

Robust monitoring of an industrial IT system in presence of structural change

Espéran Padonou, Olivier Roustant, Michel Lutz

► **To cite this version:**

Espéran Padonou, Olivier Roustant, Michel Lutz. Robust monitoring of an industrial IT system in presence of structural change. 2013. <hal-00790206>

HAL Id: hal-00790206

<https://hal.archives-ouvertes.fr/hal-00790206>

Submitted on 22 Feb 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Robust monitoring of an industrial IT system in presence of structural change

Esperan Padonou*, Olivier Roustant*, Michel Lutz†

February 22, 2013

Abstract

This paper presents an original research initiated by the monitoring needs of a semiconductor production plant. The industrial operations rely on an Information Technology (IT) system, and several time series data are controlled statistically. Unfortunately, these variables often contain outliers, as well as structural changes due to external decisions in the IT activity. As a consequence, it has been observed that the monitoring results obtained with standard techniques could be severely biased.

This paper presents some contributions to overcome such difficulties. A new monitoring method is proposed, based on robust Holt-Winters smoothing algorithm, and coupled with a relearning procedure for structural breaks detection. Such a method is flexible enough for a large-scale industrial application. We evaluate its performances through simulations studies, and show its usefulness in industrial real applications for univariate and multivariate time series. The scope of application deals with IT activity monitoring, but the introduced statistical methods are generic enough for being used in other industrial fields.

Keywords: Monitoring, Control charts, Holt-Winters, Robust statistics, Structural changes

*Ecole Nationale Supérieure des Mines, EMSE-FAYOL, CNRS UMR6158, LIMOS, F-42023 Saint-Etienne, France

†STMicroelectronics - Ecole Nationale Supérieure des Mines, PIESO/UMR 5600

1 Introduction

Manufacturing activities efficiency relies increasingly on Information Technologies (IT). This is clearly true for wafer fabrication plants, where this research is ongoing through a partnership with the company STMicroelectronics. Several activities are useful for a proper IT management (see eg. Rudd *et al.*¹⁸). This paper is focused on the monitoring activity. A careful monitoring is based on a cautious observation of the IT system. The objective is to have a close look on all data recorded and stored to track the activity of a plant IT system.

As the size and complexity of IT systems strongly increase, the viability of monitoring can only be ensured by employing automated procedures as mentioned by Dugmore *et al.*⁵ Usual monitoring is based on threshold exceeding detection (Rudd *et al.*¹⁸). However, facing the enormous diversity and expanse of the components of modern IT systems, IT experts do not have always enough knowledge or time to determine a priori the critical thresholds. Consequently, automated monitoring procedures that do not require any preliminary hypothesis about the monitored variables, are required. In this case, the monitoring activity requires the ability to distinguish normal system operations from exceptional events.

To quantitatively detect such abnormal behaviors in the fluctuation of a time series, a standard procedure is the following: 1) Model the time series to be monitored; 2) Monitor model residuals, through an appropriate control chart (see eg. Montgomery¹³, Croux *et al.*⁴). This procedure is known in IT management studies. By way of example, Hellerstein⁸ has developed an approach based on ARIMA modelling, whereas Brutlag³ and Leikis¹¹ have recourse to Holt-Winters smoothing. This paper focuses on the latter. Holt-Winters smoothing is well adapted in industrial contexts, where time series have to be modeled without any preliminary analysis. Moreover, it is a fairly good approximation for many kinds of time series, accurate enough for industrial applications (see eg. Makridakis *et al.*¹², Croux *et al.*⁴). Furthermore, we observed that the Holt-Winters model is validated for most of the time series encountered in our research field.

However, a monitoring based on the usual Holt-Winters algorithm revealed several limitations at STMicroelectronics. That is why we developed some methods to overcome these difficulties. This paper provides further details about their implementation in such a real industrial background. First of all, we observed that the usual Holt-Winters monitoring procedure could be severely biased by the presence of

outlier observations. Consequently, we implemented a robust approach, as proposed by Gelper *et al.*⁶ and Croux *et al.*⁴ Nonetheless, we noticed that the efficiency of this robust Holt-Winters control is seriously affected when structural changes occur. That is why we developed a dynamical procedure, which detects structural breaks in a time series and reinitializes the robust smoothing parameters accordingly. This dynamical procedure is an original contribution to References^{4,6}, contributing to improve the flexibility for applications in industrial fields. Finally, it should be mentioned that an industrial monitoring implies a simultaneous control of several dozens of variables. There are numerous approaches in multivariate process control (see e.g. Bersimis *et al.*² for a review), and a specific investigation of robust techniques in this context is beyond the scope of the paper. Alternatively, we show how the robust controlling procedure can be adapted to the framework of multivariate process control.

This paper is organized as follows. In Section 2, we introduce the main statistical methods that have been deployed to control the STMicroelectronics IT system: The robust Holt-Winters monitoring proposed by Gelper *et al.*⁶ and Croux *et al.*⁴; The dynamical contribution developed through our researches, improving its flexibility in changing environments. In Section 3, some performance tests are introduced. They are grounded on quantitative simulations to: 1) Compare the usual and robust Holt-Winters monitoring procedures; 2) Evaluate the structural break detection capacity of the new dynamical procedure. Lastly, Section 4 introduces some example from real industrial cases: Univariate and multivariate applications are considered.

2 Investigation of a robust monitoring for trended time series with structural changes

We recall here the main principles of Holt-Winters based monitoring, and its robust version introduced by References^{4,6}. Then, we propose a new methodology in presence of structural changes.

2.1 Monitoring based on Holt-Winters smoothing

The Holt-Winters (HW) algorithm is a popular technique used to provide short-term forecasts of a given time-series (see e.g. Makridakis *et al.*¹²). The predictions are built iteratively as a linear combination of the observed values and the prediction

obtained at last step.

