CONTROL ENGINEERING LABORATORY

Some change detection and time-series forecasting algorithms for an electronics manufacturing process

Marko Paavola, Mika Ruusunen and Mika Pirttimaa

Report A No 26, March 2005
Some change detection and time-series forecasting algorithms for an electronics manufacturing process

Marko Paavola, Mika Ruusunen and Mika Pirttimaa

University of Oulu, Control Engineering Laboratory

In a sequential manufacturing process, a product unit proceeds through different manufacturing stages. At these stages, sensors monitor the features of the unit. In this work, the information produced by the sensors is employed to detect abrupt changes in process variables as well as to forecast their future behaviour. The developed algorithms were implemented as an on-line application to a manufacturing system.

A literature survey was performed to study the most common methods utilised in change detection and time-series forecasting. The most promising methods were selected on the basis of their on-line applicability and transferability to new manufacturing lines. These methods were further evaluated with off-line data. Finally, the difference method was applied for change detection and linear regression-based method for forecasting in this case.

During both on-line and off-line tests, some satisfactory results were attained. Real-time, on-line manufacturing environment set also its requirements for the applications. In the future, the possibility of combining expert knowledge with the aforementioned methods should be examined. The information thus received could be further utilised in the preventive maintenance and quality control.

Keywords: Electronics manufacturing, fuzzy, anomaly detection, modelling, prediction.
1 INTRODUCTION

Sequential manufacturing processes are usual in many areas of manufacturing industries /45/. A product unit proceeds through a sequence of manufacturing phases, in which features (for example holes, parts and materials) are added /45/. At each stage of the process, the sensors monitor the product and produce information about the process variables /45/. This information may be utilised to detect abrupt changes at different stages of the process as well as to forecast the behaviour of the variables.

The detection of abrupt changes in signals is a classical problem in signal processing, which can be used for example for event detection /8/. However, the current change detection methods tend to be quite sophisticated in nature. Additionally, a priori knowledge about the possibilities of the changes and their distributions is often required. This may make the implementation of these methods difficult as an automatic, on-line change detection application. In this report, some approaches to overcome these difficulties in change detection in an electronics manufacturing process are presented.

Although the need to forecast the behaviour of the process variables is obvious, examples of sequential, on-line applications are hard to find in the literature for the manufacturing processes. This is particularly true when the long-term trends are of concern. In this report, two examples of on-line, time-series approaches for forecasting are presented. The first one produces an estimate for a longer period, ranging from several days to even weeks. The other produces an accurate estimate of a process capability index for the next few hours.

In the following section, some of the most common methods utilised in change detection and time-series forecasting are first briefly discussed. Then, the approaches applied to the change detection problem and time-series forecasting are introduced. Next sections report the test environment and results. The report ends with some conclusions.
2 METHODS

2.1 Literature Survey

The generic problem of detecting abrupt changes in process parameters has been widely studied, see for example /31/ and /32/. These changes may be due to a shift in the mean value (edge detection) or to a variation in signal dynamics. Furthermore, research can concern with on-line (sequential) or off-line procedures /31/, /32/. In this report, the particular interest lies in detecting the alteration of the mean value sequentially. Figure 1 gives an example of such an abrupt change in a signal.

Reference /28/ presented an on-line change detection application for an artificial vision-based quality control system. Two popular edge detection methods, linear (differentiation) and a non-linear (see /24/) were tested. Although both methods performed satisfactory, the differentiation method was selected because of its flexibility and faster execution times.

In /36/, the cumulative sum (CUSUM) –technique was modified and employed to detect the changes in the process mean on-line. The modifications enabled the identification of the instant and duration of the anomaly. The method performed well with both simulated and real data.

An on-line Shiryayev Sequential Probability Ratio Test (SSPRT) was derived and implemented for a fault detection and isolation scheme of Advanced Vehicle Control Systems in /34/. The method used the Bayesian approach and therefore, required estimates of the a priori probabilities. The SSPRT was extremely sensitive to changes.

