INTERNAL LOGISTICS OPTIMIZATION IN AUTOMOTIVE MODULAR MANUFACTURING SYSTEMS

Antonio C. Caputo
caputo@ing.univaq.it

Mario Palumbo
palumbo@ing.univaq.it

Pacifico M. Pelagagge
pelmar@ing.univaq.it

Federica Scacchia
scacchia@ing.univaq.it

University of L'Aquila, Faculty of Engineering, 67040 Monteluco, L'Aquila, Italy

Abstract. *In this work discrete event simulation techniques are utilized to examine the internal logistics issues in cellular manufacturing systems with specific reference to the automotive sector, where the production system results from the integration of functionally independent Elementary Technological Units (UTE). After characterization of logic and material flows, and taking as a reference the facility of a major Italian car manufacturer to gather the necessary field data, the system criticalities are identified and discussed. Finally, in order to satisfy the required performance levels several corrective approaches under both the managerial and capacity increase point of view are suggested and compared. As a result a greater knowledge of internal logistics issues in cellular environments is obtained and several ways to solve some of common problems to the logistic management in widespread cellular manufacturing systems are presented.*

Keywords*. Cellular manufacturing, internal logistics, automotive industry, discrete event simulation.*

1. Introduction

The automotive industry has undergone several major transformations during the last decades. Until the seventies, traditional taylorism-derived concepts dominated the design of manufacturing systems. During the '80's a major change occurred through the introduction of hard automation including heavily robotized plants equipped with automated guided vehicles and computerized numerical control workcentres. During the 90's the concept of highly integrated plants pushing the limits of Computer Integrated Manufacturing paradigm, which heavily resorted to CAM technologies, gradually evolved introducing the concept of lean manufacturing. This happened following the succesful adoption of Just In Time (JIT) techniques in Japan leading to a manufacturing practice to be known worldwide as the Toyota manufacturing system which advocated a major focus on quality, waste minimization, and the decentralization of responsabilities. At the end of the nineties the transformation was almost completed and the new paradigm was the modular factory where many of the auxiliary functions such as logistics and maintenance were outsourced leaving the mother firm with the task of final assembly, and the responsibility of overseeing the overall manufacturing process and guarenteeing standards, costs, quality, lead times and preserving the integrity of the corporate brand. This revolution paved the way, under the influence of Japanese thinking and the positive experiences of Group Technology in job shops environment, to the advent of cellular manufacturing.

Cellular production systems enable a rationalization of material flow in the production system with the possibility of several related benefits which in overall contribute to the increase of process performances as pointed out by several Authors (Agarwal et al., 1998; Farrington et al., 1998; Billo, 1998; Block, 1983; Burbidge, 1992; Monden, 1998, Shingo, 1989). Several Companies actually witnessed a decrease of setup time, WIP inventory and material handling costs, with consistent increase in product quality, material flow, machine and space utilization besides reduced rework percentage and improved employee morale as described in a survey by Wemmerlov et al. (1989).

A further evolution of this concept is now represented by the decomposition of the manufacturing system in autonomous manufacturing entities called "Elementary Technological Units" (UTE) which operate sinergistically to execute the manufacturing process. UTE are single independent organizational entities which sequentially cellularize an otherwise linear production process. Therefore, even in the context of a single-piece flow the manufacturing process is not the result of a sequence of single operations performed on a single production line but is the sequence of processes carried out in separate but strongly integrated and interdependent manufacturing cells (cellular manufacturing) each one considering the subsequent one as a customer. Each UTE is thus an autonomous manufacturing environment, performing a subset of the entire production process, provided with its managerial staff and with specific duties and responsibilities including the management of human resources inside the UTE (active operators involvement, quality assurance, training, supervisioning, safety management, job rotation etc.). Each UTE is responsible for its quality level, its maintenance and the uninterrupted flow of material while its main duties encompass process execution, management of the technical equipment, quality control and logistic tasks such as continuous line feeding, inventory management, and material handling.

In this paper the logistic problems arising in a manufacturing system organized according to the UTE paradigm are analyzed in order to highlight criticalities and suggest performance improvement measures. The analysis is carried out with reference to the automotive sector, where this kind of manufacturing system has been pionereed and is more deeply rooted. In the paper at first the general interactions among the UTE and the logistic resources (internal transportation systems, warehouses, loading docks, personnel and auxiliary equipments) are examined in order to define an overall framework. Subsequently this approach is specialized with reference to an automotive plant located in central Italy owned by a major Italian car manufacturer in order to obtain the actual field data. Following an accurate data collection and analysis phase, the logic scheme of the process is then defined. In particular the logic followed when making material calls to the warehouse and the way in which the warehouse fulfils demand have been analysed. Furthermore, the activities and process times of all logistic resources have been detailed. Afterwards, a detailed computer simulation model is developed which includes the order generation process, the warehouse management and the internal transportation tasks. The model is then used to highlight criticalities of the present logistic system and to individuate improvement strategies, also assessing their effectiveness in a comparative manner.

2. Material flows and internal logistics processes in lean manufacturing

In lean manufacturing systems adopting cellular layout, different materials management criteria may be adopted. In the present case materials are classified in six distinct groups. *Calls* for materials and replenishment of line storage areas is generally carried out by the UTE manager, who is in direct contact with internal warehouse and external suppliers according to the actual production schedule.

JIT Flow: materials directly supplied in a JIT manner by external firms located nearby the plant.

Sequenced (SEQ) Flow: model-specific materials whose demand is strictly linked to the sequence in which the various car models are assembled on the line. Such materials directly feed buffer storage areas along the line and are sequentially picked according to the assembly sequence. Supply is through fork lifts unloading external trucks arriving each two hours and remaining in dedicated external parking areas.

Tight Flow: Materials arrive daily through external trucks remaining parked in specific areas on the basis of daily orders from the UTE manager. UTE fork lifts unload periodically the parked trucks to feed the line, and load empty containers. The frequency of container rotation is about 2 hours.

Slowly Rotating Flow (SRF): materials stored in the warehouse and retrieved on the basis of specific calls from the UTE. A call implies the print out of an order sheet (OS) in the warehouse, assignment of the OS to a picker, picking by warehouse personnel utilizing fork lifts and temporary storage on the warehouse output bay, transportation by tractor trailer to an UTE exchange storage area (where an exchange of full and empty containers occurs), and moving to the UTE line by UTE fork lifts. The frequency of containers rotation is about 4-8 hours.

Fixed Frequency Flow (FFF): materials stored in the warehouse and moved at fixed frequency (usually at night) to the UTE in order to replenish daily the missing parts. Usually they consist of small sized independent demand items (screws, gaskets, bolts, etc.). The operator checks inventory in the gravity shelf along the line where they are stored, picks the missing materials at the warehouse and replenishes the line buffers.