For illustration, let us consider a time series y , observed at dates $1, 2, \dots, n - 1$. For the sake of simplicity, we assume that y is non-seasonal, though the methodology is similar. The HW algorithm is based on the assumption that y is a sum of two time-series α and β corresponding respectively to a local level (order of magnitude) and a trend. These auxiliary time series are estimated iteratively as averages of the last observation and the last predictions, weighted by two parameters λ_1 and λ_2 :

$$\hat{\alpha}_t = \lambda_1 y_t + (1 - \lambda_1) \hat{y}_{t|t-1}, \quad t = 1, \dots, n - 1 \quad (1)$$

$$\hat{\beta}_t = \lambda_2 (\hat{\alpha}_t - \hat{\alpha}_{t-1}) + (1 - \lambda_2) \hat{\beta}_{t-1}, \quad t = 1, \dots, n - 1 \quad (2)$$

Logically, the one-step-ahead forecast done at date $t - 1$ for date t is then given by:

$$\hat{y}_{t|t-1} = \hat{\alpha}_{t-1} + \hat{\beta}_{t-1}, \quad t = 1, \dots, n \quad (3)$$

which gives in particular the prediction at date n . In practice, λ_1 and λ_2 are estimated by minimizing a criterion (often the least-square criterion) based on the forecast errors:

$$E_t = y_t - \hat{y}_{t|t-1}, \quad t = 1, \dots, n - 1 \quad (4)$$

and the algorithm is initialized by a linear regression on the first m values.

As a second step, monitoring can be performed. While applying a control chart to y is not recommended, due to the violations of the usual assumption (*identically and independently distributed data*) especially if y is trended, the forecast errors E_t may be close to satisfy it. Then, assuming furthermore that E_t are normally distributed $N(0, S^2)$ the upper and lower control limits for E_t are then given by:

$$UCL = +q_{\alpha/2} * \hat{S} \quad (5)$$

$$LCL = -q_{\alpha/2} * \hat{S} \quad (6)$$

where $q_{\alpha/2}$ is the quantile of a Student distribution at level $\alpha/2$, and \hat{S}^2 is the usual variance estimator:

$$\hat{S}^2 = \frac{1}{n - m - 1} \sum_{t=m+1}^{n-1} E_t^2 \quad (7)$$

These limits intend to detect the dates that correspond to an anomaly: A value of E_t outside the interval [UCL, LCL] should be a strong indication of an abnormal behavior (for a given confidence level α). However, the limits themselves are sensitive to outliers, since the variance estimator overestimates the true variance in presence

of outliers. Furthermore, the predicted value $\hat{y}_{t|t-1}$ depends linearly on past values that may contain outliers. These problems are solved by the robust version of the Holt-Winters monitoring.

2.2 Robust monitoring based on Holt-Winters smoothing

The Robust Holt-Winters algorithm (RHW) introduced by Gelper *et al.*⁶ and Croux *et al.*⁴ considers two additional auxiliary time series: y^* , representing a *cleaned* proxy of y after outliers treatment, and σ the expected prediction error, representing a robust estimate of the forecast error E_t . To obtain a robust algorithm, large values are truncated when larger than a given threshold. More precisely, the expected errors σ_t are computed recursively by:

$$\hat{\sigma}_t^2 = \lambda_\sigma \left[\psi_k \left(\frac{E_t}{\hat{\sigma}_{t-1}} \right) \right]^2 \hat{\sigma}_{t-1}^2 + (1 - \lambda_\sigma) \hat{\sigma}_{t-1}^2 \quad (8)$$

where λ_σ is a given weight, and ψ_k is the Huber function with boundary value k :

$$\psi_k(x) = \begin{cases} x & \text{if } |x| \leq k, \\ \text{sign}(x) \times k & \text{if } |x| > k \end{cases} \quad (9)$$

The error E_t is still given by $E_t = y_t - \hat{y}_{t|t-1}$, with $\hat{y}_{t|t-1} = \hat{\alpha}_{t-1} + \hat{\beta}_{t-1}$, but the local level and trend are now estimated by using the cleaned time series y^* :

$$\hat{\alpha}_t = \lambda_1 y_t^* + (1 - \lambda_1) \hat{y}_{t|t-1} \quad (10)$$

$$\hat{\beta}_t = \lambda_2 (\hat{\alpha}_t - \hat{\alpha}_{t-1}) + (1 - \lambda_2) \hat{\beta}_{t-1} \quad (11)$$

where y_t^* is given by:

$$y_t^* = \psi_k \left(\frac{E_t}{\hat{\sigma}_t} \right) \times \hat{\sigma}_t + \hat{y}_{t|t-1} \quad (12)$$

Notice that the role of the Huber function ψ is to truncate the forecast errors E_t when larger than k times the expected prediction error $\hat{\sigma}_t$:

$$\psi_k \left(\frac{E_t}{\hat{\sigma}_t} \right) \times \hat{\sigma}_t = \begin{cases} E_t & \text{if } |E_t| < k \hat{\sigma}_t, \\ \text{sign}(E_t) \times k \hat{\sigma}_t & \text{if } |E_t| > k \hat{\sigma}_t \end{cases} \quad (13)$$

For instance, Equation (12) can be rewritten in a simpler way:

$$y_t^* = \begin{cases} y_t & \text{if } |E_t| < k \hat{\sigma}_t, \\ \text{sign}(E_t) \times k \hat{\sigma}_t + \hat{y}_{t|t-1} & \text{if } |E_t| > k \hat{\sigma}_t \end{cases} \quad (14)$$

Finally the parameters λ_1 and λ_2 , and the standard deviation S of the forecast errors are computed with robust procedures, while λ_σ is fixed to 0.3 (see References^{4,6} for more details). The algorithm initialization is also done in a robust way by repeated median regression. For that, we recommend using a period of length 7: This is short enough to assume a local linear trend and long enough to be resistant to 2 outliers.

Since the RHW smoothing is a full robust procedure, the control charts based on the errors E_t (see Section 2.1) are now resistant to outliers, which is a clear improvement to the (non-robust) HW-based monitoring. However, an adaptation is necessary in presence of a structural change.

