

DOWNTIME DATA -- ITS COLLECTION, ANALYSIS, AND IMPORTANCE

Edward J. Williams
114-2 Engineering Computer Center, Mail Drop 3
Ford Motor Company
Post Office Box 2053
Dearborn, Michigan 48121-2053, U.S.A.

ABSTRACT

Until the day when plant production personnel and equipment have no downtime, proper collection and analysis of downtime data will be essential to the development of valid, credible simulation models. Methods and techniques helpful to this task within simulation model building are described.

1 INTRODUCTION

Ford Motor Company is steadily increasing its use of simulation to improve the design of production processes, both those still on the drawing board and those currently in operation. To be valid and credible, these simulation models must include expected or actual downtime experience. Since the collection of downtime data represents heavy investments in both time and cost, it is important to recapture these investments via the benefits of using valid and credible simulation models. The following considerations, to be discussed sequentially in the remainder of this paper, all pertain to the valid modeling of downtime:

- invalidity of common simplifying modeling assumptions
- techniques of downtime data collection
- describing the downtime correctly in the modeling tool being used.

2 INVALIDITY OF COMMON SIMPLIFYING ASSUMPTIONS

The most brash assumption is to ignore downtime altogether. Unless downtime never occurs (a situation never yet seen in our process-engineering practice), omission of downtime analysis produces an invalid model. Fortunately, such an invalid model also has no credibility, and hence will not be used by management to reach wrong conclusions.

Another, more plausible, simplifying assumption is to

- observe that downtime is a certain percent of total simulated time
- run the model with no downtime
- factor its throughput downward by the percentage of downtime.

This assumption is typically unworkable for two reasons. First, very rarely does the downtime itself

pertain to the entire system being modeled. Second, the analysis outlined above applies a downtime "correction" to the throughput statistic only. In practice, performance statistics other than throughput are of concern to the user. For example, a process engineer designing line layout must determine the maximum queue length upstream from a certain operation. Hence, this simplifying assumption is best reserved for rare system-global downtimes. For example, if records show that a certain plant shuts down a given number of scheduled production days per year due to snowstorms, the computation above is well-suited to evaluate the overall productivity of the plant.

A variant of this assumption may be applied to each machine individually. For example, if a machine's cycle time is a constant x and the machine is down a fraction y of total time, this assumption models the machine's cycle time as $x/(1-y)$. This variant likewise tends to estimate global performance metrics such as throughput well, but estimate local performance metrics such as maximum queue lengths poorly.

A third simplifying assumption is "the downtime duration is a constant equal to its mean," and hence replaces a random variable representing downtime duration with that mean value. This assumption typically produces an invalid model which overestimates throughput. Downtimes markedly longer than the mean exhaust downstream buffer stock; once that stock is exhausted, downstream operations suffer unproductive time which can never be recouped. Similarly, upstream operations experience severe backup which the invalid model will fail to represent as high queue-length maxima. Vincent and Law² describe an analogous pitfall arising from replacing a processing time by its mean. A variant of this assumption models downtime with a uniform or triangular density. These densities are often useful "rough-draft" approximations for model verification. However, the uniform has no unique mode, neither the uniform nor the triangular has inflection points, and both the uniform and the triangular have finite ranges. Therefore, these densities should not remain in the model without validation that these constraints are appropriate to the downtime being modeled.

3 TECHNIQUES OF DOWNTIME DATA COLLECTION

In industrial practice, the model builder visiting the production floor must often work with non-technical personnel unacquainted with simulation analyses; in turn, those employees often have to answer questions based on scanty or disorganized data. We have encountered the following problems and devised the following countermeasures:

