.

R2R control method is based on integration of SPC and EPC, it make adjusts equipment parameters to achieve the goal of control. Generally, the design of R2R controllers follows these steps:

1. The use of design of experiment (DOE) to construct initial model, and use this initial model to calculate an optimum control foundation.
2. The R2R Controller makes parameter changes after each run.

The traditional structure of an R2R controller can be seen in Fig. 1.

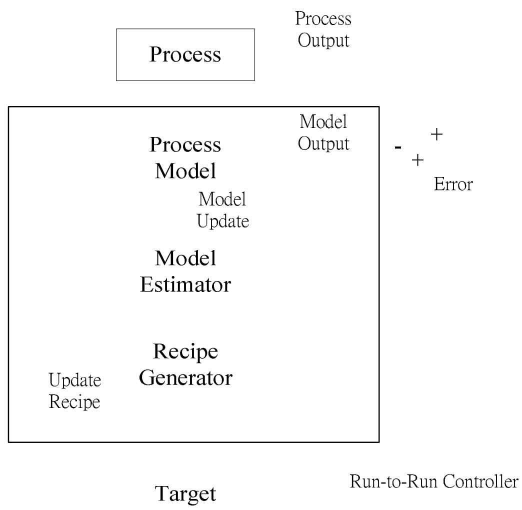

Figure 1. The traditional R2R controller (*C. Zhang et al.* 2003)

Deming (1982) used funnel experiment theories to prove that excessive control takes place in a stable system. Although this funnel experiment was a special case, the idea of using models to control and adjust is easy to understand. McGregor (1990) imitated Deming to show the necessity of processed control and the negative effects of excessive control.

In the R2R control method, processes are not stable during the initial stage, but gradually steady after some adjustments, although too much control is not required. Luo L.C. (2004) suggests adding EWMA algorithm to the R2R controller to adjust the process parameters. The result shows that when the variation of response is small, it not only steadily controls the situation, but needs fewer adjustments and reduces production costs. But if the processes have a high correlation ($\phi \geq 0.7$), the result would not be good enough. In addition, Ren J.H. (2000) says that when there is a high correlation between processes, using EWMA would not be good as self-tuning to control.

We used the self-tuning algorithm for adjustment limits within the R2R controller to explore it when the processes had a high correlation to set time adjustment limits. We expected to be able to use this procedure to reduce process parameter adjustment. After we comparing EWMA and self-tuning algorithms and combining them with adjustment limits each other to control, we found that the self-tuning algorithm made the adjustment limits in the R2R controller, regardless of how the correction between processes. This method could efficiently reduce the number of adjust process setting and not increase the variance of responses.

2. METHODOLOGY

The EWMA controller is one of the feedback control methods used by Box and Jenkins (1974). The controller used the value of real and predicted response feedback to adjust process control parameters and made the process response more close to the goal next run. The EWMA controller in SISO process is:

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

Where y_t is the quality characteristic of the process output, α is the process offset and β is the first-order term of controllable variable μ_t. In addition, ε_t in this hypothetical model conformed to normal distribution $\left(\varepsilon_t \sim N\left(0, \sigma^2\right)\right)$, but in practice, ε_t usually conformed to other distributions, like uniform distribution.

Formulate (1) could also be presented like this:

$$\hat{y}_t = a_{t-1} + bu_{t-1} \tag{2}$$

Where y_t is the quality characteristic of the process output, α is the process offset and β is the first-order term of controllable variable μ_t. But the EWMA controller would only estimate and replace the interceptor. The coefficient of b is determined by experiment prior to the actual process.

When formulate (2) was used to EWMA control, calculated error e_t needed to be between the real and estimated output:

$$e_t = y_t - \hat{y}_t \tag{3}$$

So, the revised formulate of EWMA is as follow:

$$a_t = \lambda \cdot e_t + a_{t-1} \tag{4}$$

Where λ is the predicted weight between 0 and 1. If (2) and (3) are combined into formulate (4), the new formulate is:

$$a_t = \lambda \cdot (y_t - bu_t) + (1-\lambda) \cdot a_{t-1} \tag{5}$$

The EWMA controller used formulate (5) to determine the estimated output, which is continually revised. So, the adjusting parameter was like:

$$u_t = \frac{T - a_{t-1}}{b} \tag{6}$$

Where u_t is the input at time t, T is object value, and other parameters are defined as in formulate (2).

