PRODUCTION LINE MODELING AND BALANCING: COMPARISON OF EXISTING TECHNIQUES AND PROPOSAL OF A NEW METHODOLOGY

Alessandro Silvestri ^(a), Cristina Cerbaso ^(b), Gianpaolo Di Bona ^(c), Antonio Forcina ^(d), Vincenzo Duraccio^(e)

(a) (b) (c) (d) Department of Civil and Mechanical Engineering University of Cassino and Southern Lazio
(e) Niccolò Cusano University
00166 Roma (RM) – Italy
tel. 0039-776-2994330 - Fax 0039-776-310812

^(a)silvestr@unicas.it, ^(b)c.cerbaso@unicas.it, ^(c) dibona@unicas.it, ^(d) a.forcina@unicas.it, ^(e)vincenzo.duraccio@unicusano.it.

ABSTRACT

In industrial plants, the heads of production often allocate tasks to operators without following a real methodology, but basing on experience; therefore, there are many problems when a quickly changes of capacity or constraints occur.

In the literature, there are not techniques able to provide unique and effective solutions. The starting point of this study is to explore the main methods of production line balancing, understanding their advantages and disadvantages in assigning tasks to workers with the same amount of time. Well-balanced lines allow to improve productivity and to make production lines enough flexible to absorb external and internal irregularities. The result is a better utilization of workstations, machines and people.

The main aim of the present work is to propose a new technique of balancing, in particular for assembly lines, simple, general and effective.

Keywords: operator, task, saturation, algorithm

1. INTRODUCTION

In a flow shop production, it is important to ensure a balanced allocation of the operations among workstations: no overloaded workstations (bottlenecks) rather than shortly saturated. The balancing activity is essential to guarantee the performance of the whole line (in particular, the productivity), whose speed is due to the slowest station, at least in the case of "paced" lines (production rhythm imposed), fixed or continuous (respectively we have the product replaced or in transit in a station for the execution of all tasks assigned to each workstation, during the given time, called station or cycle time).

The criteria for the assignment of tasks to each workstation, are mainly based on efficiency and productivity objectives, fragmentation of operations and specialization of tasks. Therefore, the characteristics of machineries and equipment and the employees required skills are highly specific, in order to realize a short number of actions. In many case the process modeling through simulation techniques can be realized in order to achieve the production line optimization and improvement (Falcone *et al.*, 2005; Falcone *et al.*, 2010; Falcone *et al.*, 2013).

2. WORKLOADS BALANCING ACTIVITY

The balancing activity of a production line means, in general, to allocate the operations among the stations in such a way that the cycle time for each stage (sum of the times of the operations allocated to the station) is as uniform as possible (Aase *et al.*, 2003). In practice, however, the perfect balancing is almost never reached due to constraints such as:

- cycle time;
- space;
- allocation of operations;
- qualification of operators;
- incompatibility between operations;
- management of material flows;
- precedence relationships;
- use of the operators.

The workloads balancing is a very important issue, as it influences the productivity of a line. It is necessary both in the case of designing a new line and in the case of improving the performance of an existing one.

The first step consists in identifying the so-called "bottlenecks", defined as the phases of the process where there is the lowest production capacity (De Carlo *et al.*, 2013; Duraccio *et al.*, 2006). They are due to the accumulation of material to be processed in a single location, which causes a slowdown of the production time of the whole line, and, consequently, a production capacity reduction. The reasons could be the slowness of operators and machines, excessive paths length, non-optimal position of equipment and material, that cause a loss of time due to activities of walking and picking (Falcone *et al.*, 2011; Iannone *et al.*, 2007).

3. BALANCING TECHNIQUES

Although many studies have been carried out over the years to try to get solutions to the problem, simple, effective and unambiguous, there is not yet an analytical method able to consider all the different aspects of the problem.

Those difficulties are mainly due to the different structural configurations of the modern industrial plants, which doesn't permit to study a generic approach and solution. Nowadays, also the wide range of realized products contributes to complicate the balancing activity.

Therefore, a large number of models and solutions allowing to solve specific problems of balancing, or, however, a little number of problems, according to their particular application, have been proposed.