2.3 Robust monitoring with structural changes

In this section, we focus on time series that possibly contain outliers and structural changes. To detect structural changes in parametric models, three main classes exist (Zeileis²²): F statistics, fluctuation tests and maximum likelihood scores. We choose to use a common and simple F statistics, the Chow test, since it can be easily adapted to a robust framework. It consists in splitting the sample into 2 groups: The first one before the break date and the second one after. The model parameters are estimated for both of them so that an F test be performed to judge whether they are equal or not. The Chow test is easy to use but restricted by 2 limitations. The first one is mentioned by Hansen⁷: "An important limitation of the Chow test is that the break date must be known a priori". Moreover, the exact number of changes is unknown. The second problem is a question of robustness: A break may be missed or falsely detected because of outliers. In this section, we show how the robust Holt-Winters smoothing allows to deal with these problems and deduce a strategy for structural change monitoring.

An introducing example Consider the time series in Figure 1 with an outlier at date 10 and a break at date 26. As expected, the robust algorithm is not sensitive to the outlier at date 10 contrarily to the classical smoothing: This is its main advantage. But after that, the level changes suddenly at date 26. The robust algorithm does not admit this modification quickly and many false alarms follow. Nevertheless, this specificity can be used to detect break dates. Indeed, when a structural change happens, there is a quite long period (here 26—34) of successive false alarms corresponding to successive large errors in the RHW smoothing. During this period, the predicted values $\hat{y}_{t|t-1}$ seem to exhibit a deterministic pattern (Figure 1, d): Actually, we show below (Figure 2 and Proposition 1) that they match exactly with an

exponential function. These two facts strongly suggest that the periods of successive large errors given by the robust Holt-Winters smoothing are useful for break dates detection.

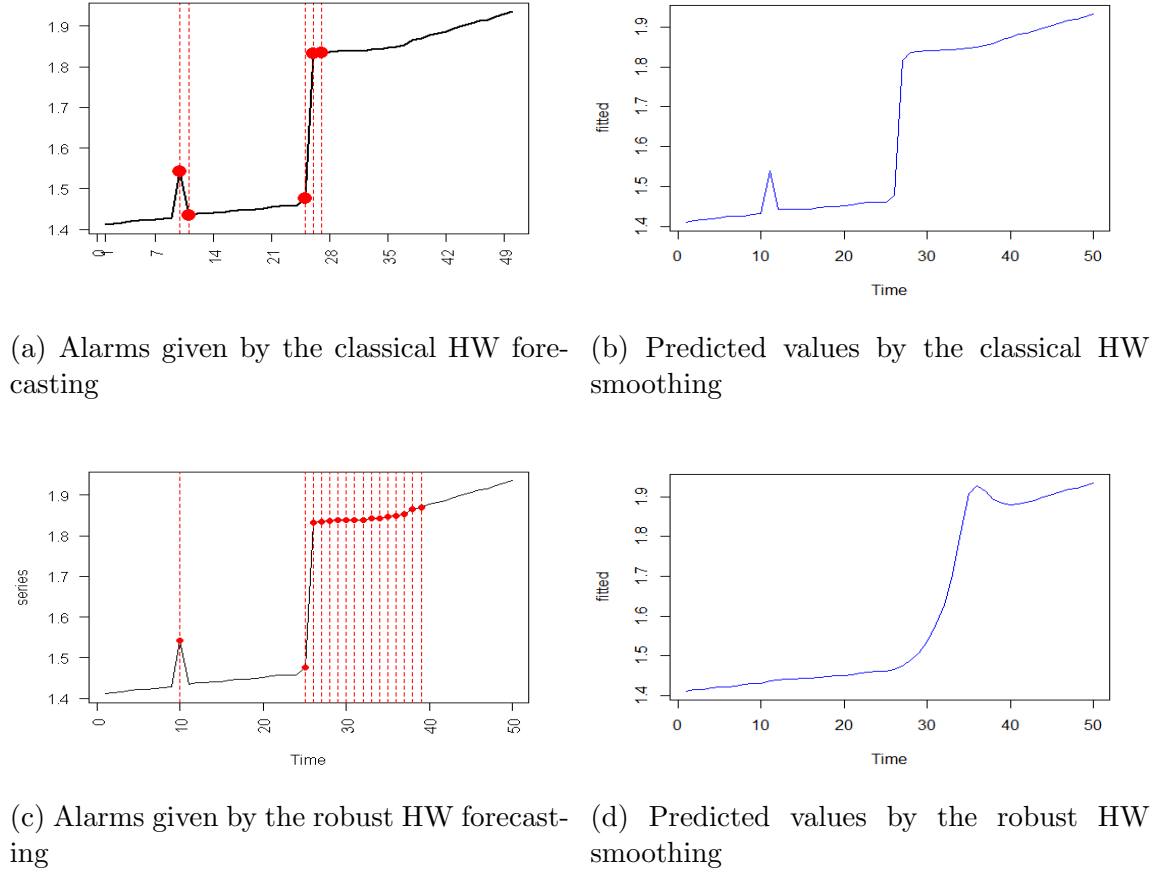


Figure 1: A real time series with outlier and structural change

2.3.1 Structural change and consecutive large errors of the robust Holt-Winters smoothing

When a structural break occurs in practice, there is a sequence of consecutive false alarms due to a succession of large errors. The robust smoothing enables to quantify the importance of these errors by comparing them to their predicted values. Let us call *relative error* the ratio $\frac{E_t}{\hat{\sigma}_t}$. A succession of large values of this ratio is a

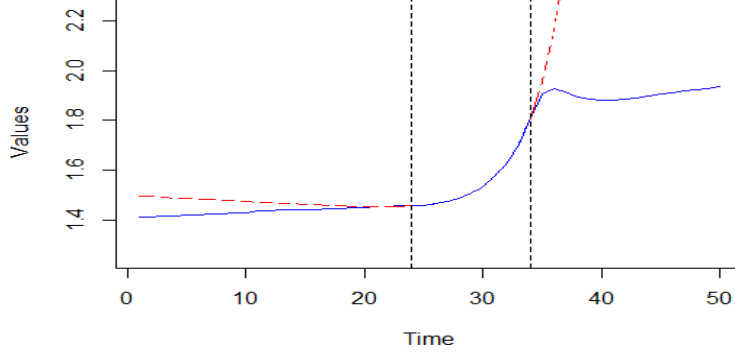


Figure 2: Zoom on Figure 1 (d): The predicted values (solid line) coincide with a function f of the form $f(t) = ar^t + bt + c$ (dotted lines) during the period $[26-34]$.