Highly Rotating Flow (HRF): Cumbersome and large sized highly rotating materials directly stored in visually monitored storage areas along the line. Line is replenished by UTE fork lifts which exchange empty and full containers.

Table 1. Material handling modality.

Material class	Picking T = from truck W = from warehouse	Order frequency F = Fixed C = on call RT = Real time	Transportation FL = Fork lift T = Trailer	Type of material call OS = Order sheet V = visual AS = assembly sequence
JIT	-	RT	FL	AS
SEQ	T	C/RT	FL	AS
TIGHT	T	F/C	FL	-
SRF	W	C	FL/T	OS
HRF	W	-	FL/T	V
FFF	W	F	FL	-

- = Not applicable or not relevant.

Materials classification is carried out according to several factors: size, the number of different codes, the availability of space for storage along the line, the parts commonality with the various models assembled and the requirement for assembly sequencing, besides the supplier lead times. The choice among sequenced and JIT flow, apart from economic considerations depends from the ratio of the delivery lead time respect the anticipation lead time of the line sequencing decisions. Usually the unit loads are moved in specific kind of containers which are owned by the main firm but are utilized by external suppliers and the internal logistics subcontractors. Material handling, therefore, implies also exchanging full and empty container with the external suppliers vehicles and/or accumulating empty containers waiting to be reutilized by the suppliers in specific park areas. Each UTE has available specific parking lots or loading/unloading docks for external trucks carrying items to the UTE. Table 1 resumes the materials classification and the adopted handling practices. From Table 1 and the above description it follows that materials are picked from either external trucks at their respective parking areas / unloading docks or from some internal warehouse (a central warehouse WH, acting as a consolidation center, or from along-the-line visually controlled open air warehouses). Materials to/from the UTE are handled resorting to fork lifts (FL) while trailers are used to move empty containers and to connect the UTE areas to the central warehouse. Main material flows among a generic UTE in the assembly area and the logistic infrastructure are thus schematized as shown in figure 1.

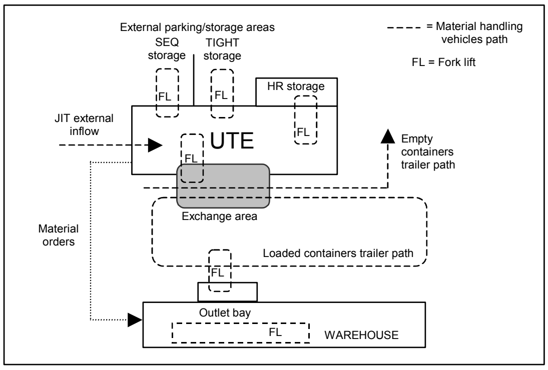

Figure 1. Scheme of the material flows structure.

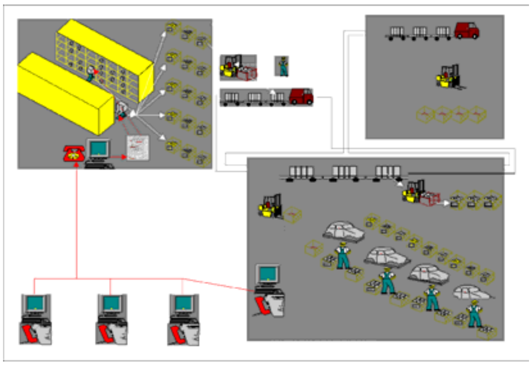

Figure 2. Example of the animated graphical interface.

SRF materials are supplied by the warehouse following a call (the order) from the UTE. An order issued at a computer terminal in the UTE results in a order sheet (OS) printed out in the warehouse. After order picking by WH personnel, generally resorting to FL, the material is deposited in the WH retrieval bay associated to each UTE, to wait for the trailer moving loaded containers to the UTE area. In proximity to the UTE, in a proper exchange area, the material is unloaded from the trailer and moved to the line resorting to FL, while empty containers are loaded on the trailer to be taken away towards collection area and finally reutilized at either the WH or at the external supplier site. Shared resources are utilized for SRF, HRF and TIGHT codes as far as the recovery of empty containers process is concerned while each materials class utilizes dedicated resources for handling of loaded containers. FL and trailers are then univocally associated to specific UTE or to the WH. In the considered manufacturing system the OS reports the due material delivery time at the line. Deliveries may be scheduled after 30, 60, 120, 240 minutes from the time of order issue according to the assigned order priority level (I, II, III, IV).

In this material handling scenario the most critical issues are those related to the supply of slowly rotating materials. In fact while other classes of codes are directly accessible to the UTE, because the warehouse or the trucks unloading bays are nearby the UTE and FL operated by UTE personnel directly handle the materials, the slowly rotating or fixed frequency codes are stored in the central warehouse which is located apart from the building housing the assembly UTE, and they are moved to the UTE area resorting to dedicated tractor trailers. Moreover an interaction with the UTE FL is required to load/unload the trailers and to move loads from the trailers to the line. In greater detail the handling process is as follows: UTE personnel monitors parts inventory at the line stations and via computer terminal issues OS to the WH assigning them a priority level. OS are printed out at the WH and warehouse personnel (two operators with FL) retrieves the requested material accumulating the containers on the outlet bays according to their final destination. Each UTE is assigned a bay and each bay is served by a trailer. The trailer driver utilizes a FL to load the ready containers waiting at the bay on the wagon of its trailer and drives the trailer back to the destination area in the shop. Once arrived at the UTE area the loaded trailer is uncoupled from the tractor and is then unloaded by dedicated FL, while the UTE FL provide to move the material towards the line. If an empty trailer from a previous travel is available in the exchange area it is coupled to the tractor and taken back to the WH. Empty containers coming from the UTE are accumulated in a separate area to be collected by a further trailer moving them to the external accumulation area.

3. Simulation approach

In this work the attention is focused on the body assembly facility of an automotive manufacturing plant. The facility is further subdivided in three functional areas: sheet manufacturing, painting, and assembly, while the production units are arranged in 40 distinct UTE, 25 of them in the assembly area. About 1200 cars/day are assembled on a two-shift organization with a throughput time in the assembly area of about 12 hours. However, the study focuses on a specific sub-section of the assembly area including 16 different UTE served by two distinct trailers, with a tractor dedicated to loaded containers and another one to empty containers, sharing three separate wagons, a single central warehouse for slow rotation and FFF codes, equipped with two FL, and two visual control open warehouses for highly rotating cumbersome codes. However, the central warehouse is common to all UTE. The selected area, while being representative of the logistic issues common to other UTE, is also particularly critical because during assembly a large percentage of codes is slowly rotating and stockouts would cause immediate line stop. In order to evaluate the performance of the internal logistics, to identify criticalities and to suggest possible improvement strategies, a simulation study was thus carried out resorting to the ARENA discrete event simulation language (Kelton et al., 1998).

