© 2010 Cenresin Publications
www.cenresin.org

DEVELOPMENT OF STOCHASTIC SIMULATOR OF CONTINUOUS REVIEW (R,Q) INVENTORY POLICY WITH DEMAND LEAD TIME AND RATIONING

Ugochukwu C. Okonkwo, and Sam Nna Omenyi
Department of Mechanical Engineering
Nnamdi Azikiwe University, Awka

ABSTRACT

Inventory demands categorized into classes of service differentiation are often avoided by researchers, because of the varieties of interacting variables which such research entails. In this paper, however, a model of a continuous review (r,Q) inventory policy having service differentiation of demands was developed, with complete graphical user interface and digital tracing. In this study, composite stepwise two dimensional graphical representation of the model was formulated, which captured the stochastic demands and stepwise state transitions of the system. The inverse transform algebraic method was applied for the generation of random numbers while next event method was used for the time advancement of the simulation clock. Traces and structured walk through were utilized for the verification of the stochastic simulation models. Batch mean method was used in determining the confidence intervals of the simulation model, with 10^5 run time and 100 replications. The developed model results were validated with the case study ANAMMCO data. On the average a savings of 18.51% demands in comparison with the conventional models was found, which indeed will result in huge cost savings in absolute terms. Beyond that, the insights from this model will increase the overall efficiency of spare parts inventory control.

Keywords: Simulation, Service Differentiation, Spare Parts Inventory, Rationing

INTRODUCTION

In today's technological environment, the indispensability of stochastic simulation for effective engineering decision making, is obvious. Lost revenues due to negligence or lack of knowledge of its use are enormous. Inventory systems are usually stochastic in nature, hence, simulating such systems, capture its obvious and latent behaviours, thereby giving insight for better decision making. The use of stochastic simulation is even more crucial when it comes to spare parts inventory systems. This is because spare parts inventory differ substantially from regular inventory systems. The key reason for this difference is that spare parts provisioning is not an end in itself, but a means to guarantee up-time of equipment. With respect to spare parts inventory, the customer's sole interest is that his systems are not down due to lack of spare parts because equipment downtime is lost production capacity. In 2006, Koudalo[1] investigated revenues in the spare parts service business over a period of one year, and he reported combined revenues of more than $1.5 trillion. Flint[2] stated that the world's spare parts inventory in the aviation industry in 1995 amounted to $45 billion at that time. Any means to downsize this stock, without decreasing customer service, would be more than welcomed by the aviation industry. Also in other industries, large amounts of money are invested in spare parts inventory and this has increased over the years. Because of these large amounts of money involved, savings of a few percents through stochastic simulation, only constitute large cost savings in absolute terms.

Nahmias and Demmy[3] were the first to consider multiple demand classes in a continuous review inventory model. They analysed an (r,Q) inventory model, with two demand classes, Poisson demand, backordering, a fixed lead-time and a critical level policy, under the crucial assumption that there is at most one outstanding order. This assumption implies that at the point when a replenishment order is triggered, the net inventory and the inventory on hand are the same. The model of Nahmias and Demmy was analysed in a lost sales context by Melchiors, et al[4].

Moinzadeh and Nahmias[5] and Moinzadeh and Schmidt[6] also addressed a problem under continuous review policy. The former considers fixed ordering costs for both regular and expedited demands. The authors developed a heuristic approach that was an extension of the standard (r,Q) policy. The latter assumed that demand and fixed ordering costs were small compared to the holding cost and therefore a one-for-one ordering policy is reasonable. In their model, they considered a single location system facing Poisson demand and having the choice of two replenishment modes, a more expensive mode with a shorter lead-time and a cheaper one with a longer lead-time. Not only current inventory is considered but also information on the outstanding demands is considered before placing an expensive emergency order. Their results show considerable savings associated with the dual supply strategy. Bradley[7] studied a related problem of dual sourcing in capacitated settings. Deshpande, et al[8] considered a rationing policy for two demand classes differing in delay and shortage penalty costs with Poisson demand arrivals under a continuous review (r,Q) environment. They defined a so-called threshold clearing mechanism to overcome the difficulty of allocating arriving orders and providing an efficient algorithm for computing the optimal policy parameters which are defined by (Q, r, K), K being the threshold level.

The latest simulation of the spare parts inventory model relevant to this was performed by Wang et al[9]. In that simulation model, a supplier had an inventory system that provided several lead-time options to its customers. The inventory replenishment lead time was a multiple of the inventory review cycle. They considered an inventory commitment problem, in which the supplier allocated its on-hand inventory to two groups of customers. When inventory ran out, the supplier backordered demand to future cycles. They formulated the inventory commitment decision using dynamic programming algorithm. The simulator also explored the optimal inventory replenishment issue and evaluated the performance of the models. Okonkwo[10], through stochastic simulation, projected the behaviour of vehicles in the maintenance workshop by capturing the maximum queue length of vehicles and average time of waiting for vehicles that entered the queue.