forewarning sign of structural change. More formally, we call t_1 a *suspicious date* for structural change if there exists an integer $p \geq 3$ such that:

$$\psi_k \left(\frac{E_{t_1}}{\hat{\sigma}_{t_1}} \right) = \psi_k \left(\frac{E_{t_1+1}}{\hat{\sigma}_{t_1+1}} \right) = \dots = \psi_k \left(\frac{E_{t_1+p-1}}{\hat{\sigma}_{t_1+p-1}} \right) = k \quad (15)$$

or

$$\psi_k \left(\frac{E_{t_1}}{\hat{\sigma}_{t_1}} \right) = \psi_k \left(\frac{E_{t_1+1}}{\hat{\sigma}_{t_1+1}} \right) = \dots = \psi_k \left(\frac{E_{t_1+p-1}}{\hat{\sigma}_{t_1+p-1}} \right) = -k \quad (16)$$

The period $[t_1, \dots, t_1 + p - 1]$ is a *suspicious period*: a period when forecasting errors remain k times higher than their expected values.

Proposition 1 During a suspicious period $[t_1, \dots, t_1 + p - 1]$, the predictions $\hat{y}_{t|t-1}$ of the robust HW smoothing are given by a deterministic and monotonic function f which does not depend on any observation posterior to date t_1 . Its form is:

$$f(t) = \frac{r^{(t+1)}}{(r-1)^2} + a.t + b \quad (17)$$

with

$$r = \sqrt{\lambda_\sigma(k^2 - 1) + 1} = 1.378405 \quad (18)$$

a and b are given by \hat{y}_{t_1} and \hat{y}_{t_1-1}

Proof See Appendix 1.

2.3.2 A new methodology for structural change monitoring

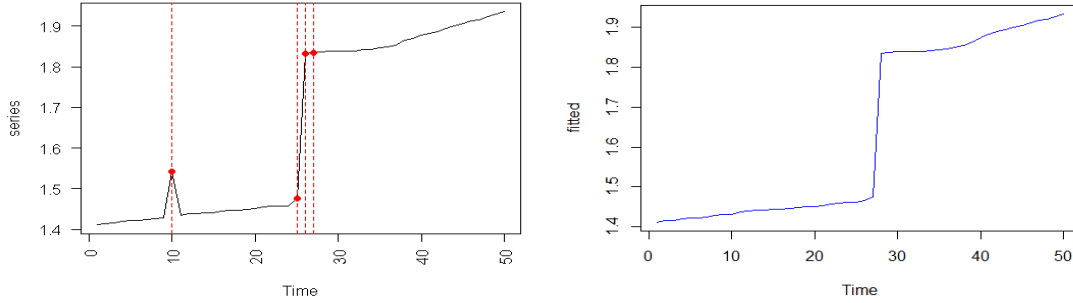
The previous observations and result about the RHW smoothing in case of a structural change suggest the following methodology that we call RHW-SC:

1. Find the suspicious dates by looking at consecutive relative errors given by the robust HW smoothing (Equations 15 and 16)
2. Apply a robust version of the Chow test to the suspicious dates detected in 1. One robust version of the Chow test consists in replacing the usual linear regression by the repeated median regression (Siegel¹⁹, Rousseeuw *et al.*¹⁷) and using a robust estimator of the residuals' variance (Croux *et al.*⁴) to compute the F statistic.

This strategy tackles the two main issues mentioned at the beginning of this section: All the possible break dates (number and locations) are automatically detected by the algorithm itself, and for a given date the statistical test for structural change is done in a robust way. In practice the methodology is applied dynamically, and when the robust Chow test is positive, the robust HW smoothing is reinitialized.

Performance of the RHW-SC methodology on the introducing example

To solve the problem raised by the introducing example, we perform dynamically the RHW-SC methodology with $p=3$. The methodology has detected only one suspicious date, namely day 26, which indeed corresponds to the structural change. Notice that only one Chow test has been used contrarily to exhaustive methods such as Quandt-Anderson that systematically test all dates. We reinitialized the robust HW smoothing at date 26, by using the repeated median regression coefficients. As a result, RHW-SC methodology remains resistant to the outlier at date 10 without generating a long sequence of false alarms after the break.



(a) Alarms given by the RHW-SC method- (b) Predicted values with the RHW-SC methodology

Figure 3: Alarms and predicted values by the robust version with structural change detection

3 Global performance tests

Some ARIMA models correspond to exponential smoothing methods (see e.g. Hyndman *et al.*¹⁰ . In particular, a Holt-Winters forecasting with smoothing parameters (λ_1, λ_2) is equivalent to an ARIMA $(0, 2, 2)$ model of parameters (θ_1, θ_2) if: $-1 \leq \theta_2 \leq 1$, $\theta_2 - \theta_1 \leq 1$ and $\theta_2 + \theta_1 \leq 1$, with :

$$\lambda_1 = 1 + \theta_1 \tag{19}$$

$$\lambda_2 = \frac{1 - \theta_1 - \theta_2}{1 + \theta_2} \tag{20}$$

Thus, we first use ARIMA $(0, 1, 1)$ time series with parameter $\theta = 0.5$. Indeed, they correspond to the special case where $\theta_2 = 0$. Next, we perform our simulations on ARIMA $(0, 2, 2)$ models with $(\theta_1, \theta_2) = (1, -0.25)$. These parameters have been estimated from an industrial time series by assuming that it comes from an ARIMA $(0, 2, 2)$ model. In this section, we compare first the performances of the RHW and the HW smoothing statically and dynamically. Afterwards, we evaluate the ability of the RHW-SC method to detect a structural change.

