Proceedings of the 2007 Winter Simulation Conference
S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, eds.

CLINIC: CORRELATED INPUTS IN AN AUTOMOTIVE PAINT SHOP FIRE RISK SIMULATION

Debra Elkins

Operations Research Group
30500 Mound Road
Mailcode 480-106-155
General Motors R&D Center
Warren, MI 48090, U.S.A.

A. Christine LaFleur

Corporate Risk Management
Adam Opel Haus
IPC D4-03
Friedrich-Lutzmann-Ring 1
Russelsheim, D-65423
GERMANY

Earnest Foster
Jeffrey Tew

Mfg Systems Research Lab
30500 Mound Road
Mailcode 480-106-359
General Motors R&D Center
Warren, MI 48090, U.S.A.

Bahar Biller
(Expert Commentator)

Tepper School of Business
360 Posner Hall, Carnegie Mellon University
Pittsburgh, PA 15213, U.S.A.

James R. Wilson
(Expert Commentator)

Department of Industrial and Systems Engineering
North Carolina State University
Raleigh, N.C. 27695, U.S.A.

ABSTRACT

General Motors (GM) has developed a first proof-of-concept simulation model to explore impacts of various fire events in automotive paint shop operations. The approach uses a chronological event tree structure to assess effectiveness of various fire protection options to reduce the potential for significant property damage and loss of production capability. For confidentiality purposes, GM has disguised the numerical data presented in this case study. GM is seeking advice from the simulation community on modeling questions related to input distribution modeling, and correlation structure among input random variables.

1 INTRODUCTION

Fire hazards are inherent to many automotive manufacturing processes such as atomized electrostatic vehicle painting, robotic welding, and metal cutting and grinding. Even small fires can significantly disrupt production by causing non-thermal (e.g., acidic smoke residue) damage to sensitive and critical computer controlled manufacturing equipment and processes. Adoption of lean manufacturing practices has minimized the inventory cushion and excess capacity available to buffer against such disruptions. Further, insurance does not eliminate this risk, primarily due to the high deductibles that many automotive manufacturers and tier-1 automotive suppliers typically carry.

GM business leaders recognize that fires are a significant risk to global operations. GM Manufacturing, Facilities and Risk Financing executives challenged the modeling team to find or develop tools and methods that would help GM better quantify and manage fire risks across its global network of manufacturing operations.

LaFleur (2007) conducted an extensive literature review on quantitative fire risk analysis methods, and found that while there is a considerable amount of work in the open literature, there are no models that directly apply to heavy manufacturing that (1) permit enterprise “roll-up” of risk measurement from the within plant manufacturing process level and (2) capture multi-facility / supply chain network interdependencies in measurement of impact. Actuarial models and financial risk portfolio approaches do not permit “drill down” from enterprise level to manufacturing process level risk measurements. Thus, the GM team chose to pursue a simulation modeling approach to assess annual risk of fires occurring in paint shop operations. Note: the GM team defines risk as a probability distribution on total annual losses (property damage + equivalent value of lost production) caused by fire events disrupting paint shop operations.

2 AUTOMOTIVE INDUSTRY FIRE DATA

The GM team is fortunate to have a wealth of internal fire data available for simulation modeling and analysis. Each fire incident is reported in a global loss reporting database developed to comply with the Sarbanes-Oxley Act of 2002. For each fire incident, the following data is collected: date and time of the fire, fire origin mapped down to the manufacturing process level, cause of ignition if known, fuels or materials involved, fire suppression modes used, amount of production downtime (if any), and estimated cost of property damage including equipment repairs or replacement costs. Using the last four years of empirical data, GM is able to assess frequency of fires for various manufacturing processes, and severity of fires based on property damage and production downtime. For paint shop operations specifically, GM has more than 50 complete fire incidents recorded. GM also has historical data of large paint shop fire insurance claims, but this data may not be totally relevant, given the major adoption of computer aided manufacturing and more complex manufacturing equipment and control system from the mid 1980's onward.

In addition, GM uses a qualitative Fire Risk Index to characterize potential for a fire to become a severe or catastrophic fire. The Fire Risk Index is a relative risk scoring method to assess manufacturing plants on six hazard categories related to fire severity including currently applicable fire protection: building and roof construction materials, supervision (fire alarms, security staff, 24-hour operations, etc.), automatic sprinklers, available water supply, exposure (presence of hazards immediately adjacent to the plant such as other manufacturing facilities, warehouses, and bulk flammable liquid storage tanks), and other special hazards (e.g., electrostatic painting, welding, or heat treatment operations).

GM also has blueprint plans for construction of new paint shops in emerging, low cost markets. Local construction codes and regulations for fire protection equipment often differ from more restrictive U.S. building codes and standards. In particular, some countries do not require building level automatic sprinklers for paint shops, as long as process level (within paint shop booths) automatic sprinklers are used. GM leadership is interested in better understanding how complying with different fire protection codes might impact annual fire risk. This modeling approach will hopefully permit the team to respond to such policy questions by characterizing risk and comparing options based on probability distributions.

3 EVENT TREE MODEL STRUCTURE

The GM team worked with GM subject matter experts (98+ years of combined industrial fire protection engineering experience) to develop a time-based event tree that defines paint shop fire scenarios of interest and captures process level details of fire protection equipment, similar to the methodology described by Barry (2002). Figure 1 shows an example event tree that enumerates ten paint shop fire scenarios of interest.

Each fork in the event tree is determined by whether or not the specified fire protection option succeeds or fails. The probability of success or failure of each fire protection option is determined from equipment specifications, fault tree reliability estimates, or expert opinion. The event tree structure yields a discrete distribution for likelihood of occurrence of each outcome scenario.

