REDUCTION OF RISK AND UNCERTAINTY IN COMPOSITES PROCESSING USING PROCESS MODELLING AND BAYESIAN STATISTICS

Göran Fernlund

Department of Materials Engineering , The University of British Columbia
309-6350 Stores Rd., Vancouver, Canada, V6T 1Z4
goran.fernlund@ubc.ca

ABSTRACT

There is often an appreciable amount of uncertainty and risk associated with the development of manufacturing processes for large integrated composites structures. This paper presents a Bayesian approach to combine partially reliable and heterogeneous information into a probability distribution function (PDF) of the outcome of a manufacturing process. The developed PDF is used to guide manufacturing decisions in order to reduce uncertainty and risk associated with the process. A case study is presented where past manufacturing experience, prototype data, and model predictions are combined to establish a PDF for the process-induced deformation of a composite structure. The PDF is used to determine the appropriate geometric compensation of the process tool so that the probability of the composites structure being within dimensional tolerances is increased, and production risk reduced.

1. INTRODUCTION

As larger and more integrated composite structures are being designed and produced in the aerospace industry, e.g. the Boeing 787 and the Airbus A350 XWB, the difficulty of manufacture increases. This leads to greater uncertainty of how to design the process in order to produce the structure within cure cycle specification, with acceptable porosity levels, and within dimensional tolerance. The short time-lines in industry today often necessitate investment in costly process equipment despite uncertainty if the equipment will be able to produce the structure within specifications. Increased uncertainty leads to increased manufacturing risk as the process may require time consuming and costly modifications or redesign before acceptable structures are produced. This has led to a need to better understand the potential outcomes of a new manufacturing process and the probabilities and risks associated with the different outcomes.

There is uncertainty whenever there is more than one outcome of an event and the true event is unknown. Uncertainty is generally measured by assigning probabilities to the different outcomes of an event. Uncertainty is often divided into irreducible and reducible uncertainty, where the irreducible component is related to phenomena which we cannot easily control. Irreducible uncertainty in the current context is inherent variability of the manufacturing process. The reducible component is related to lack of information and can generally be reduced by acquiring more and higher quality information. The information available with regards to a new manufacturing process is typically past manufacturing experience, prototype data, and possibly predictions from a process model. However, the heterogeneity and uncertainty in this information makes it is challenging to combine the information and establish a probabilistic prediction or probability distribution function (PDF) of the outcome. Risk is related to the probability of an unwanted outcome of an event and the cost associated with that outcome. Risk can be measured by a set of outcomes each with quantified probabilities and quantified costs. Risk is reduced by having a better understanding of the probabilities of the different outcomes and by adjusting the manufacturing process accordingly.

Engineering decisions in situations of uncertainty are often made ad-hoc using engineering judgment with subjective and arbitrary weighting of the evidence at hand. There is ample evidence from the fields of psychology and behavioral economics that humans in general are prone to make irrational decisions when faced with situations of uncertainty [1]. To avoid making irrational decisions in these circumstances it is often recommended that structured processes based on statistics and probability theory are used [1]. This paper presents a Bayesian approach to combining manufacturing experience, prototype data, and process simulations into a Bayesian PDF or belief function of the outcome. This knowledge is then used to adjust the manufacturing process in order to reduce manufacturing risk. The methodology is illustrated with a case study where a belief function of the process-induced deformations of a composites structure was developed. The belief function was used to determine the appropriate geometric compensation for the process tool and the risk reduction achieved.

2. BAYESIAN STATISTICS AND PROBABILISTIC PREDICTIONS

Assume that a specific outcome of a new manufacturing process that we are interested in is a random variable Y that is Normally distributed with unknown mean μ but known standard deviation σ, denoted $Y \sim N(\mu,\sigma^2)$. A Normal assumption is appropriate when there are many independent factors that contribute to variability in the outcome Y. The standard deviation σ is assumed known based on past manufacturing experience for the type of process under consideration. Our objective is to establish a probability distribution function (PDF) for the outcome Y of the process. The standard deviation σ is a measure of the irreducible uncertainty in the current case, whereas the reducible uncertainty is related to how well we can estimate the mean value μ from information available to us. As we will never be able to exactly determine the value of μ, our PFD for Y, f(y) = Probability(Y=y), will always be broader and have larger variance than $N(\mu,\sigma^2)$.

