ELSEVIER

Available online at ww ScienceDirect

Int. J. Production Economics 112 (2008) 714–722

international journal of production economics

www.elsevier.com/locate/ijpe

Simulating an agile, synchronized manufacturing system

John G.H. Carlson[a,]*, Andrew C. Yao[b]

[a]University of Southern California, Los Angeles, CA, USA
[b]California State University, Northridge, CA, USA

Received 2 January 2006; accepted 1 June 2007
Available online 24 July 2007

Abstract

Retail customers are demanding more variety, more features and quicker order response times from manufacturers. Furniture production systems have had to become more flexible to respond to the variety of styles, fabrics and patterns offered by retailers. In this make-to-order environment, the orders received must be grouped into specific, logical *batches* whose short-cycle operations require close coordination and monitoring throughout the facility. For upholstered furniture, each batch may be unique because it consolidates orders having fabrics of different colors, texture or style. In recliner chair and similar production systems, the parallel component subassembly lines must maintain synchronization within each line and between the lines for components to *simultaneously* reach final assembly. The simulation developed represents an existing production system. It generates expected outputs under conditions of *operation variability*, *queue lengths* (*buffers*) and *batch changeover* (*set-up*) *times* over a range of 3 uniform and feasible batch sizes. Thus, the real-time *status and location* of components and subassemblies consigned to a specific production batch is essential for maintaining and improving quality and utilization of personnel, space, material and other resources.
© 2007 Published by Elsevier B.V.

Keywords: Just-in-time; Synchronized/parallel assembly; Agile production; Quick response; Uniform batch size; Simulation

1. Introduction

Simulation is the most robust and realistic way of evaluating the performance of a system of multiple queues. Its primary use is to test changes in a system *before* they are implemented. Combined serial and parallel queue disciplines are difficult if not impossible to be treated by analytical methods. According to Hall (1999), testing of different probability distributions and various parameter changes found in many production systems cannot be accommodated except by simulation.

Discrete object-oriented computer simulation has been used to identify and help solve problems in an ever increasing number of applications. The ongoing research on hundreds of assembly lines at General Motors by Alden et al. (2006) has led to many simulation models and observations that have saved millions of dollars. Simulation saves considerable time and money by viewing the dynamics of a system and providing insight into and a better understanding of those dynamics. Kline et al. cites the use of simulation as an operations research tool in analyzing a hardwood processing system that produced cabinets and similar products. The simulation helped illustrate the feasibility of alternative solutions by observing the *animated* flow of

---

*Corresponding author.
E-mail address: jghc1445@aol.com (J.G.H. Carlson).

0925-5273/$ - see front matter © 2007 Published by Elsevier B.V.
doi:10.1016/j.ijpe.2007.06.008

products through the processes. Simulation can also offer genuine excitement by pre-testing ideas and introducing realistic "what-if" changes in the parameters. As Keller et al. (1991) and Spedding and Sun (1999) concluded, simulation can also be useful in enhancing a cost accounting system by evaluating *manpower*, *space* and *equipment* requirements.

Enormous amounts of money continue to be spent by companies and industries to improve **small-lot** production. McRainey (1977) observed, as have others, that manufacturers are constantly being challenged by the demands of the distribution systems for quick response and just-in-time (JIT) requirements of customers. Manufacturers and certainly their marketing personnel, seek small-lot production with processes changed over quickly from one product to another to better serve customers. However, as Katayama and Bennett (1999) conclude per the classical economic models, i.e. EOQ/EMQ, an emphasis on agility must simultaneously focus on changeover costs when producing in smaller lots. Whitehead (2000) restates an underlying principle from the JIT concept that *agile*, small-lot systems can exist in concert with *lean* manufacturing systems. Both focus on reducing waste through lower inventory investment, space savings, better material handling, and reduced changeover and processing times. Thus small-lot sizes are fundamental to flexible JIT systems and enhance superior customer service.

The simulation study from Baykoc and Erol (1998) examined the performance of a multi-item, multi-line, multi-stage JIT system and demonstrated how the systems react under different circumstances. The variability of processing time and arrival demands of subsequent operations were studied. Sianesi (1998) demonstrated that the flexibility inherent in JIT production applied to "mixed-model" systems reduces WIP inventories in make-to-order environments.