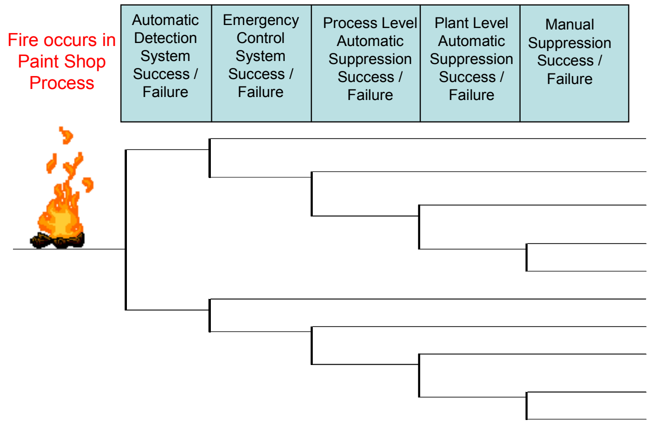

Figure 1: Scenario enumeration including process level fire protection details in a time-based event tree.

Figure 2 shows the summary results that are obtained from the event tree model shown in Figure 1. Each of the ten scenarios has a likelihood of occurring (calculated by multiplying probabilities within the event tree), and a short summary description of how the subject experts characterize such a fire event including the resulting property damage and lost production (in terms of downtime), based on success or failure of the fire protection options deployed.

Example Scenario Likelihood (calculated from event tree branch)	Scenario Outcome Description	Characterization of Property Damage and Lost Production (e.g. Downtime)
0.8	Best case: auto detection successful; emergency control system extinguishes fire	Best case; minimum property damage; minimum lost production for system check and restart
...	...	...
0.00001	Worst case: all suppression efforts fail to control fire below damage threshold	Maximum Foreseeable Loss: extreme property damage and downtime of at least 90 days due to total loss of building

Figure 2: Example event tree scenario outcomes.

4 APPROXIMATING TOTAL ANNUAL PAINT SHOP FIRE EVENT LOSSES USING MONTE CARLO SIMULATION

The modeling team used a collective risk model to obtain a simple first approximation for total annual losses in a paint shop caused by fire events. Total annual loss L_i for year i is defined by

$$L_i = \sum_{j=1}^{N_i} (X_j + Y_j), \quad (1)$$

Here N_i is the number of paint shop fire events that occur each year i, X_j is the cost of property damage per fire, and Y_j is the financial value of lost production (equivalent financial value of the lost downtime D_j) caused by the fire event. The equivalent value of lost production is given by

$$Y_j = D_j * \text{JPH} * \text{Average Net Profit per Vehicle}. \quad (2)$$

JPH is the average jobs per hour for the paint process, i.e., the process throughput rate. This calculation gives a deliberately simple estimate of the lost opportunity to use manufacturing capacity, labor, and other resources to add value by painting a vehicle. For more information on collective risk models, the interested reader should consult Embrechts et al. (1997), Klugman et al. (1998), Rolski et al. (1999), or Mikosch (2004).

The modeling team used the empirical fire incident data and Goodness of Fit testing to check that it is reasonable to assume that fires occur in paint shop operations according to a Poisson process (i.e., N_i is a Poisson random variable). Assuming that distributions for property damage X_j and value of lost production Y_j can be determined (to be discussed in a later section), it is straightforward to simulate total annual loss L_i for each year i.

Simple Monte Carlo Simulation Algorithm for Total Annual Loss from Paint Shop Fire Events:

1. Determine the number of years i to run the simulation (100 year, 500 year, 10,000 year, etc.).
2. For each year i, sample N_i ~ Poisson random variable for number of paint shop fires that occur in year i.
3. For each fire j = 1, 2… N_i,
    a. sample from the discrete event tree derived distribution to determine which of ten fire scenarios occur.
    b. sample from the property damage distribution to obtain a value of property damage X_j.
    c. sample from the downtime distribution to obtain a downtime value D_j. Convert the downtime value to an equivalent value of Y_j, the financial value of lost production. See equation (2).
4. For each year i, calculate the total annual loss L_i as given by (1) above.
5. Repeat steps 2-4 for i = 1 to number of years determined in step 1. (e.g., i = 1 to 10,000).
6. Given the simulated values of L_i, plot the empirical distribution $P\{L_i \leq x\}$ for total annual losses caused by fire events in a paint shop, or the related empirical probability density function.

5 PROPERTY DAMAGE AND DOWNTIME DISTRIBUTION MODELING

Since there were only 50+ data points (complete detailed paint shop fire incident reports), the team analyzed the data as a single set, obtained a single property damage distribution, then determined a method to relate the event tree scenario determined in Algorithm Step 3a to the value sampled from the property damage distribution in Algorithm Step 3b. The remainder of this section describes this process. A similar approach is applied to obtain and sample from the downtime distribution.

Note that no extreme or catastrophic fires were observed in the data; thus the modeling team decided to treat modeling of the right tail of the distribution separate from the "main body" of the distribution. For the "main body" of the distribution, the team selected the lognormal distribution as a reasonable fit of the property damage data, based on Goodness of Fit testing. Figure 3 shows an example fit of a property damage distribution and emphasizes the "main body" and right tail of the distribution. As a side note, a lognormal distribution is often used in practice to model property damage loss amounts in property-casualty insurance.

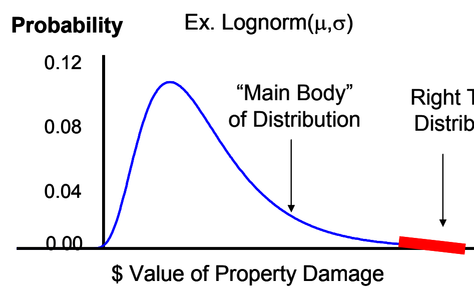

Figure 3: Example property damage distribution fit treating the "main body" separate from the right tail.