Reference /7/ presented a change detection method with learning capabilities. The algorithm utilised thresholds, which are adapted on-line to attain a certain error performance. The algorithm was supposed to be applicable to numerous uses, including industrial quality control.
In /42/, a neural network (NN) approach for on-line change detection was proposed. The anomaly detection utilised the changes in the autocorrelation (ACF) and partial autocorrelation functions (PACF) of certain time-series. A feedforward NN was first trained with a set of a typical ACF/PACF patterns. Then, for an on-line application, a moving window was utilised in defining the data set to be examined. If the examined data set reflected changes to the ACF and PACF, the network would detect them and suggest that a new parameter set should be estimated. In spite of successful results, the drawback of the approach was that the requirement for a stationary time-series had to be satisfied.

Time-series forecasting means making predictions on the basis of data consisting of one or more time-series. A time-series is a collection of observations collected sequentially in time /11/. Figure 2 gives an example of an actual and predicted time-series. Table 1 shows different approaches to forecasting with their objective /15/.

![Figure 2. An original time-series (solid line) and a prediction made of it (dashed line) /18/.

Figure 2. An original time-series (solid line) and a prediction made of it (dashed line) /18/.
Table 1. Different approaches to forecasting /15/.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecasting Series With No Trend</td>
<td>Expect for random variation, time-series remain essentially constant over time.</td>
</tr>
<tr>
<td>Forecasting Series With Trend</td>
<td>Time-series have long-term patterns of growth or decline.</td>
</tr>
<tr>
<td>Using Regression in Forecasting</td>
<td>Statistical methodology of regression analysis is applied to time-series forecasting.</td>
</tr>
<tr>
<td>Forecasting Seasonal Series</td>
<td>The average levels of time-series are affected by seasonal influences arising at regular intervals.</td>
</tr>
<tr>
<td>Forecasting Cyclical Series</td>
<td>Other than seasonal forces causes the level of time-series to oscillate above and below the trend.</td>
</tr>
<tr>
<td>The Box-Jenkins Approach to Forecasting</td>
<td>The Box-Jenkins methodology is applied to the time-series forecasting.</td>
</tr>
<tr>
<td>Monitoring Forecasts</td>
<td>The accuracy of forecasting activity is monitored in order to allow reaction and intervention when apparent faults occur.</td>
</tr>
</tbody>
</table>

Next section gives a brief review of books on time-series analysis and forecasting. Also some journals with articles on forecasting are presented.

General introductory books on time-series analysis are, e.g. /4/, /10/, /11/, /13/, /15/, /21/, /23/, /29/, /43/. Reference /41/ is a more advanced book, which is particularly strong on spectral analysis, multivariate time-series and non-linear models. Other intermediate to advanced books are /5/, /17/, /30/. The revised edition of famous book by Box and
Jenkins /6/ describes an approach to time-series analysis, forecasting and control which is based on a class of linear stochastic processes, called ARIMA models. /11/

The following books target more towards forecasting than general time-series analysis. /19/ is a general book on the topic, especially for applications in economics. Some other general texts on time-series forecasting can be found in /1/ and /35/. /3/, /12/, /16/, /33/ aim more to business and economics students. /11/

There is a useful collection of review articles in /2/. More specialized books include /20/ on structural models and /44/ on dynamic linear models, which is written from a Bayesian viewpoint. /40/ give some case studies using the latter approach together with a software package called BATS. The two main journals devoted to forecasting are the International Journal of Forecasting (published by North-Holland) and the Journal of Forecasting (published by Wiley). Papers on forecasting can also appear in many other statistical, management science, econometric and operational research journals, for example in Journal of Business and Economic Statistics, Management Science and in Journal of Econometrics. A brief review of recent developments in time-series forecasting is given in /9/. /11/

### 2.2 Applied Methods

#### 2.2.1 Change Detection

As mentioned in Section 1, the change detection methods based on statistical or Bayesian approach are quite sophisticated in nature. Additionally, as discussed in Section 2, the Bayesian approach often requires a priori knowledge of the probabilities of the changes and their distributions.