The system described in this study is more complex in that the subassemblies are produced on separate but *parallel* lines and linked to a specific mixed-product *batch*. Also the operations must be synchronized within a relatively narrow time interval. Delays of any component batches may cause all production to slow or stop. The time for a unit or batch in the system will depend on the *maximum* of the various operation and waiting times and not just on their sum. This leads to more complicated queue disciplines. It is a requirement to finish each of the dependent operations at the same time. There is little value by completing an operation or a batch early only to wait for other parallel operations to be completed. In fact, it may be disruptive and wasteful of costly resources of space, personnel and equipment.

Simulation models do require empirical data, yet reasonable estimates or sample data are helpful in identifying the empirical data needed. In the system studied, estimates were used to help develop the simulation model and generate results approximating an existing production system. Stopwatch studies or video tape gathering of real-time data may discover other variables for which the simulation model does not accommodate. On the other hand, the model verifies the fundamental logic employed in managing the system and points towards areas where constant improvement, the company credo, can enhance profitability. Large lots may appear more economical but smaller lots or batch sizes leads to less waste of space, inventory investment and better customer service. Statistical analysis of empirical data may add refinement to the results but being able to manipulate the model and ask "what-ifs" appears to offer more of a contribution to understanding a complex system.

To represent an actual, interactively constrained production system by a discrete event, animated model is a challenge. The ability of a simulation to visually represent the flow, delays and projected throughput helps understand some of the requirements for maintaining, controlling, improving and managing a fairly complex JIT system. The simulation model designed makes a number of realistic assumptions in order for production to respond to the need for small-batch production. The simulation objective:

- To discover the effect on throughputs for selected standard batch quantities as a function of operation time variability, batch changeover times and WIP buffers.

2. Upholstered furniture manufacturing

Furniture manufacturing is an industry where the lead time and retail inventory are critical to sales. The Grubb Furniture Mfg. case study Keller et al. (1991) reminds us that if customers want a particular item that is not in stock at the retailer, they still want it *now* or as soon as possible. If the lead time from the manufacturer is 8 weeks or more, they may go elsewhere. Production in small batches can contribute significantly by offering less lead time and more product variations.

The image of a large-scale, custom furniture manufacturing facility is one of a hybrid flow-shop supported by comprehensive material requirements planning (MRP), a shop-floor control system to monitor the movement of batches through the workstations and a flexible material handling system. The subassembly system for a representative manufacturer of recliner chairs was modeled for a computer simulation to study the interaction of parallel subassembly lines. These lines assemble seats, backs, bodies and control mechanisms, and merge into a final assembly. If any workstation falls behind, their limited input buffer would restrict or *block* the upstream workstations, creating a bottleneck. If its output buffer becomes depleted, it *starves* the downstream operations. The blocking and starving creates *interdependence* among the workstations on each line *and all the parallel assembly lines*.

The manufacture of *upholstered* chairs that *recline* holds a special fascination. Some recliners can swivel, rock, vibrate, message, warm and can offer pockets for phones, TV remotes and other features that customers want. The fascination extends to orders for different styles, fabrics and other features such as buttons, pockets, feet, etc. The "fascination" or challenge for the manufacturer is to combine various orders into batches and then develop feasible and economic production schedules that also meet customer expectations. The possible number of combinations can be restricted by reducing the number of styles, fabrics and colors available in each model group.

An effective production schedule for small-lot production is contingent on consolidating customer, and therefore retailer, orders into groups having common characteristics such as style and a promised delivery-date range. Synchronizing all activities to the developed schedule requires continuous visibility of the factory floor. Giving immediate attention to problems is paramount to coordinating subassemblies with the final assembly schedule.

Furniture units-in-process consume a relatively large amount of floor space especially when produced in *batches*. A manufacturer thus must focus on having a *lean* facility to keep the units moving through the plant and to the retailer as fast as possible. Because most operations are labor-intensive and each batch can be unique, production requires an experienced work force that can quickly change over from one batch to the next.

