INL/CON-07-13399
PREPRINT

Risk Analysis of the Space Shuttle: Pre-Challenger Bayesian Prediction of Failure

NASA Space Systems Engineering & Risk Management Symposium

Dana L. Kelly
Curtis L. Smith

February 2008

The INL is a
U.S. Department of Energy
National Laboratory
operated by
Battelle Energy Alliance

Idaho National Laboratory

Risk Analysis of the Space Shuttle: Pre-Challenger Bayesian Prediction of Failure

Dana L. Kelly
Curtis L. Smith

Idaho National Laboratory, P. O. Box 1625, Idaho Falls, ID 83415

Abstract

Dalal et al[1] performed a statistical analysis of field and nozzle O-ring data collected prior to the ill-fated launch of the Challenger in January 1986. The purpose of their analysis was to show how statistical analysis could be used to provide information to decision-makers prior to the launch, information that could have been expected to lead to a decision to abort the launch due to the low temperatures (~30^o F.) present at the launch pad on the morning of the scheduled launch. Dalal et al.[1] performed a frequentist analysis of the O-ring data, and found that a logistic regression model provided a relatively good fit to the past data. In the second portion of their paper, Dalal et al. propagated parameter uncertainties through the fitted logistic regression model in order to estimate the probability of shuttle failure due to O-ring failure at the estimated launch temperature of ~30^o F. Because their analysis was frequentist in nature, probability distributions representing epistemic uncertainty in the input parameters were not available, and the authors had to resort to an approximate approach based on bootstrap confidence intervals. In this paper, we will re-evaluate the analyses of Dalal et al. from a Bayesian perspective. Markov chain Monte Carlo (MCMC) sampling will be used to sample from the joint posterior distribution of the model parameters, and to sample from the posterior predictive distributions at the estimated launch temperature, a temperature that had not been observed in prior launches of the space shuttle. Uncertainties, which are represented by probability distributions in the Bayesian approach, are propagated through the model to obtain a probability distribution for O-ring failure, and subsequently for shuttle failure as a result of O-ring failure. No approximations are required in the Bayesian approach and the resulting distributions can be input to a decision analysis to obtain expected utility for the decision to launch.

When using a mathematical model, careful attention must be given to uncertainties in the model.
- Richard Feynman

Introduction

Dalal et al.[1] performed a statistical analysis of field and nozzle O-ring data collected prior to the ill-fated launch of the Challenger in January 1986. The purpose of their analysis was to show how statistical analysis could be used to provide information to decision-makers prior to the launch, information that could have been expected to lead to a decision to abort the launch due to the low temperatures (~30^o F.) present at the launch pad on the morning of the scheduled launch. Dalal et al.[1] performed a frequentist analysis of the O-ring data, and found that a logistic regression model provided a relatively good fit to the past data.

In the second portion of their paper, Dalal et al.[1] propagated parameter uncertainties through the fitted logistic regression model in order to estimate the probability of shuttle failure due to O-ring failure at the estimated launch temperature of ~30^o F. Because their analysis was frequentist in nature, probability distributions representing epistemic uncertainty in the input parameters were not available, and the authors had to resort to an approximate approach based on bootstrap confidence intervals, an approach developed by Efron.[2]

In this paper, we will re-evaluate the analyses of Dalal et al.[1] from a Bayesian perspective. Markov chain Monte Carlo (MCMC) sampling will be used to sample from the joint posterior distribution of the model parameters, and to sample from the posterior predictive distributions at the estimated launch temperature, a temperature that had not been observed in prior launches of the space shuttle. The open-source version of the WinBUGS software package[3] will be used to carry out the MCMC sampling. Uncertainties, which are represented by probability distributions in the Bayesian approach, are propagated through the model to obtain a probability distribution for O-ring failure, and subsequently for shuttle failure as a result of O-ring failure. No approximations are required in the Bayesian approach and the resulting distributions valuable inputs to a decision to launch. Also, this approach to analysis relates probability of failure directly to observables, such as temperature, providing a tangible model that engineers can use to make launch predictions.

