KEDGE BUSINESS SCHOOL
ILS2016
Information Systems Logistics and Supply Chain
6th International Conference
université de BORDEAUX

K-means Clustering Based Method for Production Flow Management Facing Rework Disturbances

Emmanuel ZIMMERMANN[1, 2, 3], Mélanie NOYEL[1, 2, 3], Philippe THOMAS[1, 2], Hind BRIL EL HAOUZI[1, 2], André THOMAS[1, 2]

[1]Université de Lorraine, CRAN, UMR 7039, Campus Sciences, BP 70239, 54506 Vandœuvre-lès-Nancy cedex, France
[2]CNRS, CRAN, UMR7039, France
[3]Acta-Mobilier, parc d'activité Macherin Auxerre Nord 89270 MONETEAU

{emmanuel.zimmermann, melanie.noyel, philippe.thomas, hind.el-haouzi, andre.thomas}@univ-lorraine.fr

Abstract. Some Manufacturing companies suffer from a high rework rate on their production. They have mostly problems with their production scheduling flow, too. Author's previous work have highlighted a panel of Key Performance Indicators (KPI) and their interactions. A simulation model has been made to obtain a cartography. It represents the evolution of the amount of delayed products depending of two indicators. This representation highlights four areas useful to decision making. The main goal of this work is to propose an automated approach to determine these areas and the relating production flow management decisions. The determination of these areas is performed by using clustering approach, when the production flow management rules associated to each area are determined by performing experimental design. This work refers to a bigger project that tries to develop a production flow monitoring and control system based on the Product Driven System (PDS).

Keywords: reworks, flow disturbance, indicators, clustering, design of experiments.

1 Introduction

The manufacturing companies producing very high quality products have sometimes to deal with heavy rework rate. Acta-Mobilier company, which offers very high quality lacquered furniture, is a good example. Thus, the company has to deal with a reworks rate higher than 30%. This rate is able to jump up to 80% for some product ranges. This is resulting from the gap between the growing customer requirements and the speed by which the company implements the new processes. Moreover, the delivery delay becomes even shorter. The purpose of the study presented in this paper is to give those companies support tools or methods. The aim is to understand their flows behavior in reaction to the non-quality rate variations and by the same way prevent the logistic problems. The flows disturbances resulting from the non-quality could not be easily observed precisely. However, they highlight behaviors that could be associated to snowball effect (Forester effect) making them hard to evaluate and to control. In the previous works [8], ARENA software was used to find out the non-quality indicator families. We were trying to identify with a better accuracy the workshop state. A combination of different indicators allows establishing a cartography of the perturbation impact on the workshop. This highlights different behavior areas depending on the predefined KPIs. The approach proposed in this paper flows two stages: First, the clustering algorithm is used to define automatically the borders of these areas. And in a second step an experimental design is performed to determine the best control method to apply for each of the cluster.

In the next section, we will make an assessment of the state-of-the-art about flows control, the KPIs and clustering algorithms. Then, a discussion about the proposed clustering algorithm and its results will be presented. The third section presents the approach used to determine the best production flow management rules for each cluster. In the end, a conclusion will be drawn and perspectives of evolution will be stated.

2 State-of-the-art

2.1 Flow monitoring disturbed by non-quality

Even if the non-quality became a well-known notion, as reflected by the countless works about it. The bound between rework and flow disturbance may not always be well highlighted. The COQ (Cost of Quality) [2] method with the use of the ghost factory, a parallel workshop whose mission is to repair defaults generated in the official one, introduced the term of non-quality already in the 50th. This workshop represents 40% of the official production plant capability. The first norm on this topic emerged in France in 1986 (X 50-126). It aims was to evaluate the non-quality costs [1]. A statistical analysis in relation to a problem of workstation highlights a vibrational phenomenon [10] mentioned in other works. Their aim was to evaluate a multiproduct workshop performance [10]. The flow disturbance of the production by reworks is especially significant if reworks occur on the downstream of the production process [6]. Therefore, on production monitoring subject, it is necessary to follow and maintain a low rework rate to be certain to simplify the initial monitoring problem [9]. However, even if the reworks rate is reduced and stabilized, the resulting disturbance on the product flow is still constant and complicates the scheduling function. In order to evaluate the reworks rate, the most used indicator is relative to the amount of defects. But the more the reworks rate increases, the higher is the probability to repair the same product many times. The challenging issue is how to quantify the impact of these repaired products several times? In this situation, it becomes hard to establish a direct link between reworks rate and flow disturbance.