3.1 Comparison tests in a static setting

We are interested in comparing 2 characteristics of the HW and the RHW methods:

1. The *power*: The probability that an outlying observation is detected
2. The *false detection rate*: The probability that a normal observation is detected. This risk is called *size* of the control chart when there is no outlier. Below, we use 0.95 as a confidence level, so the false detection rate is expected to be 0.05.

3.1.1 Simulation results for ARIMA (0, 1, 1) time series

We use the following strategy:

1. Generation of a time series: simulate an ARIMA (0, 1, 1) time series of length 160 with parameters $\theta = 0.5$. The first 60 values serve as a training sample (including $m = 7$ values for HW and RHW initializing).
2. Generation of outliers: For a fixed contamination rate R , choose randomly $160R$ dates among the 160 dates. Contaminate these observations by adding or subtracting to them a value e . Whether it is an addition or a subtraction is chosen at random. Two cases are considered for e :
 - (a) e is a fixed value among 10, 20 or 30
 - (b) "Mix": e is chosen at random uniformly among 10, 20, 30
3. Perform the HW and the RHW methods and estimate:
 - (a) Their *power*: as the percentage of the contaminated observations that are really detected
 - (b) Their *false detection rate*: as the percentage of non contaminated observations that are detected.
4. Repeat 100 times the steps 1 to 3.

The RHW and HW monitoring are equivalent when there is no outlier (Reference⁴). Their differences become significant when the data is contaminated.

On the one hand, Table 1 and Table 2 below show that the HW control chart is subject to two effects. The first one is the widening of its control limits by outliers, which tends to reduce abnormally its false alarms rate. The second one concerns the dates following these outliers: the corresponding predictions are biased and this fact tends to raise the false alarms rate. Here, the first effect is predominant. This explains the too low false alarms risk and the poor detection rate of the HW method.

On the other hand, Table 3 and Table 4 show that the RHW methodology outperforms the HW smoothing especially when the number of outliers raises and when the magnitude of these outliers is not fixed: which is a realistic case. In fact, the RHW control chart doesn't lose neither its power, nor the stability of its false alarms rate that remains stable around the theoretical value 5%.

Table 1: HW detection rate

	R = 2%	R = 5%	R = 10%
e = 10	93.5	70.0	31.0
e = 20	94	74	35
e = 30	91	72	36
Mix	72	55	33

Table 2: HW false alarms rate

	R = 2%	R = 5%	R = 10%
e = 10	3.3	3.6	2.1
e = 20	1.7	3.0	7.5
e = 30	1.9	6.0	10
Mix	3	5.3	4

Table 3: RHW detection rate

	R = 2%	R = 5%	R = 10%
e = 10	100	100	99.2
e = 20	100	100	99.9
e = 30	100	100	100
Mix	100	100	99.8

Table 4: RHW false alarms rate

	R = 2%	R = 5%	R = 10%
e = 10	5.2	4.8	6.1
e = 20	5.2	5.2	6.4
e = 30	5.2	4.8	7.1
Mix	5.1	4.9	6.1

3.1.2 Simulation results for ARIMA (0, 2, 2) time series

The same strategy as in the precedent section is used. Only the first step is modified to generate ARIMA (0, 2, 2) models with $(\theta_1, \theta_2) = (1, -0.25)$. The results are summarized below (Tables 5, 6, 7 and 8). They show that the RHW control chart remains better in term of detection. Nevertheless, its false risk becomes higher than expected even if it remains stable enough for a fixed contamination rate contrarily to the HW smoothing.

Table 5: HW detection rate

	R = 2%	R = 5%	R = 10%
e = 10	97.6	91.3	67.0
e = 20	100	92.2	74.4
e = 30	97.1	91.0	73.3
Mix	92.7	77.3	57.5

Table 6: HW false alarms rate

	R = 2%	R = 5%	R = 10%
e = 10	3.3	2.7	3.7
e = 20	2.1	4.1	8.0
e = 30	5.2	1.5	4.0
Mix	2.2	3.9	6.4

Table 7: RHW detection rate

	R = 2%	R = 5%	R = 10%
e = 10	99.8	99.0	83.0
e = 20	100	99.9	93.5
e = 30	100	99.9	95.0
Mix	99.7	99.3	91.5

Table 8: RHW false alarms rate

	R = 2%	R = 5%	R = 10%
e = 10	6.4	7.2	9.0
e = 20	7.1	7.1	8.5
e = 30	6.2	7.2	9.7
Mix	7.0	7.4	9.0

3.2 Comparison tests in a dynamical setting for ARIMA (0, 2, 2) time series

In practice, industrial variables are tracked daily; the smoothing parameters and control limits are re-estimated every day. So, we perform dynamical simulations that consist in using the same procedure as for the static setting but with updating the smoothing parameters and controls limits at each iteration.

Table 9: HW detection rate

	R = 2%	R = 5%	R = 10%
e = 10	99.6	95.4	80
e = 20	100	93.6	80
e = 30	100	97	78.6
Mix	95	70	52

Table 10: HW false alarms rate

	R = 2%	R = 5%	R = 10%
e = 10	4.6	4.3	4.2
e = 20	3.3	2.2	2
e = 30	2.5	2.2	2.4
Mix	3.0	3.2	2.0

Table 11: RHW detection rate

	R = 2%	R = 5%	R = 10%
e = 10	100	100	99.7
e = 20	100	100	100
e = 30	100	100	100
Mix	100	100	100

Table 12: RHW false alarms rate

	R = 2%	R = 5%	R = 10%
e = 10	6.3	6.8	8.6
e = 20	6.0	7.1	9.1
e = 30	6.2	6.9	9.3
Mix	6.1	7.0	7.2

Tables 11 and 12 show how suitable the HW monitoring is to update its smoothing parameters and control limits contrarily to the RHW method. Indeed, the HW detection rate is drastically improved even if the power of the RHW smoothing remains the better. Let's remark that the HW false alarm risk has raised and approaches better the theoretical value 5%. As for the RHW false alarms risk, it has decreased

to approach this same theoretical value; but this later improvement is less obvious. However, the RHW false alarms risk has become very stable and non dependant on the magnitude of outliers, which is not the case for the HW control chart. Anyway, the RHW control chart outperforms the HW one when the scale of the outliers is more realistically (randomly) chosen.