- Problem: Production workers record as a downtime interval a period of time during which the machine is performing no work.
Solution: Explain the terms "starved" -- the machine is ready to work but has no work to do, "blocked" -- the machine has finished work but has no room downstream and hence can't unload the workpiece to accommodate another, "busy" -- the machine is doing productive work, and "down" -- the machine has malfunctioned and needs service. Clarify that the last category represents a downtime interval, and that the first three categories collectively represent an uptime interval.
- Problem: Production workers record a single number representing the percent of time a machine is down.
Solution: Explain that "percent downtime" alone provides too little information -- for example "10% downtime" might indicate that a machine typically operates normally for nine minutes and then goes down for one minute, or that a machine typically operates normally for nine hours and then goes down for one hour. Among the three metrics "percent downtime," "mean time to fail" [MTTF], and "mean time to repair" [MTTR], any two determine the third.
- Problem: After downtime data is collected, it proves inadequate for cycle-based downtime modeling.
Solution: Record the number of machine cycles completed during each uptime interval, in addition to recording the duration of that interval.
- Problem: The shortest downtimes go unrecorded because recording them takes nearly as much time as repairing them.
Solution: Ideally (but expensively), assign an incremental worker to record these downtimes while the production worker repairs them (e.g., by clearing a jam). Or, in addition to collecting the downtime data logs, ask production personnel a question such as "How many downtimes lasting less than a minute do you typically fix each hour?"
- Problem: In an operation running continuously across shifts, the downtime data are inconsistently recorded and/or subdivided across shifts.
Solution: Provide recording forms and instructions common to the different people recording uptime and downtime durations across each shift. Coalesce data intervals across shift changes. For example, suppose the data logs show:
Machine A repaired at 11:40 PM (recorded by shift 1),
Shift change at 12 midnight,
Machine A went down at 12:50 AM (recorded by shift 2).
These data indicate one uptime interval of 70 minutes, not two separate uptime intervals of 20 and 50 minutes.
- Problem: In a particular modeling context, the downtime interval may need further subdivision.
Solution: Ask the following questions:
Typically, how long is a machine down before production personnel notice that it is down? Once the downtime is noticed, how long does it take needed repair resources (maintenance workers, equipment) to reach it? Then, once the repair begins, how long does it take? Non-zero answers to the first two questions indicate that the model builder must subdivide the downtime interval accordingly. For example, if the first answer is non-zero, neglecting subdivision of the downtime will lead the modeler to allocate repair resources to the entire MTTR interval, thereby overestimating the utilization of repair resources.
- Problem: The MTTF for a machine may be only weakly correlated with elapsed time.
Solution: Assess the machine operation to decide whether the MTTF should be based on elapsed time, service time, or cycles completed. For example, a machine which, whether actually operating or not, draws power from a battery, will probably have battery-recharge downtimes based on elapsed time. A polishing machine will probably have abrasive-replenishment downtimes based on service time, irrespective of whether the service time comprises long segments polishing a few large workpieces or short segments polishing many small workpieces. A drilling machine will probably have drill-bit-replacement downtimes based on cycles completed, i.e., number of holes, of uniform diameter and depth, drilled in workpieces.
- Problem: No downtime data exists for a machine (as often occurs when a process still under design

is to be modeled and the machine and its vendor are not yet chosen).

Solution: Using experience from similar situations and similar machines, develop a best-case and worst-case scenario for the downtime of the machine. When developing these scenarios, consider the following:

- MTTF may be approximately proportional to the total number of components in the machine
- MTTR may be approximately proportional to machine complexity
- if the new machine will be installed in a different plant, that plant's operating conditions, tooling, and/or maintenance practices may differ from those of the plant using the currently similar machine.

Run the model under both scenarios (sensitivity analysis, section 4) to assess the effect of changes in the reliability of this machine. If this machine thus proves to be a critical point of the system, alert candidate vendors of this criticality. Incorporate reliability-performance criteria into contractual terms.

4 MODELING CONSIDERATIONS

4.1 Choosing an Appropriate Probability Density

Since downtime (and uptime) durations oughtn't to be replaced with their means, an appropriate probability density must be included in the simulation model. The temptation to use the existing data as an empirical density should usually be avoided, because doing so tacitly assumes that any duration shorter than the sample minimum or larger than the sample maximum is impossible. This assumption is almost always untenable.

That said, the choice of an appropriate theoretical density becomes important. The following steps will assist in choosing one:

- Before undertaking calculations, plot a histogram of the available data and compare its shape with those of the candidate probability density functions.
- Compare properties of the empirical data set with those of a candidate theoretical density.
- Assess the goodness-of-fit with statistical tests such as the chi-square, Kolmogorov-Smirnov, and Anderson-Darling tests.

For example, a normal density should be avoided if its standard deviation, relative to its mean, is large enough to imply occasional durations less than zero. Also, since the mean, median, and mode of a normal density are all equal, a normal density should be avoided if these equalities conspicuously fail to hold for the sample data.

Similarly, if the sample mean and sample standard deviation are markedly unequal, an exponential density (for which these two quantities are always equal) should be avoided. Likewise, an exponential density should be avoided if the sample mode is well-removed from the sample minimum. A uniform or beta density should be avoided if no upper limit to durations is apparent, because these densities are non-zero over finite ranges.

4.2 Sensitivity Analysis

Sensitivity analysis is a method of assessing how much or how little the observable behavior of the system being modeled varies as its intrinsic properties vary. In the context of studying downtime, sensitivity analysis examines the extent of change in performance metrics such as throughput, downstream utilization, and queue-length maxima in response to changes in downtime properties such as percentage, duration, and variability of duration. For example, of two candidate machines potentially installed at a critical point of a system, the machine with smaller variance of downtime duration may greatly improve system performance even when percent downtime and average duration of downtime are equal for the two machines.