In order to evaluate both the non-quality and the flow disturbance, a lot of indicators were used. Like all of the KPIs, these ones are formulized by the triplet (purpose, unit of measure, variable), they represent the system efficiency related to a pre-established norm. In [8], we highlight some of them, as, for example, the number of defect N_{defect}, or the time to finish all the production traditionally called $C_{max.}$
Each indicator highlights only one sight of the problem. In order to outline a more global and realistic view, [8] has demonstrated the pertinence to work with a combination of these indicators. In other word, combining two or more indicators allows to sketch–up different views and makes a more complete assessment about the production workshop efficiency. For example, N_{late} represents the number of pieces that outstep the due date and N_{defect} evaluates the number of pieces which have at least presented one defect. It is frequently used in factories because it comes directly from quality control methods. This indicator is connected to the ability of the workshop to make a product well from the first time. But, alone, it does not allow understanding and prevised the effect on the flow disturbance. Thus combined to another flow indicator, it might become possible to identify one or several thresholds. Then we could determine areas in which flow disturbance becomes uncontrollable. As a result, cartography of the workshop state can be established according to the values of the different indicators.

The goal is to develop a synthetic system of indicators usable to determine the best control method to apply for the considered situation. The figure 1 shows the combination between N_{wip} and N_{defect}. Here N_{wip} represents the number of pieces in outstanding: pieces which wait in front of a busy workstation. In this chart, four behavior areas seem to appear and our goal was to determine the boundaries of these areas.

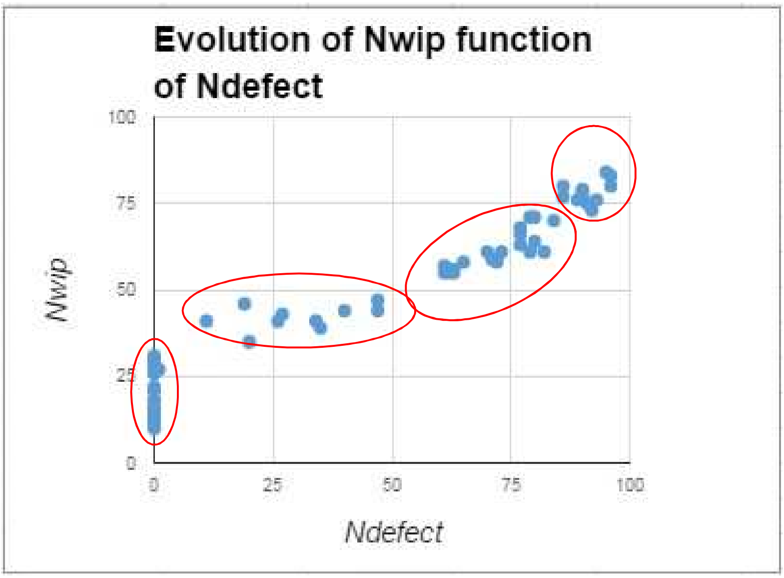


Figure 1: evolution of N_{wip} (number of pieces in outstanding) depending on N_{defect}

2.2 The bound determination by using clustering algorithms

One of the most powerful approaches to determine the membership of data to an area and to set up the limits of each of them is the clustering [4]. Many clustering algorithms exist and some of them are recapitulated in table 1 [3]. In our work, the K-means algorithm is chosen for two raisons: in our case the number of needed clusters is well known and the algorithm was simple to implement and gave a quick result. And on the other hand, the K-means algorithm aim is to gather N points of observation in K clusters with a constraint: each of the N points has to belong to the cluster for which the distance to the mean is the smallest.