A simulation model was developed by Persson and Saccani[11] in order to support a case study concerning a world player of heavy equipment. Its spare parts distribution system, configuration and allocation decisions were modeled. Stochastic simulation which was well suited for time-dependant relations was used. Supply chain simulation applied to the case study provided useful insights on the decision choices and the cost structure related to the spare parts distribution system.

THE PROPOSED MODEL IN RELATION TO THE LITERATURE

While Melchiors, et al[4] developed their model in a lost sales environment, the model to be developed with that of Nahmias and Demmy[3] considered backordering behaviour, which is more complex. The replenishment order cycles in the models of this study will be based on inventory position, taking into account the demand lead time which Nahmias and Demmy[3] did not consider. Deshpande, et al[8] considered a rationing policy for two demand classes which differ in delay and shortage penalty costs. This study will pursue approaches involving down order class and maintenance order class.

Okonkwo[10] and Persson and Saccani[11] incorporated an algorithm tool in the simulation model which had the capability of performing sensitivity analysis in some of the sensitive parameters while Wang et al[9] simulation model had the capability of performing sensitivity analysis in all the sensitive parameters of its model in the existing literature. The simulation model to be discussed just like the model by Wang et al[9], will have the capability of performing sensitivity analysis for all the sensitive parameters of the model.

MATERIALS AND METHODS

The design Engineer has the responsibility of considering many factors relating to the system that is to be simulated and taking informed decisions based on his considerations. Given this responsibility, relevant details of the Spare Parts Complex of Anambra Motor Manufacturing Company (ANAMMCO), which satisfy the basic requirements for the model, were studied. The staff of the spare parts complex as well as documents and software obtained from the company are sources of relevant information. Equipped with the required information, the real system is then reduced firstly to a composite stepwise 2-dimensional graphical representation, thereafter to logical flow diagrams, algorithms and programmes.

In designing the models, service differentiation through rationing is incorporated. This is because, in spare parts inventory, just as different customers may require different product specifications, they may also require different service levels. For instance, for a single product, different customers may have different stockout costs and/or different minimum service level requirements or different customers may simply be of different importance to the supplier by similar measures. Therefore, it can be imperative to distinguish between classes of customers thereby offering them different service.

Most companies target to achieve maximum of service level requirements while considering the aggregated demand which is much simpler to model. However, the models of this study have two classes of service differentiation through rationing at the critical level. The high priority demand represents the down demands while the low priority demand represents the maintenance demands. The high priority demands will have zero demand lead time while the low priority demands will have a positive demand lead time, in the models.

Priority clearing mechanism is used in the models because it ensures that high priority demands that are backordered are cleared promptly.

The inverse transform algebraic method is used in the simulation development. The random phenomenon has a negative exponential density function, expressed thus:

$$F(t) = \lambda e^{-\lambda t} \qquad F(x) = \int_0^x \lambda e^{-\lambda t}\,dt$$

$$F(x) = 1 - e^{-\lambda x} \qquad e^{-\lambda x} = 1 - F(x)$$

$$-\lambda x = -\ln\{1-F(x)\} \qquad x = -(1/\lambda)\ln R(x)$$

where

$R(x) = 1 - F(x)$,

$R(x)$ is a random number between zero and one and x is the variable [15].

The next event method is considered in this study because discrete event simulation models were carried out. In this method, the clock is incremented by a variable amount rather than a fixed amount. This variable amount is the time from the event that has just occurred until the next event of any kind, occurs. In other words, the clock jumps from event to event and remains constant between events.

The programming software that was used for the development of the software package is Visual.Net. The major reason for this choice among the other software is that it has the capability of programming a true client web-based application.

Well known simulation techniques were used to verify the simulation models. Firstly, running it for a known situation where the result can be easily calculated. Thereafter, *Traces* was employed. Finally, *Structured Walk Through* was conducted.

The simulation models are stochastic. The implication is that the results that will be generated for a given dataset at a given time, will not exactly be the same with the results that will be generated if the simulation is repeated using the same dataset. Hence, there is a need to check the bounds in which the results revolve. If the interval is relatively small, then the results that are generated from one replication can be used confidently. To verify this, 10^5 maximum simulation time and 100 replications was used for a given dataset.

For the validation, the developed model was made to conform with the current practice of the company (ANAMMCO) by adjusting some of their input parameters, thereafter; they were validated with the results generated by the companies software package.

MODEL DEVELOPMENT

Manual Simulation Representation

The model of this study involves multiple stochastic parameters and stepwise state transitions, hence, there appears a need to perform manual simulation in readiness for the development of the models. To highlight the extent of this multiple stochastic nature, in the design of this study, the high and the low stochastic demand parameters are filled from the same pool of inventory. However, a high priority demand can only be backordered if there is zero inventory, otherwise it is filled. On the other hand, a low priority demand can only be filled if the stock is above the critical level, otherwise it is backordered. The low priority demands are not satisfied immediately. They give a constant demand lead time before the demands are filled. As a result of the stochastic nature of the demands, these will be transmitted directly to the generation of replenishment orders and clearing of backorders. Also, a replenishment order only occurs when the stock has reached reorder level. Furthermore, the manual simulation has the capability of both bulk demand and bulk replenishment of spare parts. Besides, the probability of partial filling which naturally arises for bulk demands of spare parts, is observed in this case but not filled except the customer reduces his demand so as to clear the stock. Hence, it is either the demands are completely filled or completely backordered.