However, if there are not suitable models to be applied, customized models are developed, as needed, by the managers who are experts in the production line.

Below the most common methods existing in the literature will be analyzed and a new technique will be proposed, in order to maximize the saturation of the operators and the utilization of the workstations. The new methodology integrates and improves existing methodologies, allowing to obtain an unique configuration of an assembly line (Askin and Zhou, 1997).

3.1. Salveson's Criterion

Sometimes, there are simple problems of balancing production lines. In particular, it happens when there are not typical basic constraints described above, commonly present in practice, especially precedence constraints between operations.

In this specific case, the optimum solution is achieved simply by minimizing the number of stations of the line, choosing the best combinations of operations able to reduce the operator/machine cycle time and the number of utilized stations.

In this case, *Salveson's criterion* is helpful. The iterated procedure allows finding the smallest number of workstations that meets the production requirements.

3.2. Kottas - Lau Method

The *Kottas - Lau* algorithm is a very valid criterion for the industrial reality. This methodology allows to obtain a good balancing degree of the line and seeks to optimize the total cost of production. The algorithm considers the operations execution times as stochastic variables and describes them through a normal distribution with mean value and standard deviation assigned.

The categories of considered costs are:

- labor costs: decrease increasing the amount of work assigned to each operator, thanks to the reduction of the number of workers needed to complete the processing cycle;
- costs of non-completion: if many tasks are assigned to the operator, the likelihood he does

not complete his work and, consequently, the related costs increase.

The correct balancing of a production line permits to reach the best compromise between the two types of costs listed above, such as to minimize the total cost of production (Figure 1). The trends of costs are shown in Figure 2.



Figure 1: Block diagram of *Kottas – Lau* algorithm - I_k is the completion out of line total cost of all the operations which, due to precedence constraints, can not be performed if the k-th operation is not completed.



Figure 2: Costs Trends

3.3. LCR and KWM

The acronyms LCR and KWM, or *Largest Candidate Rule* and *Kilbridge & Wester Method*, identify two methods often used in industrial practice to balance production lines, solving relatively simple problems thanks to their simplicity of application. In particular, they use a single parameter, the operations execution time, and the diagram of precedence to assign tasks to stations. Even if the balancing activity is based on only one key factor, the precedence constraints between operations and the maximum cycle time of the station have always to be taken into account.

3.3.1. Largest Candidate Rule (LCR)

In the *LCR* method, the iterative process used to search for the balancing solution is only based on the task time, that is the execution time of the considered operations. The procedure is described below and is particularly quick and easy to be applied:

- 1. build a table containing the tasks, their execution time and their immediate predecessors;
- 2. order the operations according to the execution time, placing those with longer duration at the top;
- 3. assign tasks in the obtained order and in accordance with the priorities and the maximum cycle time the station supports;
- 4. repeat the previous step until the available operations are finished.

3.3.2. Kilbridge&Wester Method (KWM)

Heuristic methods can be used to solve the balancing problem (Scholl *et al.*,2006). The *KWM* method is a heuristic algorithm that selects the operations to be assigned to stations, simply according to their position in the diagram of priorities. This method solves a problem of the *LCR* procedure, where the operations at the end of the diagram of priorities may be the first to be assigned just because the value of their execution time is greater.

In the *KWM* method, the operations are positioned in the columns, according to the diagram shown in Figure 3, and, subsequently, are organized and ordered in a table (Table 1) according to their column belonging.



Figure 3: Diagram of priorities

Tabel 1: KWM operations table

Operation	Column	Task Time (min)	Total Task Time of Column (min)
1	Ι	6	6
2	II	2	
3	II, III	4	
4	II, III	9	
5	II, III	2	17
6	III	2	2
7	IV	3	
8	IV	6	9
9	V	5	
10	V	5	10
11	VI	3	3

If an operation can be included in more than one column, all the columns where the operation appears, must be specified in order to show the portability of that specific operation. Once the table has been completed, the iterative procedure for the tasks assignment is performed, starting with those belonging to the first column until it reaches the maximum cycle time.

If the operations are not assigned in a column due to exceeding the maximum allowable cycle time, it will need to open a new station. Therefore, it cannot be assigned operations belonging to a next column if there is still a previous one containing unassigned operations.

