Application and reliability of change-point analyses for detecting a defective stage in ingrated cicuit manufacturing

P. Besse*, C. Le Gall[1]
Motorola Semiconducteurs S.A.S.
Avenue Eisenhower, 31023 Toulouse Cedex
*Laboratoire de Statistique et Probabilits, UMR CNRS C5583
Universit Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex

Abstract

This paper presents a reliable method for highlighting a defective stage within a manufacturing process when the existence of a failure is only known at the end of the process. It was developed in the context of integrated circuit manufacturing, where low costs and high yields are indispensable if the manufacturer is to be competitive. Change detection methods were used to point out the defective stage. Two methods were compared and the best chosen. Thanks to this approach, it was possible to solve some yield problems for which the engineers' investigations were far from the real cause of failure. However, there is a strong requirement to assess the reliability of the suspicions cast on the incriminated stage, otherwise engineers could be made to do useless work and time could be wasted looking into events that are not the true cause of failure. Two complementary tools were implemented for this reliability assessment and their efficiency is illustrated by several examples.
Key-words: Bagging, change-point analysis, Markov Chain Monte Carlo, optimal segmentation, quality control, shuffle of a deck of cards.

[1]This work was partially supported by a grant of the French education and research ministry (CIFRE convention)

1 Introduction

Integrated circuit manufacturing is quite complex. It requires several hundreds of physical and chemical stages. According to the chip size, several hundreds of chips are manufactured simultaneously on a silicon wafer. The sequence of process stages is performed on sets of wafers called lots. At the end of the manufacturing process, chip functionalities are checked. If a chip fails one of the functionality tests, it is rejected automatically. This reduces yield because the chip cannot be sold. The number of marketable chips out of the number of chips on the wafer is the yield of the wafer. IC manufacturing is very expensive, especially because of the facilities required and costs are rising as components become more and more complex and require more and more meticulous process stages. For this reason, when the number of lots processed is very large, a decrease of yield must be explained and removed as fast as possible.

In a companion paper, Bergeret and Le Gall (2003) propose an approach to highlight the process stage associated with low-yield lots. Only, the manufacturing date of each lot at each stage is required because the approach is based on lot mixing during the manufacturing process. Many statistical techniques have been tried to solve this issue. The first purpose of this paper is to consider the mathematical background of two techniques and to compare them. Then, it deals with a very strategic objective that concerns the reliability of the process stage highlighted as a likely defective stage. It is most important not to diffuse false alarms that would make useless work for engineers and also delay the solving of yield issues.

Process stages work on exactly the same lots but with different tools. At each process stage, each lot is assigned to one tool to be processed. Usually, production management software assign the lots to an available tool. If the wafer fabrication facility is heavily loaded, there are waiting times before the lot is processed and these times can vary from one tool to another. Maintenance on tools can also cause waiting times. The processing order of lots therefore varies from one stage to another and this is known as lot mixing. This lot mixing gives valuable information. Let us assume (Figure 1a) that a stage generates 5 good lots (low rate of rejected chips) in a row and then 5 bad lots (high rate of rejected chips) in a row. It is likely that this stage is the defective stage because a sudden deterioration has occurred: before this deterioration, all the lots were good, and after, all the lots were bad regardless of the tool used to process the lots. Another process stage is given in Figure 1b. Here, the bad lots were not processed consecutively. In fact, between stages (a) and (b), the difference is the manufacturing order of each lot. The 10 values of the response variable are the same in the two graphs (i.e. the y-coordinates are the same), but they are differently ordered. The idea is to use this information to detect the defective stage by finding where the drifts between good and bad lots are the most significant.

A graphical approach is not always efficient for solving real issues because usually, even for the defective stage, the transition between good and bad lots is not clear. For this reason, change detection methods proposed by Green (1995) and Lavielle (2002) have been used. According to the definition of a defective stage, a curve can more easily be fitted to a defective stage than to a non-defective stage because, at a non-defective stage, there is more variation than at a defective stage. Using the Root Mean Squared Error (RMSE) as a goodness of fit criterion, the defective stage should be identified as

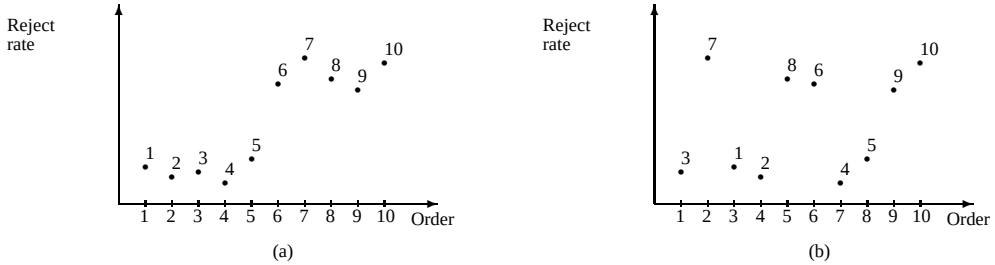

Figure 1: *Electrical reject rate versus lot manufacturing order at two process stages: (a) a defective stage, (b) a non-defective stage*

the stage with the best-fitted curve.

For the following, it is very important to retain the notion that a stage is called a defective stage only if at least one drift between low and high yielding lots is observed regardless of the tool used at the stage. The method is not able to highlight a stage where there is only one faulty tool because, in this case, the graphical representation of the stage is similar to Figure 1b: there are bad lots only when the faulty tool is used. In this case, another statistical method is more appropriate to solve the yield issue (Bergeret et al., 2003).

To prevent false alarms, the suspected stages are compared with uniformly mixed stages. If there is a considerable distance between the suspected stages and the uniformly mixed stages, this means that their lot mixing is probably unusual and that there may be a real failure at one of these suspected stages. Theory on the shuffling of a deck of cards (Diaconis, 1988) is required to make such a comparison and to obtain some theoretical information about the mixing time necessary to achieve uniform mixing. Nevertheless, the suspected stages may be little different from the uniformly mixed stages for two reasons. The first is simply that the suspected stages are not defective, and the results are really false alarms. The second reason is that the failure may be quite complex, making the lot mixing at the defective stage very similar to that of a uniformly mixed stage. Even so, the defective stage may be found because, within the manufacturing process, bad lots are the least mixed at this stage. For this reason, the bagging methodology proposed by Breiman (1996) has been adapted to test the null hypothesis of the non-existence of a defective stage within the manufacturing process.