Table 1 Space shuttle field O-ring thermal distress data

Flight	Distress[1]	Temp (^{o}F.)	Press (psig)
1	0	66	50
2	1	70	50
3	0	69	50
5	0	68	50
6	0	67	50
7	0	72	50
8	0	73	100
9	0	70	100
41-B	1	57	200
41-C	1	63	200
41-D	1	70	200
41-G	0	78	200
51-A	0	67	200
51-C	2	53	200
51-D	0	67	200
51-B	0	75	200
51-G	0	70	200
51-F	0	81	200
51-I	0	76	200

51-J	0	79	200
61-A	2	75	200
61-B	0	76	200
61-C	1	58	200

[1]Thermal distress is defined to be erosion of the O-ring or blow-by of hot gases. The table shows the number of distress events for each launch. There are six field O-rings on the shuttle, so the number of distress events is an integer in the interval [0, 6].

Stochastic Model for O-ring Distress

There are six O-rings on the shuttle, so during each launch, the number of distress events, defined as erosion or blow-by of a primary field O-ring, is modeled as binomial with parameters p and n = 6: X ~ binomial(p, 6).

In this model, p is a function of both temperature and applied leak-test pressure. The canonical link function is the logit function:

$$logit(p) = ln\left(\frac{p}{1-p}\right) \quad (1)$$

Following Dalal et al.[1] we consider two potential explanatory models:

1) logit(p) = a + b*temp + c*press
2) logit(p) = a + b*temp

The WinBUGS script for the first model, which includes both temperature and pressure as explanatory variables, is shown in Table 2. Diffuse normal priors were used for the coefficients in this model to allow the numerical results to be compared with the maximum likelihood estimates and confidence intervals obtained by Dalal et al.[1]

Table 2 WinBUGS script for logistic regression of primary O-ring distress on temperature and pressure

```
model {
for(i in 1:K)    {
        distress[i] ~ dbin(p[i], 6)
logit(p[i]) <- a + b*temp[i] + c*press[i] #Model with temperature and pressure
        }
distress.31 ~ dbin(p.31, 6) #Predicted number of distress events for launch 61-L
logit(p.31) <- a + b*31 + c*200
#Prior distributions – diffuse normal distributions about zero
a ~ dnorm(0, 0.000001)
b ~ dnorm(0, 0.000001)
c ~ dnorm(0, 0.000001)
}
```

```
inits
list(a=5, b=0, c=0) #Chain 1
list(a=1, b=-0.1, c=0.1) #Chain 2
```

Five thousand burn-in iterations were performed, followed by 100,000 iterations to estimate the parameters. The table below shows the posterior mean, standard deviation, and symmetric 95% interval for each of the parameters in the logistic regression model for p. The marginal posterior distributions for a and b are approximately normal with the listed posterior means and standard deviations.

Table 3 Summary posterior estimates of logistic regression parameters, temperature and pressure included as explanatory variables

Parameter	Mean	Standard Dev.	95% Interval
a (intercept)	2.24	3.74	(-4.71, 9.92)
b (temp. coeff.)	-0.105	0.05	(-0.20, -0.02)
c (press. coeff.)	0.01	0.009	(-0.004, 0.03)

The model that includes both temperature and pressure predicts about four distress events at 31^{o} F., the approximate temperature for the disastrous launch of the Challenger.

We next consider a simpler model, in which only temperature is included as an explanatory variable for the logistic regression. The WinBUGS script for this model is shown in Table 4.

Table 4 WinBUGS script for logistic regression of primary O-ring distress on temperature

```
model {
for(i in 1:K)    {
          distress[i] ~ dbin(p[i], 6)
          logit(p[i]) <- a + b*temp[i] #Model with temperature only
          }
distress.31 ~ dbin(p.31, 6)
logit(p.31) <- a + b*31
a ~ dflat() #Diffuse priors over real axis
b ~ dflat()
}

Inits
list(a=1, b=0.1) #Chain 1
list(a=10, b=-0.1) #Chain 2
```

One thousand burn-in iterations were required for convergence, followed by 100,000 iterations to estimate the parameters. The table below shows the posterior mean, standard deviation, and symmetric 95% interval for each of the parameters in the logistic regression model for p.