3.3 Structural change detection performance tests

To evaluate the RHW-SC methodology, we use the following strategy:

1. Generation of a time series: simulate an ARIMA (0, 2, 2) time series of length 160 with parameters $(\theta_1, \theta_2) = (1, -0.25)$. The first 60 values serve as a training sample (including $m = 7$ values for RHW initializing).
2. Generation of outliers: For a fixed contamination rate $R = 0.05$, choose randomly $160R$ dates among the 160 dates. Contaminate these observations by adding or subtracting to them a fixed value $e = 5$. Whether it is an addition or a subtraction is chosen at random.
3. Generation of structural changes: choose randomly 1 date among the 100 last values. From this date to the last one, add a linear function: $A * DATE + B$ to the time series.
4. Perform the RHW-SC methods with 0.95 as confidence level for the robust Chow test and compute:
 - (a) The detection rate for structural changes
 - (b) The false detection rate for structural changes.
5. Repeat 1000 times the steps 1 to 4.

Table 13: Detection rate

	B = 0	B = ± 50
A = 0	-	95
A = ± 50	86	95

Table 14: False detection rate

	B = 0	B = ± 50
A = 0	-	6
A = ± 50	4	5

The RHW-SC methodology is efficient enough for structural change detection especially when a shift happens. It loses in efficiency when there is no shift. But in

these latest cases, the RHW period of successive false alarms is short. The RHW-SC false detection rate corresponds to the theoretical value expected since the confidence level used for the Chow test is 95%.

4 Applications

The RHW and the RHW-SC methods are performed daily on many indicators of the information system of a company. In general, the results are satisfactory. Here, we present some of these examples and a multivariate case.

4.1 Examples of univariate time series

This first example is illustrative of structural changes detection by the RHW-SC methodology.

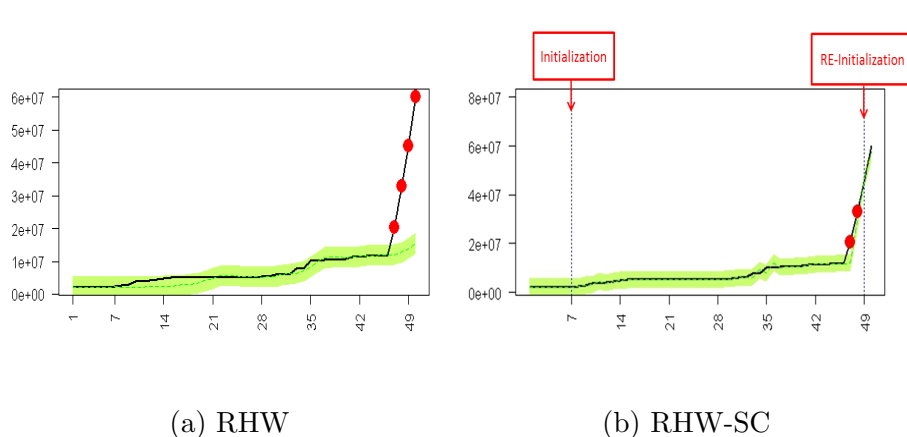


Figure 4: An industrial time series with a structural change

The time series of Figure 4 is subject to one slope change at date 47, which leads to four false alarms when the RHW methodology is performed (4a). This change is detected 3 days later by the RHW-SC monitoring, whence a reinitialization and reduction of the number of false alarms. The second example shows an extension of the RHW smoothing to seasonal time series (Reference⁶)

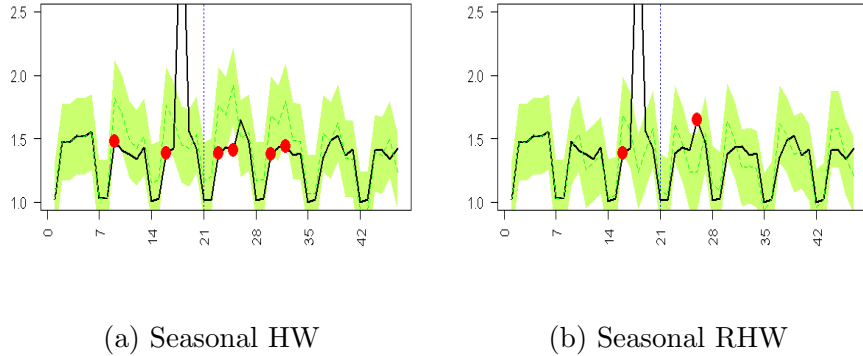


Figure 5: An industrial and seasonal time series

Figure 5 points out the sensitivity of the HW method to outliers. The abnormal observation of the third week has deformed the seasonal component. Because of the overestimation of this seasonality, the 26th observation that is really aberrant is not detected. Moreover, it leads to four false alarms the two following weeks. These problems no longer exist with the seasonal RHW smoothing.

4.2 The multivariate case

In our company, several indicators are tracked daily. So, monitoring them separately leads to false alarms every day. This is foreseeable given that the theoretical false alarms risk tends to 100% when the number of independent variables approaches 100: Whence the necessity to perform a multivariate monitoring. Notice that there are a lot of possible approaches in the multivariate framework (see e.g. Bersimis *et al.*²), and our aim is neither to do a comparative study nor to propose a best one. Rather, we show how the new robust methodology can be adapted to the multivariate case. Thus, as an example, we constitute groups of variables. For each group of p variables we use the following strategy :

1. Perform the RHW smoothing for each variable of the group.
2. Use a robust Hotelling T2 control chart to analyze simultaneously the p vectors of residuals given by step 1.