The developed model has been built on four distinct submodels: order generation, warehouse operation, full containers trailer, empty containers trailer.

3.1 Order generation submodel.

This section simulates the generation of orders from the various UTE to the central WH. Considering that more than 5000 codes are managed and that they are frequently updated, a simulation of each single code generation was unfeasible, also being scarcely useful because the utilization of material handling resources is fairly indedependent on the peculiar code handled as far as containers of similar size are utilized. An analysis of the global number of orders generated during a shift by each UTE, and their priority distribution has been then carried out. It should be noted that orders coming from all UTEs have been considered to simulate the actual WH resource utilization because the WH serves all the UTE, while all the other resources are dedicated to the group of 16 UTE taken in consideration. Order generation is simulated resorting to an unstationary Poisson process, utilizing an exponential distribution of interarrival times. To account for the widely variable frequency distribution of orders during a shift, orders are ideally issued at the maximum rate and their probability of actually being released to the shop floor and entering the WH depends on the ratio of the actual average arrival rate in the considered time period to the maximum arrival rate (*thinning process*). Orders generated by different UTE are then combined in a single queue represented by the WH printer.

3.2 Warehouse operation submodel

This model simulates the printing of the OS related to incoming materials calls, the assignment of OS to pickers, the container picking carried out by one of the two existing FL operators, and the final delivery of containers to the WH outlet loading bay. Orders should be fulfilled according to their priority. However, it has been observed that, owing to the fact that the majority of orders are issued with top priority, WH operators treat the orders simply on a First In First Out basis, therefore assigning the same priority to all orders. Experimentally derived service time delays distributions have been utilized for the picking operation and the outlet bay loading operation. Resource utilization for orders coming from UTE other than those subject of the simulation is accounted for as already mentioned.

3.3 Full containers trailer submodel

The trailer transporting full containers is in charge of delivering the containers made available at the warehouse loading bay to the assembly area UTE. The warehouse has five bays, dedicated to different sections of the facility. The simulated UTE are uniquely served by a single bay. Therefore, no interference exist at bay level with the operation of the remaining of the plant. The trailer is composed of a tractor and several wagons. The wagon resource, three of which are currently available, is shared with the trailer conveying empty containers from the UTE to the external accumulation point. However, utilization of the wagons from the full containers trailer is given a priority in order to avoid disruption of line feeding. Transportation phases include: coupling to the tractor of one or more empty wagons in proximity to the UTE loading/unloading area, travel to the WH, loading of the wagons at the WH loading bay, travel to UTE area, uncoupling of the wagon, unloading of the wagon resorting to local fork lifts, waiting for an available empty wagon to be coupled to the tractor and start of a new cycle. Number of transported containers may vary, but loading at the WH may not exceed the capacity of the available wagons, while a travel towards the UTE is started only when a minimum of containers is available to reduce the number of travels.

3.4 Empty containers trailer submodel

This model generates empty containers in the assembly area and simulates their transferral to the accumulation point resorting to the dedicated trailer. The tractor, as already said, shares the wagons resource with the full loads trailer. The working cycle is as follows: coupling of a loaded wagon to the tractor in the assembly area, travel towards the accumulation point, waiting for the unloading to be completed, travel back to the assembly area, uncoupling of the empty wagon. Usually one wagon is utilized by the full loads trailer, one by the empty loads trailer and the other is seized by the first available tractor provided that the priority constraint is satisfied. The empty loads trailer travels always with a predefined number of empty containers, therefore the travel starts as soon as the required number of empty containers is available. In order to monitor system status and performances, apart from the resource utilization automatically computed by the ARENA model, a number of performance variables have been identified: length of queue and orders waiting time at WH printer, number of orders already late at the time of OS printout, average delay at the printer output subdivided by priority level, length of printed orders queue waiting to be served by WH operators and printed orders waiting time, average delay after picking subdivided by priority level, number of full containers deposited at the WH outlet bay past the due delivery time, queue length and waiting time at the WH outlet bay, number of loaded container delivered at the assembly lines, number of late deliveries at the lines, average delivery lateness subdivided by priority level, number of full containers transported by the trailer in each travel, number of stockouts at WH bay (the times the trailer had to wait for the minimum number of containers to be ready for transportation).

As the performances are heavily influenced by the transient but cyclical nature of the shifts, long run average of the performance parameters were considered to be scarcely significant and the use of terminating simulations has been made. Replications have carried out including a warm up period corresponding to one shift in order to make the system start with a typical number of full containers waiting at the WH outlet bay so that the containers usually remaining at the end of a shift and present at the start of a new cycle are represented instead of resetting the system totally totally. A proper number of replications (about 200) has been specified in order to obtain results within specified confidence limits. Model validation has been performed by observing the model animation (figure 2), by tracing the entities flow during simulation and by verifying the overall consistency of the simulated values and the experimentally observed ones with reference to the number of containers handled during each shift, the number of containers delivered to the assembly lines, the number of containers at the WH outlet bay at shift end, the number of travels of the trailers, and the delivery lateness at the UTE, obtaining satisfactory results.

4. Data collection and model characterization

While developing the model a gata dathering campaign was carried out in order to characterize the material flows and to evaluate service times and duration of handling activities. At first the monitoring of slowly rotating codes calls to the WH during each shift was performed in oder to assess the frequency of orders to the WH and the percentage distribution of their priority level. Obviously a high percentage of urgent requests may indicate an undersizing of the material handling equipment or an uncorrect order release practice. Materials priority is determined at line level by the amount of inventory of components to assemble that can be stored or are at the moment available on the shop floor along the line. Materials available in amounts enabling an autonomy of half a shift or more are given the lowest priority. In average it resulted that 67% of orders were top priority (level I), 17% and 13% were respectively high and medium priority (levels II and III) while only 4 % of orders were low priority (level IV). The order arrival rate was monitored over 15 time buckets of 30 minutes showing a highly uneven pattern with a very high concentration of orders in the first hour of work as shown in figure 3. This causes a saturation of WH and internal transportation resources which leads to the impossibility of satisfying the delivery times and prompts line operators to further increase the percentage of urgent orders. Moreover, this leads to an accumulation of WIP during the first hours of each shift. The distribution of process times and material handling variables have been determined on the basis of experimental observation as shown in Table 2. In particular the warehousing operations have been investigated (OS printout, material retrieval, trailer loading times etc.). It has been also monitored the delay of materials delivery respect the due time as computed from the order issue time and its priority class. This has been carried out in particular for a specific UTE, in which more than 90% of materials were slow rotation codes supplied from the WH.

