

ANALYZING PROCESSES WITH HANQ

Hank Neimeier and Clem McGowan, The MITRE Corporation

Abstract

We view a process as a network of activity nodes that each take time and consume resources. Products and information flow between the nodes. Hands-on Analytic Network Queuing (HANQ) is an approach and a computational tool to analyze the performance of such processes. Currently HANQ is a set of computational blocks implemented as a library for the Extend™ simulation language product. HANQ is based on a new computational paradigm (analytic risk evaluation and analytic queuing) that overcomes several of the limitations of discrete event simulation. Using HANQ one can compute the risk of an overall process's time, resource utilization, and costs (expressed as probability distributions) given the risks of the individual activities. Deterministic solutions simplify sensitivity and uncertainty analyses. HANQ supports activity based costing (ABC) and can handle time varying workloads and service rates. HANQ integrates performance, cost, and risk in a single model.

Keywords: business process reengineering (BPR), IDEF0, analytic risk evaluation, analytic queuing, activity based costing (ABC),

1. Introduction

Hands-on Analytic Network Queuing (HANQ) is a new simulation paradigm that overcomes several of the limitations of discrete event simulation. It is based on the combination of analytic queuing networks and analytic risk modeling. The analytic queue technique gives an approximate transient solution to the general interarrival time and general service time single server queue. Analytic risk evaluation is based on the beta distribution. It provides the entire uncertainty probability distribution vice an uncertain estimate of the mean. The beta distribution can be fit based on the minimum, mean, maximum, and standard deviation statistics. In a complex results calculation, all that is required is to keep track of these statistics as the calculation proceeds. At any point in a calculation, the probability distribution of the result can be derived by fitting a beta distribution based on the four statistics. When analytic queuing is combined with analytic uncertainty, modeling dynamic uncertainty analysis becomes feasible. The time varying uncertainty distribution in resulting measures of effectiveness can be calculated at any specified time or over any user specified

time interval. This new capability is not available in discrete event simulation.

Simulation can be viewed as a network of interrelated queues. Thus the new paradigm has wide applicability. Its deterministic solution greatly simplifies sensitivity and uncertainty analysis of complex many parameter models. The factor effects in a many factor model are very difficult to obtain in discrete event simulation since they are masked by the stochastic simulation uncertainty. In discrete event simulation the causal chain between input parameter change and resulting output measure effect is broken. The stochastic random number generation process, requires many runs to see the effect of an input parameter change. In training simulations, this random learning effect can be a problem. Our proposed deterministic technique overcomes this problem.

Discrete event simulation requires a time period of many simulation events to determine sample statistics. Our technique provides instantaneous statistics. This feature, and the analytic uncertainty analysis technique, makes dynamic risk analysis feasible. A simple five node queuing network example is presented. The network has a time varying uncertain input workload. Rejects from the inspection activity are reworked and fed back to inspection. The time varying total processing delay distribution for all process activities is calculated using the proposed technique.

2. Hands-on Interactive Model Development Concept

Object oriented programming can be used to develop a real time interactive model development capability. The objects are an input workload generate block, an activity block, and a output summary block. A microcomputer with the EXTEND™ simulation package and simple HANQ library are the required hardware and software. The workload, activity, and output blocks are in the library, and are directly accessible from the menu bar. The user brings an IDEF0 activity model of the process (or sequenced list of process activities) to the modeling session. Estimates of input workload and activity processing times are also brought. A facilitator is available for any needed user training or prompting. The user builds the model by selecting, placing and labeling activity blocks from the library menu. These are wired

together based on the work flow sequence. The IDEF0 arrows are a guide to wiring the activities. A generate block is added before the first activity and wired to it. Finally an output block is included for plotting and beta distribution fitting. The model is run on the historical workload data to see if historical performance data is reasonably approximated. With model validation, several alternative processes are created and executed. The most appropriate is selected for further detailed study and implementation.

2.1 Example Model

We start with a description of a process as a network of activities. Process reengineering teams often create such "process maps" for both the current (As-Is) process and the proposed future (To-Be) process. Figure 1 is an illustrative IDEF0 process model of a system that takes in "parts" and "orders" and produces "shipped products". The process has five activities in an assembly line-type network with a feedback loop to do rework.