Table 5 Summary posterior estimates of logistic regression parameters, temperature included as explanatory variable

Parameter	Mean	Standard Dev.	95% Interval
a (intercept)	5.225	3.16	(-1.00, 11.48)
b (temp. coeff.)	-0.12	0.049	(-0.22, -0.025)

Dalal et al.[1] also examined a model that is quadratic in temperature. Specifically, they analyzed the following model for distress probability, where $\bar{t}$ is the average of the temperature readings in the data:

$$log\,it(\,p\,) = a + b(t - \bar{t}) + c(t - \bar{t})^2 \qquad (2)$$

The WinBUGS script for this model is shown in Table 6. We analyzed this model with 1,000 burn-in iterations, followed by 50,000 iterations for parameter estimation. The posterior distributions of the parameters are summarized in Table 7.

Table 6 WinBUGS script for logistic regression of primary O-ring distress quadratic in temperature

```
model {
for(i in 1:K)    {
  distress[i] ~ dbin(p[i], 6)
  logit(p[i]) <- a +b*(temp[i] - temp.mean) + c*pow(temp[i] - temp.mean, 2)
   }
temp.mean <- mean(temp[])
a ~ dnorm(0, 0.000001)
b ~ dnorm(0, 0.000001)
c~ dnorm(0, 0.000001)
}

Inits
list(a=-3, b=0.05, c=0.005)
list(a=0, b=-0.05, c=-0.005)
```

Table 7 Summary posterior estimates for logistic regression model for primary O-ring distress, quadratic model in temperature

Parameter	Mean	Standard Dev.	95% Interval
a (intercept)	-3.25	0.52	(-4.37, -2.32)
b (temp. coeff.)	-0.10	0.08	(-0.27, 0.03)
c (press. coeff.)	0.003	0.006	(-0.01, 0.01)

We use two measures to check model validity. The first uses a summary statistic based on the posterior predictive distribution, and is a measure of how well the model can replicate the observed data. We refer to this measure as a Bayesian p-value. This measure is described in more detail in Gelman et al.[4] We refer to this measure as a p-value even though, as pointed out by Bayarri and Berger,[5] this measure does not share the asymptotic properties of the frequentist p-value. The WinBUGS script excerpt in Table 8 shows how this is calculated. The second model-validation measure we will use is the deviance information criterion (DIC), a Bayesian analog of the Akaike information criterion (AIC) used by frequentists. For details see Spiegelhalter et al.[6] DIC is a measure of relative validity; the model with the lowest (best) DIC may still be a poor model from the standpoint of being able to replicate the observed data.

Table 8 Portion of WinBUGS script for calculating Bayesian p-value

```
model {
for(i in 1:K)    {
distress[i] ~ dbin(p[i], 6)
logit(p[i]) <- a + b*temp[i] + c*press[i] #Model with temperature and pressure
        distress.rep[i] ~ dbin(p[i], 6) #Replicate from posterior predictive
distribution
        diff.obs[i] <- pow(distress[i] - 6*p[i], 2)/(6*p[i]*(1-p[i]))
        diff.rep[i] <- pow(distress.rep[i] - 6*p[i], 2)/(6*p[i]*(1-p[i]))
        }
chisq.obs <- sum(diff.obs[]) #Observed summary statistic
chisq.rep <- sum(diff.rep[]) #Replicated summary statistic
p.value <- step(chisq.rep - chisq.obs) #Mean of this node should be near 0.5
}
```

The model with only temperature predicts essentially the same number of distress events as the two more complex models. The DIC is nearly the same for all the models; the simplest model with only temperature as an explanatory variable has a slightly larger Bayesian p-value than the model with both temperature and pressure, and is essentially the same as the model that is quadratic in temperature. Because the simplest model is essentially equivalent to the more complex ones, we would recommend it for predictive analyses.