Table 2. Experimental distribution of significant material handling variables.

Parameter	Distribution	Mean	Dev std
Time required by WH printers to release the OS printout	0.35 + LOGN(0.106, 0.0417	0.457 min	0.0495
WH order picking time by FL operator	NORM(1.88, 0.577)	1.88 min	0.577
Time to load a container on the trailer	0.16 + LOGN(0.863, 0.344)	1.02 min	0.326
Number of loads on a trailer travel	5.5 + WEIB(6.12, 1.56)	11	3.52
Delay in material delivery at the UTE	20 + 180*BETA(1.41, 1.26)	115 min	46.9
Interarrival time of empty containers to the trailer	GAMM(1.53, 1.26)	1.92 min	1.84
Trailer parking time at the accumulation point for unloading empty containers	2 + WEIB(15.8, 2.03)	16 min	7.18

Table 3. Values for other material handling variables.

Parameter	Value
Trailer travel time (shop area to WH)	3.5 min
Time for coupling/uncoupling the trailer	0.5 min
Time for unload the trailer at the UTE	10 min
Number of trailer travels per shift (loaded containers)	[8,13]
Trailer travel time (to/from empty containers storage area)	5 min
Capacity of the empty containers trailer (according to container size)	[21,25]
Number of containers remaining at WH bay at end of each shift	[0,13]

Deterministic point values or ranges of uniform probability have been instead determined for the other parameters as shown in Table 3. No data related to failures and repair times have been included because of the large number of back up vehicles available in the facility.

5. Model evaluation

In order to evaluate the actual system performances in terms of Work In Process (WIP), number or orders late and delivery lateness, throughput time and resource saturation, a number of simulations have been made adopting the current capacity levels and management practices.

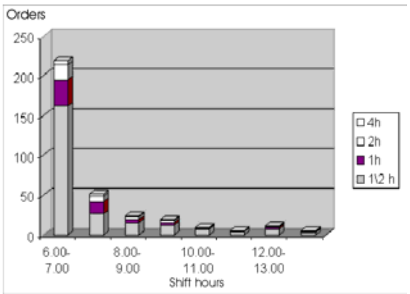

Figure 3. Trend of order frequency during a 8-hr shift, according to priority level.

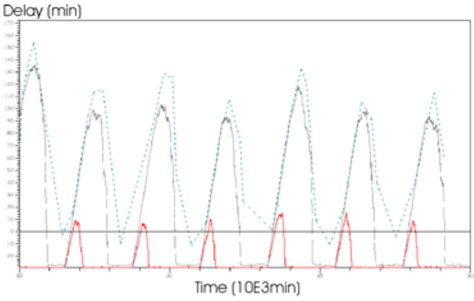

Figure 4. Trend of delays (red = printing, black = picking, green = UTE arrival) for top priority orders.

As a result the following criticalities were pointed out:

- excessive delay in orders fulfilling,
- high WIP,
- non homogenous resource utilization during the work shift, with oversaturation in the first hours and undersaturation in the remaining hours.

As an example Figure 4 shows a typical trend of delays at OS printing, warehouse picking and at UTE delivery in the case of top priority orders which should be served within a 30 minutes time span. Figure 5 instead shows tipical WIP curves of OS waiting to be printed, printed orders waiting to be picked and containers at the WH outlet bay.

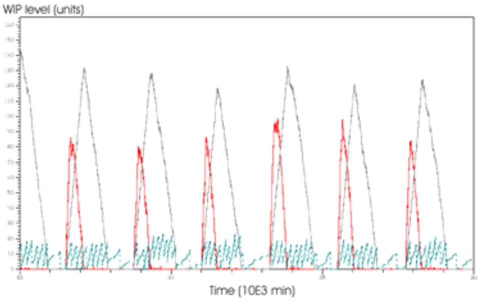

Figure 5. Trend of WIP (red = at printer, black = at picking, green = WH outlet).

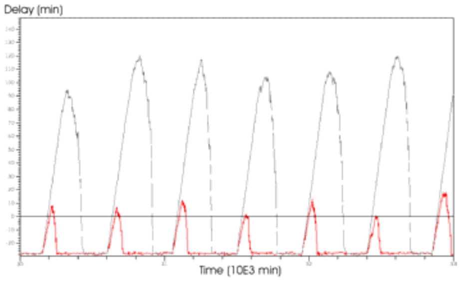

Figure 6. Delay after picking of top priority orders (black = 2 pickers, red = 5 pickers).

As far as delays are concerned it can be pointed out that the delay at the WH is the sum of delay at OS printing, delay due to picking operations and delay due to waiting at the loading bay (caused by absence of the trailer and its loading operations), while the delay in arrival at the UTE is the sum of delays at the WH plus travel time, other delays during transportation, delay associated to unloading and transferral to the line.

It is evident that delays and WIP are concentrated in the first hours of the shifts when the system is oversaturated by order calls from the assembly lines, especially for urgent orders. This also shows that the system is slow (high WIP) and the capacity is apparently undersized. Responsibility of the situation is also the uncorrect practice of concentrating orders at shift start and giving them the status of high priority as experimentally verified (figure 3). However such amount of delays is not tolerable especially in a lean manufacturing environment because it may put at risk the continuity of the assembly operations. Furthermore, being the logistic services outsourced, the manufacturing plant may issue penalties to the logistic provider for each order late.

6. Analysis of single effects improvements

After highlighting the existing criticalities, several changes have been suggested either from the managerial point of view and under the resource sizing aspect in order to reduce process times and increase the capacity. The effects of such modifications have been examined resorting to the simulation model in order to assess the system response and to evaluate their effectiveness. The following measures have been analysed in detail:

Capacity criteria

- Resource augmentation (WH printer, WH fork lifts, trailers wagons)
- Automated WH

Managerial criteria

- Logic of orders scheduling at the WH
- Trailers management
- Rules of order launch from UTE
- Rules for WH fork lift assignment
- Special rules for urgent orders

6.1. Resource capacity increase: WH printer

To eliminate the excessive queue of orders to be printed at the WH terminal, which delays the start of the picking operation, a possible solution is the substitution of the current printer with a faster one. In fact it may happen that an order is already late as soon as it is printed. A printer with an average printing time of 10 s per order is hypothesized. Nearly real time printing is thus possible with practical elimination of the queue and negligible printig delays of no more than few minutes. However, the picking operation now results to be the bottleneck. In overall the situation is not actually improved as no decrease of the delay at WH outlet is observed. The benefit got from a new printer could be enjoyed only if an increase of pickers throughput may be obtained, otherwise it would only consist of an improved visibility of the real time order status of the assembly plant.