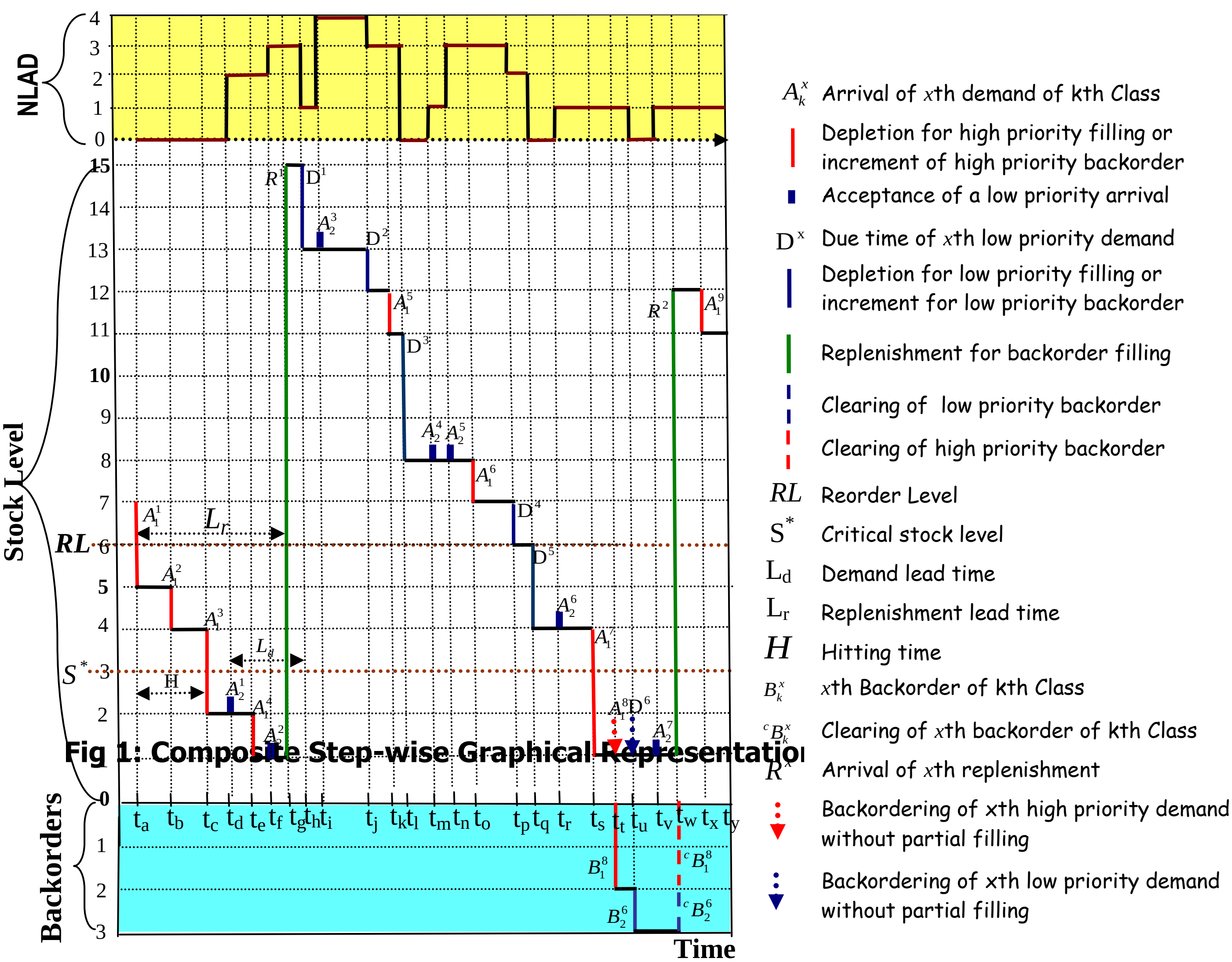


Fig 1: Composite Step-wise Graphical Representation

Fig 1 is the composite stepwise 2-dimentional graphical representation of the stochastic behaviour of the system for the model. Specifically, the composite graph captured the following:

1. The stock level behaviour. This involves the impact of high and low (acceptance and due time) priority arrivals, as well as the arrival of replenishment order. It also shows how the behaviour of the system when partial filling is not allowed.
2. The backorder behaviour. This includes the backordering and its clearing, through priority clearing mechanism.
3. The number of low priority arrivals that are not yet due (NLAD) as well as what causes the stepwise increment and decrement.