Table 9 Model validation results for logistic regression models for primary O-ring distress

Explanatory Variables	DIC	Bayesian p-value
Temperature and pressure	36.58	0.19
Temperature	35.75	0.21
Quadratic in temperature	37.18	0.20

Probability of Shuttle Failure

The probability of shuttle failure is given by the joint probability of a) primary O-ring erosion, b) primary O-ring blowby, c) secondary O-ring erosion, and d) secondary O-ring failure. The above analysis examined the probability of a ∪ b, because we defined O-ring distress as either erosion or blowby. Following Dalal et al.[1], we use p_a, p_b, p_c, and p_d to denote the probabilities of (a) – (d), conditional upon the preceding events. The probability of failure of a field joint is then given by the product of these conditional probabilities:

$$p_F = p_a p_b p_c p_d \qquad (3)$$

There are six field joints on each shuttle. Assuming the joint failures are independent and identically distributed, then the probability of shuttle failure due to field joint failure is the probability that at least one of the six joints fails:

$$p_{sh} = 1 - (1 - p_F)^6 \qquad (4)$$

Note that this assumption is highly questionable; joint failures are likely to be dependent, and thus we are likely calculating a lower bound on the shuttle failure probability.

We now turn to estimating each of the inputs to Equ. (3). For p_a, we have the primary O-ring erosion data given in Table 10 below.

Table 10 Data on primary O-ring erosion for shuttle field joints

Flight	Erosion	Temp (^{o}F.)	Press (psig)
1	0	66	50
2	1	70	50
3	0	69	50
5	0	68	50
6	0	67	50
7	0	72	50
8	0	73	100
9	0	70	100
41-B	1	57	200
41-C	1	63	200
41-D	1	70	200
41-G	0	78	200
51-A	0	67	200
51-C	2	53	200
51-D	0	67	200
51-B	0	75	200
51-G	0	70	200
51-F	0	81	200
51-I	0	76	200
51-J	0	79	200
61-A	0	75	200
61-B	0	76	200
61-C	1	58	200

Dalal et al.[1] fit a logistic regression model for p_a using this data, with temperature and leak-test pressure as explanatory variables. They concluded that pressure was not a significant variable, but kept it in the model because NASA engineers had thought that it would be an important predictor of erosion. We performed a Bayesian analysis of this model, as above.

The table below shows the posterior mean, standard deviation, and symmetric 90% interval for each of the parameters in the logistic regression model for p. These results compare well with the MLEs and 90% bootstrap confidence intervals obtained by Dalal et al.[1] Zero is near the center of the marginal posterior distribution for the coefficient of pressure, indicating that pressure is not a significant explanatory variable, as was concluded by Dalal et al.[1] Note also the larger Bayesian p-value for the model with temperature alone, and the slightly smaller DIC. Together, these suggest that the model without pressure as an explanatory variable is better able to replicate the observed data. We will estimate the probability of shuttle failure both with pressure included, and with the simpler model that only includes temperature as an explanatory variable.

Table 11 Summary posterior estimates of logistic regression parameters for primary O-ring erosion, temperature and pressure included as explanatory variables

Parameter	Mean	Standard Dev.	90% Interval
a (intercept)	8.38	5.38	(0.43, 17.46)
b (temp. coeff.)	-0.19	0.07	(-0.31, -0.08)
c (press. coeff.)	0.004	0.01	(-0.01, 0.02)

Table 12 Model validation results for logistic regression models for primary O-ring erosion

Explanatory Variables	DIC	Bayesian p-value
Temperature and pressure	26.84	0.34
Temperature	24.91	0.50

We next consider the conditional probability of primary O-ring blowby, given erosion, p_b. Of the seven field joints that exhibited erosion, only two also exhibited blow-by. This is too sparse a sample for regression modeling, so we follow Dalal et al.[1] and estimate p_b by pooling the data from field O-rings with data from nozzle O-rings, which exhibited similar performance (5 blow-by events in 17 erosion events). This gives a total of 7 blow-by events in 24 erosion events with which to estimate p_b. Dalal et al.[1] chose a uniform(0, 1) prior for p_b; a Jeffreys prior, which is a beta(0.5, 0.5) distribution, would be a more standard choice in PRA. However, the extra bias introduced by the uniform prior is minimal in this case, so we will retain the uniform prior for p_b.