Among existing robust Hotelling T2 control charts (see e.g. Rousseeuw,¹⁵ Alfaro *et al.*¹) we consider here the computation of the confidence ellipsoid with the Minimum Covariance Determinant criterion introduced by Rousseeuw *et al.*^{15,16} . There

are two identified difficulties: A poor orientation of the confidence ellipsoid and an underestimation of its size. To face these problems, we use recent solutions found in the literature. Firstly, the orientation is improved by choosing a subset of the 75 % *best points* for the MCD criterion instead of 50% (Huber *et al.*⁹). Secondly, the size estimation is improved by using two correcting factors: One asymptotic factor of consistency to the chi square distribution (Rousseeuw *et al.*¹⁷) and one empirical result for small samples (Pison *et al.*¹⁴).

Now, we present an example with a group of $p = 4$ variables. Each day, the chi square statistic is computed for this group. The result is plotted in Figure 6. This control chart has successfully detected the 6 most important outliers but does not say where the problems come from. This is the well known problem of multidimensional *out of controls* interpretation. As an example, let us focus on the alarm on date 56. Its cause can be known by looking at each variable separately. Then, Figure 7 shows that the variable BIN2 is mainly responsible for it.

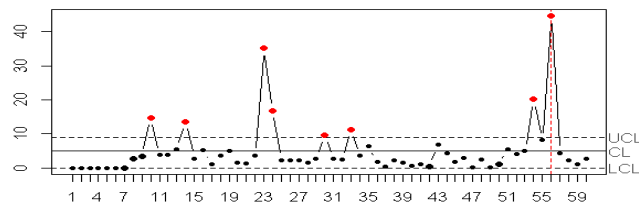
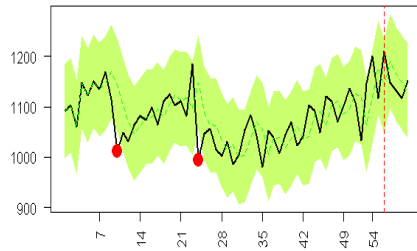
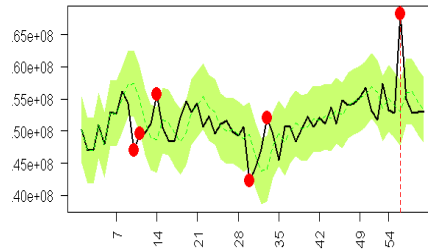


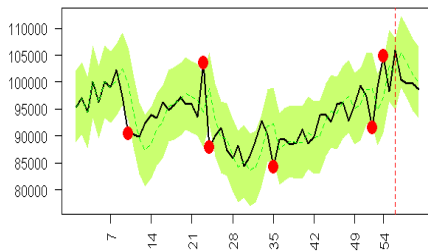
Figure 6: Chi squared daily statistics for a group of 4 variables



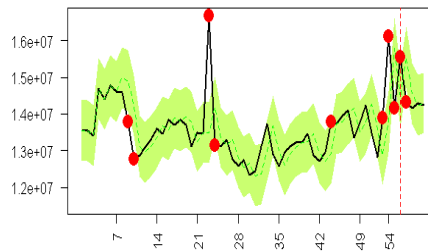
(a) Variable BIN1



(b) Variable BIN2



(c) Variable BIN3



(d) Variable BIN4

Figure 7: RHW smoothing for the four variables of the group

Nevertheless, this solution is not realistic for an industrial use because it produces too many graphs that need to be examined. Among the numerous existing criteria to interpret multivariate signals, the partial relative contributions mentioned by Montgomery¹³ are very popular for their efficiency. In Figure 7, these relative contributions at date 56 confirm the influence of BIN2.

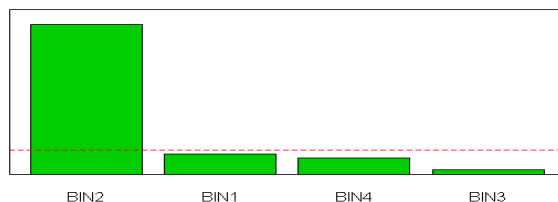


Figure 8: Partial contributions to variability at date 56

5 Conclusion

In this paper, we have studied a monitoring based on robust Holt-Winters smoothing, as proposed by Gelper *et al.*⁶ and Croux *et al.*⁴ Based on an industrial application of this method at STMicroelectronics and simulation studies, we can confirm its high robustness. However, as poor results were noticed in case of structural changes, we have also proposed and evaluated an improved dynamical approach, for a better integration of changing environments. The efficiency of this robust and dynamical method has been demonstrated on real univariate and multivariate STMicroelectronics case studies. This contribution has been developed as an improvement for the STMicroelectronics IT system monitoring, but it could be applied to any other industrial applications, where time dependent variables has to be statistically controlled.

For further research, there are several interesting outstanding questions. Firstly, some tuning choices could be investigated. Hence, the truncation parameter was set to $k = 2$ and the smoothing parameter λ_σ to 0.3 (still proposed by References^{4,6}). These values could be discussed: Do they always provide optimal results or should they be contextually adapted? Secondly, when structural changes occur, we decided to reinitialize the Holt-Winters parameters. An alternative could be to use dynamical smoothing parameters that may change over time, as proposed by Williams²¹ or Taylor²⁰. Finally, in our methodology, we chose a Chow test to detect structural breakpoints, since it is easily adapted to the robust framework. Nonetheless, as mentioned earlier, other approaches exist (other F statistics, fluctuation tests, maximum likelihood scores). Their potential of application in our monitoring procedure and a comparative study or their performance may bring to interesting further insights.

