Proceedings of the 2001 Winter Simulation Conference
B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, eds

QUANTIFYING SIMULATION OUTPUT VARIABILITY USING CONFIDENCE INTERVALS AND STATISTICAL PROCESS CONTROL

Amy Jo Naylor

Corning Inc.
Corning, NY 14831, U.S.A.

ABSTRACT

Two types of variability can occur in model output: variability between replications and variability within each replication. The objective of the model combined with the type of output variability determines which tool is more appropriate for output analysis. Many output analysis techniques are used to translate simulation model results into a format that answers the model objective. This paper compares two tools for output analysis: confidence intervals and statistical process control. Each tool quantifies a different type of variation from the model results. As such, statistical process control is applied beyond monitoring the consistency of run data. A supply chain example with one factory, multiple parts, and multiple distribution centers is used throughout the paper to illustrate these concepts.

1 INTRODUCTION

Simulation models answer questions for processes that contain variability and interactions. A clearly defined objective directs the scope, level of detail and model development for any simulation model (Sanchez 1999 and Centeno and Reyes 1998). To answer the objective properly, the methodology used to analyze the output must be carefully selected. Many techniques for simulation model output analysis are described in Sanchez (1999), Centeno and Reyes (1998), Goldsman and Tokol (2000), and Alexopoulos and Seila (2000).

The commonality of all simulation output analysis methodologies is that an expected interval is generated, *not* a static number. Intervals are the result of the variation in the model (Sanchez 1999). One type of simulation output variation is the variability *between* replications. In this instance each replication generates one representative value. Examples are cumulative data, such as a count or rate. The second type of simulation output variation is variability *within* replications. In this case values are generated at regular time units during the replication. Many capacity objectives require understanding within replication variation.

This paper focuses on recognizing the type of output variation and the appropriate tool to analyze it. While many tools are available for analysis, confidence intervals and statistical process control (SPC) are demonstrated. These concepts are described throughout the paper using a supply chain model developed by the author. The paper first describes general conflicting objectives in any supply chain system. Next, the objectives, workflow, and model development are described for the example. The two objectives are reviewed for their type of variation and are answered. Comments are provided on how to interpret out of control charts. Additional examples are provided to further clarify the differences between the two types of output variability. Finally, the i. i. d. issue, specifically autocorrelation, is mentioned.

2 CONFLICTING SUPPLY CHAIN OBJECTIVES

Every supply-demand chain must overcome the variation inherent in production and delivery systems. Typically, inventory is held to buffer this variability. (Various buffers and their relation to variability are well described in Standard and Davis (1999).) However, holding inventory comes at a high cost. On hand stock represents investment in a desired customer service level. Additional money must also be spent on handling costs, overhead storage costs, and the damaged or obsolete units in inventory. Reducing inventory as low as possible while simultaneously maintaining an identified level of product availability for the customer can be quite a challenge. Simulation is ideal for providing insights into supply production, storage, and distribution systems that are inherently replete with variability and interactions. However, before the supply-chain model can be developed, an objective must be clearly defined.

3 OBJECTIVES AND WORK FLOW

The objective for this example is two-fold: given specific maximum inventory target levels, identify the expected effect on customer level of service (LOS) and the daily expected total inventory value. Reviewing these objectives throughout the multiple model development stages will ensure the appropriate level of detail is achieved and the correct data are identified and collected. The supply process being modeled is diagrammed in Figure 1.

This supply chain is a pull system where the daily reductions in on hand stock cause a ship quantity to be identified according to a predetermined schedule. As a result of the ship quantity being removed from the appropriate kanban, manufacturing builds an equal number of units to replenish the kanban. Ship quantities are determined by subtracting the sum of the in transit and on hand units from the maximum inventory level and then adding any backorders for each part at each distribution center (DC).

The variability in daily customer demand drives the system and therefore the model. Daily demand data for each part going to each DC were located and analyzed (including fitting to a distribution). Daily demand is the catalyst for activity throughout the supply chain.

Interactions are part of the system due to the supplying structure: one manufacturing site supplies many parts to many different DCs. For this example, one factory supplies three products to four different DCs. Interactions occur in three ways: one production line in the factory assembles two of the three parts, one factory supplies four DCs, and the total of all three products shipped to each DC is constrained by truck capacity.