We turn next to the conditional probability of secondary O-ring erosion, given blow-by of the primary O-ring, p_c. Again, there was very little data with which to quantify p_c. Dalal et al.[1] pooled data from field and nozzle O-rings, yielding two events out of seven in which primary O-ring blow-by led to erosion of the secondary O-ring. Again, Dalal et al.[1] used a uniform prior for p_c.

No events existed in which a secondary O-ring had failed following primary O-ring erosion and blow-by, followed by secondary O-ring erosion. Therefore, Dalal et al.[1] set

p_d equal to p_b. This correlates the state of knowledge of p_d with that of p_b, and changes Equ. (3), giving the probability of field joint failure, to

$$p_F = p_a p_b^2 p_c \qquad (5)$$

Table 13 shows the complete WinBUGS script used to estimate each of the terms in Equ. (5) and to propagate the uncertainties represented by the posterior distributions of the conditional probabilities in this equation to obtain the distribution for shuttle failure probability as a result of O-ring failure.

Table 13 WinBUGS script used to calculate probability of shuttle failure as a result of field joint failure

```
model {
for(i in 1:K)    {
  erosion.prim[i] ~ dbin(p.a[i], 6)
  logit(p.a[i]) <- a + b*temp[i] + c*press[i]
  }
blowby.erode ~ dbin(p.b, n.erode.blby) #Binomial dist. for primary blowby, given
erosion
n.erode.blby <- 24 #Pooled field and nozzle O-ring data
p.b ~ dunif(0, 1) #Prior used by Dalal et al.
erode.sec ~ dbin(p.c, n.erode.sec) #Binomial dist. for sec. erosion, given primary
erosion and blowby
n.erode.sec <- 7
p.c ~ dunif(0, 1)
p.F.31 <- p.a.31*pow(p.b, 2)*p.c #Probability of field joint failure at 31 deg. F
p.F.60 <- p.a.60*pow(p.b, 2)*p.c #Probability of field joint failure at 60 deg. F
p.sh.31 <- 1 - pow(1-p.F.31, 6) #Probability of shuttle failure at 31 deg. F
p.sh.60 <- 1 - pow(1-p.F.60, 6) #Probability of shuttle failure at 60 deg. F
logit(p.a.31) <- a + b*31 + c*200
logit(p.a.60) <- a + b*60 + c*200
a ~ dnorm(0, 0.000001) #Diffuse priors on logistic regression coefficients
b ~ dnorm(0, 0.000001)
c ~ dnorm(0, 0.000001)
}

data
list(blowby.erode=7, erode.sec=2)

inits
list(a=5, b=0, c=0) #Logistic model for temp and press
list(a=1, b=-0.1, c=0.1)
```

This script was run with 1,000 burn-in samples, followed by 100,000 samples for parameter estimation. The results for shuttle failure probability are summarized in Table 14. As can be seen, the probability of shuttle failure is essentially independent of any effect due to leak-test pressure, and is significantly higher at the lower of the two temperatures. The posterior distributions for shuttle failure probability at the two different temperatures are shown below.

Table 14 Summary of posterior distributions for shuttle failure due to field joint failure

Explanatory Variables	Mean		90% Interval	
	31° F.	60° F.	31° F.	60° F.
Temperature and pressure	0.163	0.02	(0.03, 0.39)	(0.0035, 0.07)
Temperature	0.165	0.02	(0.03, 0.39)	(0.0035, 0.07)