Since the team had no quantitative fire incident data to model the right tail of the distribution, the team leveraged subject matter domain knowledge about total cost and time to recover or replace a paint shop given a catastrophic fire (e.g., Maximum Foreseeable Loss (MFL) Scenario). The team modeled the right tail using a triangular distribution based on input parameters from subject matter experts. Future modeling work could include evaluating impact of heavier tail distributions such as Extreme Value or Pareto distributions. The team including subject matter experts defined minimum, maximum and most likely values of property damage, assuming that the catastrophic MFL scenario occurs. The subject experts used the Fire Risk Index and blueprint plans described in Section 2 to calibrate estimations of worst case scenarios. The modeling team also collected some publicly available benchmark data on other automotive paint shop fires to help experts with estimating property damage values and downtimes (Industrial Fire World Magazine, 2006).

Once the two distribution pieces are obtained, they are "patched" together to cover the domain of the property damage random variable (see Figure 4 below). Referring back to Figure 2, where the discrete probability distribution on scenarios is obtained for the example event tree, note that the worst case maximum foreseeable loss (MFL) scenario has a probability of 0.00001. Observe that the total probability estimate for a MFL scenario is approximately 0.00002 (since there are 2 MFL scenarios, 1 in the top half of the event tree when the detection system works, and the other in the bottom half of the event tree when the detection system fails). Using this probability value, we rescale the area under the right tail in the triangular distribution piece to be 0.00002, and the area under the "main body" distribution to be 0.99998.

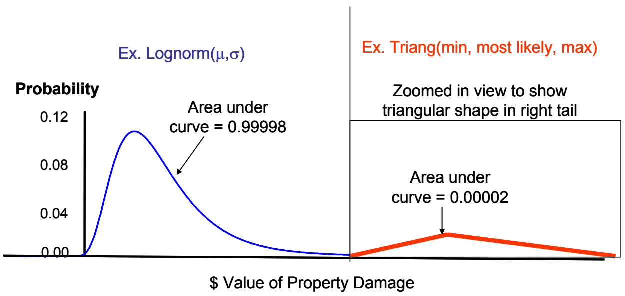

Figure 4: Patching together distribution pieces to cover the domain of the property damage random variable.

At this point, the team has obtained a single property damage distribution (as a special case of a mixture distribution), but needed a way to make sure that based on the scenario outcome that occurs, the correct domain of the random variable is sampled from. The modeling team and experts agreed that a really simple way would be to truncate the domain for the different scenarios (as shown in Figure 5). Thus, given that a specific scenario occurs, the associated random variable is sampled from the appropriate portion of the distribution, by constraining the random variable (using acceptance – rejection) to have a value falling within the lower and upper limits of the section of the domain assigned to the scenario.

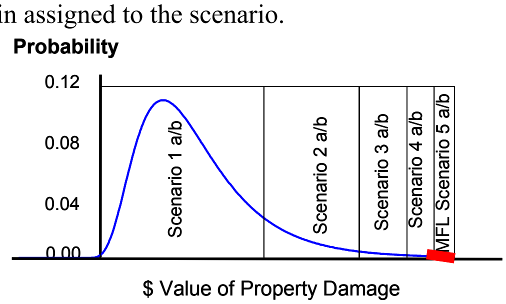

Figure 5: Assigning scenario domain limits within the fitted property damage distribution.

Each related pair of scenario outcomes in the top half or the bottom half of the event tree is assumed to have the same distribution and domain bounds on potential property damage. However, the event tree itself does have slightly different probabilities on the top and bottom half of the event tree, which do change the scenario likelihood of occurrence probabilities slightly. Our reason for doing this is to keep the model inputs simple. The automated detection system is defined as infrared camera detection, or visual human detection of flame within 2 minutes. If either detector (camera or human) is successful, the emergency control system can be deployed. The emergency control system includes automatic shutoffs of power, conveyor, paint and thinner pumps, and automatic paint system equipment interlocks. Note that there is a very short time window of 2 minutes before process level automatic fire suppression is engaged. Due to this short time window, we felt that there would only be minor differences (that could be ignored in a first rough model) in the probability distributions for property damage and downtime, depending on success or failure of each suppression option deployed.

6 QUESTIONS FOR THE SIMULATION COMMUNITY

Question 1: Can the expert community suggest any other methods to relate a specific event scenario outcome to property damage and downtime distribution sampling? Is there a better way to break up or divide the domain in the property damage distribution (and similarly for the downtime domain) rather than the approach used in Figure 5? In particular, the modeling team would like the expert reviewers to comment on advisability of using a similar approach to defining sampling distributions, but allowing the scenario domain bounds to perhaps overlap. The modeling team thinks that the hard bounds on the random variable domains are unnecessarily causing multiple peaks to appear in the example output density function plots shown in Figures 6 and 7.

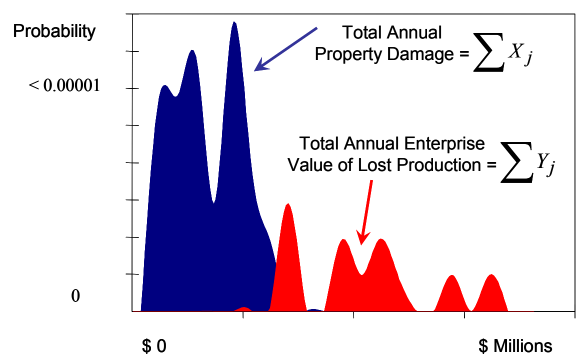

Figure 6: Empirical probability density overlay graph comparing total annual costs of paint shop fires.