The components and subassemblies of a large variety of chairs require even more coordination, in a sense, than automobile production. The synchronized flow of unique batches through the work centers appears more complex because of a much greater variety of styles and fabrics available and the greater reliance on operator skills. Technology has helped in some areas such as cutting the layers of fabric with numerically controlled laser cutters. However, the number of fabric/style/color combinations available plus special features can be overwhelming unless grouped into manageable style configurations.

Exhibit 1 shows the queuing network of the combined serial/parallel processing of chair components. In this network, if the final operations on each branch are completed and the batches are joined simultaneously at final assembly, the system is assumed to be *synchronized*. Detailed synchronization may be desired at each of the parallel operations such that if a batch incurs difficulty, alternatives can be considered. The alternatives include stop the total system or allow parallel operations to slow down until the problem batch can catch up or bypass the affected batch in the system. The latter can be a very disruptive activity.

3. Operation details

Distinction is made between *small-lot* production of homogeneous items and *small-batch* production wherein fabrics within a batch may have different characteristics of color, pattern or texture. In batch production, items within a batch are to be processed through their successive operations in the sequence dictated by the cutting operation. Thus, subassemblies for seats, backs and arms processed on parallel subassembly lines are linked. They must also maintain this same item sequence for the *matching* to occur at final assembly line. This is quite different from most conventional linear assembly operations. The **size of the batch** is a dominant factor affecting output, and for work–load balance, material handling and floor control purposes, **it must be a constant**.

Mobile racks or carts are characteristically used to move batches between operations, with each rack

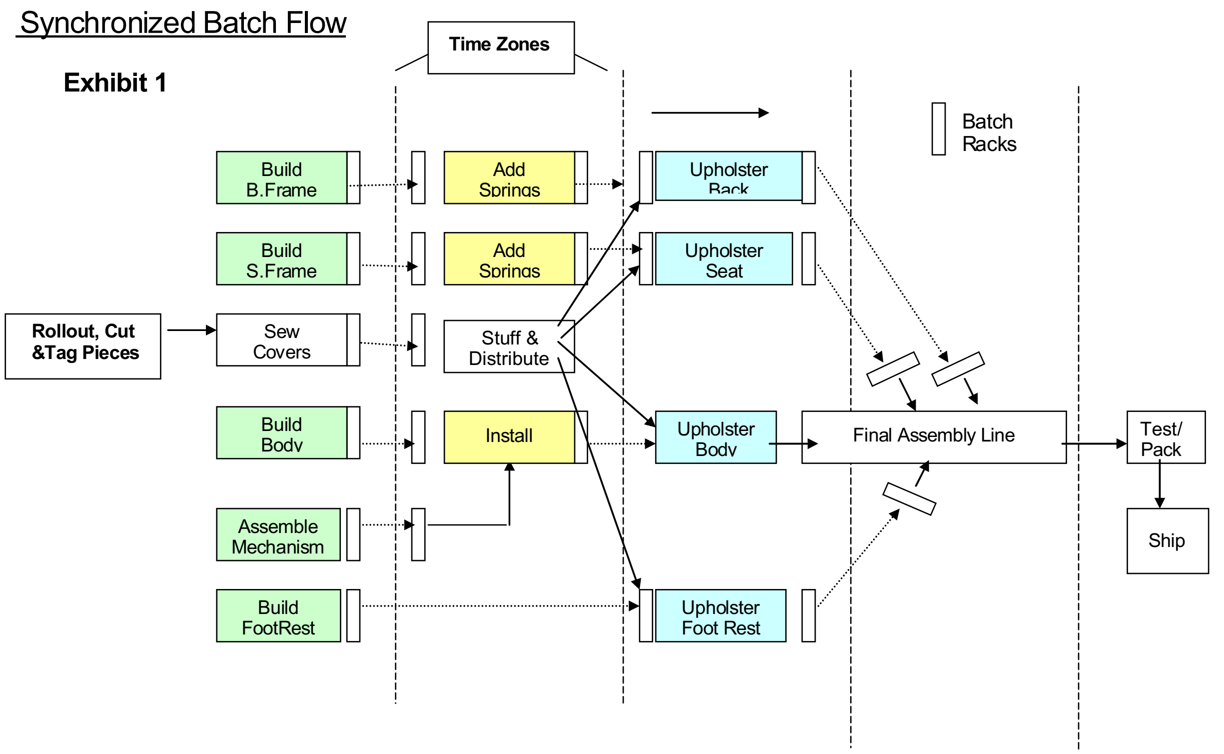

Exhibit 1.