We will assume that we have three sources of information relevant to the problem: past manufacturing experience, prototype data, and model (simulation) results. We start by creating a Normally distributed PDF for the unknown mean μ, $g(\mu)$ based on past manufacturing experience

$$g(\mu) = N(m, s^2) \qquad (1)$$

The objective here is to provide a best estimate of the expected mean value m and standard deviation s for the value of μ, based on past experience. For guidance, select m as the most likely outcome based on experience and s = [(highest possible value) – (lowest possible value)]/6 [2]. This means that we believe that there is a 99.7% probability that μ lays between that highest and lowest possible value. The function $g(\mu)=N(m,s^2)$ is a mathematical representation of our entire prior belief of μ based on past manufacturing experience.

Bayesian statistics allows us to update our PDF $g(\mu)$ based on prototype data and model predictions as they become available. If the prototype data D or model predictions M, are considered random samples from the population of interest, each with *n* data points and average value $\bar{y}$, Bayesian statistics gives that our updated PDF, given the data D or model predictions M $g(\mu \mid D \text{ or } M)$, is also Normally distributed but with mean m_i and standard deviation s_i [2].

$$g(\mu \mid D\,or\,M) = N(m_i, s_i^2); \quad m_i = \frac{m\sigma^2 + n\bar{y}s^2}{\sigma^2 + ns^2}; \quad s_i = \frac{\sigma s}{\sqrt{\sigma^2 + ns^2}} \qquad (2)$$

The mean and standard deviation of this updated distribution are dependent on both the known standard deviation σ, the prior information (m,s^2) as well as the new information $\bar{y}$ and n. Equation (2) represents a probabilistic fusion of this information. The larger the number of data points n the more weighted m_i becomes towards $\bar{y}$ and the less important the prior belief is.

When performing process simulation of a manufacturing process, multiple, independent analyses are often performed for model validation and to develop confidence in the predictions [REF]. These independent analyses may be treated the same way as experimental data, and be used directly in eq. (2). However, different analyses are seldom truly independent as they often share some fundamental data or underlying assumptions. If many simulations are performed an unrealistic weight will be assigned to the model predictions if they are treated the same way as experimental data. To avoid this, it is recommended that a reasonable standard deviation for the model that corresponds to past agreement between model predictions and experimental data is established. Sample standard deviation s is related to the number of data points n and the population standard deviation σ through $n = (\sigma/s)^2$ for a random sample from a Normal distribution. With $\bar{y}$ and n established for the model predictions, the PDF given the model predictions $g(\mu\,|\,M)$ can be updated according to eq. (2). With the updating scheme shown in eq. (2) it doesn't matter in what order updating is done, data first or model first, or if data and model results are pooled together. However, will use subscripts M and D to denote the data set the updating was based on as the pedigree and reliability of the different data sets may be different.

The presented equations are based on that experimental data and model predictions can be considered random samples from the population of interest, which may not always be true. The population of interest is often a production part, which has not yet been produced. Experimental data are taken from a prototype, which often is a sub-scale simplified version of the part of interest. The prototype is often laid-up, bagged, and processed under conditions that are not identical to the final production part. Similarly a model is a simplified mathematical description of the production part that may not contain all the details and physics of the final production part. To address that neither prototype data nor model predictions may not fully represent the final production part a probability p ($0 \le p \le 1$) that the data p_D and model p_M predictions actually are random samples from the population can be introduced [4]. If the data or model predictions are representative of the population, the PDF should be updated according to eq. (2), if not, no updating should take place and the prior PDF should remain. The updating scheme for partially reliable data sources is shown in Figure 1.