6.2. Resource capacity increase: WH pickers increase

Previous analysis showed that even if the capacity bottleneck at the printer terminal may be easily removed, order pickers capacity, which would be correctly utilized only in case orders arrived at an uniform rate during the shift, is undersized to cope with the peak of activity experienced during the first part of the shift. As an example the number of OS waiting to be picked grows up to around 170 while the waiting times total up to 150 minutes per order in the more congested periods. In fact while the peak orders arrival rate is about 300 per hour, the average pickers velocity is 75 orders per hour. An increase in the number of pickers from 2 to 5 is thus analysed. This measure is highly effective, in fact the queue of orders to be picked is nearly eliminated. It follows that the delay cumulated after picking and the number of orders leaving late the WH is strongly reduced. Figure 6 as an example shows the average delay after picking for the top priority order. Delays are in average negative, while the maximum delay at peak hours is in the order of less than 10 minutes as compared to more than 100 minutes in the case of only 2 pickers even if it is not eliminated at all (an average of 65 top priority orders still result late). The situation of lower priority orders is even better with always negative delays. However, as in the previous case, this implies the botleneck is shifted at the loading bay owing to insufficient capacity of the transportation system, witnessed by a significant WIP increase (figure 7) and correspondingly a higher waiting time of containers at the loading bay. However, the utilization of the trailer increases as the times the trailer remains idle at the WH waiting for the minimum number of containers to accumulate is reduced from an average of 7 to an average of 2 per shift. This has a positive effect because the delay at the time of arrival of materials at the UTE is somewhat reduced. The added cost of operators and forklift per shift is about 450 Euro, and at current car manufacturing rate it translates in an added logistic charge of about 0.9 Euro/vehicle. Therefore the economic justification of this measure in case a penalty is applied for each order delivered late should be verified comparing the saving connected to the reduced number of late deliveries to this added cost.

6.3. Resource capacity increase: increase of number of wagons of the full loads trailer

In the real system and during the simulation it has been observed that the available wagons transporting full and empty containers are nearly totally saturated. This causes a delay of the tractor in the UTE area because it has to wait for the wagons to be available in order to couple them and start a new cycle. Adding a new wagon (i.e. passing from 2 to 3) enables to nullify this wait. However, if the WH throughput is not increased this does not result in a lower delay of delivery at the UTE. On the contrary a slight increase of the number of times the trailer is idle at the WH bay is observed. However, addition of a new wagon is much cheaper than adding a further tractor in the current situation.

6.4. Automated warehouse

It is also hypotesized to reduce the throughput time of the warehouse by adopting different picking/storage criteria and automated data management to eliminate the OS printing and assignment to the picker phases and simultaneously reduce the average picking time from 1.88 minutes to 1.4, 1 or even 0.6 minutes. Figure 8 shows, in the case of priority I orders, the progressive reduction of the delay cumulated by the orders after the picking operation is terminated. However, even if the results are interesting this solution is particularly capital intensive.

6.5. Logic of orders scheduling at the WH

Orders issued to the WH are associated with a priority level which defines the time bucket in which they should be fulfilled. However, since about 64% of total orders are concentrated in the first hour of a shift and 66% of them are top priority, the WH has adopted the practice of fulfilling orders in FIFO logic, neglecting the priority indication. It has been therefore investigated what would happen should the priority logic had been strictly followed, therefore fulfilling orders according to their due time determined by the launch time at UTE level and their priority class. Figure 9 and 10 show the average delay of highest priority orders (fig. 9) and lowest priority (fig. 10) comparing the situation in which the FIFO criterion or the earliest due time is adopted. It results that adoption of the earliest due time rule slightly reduces the average delay of top priority orders but significantly increases that of low priority orders which, being previously fulfilled early on average now may also result late. Therefore no net improvement may be expected simply changing the criteria adopted by the warehouse to fulfill orders.

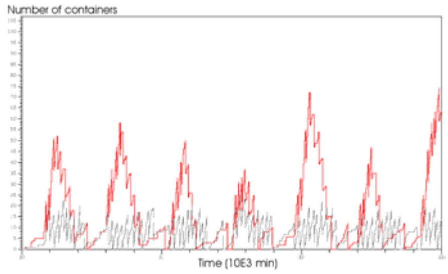

Figure 7. WIP at WH outlet loading bay (black = 2 pickers, red = 5 pickers).

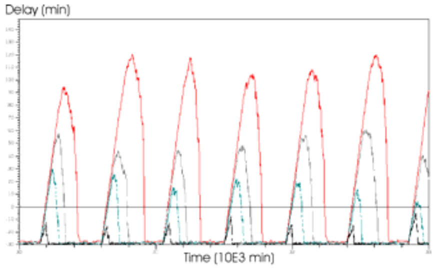

Figure 8. Delay of top priority orders at end of picking phase in the WH (average picking time per load: red = current situation, grey = 1.4 min, green = 1 min, black = 0.6 min).

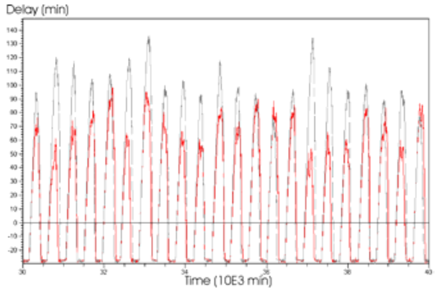

Figure 9. Comparison of average delay of high priority orders (black = FIFO, red = earliest due time).

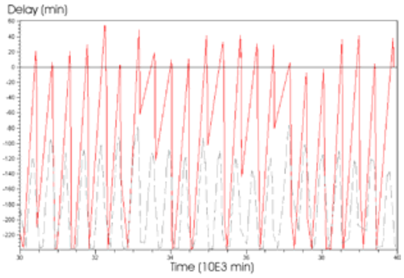

Figure 10. Comparison of average delay of low priority orders (black = FIFO, red = earliest due time).