holding one batch of a component such as a seat or back. The racks are identified and their contents, location and status maintained in a database system. Mobile racks allow the utmost in flexibility and preserving the item sequence within the batches at each workstation. With a short manufacturing cycle of only a few hours, the product variations and other exigencies provide a challenging experience for operator and production control personnel.

Exhibit 1 represents a recliner-chair production system. The operations can be partitioned into time zones, from left to right, to show the major points where synchronization can be monitored. As shown, the frame assemblies are built and then upholstered with the fabrics arriving from the cutting and sewing operations. The body frame is built, a reclining mechanism installed and then upholstered in the same sequence as the other component subassemblies were made. The body is moved directly to final assembly where the subassemblies are installed at an appropriate point and, of course, in the same sequence. The completed chair is inspected, tested, boxed and loaded on delivery trucks assigned to specified routes. The operations are described as follows:

1. Fabric sheets are rolled out, cut to length and laser cut into component pieces. The cutting operation *defines the batch* and *processing sequence* for units within a batch.
2. The cut pieces are tagged and sewn into several “covers” for seats, backs, body, etc.
3. Sewn covers, in *a prescribed sequence*, are distributed to the subassembly areas.
4. Wood frames are built in batches, set on carts and moved to the appropriate lines.
5. Frames are upholstered in the *prescribed sequence* and replaced on the cart.
6. Each body is upholstered in prescribed sequence and starts down the final assembly line.
7. The matching subassemblies are installed on the body in the *prescribed sequence*.
8. The units are tested, inspected, boxed and loaded on assigned, waiting delivery trucks.

Synchronization is achieved when all the subassembly carts have arrived at final assembly.

Delayed arrivals will cause loss of time and reduce the output per day expected.

4. The Simulation

The purpose of the simulation was to develop a model to pretest changes in a production system being considered by management. Several years ago they reduced the batch size, which varied from 20 to 50 units per batch, to a standard quantity of 18 units per batch. Given that smaller batches offer greater flexibility to respond to style changes, schedule changes, improved quality control and customer service, does further reduction in batch size significantly affect output, production costs and other factors including impacting the skilled labor force?

The model developed attempts to emulate this complex system. Predicting throughput for a single *linear flow system* with 3 or more manually controlled sequential workstations, with or without product changeovers, with or without WIP buffering, with or without random delays is not as difficult as in combined serial/parallel systems. Effective use of queuing and statistical models may preclude the need for simulation in these simpler cases, although the advantages of simulation in providing insight are lost. When there is interaction between the workstations in the case of 2 or more *parallel* subassembly lines, the model is more complex. Synchronization at points of coordination is required because of space, quality or time constraints. A simulation, using realistic estimates and assumptions, can contribute insight into several operating areas including communication, monitoring and supervision needs.

Before conducting an extensive and detailed data-gathering effort for the simulation, preliminary estimates of operation, changeover and move times are helpful to understand what data are needed and, more importantly, what data are not needed to represent the system. Some assumptions made for this initial simulation to represent the system are as follows:

> A system startup distortion is mitigated by discarding the first 100 of each 2000 cycle runs.
> For balancing, the same mean and variance parameters were assumed for all operations.
> Simulated times for the *batches* are generated using normal distribution parameters.
> Batch changeover times were applied equally to all subassembly operations.
> Queuing buffers between all workstations were held constant at 0, 1 or 2 batches.
> Partial batches are not permitted.

The simulation computes the average time per unit from processing batches of 18, 15 and 12 units per batch under different constraints of buffers, operation variability and changeover times:

Again, the purpose of the simulation model is to determine the impact on throughput when the following parameters are changed.