As expected, lost production is less probable, but can dwarf property damage in value if it occurs.

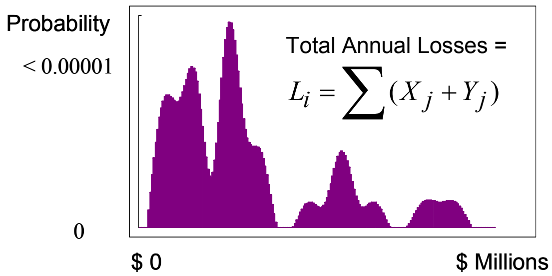

Figure 7: Empirical probability density function for total annual losses caused by fires in a paint shop

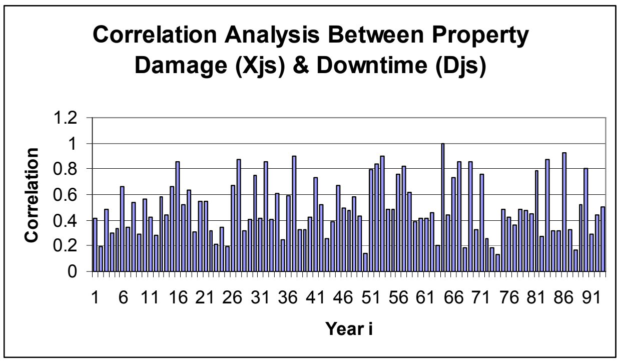

Figure 8: Plot of positive random correlation between property damage and downtimes for each year *i.*

Question 2: Is there a better way to introduce correlation between the two key random variables of property damage and downtime (and related value of lost production)? Is there a methodology such that we can specify (and control) the amount of correlation as an input parameter to the simulation?

GM subject experts commented that for paint shop fires, if there is large property damage, there will often be long associated downtimes. The modeling team did explore the empirical data, and found no correlation between property damage and downtimes for the reported paint shop fires (many of which fall in the best case minimum property damage, short downtime scenarios). Note that in other manufacturing processes that the team intends to model for next steps, it is possible to have large property damage and minimal downtime, and conversely, minimal property damage but long downtime.

Response to Comments from one of the Simulation Expert Reviewers: We agree with one of the reviewers that by sampling the property damage and downtime distributions with domain bounds defined by scenario occurrence, we have included positive correlation between these two random variables. We believe for paint shop fires, low property damage loss values should be associated with short downtimes (and thus low values of lost production), and conversely high property damage loss values will be associated with long downtimes (and high values of lost production). Figure 8 shows that there is clearly positive but random correlation for each year *i*, between the sampled property damage values (X_j's) and the sampled downtime values (D_j's). Based on these considerations, we ask Question 3, as follows:

Question 3: Is there a better way to represent the variables and their correlation? Is there an easier way to implement this procedure and understand the annual loss exposure from property damage vs. enterprise value of lost production?

7 COMMENTS BY BAHAR BILLER

First I comment on the derivation of the probability density function in Figure 4 and then suggest an alternative representation via the use of a flexible family of distributions for event tree scenario outcomes in Figure 2. Then I talk about incorporating correlation between two different random variables, i.e., property damage and downtime, with arbitrary marginal distributions.

The team chooses marginal distributions for property damage and downtime using historical data drawn from 50+ detailed incident reports. Using the stochastic input-modeling tools, the team identifies the lognormal distribution as a good fit for the stochastic process underlying the property damage (Figure 4). Due to the lack of any catastrophic events in the historical data set, the team collects expert opinion on the value of the property damage in the form of minimum, most likely, and maximum values of the random variable of interest and incorporates the collected expert opinion into the right-tail of the fitted lognormal distribution in the form of a triangular distribution to represent the maximum foreseeable loss (Figure 3 and Figure 4).

A close look at the existing stochastic input-modeling literature shows that expert opinion on the value of the maximum foreseeable loss can be obtained using the following two methods: (a) mean and variability method and (b) breakpoints method. The former method requires the specification of a lower bound, an upper bound, a value for mean, and a value for variability. Thus, it fails to translate any expert opinion into asymmetric distributional shapes with arbitrary tail characteristics. On the other hand, the breakpoints method is particularly useful for modeling quantities with a large number of possible outcomes, enabling the use of as many breakpoints as can be confidently obtained especially near the extremes. What the GM team describes in Figure 4 is a special application of the latter method where the break points correspond to the minimum, most likely, and maximum values of the random variable of interest. The triangular distribution is clearly a natural selection. An alternative selection for the underlying probability distribution is the pert (beta) distribution that is constructed with the information set identical to that of the triangular distribution and is typically used to model the activity times in project management problems (see Figure 9 for the example plots of triangular and pert distributions).

The reason I recommend the pert distribution as alternative to the triangular distribution is its ability to provide smoother representation for the extreme tails.