Table 1: Strengths & weaknesses of some unsupervised learning clustering algorithms [4]

Algorithms	Strengths & weaknesses
K-means / K-methods	+ Allow to draw the cluster's borders + Algorithm already implemented in many calculating software + Give reasonably good results with few calculation time - The number of clusters is a parameter and not a result - Changing the number of cluster can arbitrary modify the points owning to each cluster - There is a moderate risk to fall in a local minimum - The result is completely dependent of the initialization parameters and the chosen distance
Agglomerative hierarchical	+ Monotonicity property - Unappropriated for a lot of data
Divisive hierarchical	- Rarely used Could give better results than agglomerative for a few number of clusters

Vector quantization	- Reserved only for image compression
Self-organizing maps	+ Low operating rate + Maintains the topological bounds -Unmodifiable neighborhood during operation

The idea is to use clustering algorithm on indicators values (N_{wip}, N_{defect}, C_{max}, N_{late}…) on the considered process during production. The objective is to determine the different functioning conditions. In a second step, the goal is to determine the best control rule to apply for each of these functioning conditions.

2.3 Control rules

The monitoring rule has a huge impact on the flow of products in manufacturing. For example, whereas SPT (Shortest Processing Time) privileges the shortest tasks which consume few time, EDD (Earliest Due Date) focuses on the products with the nearest due date and that without regards on the required time to achieve them. As soon as the repair rate becomes important, the most suitable control rules aren't necessarily the same regarding to the situations. There are at least 150 different control rules. FIFO (First In First Out), SPT, EDD and CR (Critical Ratio) are really frequently used rules in workshop and are considered as benchmark for simulating research [8] where the repair rate varies. Lots of work in the semi-conductor sector show that when there are huge quantity of production and repairs a specific rule like FSVCT (Fluctuation Smoothing of Variance of Cycle Time) can provide better results [7]. However, the factor "production volume" and the repairs rate inferred are really different from our context. It seems that generally, in workshops with many repairs, SPT, EDD and FIFO rules are the most frequent, even if SPT seems to be the most efficient. Currently, FIFO and EDD rules are the ones principally applied by Acta-Mobilier. Actually the company must deal with short due date that's why it focuses especially on the EDD rule.

3 The areas determination

The data used to establish the cartography, were produced by simulation with the software ARENA which gave us the possibility to record as many evaluation points as we needed. Another point is that it allows to test the system in the critical conditions of use.

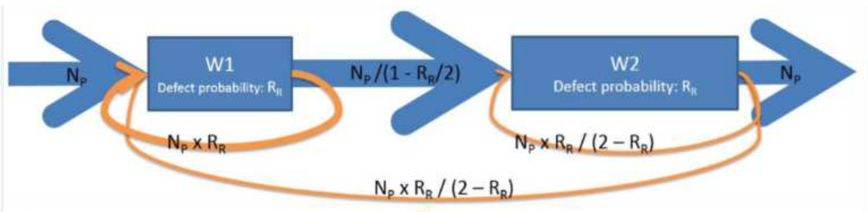


Figure 2: Model of the system used for the tests

The proposed model is based on a system with 2 workstations (W_1 and W_2) which have the same defect rate (R_R). In figure 2: N_p represents the number of produced pieces and the formula on each arrow represents the probability to reach the next step. The system variability is that the evolution of the rate of defectuous pieces producted and the outstanding rate are completely random. These ones are represented by sinusoides with high but different periodicity. This way, after some temporal periods, we are certain to evaluate the most possibles cases. Scenarii were launched for 3 different values of N_{defect}, N_{wip} and monitoring rules: EDD, FIFO and SPT.
This system will allow obtaining a global sight of the workshop state in real-time, the reworks impact and/or the non-quality of the flow. We studied the behavior depending on the progressive rise of the reworks rate. But flow disturbance stays something difficult to measure.

The K-means algorithm has been used in order to well define the behaviors areas. Data processing was made with SCILAB which contains a standard version of the K-means in its functions library [11]. The chosen settings are: 4 clusters and the Euclidian distance.
The figure 3 shows the result obtained on the data by the use of K-means algorithm with the settings described before. The bounds between the different areas is also drawn. This figure shows that the four areas are well identified and the bound between these areas may be determined. The second steps consist to determine the best rule to use for each identified operating points (each areas).