However, the applied method had to be flexible and easy-to-implement to new operating conditions. Therefore, detecting changes on the basis of the data without any a priori assumptions was an important issue. It would reduce the uncertainties involved with the a priori estimates as well as initial tuning requirement. On the basis of these criteria, three different change detection methods were chosen for further development, namely:

1. Difference Method
2. Possibilistic Approach
3. Model Based Approach

As mentioned in Section 2.1, the difference method is flexible and efficient according to /28/, where it was calculated on the basis of the sequential gray-scale values of image pixels. However, the changes to be detected were slower than those of consecutive parameter values in this case. Therefore, the difference method was modified as follows:

\[ \Delta_t = |x_t - x_{t-n}|, \]  

(1)
where $\Delta$ is the difference, 
$x_i$ is the current value of the parameter and 
$x_{i-n}$ is the lagged value.

An alarm is triggered if

$$\Delta_i \geq \text{ThVal}, \quad (2)$$

where $\text{ThVal}$ is a threshold value.

The difference is calculated when a new measurement is received. For each variable, the lag $n$ and the threshold value $\text{ThVal}$ must be predefined. The absolute value in (1) enables the use of single threshold value for both positive and negative changes.

The developed possibilistic approach for change detection starts from fuzzified distributions of parameter values from the longer and shorter periods. A change is detected if these two distributions differ significantly from each other. The difference is monitored utilising a certain index, which is defined from the intersection of the distributions. The index can receive values from interval $[0,1]$, which present the degree of affinity. The similarity increases as the index value increases. In the extremes of the interval, index value “zero” suggests that the distributions are completely different and “one” that the distributions are perfectly similar. The distributions and the formation of the index value are presented in Figure 3. Motivation for the use of fuzzy approach can be found for example in /14/.

As seen in Figure 3, triangular membership functions (MF) were utilised in fuzzification of distributions. The MF type was selected mainly because it produced more accurate index values. Also the interpretation is straightforward. For the vertices of the MFs, three mean values were calculated, that is:

1. Mean of the distribution ($M_{\text{Center}}$)
2. Mean of the values lower than $M_{\text{center}}$ ($M_{\text{Lower}}$)
3. Mean of the values higher than $M_{\text{center}}$ ($M_{\text{Higher}}$)

Now, the vertices are defined at $(0, M_{\text{Lower}})$, $(1, M_{\text{Center}})$ and $(0, M_{\text{Higher}})$ for both distributions. The updating happens in “a moving window” manner: When a new measurement comes, the oldest one is removed and the calculations are repeated. An alarm is triggered if the difference between the distributions is significant enough, i.e.:

$$I \leq \text{ThVal}, \quad (3)$$

where $I$ is the index describing the similarity and $\text{ThVal}$ is the threshold value.
Note that the absolute distributions are not of concern in this approach. Instead, the proportional change of the short-term distribution compared to the long-term distribution is important for the change detection purpose.

Figure 3. The fuzzified distributions of the parameter values from long (solid line) and short (dashed line) periods. The index describing the similarity is defined from the intersection of the distributions (horizontal line).

As mentioned above, the model-based approach included both autoregressive and linear regression model structures. In general, deviations in the model accuracy, parameters, outputs or state variables can detect whether the process conditions have changed /22/. In this case, the change detection utilised the cumulative sum of the modelling error. It is defined as:

\[ E(j) = \sum_{i=1}^{n} (y(j-i) - \hat{y}(j-i)), \]  

(4)

where \( E(j) \) is the cumulative sum of the modelling error, \( y(j) \) is the actual process output and \( \hat{y}(j) \) is the modelled process output. The index \( i \) in (4) specifies how many previous modelling errors are included in the sum. An alarm is triggered if

\[ E(j) \geq ThVal \]  

(5)

The autocorrelation plot determines if a time-series model describes the data appropriately /37/. Typical examples of different cases of autocorrelation plots (data from
random process, weak and strong autocorrelation, sinusoidal data) can be found in /37/. Compared to the autocorrelation plots of different parameters (Figure 4), it was concluded that the parameters 1-4 and have strong autocorrelation and parameters 5-6 moderate autocorrelations. Therefore, the autoregressive (AR) model structure could be used.