4 MODEL DEVELOPMENT

Three parts going to four DCs results in a requirement to track twelve data streams. Each data stream represents a supply part and DC pair or set. Each set is characterized by a probability distribution, defined to model the variability in the observed daily demand for the supplied parts, and a maximum inventory level, which is used for shipment calculations. For each simulated day, the probability distribution identifies a unique demand quantity to be removed from the on hand balance for each part-DC set. If it is the day before the scheduled ship day to the DC, then a ship quantity is calculated for each part at that DC, parts are withdrawn from the appropriate kanban, and the factory refills as much of the kanbans as possible at its defined production rate.

Data for each part-DC set are tracked and collected daily, including the ship quantity for the day before shipment, the quantity in transit, the on hand balance at the DC, the backorder quantity at the DC, the number of days demand is met at the DC (# Delivered Days), and the number of days demand is not met at the DC. The last two values are used to calculate the overall customer LOS for each part-DC set.

The performance measure for the first objective, LOS, is calculated for each part-DC set using the formula *LOS = # Delivered Days/# Total Days*. The total inventory relationship is calculated as *Total Inventory = Quantity In Transit + On Hand Balance – Backorder Quantity* to answer the second objective. The LOS and Total Inventory Quantity metrics differ in that the former generates one value for each replication while the latter generates a data point each time unit (day) of the replication. LOS is only the final ratio; it is not relevant to know the LOS half way through or at any point except at the end of the simulation replication. As such, the interval to answer the first objective will reflect between replication variation. For the Total Inventory Quantity, each day of the replication is important to understand the variation during the time period (replication length) thus the interval answering the second objective will reflect the variation within the replications.

5 ANSWERING OBJECTIVE 1: CUSTOMER LOS

The first objective for this model is to identify the expected effect of specified maximum inventory levels on customer LOS. Each replication yields one LOS value. It is impor-

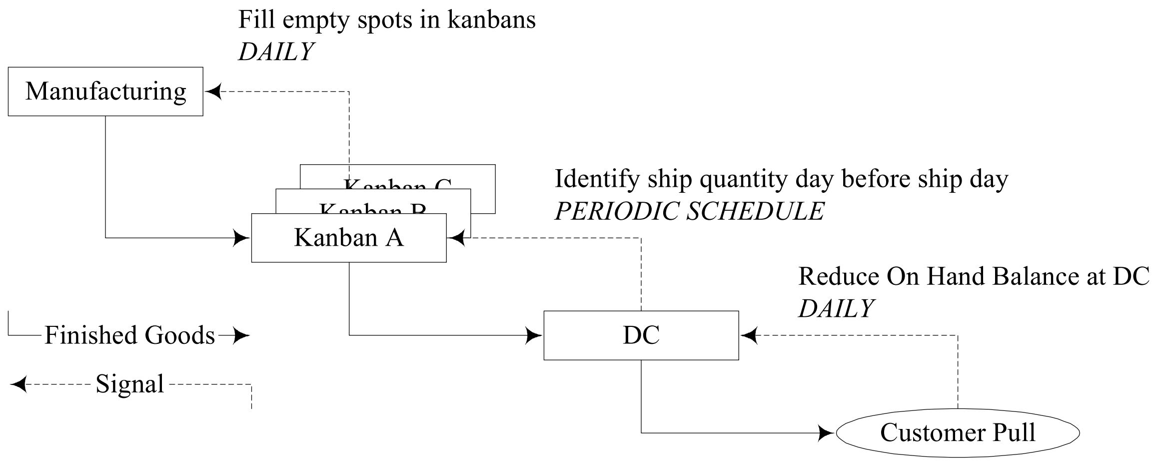

Figure 1: Supplying Process for One Distribution Center

tant to note that this one value is the actual value calculated for the entire length of the replication; it is *not* an average for the replication. The variation in LOS is observed solely from replication to replication. This one source of variability is captured as the standard deviation of the LOS values across all replications. Therefore, calculating a confidence interval for the expected span of customer LOS for the indicated maximum target level is appropriate. The confidence interval incorporates the variation between the replications and matches the type of variation found with LOS model data.

For example, when a 95% confidence for a specific part-DC set's maximum inventory quantity of 970 units is used, the expected LOS is between 95.3% and 98.9%.