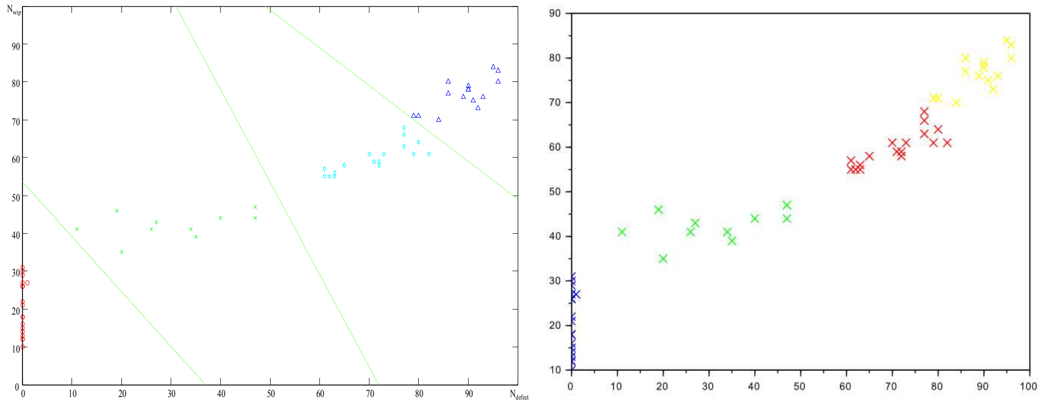


Figure 3: Result obtained with a 4-Means algorithm and a Euclidian distance

4 Which control rules for each cluster?

An experiment design involves that several physical or numeric experiments have to be made. Settings have to be chosen and the values they can take have also to be set up. All of the possible settings combinations have to be tested, and then results have to be analyzed. The chart shows each of the settings and highlights the selected values for the experiments and a curb for every interaction (an interaction is the representation of the combined effects of 2 setting). After that, with a simple read of the chart, the best solution to adopt can be deduced.

The definition of levels for each factors is described in Table 2.

Table 2: Levels for factors in the experimental design factors

Factors	Levels		
NDefect	Low (30%)	Average (45%)	High (60%)
NWIP	Low	Strong	Medium
Control Rule	EDD	FIFO	SPT

The produced simulations take into account all of the combinations of the 3 values of N_{wip}, N_{defect} and the control rules. The obtained results about the number of overdue on the production chain allowed to establish a design of experiments and (Figure 4 shows the factor effect graphs), this one allows to visualize which monitoring rule is the best to apply for a given configuration of N_{wip} and N_{defect}. This analysis is the key to finally associate the « best » or the less bad possible monitoring rule to every cluster.
To associate control rules to the map established previously, we made an experiments plan with the challenge to identify the influence of variables N_{defect}, N_{wip} and the monitoring rule on the system.