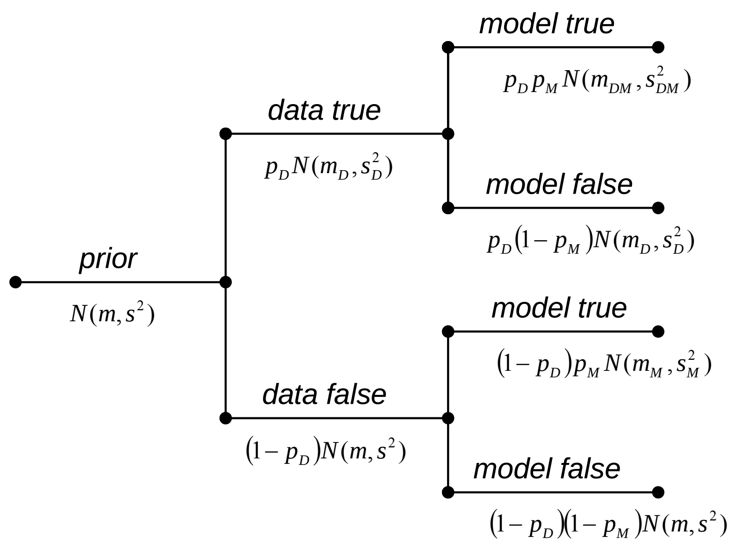

Figure 1. Diagram of the updating process for partially reliable data and model predictions.

Conditional probability analysis then gives the following PDF for the population mean μ given Data, Model, and probabilities p_D and p_M

$$g(\mu \mid D, M, p_D, p_M) = p_D p_M N(m_{DM}, s_{DM}^2) + p_D(1 - p_M)N(m_D, s_D^2) + (1 - p_D)p_M N(m_M, s_M^2) + (1 - p_D)(1 - p_M)N(m, s^2) \tag{3}$$

The subscripts D, M, and DM denotes updating based on data only, model only, and both data and model, respectively. Equation (3), represents our entire belief of the population mean μ given past experience, data, model predictions, and applicability of data and model to the problem at hand. To develop a PDF for the outcome f(y), given the uncertainty in μ given by eq. (3), we have that

$$f(y) = P(Y = y \mid D, M, p_D, p_M) = \int_{-\infty}^{+\infty} f(y \mid \mu) \cdot g(\mu \mid D, M, p_D, p_M) d\mu \tag{4}$$

where $f(y|\mu) = N(\mu,\sigma^2)$. Integration, or marginalization, of eq. (4) gives that the resulting PDF *f(y)* is

$$f(y) = p_D p_M N(m_{DM}, \sigma^2 + s_{DM}^2) + p_D(1 - p_M)N(m_D, \sigma^2 + s_D^2) + (1 - p_D)p_M N(m_M, \sigma^2 + s_M^2) + (1 - p_D)(1 - p_M)N(m, \sigma^2 + s^2) \tag{5}$$

Thus the PDF for our belief of the outcome of the process *f(y)* is a linear combination of four Normal distributions of the same form as for the PDF for the unknown mean μ (eq 3), but with larger variance due to the variance of μ itself σ^2.

We can now calculate some important properties of our belief function of the outcome. The expected or mean value $\overline{m}$ of our belief function *f(y)* can be calculated to be

$$\overline{m} = p_D p_M m_{DM} + p_D(1 - p_M)m_D + (1 - p_D)p_M m_M + (1 - p_D)(1 - p_M)m \tag{6}$$

Similarly, the variance $\overline{s}^2$ is

$$\overline{s}^2 = p_D p_M \left[(\overline{m} - m_{DM})^2 + (\sigma^2 + s_{DM}^2)\right] + p_D(1 - p_M)\left[(\overline{m} - m_D)^2 + (\sigma^2 + s_D^2)\right] + (1 - p_D)p_M \left[(\overline{m} - m_M)^2 + (\sigma^2 + s_M^2)\right] + (1 - p_D)(1 - p_M)\left[(\overline{m} - m)^2 + (\sigma^2 + s^2)\right] \tag{7}$$

The first term within each square bracket represents the between sample variance whereas the second term represents the within sample variance. Equations (5-7) summarize our belief of the actual distribution $N(\mu,\sigma^2)$ given past experience, data, and model predictions. As the standard deviation of a distribution is related to the uncertainty in the outcome we will use it as a measure of how well we have managed to reduce the uncertainty using prior information, data and model predictions. We will define the effectiveness E of our belief model as the ratio of the irreducible standard deviation σ to the standard deviation of our belief model $\overline{s}$

$$E = \frac{\sigma}{\overline{s}} \tag{8}$$

Thus the smaller we can make $\bar{s}$, the higher the effectiveness E of our belief model ($0 \le E \le 1$). In the special case when we believe that data and model predictions are true random samples of the population, $p_D = p_M = 1$, eqs. (5-8) reduce to

$$f(y) = N(m_{DM}, \sigma^2 + s_{DM}^2) \tag{9}$$

$$\overline{m} = m_{DM} \tag{10}$$

$$\bar{s}^2 = \sigma^2 + s_{DM}^2 \tag{11}$$

$$\begin{aligned} E(n=0) &= \frac{1}{\sqrt{1+(s/\sigma)^2}} \\ E(n \ge 1) &\approx \sqrt{\frac{n}{n+1}} \end{aligned} \tag{12}$$

where n is the total number of data and model predictions pooled together. Equation (12) shows that without any data or model predictions the effectiveness E is typically low ($s > \sigma$) resulting in a large uncertainty in the outcome. However, even with a small amount of data n, the effectiveness increases rapidly. However, the effectiveness drops when p_D, p_M are less than one as will be illustrated in the case study below.