6 ANSWERING OBJECTIVE 2: EXPECTED INVENTORY VALUE

The second objective for this model is to identify the expected effect of maximum inventory levels on the daily total inventory value. The inventory position changes every day in reaction to the variable daily demand and the shipping schedule. Each of the ten replications was run for 20 days beyond the warm up period. Merely taking the average total inventory for each replication and calculating a confidence interval for the averages yields an artificially narrow confidence interval that does not fully represent the actual daily variation in inventory value. By applying a standard confidence interval for inventory, the variability between the replication averages is reflected and the variability within each replication is lost.

For example, calculating the confidence interval for inventory position for the same part-DC set yields a range of 811 to 843 units. Multiplying by the cost per unit of \$160 yields an expected inventory value for this one set to range from \$129,760 to \$134,880. Does this interval answer the objective? If the objective were to identify the expected *average* inventory value then this would be an acceptable result. However, that is not the objective. Does this interval truly represent the daily variation represented in the inventory value? It does not. In fact, it is highly likely that inventory value will oscillate above and below this interval quite regularly. Indeed, the daily inventories generated by fitting each part-DC set's daily demand to a probability distribution are severely truncated by not accounting for the day to day variation of inventory position. A confidence interval loses all of the variation observed within each replication.

A tool that accounts for within replication variation is SPC. In this example the Shewhart control charts are applied. There are two phases to this control chart methodology. First, a dispersion (typically range or standard deviation) chart is generated to see if the variability is equal for each replication. If the replication to replication variation is equal (i. e. within the calculated control limits or in control), then the average chart and the expected limits are plotted to identify if the replications are equivalent or different. The single most important feature of using the control chart methodology is that day to day variation, which is found within each replication, drives both sets of limits. Understanding and quantifying this variation is essential to properly answering the inventory value objective.

Figure 2 is the dispersion chart for the same part-DC set discussed earlier. It shows that the within replication variation is equal across all ten replications. For the example this means that the standard deviation of the total inventory quantity for Part C at DC 4 is expected to vary between 84 and 168 units during any month.

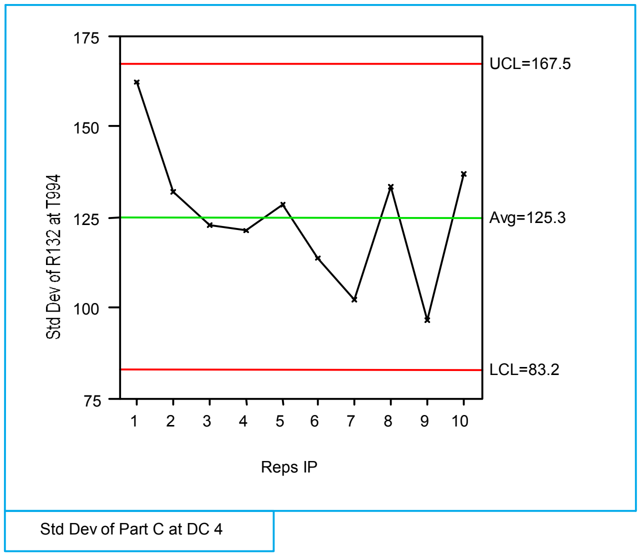

Figure 2: SPC Dispersion Chart for Part C at DC 4

Figure 3 is the average chart of the same part-DC set. The average within replication standard deviation for the total inventory quantity is applied to the grand average to identify the limits. Interpreting the model results this way predicts a daily inventory of 768 to 886 Part C units at DC 4. The resulting range of expected inventory value for this part-DC set is \$122,880 to \$141,760. The second objective is answered completely by plotting all twelve sets of charts, summing all of the limits, and calculating the inventory value.

The mechanics of constructing control charts can be found in many texts (Wheeler and Chambers 1992) and on the World Wide Web (NIST/SEMATECH).