To build a HANQ model the analyst needs a microcomputer with the Extend™ simulation package and the simple HANQ three block library which we have built. The workload, activity, and output blocks are in the library, and are directly accessible from the menu bar. To build a model the user selects, places and labels activity blocks from the library menu. The user wires these together based on the IDEF0 arrows. A generate block is added before the first activity and wired to it. Finally an output block is included for plotting and beta distribution fitting. If possible, the model is run on historical workload data to see if historical performance data is reasonably approximated. With model validation, several alternative processes can be created, executed, and compared. Wherever data is not available (e.g., a new improved process activity, a future workload) the user should estimate the input workload and activity processing times.

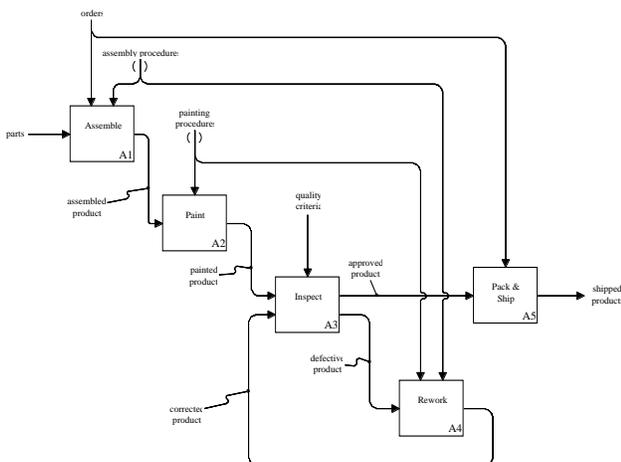


Figure 1 IDEF0 Process Model

Figure 2 shows two views of the figure 1 production process. The top view displays node statistics while the bottom view shows cumulative statistics from the start of the process.

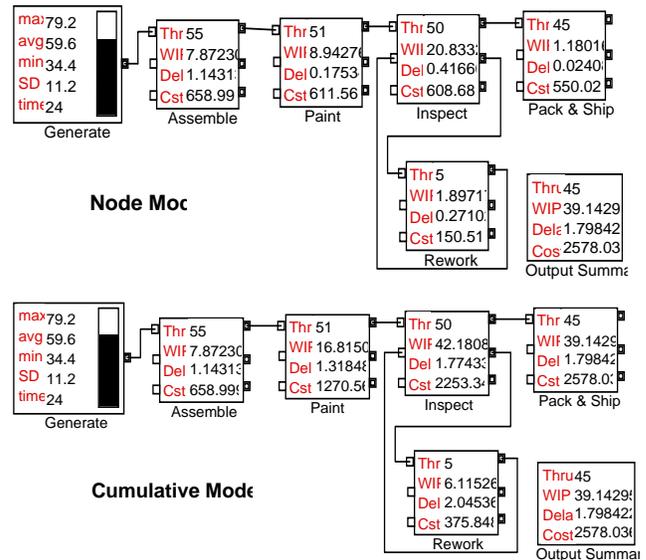


Figure 2 Animated Process Model

The production process is composed of five activities: assemble, paint, inspect, rework, pack and ship. Painted products that fail inspection go through rework. This illustrates a feedback loop. The process model is composed of three block types:

- Generate that inputs the time varying workload statistics
- Activity that represents the process with service rate statistics, fixed cost, variable cost and storage cost
- Output summary that plots historical process performance, and fits beta distributions to process statistics

The following paragraphs describe each of these block types and display their dialog boxes. A dialog box is opened up by double clicking on a block icon. Data inputs and calculation results are displayed in the dialog box. In animation mode, selected statistics are displayed directly on the model icon.

The generate icon displays the uncertain workload statistics for each time step of the model. Maximum, average, minimum, and standard deviation of product demand in jobs per month are displayed on successive lines. The bottom line indicates time and the bar to the right dynamically plots average demand arrival rate. In node mode, the activity blocks show throughput, work in process, node delay, and activity cost on successive lines. Throughput is the rate that jobs are finished and leave the activity for the next one. Work in process is the number of jobs in the activity backlog, or

being serviced by the activity. When arrival rate to an activity exceeds its service rate capacity, jobs join a backlog awaiting processing. Delay is the average activity delay including backlog and service times. Cost is the total cost for that activity per month. It includes fixed cost, variable cost times throughput, and storage cost times delay and work in process components. The final output summary block on the lower right shows process statistics. Throughput, work in process, delay, and cost are shown for all activities from assemble to pack and ship. In cumulative mode the cumulative values of work in process, delay, and cost are shown on the activity icons. The generate and output summary blocks are the same. Note that in this mode the output summary block values are identical to those of the last pack and ship activity.