3.4. Ranked Positional Weight Method (RPWM)

The *Ranked Positional Weight Method* is a heuristic technique for lines balancing, introduced and developed by Helgeson and Birnie.

As the name implies, this method is based on the so-called "positional weights", consisting of coefficients assigned to all the operations to enable their assignment to the line stations.

The positional weight of an operation is the sum of its task time and all other operations directly following in the diagram of priorities. For example, referring to Figure 4 which shows a simple graph of priorities, the positional weight of operation 2 is given by the sum of its execution time and those of operations 4 and 5 that follow number 2.



Figure 4: Example of diagram of priorities

Subsequently, the operations are included in a list and ordered according to the decreasing positional weight.

The method assigns to open stations at first the activities with a greater positional weight, always

respecting the maximum cycle time that station can support. If it exceeds that time cycle, a new station will be opened.

According to the heuristic *RPWM*, the greater is the number of operations available for the assignment to the various stations, the greater will be the probability that at least one of them is suitable for a particular workstation. Following this logic, it's possible to exploit the most of the time cycle of each workstation, finding appropriate operations in terms of execution time. This implies a smaller number of stations to be opened and a minimization of associated downtime.

3.5. Heuristics COMSOAL

The Computer Method of Sequencing Operations for Assembly Lines, alias COMSOAL, is another heuristic method able to find possible solutions, but not necessarily optimal, also to balance lines with a great number of operations. The COMSOAL method uses a search strategy of assignable operations which allows to obtain a large number of admissible solutions. Randomly and not according to a predetermined criterion it selects and assigns the available tasks. This approach is useful when the line optimization requires solutions which are not necessarily limited only to obtain the minimum number of stations (Becker *et al.*, 2006; Boysen *et al.*, 2008).

3.6. FABLE algorithm

The *FABLE* algorithm is structurally similar to *COMSOAL*. Its application involves the construction of a "tree" of possible solutions that also includes the optimal one.

To expedite an acceptable solution, it builds a tree branch at a time. Such construction is called "laser search". In addition, the *FABLE* algorithm uses a technique called "backtracking", usually used to look for solutions in different areas, where it is necessary to comply with specific constraints.

This technique, applied in tree structures such as the FABLE algorithm ones, develops the branches of the solutions, keeping track of all the nodes and branches previously visited, so as to be able to go back if the process in the current branch does not lead to valid results.

3.7. Elmaghraby criterion

A further criterion, developed by Elmaghraby, is very similar to the *Ranked Positional Weight method*, about logical characteristics and operational.

The heuristic *Elmaghraby* is particularly well suited to solve the balancing problems of simple lines. The development of the algorithm is mainly based on two key points:

- construction of the matrix of priorities;
- assignment of operation position coefficients.

The matrix of precedence "P" consists of the elements p (h, k), which, in case of precedence

constraint between the operations "h" and "k", take the value p (h, k) = 1, whereas in case of absence of sequential constraints, take the value p (h, k) = 0.

The position coefficient, however, is equivalent to the positional weight of *RPWT* criterion, given by the sum of the execution time of the operation i (i = 1...n) and the time of all following operations bound to it.

3.8. Imposed Operators Maximum Degree of Saturation Method

The method called *Imposed Operators Maximum Degree of Saturation Method* considers as binding, the relationships of precedence between the operations and considers also a coefficient, indicating the maximum degree of saturation that a station can support. It defines the *degree of saturation* of the operator (*DS*):

$$DS = \frac{\sum_{i=1}^{3} T_i}{TC}$$
(1)

where:

- T_i is the execution time of the operation "i";
- S is the set of operations assigned to the operator;
- TC is the cycle time.

This methodology is based on the respect for the inequality:

$$DS \le \alpha$$
 (2)

and

$$0 < \alpha \le 1 \tag{3}$$

taking into account the precedence constraints between operations, where α is the imposed degree of saturation. The logic used in the algorithm is summarized in the block diagram shown in Figure 5.



Figure 5: Imposed Operators Maximum Degree of Saturation Method - Block diagram

In the Block diagram, the set A contains the operations already assigned, while the set B contains the operations to be assigned.