3. CASE STUDY

Cured composite structures seldom have the same dimensions as those of the tool they were processed on due to mismatch in thermal expansion, cure shrinkage, and residual stress build-up during processing. This process-induced deformation often requires a geometric compensation of the tool so that the processed part meets dimensional tolerances. In the presented case (for more details see [3]), the problem was to determine an appropriate geometric tool compensation for a large number of different sized composite parts that can be described as curved L-angles with cut-outs, shaped as circle sectors (Figure 2). The parts are made of 180°C cure carbon-epoxy prepreg and are processed on a female metal tool in an autoclave.

Manufacturing of this structure represented significant risk as it was not obvious beforehand what the significant modes of process-induced deformation and their magnitudes would be, nor their dependence on process conditions such as the tool material. If process-induced deformation would turn out to be significant and no compensation of the tooling had been performed there would be a significant cost to address the problem after the fact through increased assembly cost or redesign of the tool. To determine the significant modes of deformation and their magnitudes, a coordinated approach with process analysis and prototype testing was undertaken. Two-dimensional and three-dimensional finite element analysis as well as closed-form process analysis, all indicated that the main modes of deformation for the structure were “spring-down” and tab “spring-in” and (see Figure 2). Tab “spring-in” was assessed as the only deformation mode that required tool compensation as the restoring forces required to prevent the spring-down were low. These analytical findings were confirmed by experimental measurements on prototype parts. Table 1 shows the predicted and measured tab spring-in from four different models and experimental measurements.

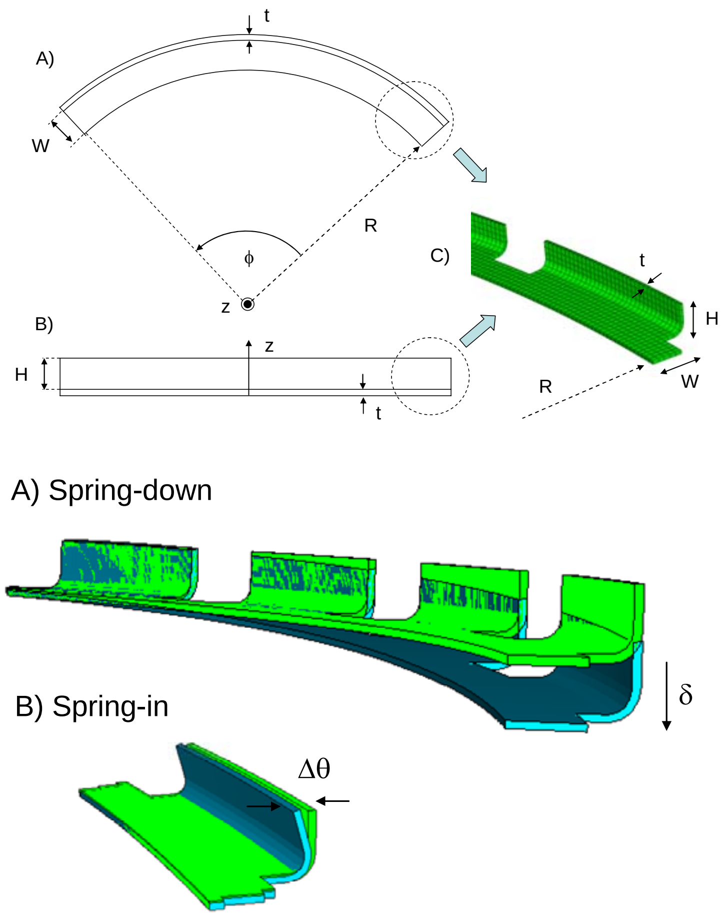

Figure 2. Top) Schematic of composite part; Bottom) Predicted process-induced deformations "spring-down" and "spring-in" from FE analysis.