2.2 Generate Block: Dynamic Uncertain Workload Input

Figure 3 shows the generate block dialog. The user inputs the arrival rate (jobs/month) range for the workload graph and then clicks on the edit button. This brings up the figure 4 plot.

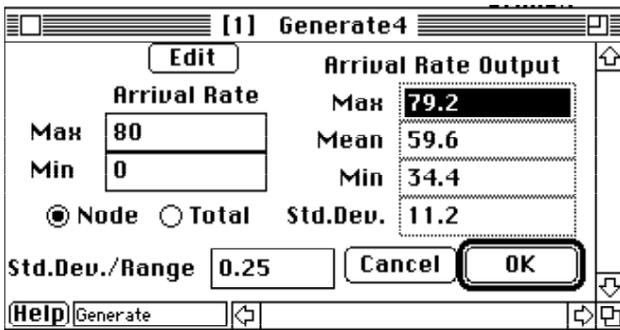


Figure 3. Generate Block Dialog

The user can sketch in the dynamic minimum, mean, maximum, and standard deviation statistics over the model run time. A table of numerical statistics values is provided below the graph. Numerical values can also be entered via this table. The dynamic maximum, mean, minimum, and standard deviation values are shown to the right of the dialog box and on the generate icon face. Node and total (cumulative) radio buttons are provided to set the model display mode. Note that there is a significant rising seasonal pattern to the workload. The model thus provides the capability to input a dynamic workload uncertainty distribution.

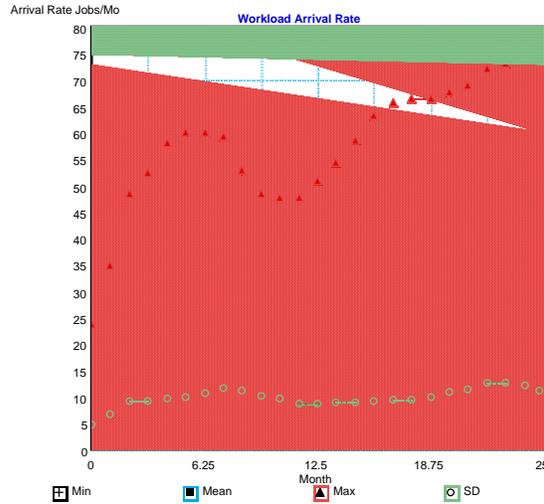


Figure 4. Workload Statistics Versus Time (Edit Button)

2.3 Activity Block: Process Input and Calculated Outputs

Figure 5 shows the inspection activity block dialog.

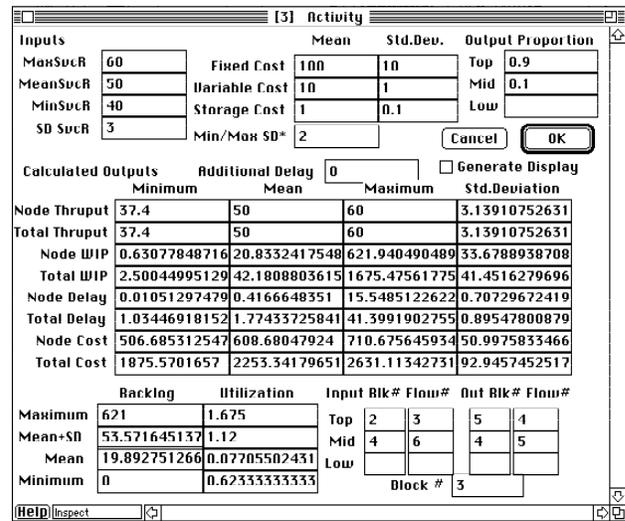


Figure 5. Inspection Activity Block Dialog

User process inputs are shown above and calculated outputs are shown below. The activity capacity is measured by its service rate in jobs per month. The uncertain capacity is input in the upper left as maximum, mean, minimum, and standard deviation in service rate. Activity costs are input to the right. Fixed cost per month, variable cost per job, and storage cost per job month are input in successive cells. Uncertainty in these costs is specified with the standard deviation cells to the right. The minimum maximum range in costs is specified in the "Min/Max SD*" below the costs. In the case displayed, the minimum cost is two standard deviations below the

mean, and the maximum cost is two standard deviations above. Output proportion is input to the right of the cost standard deviation inputs. In this case ninety percent of the products pass inspection and are routed to the pack and ship activity. Ten percent of the products fail inspection and are routed to rework. After rework they are inspected again. Below the cost inputs is a cell for additional delay. This is added to the service time and backlog delays when calculating node delay. In this example all additional delays are set to zero. To the right is a "Generate Display" check block. Checking this box stores the cumulative activity throughput, work in process, delay, and cost results for later plotting by the output block. Normally this is checked on the last model activity. However, this provides the capability to output results to an intermediate process activity.