4. PROPOSAL OF A NEW METHODOLOGY

The proposed innovative methodology, named *Maximum Degree of Saturation Method (MDS)*, has as its starting point the *Imposed Operators Maximum Degree of Saturation Method* and aims to saturate workstations as much as possible, minimizing a variable that keeps track of the time available to be committed, in accordance with the precedence constraints and other restrictions.

In particular, this methodology could be used for production lines processing a single type or different types of products. An important goal is the reduction of the desaturation percentage and, consequently, the number of stations.

Also in this case, such as for *Imposed Operators* Maximum Degree of Saturation Method, the Degree of Saturation DS is defined as:

$$DS = \frac{\sum_{i=1}^{S} T_i}{TC}$$
(1)

Maximum Degree of Saturation Method (MDS) finds applications in assembly lines for which the movement of an operator or a robot from a workstation to another is a constraint (Agnesis *et al.*, 2003).

In this case, the maximum available time to complete all transactions within each location corresponds exactly to the "takt time", dictated by the market demand.

Figure 6 shows the block diagram of the proposed algorithm operation.

The application of the *Maximum Degree of* Saturation Method is quite simple.

First, it starts by opening a new station, being empty the set A of already assigned operations. Then, it identifies the set B of the possible operations sequences, respecting the precedence constraints. Selecting these combinations, it must bear in mind that the available operations are those for which the previous activities have already been assigned to a workstation or have already been taken into account in a possible location.

At this point, it calculates the *Degree of Saturation* (DS) for each identified combination and it compares with the Imposed *Degree of Saturation* α . The choice of the value α (between 0 and 1) must take into account the constraints and it is also necessary to remember that a high degree of saturation reduces the cost of the line, but increases the cost of non-completion (in this sense it is possible to integrate also the considerations arising from the *Kottas-Lau Method*).

All combinations of operations not satisfying the relation:

$$DS \le \alpha$$
 (2)

will be excluded, while the others will be taken into account.



Figure 6: Block diagram of functioning of Maximum Degree of Saturation Method

After selecting combinations of operations satisfying both precedence constraints that the relation, one wonders about the possibility of having two or more sequences of activities with the highest degree of saturation.

If not, it assigns to the current workstation the combination of operations with the highest DS and, later, if the set A still contains operations, it opens a new workstation, otherwise the balancing operation ends.

If so, it calculates the positional weight of each transaction, as the sum of the times of the single operations following the activity, including the operation in question:

$$PW_i = \sum_{i=1}^{n} TP_i \tag{4}$$

This approach refers to a technique previously presented, namely the *Ranked Positional Weight Method*, in which the operations are assigned to workstations based on the values of the positional weights: the operation to be assigned is the one that presents a positional weight greater than the others, always respecting the precedence relations.

To give priority to these operations means to reduce the non-completion impact when the production line stops for various reasons.

In particular, stopping the production flow, the operations (in accordance with the precedence diagram and therefore with the technological and assemblability constraints) with a greater execution time could be completed, while the remaining ones, with lower execution times, would be completed out of line at less cost.

Continuing the explanation of the algorithm logic shown in the block diagram, after calculating the positional weight of each operation, it assigns the sequence containing the task with the highest positional weight and then it opens a new workstation.

In case there were multiple combinations of operations with the same saturation degree (the highest), and operations with the same highest weight positional, it gives precedence to the combination containing the operation with the second highest weight positional; whereupon it opens a new station.

If also the latter is equal between two or more stations, it chooses the sequence of operations with the third highest positional weight and so on.

Then it will open a new workstation and then it will continue with the production line balancing. This approach is reasonable, since it prefers to terminate within the line operation in which are related activities engaging more time to be completed, rather than one to which follow operations less onerous in terms of time. This is explained by noting that in the event of a production line shutdown, if it need to finish the job and it has the possibility to complete operations out of the line, surely the activities with a lower duration will have a lower cost than operations with higher running times.