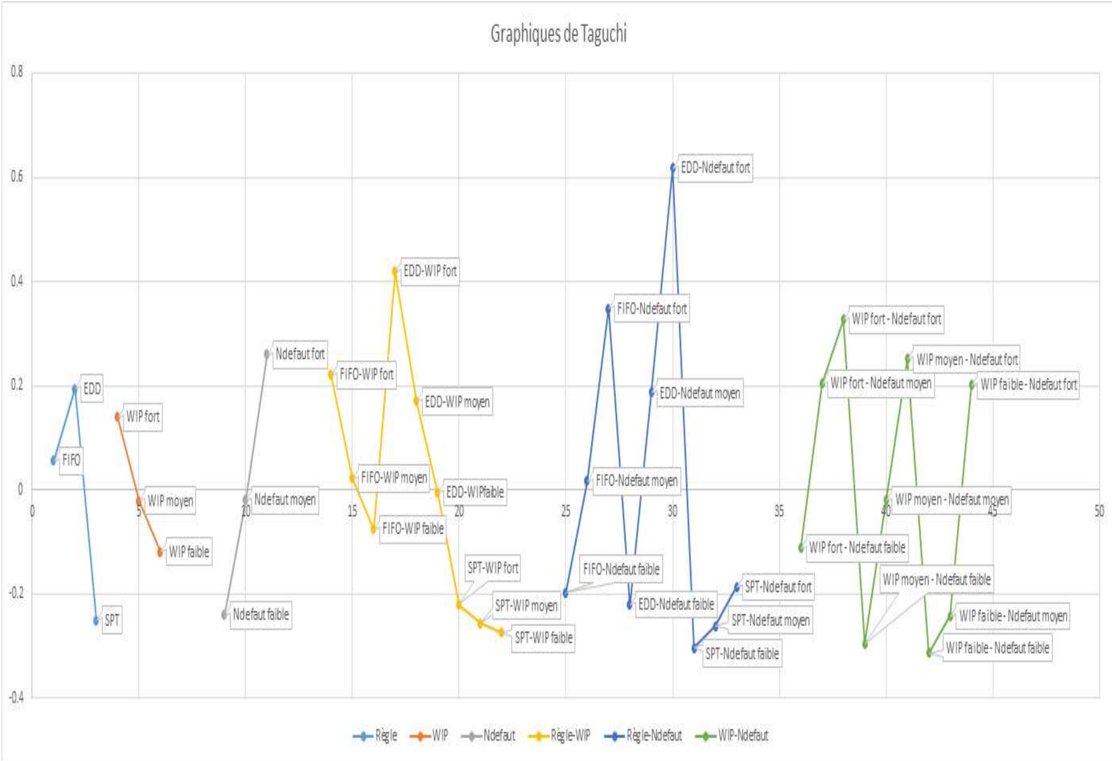


Figure 4: Experimental Design with management flow rules, N_{wip}, N_{defect} as factors and their interactions.

The first 3 curves show the impact of three factors individually. Analyzed alone, they confirm that the best case, for not having delays, uses SPT rule in conditions of low Work in Process and with few defects. However, when looking at interactions (3 right curves), we realize that in conditions of Work in Process and repair rate imposed, the most suitable management flow rule is not always the same. We can construct the table 3 having the most appropriate rule to a position given by the value of the two indicators considered.

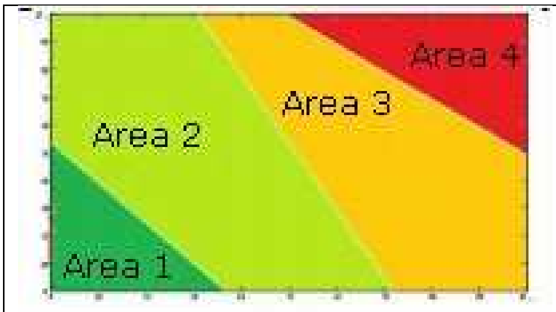
Table 3: Best control rule regarding to the imposed indicator values

		N_{Defect}		
		30%	45%	60%
Wip	High	SPT	SPT	SPT
	Average	EDD	SPT	SPT
	Low	EDD	EDD	SPT

The corresponding limit for the changeover between EDD and SPT is the central green curve in Figure 3. The other two curves are only state of saturation but does not justify the passage of a control rule to another one. It may instead be associated with common sense warnings or recommendations as shown in Table 4.

From an Acta-Mobilier's operator point of view, the most appropriate rule for the operation in the workshop is the FIFO rule with a batch processing in the order of their arrival, without any additional information. The EDD rule also applies easily since the departure dates are clearly marked on organized pallet. This way the operator can choose any. However, the SPT rule is more difficult to implement because it depends on the working time which must be known before choosing the next work to be performed. This information which is available in the ERP may not be displayed directly on the follower sheet items because there is one per workstation / phase range. The implementation of the latter rule therefore requires the presence of an efficient information system. One can possibly imagine the presence of this information directly on the product itself, thanks to Auto-ID technologies, which will return it at the right time by consensus or with neighboring products.

Table 4: Work recommendation depending on workshop saturation areas

	Area 1	Area 2	Area 3	Area 4
Advocated Monitoring rule	EDD	EDD	SPT	SPT
Recommendation		Increase rework workshop capacity	Assign someone to the follow-up of the pieces behind schedule	Define batches to prioritize

The recommendation of the Area 2 aims to quickly reduce the number of ongoing repairs and thus move the point representing the state saturation of the workshop to the left to return in Area 1. Once the situation returns to Area 1, the workstation can return to a normal capacity for the repair.
The recommendation of Area 3 has the objective of act more on the number of delays. The combined action on delays and repairs will tend to move the point simultaneously down and to the left with the objective to reach the Area 2 rapidly.
The recommendation of Area 4 rather palliative as the situation is serious. It tries to "save" lots of great importance in treatment of temporarily abandoning the others. This area, materialized in red on the map, is ideally never reached.