Calculated model outputs are shown below the input section, Minimum, mean, maximum, and standard deviation statistics are shown in successive columns. Alternating node and cumulative total rows are shown for each parameter. The later cumulative total is the sum of parameter values of all node activities preceding the node plus the node value. Throughput, work in process, delay, and cost parameters are displayed in successive row pairs. Backlog and utilization statistics are shown in the bottom left of the dialog box. These values are shown for maximum, mean plus standard deviation, mean, and minimum arrival rate loading of the activity. To the right input and output connector link and block numbers are shown. These can be ignored, since they are automatically generated when the user graphically links the blocks together.

2.4 Output Summary Block

The output summary block is in the lower right of the model displayed in figure 2. It presents cumulative statistics for the activity with the "Generate Display" check box selected. In our case this is the last pack and ship activity. Figure 6 displays the dialog box.

History	Minimum	Mean	Maximum	Std.Dev.
Thruput	33.66	45	65	7.57537119649
WIP	2.82475212177	39.1429531443	1716.30151633	52.826050055
Delay	1.04330529176	1.79842217308	46.0297116166	1.18734554948
Cost	2145.84612845	2578.03604093	3010.2259534	95.5118755548

#Cells 36 Cancel OK

Figure 6. Output Block Dialog

Total process throughput, work in process, delay, and cost are shown in successive rows. Minimum, mean, maximum, and standard deviation statistics are shown in

successive columns. Buttons to the left produce plots of results.

The first "History" button produces the figure 7 plot.

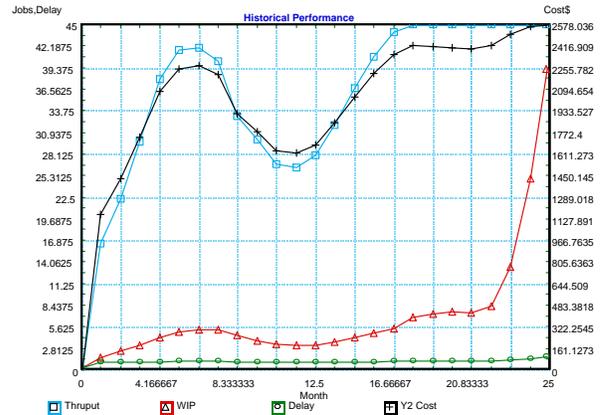


Figure 7. Historical Performance (History Button)

Note the dynamic changes in throughput, work in process, delay, and cost. Cost is plotted relative to the right axis, while the other parameters are plotted relative to the left axis.

The buttons below the "History" button fit beta distributions to the statistics at the end of the simulation run. Figures 8-10 show results for throughput delay and cost. Both density and cumulative distribution functions are shown.