As indicated above, these sensitivity analyses are valuable when no downtime data are available. Comparing system performance under best-case and worst-case scenarios assesses the criticality of downtime performance at a specific point within the system. The greater this criticality, the greater the attention that should be devoted to increasing the accuracy of downtime estimation at that point.

Additionally, sensitivity analyses, in keeping with the "what-if" gaming abilities of simulation, provide accurate assessment of the return on various investments proposed for downtime-performance improvement. Such proposed investments might include capital expenditure for equipment with shorter downtime durations, less variable downtime durations, or longer uptime durations. Competing proposals might involve increasing payroll costs to accommodate hiring additional and/or more highly trained repair crews to improve downtime performance, or increasing outsourcing costs for contracting work externally to the system during its downtime intervals.

4.3 Modeling System Behavior During Downtime Intervals

Modern model-building tools and languages allow the modeler a variety of options for modeling system behavior during downtime. To use these software capabilities effectively in the building of a valid, credible model, the modeler must ask the system experts questions such as these:

- Can an interval of downtime for a given machine begin at any time, or only when that machine is busy (in contrast to blocked or starved)?
- When a downtime interval begins during machine-busy time, can the machine finish the workpiece currently occupying it?
- If the answer to the immediately preceding question is "no," as it usually is in practice, does the workpiece become scrap immediately, await the end of the downtime interval, or get routed to backup processing?
- If the answer to the immediately preceding question is either of the last two alternatives, does the intervention of the downtime leave the remaining processing time required by the interrupted workpiece unchanged, or increase that requirement?
- When workpieces approach a downed machine from upstream, do they accumulate behind it or get routed elsewhere? The answer may be a composite of these possibilities; for example, after a certain amount of backup has gathered, additional arrivals may be sent to a subcontractor.
- Can separately specified downtimes attributable to different causes overlap? For example, a machine may be undergoing a downtime based on cycles (e.g., change of drilling bit) at the time a downtime based on elapsed time is scheduled to begin (e.g., recharge batteries). The modeler must check whether these downtimes should run consecutively or concurrently.

After the above questions have been asked and answered, the modeler must, while building a model in the simulation language or package of choice, study its documentation thoroughly to assure an accurate match between the workflow of the system and the corresponding workflow in the model. Achieving accuracy of this match represents the task of model verification -- checking that the model's behavior on the computer matches the modeler's expectations.

5 SUMMARY AND OUTLOOK

During the rapid rise in simulation modeling usage at Ford Motor Company during the past few years, production and process engineers have become increasingly aware that valid downtime modeling is an essential ingredient of valid, credible models. Each of the following is in turn an essential ingredient of valid downtime modeling:

- avoidance of oversimplifying assumptions
- careful attention to downtime data collection
- accurate probabilistic characterizations of empirical data sets
- correct usage of simulation software in modeling process logic in the face of downtime.

Planned developments include implementation of automated downtime data collection, increased archival and sharing of downtime data among corporate components, and development of spreadsheet macros to smooth the interface between data collection and simulation software.

ACKNOWLEDGMENTS

Drs. Hwa-Sung Na and Sanaa Taraman of Ford Alpha, Ken Lemanski of Ford Product and Manufacturing Systems, and Dr. Onur Ulgen, president of Production Modeling Corporation and professor of Industrial and Manufacturing Engineering at the University of Michigan - Dearborn, all made valuable criticisms toward improving the clarity of this paper.

REFERENCES

- Grajo, Eric S. 1992. An Analysis of Test-Repair Loops in Modern Assembly Lines. In *Industrial Engineering*, 54-55. Production Modeling Corporation, Dearborn, Michigan.
- Law, A. M. and W. D. Kelton. 1991. *Simulation Modeling and Analysis*, 2d ed. New York: McGraw-Hill.
- Vincent, Stephen G. and Averill M. Law. 1993. UniFit II: Total Support for Simulation Input Modeling. In *Proceedings of the 1993 Winter Simulation Conference*, ed. Gerald W. Evans, Mansoor Mollaghasemi, Edward C. Russell, and William E. Biles, 199-204. Averill M. Law & Associates, Tucson, Arizona.

AUTHOR BIOGRAPHY

EDWARD J. WILLIAMS holds bachelor's and master's degrees in mathematics (Michigan State University, 1967; University of Wisconsin, 1968). From 1969 to 1971, he did statistical programming and analysis of biomedical data at Walter Reed Army Hospital, Washington, D. C. He joined Ford in 1972, where he works as a computer analyst supporting statistical and simulation software. Since 1980, he has taught evening classes at the University of Michigan, including both undergraduate and graduate simulation classes using GPSS/H, SLAM II, or SIMAN.