Note that it may happen that not all combinations of operations to be compared for the highest positional weight contain the same number of activities. For example, it might have a combination with a single operation, and another one with two. In this case, both having the same saturation degree and the same highest positional weight, the analyst should look for the sequence with the operation that has the second highest weight, but it is evident that the sequence composed of a single task would be automatically excluded. If a situation like this occurs, it will be taken into account the combination presenting the activity to which is linked the largest number of operations. This reasoning stems from the fact that non-completion of this operation within the line involves higher costs because it should be carried out of the line all the other activities in addition to it and it might need more labor, equipment, resources, etc.

In case the last open station is forcibly shortly saturated, it might delete an already assigned task from a workstation with a very high degree of saturation (in accordance with the precedence constraints) and assign it to the last station. In this way, it would perform a workloads balancing to avoid that there are stations with a very high degree of saturation, and a station (the last one that has been opened) with a very low degree of saturation. Obviously, a situation like this is inconceivable in real cases, because it would mean, in the case of manual production line, one of the operators with little work, i.e. with a very high percentage of desaturation.

The latter observation is very important because it must always be remembered that balancing a production line means allocate the various operations in such a way as to balance the workload of each operator, minding the goal to pursue.

There will be cases where it will be evident the saturation lack of the last open station. Other times, however, it is uncertain about what to do and the decision will be taken by the person making the line balancing, based on his own experience and on the production cycle knowledge.

Wanting to make a direct comparison between the examined methods, it can refer to Table 2, which summarizes the main features of each technique, including the developed new one.

	Preceden ce relations	Cycl e time	Producti on costs	Takt time	Backtrac king Task	Same probabil ity of tasks selection	Probabilis tic distributi on of task time	Generati ng multiple solutions
Kottas Lau	\checkmark	\checkmark	\checkmark	\checkmark	-	-	\checkmark	-
K&W	\checkmark	\checkmark	-	\checkmark	-	-	-	-
LCR	\checkmark	\checkmark	-	\checkmark	-	-	-	-
RPW M	\checkmark	\checkmark	-	\checkmark	-	-	-	-
Comsoal	\checkmark	\checkmark	-	\checkmark	-	\checkmark	-	\checkmark
Fable	\checkmark	\checkmark	1	\checkmark	\checkmark	ŀ	-	\checkmark
Salveson	-	\checkmark	-	\checkmark	-	-	-	\checkmark
Elmaghra by	\checkmark	\checkmark	-	\checkmark	-	-	-	-
Imposed Operators Maximum Degree of Saturatio n Method	V	V	-	\checkmark	-	V	-	V

Table 2 – Comparison between the different balancing techniques

Maximum								
Degree of								
Saturatio					-	-	-	-
n								
n	V	V	N	N	-	-	-	-

The Maximum Degree of Saturation Method (MDS) provides as output a single solution, which turns out to be the optimal one, as it aims to minimize the coefficient of balancing delay. Furthermore, for a not excessive number of operations to be placed within the workstations, it may be applied manually. That makes this technique quite practical in the business reality.

As shown from Tab. 2, all of the discussed techniques, including the one developed, present as key points the cycle time and the takt time, ie the cycle time available within each station.

This last parameter is very important as it determines the production rate, set depending on the market demand.

In general, it is very difficult, if not impossible, to state with certainty that a method is better than another in searching the optimal solution for the problem of balancing a production line.

Surely the developed technique combines the easiness of application to the effectiveness of the result.

5. CONCLUSIONS

The main features of the *Maximum Degree of Saturation Method* are the following:

- it is easy to be applied, since at each step it only requires the verification of compliance with the precedence constraints and relation (2). Also, the evaluation of operations positional weights is very easy, being a simple sum of the activities task time;
- for a limited number of operations, the technique does not require to be implemented using the computer, since it is even easy to be used manually;
- using deterministic times, it does not consider the randomness of the operations execution times. This might seem like a disadvantage, but many business realities are based on unique values of operations task time, deriving from studies carried out by Times and Methods office. These deterministic times then become imposed times for the operator to perform his activities;
- requires a proper choice of the coefficient α (which not necessarily must be the same for all the workstations);

The *Maximum Degree of Saturation Method* has considerable advantages, as:

• it determines a unique and optimal solution, by selecting in advance the best combination of operations which saturates the available cycle time for each workstation;