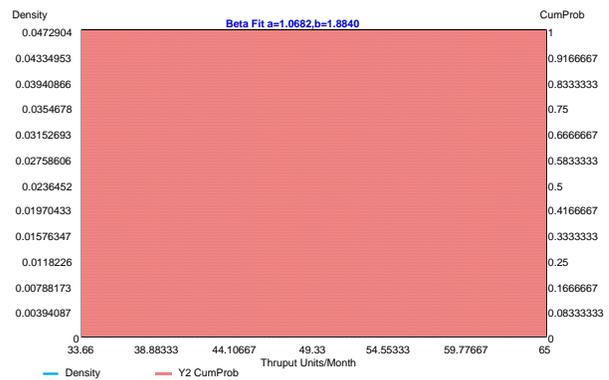


Figure 8. Process Throughput Probability Distribution

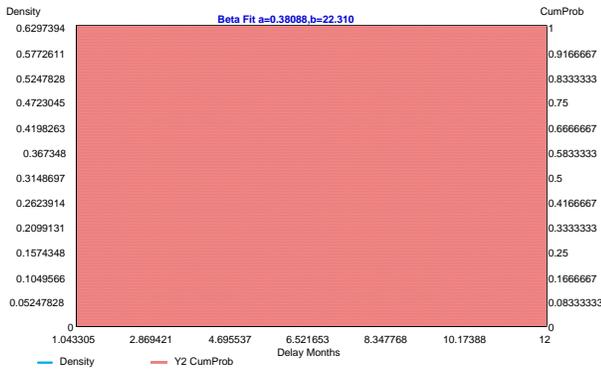


Figure 9. Total Processing Delay Probability Distribution

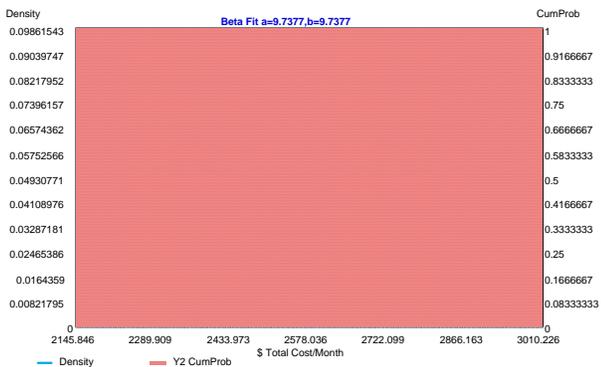


Figure 10. Total Cost Per Month Probability Distribution

Note the different beta density function shapes. The fit beta a and b parameters are calculated based on the minimum, mean, maximum, and standard deviation statistics. The output risk distributions are nonlinear functions of uncertainty in:

- Workload demand arrival rate
- Activity service rate
- Activity cost uncertainty (fixed, variable, storage)

The analytic queuing process is highly nonlinear and greatly impacts the output risk distributions. The next section and paper [ref.1] outlines analytic queuing. The distributions allow one to give confidence statements about process delay, throughput, work in process, and cost at any time in the process. This provides a dynamic risk evaluation capability for the first time. Note that these estimates include all the process interactions between variables. Most present risk analysis techniques treat processes linearly and do not account for parameter and process interactions.

2.5 Transient Analytic Queue Approximation

The previous model is based on two new paradigms, transient analytic queuing, and analytic risk evaluation. More details are found in references 1 and 2.

The mean steady state queue wait for a G/G/1 queue is given by:

$$\text{Wait} = \tau \rho (C_a^2 + C_s^2) / (2(1-\rho))$$

Where:

τ = service time

C_a^2 = square coefficient of variation in inter-arrival time
(variance in inter-arrival time divided by the mean inter-arrival time squared)

C_s^2 = square coefficient of variation in service times

ρ = utilization (service time divided by inter arrival time)

The relaxation time or time to reach steady state in discrete event simulation is two times the above wait divided by one minus the utilization. At high utilization the time to reach steady state is far longer than the time a dynamic workload remains close to a given value (the relaxation time equation has a $(1-\rho)^2$ term in denominator). In the limit, it takes infinite time to reach steady state at unity utilization. At low utilization's (<.8) the steady state wait equation gives reasonable answers. Between .8 and unity utilization we calculate a corrected utilization adding one half the utilization above .8 to .8 $(.8+(u-.8)/2)$. Thus the highest utilization used in the steady state delay equation is .9. More complex models have been developed accumulating time at different utilizations and correcting for the relaxation time, however the improvement in accuracy does not justify the added calculations.

When utilization exceeds unity, the excess over unity is added to work backlog. Conversely when utilization is less than unity, the work backlog is reduced. The mean node delay is the queue wait, plus the service time, plus the service time multiplied by the number of jobs in the queue backlog. The coefficient of variation in service time (C_s = standard deviation in service time divided by mean service time) can be calculated from service observations. The coefficient of variation in inter-arrival times to the process can be calculated from the times between work requests. The coefficient of variation of inter-arrival times to the second node is based on the utilization of the first node. At high utilization it is identical to the first node coefficient of variation of

service times. At low first node utilization it is close to the first node inter-arrival coefficient of variation. In between we use a utilization weighted sum of the two coefficients of variation (squared utilization weight).