In next sections, some theoretical tools are described in detail. Section 2 deals with the change detection methods that are used to identify the defective stage. Two methods are studied and compared: a Bayesian method with the use of a reversible Markov Chain Monte Carlo computation (Green, 1995) and a method that uses an optimal segmentation of a random process (Lavielle, 2002). When stages have been suspected by one of these statistical methods, the results are sent to the engineers. There is a strong requirement not to diffuse false alarms. Hence, in section 3, we focus on the prevention of false alarms. Some aspects of card shuffling theory (Diaconis,

1988) and bagging methodology (Breiman, 1996) are introduced. Finally, in section 4, three real examples are presented to illustrate the efficiency of this approach.

2 Detecting changes in data

In the following, two variables are considered: Y will be the response variable and X the explanatory variable. There are n bivariate observations (x_i, y_i). In the framework of this study, X is a set of n points represented on a regular grid on $[0, 1]$. This actually represents the manufacturing order of the lots at each process stage. The response variable, Y, can be the yield of the lots, for example. The Y distribution is similar for all the process stages; the only difference is that Y values are not associated with the same X values, expressing the fact that the lots have been mixed. A Y by X plot is observed for each process stage. For some process stages, changes in the trend are observed. Then, the aim is to fit a curve to each process stage to find the stages where there are the most significant changes. Y is related to X by a function f :

$$y_i = f(x_i) + \epsilon_i, i = 1, ..., n \ (1)$$

with ϵ_i, zero-mean random errors.

Many parametric and non-parametric techniques are available to estimate f. Some, such as the smoothing spline, have been tested (Bergeret and Besse, 1999). Results were quite good but it appeared that this depended on a good choice of the smoothing spline parameter. In the present work, only the fits by a piecewise polynomial function, such as a step function, are used. Therefore, it is necessary to choose the number of pieces, also called knots, and their position. These knots should coincide with time instants of change in Y values at each process stage. Two methods have been used and compared. The first one is a Bayesian method that consists of fitting Y by a piecewise polynomial function with, at the beginning, an unknown number of knots at unknown locations. These parameters have to be inferred. A reversible jump Markov Chain Monte Carlo (Green, 1995) computation is used. The second consists of an optimal segmentation of a random process (Lavielle, 2002). It requires an optimization of a contrast function, e.g. a likelihood function, penalized by the number of knots chosen. It allows changes in the mean and/or the variance of Y to be found.

2.1 A Bayesian Fit

In this case, f is assumed to be a piecewise polynomial function with an unknown number of knots at unknown locations. These parameters have to be inferred. The random errors in equation (1) are assumed to come from a Gaussian distribution with σ an unknown constant for the standard deviation. The true model of fit is unknown but belongs to the class of models M_0, M_1, ... where M_k denotes the model with exactly k interior knots. The overall parameter space Θ is a countable union of subspaces $\Theta = \bigcup_0^\infty \Theta_k$ where Θ_k is a subspace of the Euclidean space $R^{n(k)}$. The space $R^{n(k)}$ is an $n(k) = k + 1$ dimensional parameter space corresponding to model M_k. For k interior knots, the vector parameter θ^k is $(k + 1)$ dimensional with the k locations of

knots and σ^2, the variance of random errors. See Green (1995) for further mathematical details.

The joint distribution $f(k, \theta^k, y)$ is modeled as

$$p(k, \theta^k, y) = p(k)p(\theta^k|k)p(y|k, \theta^k)$$

i.e. as the product of the model probability, the parameter prior (usually a poisson distribution) and the likelihood. The target distribution is the joint posterior distribution of the pair (k, θ^k). Indeed, when these parameters are known, the piecewise polynomial function can be estimated by least square estimations. It is necessary to sample from the joint posterior distribution because analytic or numerical analyses are not efficient in this case. The full Bayesian model of this joint posterior distribution, given the data, is written as

$$p(k, \theta^k|y) = \frac{1}{Z}p(k)p(\theta^k|k)p(y|k, \theta^k)$$

where Z is a normalizing constant. The sampling was done by using a class of Metropolis-Hastings algorithms (Metropolis, 1953 and Hastings, 1970). These algorithms are able to construct a Markov chain that converges to a unique stationary distribution that coincides precisely with the target distribution. States of this chain are described by the number of knots and their position. The transistions of the chain consist in the addition, deletion or movement of a knot. Usual Metropolis-Hastings algorithms cannot be used to build such a Markov Chain because the dimension of the Markov chain states varies with the transitions of the chain. For this reason a reversible jump Metropolis-Hastings algorithm is required (Green, 1995). This algorithm builds an aperiodic and irreversible Markov chain that converges to a unique stationary distribution. Furthermore, with these properties, the ergodicity theorem can be used. See Denison et al. (1998) for further details and a C-code to implement this algorithm.

The aim of the algorithm is that the chain converge to its stationary distribution. When the Mean Squared Error (MSE) associated with the model settles down, after several iterations called "burn-it" iterations, it is assumed that the stationary distribution has been reached, the ergodicity theorem is then applicable. At this time, each new iteration of the algorithm is a part of the sample of the joint posterior distribution of (k, θ^k), given the data Y. Then, the ergodicity theorem is applied, and finally the model of fit is the average of the models generated by these iterations. Consequently, for each stage, f is estimated by a smoothed curve instead of a step function chosen, in this work, as a piecewise polynomial function f. For the following, it is important to note that the random erros are assumed to come from a Gaussian distribution with σ an unknown constant for the standard deviation even if yield variables are traditionally modeled by more complicated distributions.