6.6. Management of trailers cycles

It has been already pointed out that the transportation of containers from WH to the UTE introduces further delays, apart from the transportation time, connected to the availability of the trailer. It has also been stated that a part of the delay the trailer experiencies at the start of a new cycle may be reduced increasing the number of wagons. However, further improvements may be pursued adopting different strategies for trailer cycles management. Major delays in fact come also from the fact that the trailer driver at the WH has to load by himself the ready containers on the vehicle. It is thus suggested to adopt a further wagon as already hypotesized, totalling 4 wagons shared with the trailer of empty containers, and modify the full loads trailer management by charging WH pickers to load containers directly on the wagon instead of delaying the trailer cycle as happens when this duty is assigned to the trailer driver, leaving an empty wagon at the WH bay in order to be loaded by WH pickers when the trailer is executing its transportation towards the UTE. Therefore the new cycle, enabled by the addition of a further wagon which, actually, decouples the cycles of the full and empty loads trailer, is as follows. The tractor of the full loads trailer couples an empty wagon in the UTE area and travels to the WH. Arrived at the WH uncouples the wagon which remains there to be subsequently loaded by WH pickers, couples the wagon previously left at the WH and loaded in the meantime by the WH pickers, starting a new travel towards the UTE. When arrived at the UTE it uncouples the loaded wagon and couples the unloaded one starting the cycle again. According to this criterion the trailer cycle time is reduced and the frequency of travels increases and, as the overall number of containers transported remains unchanged, the number of transported containers per travel is reduced as is the WIP at the WH outlet bay (figure 11). The delivery delay at the UTE is also reduced. Utilizing more than four wagons is not justified because even with four wagons the loading capacity of the wagons is underutilized. As a results the following benefits have been assessed on the basis of the simulation: the average waiting time of the containers at the WH bay is reduced by 11 minutes, containers queue length at WH outlet bay is reduced by more than 3 units (nearly a 50% cut). As far as the delivery delay of top priority orders at the UTE is concerned the average number of late deliveries (now 83) is reduced by 6 units and the average delay time is reduced of 10 minutes respect the previous value of 70. No significant benefit is instead observed for the other priority classes. As an added economic benefit a saving concerning the elimination of the fork lift utilized by the trailer driver to load the containers is got.

6.7. Rules for order generation at the UTE

In figure 3 the typical pattern of order generation at UTE level has been shown. The concentration of orders at start of the shifts causes saturation of the entire logistic system, long delivery delays and excessive WIP, while resources remain undersaturated during the secon half of the shift. Furthermore, high priority is assigned to most of the orders so that the WH personnel can no longer discriminate among really urgent orders and the entire priority criterion is vanified. Surely the human factor in this case is paramount, infact two distinct factors cause this trend of order emission: at first the line manager fears that material shortage could stop the line. Therefore, believing that the logistic chain is unreliable, he is prompted to accumulate early all the required inventory. Second, the logistic personnel at UTE level (fork lift operators) prefer to advance as much as possible their work load in the first half of thee shift. Proper education of operators will be invaluable in solving this mental approach. However, a new role could be introduced: that of the *orders solicitor*, an operator in charge of controlling the actual inventory at the line and consequently issuing orders to the WH, who has no direct interest of accumulating inventory earlier than required being independent from the UTE management. The cost of the *order solicitors* should be the compared with the savings from reduced penalties thanks to the lower number of late deliveries in order to assess the feasibility of this further measure.

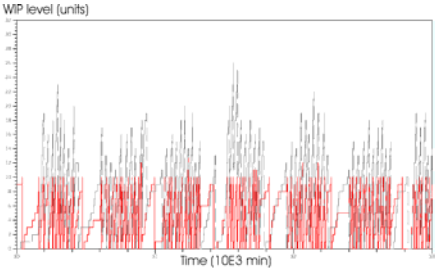

Figure 11. WIP at WH outlet loading bay with modified trailer cycle (black = old cycle, red = new cycle).

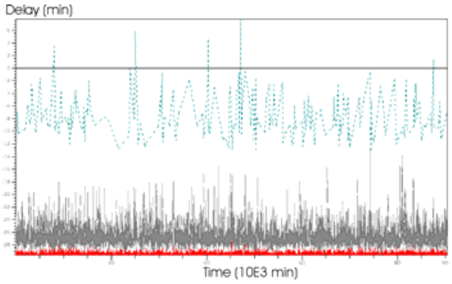

Figure 12. Delays with modified order release criterion (red = delay at printer, black = delay after picking, green = delay at UTE delivery).

In order to evaluate the system behaviour with a different pattern of order generation a criterion of balanced and evenly distributed order generation has been introduced:

- orders are issued at a constant rate (as constant is the assembly line pacing) over the entire shift duration (about 46 orders/hr result),
- only two levels of order priority exist: urgent orders (30 min delivery) and standard orders (120 min delivery);
- urgent orders, being the exception are only a small percentage (10%) of the total.

The effect of this “lean” order issue criterion is a drastic reduction of WIP at the WH (shorter queues at printer, and picking) and nearly total elimination of WH delays (figure 12), thus witnessing that current WH sizing is adequate to manage a balanced workload and that, since most orders are early (average advance on priority orders is 26 minutes, while on standard orders is 116 minutes), the WH could be actually managed in a JIT manner if a smooth flow of orders existed. In fact while the average order arrival rate is 46 orders/hr, the WH troughput is about 65 orders/hr. However, there is no significant decrease of *average* values of WIP and waiting time at the outlet bay, while the trend is substantially different: in the new system the istantaneous value are constantly around the average with a small coefficient of variation, while in the previous system the istantaneous values were strongly fluctuating still maintaining roughly the same average (high coefficient of variation). Accumulation of WIP and possible delivery delay of orders being early at the WH outlet bay is imputable to the trailer, which generates an average delay of about 20 minutes on all orders. In fact the average throughput of the trailer is about 35 containers/hr. The utilization of the trailer, however, changes somewhat: the number of travels increases slightly and the containers per travel reduce, while the number of times the trailer is idle at the outlet WH bay because there are not enough containers increases. At the same time the waiting time reduces so that the resource utilization is nearly the same (about 90%). However, the quick response of the WH roughly offsets the delay induced by the trailer bottleneck so that in average the urgent orders are delivered to the line 5 minutes early (figure 12) and only a small fraction of such orders is late. Standard orders instead are in average 90 minutes early at the line. In overall the system is slightly undersaturated, the resource utilization level is uniform over the shift length and orders are nearly constantly early, showing that this management leverage is the most effective of all corrective measures. As far as the *orders solicitors* required to guarentee this more correct order generation criterion are concerned, three of them would be necessary and a net economic gain would be obtained from reduction of late deliveries if the penalty for each late delivery was higher than 4.4 Euro.