The self-tuning controller is an on-line estimating method. Del Castillo (1996) has already regarded this as a kind of feasible R2R control method in a SISO system. The SISO model is called the contracted MIMO model, and can calculate the trend of the responses, and compensate for the errors to the target value. The model is described as follows:

$$y_t = \phi y_{t-1} + \alpha + \beta\mu_{t-1} + dt + N_t \tag{7}$$

Where y_t is the quality characteristic of the process output, α is the process offset and β is the first-order term of controllable variable μ_t. D_t is the disturbance term. If the process has a deterministic trend, then $D_t = dt + N_t$. Where d is the deterministic drifting rate per hour, N_t is the colored noise term in the form of an ARIMA model. If N_t follows ARIMA (1,0,1), then $N_t = \frac{1-cz^{-1}}{1-wz^{-1}}\varepsilon_t$. Where c is the first-order moving average (MA) coefficient and w is the autoregressive (AR) coefficient. The noise term ε_t is assumed to be white noise, it should conform to normal or uniform distribution. We used S-plus software to simulate the four different control methods. First, two simulations were used: EWMA and Self-tuning algorithms to control the process. These part was that traditional R2R control, that adjust parameters on each run. The other two methods used joined the adjustment limits, and only adjusted the parameters when the process response was over or less than the parameters.

In this article, adjustment limits were set at the constant value, and this value was calculated by the target value and the receivable variance of the process. In practice, the target value and the receivable variance of the process can be set according to historical data.

$$(UAL, LAL) = (T + k\hat{\sigma}, T - k\hat{\sigma}) \tag{8}$$

UAL was the upper adjustment limit and LAL was the lower adjustment limit. The target value T was set at 0 in this article, and k was a receivable variance greater than 0. If k was set near to 0, the limits' width would be narrow and vice verse. The adjustment limits' width was affected the variance of responses. We set k to eight different setting to make comparison. The estimator of process variance was $\hat{\sigma}$. In formulate (7), the source of variance only was ε_t, so $\hat{\sigma}$ was ε_t's variance in this article $(\hat{\sigma} = 1)$.

The aim of this simulation was to find the influence of the process and the process's response when joined adjustment limits into R2R control. But how to set the adjustment limits? We used two different valuation indices (AAI & MSE) to evaluate the result of the joined adjustment limits in the R2R control.

1. Average Adjustment Interval (AAI):

AAI is the number of runs between parameter adjustments. For example, in a 100-run process we adjusted parameters 10 times, so the AAI of this control was 10. If the AAI became smaller, that meant the process parameter needed to be readjusted more often.

2. Mean square error (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_t)^2 \tag{9}$$

Where y_i was the actual response, y_t was the target and *n* was the number of runs. If the MSE was little, this meant the variance of response was small. In the other words, the process was stable and better controlled.

After we joined adjustment limits, the R2R controller would not adjust the process parameter on each run, so the variance of response often became bigger. In order to check if this method made a variance increase or not, we used the F-test to verify the MSE of a traditional R2R control and joined the adjustment limits to the R2R control.

$$F = \frac{MSE_1}{MSE_2} \tag{10}$$

We use F value to determine the influence of the response's variance in different control methods and parameter settings.

Adjustment limit settings affect production costs and quality. So, we used two indexes (AAI & MSE) to find reasonable adjustment limits, hoping to reduce adjustment times, further and manufacturing costs.

3. ANLYSIS AND RESULT

All figures should be positioned at the top of the page when possible. All figures should be numbered consecutively and captioned; the caption should be centered under the figure as shown in Figure 1. All text within the figure should be no smaller than 9pt. There should be a minimum of two line spaces between figures and text.

We used the EWMA algorithm, the self-tuning algorithm, the EWMA algorithm with adjustment limits, and the self-tuning algorithm with adjustment limits of an R2R controller. We simulated 200 different processes to correlate data in these four situations using the computer program s-plus. We set the environmental parameters at $\alpha = 2.0$; $\beta = 2.0$; $c = 0.5$; $w = 0.2$; $T = 0$; $\lambda = 0.1$, and the correlation degrees were $\phi = 0.01$; 0.05; 0.1; 0.3; 0.5; 0.7; 0.9; 0.95; 0.99. We set the width of the adjustment lines were ± 0.02; ± 0.025; ± 0.05; ± 0.1; ± 0.3; ± 0.5; ± 0.7; ± 1, and the initial parameters of self-tuning were $(\phi, \beta, d) = (1, -1, -1)$. All these simulation were set separately.

In table 1, the responses' MSE became bigger with an increase in the process correlation degree using the EWMA algorithm, and showed better control in the lower process correlation. After drop the unstable part during the initial stage, the self-tuning control had smaller MSEs than the EWMA in the higher correlation degrees ($\phi \geq 0.7$). But, the EWMA was better of calculating simple and smaller MSEs in the lower correlation degrees ($\phi < 0.7$). In table 2, we used the EWMA algorithm to combine the adjustment limits in the processes which had higher correlations, The responses' MSE were increased appreciably than the lower correlations. Even if we used EWMA algorithm to combine adjustment limits to the responses' MSE became bigger in the higher correlation. But, if we used the self-tuning combine adjustment limits, the high and low correlations were steady. (adjust demarcation line $\leq \pm 0.5$).