Table 1. Results from process analysis (model) and prototype testing (data).

Data source	Spring-in (deg)
Model 1	0.78°
Model 2	0.83°
Model 3	0.76°
Model 4	0.83°
Model mean =	**0.80°**
Data 1	0.87 °
Data 2	0.93 °
Data 3	0.84 °
Data 4	0.86 °
Data mean =	**0.88°**

With the data presented in Table 1 we can use the methodology presented in the previous section to establish a probability distribution function for the spring-in. However, before we start we need to define how accurate we need to estimate the spring-in and the inherent variability in the process, σ. We will assume that the standard deviation for spring-in of the process $\sigma = 0.03°$ and that we need the spring-in of the manufactured parts to be within $\pm 0.1°$ of engineering specifications to not incur any additional cost in assembly of the structure. Thus, if the outcome of the process is normally distributed, $N(\mu,\sigma^2)$, less than 0.1% of all parts will be outside dimensional

tolerances if the tool is compensated exactly for the population mean μ (assuming that the mean is around 1°).
Based on prior experience from similar parts, material systems and processes, our best estimate of the tab spring-in is 1° and we are very confident (99.7% sure) that the spring-in is between 0.55 and 1.45°. If this information is converted to a normal distribution of the form in eq. (1) we have that m = 1°, and s = 0.15°. This is our "prior" estimate of the outcome before we have received any experimental data or model results. Based on past experience of how reliable our model and our measurement techniques are we assign the following probabilities for the reliability of data and models $p_D = p_M = 0.9$. Table 2 shows the relevant parameters and calculated results using the presented methodology, and Figure 2 shows a comparison of the prior $N(m,\sigma^2 + s^2)$ and posterior eq. (5) estimates of the outcome of the process. The required tolerance interval +/- 0.1° has also been indicated and centered around the mean value of both distributions in Figure 2.

Table 2. Input data and results from Bayesian analysis.

Parameter	Value	Parameter	Value
m =	1	m_D =	0.88
s =	0.15	s_D =	0.01
σ =	0.03	m_M =	0.80
p_D =	0.9	s_M =	0.01
p_M =	0.9	m_{DM} =	0.84
$\bar{y}_D$ =	0.88	s_{DM}=	0.01
n_D =	4	$\bar{m}$ =	0.84
$\bar{y}_M$ =	0.80	$\bar{s}$ =	0.04
n_M =	4	E =	71%

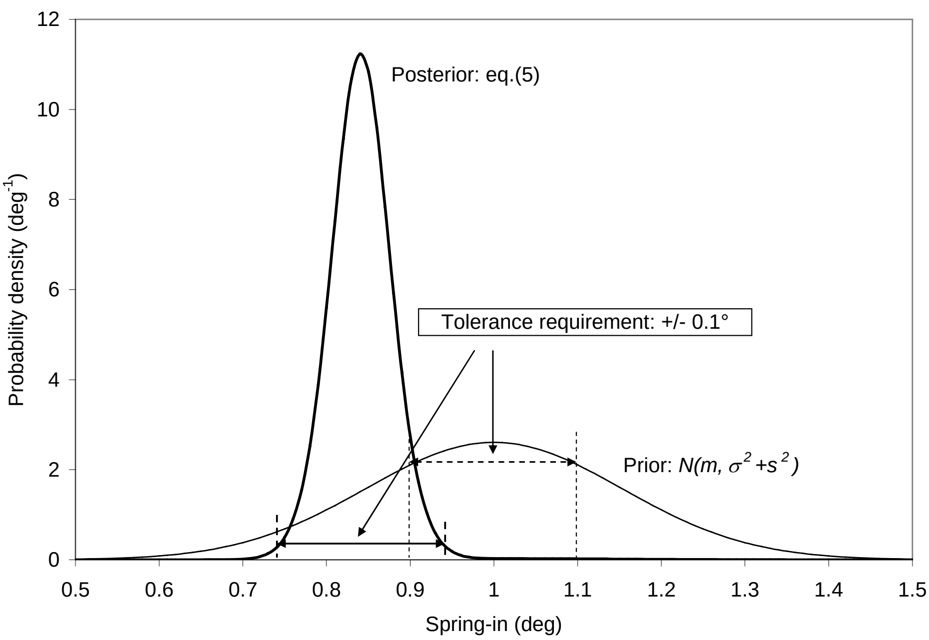

Figure 3. Comparison of prior and posterior probability density distributions for the outcome of the process.