2.2 Optimal segmentation of a random process

In this case, Y is assumed to be a non-stationary real process. Moreover, Y is assumed to be piecewise stationary: there are instants $(t_k, k \geq 0)$ such that (Y_{t_k+1}, ..., $Y_{t_{k+1}}$) is stationary for all $k \in \mathbb{N}$. For this reason, a family of instants of change $(t_k, k \geq 0)$ is

looked for. Let $R = (R_1, ..., R_n)$ be a random vector defined by

$$R_i = \begin{cases} 1 \text{ if there exists a } k \text{ such that } i = t_k, \\ 0 \text{ otherwise.} \end{cases}$$

Hence, it is equivalent to detect the instants of change or to recover the vector R (Lavielle, 1998). The most likely value of R given a realization y of Y is then the target value of the study. Let $\pi(r) = P(R = r)$ be the prior probability of having the configuration r of R. R is defined by a sequence of independent Bernouilli variables: $\pi(r) = \lambda^{\sum_{i=1}^{n} r_i}(1-\lambda)^{n-\sum_{i=1}^{n} r_i}$ where λ is a real parameter between 0 and 1. In fact, λ is the probability that a change appears at any instant i. The maximum a posteriori estimate is obtained by maximizing the conditional probability $P(R = r|Y = y)$. In the framework of this paper, it is assumed that the distribution of the process Y is parametric. For example, it is assumed that Y is a step function and, between two changes, the model of fit is the average of the values observed. For a given configuration r of K segments, let $\theta = (\theta_1, ..., \theta_K)$ be the sequence of parameters such that θ_k is the parameter vector associated with the kth segment. Let $h(.|r, \theta)$ be the density function of the distribution of Y conditionally to $R = r$. Then θ can be estimated simultaneously with R by maximizing the a posteriori distribution of R and

$$(\hat{r}, \hat{\theta}(\hat{r})) = arg \max_{r,\theta} h(y|r, \theta) * \pi(r).$$

In the following, $\hat{\theta}(\hat{r})$ is estimated by maximizing the log-likelihood $l_y(r, \theta)$, then

$$\hat{r} = arg \min_{r \in \{0,1\}^n} \{-l_y(r, \hat{\theta}(r)) + \alpha \sum_{i=1}^{n} r_i\} \text{ where } \alpha = log(\frac{1}{\lambda} - 1).$$

The log-likelihood is related to the fit of the observations y, whereas the second term is related to the number of pieces. The parameter α gives the relative weight related to the two terms. If α is small, several changes can be detected and there are few omissions. Alternatively, if α is large, few changes are detected and one may be omitted. Lavielle (2002) uses a dynamic algorithm that gives the global minimum rapidly. When the changes have been found, the model of fit is usually a step function. This method allows changes in the mean and/or the variance of y to be detected. Nevertheless, in this paper, only the detection of changes in the mean is used.

2.3 Conclusion

Both methods can be used to look for changes in data. However, they do not proceed with the same way. The first one uses a Markov chain and its properties to find the number and the position of the changes with a modified version of a Metropolis-Hastings algorithm. The second uses an optimization of contrast functions, e.g. a likelihood function, penalized by the number of changes. Both are efficient to find changes in data but the use of the ergodicity theorem provides a better model of fit and then a better response to the problem proposed in this paper. Furthermore, this method requires only the choice of the poisson parameter of the prior probability concerning the number

of knots and results are quite stable regardless of the value of this parameter (Bergeret and Le Gall, 2003). The alternative, the optimization of the penalized contrast function, requires the choice of several parameters, like the penalization constant α, the minimum size of the segments and the maximum number of segments authorized. In the framework of the application proposed in this paper, the choice of these parameters has too much influence on the results and the method is not always efficient to find the real defective stage. However, with an "objective" choice of the optimization parameters, i.e. a reasonable choice of the constant α, the minimum size of the segments and the maximum number of segments (see section 4 for examples), in some cases, it may appear that no changes are detected for all the process stages. This result could denote that there are actually no changes to detect because there are no real defective stages within the manufacturing process. In these cases, the Bayesian method involves the fit of the process stages by a smoothed curve making some process stages suspicious according to their goodness of fit. Hence, while the Bayesian method is really efficient to find a defective stage, it may also involve false alarms, which the algorithm of optimal segmentation may prevent.

3 Reliability of the results

This reliability is only measured in the framework of the Bayesian method because it was the one that provided the better results in highlighting the defective stage in different real situations. Nevertheless, the approach can be applied with any method. The only requirement is to have a criterion to compare lot mixing between the process stages. The stage where bad and good lots are the least mixed must be suspected as a defective stage. It is easier to fit such a stage rather than a non-defective stage where there are many variations that can only be fitted with difficulty. The RMSE ($RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y(t_i) - f(t_i))^2}$) between the actual data y and the model of fit f measures the goodness of fit. The smaller it is, the more accurate is the fit. For this reason, the defective stage is expected to have the lowest RMSE. So, in this case, the RMSE is the criterion to compare the lot mixing. It is important to note that the RMSE comparison is done with the same value of the parameter λ for each stage, otherwise the RMSE should be adjusted for the number of parameters in the model. In section 4, there are details about the choice of the parameter λ. Whether the stage is defective or not, a smoothed curve is drawn from the MCMC computation and RMSE are assessed for each stage. Thus, even if there is no defective stage, the stage with the lowest RMSE will be suspected.

In practice, a yield decrease is not always due to a defective stage. For instance, the root cause may also be a defective tool. In this case, the approach proposed in this paper with the Bayesian method is not efficient and leads to false alarms. Other statistical methods have to be used, see Bergeret et al. (2003) for further details. So, when a yield decrease appears with no prior knowledge on the root cause, the method is used without knowing if there really is a defective stage. For this reason, prevention of false alarms is strongly required. In the following, if a real defective stage exists, it will be supposed that the Bayesian method is able to put this stage among the best-

fitted stages. This is the case in practice. Prevention is based on two complementary approaches. The first approach based on the theory of the shuffle of a deck of cards (Diaconis, 1988) consists of a comparison of the lot mixing at the best-fitted stages with a uniform mixing. This comparison is able to identify cases with quite a simple failure, but if the failure is more complex, it is difficult to conclude between a real failure and a false alarm. Another approach is then required to choose between real failure and false alarm. This approach consists of testing the null hypothesis that there is no actual defective stage. It is based on bagging methodology (Breiman, 1996). It will be demonstrated that, if the distribution of the criterion used to highlight the defective stage with the bagging methodology is normal, the null hyptohesis is not rejected.

3.1 Mixing time

The manufacturing process can be compared to a deck of cards: the lots are the cards, process stages are the same deck of cards and the manufacturing process implies shuffling the cards. The shuffle observed is essentially a sequence of insertions of lots. Diaconis (1988) lists several specific shuffles. Two are particularly similar to the lot mixing observed during the manufacturing process.

Transposition: Pairs of cards are randomly transposed repeatedly. It has been shown (Diaconis, 1988) that the deck is well mixed after $k = \frac{1}{2}nlogn + cn$ shuffles. The constant c comes from the inequality $||P^k - U|| \leq aexp(-2c)$ with P the probability distribution of the mathematical model describing the shuffle by transposition, U the uniform density and a a universal constant. This inequality assesses how many shuffles are required to reduce the distance from a uniformly mixed deck.