Batch size	3 levels; 18, 15 and 12 units per batch.
Buffers permitted (WIP)	3 levels; 2, 1 and 0 batches per workstation.
Batch time variability	2 levels of standard deviation.
Batch changeover times	4 levels.

The simulation generates the expected throughput for each of the 72 ($3 \times 3 \times 2 \times 4$) combinations of the above. The simulation design would grow exponentially in complexity as more features are included such as different changeover times per workstation, etc. The objective was to test throughput sensitivity using *3 feasible batch sizes* when the number of buffers, operation variability and changeover times are varied.

5. Results

The primary variable to be tested is the **batch size**. Each of 3 batch sizes is tested using the simulation model to generate the average time per batch under conditions of batch time variability, maximum buffers permitted and changeover times at each of 14 operations. The results are expressed as long-run, average daily outputs in units, i.e. chairs, as calculated from the simulations of batch times and compiled in Table 1. This comprehensive matrix permits the effect on output to be examined for any combination of the selected parameter values chosen.

The **best** long-run daily throughput is achieved where the changeover time is assumed 0, **2 buffers** (batches) are allowed at each workstation and the operation variability is the least. From the *average* time per unit generated after 2000 runs and shown in column 4, the throughput expected for an 8 h day is **395** units for batches of **18 units**. For the **worst**

Table 1

	Buffers and units per batch	Batch SD	Batch SD	Simulated output per day for selected buffers							
				Batch sizes, batch SDs and changeover (setup) times							
		0.5 min.	2.0 min.	Output in units per day Changeover (min./batch/oper.)				Output in units per day Changeover (min./batch/oper.)			
		Avg. min. per unit*		0.0	0.5	1.0	2.0	0.0	0.5	1.0	2.0
	2 Buffers										
1	18	1.214		395	387	378	362				
2			1.253					383	375	367	352
3	15	1.242		386	376	367	349				
4			1.266					379	369	360	343
5	12	1.280		375	363	352	332				
6			1.297					370	359	348	328
	1 Buffer										
7	18	1.235		389	380	372	357				
8			1.270					378	370	362	348
9	15	1.263		380	370	361	344				
10			1.291					372	362	354	337
11	12	1.301		369	357	347	327				
12			1.328					361	350	340	321
	No Buffer										
13	18	2.263		212	210	207	202				
14			2.283					210	208	205	200
15	15	2.295		209	209	203	198				
16			2.313					208	205	202	196
17	12	2.343		205	201	198	191				
18			2.359					203	200	197	190
	1	2	3	4	5	6	7	8	9	10	11

For example, cell (1,4) = average units/day with 2 buffers/station = **395 units per day (best)**; for example, cell (18,11) = average units/day with 0 buffers/station = **190 units per day (worst)**.

*Average min/unit generated from 2000 + cycles (batches) of the simulation model.

practical case scenario, the throughput averages **321** units per day from large **variability** of batch times, **single** buffered workstations having long, 2 min batch changeover times **and** the smaller batch size of **12 units**. The difference in throughput from the lowest output (321) to the highest (395) is approximately **23%** (*0 buffer scenarios were dismissed as impractical*).

Table 1 details are as follows: column 1 identifies the 3 *batch sizes*, 18, 15 and 12 units and grouped under 2, 1 or 0 *buffers* or WIP batches, an *average time* per batch is derived from a simulation run of 2000 or more batches and having a batch SD of 0.5 min at each operation in the system. Column 2 is the average time *per unit* from dividing the average batch time by its batch size (18, 15 or 12). Column 3 is the same but with a batch SD of 2.0 min. Column 4, the average, long-run output per day, results from dividing the average minutes per unit from column 2 into a 480 min day. Columns 5–7 yield average outputs per day after deducting the time lost from changeovers from batch to batch. Exhibit A shows impact of batch size on time per unit for 2 levels of task variability. Exhibit B reflects the impact of batch size, variability and number of buffers on output per day.