The time intervals points for this manual simulation are t_a, t_b, t_c, . . . t_y. At a closer observation, this manual simulation does not witness any change of behaviour in-between adjacent time intervals. However, change occurs at every interval point except t_y which is the arbitrary ending time. Hence, it shows that it is a discrete event simulation. Although t_a is an arbitrary starting time, but it coincided with the arrival of a high priority job. For easy comprehension of Fig 1, the behaviour of the system at some of the time points are hereby presented with explanation.

Time t_a: This is the arbitrary starting time of the manual simulation. This coincided with the arrival of the first high priority, 2 SKU demand, causing the stock level to deplete from 7 to 5. This demand reduced the stock level beyond the reorder point R and so an order is also placed at this time.

Time t_d: Arrival of the first low priority, 2 SKU demand but the stock level is unchanged because the due date has not arrived. An increment of NLAD to 1.

Time t_g: Arrival of the replenishment order of 14 SKU after a demand replenishment lead time of L_R placed at time t_a, this increased the stock level to 15.

Time t_h: Due time of the first low priority, 2 SKU which is filled, causing subsequent depletion of stock from 15 to 13.

Time t_t: Arrival of the eight high priority, 2 SKU demand, the entire 2 SKU is backordered because the icon activating partial filling is not enabled. This resulted in the stock level to still remain at one.

Time t_u: Due time for the sixth low priority, 1 SKU demand. The 1 SKU that is remaining in the stock cannot be used to fill this demand because of two reasons there is a backorder at that moment. Even if there is no backorder, the 1 SKU that is remaining is reserved for high priority demands since the critical level is 3. Also, NLAD is reduced to zero.

Time t_v: Arrival of the seventh low priority, 2 SKU demands, causing the NLAD to move to 1 but no change in the stock level.

Time t_w: Arrival of the second replenishment order of 14 SKU. 3 SKU was used to clear the 3 backorders while 11 SKU was used to increase the stock level from 1 to 12.

Time t_x: Arrival of the ninth high priority, 1 SKU demand and depletion of the stock level to 11.

Time t_y: This is the arbitrary termination point of the manual simulation. There is no change on any of the three modules.

Stochastic Simulation Model Representation

The notations used in the Model are tabulated below:

Table 1 Notation of the Stochastic Simulation Model 3

S/N	Code	Meaning
1.	λ_1	High Priority Arrival Rate
2.	λ_2	Low Priority Arrival Rate
3.	β_1	Fill rate of high priority demand
4.	β_2	Fill rate of low priority demand
5.	CNHD	Cumulative Number of High priority Demand
6.	CNLD	Cumulative Number of Low priority Demand
7	CNB_1	Cumulative Number of High Priority Backorder
8	CNB_2	Cumulative Number of Low Priority Backorder
9.	DTLD	Due time of the Low Priority Demand
10.	FHD	Filled High Priority Demand
11	FLD	Filled Low Priority Demand
12.	L_r	Replenishment Lead Time
13.	L_d	Demand Lead Time
14.	$MQLB_1$	Maximum Queue Length of High Priority Backorder
15.	$MQLB_2$	Maximum Queue Length of Low Priority Backorder
16.	$MRTB_1$	Mean Response Time of High Priority Backorder
17.	$MRTB_2$	Mean Response Time of Low Priority Backorder
18.	NB_1	Number of High Priority Backorder
19.	NB_2	Number of Low Priority Backorder
20.	NROT	Number of Replenishment Order in Transit
21.	NTHD	Next Time of High Priority Demand
22.	NTLD	Next Time of Low Priority Demand
23.	NTRO	Next Time of Replenishment Order
24.	MS^*	Maximum Critical Level
25.	MST	Maximum Simulation Time
26.	MSTE	Maximum Simulation Time Exceeded?
27.	PI	Physical Inventory
28.	QR	Quantity Replenished
39.	QDI	Quantity Demanded Interval
40	RL	Replenishment Level
41	S^*	Critical Stock Level

A typical flowchart representation of the stochastic simulation of the model is shown in Fig 2. The flowchart model has 57 modules. These modules are all labeled and linked together. The flowchart has four typical event channels. The event channel of *next time of high priority arrival* started with module 9 and ended with module 20. The event channel of *next time of low priority arrival* is from module 24 to 27 while that of *due time of low priority* is from module 29 to 41, respectively. The event channel of *next time of replenishment order* is from module 42 to 51.

This Model is capable of performing multiple sensitivity analysis. However, Fig 2 shows a flowchart capable of performing sensitivity analysis of λ_1 only. The flowcharts of this model that perform other sensitivity analysis are essentially the same, the only modules in Fig 2 that are to be changed to adapt to the particular sensitivity analysis requirement are modules 1, 2, 55, 56 and 57. Specifically, this model will be capable of performing six additional sensitivity analysis.