Borel's shuffle: The top card of a deck is removed and inserted at a random position, then the bottom card is removed and inserted at a random position. This is repeated k times. It has been shown (Diaconis, 1988) that the deck is randomly mixed after $k = nlogn + cn$. If the shuffle consists of repeatedly removing a card at random and replacing it at random, the deck is randomly mixed after the same number k of shuffles.

The lot mixing observed during the manufacturing process is similar to Borel's shuffle. Nevertheless, a deep study of the manufacturing process showed that, from one stage to another, there are far less than $nlogn + cn$ shuffles of lots, where n is the number of lots at each process stage. That means that the sequence of lots observed between neighboring stages is quite similar and that the sequence begins to be different only when the stages are distant. This implies that the neighboring stages are quite similar, making their fit by a smoothed curve quite similar too. For this reason, one cannot easily suspect only the best-fitted stage but rather a small set of best-fitted stages that must be neighboring stages. To assess if the suspected stages are uniformly mixed or not, simulations of new process stages are required. We chose to simulate lot mixing between two successive process stages as observed during the manufacturing process. Thus, these simulated process stages may be considered as potential stages that may

complete the real manufacturing process or that may form a new manufacturing process. First, the aim is to assess how far the real manufacturing process and its defective stage are from uniformly mixed stages. Second, the aim is rather to compare the real manufacturing process with simulated manufacturing processes. Then, it is possible to assess if the RMSE minimum observed in the real manufacturing process is really low compared to the minimum observed in each simulated manufacturing process.

During the manufacturing process, between two neighboring stages, one can observe the number M_i of lots that have moved and the distance D_i of the moves, for $i = 1, ...,$ (number of stages-1). In fact D_i is the range between the current position of the lot i at the current stage and the previous position of the lot i at the previous stage among the sample of lots. Because of some particularities in the tools, the lot mixing is not similar from one stage to another: the M_i and D_i differ from one stage to another. These parameters are estimated from the pairs of neighboring stages. Let D be the average of the D_i and let M be the average of the M_i. First, only new stages are simulated to complete the manufacturing process. This means that around 1000 stages before the first real stage of the manufacturing process are generated and around 1000 stages after the last real stages are generated. Then, the progression towards a uniformly mixed stage can be visualized. A process stage is generated from another stage by moving M lots. These M lots are randomly chosen. They are inserted at the places chosen by a poisson distribution with D as parameter (see the following algorithm for further details). More than one insertion is made from one stage to another. Thus, it can be expected that uniformly mixed stages are achieved before $nlogn + cn$ generated stages as supposed with Borel's shuffle. To simulate new manufacturing processes, the following algorithm is used.

1. As the starting stage, generation of a random stage from any stage of the process.

   Let n be the number of lots processed by each process stage. After $\frac{1}{2}nlogn + cn$ transpositions of 2 lots uniformly chosen, lots are expected to be uniformly mixed. It is very important to generate a uniformly mixed stage as the starting stage because of the requirement not to influence the simulations by a particular configuration of lots.

2. Generation of a new manufacturing process

   Process stages are generated as observed during the real manufacturing process: they are generated by pairs from the pairs of neighboring stages observed during the manufacturing process, i.e. with the same particularities (the same D_i and M_i). Hence, one can expect the manufacturing process generated have a structure close to that of the real manufacturing process. To generate a process stage from its preceding neighbor, a sub-algorithm is used:

   (a) A lot is chosen from a uniform distribution.

   (b) Although the number of possible insertions is finite, the chosen lot is inserted at the position drawn from a poisson distribution. The value D_i is used as the poisson parameter. Hence, on average, one can expect to observe the same moves as in the real process.

(c) Repetition from step (a) according to the average number M_i of lots which have moved between the current couple of neighboring stages.

3. Use of the Bayesian method to fit each stage of the manufacturing process generated and keeping of the RMSE of each process stage simulated.

4. Repetition from step 2 for several iterations.

Then, the RMSE distribution of the actual process can be compared to the distribution of the simulated processes and it can be checked if the RMSE minimum of the real manufacturing process is really low.

3.2 Test of the actual existence of a defective stage

We shall take the null hypothesis "there is not one defective stage within the manufacturing process" and the alternative hypothesis "there is one defective stage". The null hypothesis means that all the stages could be ranked at any position between 1 and N, where N is the number of stages. So, the rank of each stage could be drawn from a discrete uniform distribution from the set $\{1, ..., N\}$. Process stages are defined from the same set of lots, the only difference is the order of the lots in the set. From this set of lots, many bootstrap sets of lots are generated. This means that some lots are replicated and some others are removed. This bootstrap sampling leads to a slight perturbation at each process stage. Under the null hypothesis, it is expected that, for each bootstrap set of lots, stages are also ranked at any position between 1 and N. The average rank of each stage is assessed from the bootstrap iterations. According to the Central Limit Theorem and under the null hypothesis, the average rank of each stage could be drawn from the Gaussian distribution. Thus, the test of the null hypothesis amounts to a test for normality: the null hypothesis is rejected if the normality test rejects the hypothesis that the distribution of the average ranks is normally distributed. It is then possible to assess the probability of concluding that there is a defective stage within the manufacturing process when in fact there is not. On the other hand, the distribution of the average ranks under the alternative hypothesis is difficult to define, making the assessment of the test power complicated but feaseable.

This approach is widely inspired by the bagging methodology proposed by Breiman (1996) in the context of the classification and regression tree (Breiman et al., 1984). Input variables are all the process stages and the response variable is the rank of the process stages after using the Bayesian fit. The aggregate response variable is the average rank of each stage obtained after all the boostrap iterations.

Beside the statistical test, the analysis of the average ranks gives further valuable information. Indeed, the stage with the lowest RMSE is ranked at the top. Thus, the lowest average ranks are for the process stages frequently ranked at the top positions. If these process stages do not coincide with the stages initially suspected, the results must be appreciated as non reliable. Furthermore, because lot mixing between neighboring stages is quite low, it can be expected that the low average ranks are associated with neighboring stages.