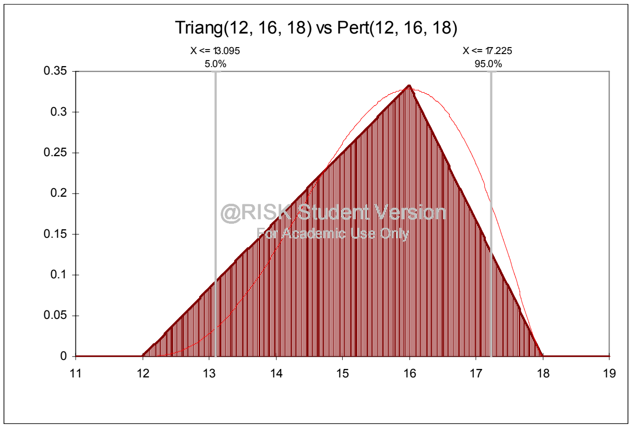

Figure 9: Triangular and pert distributions

Now I switch my attention to Figure 5, in which the team assigns scenario domain limits within the fitted property damage distribution using the scenario likelihoods in Figure 2 and samples values for the property damage from the portion of the probability distribution associated with the specific scenario. Since the team samples a value for the property damage and a value for the downtime depending on the scenario of interest, my recommendation is that the team designs its experiments in a way that a specific marginal distribution is chosen for each scenario from a flexible family of distributions. In other words, an alternative to patching together distribution pieces to cover the domain of the property damage random variable might be the scenario-dependent use of a flexible family of distributions with the ability of capturing a wide variety of distributional shapes. One such flexible family, which has been used in a variety of simulation applications, is the Johnson translation system (Johnson 1949a; Johnson, Kotz, and Balakrishnan 1994). The team might especially find the use of the Johnson bounded distribution very beneficial as it would provide the team great deal of flexibility in defining scenario-dependent distributional properties. More specifically, the Johnson bounded distribution is a four-parameter distribution whose two (location and scale) parameters define the lower bound and the upper bound for the values of the random variable of interest, i.e., level of variability, and whose two shape parameters define the level of skewness in the shape of the marginal distribution. Thus, by using the same functional form, but with different parameters, for each scenario of interest, the team can study the impact of scenario-dependent variability and scenario-dependent skewness on the output performance measures. Notice that the scenario domains limits within the fitted property damage distribution as illustrated in Figure 5 makes it difficult for the team to change the distributional properties of different scenarios independent of each other, further complicating the execution of a comprehensive sensitivity analysis. This difficulty is overcome via the use of a flexible distribution for each scenario of interest.

The identification of a marginal distribution from the Johnson translation system for the scenarios characterized in Figure 2 would be a straightforward task if there existed data sets large enough to utilize the stochastic input-modeling tools to the fullest extent. A good reference for fitting Johnson marginal distributions to independent data is Swain et al. (1988). Recall that the historical data of the GM team is composed of 50+ detailed incident reports. Therefore, the construction of the input models for the scenario outcomes characterized in Figure 2 requires the wide use of the opinion of subject matter experts. It might also be possible to start the construction of the scenario-dependent Johnson marginal distributions with the probability density function provided in Figure 5. By controlling the level of variability and the level of weight in the tails separately, the scenario-dependent portions of the fitted property damage distribution can be matched as closely as possible by a series of distributions from the Johnson bounded family. This might additionally allow the GM team to have overlapping scenario domain bounds.

Finally, I recommend the use of the Normal-To-Anything (NORTA) process designed specifically for random vectors (Cario and Nelson 1997) to represent scenario-dependent correlations between property damage and downtime. More specifically, the values of property damage and downtime of any specific scenario can be sampled with a correlation prespecified by the team by simply transforming a multivariate normal base random vector into the desired process via the use of the inverse cumulative distribution function. The team can find the implementation details of this process in the survey paper on multivariate input processes (Biller and Ghosh 2006) as well as background on the extension of the univariate Johnson bounded distribution to the two-dimensional Johnson distribution in Johnson (1949b), Johnson, Kotz, and Balakrishnan (1994), and Biller and Nelson (2003).

8 COMMENTS BY JAMES R. WILSON

The GM authors pose some very interesting questions about how to perform simulation input modeling when relatively limited data must be supplemented by expert knowledge of the application domain so as to build an adequate probabilistic model of the process by which an automotive paint shop fire results in a pair of performance measures $[X, D]$, where X denotes the dollar value of the associated property damage and D denotes the corresponding lost downtime. There are five different scenarios for which we must model the joint distribution of $[X, D]$ and specify an algorithm for generating random samples from that distribution. In each scenario, the authors also seek a method for systematically investigating the effects of varying levels of stochastic dependence between X and D, ranging from the case in which X and D are independent to the case in which X and D exhibit a strong degree of dependence.

In this reply to Elkins, LaFleur, Foster, and Tew, I will propose a hybrid approach to input modeling in which both data and expert knowledge are used to build a rough empirical model of the random vector $[X, D]$ that can be readily incorporated into the authors' simulation model of automotive paint shop fires.

8.1 Reply to Question 1

The authors have records on about 50 fires in GM's paint shops, and they use this data set to fit lognormal probability density functions (p.d.f.'s) to the marginal distributions of X and D separately. For simplicity we focus first on the marginal p.d.f. of X; a similar discussion will also apply to the marginal p.d.f. of D.

The authors state that scenario 5 is a catastrophic fire or maximum foreseeable loss (MFL); and the probability is 0.00002 that a given fire will evolve according to scenario 5. For the other four scenarios, the authors partition the interval $[0, x_{0.99998})$ (that is, the range of possible values of property damage from \$0 up to but not including the MFL, or the 0.99998 quantile of X) into subintervals corresponding to scenarios 1, 2, 3, and 4, respectively. For the conditional marginal p.d.f. of X given that scenario 5 has occurred, the authors fitted a triangular distribution to expert estimates (knowledge or opinion) of the minimum, maximum, and most likely (modal) values of property damage when a catastrophic fire occurs. For scenario k (where $k = 1, \ldots, K$, and K denotes the total number of potentially different distributions that may be required for different scenarios), let $[a_k, b_k)$ denote the range of possible values for X when scenario k occurs. Currently the authors are assuming that $K = 5$, and that

$$a_1 = 0 < b_1 = a_2 < b_2 = a_3 < \cdots < b_4 = a_5 < b_5 < \infty; \quad (3)$$

and the authors ask first if there is a better way to divide the space of possible values of X to allow, for example, some overlap of the subintervals corresponding to different scenarios.