We found that joining adjustment limits in the EWMA algorithm could reduce the number of adjustment times, but not make responses' MSE increase (see table 1 and table 2). However, if the process correlations were high (ϕ =0.7), we not only used EWMA algorithm, but also the combined it with the adjustment limits. The unstable response made the MSE high. In addition, when we used the self-tuning controller, there was an unstable phenomenon in the initial responses, but several rows later the responses converged to a stable state. We used the self-tuning controller after eliminating the unstable responses to get the same result when using the EWMA controller if the process had lower corrections (ϕ =0.7). In other words, while processes had lower corrections, whether we used the EWMA algorithm or the self-tuning algorithm to control, we got a good control result. But, if the process had high corrections, we used the self-tuning control to get smaller responses' MSE. During this time, we used the self-tuning control to get better control results than used the EWMA control.

Although we used the self-tuning to combine the adjustment limits, we got the same control effect and reduced the number of times to adjusting the processes parameters. But, the adjustment limits could not be set up perfunctorily. Under this set of process parameters, we found the adjustment limits misfit over ± 0.1 when processes corrections were very low $(\phi < 0.1)$ or very high $(\phi > 0.7)$ in table 2. When if the limits were set wider, the number of times we adjusted would be less. According to the parameters set in our research, adjustment limits could be set to ± 0.05, which reduced the number of times we had to adjust parameters, but without increasing the responses' MSE. Table 3 compares these four kinds of controls.

Table 1. EWMA and Self-tuning controls in different correlation degrees

correlation degrees	EWMA algorithm		Self-tuning algorithm	
	$\bar{y}$	MSE	$\bar{y}$	MSE
ϕ =0.01	0.1008	0.4985	-0.0416	0.7016
ϕ =0.05	0.1031	0.5324	-0.0640	0.7282
ϕ =0.1	0.1056	0.4900	-0.0730	0.7850
ϕ =0.3	0.1070	0.4801	-0.0593	0.7239
ϕ =0.5	0.1093	0.6052	-0.0605	0.7297
ϕ =0.7	0.0965	0.8715	-0.0561	0.7043
ϕ =0.9	0.1292	1.4021	-0.0520	0.7255
ϕ =0.95	0.0929	3.5236	-0.0417	0.8307
ϕ =0.99	0.0755	10.8782	-0.0488	0.6588

Table 2. The MSE of EWMA and Self-tuning controls in different correlation degrees and adjustment limits

Adjustment lines / correlation		± 0.02	± 0.025	± 0.05	± 0.1	± 0.3	± 0.5	± 0.7	± 1
ϕ =0.01	E	0.444	0.494	0.483	0.507	0.481	0.418	0.49	0.432
	S	0.826	0.607	0.778	0.891	0.994	1.258	1.628	1.794
ϕ =0.05	E	0.449	0.467	0.496	0.472	0.493	0.524	0.553	0.536
	S	0.576	0.65	0.756	0.842	1.077	1.312	1.379	1.855
ϕ =0.1	E	0.481	0.46	0.499	0.474	0.498	0.509	0.53	0.489
	S	0.801	0.689	0.728	0.761	1.124	1.41	1.426	1.82
ϕ =0.3	E	0.559	0.449	0.479	0.526	0.477	0.498	0.525	0.488
	S	0.722	0.82	0.697	0.862	1.44	1.286	1.666	1.618
ϕ =0.5	E	0.549	0.586	0.598	0.583	0.601	0.543	0.598	0.557
	S	0.688	0.709	0.736	0.679	1.162	1.559	2.071	2.514
ϕ =0.7	E	0.742	0.713	0.758	0.713	0.783	0.785	0.755	0.789
	S	0.732	0.708	0.789	0.862	1.064	1.32	1.683	2.128
ϕ =0.9	E	1.569	2.086	1.935	1.743	1.361	2.777	1.65	2.272
	S	0.675	0.662	0.806	0.723	1.413	1.855	2.818	3.124
ϕ =0.95	E	4.039	3.526	4.357	2.457	5.058	3.152	4.468	6.36
	S	0.772	0.664	0.894	0.851	1.729	1.94	2.253	5.594
ϕ =0.99	E	23.564	30.111	12.836	11.899	18.19	13.456	16.179	10.498
	S	0.735	0.694	0.842	0.974	1.245	2.636	3.559	7.59