Figure 4. Autocorrelation coefficient plots

As the autocorrelation function indicated that the AR model was appropriate, the partial autocorrelation plot could be used to find out the order of the model (number of lags) /38/. The partial autocorrelation functions of the parameters are presented in Figure 5. It was concluded that, for parameters 1-4, the modelling could start with two lags. For parameters 5 and 6 a higher order model could be more suitable.

To test the AR model-based change detection, an ARX-model was constructed for parameter 1. Several lags were tried out and, finally, lags 5 and 10 were selected as the inputs of the model. The model is of the following form:

\[ \hat{y}(t) = B(q)u(t) + e(t), \quad (6) \]

where \( \hat{y}(j) \) is the calculated output, 
\( B(q) \) contains the coefficients for lagged variables and 
\( e(t) \) is an error term.

For the developed model, the \( B_1(q) \) was identified as 0.01739 and \( B_2(q) \) as 0.9813. The error term \( e(t) \) was defined as zero.
2.2.2 Forecasting

As mentioned in Section 2.2.1, one objective of the forecasts was to predict the parameter values in the long term. For this purpose, the linear regression model was used. Second, the process capability index ($C_{pk}$) was predicted utilising autoregressive linguistic equation-based model /25/. 

Figure 5. Partial autocorrelation function plots of the parameters.
The linear regression model was employed because linear trends could be observed in the data, especially with the first four parameters. The linear model was as follows:

\[ \hat{y} = at_n + b, \]  

(7)

where \( \hat{y}(j) \) is predicted output, 
\( a \) is the slope, 
\( t_n \) is a future time instant and 
\( b \) is a constant.

The parameter \( a \) in (7) is calculated on the basis of the data utilising the Method of Semiaverages as presented in /15, 175-176/:

\[ a = \frac{\bar{x}_2 - \bar{x}_1}{\bar{t}_2 - \bar{t}_1}, \]

(8)

where \( \bar{x}_2 \) is the average in early half of the series, 
\( \bar{x}_1 \) is the average in the later half of the series 
\( \bar{t}_2 \) is the average time value in the early half of the series 
\( \bar{t}_1 \) is the average time value in the later half of the series.

The constant \( b \) is calculated as a mean of the previous \( n \) data points:

\[ b = \frac{\sum_{i=1}^{n} x_{j-i}}{n}, \]

(9)

where \( x \) is the parameter value, 
\( n \) is the number of data points and 
\( j \) is an index referring to the current value of parameter.

As long as parameters remain unchanged, the future time instant is updated sequentially, i.e.:

\[ t_n = t_{n-1} + 1, \]

(10)

where \( t_n \) is the updated future time instant and 
\( t_{n-1} \) is the previous future time instant.

For more information about the linear regression models refer to /15/. The parameters were recalculate if the cumulative sum of modelling error (4) exceeded the threshold value as presented in (5). Therefore, as mentioned above, the method was assumed to apply to the change detection purpose. On the other hand, the parameters were also
recalculated if some other change detection method mentioned in Section 2.2.1 signalled an anomaly.

As mentioned above, also the process capability index \( C_{pk} \) was predicted. The \( C_{pk} \) is a standard term used throughout the industry to represent the quality production ability of the process /39/. The AR model structure was utilised because it was supposed that if the autocorrelation exists in the parameter values it would also exist in the \( C_{pk} \) values. Additionally, linguistic equations were applied to the model structure because of its ability to take non-linearity into account. Applications of linguistic equations to process control and monitoring tasks can be found in /25/, /26/, /27/. Detailed information about the \( C_{pk} \) index is in /39/.