As the starting point for my response to the authors' first question, I want to reexamine the authors' method for handling scenario 5. If m_5 denotes the modal dollar value (most likely value) of the property damage due to a catastrophic fire, then the authors propose using a triangular distribution with minimum a_5, mode m_5, and maximum b_5 as their model for the conditional distribution of X given a catastrophic fire. For the reasons detailed in the input-modeling tutorial for this conference (Kuhl et al. 2007), I suggest that an attractive alternative input model in this situation is the generalized beta distribution with minimum a_5, maximum b_5, and shape parameters

$$\theta_1 = \frac{\omega^2 + 3\omega + 4}{\omega^2 + 1} \text{ and } \theta_2 = \frac{4\omega^2 + 3\omega + 1}{\omega^2 + 1}, \quad (4)$$

where

$$\omega = \frac{b_5 - m_5}{m_5 - a_5}. \quad (5)$$

If C denotes the scenario number, which is also a random variable on each replication of the authors' simulation, then I propose the following generalized beta conditional p.d.f. for X given $C = 5$:

$$f_{X|C}(x \mid 5) = \frac{(x - a_5)^{\theta_1 - 1}(b_5 - x)^{\theta_2 - 1}}{B(\theta_1, \theta_2)(b_5 - a_5)^{\theta_1 + \theta_2 - 1}} \text{ for } a_5 \le x \le b_5, \quad (6)$$

where

$$B(\theta_1, \theta_2) = \int_0^1 z^{\theta_1 - 1}(1 - z)^{\theta_2 - 1}\, dz \text{ for } \theta_1 > 0, \theta_2 > 0 \quad (7)$$

is the beta function.

For the other four scenarios, I suggest an approach similar to Equations (4)–(7), possibly adapted to take into account other sample information or expert knowledge. Suppose that in scenario k, the fire expert selects (i) a lower limit a_k for X which may overlap with scenario $k-1$ in the sense that we may have $a_k < b_{k-1}$; and (ii) an upper limit b_k for X which may overlap with scenario $k+1$ in the sense that we may have $a_{k+1} < b_k$. Such overlapping is impossible with the authors' partition (3) of the space of possible values of a single lognormal distribution; but if we model the conditional p.d.f. $f_{X|C}(x \mid k)$ as a generalized beta p.d.f. on the interval $[a_k, b_k]$, then the only remaining issue is how to specify the shape parameters θ_1 and θ_2.

If the mode m_k is specified by the expert, then Equations (4) and (5) may be used. If we have available the sample mean $\bar{X}$ and sample variance S^2 for a reasonable number of observations of X in scenario $C = k$, then a moment-matching approach to modeling the conditional p.d.f. of X given $C = k$ is to take

$$\theta_1 = \frac{\omega_1^2(1 - \omega_1)}{\omega_2^2} - \omega_1 \text{ and } \theta_2 = \frac{\omega_1(1 - \omega_1)^2}{\omega_2^2} - (1 - \omega_1) \quad (8)$$

in Equation (4), where

$$\omega_1 = \frac{(\bar{X} - a_k)}{(b_k - a_k)} \quad \text{and} \quad \omega_2 = \frac{S}{(b_k - a_k)}. \tag{9}$$

This approach would give the authors the flexibility to allow overlapping subintervals defining the space of X for different scenarios while also enabling the authors to exploit the full flexibility of the generalized beta distribution to model the main features of the conditional p.d.f. $f_{X|C}(x \mid k)$ for $k = 1, \ldots, K$.

This approach can be adapted to other combinations of sample information and expert knowledge. For example, if the mode m_k and the sample mean $\bar{X}$ are to be used to specify the p.d.f., then we can solve Equations (4) and (5) of Kuhl et al. (2007) simultaneously for the shape parameters θ_1 and θ_2. From this discussion, the advantages of using the generalized beta distribution for rapid input modeling of X for each scenario should be clear.

If greater flexibility is needed to represent adequately the shape of the conditional p.d.f. of X given $C = k$ for some scenario k, then I suggest the authors consider using the Johnson translation system of distributions or the Bézier distribution family; see Kuhl et al. (2007). All of the above comments also apply to the conditional marginal p.d.f. of D given $C = k$ for $k = 1, \ldots, K$. In the next subsection, I will address the problem of modeling the conditional joint distribution of the random vector $[X, D]$

As a final remark about the authors' Question 1, I should add that the multimodal character of the overall unconditional marginal distributions of X and D is mainly due to the mixture of the conditional marginal p.d.f.'s,

$$f_X(x) = \sum_{k=1}^{K} \Pr\{C = k\} \cdot f_{X|C}(x \mid k), \tag{10}$$

that results from summing over all possible scenarios. I believe that the multimodality of the unconditional marginal p.d.f.'s of X and D will be one of the main features of these p.d.f.'s no matter what models are used for the conditional marginal p.d.f.'s of X and D.

8.2 Reply to Question 2

The authors' second question concerns how to model stochastic dependence between the components of the random vector $[X, D]$. Although the authors state that they observed no such dependence in the scenarios involving minimal property damage or short downtimes, they seek to take account of the experts' assertion that in fires involving substantial property damage, the associated downtimes will also be substantial. Because the setup described in the previous subsection uses a separate input model for each scenario, we will be able to reconcile the authors' observations about the lack of dependence between X and D in light-damage scenarios with the expert's knowledge about the nonnegligible dependence between X and D that is present in heavy-damage scenarios.