References

1. J.L. Alfaro and J.F. Ortega. A robust alternative to Hotelling's T2 control chart using trimmed estimators. *Quality and Reliability Engineering International*, 24(5):601–611, 2008.
2. S. Bersimis, S. Psarakis, and J. Panaretos. Multivariate statistical process control charts: An overview. *Quality and Reliability Engineering International*, 23(5):517–543, 2007.
3. J.D. Brutlag. Aberrant Behavior Detection in Time Series for Network Monitoring. *Proceedings of the 14th Systems Administration Conference (LISA 2000)*, 2000.
4. C. Croux, S. Gelper, and K. Mahieu. Robust control chart for time series data. *Expert Systems with Applications*, 38(11):13810–13815, 2011.
5. J. Dugmore and S. Lacy. *Capacity management*. British Standards Institution, 2005.
6. S. Gelper, R. Fried, and C. Croux. Robust forecasting with exponential and Holt-Winters smoothing. *Journal of Forecasting*, 29(3):285–300, 2010.
7. B. Hansen. The new econometrics of structural change: dating breaks ins US. Labor Productivity. *Journal of Economic Perspectives*, 15:117–128, 2001.
8. J.M. Hellerstein. Quantitative Data Cleaning for Large Databases. United Nations Economic Commission for Europe (UNECE), 2008.
9. M. Hubert and K. Van Driessen. Fast and robust discriminant analysis. *Computational Statistics and Data Analysis*, 45:301–320, 2004.
10. R.J. Hyndman, M.L. King, I. Pitrun, and B. Billah. Local Linear Forecasts Using Cubic Smoothing Splines. *Australian and New Zealand Journal of Statistics*, 47(1):87–99, 2005.
11. C.A. Leikis. Consolidated Capacity and Performance Reporting. In *Int. CMG Conference*, pages 527–534, 2007.
12. S. Makridakis, S. Wheelwright, and R.J. Hyndman. *Forecasting, methods and applications*. Wiley, 3rd edition, 1998.

13. D.C. Montgomery. *Statistical Quality Control: A Modern Introduction*. John Wiley & Sons, 2012.
14. G. Pison, S. Van Aelst, and G. Willems. Small Sample Corrections for LTS and MCD. *Metrika*, 55:111–123, 2002.
15. P.J. Rousseeuw. Least Median of Squares Regression. *Journal of American Statistical Association*, pages 871–880, 1984.
16. P.J. Rousseeuw and K. Van Driessen. A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics*, 41:212–223, 1998.
17. P.J. Rousseeuw and A.M. Leroy. *Robust Regression and Outlier Detection*. Wiley, 1987.
18. C. Rudd and V. Lloyd. *Service Design, Itil, Version 3*. Stationery Office Books, 2007.
19. A.F. Siegel. Robust regression using repeated medians. *Biometrika* 69, pages 242–244, 1982.
20. J.W. Taylor. Smooth transition exponential smoothing. *Journal of Forecasting*, 23:385–394, 2004.
21. T.M. Williams. Adaptive Holt-Winters forecasting. *Journal of Operational Research Society*, 38(6):553–560, 1987.
22. A. Zeileis. A Unified Approach to Structural Change Tests Based on ML Scores, F Statistics, and OLS Residuals. *Econometric Reviews*, 24:445–466, 2005.

Acknowledgements

The authors would like to thank the company STMicroelectronics for giving the opportunity to work on this research project, and especially X. Ambrosioni and P. Palazon, for their support. We also acknowledge the participants of the ENBIS-12 conference (Ljubljana, 9-13 September 2012), for their relevant remarks on our first oral presentation of this work.

Appendix

We prove the Proposition presented in Section 2.3.1: If $t_1, t_1 + 2, \dots, t_1 + p - 1$ is a suspicious period, then the sequence $\hat{y}_{t_1}, \hat{y}_{t_1+1}, \dots, \hat{y}_{t_1+p-1}$ is given by a deterministic and monotonic function f which does not depend on any observation posterior to date t_1 .

Proof

The period $t_1, t_1 + 1, \dots, t_1 + p - 1$ is defined by either Equation 15 or Equation 16. Without loss of generality, consider Equation 15. By combining this definition 15 and Equation 8, we obtain the equation:

$$\hat{\sigma}_t^2 = (1 + \lambda_\sigma(k^2 - 1))\hat{\sigma}_{t-1}^2 \quad t_1 + 1 \leq t \leq t_1 + p \quad (21)$$

which shows that the predicted errors $\hat{\sigma}_t$ follow a geometric progression with common ratio,

$$r = \sqrt{1 + \lambda_\sigma(k^2 - 1)} \quad (22)$$

Consequently, by denoting $t_0 := t_1 - 1$, the date before t_1 , the predicted errors are given by:

$$\hat{\sigma}_t = r^{t-t_0}\hat{\sigma}_{t_0} \quad t_1 + 1 \leq t \leq t_1 + p \quad (23)$$

With the common choice $k = 2, \lambda_\sigma = 0.3$, we have $r \approx 1.378 > 1$. Thus, the expected error goes increasing exponentially. Furthermore, the cleaned time series y_t^* becomes :

$$y_t^* = k.r^{t-t_0}\hat{\sigma}_{t_0} + \hat{y}_t \quad (24)$$

Now, relying on the equivalence with ARIMA(0,2,2) model (see e.g. Hyndman¹⁰), the forecast values of the Holt-Winters smoothing with parameters (λ_1, λ_2) follow the recursive scheme:

$$\hat{y}_{t+1} = (2 - \theta_1) y_t^* + \theta_1 \hat{y}_t - (1 + \theta_2) y_{t-1}^* + \theta_2 \hat{y}_{t-1} \quad (25)$$

where θ_1 and θ_2 are the parameters of the corresponding ARIMA(0, 2, 2) model (See Equations 19 and 20). Given Equation 24, this scheme becomes:

$$\hat{y}_{t+1} - 2\hat{y}_t + \hat{y}_{t-1} = P(t) \quad (26)$$

with:

$$P(t) = k[(2 - \theta_1)r - (1 + \theta_2)]r^{(t-1-t_0)}\hat{\sigma}_{t_0} \quad (27)$$

This is a linear equation, whose solutions are given by the sum of the solutions of the homogeneous linear equation and a particular solution:

$$\hat{y}_t = f(t) = \frac{r^{t+1}}{(r-1)^2} + a.t + b \quad (28)$$

The values of the constants a and b are imposed by \hat{y}_{t_1} and \hat{y}_{t_1+1} . This shows indeed that the predicted values are purely deterministic and increasing exponentially during the suspicious period.