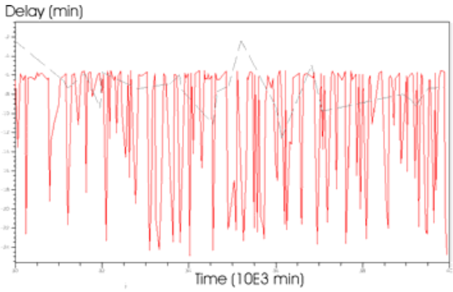

Figure 13. Urgent orders delivery delays with separate management (red = separate delivery, black = common delivery).

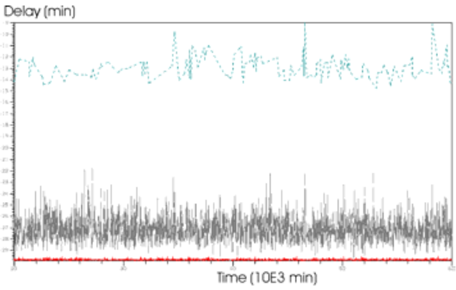

Figure 14. Delays with combined scenario (red = delay at printer, black = delay after picking, green = delay at UTE delivery).

6.8. Rules for pickers assignment

Provided that the level strategy is adopted for orders generation, the logic of assigning orders to WH pickers may be further rationalized. For example one of the two pickers may be devoted only to urgent orders, and the second one only to standard orders (the first picker may be assigned to standard orders only when urgent orders are lacking). This assignment criterion enables to totally eliminate any delay at containers delivery.

6.9. Separate management of urgent orders

A further improvement, following the new order issue criterion, may be pursued when the urgent orders are managed in a separated manner from the standard orders. Physical separation of the two material flows while still reducing the delay of urgent orders (actually further increasing the earliness) enables smoother management of flows, an improved visibility and traceability of urgent orders status and a greater flexibility in their handling enabling to cope more effectively with unexpected situations.

According to this further criterion the WH is equipped with two distinct printers, one for urgent orders and one for standard, the two pickers are dedicated to the two classes of orders as above, and a separate outlet bay is installed for urgent orders. While the containers of standard orders are transported by the same trailer, a smaller sized transporter (with a capacity of three containers), driven by the UTE foreman, is utilized exclusively to move the urgent containers. This dedicated transporter will wait at the loading bay to saturate its capacity unless the order earliness drops below ten minutes. Simulation results show that in average nine travels per shift will be required and that overall WIP level will decrease. No cost increase will result because the transporter is already assigned to the UTE foreman for his movements within the shop floor. Some simulation results are shown in figure 13. As expected the reduction of WIP at the standard order bay implies changing also their delivery delay, i.e further increasing their earliness.

7. Performance analysis of an optimized solution

From the previous analysis results that managerial interventions are much more effective than capacity increase. Therefore an optimal approach could be that of revising the order release procedure, thus reducing the system variability, and increase tha capacity of the critical resources. This kind of superposition of effects will be investigated in the following by assuming an *optimized* logistic system characterized by:

- faster WH printer,
- modified full loads trailer cycle (4 wagons and loading by the WH pickers),
- modified WH pickers assignment (one assigned to urgent and standard orders, the other to standard orders),

operating in the revised reference scenario characterized by:
- constant order arrival rate,
- two priority classes (10% urgent and 90% standard orders).

As shown in figure 14 the delay performances are further improved respect the sole adoption od the revised order release procedure (figure 12). Moreover the cost of this solution is further decreased because the cost of *orders solicitors* is partially offset by the saving of the fork lift utilized by the trailer driver for loading. The robustness of this optimized solution may be then assessed hypothesizing variation of the manufacturing scenario assumed up to this point. This has been made changing the order generation pattern passing from an even distribution during the shift to an esponentially decreasing one, by changing the ratio of urgent to standard orders, by changing the temporal concentration of urgent orders within the shift duration and by increasing the production volume.

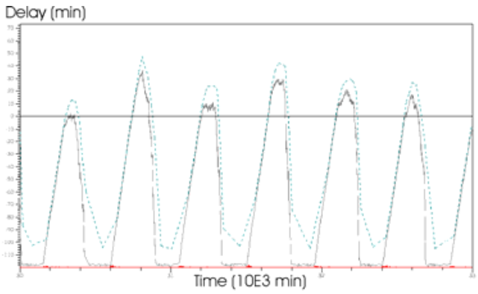

Figure 15. Standard orders delay with exponential generating rate (red = delay at printer, black = delay after picking, green = delay at the UTE).

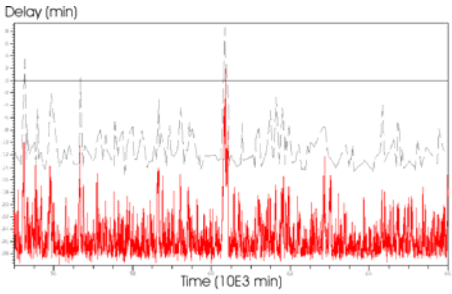

Figure 16. Delays of urgent orders (50% of the total) at UTE delivery (grey) and at WH outlet (red).

7.1. Effect of order generation pattern

When an order generation pattern having a peak at start of the shift and exponentially decreasing (as happens currently) is assumed instead of the uniform order generation rate suggested above, a delay in fulfillment of standard orders immediately follows (Figure 15), while urgent orders still behave satisfactorily with only some occasional delay limited to no more than 5 minutes. Figure 15 also shows that the main cause of delay is imputable to the saturation of the WH pickers. However, this may be solved by increasing the number of pickers. Simulations showed that passing from two to three pickers (one still preferentially dedicated to urgent orders) enables all standard orders to arrive at least 30 min early at the assembly line. Some occasional delay still may exist for urgent orders, even when the number of pickers is brought to four, but it may be eliminated by enabling all pickers to contribute to the handling of urgent orders (i.e. the three pickers serve both kind of orders with priority to urgent ones)

7.2. Effect of order priority

The ratio of urgent to total orders has been gradually increased from the hypothesized 10% to 50% still assuming a constant arrival rate. This change obviously leaves unaffected the standard orders handling. Simulation showed that delays at delivery start to show when the percentage of urgent orders rises above 40% which, in a correct production environment, is highly unlikely. Figure 16 as an example shows the allocation of delays at two stages of the handling process for an urgent orders percentage of 50%. The difference of the UTE delivery delay and the WH outlet delay is imputable to the trailer delay. Such delays may be eliminated, apart by speeding the trailer cycle, by assigning both pickers to urgent orders as verified through simulation. Summing up the optimized system may withstand a percentage of urgent orders up to 30% without introducing delays, beyond this limit some countermeasure should be gradually taken as an example by reassigning pickers.