Risk is related to the probabilities and cost associated of different outcomes of a variable process. We will use a very simple definition of risk, i.e., risk is the probability of failure multiplied with the cost of failure. The cost of failure typically increases the further you are off target but in the following we will use a conservative flat cost of failure, which implies that the risk is proportional to the probability of failure. Figure 3 shows that with only prior information based on experience our best decision would be to compensate the process tool for a 1° spring-in. This gives us a probability of failure of 51 % if our tolerances are ± 0.1°. After gathering prototype data and model predictions our best estimate of the mean value of the spring-in is $\overline{m}$ = 0.84° with a standard deviation $\overline{s}$ = 0.04. With this information at hand our best decision is to compensate the tool for a spring-in of 0.84° which would reduce the probability of failure from 58% down to 1.5%. The probability of failure is calculated by integrating eq. (5) from -∞ to $\overline{m}$ -0.1° and from $\overline{m}$ +0.1° to +∞. The probability of failure is represented by the tail areas outside the tolerance requirements shown in Figure 3.

4. DISCUSSION

The results presented in Table 2 and Figure 3 are not particularly surprising. The predicted average spring-in is essentially the average of the model predictions and the experimental data as the weight of the prior given eight data points and the small σ/s ratio in the fusion of information is very small. To study how the predicted average spring-in, the model effectiveness, and the probability of failure is affected by the reliability of data and model predictions the analysis was repeated with different values of $p_D = p_M$ from 0 to 1, see Figure 4.

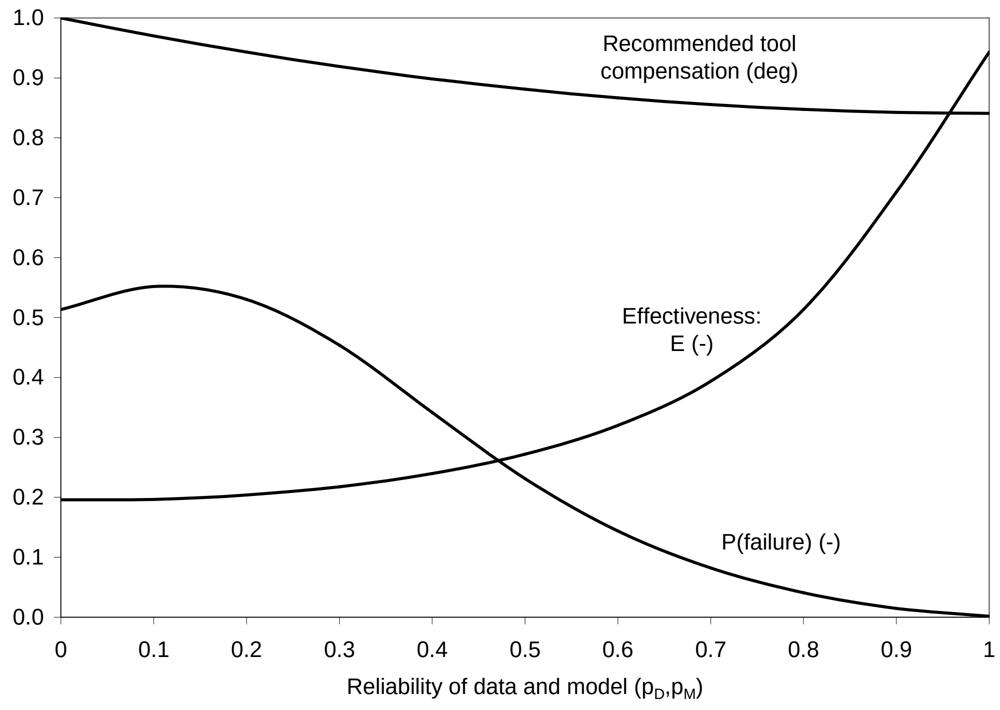

Figure 4. Recommended tool compensation, model effectiveness E (-) and probability of failure P(failure) as a function of the reliability of data and model predictions.

The figure shows that even when there is zero reliability in the data and model predictions (p_D=p_M=0), the effectiveness is close to 20% based on the information contained in the prior. The recommended tool compensation is the mean of the prior, 1°, and the probability of failure is 51%. It is interesting to note that even with data and model results with low reliability, the standard deviation of the belief function and the probability of failure is reduced compared to the case when no data or model results are available.