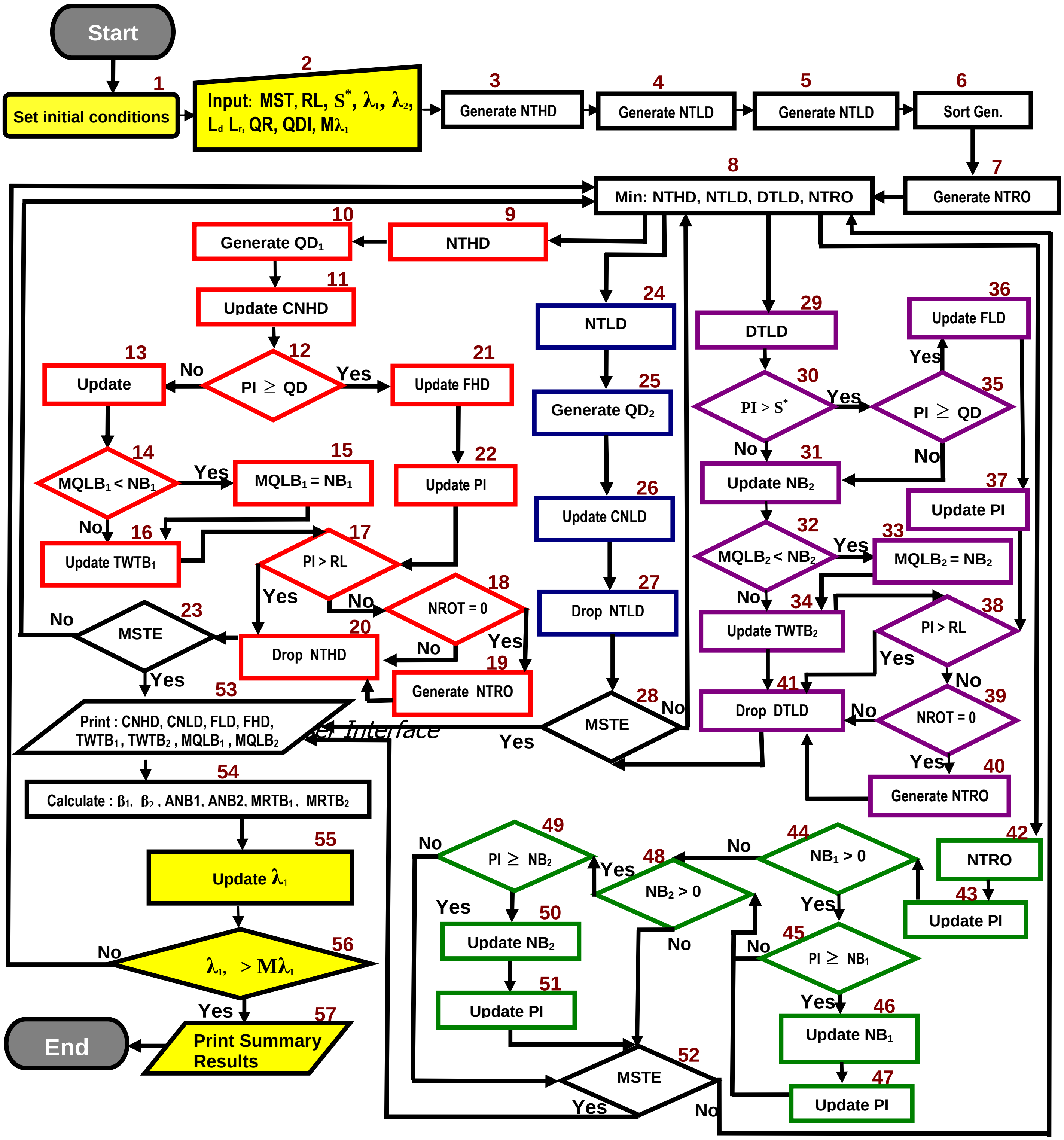


Fig 2: A Typical Flowchart of the Model (*for λ_1 Perturbation only*)