7.3. Effect of urgent orders concentration

In this case the same number of urgent orders previously assumed (10% of the total) has been concentrated in time windows gradually reducing from 2.5 h to 0.5, instead of being considered uniformly distributed along the shift. According to simulations results, a concentration up to peak periods as short as 2 hours may be withstood by the system while for narrower time windows delivery delays at the UTE may show (Figure 17). In case of urgent orders concentrated in time spans shorter than 2 hours, particularly effective may be the adoption of the separate handling circuit for urgent orders with a dedicated transporter as previously discussed. This solution enables to nearly eliminate UTE delivery delays even for urgent orders concentrations periods 30 minutes long, giving in average an anticipation of urgent orders delivery respect the planned due time of about 15 minutes.

7.4. Effect of production increase

A production increse ranging from 10 to 50% has been finally hypothesized thus increasing the burden on the internal logistic system owing to a greater number of materials orders to be fulfilled and a larger number of empty containers to be managed. The consequent delays at UTE delivery of urgent orders are shown in figure 18. According to simulations the system remains stable only for production increases up to 30%. Beyond this limit the saturated resources cause a WIP exlosion and consequent a delivery delay owing mainly to the WH saturation, while the trailer may adapt itself to the increased production level by making more travels per shift and by transporting a higher average number of containers per travel. Greater production level increases may be faced by resorting to WH automation as previously analysed in order to reduce cycle time. However, a separate picking for urgent and standard codes with picking times reduced to 1.4 or 1.6 minutes per load may eliminate delays at UTE delivery even for 50% production increase (Figure 19). However, similar results could be also obtained by increasing the number of pickers of a single unit.

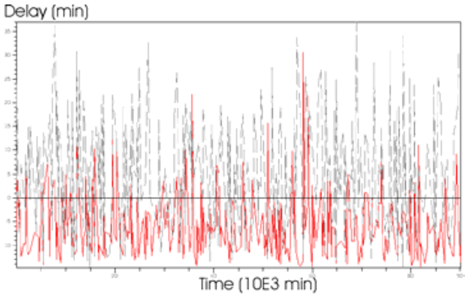

Figure 17. Delays of urgent orders at UTE delivery according to peak period duration (black = 0.5, red = 1 hr).

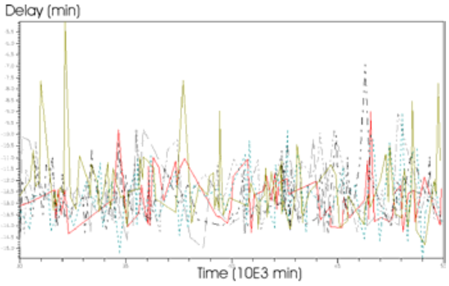

Figure 18. Delays of urgent orders at UTE delivery according to production increase (red = reference case, blue = +10%, dashed grey = +20%, grey, +30%, black = +40%, green = +50%).

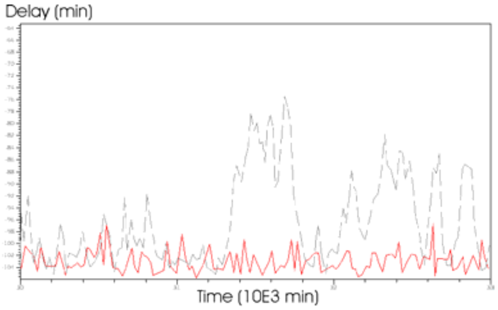

Figure 19. Delay at final delivery with a production increase of 50% (unit WH picking time: red = 1.4 min, grey = 1.6 min)

8. Conclusions

In this work discrete event simulation techniques are utilized to examine the internal logistics issues in cellular manufacturing systems with specific reference to the automotive sector, where the manufacturing system results from the integration of several functionally independent Elementary Technological Units. In particular an in depth analysis of the performance of an internal logistics system for the resupply of an automotive production line in a cellular manufacturing environment has been carried out. While the considered case study enabled to focus on the problems of a real system providing invaluable field data, the analysis has a general value because problems common to the internal logistic management in widespread cellular manufacturing systems have been encountered and solved.

The use of a simulation model enabled to thoroughly examinate material flows and system performance pinpointing critical issues and enabling to devise and compare improvement strategies. The effectiveness of such strategies has been also verified by changing the scenario conditions resorting to a sensitivity analysis. In particular the managerial logic has been greatly improved, especially acting on the order release logic from the cells to the warehouse, on the way the transportation resources are utilized and on the priority rules the warehouse adopts to fulfil orders.

Furthermore the effect of changing the capacity of logistic resources has been explored. Therefore useful insights have been obtained in the working of logistic system, while the management has obtained a useful tool for decision making. As a result a set of best combinations of managerial changes and capacity increases has been defined which may improve system's performances in a cost-effective manner and may be implemented at low cost, confirming the validity of the proposed methodological approach. It has been shown that for a lean supply system to work properly great care should be taken by managerial control to ensure that correct order generation patterns are followed, and that by minor changes in the service organization significant performance improvement and responsiveness to modified production scenarios may be obtained. The revised system implies mainly a change in the working logic of the warehouse and the transportation personnel, while a capacity increase, in particular of warehouse pickers, is only required when a strong increase of production level or uneven concentration of urgent orders are expected. As a result of the study a greater comprehension of the working of internal logistic systems in cellular manufacturing environments has been obtained and several practically applicable improvement approaches have been identified.

9. References

Agarwal, A. & Sarkis, J., 1998, A review and analysis of comparative performance studies on functional and cellular manufacturing layouts, Computers Ind. Eng., vol. 34, n. 1, pp. 77-89.

Billo, R.E., 1998, A design methodology for configuration of manufacturing cells, Computers Ind. Eng., vol. 34, n. 1, pp. 63-75.

Block, J.T., 1983, Cellular manufacturing systems reduce set up time, make small lot productions economical, Industrial Engineering, November, pp. 36-48.

Burbidge, J.L., 1992, Change to group technology: Process organization is obsolete, Int. J. Prod. Res., vol. 30, n. 5, pp. 1209-1219.

Farrington, P.A. & Nezemetz, J.W., 1998, Evaluation of the performance domain of cellular and functional layouts, Computers Ind. Eng., vol. 34, n. 1, pp. 91-101.

Kelton, D.W., Sadowski, R.P., Sadowski, D.A, 1998, Simulation with Arena, McGraw Hill.

Monden, Y., 1998, Toyota Production System: An integrated approach to just in time, 3rd ed., Engineering and Management Press.

Shingo, S., 1989, A study of the Toyota Production System, Productivity Press, Portland, Oregon, USA.

Wemmerlov, U. & Hyer, N.L., 1989, Cellular manufacturing in the US industry: A review of users, Int. J. Prod. Res., vol. 27, n. 9, pp. 1511-1530.