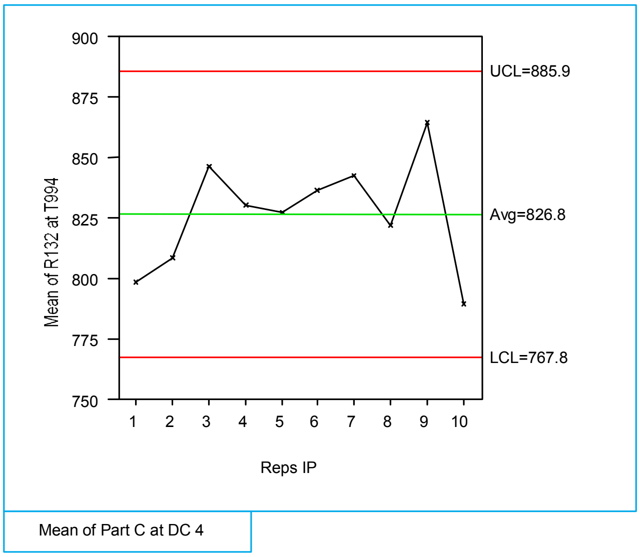

Figure 3: SPC Average Chart for Part C at DC 4

Using confidence intervals versus SPC yields very different results for the second objective (Table 1). As expected, the total inventory position across the entire supply chain yields a much wider range using SPC since the within replication variation is included. In this example, the interval is five times wider using SPC than when 95% confidence intervals are calculated.

Table 1: Comparing Supply Chain Intervals for Total Inventory Position

	95% Confidence Interval	SPC Interval
Units	(5,267, 5,436)	(4,928, 5,772)
Cost	($842,720, $869,760)	($788,480, $923,520)

Is one interval more correct than the other? Does it matter which interval is reported to answer the objective? The answer to both of these questions is yes. The SPC interval is the more appropriate result since it includes the within replication variation which corresponds to the natural day to day variation in the system. Answering the objective with the proper interval is important because, by encompassing the day to day variation, the simulation results are more credible. When the customer compares actual results to the SPC interval they are more likely to see the actual total inventory quantities within the interval thus gain additional appreciation for the credibility and usefulness of simulation models. This is more likely to occur because the SPC interval includes the day to day variation instead of the variation observed by using only the averages.

7 INTERPRETTING OUT OF CONTROL CHARTS

What if either chart is unpredictable (out of control)? This happens when a replication measure exceeds either the dispersion or average limits. To identify one potential cause, determine if the input data was in control. It is possible that the input data can have a probability distribution but its variation makes it unpredictable. In this case a process is developed and modeled from an unpredictable data stream. While the desired state is to have in control variation – and eventually to reduce variation – sometimes processes must be developed to handle out of control variability. In the supply chain example, the direct cost for out of control variation is a greater buffer of parts (greater maximum inventory levels).

7.1 Out of Control Dispersion Chart

The dispersion value (for example, range for small n or standard deviation for larger n) is calculated for each replication. An out of control dispersion chart – when a replication's dispersion exceeds the limits – means the within replication variation is not equal; variation is unpredictable from replication to replication. Technically, an out of control dispersion chart means that the model's objective cannot be answered definitively since any resulting average chart is not valid.

7.2 Out of Control Average Chart

An out of control average chart – when the average of any replication exceeds the limits – indicates that the variation between the replications is greater than the variation within the replications. What this means to interpreting the model's results depends on the objective. For the inventory scenario used, an out of control average chart indicates that the limits for inventory value are only an estimate; it is expected that actual inventory value will exceed the identified limits.

8 ADDITIONAL APPLICATIONS

Determining which of the two techniques – confidence intervals or SPC – should be used depends on identifying a specific objective and recognizing the type of variation in the output data for that objective. This approach can be applied to many different applications as indicated in Table 2. Included are output and replication length modeling decisions that also determine which type of analysis is appropriate. ('CI' represents 'Confidence Interval'.)

Table 2: Application Examples

Objective	Output	Replication Length	Output Variation	Analysis
Identify shipping dock capacity; dock is empty at the beginning and end of each day (terminating model)	Total daily time dock is busy	One day	Day to day variation is between replications	CI
Identify shipping dock capacity; day can begin and end with trucks and/or units in the dock (non-terminating model)	Total daily time dock is busy each day	One month	Day to day variation is within replication	SPC
Identify product line capabilities required to meet variable customer demand (pull signal) from many locations (or for many products on one line)	Total quantity of units required each day	One month / quarter	Day to day variation is within replication	SPC
Identify the expected number of days production is not met per month	Total number of days did not make enough units	One month	Month to month variation is between replications	CI
Identify the expected warehouse capacity needed	Total number of units in the warehouse each day	Six months	Day to day variation is within replication	SPC
Identify the expected rate for how often the quantity of units at a warehouse is less than its safety stock level	Proportion of days inventory is below safety stock level during one month	One month	Month to month variation is between replications	CI