• it reduces the number of stations to be opened and the costs associated with this operation, having regard that the number of stations appears in the numerator of the *coefficient of delay balance* formula, defined as:

$$D = \frac{\left(N_{s} \times TC_{\max}\right) - T_{k}}{\left(N_{s} \times TC_{\max}\right)} \times 100$$
 (5)

where

- N_s is the number of stations obtained by applying a specific balancing technique;
- TC_{max} is the maximum cycle time of the line, among all stations;
- T_k is the total task time of the operations;
- it presents a method for choosing the operations to be assigned to workstations based on multiple levels, giving priority to saturation and then to the activity position within the priorities diagram, integrating multiple techniques in a single balancing procedure (*Kottas Lau; Ranked Positional Weight Method; Imposed Operators Maximum Degree of Saturation Method*).

A disadvantage in the use of this technique arises when the operations to be placed in the various stations are quite numerous. In this case, the searching of the combinations of operations available to assign is very laborious, since the number of sequences achievable grows with the number of activities.

The next step will be the application of the developed method (MDS) to real case studies in order to validated is goodness and generality.

REFERENCES

- Aase, G.R., Schniederjans, M.J., Olson, J.R., 2003. U-OPT: An analysis of exact U-shaped line balancing procedures. *International Journal of Production Research*.
- Agnetis, A., Alfieri, A., Nicosia, G. 2003 Part batching and scheduling in a flexible cell to minimize setup costs. 2003. *Journal of Scheduling 6 (1), pp. 87-108.*
- Askin, R.G., Zhou, M., 1997. A parallel station heuristic for the mixed-model production line balancing problem. *International Journal of Production Research*.
- Becker, C., Scholl, A., 2006. A survey on problems and methods in generalized assembly line balancing. *European journal of operational research*.
- Boysen, N., Fliedner, M., Scholl, A., 2008. Assembly line balancing: Which model to use when? *International journal of production economics.*

- De Carlo, F., Arleo, M.A., Borgia, O., Tucci, M., 2013. Layout design for a low capacity manufacturing line: A case study. *International Journal of Engineering Business Management*. Volume 5, Issue SPL.ISSUE
- Falcone, D., Duraccio, V., Silvestri, A., Di Bona, G., 2005. Technical and economical analysis of the layout of a palletization plant through simulation techniques. Summer Computer Simulation Conference 2005, SCSC 2005, Part of the 2005 Summer Simulation Multiconference, SummerSim 2005, pp. 387-391.
- Duraccio, V., Falcone, D., Silvestri, A., Di Bona, G., 2006. Improvement of performances of an optical system for defectiveness survey in a company of the automotive field. *Summer Computer Simulation Conference 2006, SCSC'06, Summer Simulation Multiconference, SummerSim'06*, pp. 314-318.
- Falcone, D., Silvestri, A., Di Bona, G., Forcina, A., Pacitto, A., 2010. Study and modelling of very flexible lines through simulation. (2010) *Proceedings of the 2010 Spring Simulation Multiconference - Emerging M and S Applications in Industry and Academia Symposium, EAIA, pp.* 11-16.
- Falcone, D., Silvestri, A., Forcina, A., Pacitto, A., 2011. Modeling and simulation of an assembly line: a new approach for assignment and optimization of activities of operators. MAS (The International Conference on Modelling and Applied Simulation), Rome, 12-14 September 2011.
- Falcone, D., Silvestri, A., Di Bona, G., Forcina, A., Cerbaso C., Duraccio V., 2013. Technical and economic verification of the convenience in reengineering a production line using simulation techniques. 25th European Modeling and Simulation Symposium, EMSS 2013, pp. 327-332, September 25-27, Athens, Greece. Code 100159.
- Iannone, R., Miranda, S., Riemma, S., 2007. Supply chain distributed simulation: An efficient architecture for multi-model synchronization. *Simulation Modelling Practice and Theory*, pp. 221-236, Volume 15, Issue 3.
- Scholl, A., Becker, C., 2006. State-of-the-art exact and Heuristic solution procedures for simple assembly line balancing. *European Journal of Operations Research*.