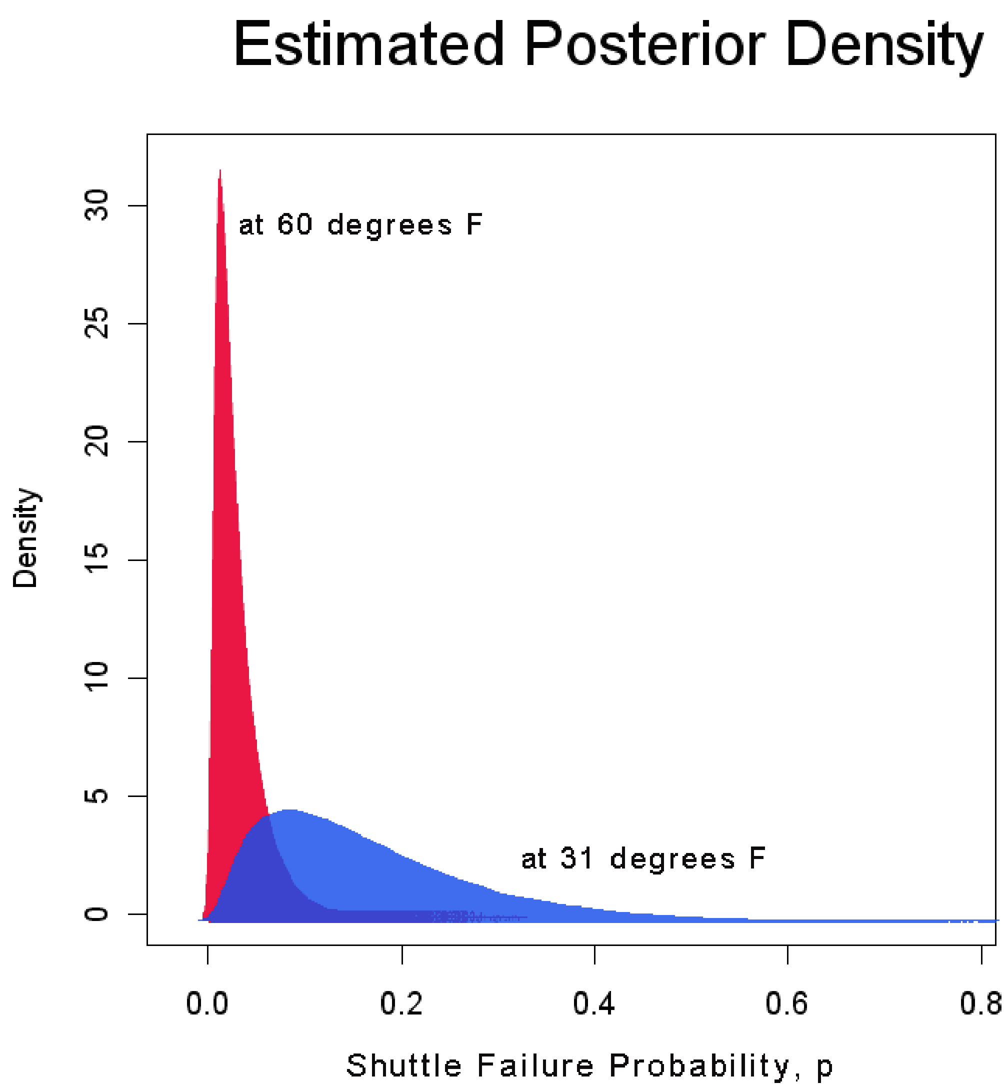

Figure 1 Posterior distribution for shuttle failure probability at 31° and 60° F.

Incorporating Uncertainty in Launch Temperature

In the Bayesian framework, it is straightforward to incorporate uncertainty about the launch temperature to represent the decision-maker's state of knowledge in advance of the launch. This is a distinct advantage over the frequentist framework. To illustrate this concept, we will treat the temperature at launch as unknown, as it would be during the launch planning stages. We will assume the average launch temperature in January is equal to the average low temperature (from www.weatherbase.com) of 52° F. We will further assume that the lowest reported temperature during January (26° F.) represents a difference of two standard deviations. We will thus take the predicted launch temperature as being normally distributed with a mean of 52° F. and a standard deviation of 13° F.