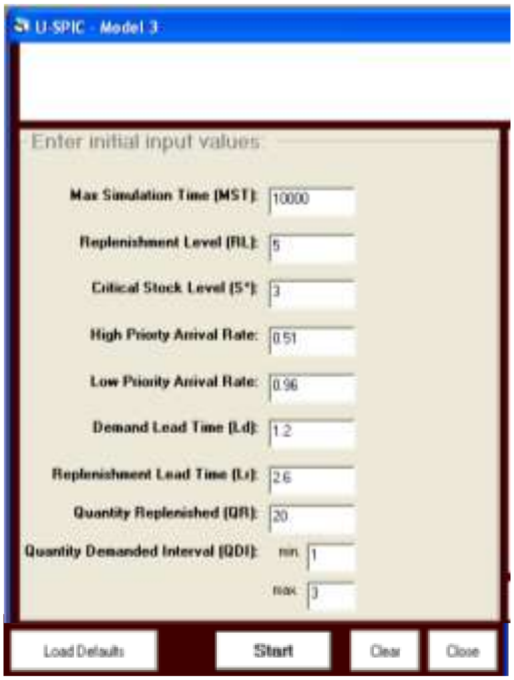


Fig 3: Model Input Dialogue Box

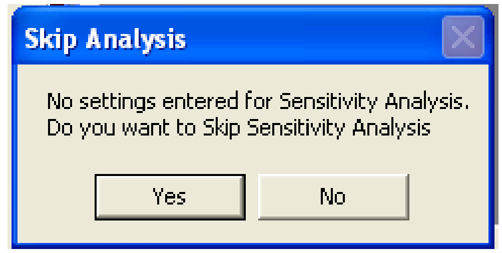


Fig 4: Start/Continue Simulation Dialogue Box

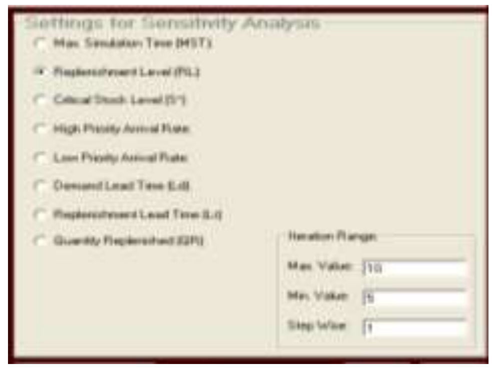
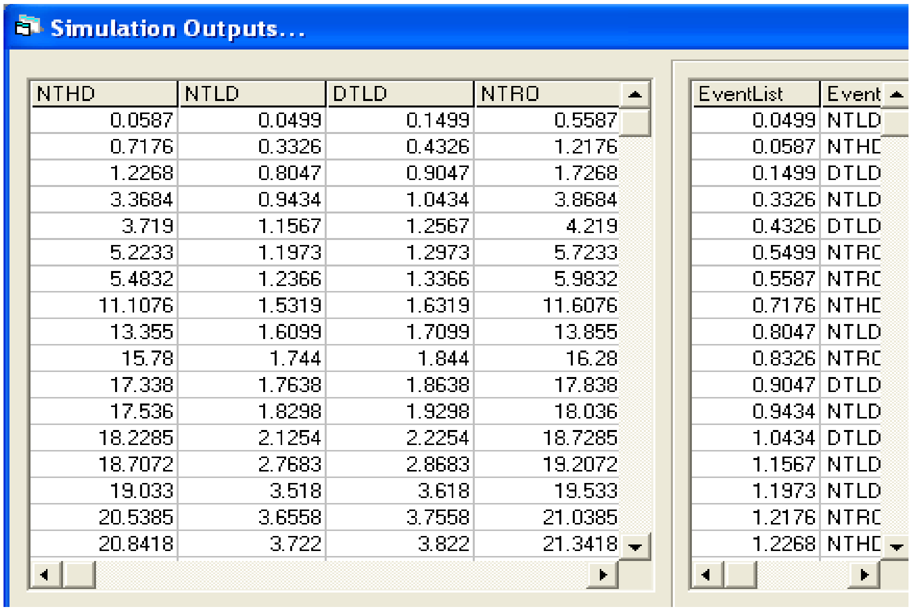


Fig 5: Model Sensitivity Analysis Dialogue Box

The input dialogue box of this model is shown in Fig 3. In this dialogue box, the required input parameters for the simulation are keyed in, one after another, in line with the dataset to be simulated. To use the default values, the defaults button is clicked and the values are automatically supplied. The Start button prompts the computer to start simulation, which invokes a direction dialogue box shown in Fig. 4. This setting of this sensitivity analysis dialogue box is completely flexible. At this instant, the parameter that is selected for sensitivity analysis is the critical stock level, while the minimum and maximum iteration values are 2 and 10, respectively. The step wise value is 1. In other words, the value of the critical stock level changes thus; 2, 3, 4, …, 10 while other input parameters remain constant. At each change of the critical stock level, the simulation is run and its outputs are displayed. Similarly, any of other sensitivity parameters can be chosen while any value can be inputted in the iteration range and step wise change.

After setting the values for the sensitivity analysis, the start button is clicked and the simulation run commences. During the simulation, the system first of all generates all the NTHD, NTLD, DTLD, NTRO which are independent and identically distributed, as is shown in the display simulation outputs window of Fig 6. The first four columns show generated events for the simulation run for NTHD, NTLD, DTLD and NTRO. Thereafter, the software sorts all the values of the generated events time for the four events in ascending order. The sorted values and the corresponding events which each value represents shown in the fifth and the sixth columns, respectively, were also shown in Fig 6.