To model the random vector $[X, D]$ with dependent components, I propose taking the opposite approach from that taken in Section 4.2 of Wilson (1997). We will start with a bivariate standard normal random vector $[Z_1, Z_2]$ whose components have correlation ρ; see Section 3 of Wilson (1997). Reversing the usual procedure in which we seek a normalizing translation (transformation) that yields $[Z_1, Z_2]$, in the current situation we seek an (inverse) transformation of $[Z_1, Z_2]$ that yields a bivariate beta random vector $[X, D]$,

$$[X, D] = \left[F_{X|C}^{-1}\left\{\Phi(Z_1)\middle|k\right\}, F_{D|C}^{-1}\left\{\Phi(Z_2)\middle|k\right\}\right], \tag{11}$$

where: (i) the generalized beta marginal cumulative distribution functions (c.d.f.'s) $F_{X|C}(x \mid k)$ and $F_{D|C}(d \mid c)$ have inverses that can be readily approximated using, for example, Equation 26.5.22 of Abramowitz and Stegun (1965); (ii) the standard normal c.d.f.,

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(-t^2/2\right) dt \text{ for all } z, \tag{12}$$

is readily approximated using, for example, Equation 26.2.19 of Abramowitz and Stegun (1965); and (iii) the bivariate standard normal random vector $[Z_1, Z_2]$ with correlation ρ has joint p.d.f.

$$\varphi_{Z_1, Z_2}(z_1, z_2; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{z_1^2 - 2\rho z_1 z_2 + z_2^2}{2(1-\rho^2)}\right) \tag{13}$$

for all z_1 and z_2.

The only thing remaining to complete our specification of the procedure for generating the random vector $[X, D]$ is to specify the method for generating the standard normal random vector $[Z_1, Z_2]$. Virtually every simulation language has a mechanism for generating independent standard normal random variables; and this is used to generate $Z_1 \sim N(0,1)$. Then by the method of conditional distributions, Z_2 is generated from a normal distribution with mean ρZ_1 and variance $1 - \rho^2$; and the resulting random vector $[Z_1, Z_2]$ is fed into Equation (11) to yield the random vector $[X, D]$.

The conditional correlation $\rho_{XD}(k)$ between the components of $[X,D]$ given scenario $C=k$ may be evaluated from

$$\operatorname{Cov}\left(X,D\middle|C=k\right)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}F_{X|C}^{-1}\left\{\Phi(z_1)\middle|k\right\}F_{D|C}^{-1}\left\{\Phi(z_2)\middle|k\right\}\cdot\varphi_{Z_1Z_2}\left(z_1,z_2;\rho\right)dz_1\,dz_2-\operatorname{E}\left[X\middle|C=k\right]\cdot\operatorname{E}\left[D\middle|C=k\right]\qquad(14)$$

and

$$\rho_{XD}(k)=\frac{\operatorname{Cov}\left(X,D\middle|C=k\right)}{\sqrt{\operatorname{Var}\left[X\mid C=k\right]\cdot\operatorname{Var}\left[D\mid C=k\right]}}.\qquad(15)$$

To evaluate systematically the effects of increasing the conditional correlation $\rho_{XD}(k)$ between the components of the random vector $[X,D]$ given scenario $C=k$, I suggest tabulating and plotting $\rho_{XD}(k)$ as a function of the parameter ρ for, say, $\rho=0,\ \pm0.05,\ \pm0.10,\ \ldots,\ \pm0.95$. This will reveal the range of possible correlations between X and D in scenario k that can be achieved with the modeling and simulation approach detailed in Equations (11)–(15). Moreover by linear interpolation in this plot, we can estimate the levels of ρ that are required to run scenario k with the following prespecified levels of correlation between X and D:

$$\rho_{XD}(k)=0,\pm0.05,\pm0.10,\ldots,\pm0.95,\qquad(16)$$

where it must be recognized that not all the values listed in Equation (16) may be achievable. By performing multiple replications of scenario k using the achievable levels of $\rho_{XD}(k)$ specified in Equation (16), the authors should be able to obtain a clear idea of the effect of correlation between X and D on their risk analyses for each scenario separately as well as for the overall unconditional joint distribution of X and D taken over all scenarios.

I hope that these remarks have been responsive to the authors' questions and that some of these suggestions will prove to be useful in the application at hand.

ACKNOWLEDGEMENTS

The GM authors would like to thank the following subject matter experts for their many valuable insights on this problem: Thomas P. Hunter and P. Kyle Weddle, GM Corporate Risk Management, J. William (Bill) Sheppard, Wallie D. Williams, Michael D. Throop, GM Global Security Fire Protection, and Dr. Arvind Atreya, University of Michigan, Department of Mechanical Engineering

REFERENCES

Abramowitz, M., and I. A. Stegun. 1965. *Handbook of mathematical functions.* New York: Dover Publications.

Barry, T.F. 2002. *Risk-informed, performance-based industrial fire protection: an alternative to prescriptive codes*. TFBarry Publications. Available via <www.fireriskforum.com> [accessed May 1, 2007].

Biller, B., and S. Ghosh. 2006. Multivariate input processes. In *Handbooks in Operations Research and Management Science: Simulation*, ed. B. L. Nelson and S. G. Henderson . Amsterdam: Elsevier Science.

Biller, B., and B. L. Nelson. 2003. Modeling and generating multivariate time-series input processes using a vector autoregressive technique. *ACM Transactions on Modeling and Computer Simulation* 13: 1–27.

Cario, M. C., and B. L. Nelson. 1997. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Working Paper, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL.

Embrechts, P., Kluppelberg, C., and T. Mikosch. 1997. *Modelling extremal events for insurance and finance*. Springer Verlag.

Industrial Fire World Magazine. Online Incident Logs Database. Available via <www.fireworld.com> [accessed January 2006].

Johnson, N. L. 1949a. Systems of frequency curves generated by methods of translation. *Biometrika* 36: 149–176.

Johnson, N. L. 1949b. Bivariate distributions based on simple translation systems. *Biometrika* 36: 297–304.