The \( C_{pk} \) was calculated using distributions of 300 measurements. The lags of the model were defined utilising the knowledge about lags of the parameters (see Section 2.2.1) and trial-and-error –method. The identified model was:

\[
a_1 C_{pk}(t) + a_2 C_{pk}(t - 3) + a_3 C_{pk}(t - 5) + bias = 0, \quad (11)
\]

where \( a_1, a_2 \) and \( a_3 \) are model coefficients,
\( C_{pk} \) is the capability index and
\( bias \) is an error term.

The capability index can be predicted three steps forward by solving the equation as follows:

\[
C_{pk}(t + 3) = \frac{a_2}{a_1} C_{pk}(t) + \frac{a_3}{a_1} C_{pk}(t - 2) + bias
\]
3 RESULTS AND DISCUSSION

3.1 Test Environment

At first, the tests were run off-line utilising a personal computer and Matlab® R12.1 development environment. From the beginning, however, the structure of the program was designed according to the on-line requirements. Therefore, the system includes an automatic procedure for removal and replacement of invalid measurements. Before the change detection and forecasting, smoothening of the data had to be carried out employing moving average-technique. Figure 6 shows the flowchart of the programmed software.

As mentioned above the easy configurability of the system to other production lines was important. Therefore the software modules were designed to be as flexible and independent as possible. In the on-line application, the interface to measurement software was designed utilising Dynamic Data Exchange (DDE). It employed the hot link method, thus enabling the automatic execution of the programmed software each time a new measurement was received.
3.2 Performance of Applied Methods

3.2.1 Change Detection

This Section shows test results for three different change detection methods. The test material included a data set of 113579 measurements. The first 60000 data points were for training and the rest for validation. Visual selection from the material revealed 14 significant changes. The data and the changes are in Figure 7.

The difference method performed quite well, the number of detected anomalies was 13 and there was only one missed change and one false alarm. The lag $n$ and the threshold value $Thval$ (see Section 2.2.1) were defined as 100 and 0.055 respectively, utilising the trial-and-error method. The missed change was number 1 in Figure 7. Lower threshold would have revealed this anomaly, but this solution suffers from higher number of false alarms. The false alarm occurred near changes five and six, maybe because of increased variance in the data. The results attained with the difference method are presented graphically in Figure 8.
The results of the possibilistic approach are in Figure 9. The short-term distribution consisted of 300 measurements. The long-term distribution consisted of all the data points collected after last detected change.

In this case, only four changes of 14 were correctly detected. The number of false alarms was three, higher than with the difference method. It was assumed that the fast changes did not affect the short-term distribution sufficiently, which may explain the weak
performance of the method. The fact that the false alarms are caused by long term upward or downward trend in the measured parameter value seemed to support that viewpoint. Although several sizes of distributions were tested the performance of the method could not be notably improved.

As mentioned in Section 2.2.1, the model-based approach was implemented utilising autoregressive (AR) model structure, and the lags were defined at 5 and 10. However, as the model was identified with data from completely different period than the test data, a constant deviation between the model output and measured values was observed. Therefore, a bias term was included to the AR model structure.

The model-based change detection method produced slightly better results than the possibilistic approach. The number of correct detections was seven and only one false alarm was produced. The results are presented in Figure 10.

Figure 10. The detected changes with the autoregressive model and Cusum. The threshold value is marked with a line at 0.13.

It could be noticed that the model-based approach, opposite to the fuzzy distributions, performed satisfactory in the case of the fastest changes. This may be due to the model structure: as mentioned above, the prediction horizon was only five steps forward. Therefore, the modelled values followed the actual ones with the slower changes. In the future, the performance of the method may be improved by increasing the range of forecast. Also the adaptation of the model should be of concern because, as mentioned above, the parameters of the model may change in process with time.

The results attained with different change detection methods are collected in Table 2. It is clear that the difference method performed significantly better than the other tested approaches. Additionally, utilising parameters \( n \) and \( ThVal \), the method can be easily
tuned to detect both the fast, stepwise changes as well as slightly slower ones. Therefore, it is recommended for the future use.