NTHD	NTLD	DTLD	NTRO
0.0587	0.0499	0.1499	0.5587
0.7176	0.3326	0.4326	1.2176
1.2268	0.8047	0.9047	1.7268
3.3684	0.9434	1.0434	3.8684
3.719	1.1567	1.2567	4.219
5.2233	1.1973	1.2973	5.7233
5.4832	1.2366	1.3366	5.9832
11.1076	1.5319	1.6319	11.6076
13.355	1.6099	1.7099	13.855
15.78	1.744	1.844	16.28
17.338	1.7638	1.8638	17.838
17.536	1.8298	1.9298	18.036
18.2285	2.1254	2.2254	18.7285
18.7072	2.7683	2.8683	19.2072
19.033	3.518	3.618	19.533
20.5385	3.6558	3.7558	21.0385
20.8418	3.722	3.822	21.3418

EventList	Event
0.0499	NTLD
0.0587	NTHD
0.1499	DTLD
0.3326	NTLD
0.4326	DTLD
0.5499	NTRO
0.5587	NTRO
0.7176	NTHD
0.8047	NTLD
0.8326	NTRO
0.9047	DTLD
0.9434	NTLD
1.0434	DTLD
1.1567	NTLD
1.1973	NTLD
1.2176	NTRO
1.2268	NTHD

Fig 6: Simulations Output Event List

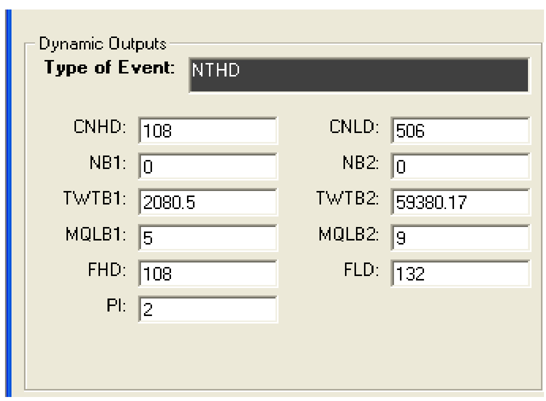


Fig 4.7: Dynamic Digital Tracing Output of Event type

During simulation, if the speed of the simulation is reduced from 2.4 x 10^9 Hz to 1 Hz, The digital tracing of the simulation run can easily be observed as the variables are updated as shown in Fig 7.

Intermediate Simulation Output:

Low Priority Arriv.	CNHD	CNLD	FHD	FLD	TWTB1	TWTB2	MQLB1	MQLB2
0.25	15869	2376	15242	1787	168.18	81.34	1	3
1.25	15437	12672	14344	7853	333.35	2787.88	1	4
2.25	14978	24327	13324	11353	635.14	11696.99	2	5
3.25	15709	30012	13298	11933	1140.80	21035.18	3	6
4.25	16985	42927	13506	13028	2163.64	43072.00	4	7
5.25	15698	59563	11910	13497	2822.38	81914.60	5	8

Fig 8: Intermediate Simulation Output Results

Final Simulation Output:

Low Priority Ar	B1	B2	ANB1	ANB2	MRTB1	MRTB2
0.25	0.9605	0.7521	0.0616	0.0595	1.0009	0.1381
1.25	0.9292	0.6197	0.1104	0.4629	0.6220	0.5785
2.25	0.8896	0.4667	0.1722	1.1774	0.4731	0.9016
3.25	0.8465	0.3976	0.2395	1.9578	0.3841	1.1635
4.25	0.7952	0.3035	0.3195	3.3566	0.3050	1.4406
5.25	0.7587	0.2266	0.3764	3.9029	0.2683	1.7782

Fig 9: Final Simulation Output Results

At the end of each stepwise increment, the output results generated for each perturbation is registered in the intermediate and final simulation outputs. Figures 8 and 9 show a print screen of intermediate and final generated outputs results having six perturbations.

Determination of the Confidence Intervals

A maximum simulation time of 10,000 time units for 100 replications were executed in for the model, using the following dataset [**RL = 5, S^* = 3, λ_1 = 9.5, λ_2 = 6.27, L_d = 1, L_r = 3, QR = 12 and QDI = (1-3)**] respectively. The minimum and maximum values of the fill rates for both priorities of the 3 were tabulated and are shown in Table 2.

Table 2: Maximum and Minimum Results for High and Low Priority Fill Rates

Fill Rate	Model	
	Maximum	Minimum
β_1	0.8798	0.8609
β_2	0.7494	0.7301

From Table 2, the associated confidence intervals of the simulation results which are the difference between the maximum and minimum values is 1.87×10^{-2} and 1.93×10^{-2} for high and low priority fill rates, respectively, which is reasonably acceptable.