2.6 Analytic Risk Evaluation

Analytic risk evaluation is based on the beta distribution. The sums and products of beta variables are approximately beta distributed. Reference [1] documents the minimal error from approximation. One does not have to resort to computationally intensive convolution or mellin transforms to calculate the sums and products of random variables. All that is required is to calculate minimum, mean, maximum, and standard deviation statistics. Cumulative statistics along the process activity path can be calculated. The beta density distribution is bounded by high (max) and low (min) values.

$$P = C \left\{ \frac{(x-\min)}{(\max-\min)} \right\}^{a-1} \left\{ 1 - \frac{(x-\min)}{(\max-\min)} \right\}^{b-1}$$

Where:

P is the probability density

C normalizes the area under the distribution to unity

$$(C = 1 / \int_0^1 x^{a-1} (1-x)^{b-1} dx)$$

min is the minimum variable value

max is the maximum variable value

The beta distribution can be fit based on the minimum, maximum, mean and variance. The fit parameters a and b are based on a range variable r and a skewness variable s, and are defined in standard terms as:

$$r = \text{variance} / (\max - \min)^2$$

$$s = (\text{mean} - \min) / (\max - \min)$$

$$a = s^2 (1 - s) / r - s$$

$$b = s (1 - s) / r - 1 - a$$

If $(a-1)(b-1) > 0$ then the beta distribution has a mode at

$$\min + (\max - \min) (a - 1) / (a + b - 2)$$

Figure 11 shows examples of different beta distribution shapes. For the mean midway between the minimum and maximum values, the distribution is symmetric with equal a and b parameter values. Unity a and b parameters yield a uniform distribution (1,1). Values less than 1 lead to a "U" shaped distribution (.5,.5). With a and b equal 2 (2,2) the distribution has a

parabolic shape. At higher a and b values (6,6; 12,12) the distribution has a shape similar to a normal distribution. The lower the standard deviation relative to the range, the higher the a and b parameter values, and the more peaked the distribution shape. If a is greater than b then the distribution has a negative skew (9,2). Conversely, if a is less than b the distribution has a positive skew. Triangular distributions (1,3) and left (.5,3) and right "J" shaped distribution shapes are also possible.

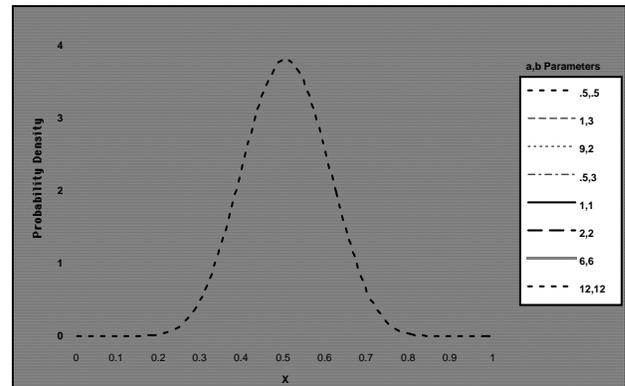


Figure 11. Beta Distribution Shapes

An important part of examining COTS in a BPR effort is identifying and evaluating COTS applications suitable for the organization's business process. These COTS applications are frequently associated with a best practice. Business COTS applications offer the essentials of a standard solution.

4. Conclusions

Hands-on analytic network queuing is a new technique that provides the capability for real time interactive model development. It is based on an approximate analytic solution to the general queuing problem which can handle time varying workloads. The deterministic solution simplifies sensitivity analysis. Complex, many factor problems are now solvable for the first time. Processes are composed of networks of activities. The activity nodes take time to perform their task and consume resources. The network links carry product and resource flows. A model can be constructed by wiring together two types of building blocks: activity and generate. The generate block originates the process workload. The process itself is just a linked set of activity blocks.

Analytic risk evaluation is based on the beta distribution. It provides the entire uncertainty probability distribution vice an uncertain estimate of the mean obtained by discrete event simulation. The beta distribution can be fit based on minimum, mean, maximum, and standard deviation statistics. When analytic queuing is combined with analytic risk

evaluation, dynamic risk evaluation becomes feasible for the first time. The time varying output uncertainty distribution can be calculated at any specified time or over any user specified time interval. This new capability is not available in discrete event simulation.

References

[1] Neimeier,H. *Analytic Queuing Network*, 1994 International System Dynamics Conference Proceedings, Stirling Scotland, August 1994.

[2] Neimeier,H. *Analytic Uncertainty Modeling*, 1994 International System Dynamics Conference Proceedings, Stirling Scotland, August 1994.

[3] Neimeier,H., Gullledge,T., *Functional Economic Analysis of Purchasing at MITRE*, 1994 International System Dynamics Conference Proceedings, Stirling Scotland, August 1994.