5 Conclusions

This paper addresses a challenging issue for factories suffering from a high rework rate. The proposed approach has as objective to follow and to give a help decision tool in order to decrease the impact of rework on the factory performances. The main contribution is to combine two methods: Clustering and Experiment plan. Based on actual data from Acta-Mobilier and an ARENA simulation model, four clusters of data that correspond to four workshop saturation zones were identified. The use of an experiment plan performed on the same data, leads to identify two best rules for different situations and distribute them on four areas. The main advantage of this method is its simplicity and the visual aspect suitable for intuitive use on the shop floor. It is clearly oriented towards the easy and fast application, and to communication and empowerment of workers. Based on such approach and the use of PDS concept, an auto-control system and dynamical scheduling flows could be created. Indeed, this work refers to a bigger project trying to develop a production flow monitoring helping system based on the Product Driven System concept and real-time visuals indicators combination. This way, the product and the map based on the indicators described before will become an agent of the global system. Together they will impact on the scheduling and the dynamical re-scheduling by influencing the management flow rule to use and the scheduling horizon.

References

1. Abouzahir O., Gautier R., Gidel T.: Pilotage de l'amélioration des processus par les coûts de non-qualité. 10ème séminaire CONFERE, pp. 3-4. (2003)
2. Feigenbaum A.V.: Quality control. Mc Graw-Hill. (1951)
3. Hastie T., Tibshirani R., Friedman J.: The elements of statistical learning, data mining, inference and prediction. Springer. (2001)
4. Jain A.K, Murty M.N., Flynn P.: Data clustering: a review. ACM Computing Surveys, 31, 264–323. (1999)
5. Kuhl M.E, Laubisch G.R.: A Simulation Study of Dispatching Rules and Rework Strategies in semiconductor manufacturing. IEEE Conference on Advanced Semiconductor Manufacturing, ASMC '04. (2004)
6. Love P. E.D., Li H., Mandal P.: Rework: a symptom of a dysfunctional supply-chain. European Journal of Purchasing & Supply Management, Vol. 5, pp. 1-11. (1999)
7. Mittler M., Schoening A.K.: Comparison of dispatching rules for semiconductor manufacturing using large facility models. Proc. of the 31st Winter Simulation Conference. (1999)
8. Noyel M.: Contrôle intégré du pilotage et de la qualité des produits, application à la société Acta-Mobilier. Thèse de doctorat de l'université de Lorraine. (2015)
9. Noyel M. Thomas P., Charpentier P., Thomas A., Brault T.: Implantation of an on-line quality process monitoring. 5th International Conference on Industrial Engineering and Systems Management IESM'13, Rabat, Maroc, 28 – 30 octobre. (2013)
10. Skidmann A., Schweitzer P. J., Nof S.Y.: Performance evaluation of a flexible manufacturing cell with random multiproduct feedback flow. International Journal of Production Research, Vol. 23, pp. 1171-1184. (1985)
11. https://atoms.scilab.org/toolboxes/nan/1.0
12. Steinbach, M., Karypis, G., & Kumar, V. A comparison of document clustering techniques. In KDD workshop on text mining (Vol. 400, No. 1, pp. 525-526). (2000)
13. Jain, A. K. Data clustering: 50 years beyond K-means. Pattern recognition letters, 31(8), 651-666. (2010)
14. Goupy, J. Pratiquer les plans d'expériences. Paris. Dunod. (2005)
15. Poirier, J. Analyse de la variance et de la régression. Plans d'expérience. Ed. Techniques Ingénieur. (1993)
16. Mohanasundaram, K. M., Natarajan, K., Viswanathkumar, G., Radhakrishnan, P., & Rajendran, C. Scheduling rules for dynamic shops that manufacture multi-level jobs. Computers & Industrial Engineering, 44(1), 119-131. (2003)