Percentage Savings Analysis and Optimizations

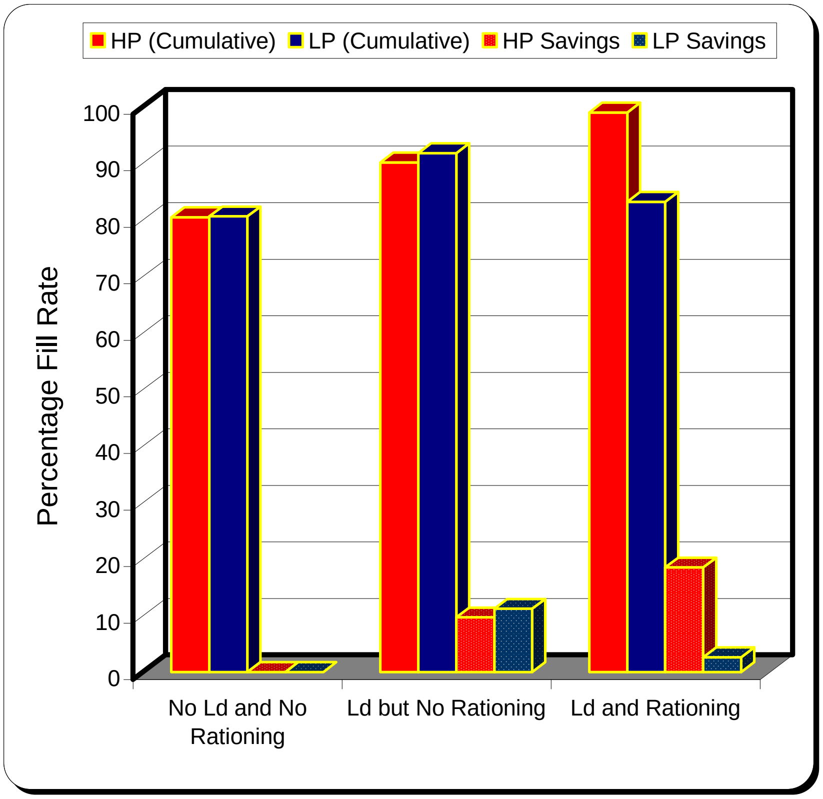


Fig 10: Percentage Savings for Different Priorities of Demands and Optimization Levels

The percentage savings analysis was done in three optimization levels. In the first level, both the demand lead time and the application of service differentiation through rationing were not recognized. The minimum fill rate that will satisfy both priorities demand, considering the aggregated demands in this case was determined. This is a reflection of the current practice in the company. It should be noted that the input values used for critical stock level and demand lead time at this level is zero. In the second level, the demand lead time was recognized; however, the provision of service differentiation through rationing was not applied. In this case both the high priority demands and low priority demands have the same critical stock level which is zero, but a positive demand lead time. From these fill rates, the generated percentage savings are 9.67% and 11.20% for high and low priorities demands, respectively. Although both have a higher fill rates but it can be observed from the fill rate that the low priority demand is having a higher fill rate (service level) while a high priority demand is having a lower fill rate. This is unacceptable.

Hence, another optimization which is rationing is applied. This is done by increasing the critical stock. Finally, in the third level, both the demand lead time and also service differentiation through rationing was applied. From the results generated, the high priority fill rate now shoots over the low priority, making a total gain of 18.51% percentage savings while the low priority fill rate dropped from the previous 11.20% to 2.60%. Notwithstanding the drop of low priority, this last optimization is accepted because it gave a higher fill rate to a high priority demand, which is the major objective of rationing.

CONCLUSION

This study has succeeded in developing a stochastic simulator of continuous review (r,Q) inventory system, designed to be operated through a graphical user interface environment. With this in place, researchers, planners/managers and modelers that are non programmers can leverage advanced constraint-based solving technology, to ensure that inventory planning control is feasible. They can be confident of the ability to execute their plans and meet demand commitments to customers, through quick and easy execution of programme, without any consultation to programmers. Similarly, the cost savings and insight from the study cannot be overemphasized.

REFERENCES

1. Koudalo, P. 2006. *The service revolution in global manufacturing industries*, Deloitte Research.

2. Flint, P. 1995. *Too much of a good thing, better inventory management could save the industry millions while improving reliability*, Air Transport World 32, 103–107.

3. Nahmias S. and W. Demmy. 1981. "*Operating characteristics of an inventory system with rationing. Management Science".* 27:1236-1245,

4. Melchiors, P., R. Dekker, M.J. Kleijn, 2000. "*Inventory rationing in an (s,Q) inventory model with lost sales and two demand classes".* Journal of the Operational Research Society 51, 111–122.

5. Moinzadeh, K., Nahmias, S, 1988. "*A continuous review model for an inventory system with two supply modes*". Management Science, 34, 761–773.

6. Moinzadeh, K. and Schmidt, C. 1991. "*An (S-1,S)Inventory System With Emergency Orders".* Operations Research, 39, 308–321.

7. Bradley, J. 2003. "*A Brownian Approximation of a Production-Inventory System with a Manufacturer that Subcontracts".* Working paper, Jonhnson Graduate School of Business, Cornell University.

8. Deshpande, V., M.A. Cohen, K. Donohue. 2003. "*A threshold inventory rationing policy for service-differentiated demand classes". Management Science* 49, 683–703.

9. Wang, H. et al, 2009. "*Inventory Models with Alternative Delivery Lead Times, Demand Backlogging, and Priority Rules".* Social Science Research Network. Available at SSRN: http://ssrn.com/abstract

10. Okonkwo U. C. 2009 *Simulation, a Tool for Engineering Research through Sensitivity Analysis of a Stochastic Queueing Model Simulator,* Journal of Science and Technology Research, Vol. 8, No 3, 34 – 40.

11. Persson, F., Saccani, N. 2007. "*Advances in Production Management Systems*" 1F1P International Federation for Information Processing, Volume 246, (Boston: Springer), pp. 313–320.