This requires only a slight modification to the WinBUGS script shown in Table 13. Table 15 shows the revised script. The predicted mean shuttle failure probability due to O-ring failure is now 0.08, with a 90% credible interval of (0.001, 0.27). This illustrates the value of knowing the launch temperature; it refines the estimate of shuttle failure probability significantly, providing more information to the decision-maker.

Table 15 WinBUGS script for predicting shuttle failure probability when launch temperature is unknown

```
model {
for(i in 1:K)    {
        erosion.prim[i] ~ dbin(p.a[i], 6)
logit(p.a[i]) <- a + b*temp[i] + c*press[i] #Model with temperature and pressure
        erosion.prim.rep[i] ~ dbin(p.a[i], 6)
        }
blowby.erode ~ dbin(p.b, n.erode.blby) #Binomial dist. for primary blowby, given
erosion
n.erode.blby <- 24 #Pooled field and nozzle O-ring data
p.b ~ dunif(0, 1) #Prior used by Dalal et al.
erode.sec ~ dbin(p.c, n.erode.sec) #Binomial dist. for sec. erosion, given primary
erosion and blowby
n.erode.sec <- 7
p.c ~ dunif(0, 1)
p.F.pred <- p.a.pred*pow(p.b, 2)*p.c
p.sh.pred <- 1 - pow(1-p.F.pred, 6)
erosion.prim.pred ~ dbin(p.a.pred, 6)
logit(p.a.pred) <- a + b*temp.pred + c*200
temp.pred ~ dnorm(52, 0.006)
a ~ dnorm(0, 0.000001)
b ~ dnorm(0, 0.000001)
c ~ dnorm(0, 0.000001)
}
```

Conclusions

We have demonstrated the value of developing models for unobservable parameters, such as O-ring failure probability, in which the unobservable parameter is a function of measurable parameters such as temperature and leak-test pressure. Incorporating such explanatory variables into the model helps to foster communication between risk analysts and system engineers, who are often more comfortable working with measurable quantities.

Bayesian estimation of the parameters in such models has been extremely difficult in the past, and has necessitated complex approximation methods, such as bootstrapping, to propagate parameter uncertainty through the model. However, the advent of easy-to-use, powerful, open-source software such as WinBUGS has made this type of analysis quite tractable, even to nonspecialists.

The Bayesian framework is particularly suited to risk-informed decision-making as it allows uncertainties in observable launch parameters such as temperature to be propagated through the model. The decision-maker can easily see the refinement in model estimates obtained by gathering additional information. The Bayesian framework is also well suited to model validation, an important but often overlooked aspect of risk analysis. This validation step also aids dialog between risk analysts and system engineers. In our example, system engineers thought leak-test pressure would be an important predictor of primary O-ring erosion, but this turned out not to be the case. This outcome could be fed back to the system engineers, who in turn might find this insight a valuable addition to their knowledge.

References

1. Dalal, S. R. et al., “Risk Analysis of the Space Shuttle: Pre-Challenger Prediction of Failure,” Journal of the American Statistical Association, Vol. 84, No. 408 (Dec. 1989), pp. 945-957.
2. Efron, B., “Bootstrap Methods: Another Look at the Jackknife,” The Annals of Statistics, 7 (1979), pp. 1-26.
3. Lunn, D. J., et al., “WinBUGS - A Bayesian Modelling Framework: Concepts, Structure, and Extensibility,” Statistics and Computing (2000), **10**, 325-337.
4. Gelman, A. et al., “Posterior Predictive Assessment of Model Fitness via Realized Discrepancies,” Statistica Sinica **6** (1996), 733-807.
5. Bayarri, M. J. and Berger, J., “Quantifying Surprise in the Data and Model Verification,” in Bayesian Statistics 6, J. M. Bernardo et al., eds., pp. 53-82, Oxford University Press, 1998.
6. Spiegelhalter, D. J., et al., “Bayesian Measures of Model Complexity and Fit,” Journal of the Royal Statistical Society B (2002), **64**, Part 4, pp. 583-639.