As expected, the average time per unit at each operation *increases as the batch size is reduced*. This is due to the increase in *relative* variability of the smaller batches and prorating the changeover time over fewer units. Larger batches reduce the effect of variability of individual units (from statistics, the batch SD is equal to the unit SD multiplied by the square root of the batch size). If the variability of

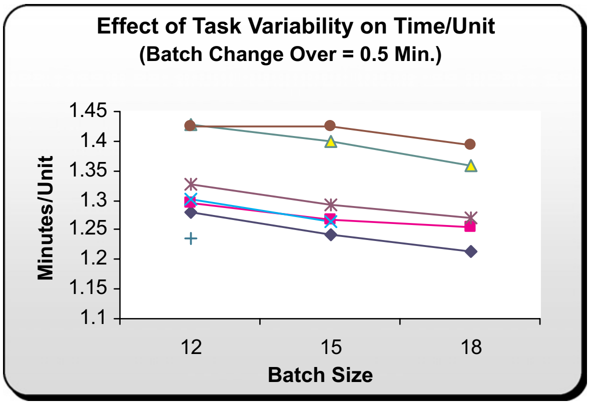

Exhibit A

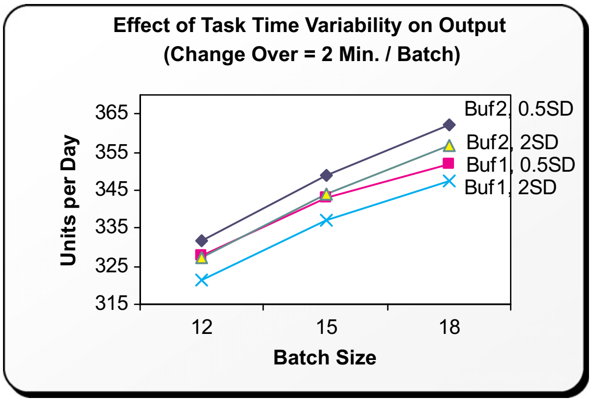

Exhibit B

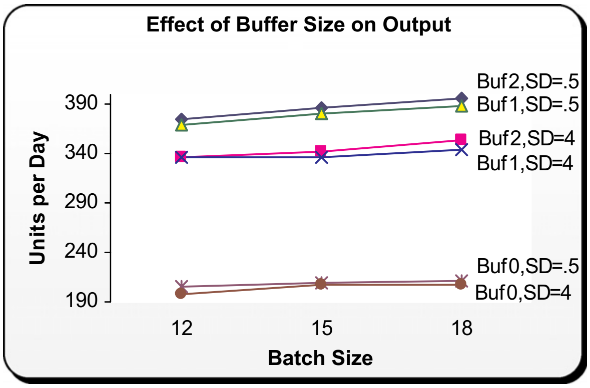

Exhibit C

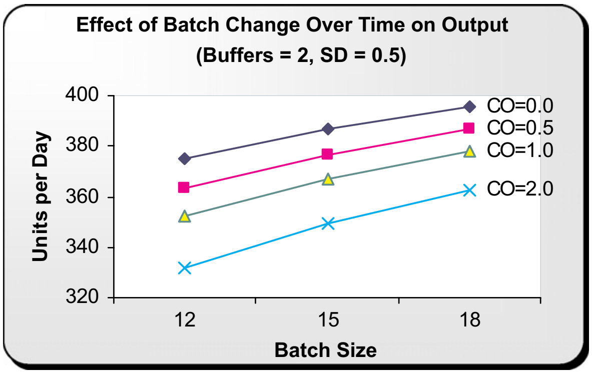

Exhibit D

Exhibit A–D.

individual units can be decreased, the batch variability and inter-operation delays would decrease and throughput is increased. This is difficult to accomplish in the real environment given the manual skills involved, the variety of products and the short cycle time.

Reducing the units in the batches from 18 to 12 yields a 5% reduction in average output [e.g. (395−375)/395]. As stated earlier, however, smaller batches require less floor space, provide more flexibility and shorter delivery times.

The case of 0 buffers is depicted in Exhibit C. Data from Table 1 indicate that the throughput would be reduced by approximately 54–48% if moving batches *directly* between workstation is attempted. Direct pass, i.e. no WIP buffer batches, between manually controlled operations, is impractical, costly and inefficient. Workstations are either starving for work or overwhelmed, causing delays. At least one batch should be available to decouple the operations and prevent delays and lost output.