Table 1: *Ranking of the process stages according to ascending RMSE. The column entitled Rank gives the rank of the stage during the manufacturing process. The column entitled RMSE is the value of the RMSE related to each stage.*

Example 1			Example 2			Example 3		
Stage	Rank	RMSE	Stage	Rank	RMSE	Stage	Rank	RMSE
DIF	10	0.033963	NETTP	11	0.344	REC	56	0.110
REO	11	0.034284	EPI	7	0.345	TAB	80	0.111
MAS	8	0.035265	EPIOX	8	0.347	MSK	60	0.112
GRA	9	0.036366	MSK	9	0.349	MES	82	0.113
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	PAS	77	0.132
...	...	...	...	...	...	...	...	...
DEP	64	0.053087	NETTX	67	0.387	IMP	35	0.133

4 Application to IC manufacturing process

A number of real yield issues have been analyzed. Three typical yield issues with known solutions are presented in this paper. The first is an issue with quite a simple root cause for failure. The second is an issue with a rather complex failure root cause. Without the use of the method proposed in this paper, a lot of time would have been wasted on solving this yield issue, because the first engineering investigations were far away from the actual defective stage. The third issue was due to a faulty tool at a process stage. This process stage cannot be appreciated as defective according to the definition of Figure 1, so the results given by the Bayesian fit will be false and must be appreciated as non-reliable.

4.1 Example 1

On a mature technology, a severe drift appeared on a Process Control test generating yield losses on several products. The response variable Y is the value of the Process Control test. A high value of Y is related to a bad lot, i.e. a lot with many rejected chips. Further process investigations showed that the bad lots had been processed at one stage with a specific batch of raw material.

Search for the defective stage: Bayesian fit

When all the process stages have been fitted by the Bayesian fit, they are ranked according to their RMSE (Table 1). The best-fitted stages, that are top ranked, must be suspected. The real defective stage DIF is ranked at the top. The ranking of the top ten stages is quite homogeneous. This means that the stages with the lowest RMSE are neighboring stages. The bottom ranked stage is related to the stage with the greatest RMSE.

The real defective stage DIF (top ranked), and the bottom ranked stage DEP are

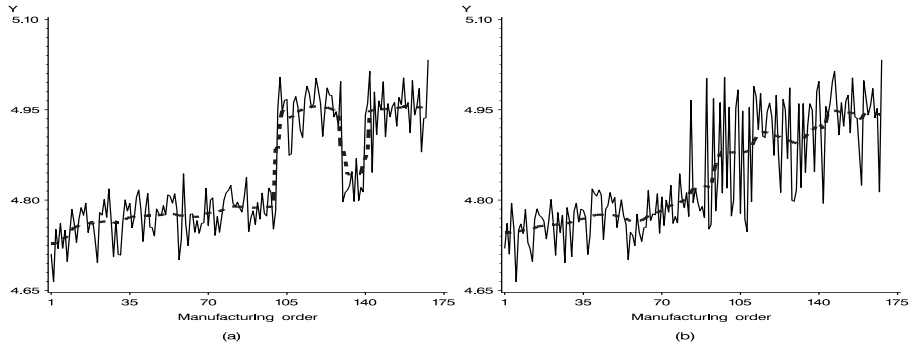

Figure 2: *Bayesian fit of the actual defective stage DIF top ranked by the RMSE (a) and the bottom ranked stage DEP (b) (solid line = real data, dotted line = model of fit).*

compared Figure 2: at the real defective stage, all the bad lots are gathered together. The few good lots among the sequence of bad lots are due to a temporary change of the raw material used.

Reliability of the results

First, we can check if the optimal segmentation algorithm confirms the presence of instants of change at the suspected stages. The penalization parameter α is fixed to 0.2. This means that a large number of knots is not penalized. The minimum size of the segments is fixed to 5 and the maximum number of segments authorized is 6. Consequently, it is supposed that, at the defective stage, there are at most 3 sequences of bad lots within 5 lots at least. Results are quite sensitive to this choice but this configuration is representative of most defective stages and the values will remain the same for the next two examples. In this case, whatever the stages of the manufacturing process, only one time instant of change is detected (Figure 3). However, as the real failure was in the real defective stage, the penalized contrast function is better optimized for this stage and its neighbors than for the other stages. This time instant of change detected at the defective stage coincides with the change of process raw material.

From the graph of the best-fitted stages, one can reasonably expect that its lot mixing is far away from a uniform mixing of lots. This is largely confirmed in Figure 4a. The real stages are represented from x-coordinates 1 to 77 according to the manufacturing order. The number 77 is related to the number of process stages within the manufacturing process of this example. The other x-coordinates are for the simulated stages that complete the real process (–993 to 0 for the stages prior to the first real stage, 78 to 1066 for the stages after the last real stage). Because the defective stage was really particular, a lot of simulated stages are required to move away from it and to come up to uniformly mixed stages. In Figure 4b, the real manufacturing process is compared to simulated manufacturing processes. The real process is represented by the x-coordinates 1 to 77, and the other simulated processes are represented by the next 77 observations. It is also obvious that the real manufacturing process is different from

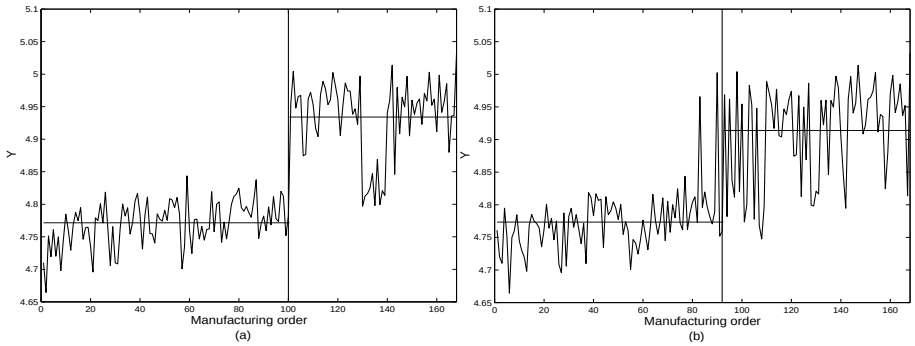

Figure 3: *Fit by optimal segmentation of the actual defective stage DIF (a) and the bottom ranked stage (b) (λ = 5, curve = real data, horizontal lines = model of fit).*