Johnson, N. L., S. Kotz, and N. Balakrishnan. 1994. *Continuous multivariate distributions, volumes I and II*. New York: John Wiley and Sons.

Klugman, S., Panjer, H., and G. Willmot. 1998. *Loss models: from data to decisions*. John Wiley & Sons.

Kuhl, M. E., E. K. Lada, N. M. Steiger, M. A. Wagner, and J. R. Wilson. 2007. Introduction to modeling and generating probabilistic input processes for simulation. In *Proceedings of the 2007 Winter Simulation Conference*, ed. S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, to appear. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers. Available online via <www.ise.ncsu.edu/jwilson/files/kuhl07wsc.pdf> [accessed September 9, 2007].

LaFleur, Chris. April 2007. Enterprise-level industrial fire risk modeling and analysis for automobile manufacturing facilities. D. Eng. Thesis, University of Michigan, Ann Arbor, Michigan.

Mikosch, T. 2004. *Non-life insurance mathematics: an introduction with stochastic processes*. Springer.

Rolski, T., Schmidli, H., Schmidt, V., and J. Teugels. 1999. *Stochastic processes for insurance and finance*. John Wiley & Sons.

Swain, J. J., S. Venkatraman, and J. R. Wilson. 1988. Least-squares estimation of distribution functions in Johnson's translation system. *Journal of Statistical Computation and Simulation* 29: 271–297.

Wilson, J. R. 1997. Modeling dependencies in simulation inputs. In *Proceedings of the 1997 Winter Simulation Conference*, ed. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson, 47–52. Available online via <www.informs-sim.org/wsc97papers/0047.PDF> [accessed September 9, 2007].

AUTHOR BIOGRAPHIES

DEBRA ELKINS, PH.D. is a Staff Researcher with expertise in Enterprise Risk Modeling at the General Motors R&D Center in Warren, Michigan. Her research interests include risk modeling for enterprise operations, manufacturing and supply chain vulnerability analysis and disruption consequence modeling, applied probability/stochastic processes, statistics, and enterprise scale simulation. Debra has served as an industry technical expert on enterprise risk for non financial services for the Department of Homeland Security, the National Academies, and the Internal Revenue Service. She is also serving as a North Carolina State Enterprise Risk Management Initiative 2006-2007 Industry Fellow. Most recently, Debra was appointed to serve a three year term on the National Academies Board on Mathematical Sciences and its Applications. Debra can be contacted at <debraelkins@wowway.com>

A. CHRISTINE LAFLEUR, D.ENG., P.E. is the Regional Risk Manager for Europe, Africa, and the Middle East, for General Motors Corporate Risk Management. Chris recently completed her D. Eng. in Manufacturing at the University of Michigan. She also holds a M.S. in Fire Protection Engineering from the University of Maryland, and a B.S. in GeoMechanical Engineering from the University of Rochester. Chris is a registered Professional Engineer in Virginia and Michigan She is a Member of National Fire Protection Association and the Society of Fire Protection Engineers (National and Metro Detroit Chapters), and a Principal Committee Member on NFPA-13, Sprinkler Discharge Criteria Committee. Her current research interests are performance based industrial fire protection and modeling and analysis to support enterprise fire risk management. Chris can be contacted at <Chris.Lafleur@de.gm.com>.

EARNEST FOSTER, PH.D is a Senior Researcher with General Motors Research and Development Center in Warren, Michigan. His interests include the General Motors Global Manufacturing System (GM-GMS), multivariate statistical process control, statistical process control for time variables, and process-oriented approaches to variation reduction. He also serves as the Membership Chair for General Motors Research and Development Center Chapter of the Sigma Xi Scientific Research Society. Earnest attended the University of Michigan, Ann Arbor and received the B.S. and M.S.E degrees in Industrial and Operations Engineering. He holds a Ph.D. in Industrial Engineering from the Pennsylvania State University. Earnest can be contacted at <Earnest.Foster@gm.com>.

JEFFREY TEW, PH.D. is a GM Technical Fellow and Group Manager of the Manufacturing Enterprise Modeling Group in the Manufacturing Systems Research Lab at General Motors' R&D Center in Warren, MI. Currently, Dr. Tew is an Adjunct Professor of Supply Chain Management at the Georgia Institute of Technology and a Visiting Professor of Industrial Engineering at Tsinghua University in Beijing. He was Coeditor of the Proceedings of the 1994 Winter Simulation Conference and is a past President of the INFORMS College on Simulation. He is the General Chair of the 2007 Winter Simulation Conference. He received a B.S. in mathematics from Purdue University in 1979, an M.S. in statistics from Purdue University in 1981, and a Ph.D. in industrial engineering from Purdue University in 1986. His current interests include the application of operations research and information technology tools to large-scale logistics (supply chain) systems and e-commerce. He is a member of Alpha Pi Mu, ACM, ASA, IIE, The Institute for Mathematical Statistics, INFORMS, SCS, and Sigma Xi. He can be reached via e-mail at <jefftew2002@yahoo.com>.

BAHAR BILLER is an assistant professor in the Tepper School of Business at Carnegie Mellon University. Her research interest lies in the area of computer simulation experiments for stochastic systems and more specifically, in the simulation methodology for dependent input processes with applications to financial markets, global supply chains, and telecommunication systems. Her web page can be found via <www.tepper.cmu.edu>.

JAMES R. WILSON is a professor in the Edward P. Fitts Department of Industrial and Systems Engineering at North Carolina State University. He is a member of AAUW, ACM, and ASA; and he is a Fellow of IIE and INFORMS. His e-mail address is <jwilson@ncsu.edu>, and his Web page is <www.ise.ncsu.edu/jwilson>.