### Table 2. The performance of the tested change detection methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>Correct Detections</th>
<th>Missed Detections</th>
<th>False Alarms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference method</td>
<td>13</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Model-based approach</td>
<td>7</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Possibilistic approach</td>
<td>4</td>
<td>10</td>
<td>3</td>
</tr>
</tbody>
</table>

#### 3.2.2 Forecasting

As mentioned in Section 2.2.2, the long-term trends were predicted utilising linear regression model. A reliable forecast requires a sufficient amount of data points. After data collection, the forecast was made until a change was detected, either by the methods discussed in Section 2.2.1 or by the forecasted error (see Section 2.2.2. An example of the linear regression based forecasting is presented in Figure 11.

![Figure 11. The values of forecasted and measured parameter. Forecast uses the linear regression method.](image)

A deviation between forecasted and measured parameters can be observed in Figure 11. These deviations are due to the fact that it was not possible to tune the change detection of the linear regression model more sensitive, because it would have been resulted to constantly changing, unsteady forecasts. It should be noticed, however, that the deviations in Figure 11 are quite small. In general, the main problem with the off-line test set-up is the relatively high number of data points required (approximately 4000) to produce a reliable prediction.
For on-line tests, the number of data points was only 200 and this enabled faster forecasting capability. The results are in Figure 12.

![Figure 12. Predicted (straight line) and actual parameter values, on-line tests.](image)

The on-line tests were carried inside a short period and, therefore, there were no major changes in the parameter values. The stepwise change in the prediction near 5000 demonstrates the use of cumulative sum (CUSUM) in prediction validation (see Section 2.2.2). As seen from Figure 12, the predicted values start to deviate from the actual ones near 3500 and therefore, the CUSUM increases until it finally exceeds the threshold and the trend is recalculated.

In general, the requirements stated for on-line application in Section 3.1 were met. The implementation of the software to production line proceeded rapidly and it was not necessary to make any reconfigurations to the structure of the program.

As mentioned in Section 2.2.2, the process capability index ($C_{pk}$) was predicted utilising linguistic autoregressive (AR) model. The results with the test data from the period of four weeks are presented in Figure 13. The model could be used to forecast the $C_{pk}$ about 45-120 minutes forward. It was assumed that the model could be utilised in forecasting the near future of the process state because it could manage the relatively big changes in
the $C_{pk}$ quite well. For example, between data points 100 and 200 there is a long-term (about three days) decreasing trend in the capability index. The trend ends at $C_{pk}$ of approximately 1.5 which is not satisfactory according to /39/. Control action may have stopped the weakening of $C_{pk}$.

Figure 13. The actual (blue line) and the predicted (red line) process capability indices.
4 CONCLUSIONS

In this report, some algorithms for change detection and time-series forecasting were presented and applied to an electronics manufacturing process. The structure of the employed methods was relatively simple. This in turn, compared to more sophisticated methods, may allow faster implementation and easier reconfigurability to other production lines according to results. Also the requirements of real-time, on-line manufacturing environment were considered in the applications.

Three different change detection methods were tested. Of the approaches, the difference method performed best: 13 out of 14 anomalies were detected and only one false alarm was produced. Additionally, the method was not restricted to only one type of change but could detect both slower and more rapid changes. Although some parameters have to be experimentally predefined, the difference method seems to be the most promising approach to be used for change detection in this case.

The linear regression model was utilised to predict long-term trends in parameter values. In both on-line and off-line tests, some satisfactory results were attained. The tuning of the methods, however, was experienced to be time-consuming and the method performed weakly in the presence of large variations in the data. Therefore, before further development, it should be carefully considered if the approach produces additional value to the process monitoring task.

A linguistic, autoregressive model was developed to produce short-term forecasts of the process capability index ($C_{pk}$). The model produced quite accurate predictions and hence, it was concluded that it could be used to predict the near future of the process.

In the future research, the possibility of combining expert knowledge with the aforementioned methods should be examined. The information thus received could be further utilised in the preventive maintenance and quality control. For this, however, systematically collected data and expert knowledge from longer time is required.
5 REFERENCES