9 I. I. D. ISSUES

The within replication simulation model output is typically not independent and identically distributed (i. i. d.) (Goldsman and Tokol 2000, Sanchez 1999, Centeno and Reyes 1998, Alexopoulos and Seila 2000). Specifically, the data can be collected so that they are identically distributed but they are rarely independent. As such, autocorrelation issues for the data within the replication are relevant. However, using SPC the within replication variation is still calculated for the dispersion chart. Opinions vary on whether this is valid. For example, Wheeler and Chambers (1992) contend that while virtually all production data is autocorrelated the control chart methodology still provides valuable insights to make decisions. Excessive autocorrelation actually tightens the control limits thus making the data more susceptible to identifying out of control behavior and, in the application for this paper, causes the interval that is reported to be reduced. Completely opposite of Wheeler and Chambers are Standard and Davis (1999) who state that most control chart methodologies assume i. i. d. data. They continue that autocorrelation violates SPC assumptions. This controversy regarding the appropriateness of statistic measurements for control charts in respect to the lack of data independence is entrenched and beyond the scope of this paper.

There are at least two ways through this quandry. First, if there are a small number of n in the subgroup (roughly $n < 20$) then range can be used for the dispersion chart. Second, Standard (1997) proposes a methodology for SPC with autocorrelated data.

10 CONCLUSION

There are many steps to define, develop, verify, and analyze a simulation model. An additional step is to recognize the type of variation that is requested in the objective and then use the appropriate tool to capture this variation during the output analysis. This important concept of recognizing and measuring the appropriate form of variation is demonstrated in this paper. Confidence intervals are recommended for metrics that have one data point for each replication. This type of measure is typically cumulative. Confidence intervals quantify the between replication variation. Statistical process control is applied to metrics that vary regularly (from time unit to time unit) within a replication. Capacity related questions are frequently answered with SPC methods since they quantify within replication variation.

ACKNOWLEDGMENTS

Appreciation is extended to Norman L. Reitter and Michael E. Zak for their insights. Thank you to Lois A. Eckstrom for encouraging continual questioning of variation and to Deborah J. Medeiros for molding simulation mechanics into objective driven, thought provoking opportunities.

REFERENCES

Alexopoulos, C. and A. F. Seila. 2000. Output analysis for simulations. In *Proceedings of the 2000 Winter Simulation Conference*, ed. J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 101-108. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Centeno, M. A. and M. F. Reyes. 1998. So you have your model: what to do next a tutorial on simulation output analysis. In *Proceedings of the 1998 Winter Simulation Conference*, ed. D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan, 23-29. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Goldsman, D. and G. Tokol. 2000. Output analysis procedures for computer simulations. In *Proceedings of the 2000 Winter Simulation Conference*, ed. J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 39-45. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

NIST/SEMATECH. Univariate and multivariate control charts. `http://www.itl.nist.gov/div898/handbook/pmc/section3/pmc3.htm`, accessed March 27, 2001.

Sanchez, S. M. 1999. ABC's of output analysis. In *Proceedings of the 1999 Winter Simulation Conference*, ed. P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, 24-32. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Standard, C. 1997. Process control in the presence of autocorrelation. In *6th Industrial Engineering Research Conference Proceedings,* ed. B. Bidanda and S. Jagdale, 101-106. Norcross, GA: Institute of Industrial Engineers.

Standard, C. and D. Davis. 1999. *Running today's factory: a proven strategy for lean manufacturing.* Cincinnati: Hanser Gardner Publications.

Wheeler, D. J. and D. S. Chambers. 1992. *Understanding Statistical Process Control*. 2d ed. Knoxville: SPC Press.

AUTHOR BIOGRAPHY

AMY JO NAYLOR is a Process Engineer at Corning Inc. The work for this paper was completed while a Production Engineer at Xerox Corporation. She received her BA in Mathematics from Alfred University in 1993 and her MS in Industrial Engineering and Operations Research from the Pennsylvania State University in 1998. Her email address is `AmyJo@reitter.net`.