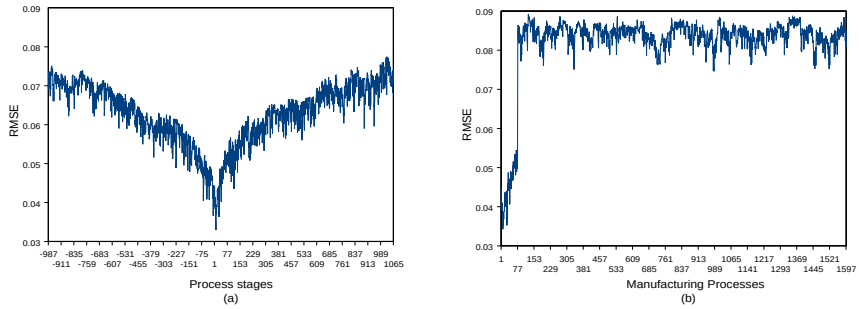

Figure 4: *Comparison of the real manufacturing process with simulated process stages (a). Comparison of the real manufacturing process with simulated manufacturing processes (b). The real manufacturing process includes 77 process stages. (Example 1)*

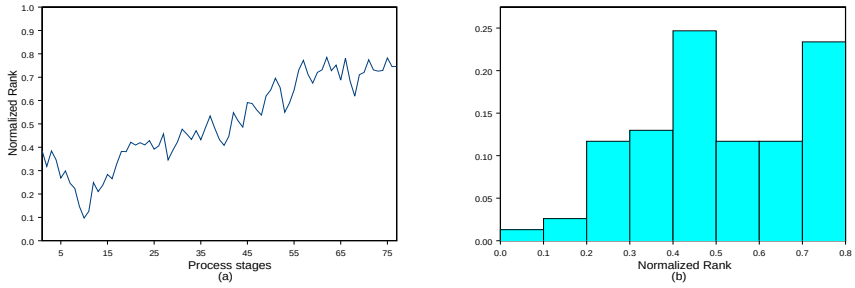

Figure 5: *Reliability of the results (Example 1).*

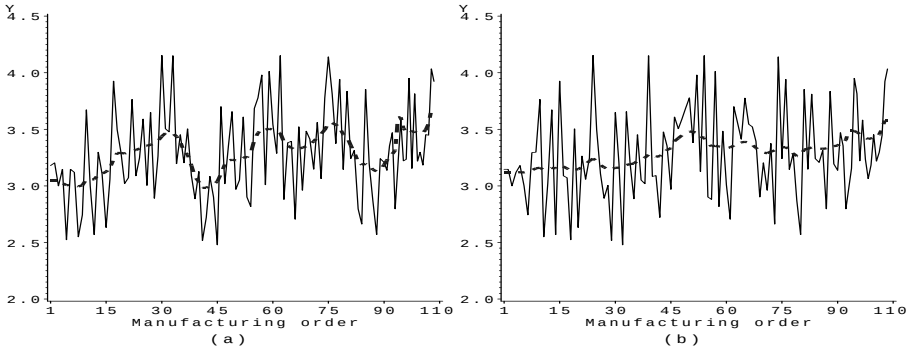

Figure 6: *Bayesian fit of the actual defective stage NETTP top ranked by the RMSE (a) and the bottom ranked stage NETTX (b) (solid line = real data, dotted line = model of fit).*

any simulated manufacturing process, and the RMSE minimum of the real process is really low. The shift between the real manufacturing process and the simulated processes is due to the random stage used as the starting stage for the process simulations. It reinforces the idea that both the defective process stage and the real manufacturing process are far away from uniformly mixed stages. Thus, we can conclude that a particular process event probably appeared during the manufacturing process at the suspected stage and this event seems to have influenced most real process stages.

Figure 5a represents the average rank of all the stages achieved after the bootstrap iterations. The lowest average ranks are for the real defective stage (the tenth in the manufacturing process) and its neighbors. The average ranks are divided by the number of process stages because this number may vary according to the yield issue. Then the normalized average rank is between 0 and 1 and a likely defective stage must have a low normalized rank. The Shapiro-Wilk test, the chosen test for normality although the average ranks corresponding to adjacent stages are strongly positively correlated, rejects the null hypothesis of the absence of a defective stage with a p-value of 0.0008. It confirms the presence of one defective stage within the manufacturing process. See Figure 5b for the distribution of the average ranks related to all the stages.

To summarize, all the preventive tools show with great confidence that the best-fitted stage highlighted by the Bayesian fit is probably the actual defective stage.

4.2 Example 2

A periodically high reject rate at an electrical test occurred on a high-production-volume product. The response variable Y is this rate of reject. Hence, a high value of Y is related to a bad lot. This problem was quickly solved thanks to the method proposed in this paper. Further process investigations showed that low yielding lots were associated with a process change that appeared at the stage suspected by the method. Actually, the process change revealed an interaction with another process stage. For this reason, the yield losses were periodic: series of high yielding lots were followed by series of low yielding lots.

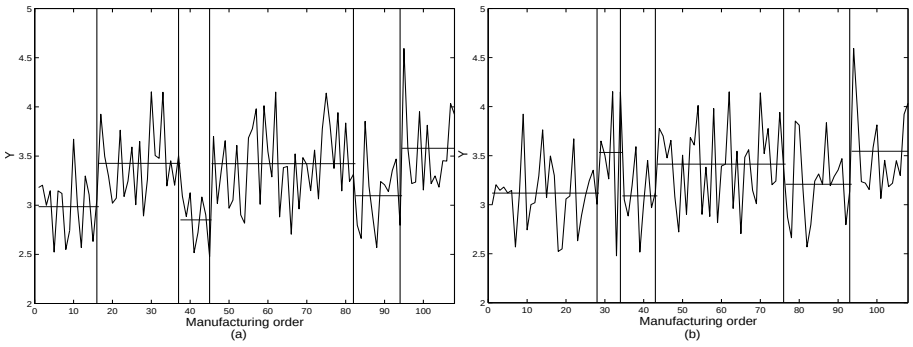

Figure 7: *Fit by optimal segmentation of the actual defective stage NETTP (a) and the bottom ranked stage (b) (curve = real data, horizontal lines = model of fit)).*