In the system modeled, buffers of 2 batches yield only about 1.7% more output than single buffers. However, the “extra” WIP batches may avoid queuing delays. The single buffer system would need less floor space and allow greater control by increased attention given to the location and status of the batches.

If the variability, expressed as batch SDs, could be reduced from 2 to 0.5 min per batch, throughput could increase from 367 to 378 units or approximately a 3% increase for the present batch size of 18. For smaller batches, the improvement is about 2%, i.e. 360–367 units per day for batches of 15 and 12, respectively. This may not appear significant but in the long run, reduced variability can contribute significantly to profitability by reducing delays and thus the average time per unit.

Exhibit D depicts the rather obvious result that output per day increases as operation changeover time is decreased. The synchronization constraint in the simulation determines that the subassembly batch components arriving last control the cycle time and cause delays for the other subassembly lines. Given the same variability, if the changeover time can be reduced from 2 to 0 min per batch, output can increase about **10%**, i.e. 347–387 units per day. For a batch size of 12, output would increase about **8%**, i.e. from 327 to 357 units per day.

6. Summary

The graphs and data demonstrate that further reduction in batch size would reduce production output. Reducing the queue size from 2 to 1 buffer (batch) would have a very small impact on output. However, synchronized output from an interactive set of manually controlled assembly line operations demands a buffered system. Less operation time variability and less changeover time between batches would increase output as would be expected. Without cost data such as cost of space, alternative material handing methods and estimated cost benefits of shorter delivery cycles, an economic model on the order of EMQ (economic manufacturing quantity) would not contribute to understanding the real system.

The synchronization required between parallel subassembly lines can be extreme in the sense that if one subassembly line became a bottleneck at any of its operations, all the other assembly lines would be affected. If this requirement is relaxed such that synchronization is necessary only at the last of the subassembly operations, as in the simulation model presented, throughput will increase. However, synchronization at each interim stage offers better control of **quality** and suggests opportunities for product improvements and improved production methods. It can help avoid delays and can anticipate problems.

In the production system described, the tradeoffs are between small-lot flexibility to better serve customers plus closer control of quality and production methods versus economies-of-scale such as *cutting* larger batch quantities and less lost time from fewer changeovers and less material handling. Smaller batch sizes and resulting larger variability would make the system much more sensitive to disruptions. With smaller batches also, troubles may be detected earlier and fewer batches would be impacted until problems are resolved.

The production system described is comparable to other JIT (just-in-time) systems used so successfully in the automotive and other industries. There are instances in JIT systems that the **best** lot quantity is found by trial and error and dependent on what the system could **tolerate** in terms of changeover delays, processing times, buffers and material handling. A simulation model that accurately represents a real system and is quick and easy to use can avoid expensive and distracting experimentation on the production floor. Simulation helps understand the real system and allow users to explore alternatives. The simulation in this case helped understand the impact of batch size and variability on system performance. It also demonstrated the importance of buffers to protect system performance.

The **batch size** is the key to an integrated, flexible, synchronized parallel customer-oriented assembly system. Small batches can better accommodate changes in schedules, changes in methods, changes in materials handling and changes in product configuration. A cost-effective production schedule is contingent on consolidating customer orders into batches that also recognize the distribution system. Small batches can reduce delivery time, thereby improving sales. Retailers and customers have become less tolerant of long delivery times, delays or postponements. In furniture production, synchronizing all activities with the cutting operation, maintaining complete on-line visibility of the factory floor plus giving immediate attention to problems has become a requirement. Long-term and short-term system balancing requires constant review of the facilities, methods, technologies, and education in JIT and quality management principles and techniques. A real-time visual monitoring system may be helpful to digest the comprehensive data generated in real time.

Further studies into production systems represented by this simulation can be entertained. The economic and behavioral impact of reduction in the lot sizes could include studies of within-lot learning, applicable communication technologies, etc. Mixed model production algorithms could be expanded to assist in the dynamic decision-making of batch constituents in complex batch-oriented, labor-intensive, market-oriented environments depicted by this simulation.