One interesting feature of the developed model is that it implicitly addresses coherence of information. It is fairly well established that when we in general receive information from several partially reliable sources our confidence in the truth of the information is dependent on: the plausibility of the information, the reliability of the sources, and the coherence of the information [5]. The first two factors are explicitly addressed in the development of the current model whereas the last one is only implicitly addressed. If the reliability of our data and model predictions were 100% it would not matter if all information comes from one source or if half of the data comes from each of the two sources. However, if the information sources are only partially reliable, our confidence in the information increases if two independent sources present us with information that support each other rather than getting all the information from one partially reliable source. To examine how the model deals with this, three variations on the presented case study was examined, see Table 3. Column 1 shows the baseline predictions presented above. Column 2 shows the predictions if the both data and model predictions were 100% reliable. Both these results can be read off Figure 4. Column 3 shows the results if instead of having two data sources with a combined average value of 0.84° and 8 data points we only had one data source. Column 4 is identical to column 3, except that the data sources now are 100% reliable. The Table shows that if the data sources are 100% reliable, it doesn't matter if we have one or two data sources. However, if the data sources are partially reliable, the standard deviation of our belief function and the probability of failure are substantially less if we have two independent data sources.

Table 3. Effect of coherence and reliability of information on probability of failure.

Parameter	**Baseline case: 2 data sources (p_D=p_M=0.9)**	**Variation 1: 2 data sources (p_D=p_M=1.0)**	**Variation 2: 1 data source (p_D=p_M=0.9)**	**Variation 3: 1 data source (p_D=p_M=1.0)**
p_D =	0.9	1	0.9	1
p_M =	0.9	1	0.9	1
$\bar{y}_D$ =	0.88	0.88	0.84	0.84
n_D =	4	4	8	8
$\bar{y}_M$ =	0.80	0.80	-	-
n_M =	4	4	-	-
E =	71%	94%	40%	94%
P(failure) =	1.5%	0.2%	7.1%	0.2%

Two parameters in the model that may appear somewhat arbitrary are the reliability of data and model predictions p_D, and p_M. Traditional statistics does not deal with partially reliable information and when the reliability is 100% and there is no prior information the current model reverts to that of traditional statistics. The reason that both data and model predictions are only partially reliable stems from that none of them are true random samples from the population of interest, which is a future event we typically don't have access to. If this partial reliability is neglected an unreasonable confidence in the outcome of the process is established. It is difficult to fully objectively establish the reliability of the data sources p_D and p_M as they depend on how well our experimental and mathematical models represent a future event. Past agreement of prototype data and model predictions with production data is probably the best basis for assessment of the reliability of data and models.

Engineering is often about making decisions in situations of uncertainty. We typically use "engineering judgment" to combine partially reliable information from various information sources to establish a belief in the outcome of an uncertain event. What actually is going on in this "judgment process" is a bit of mystery and presented with the same information different people come to different beliefs and conclusions. The presented model, where a Bayesian probability distribution function is used as a measure of our state of knowledge, can hopefully help and guide us towards making more rational, more objective, and better decisions in situations of uncertainty and risk in composites manufacturing.

ACKNOWLEDGMENTS
The author acknowledges the contribution to this work by his colleagues at the UBC Composites Group, Boeing Canada Technology Winnipeg, The Composites Innovation Centre in Winnipeg, and Convergent Manufacturing Technologies. Partial funding was provided by The Natural Sciences and Engineering Research Council of Canada.

REFERENCES

1. Dawes R, Reid H. "Rational choice in an uncertain world: the psychology of judgment and decision making". *Sage Publications, Inc*; 1st edition, 2001.
2. Bolstad W.M. "Introduction to Bayesian Statistics". *John Wiley & Sons Inc*., 2004
3. Fernlund G., Floyd A., McKay S. Process analysis and tool compensation for Curved composite L-angles. *6th Canadian International Conference on Composite Materials*, Winnipeg, August 14-17 (2007).
4. Maskell S. A Bayesian approach to fusing uncertain, imprecise and conflicting information. *Information fusion*. In press 2007.
5. Bovens L, Hartmann S. "Bayesian Epistemology". *Oxford University Press*, 2004.