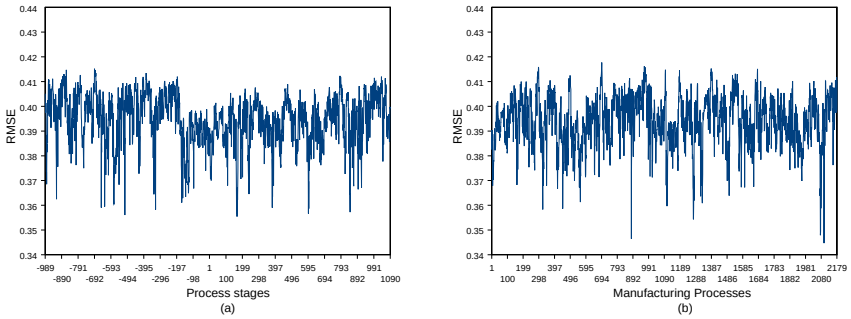

Figure 8: *Comparison of the real manufacturing process with simulated process stages (a). Comparison of the real manufacturing process with simulated manufacturing processes (b). The real manufacturing process includes 100 process stages. (Example 2)*

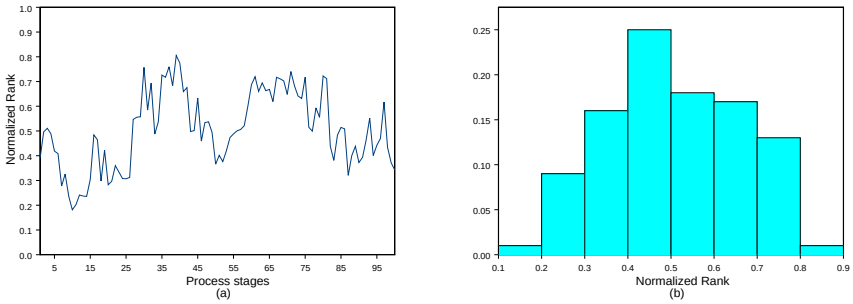

Figure 9: *Reliability of the results (Example 2).*

Search for the defective stage: Bayesian fit

The real defective stage NETTP is top ranked because it is the best-fitted (Table 1). As in the previous example, the ranking is quite homogeneous.

The real defective stage NETTP and the bottom ranked stage NETTX are compared Figure 6. The graph of the real defective stage is very different from the theoretical graph of a defective stage. This is due to an interaction between the process change and another process stage: there are some periods with a high reject rate and some periods with a low reject rate. However, compared to the bottom ranked stage, instants of change are well identified thanks to the model of fit (dotted line). Due to the variations of the reject number at the electrical test, Y is a logarithmic transformation of the reject number. This explains the Y values in the graphs.

Reliability of the results

The optimal segmentation algorithm is able to identify the time instants of change at the defective stage, Figure 7a. Some instants of change are also detected at the other process stages, Figure 7b. However, the penalized contrast function is better optimized in the case of the defective stage and its neighbors. So, both methods converge to the same results.

As previously observed, the graph of the most suspect stage, which is the real defective stage, is very different from the graph of a theoretical defective stage, that is to say a stage where good and bad lots are not mixed and where there is a shift between series of good lots and series of bad lots. Hence, before diffusing results to the engineers, there is a strong requirement to check whether the results are really due to failure at one stage. Figure 8a shows that all the real process stages seem to be lost in all the simulated stages. It is the same for the real manufacturing process among all the simulated processes, Figure 8b. This certainly means that the real process stages were nearly uniformly mixed stages. Furthermore, the RMSE minimum of the real manufacturing process is far away from the lowest. Figure 9a, however, shows the best-fitted stage at the eleventh position in the manufacturing process, and some neighboring stages have the lowest values of the average ranks achieved after the bootstrap iterations. This means that these stages are the most frequently top ranked. The graph is less easy to interpret than Figure 5a, because the failure is more complex. Nevertheless, the other groups of stages with a low RMSE are either too low, i.e. there is only one stage with a low RMSE, or the RMSE is greater. The Shapiro-Wilk test confirms the presence of one defective stage because it rejects the null hypothesis with a p-value of 0.009. See Figure 9b for the distribution of the average ranks related to all the stages. So, the results could finally be appreciated as reliable and communicated to the engineers with confidence.

4.3 Example 3

Severe yield losses appeared on a product due to one defective process tool at one process stage. These yield losses were strongly correlated with a Process Control Test used as response variable Y. A high value of Y is related to a bad lot. The Kruskal-

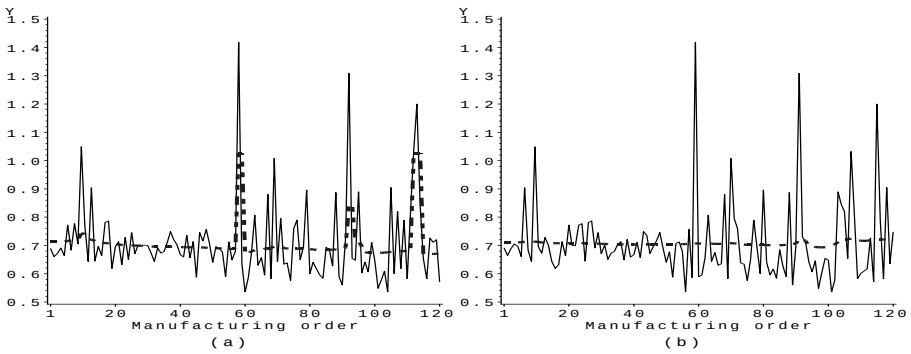

Figure 10: *Bayesian fit of the top ranked stage REC (a), and the stage with the faulty tool PAS (b) (solid line = real data, dotted line = model of fit).*

Wallis test implemented in a statistical dashboard (Bergeret et al., 2003) identified one tool at a process stage that processed all the bad lots. As there are many tools at this stage, the bad lots were sparse among all the good lots processed by efficient tools. Furthermore, there is no linear drift between good and bad lots because defective and not defective tools were used simultaneously. For this reason, the graph related to this stage is not similar to the theoretical graph of a defective stage and the use of the method proposed to find a defective stage is not suitable. In these circumstances, it is interesting to check if our approach concerning the assessment of the reliability is able to identify false alarms. In this case, it is very important to note that there was not one defective stage as defined in this paper, there was only one faulty process tool at a process stage.

Search for the defective stage: Bayesian fit

The PAS stage, where there was the defective tool, is not in the top ten ranked (Table 1). It is only ranked in 70th position. In addition, the ranking of the top ten stages is not very homogeneous.

The PAS stage and the stage with the lowest RMSE (REC) are compared in Figure 10. Both graphs are different from the theoretical graph of a defective stage. In addition, at the PAS stage, no rupture is really detected.