4. CONCLUSION

In our research, whether the processes' correlation were high or low, the adjustment limits were not bigger than ± 0.1. If adjustment limits were set too big, it would reduce the number of adjustments. But it would make the response's MSE bigger. The adjustment limits were not effective. Therefore, setting suitable adjustment limits reduced the number of adjustment, but did not increase the responses' MSE. When the adjustment limits started, the responses had a bigger MSE in the initial stage and needed to adjust each run. But after several runs, the processes became stable, and it was unnecessary to adjust each run. Users could choose several historical response values and the estimate response value of the next run to estimated the responses' MSE (choose the *t* to *t*-4 run response values and the t+1 run's $\hat{y}$ to estimate the MSE). When the MSE stabilizes (MSE≤ 1), it can start the adjustment limits mechanism, and adjust the process parameters in the next run when the calculation response is over the adjustment limit. If MSEs not stable, we adjusted the process parameters in each run, until the MSEs were acceptable.

Control methods	Strengths	Weaknesses	Applicable situation
EWMA controller	Although the responses' MSEs were similar to the MSEs in the self-tuning control in lower correlation processes ($\phi \leq 0.7$), the EWMA controller is easier to use.	The response's MSEs were too big (MSE>1.4) in the higher correlation processes ($\phi > 0.7$).	Use in the lower correlation ($\phi \leq 0.7$) processes.
EWMA controller combined with adjustment limits	In the lower correlation ($\phi \leq 0.7$) processes, it used adjustment limits to reduce the number of adjustment.	In the higher correlation processes ($\phi > 0.7$), it could not find a particular adjustment limit to reduce the number of adjustment but did not increase the responses' MSE.	Use in the lower correlation ($\phi \leq 0.7$) processes.
Self-tuning controller	It got a smaller responses' MSE than the EWMA control in the higher correlation ($\phi > 0.7$) processes. In other words, it stabilized.	Although it controlled well in the lower correlation processes, this method is complicated, and relatively inefficient.	Use in any correlation process, especially in the higher correlation ($\phi > 0.7$) processes.
Self-tuning controller combined with adjustment limits	1. If the adjustment limits were set up correctly, it could control better than the self-tuning controller in the higher correlation ($\phi > 0.7$) processes. 2. It could reduce the number of adjustment.	Although it controlled well in the lower correlation processes, this method is complicated, and relatively inefficient.	Use in any correlation process, especially in the higher correlation ($\phi > 0.7$) processes.

The self-tuning control combined with adjustment limits reduced the number of adjustment but did not increase the responses' MSEs. It satisfied the customer's requirements. The number of adjustments were reduced which lower production. In this application, the process engineer could use the EWMA or Self-tuning controller combined with adjustment limits to reduce the number of adjust process parameters in the R2R controller. In addition, the engineer could use the historic responses to appraise the start times of the adjustment limits to improve efficiency.

5. REFERENCES

1. Box, G. E. P., T. Kramer, "Statistical process monitoring and feedback adjustment- a discussion," *Technometrics*, 34(3), 251-267(1992).
2. Box, G. E. P., Coleman, D. E., and R. V., Baxley, "A comparison of statistical process control and engineering process control," *Journal of Quality Technology*, 29, 128-130(1997).
3. Del Castillo, "A multivariate self-tuning controller for run-to run process control under shift and trend disturbances," *IIE Transactions*, 28, 1011-1021(1996).
4. Del Castillo, A. M., Hurwitz, "Run-to-Run Process Control: Literature Review and Extension," *Journal of Quality Technology*, 29(2), 184-196(1997).
5. Deming, W. E., *Quality, Productivity and Competitive Position*, Center for Advanced Engineering Study, Massachusetts Institute of Technology, Cambridge, MA(1982).
6. Tsung, F., J., Shi, "Integrated design of run-to-run PID controller and SPC monitoring for process disturbance rejection," *IIE Transactions*, 31, 517-527(1999).
7. MacGregor, J. F., "A Different View of the Funnel Experiment," *Journal of Quality Technology*, 22(4), 255-259(1990).
8. Sachs, E., Hu, A. and A., Ingolfesson, "Run by Run process control: combining SPC and feedback control," *IEEE Transactions on Semiconductor Manufacturing*, 8(1), 26-43(1995).
9. Fan, S. K., Jiang, B. C., Jen, C. H., and C. C., Wang, **"**SISO run-to-run feedback controller using triple EWMA smoothing for semiconductor manufacturing processes" International Journal of Production Research, 40(13), 3093-3120(2002).
10. Zhang Chang, Deng Hao, and John S. Baras, "Run-to-run control methods based on the DHOBE algorithm," *Automatic*, 39, 35-45(2003).