Reliability of the results

With the same parameters as in the previous examples, the optimal segmentation algorithm does not manage to detect time instants of change for each stage of the manufacturing process (Figure 11). This result could mean that there are no changes to detect because there is no real defective stage.

If this were the case, the comparison with uniformly mixed stages would have to confirm it. Figure 12a and Figure 12b show that the manufacturing process and the suspected stages are already uniformly mixed. Furthermore, Figure 13a is not able to highlight a neighborhood of stages that have low average ranks compared to the

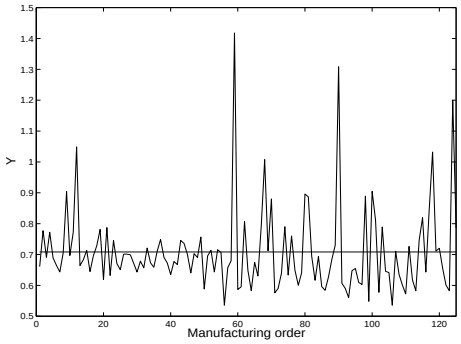

Figure 11: *Fit by optimal segmentation of the process stages: each stage are fitted by a line put at the average of the Y observations (curve = real data, horizontal line = model of fit).*

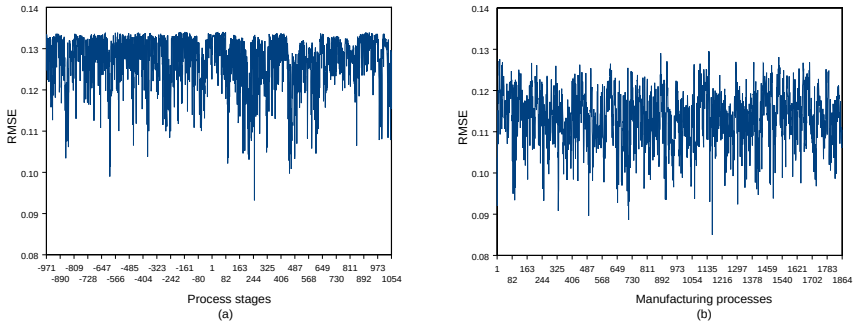

Figure 12: *Comparison of the real manufacturing process with simulated process stages (a). Comparison of the real manufacturing process with simulated manufacturing processes (b). The real manufacturing process includes 82 process stages. (Example 3)*

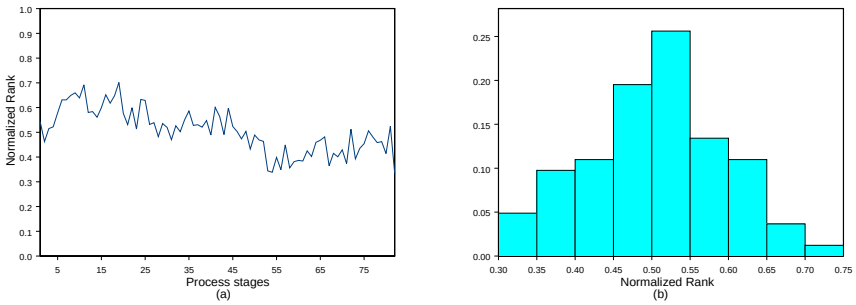

Figure 13: *Reliability of the results (Example 3).*

other stages and the lowest values of the average ranks are really high (around 0.4) compared to the previous cases where there was actually a defective stage (below 0.2). This means that the results are not reliable. In addition, the Shapiro-Wilk test with a p-value of 0.15 does not reject the null hypothesis of the absence of one defective stage within the manufacturing process. See Figure 13b for the distribution of the average ranks related to all the stages. So, in this case, with a great confidence, the results are appreciated as non-reliable and are not diffused to the engineers. In practice, other tools (Bergeret et al., 2003) were used to solve the yield issue.

5 Discussion

From the previous examples, we can conclude that the Bayesian method is efficient to highlight the defective stage. Optimal segmentation is also efficient but its use requires more parameters to be fixed than the Bayesian method does. For the application proposed in this paper, there is a strong need to propose a method with the least possible subjectivity. For this reason, the Bayesian method is preferred. However, the Bayesian method is likely to involve false alarms that are unacceptable for integrated circuit manufacturing. The complementary use of the optimal segmentation method, card shuffling theory, bagging methodology and also experience are able to prevent such false alarms. Today, this approach has become a strategic tool for yield improvement (Bergeret et al., 2003). So far, two critical yield crises have already been solved when engineers' investigations were far from the real failure cause.

References

Bergeret F., Besse P.C. (1999), "Recherche d'une étape défaillante dans un procédé industriel : quand le désordre est source d'information", *Revue de Statistique Appliquée*, vol. XLVII, no. 3, pp. 27-37.

Bergeret F., Le Gall C. (2003), "Yield improvement using statistical analysis of process dates", *IEEE transactions on Semiconductor manufacturing*, vol. 16 (3), pp. 535-542.

Bergeret F., Chandon Y., Le Gall C. (2003), "Statistical Toolbox for Yield Improvement", *Global Semiconductor*, pp42-47 .

Breiman L., Friedman J., Olshen R., Stone C. (1984), *Classification and Regression Trees*, Wadsworth & Brooks.

Breiman L. (1996), "Bagging Predictors", *Machine Learning*, vol. 26, no. 2.

Denison D.G.T., Mallick B.K., Smith A.F.M. (1998), "Automatic Bayesian curve fitting", *Journal of the Royal Statistical Society*, vol. B 60, no. 2, pp. 333-350.

Diaconis P. (1988), *Group Representations in Probability and Statistics*, IMS, Hayward, CA

Green P.J. (1995), "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination", *Biometrika*, vol. 82, pp. 711-732.

Hastings W.K. (1970), "Monte Carlo sampling methods using Markov chains and their applications", *Biometrika*, vol. 57, pp. 97-109.

Lavielle M. (1998), "Optimal segmentation of random process", *IEEE Transactions on signal processing*, vol. 46, no.5, pp. 1365-1373.

Lavielle M. (2002), "On the use of penalized contrasts for solving inverse problems. Application to the DDC (Detection of Divers Changes) problem", submitted.

Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller E. (1953), "Equations of state calculations by fast computing machine", *Journal of Chemical Physics*, vol. 21, pp. 1087-1091.