RESEARCH   **Open Access**

Credibly reaching a reliability target using a model initially constructed by expert elicitation

Lawrence E Pado

Correspondence:
lawrence.e.pado@boeing.com
The Boeing Company, 8900 Frost
Ave., Bldg 245 MC S245-1260, 63134
Berkeley, MO, USA

Abstract

The Defense Advanced Research Projects Agency Defense Science Office (DARPA/ DSO) is sponsoring Open Manufacturing (OM), an initiative to develop new technologies, new computational tools, and rapid qualification to accelerate the manufacturing innovation timeline. Certification Methodology to Transition Innovation (CMTI), an OM program, has developed a methodology to quantify the effect of manufacturing variability on product performance to address the risk to cost and performance associated with failure to take manufacturing capability and material and fabrication/assembly variation into account early in the design process. An important aspect of this program is the use of Bayesian networks (BN) to evaluate risk. The BN is used as a graphical representation of the contributing factors that lead to manufacturing defects. The reliability of the final product is then analyzed using the contributing factors. There are many types of programs where there is little relevant data to support the probabilities needed to populate the BN model. This is very likely the case for new programs or at the end of long programs when obsolescence challenges servicing a product when original vendors are no longer in business. In these cases, probabilities must be obtained from expert opinion using a technique called expert elicitation. Even under objective 'Good Faith' opinions, the expert himself has a lot of uncertainty in that opinion. This paper details an approach to obtaining credible model output based on the idea of having a hypothetical expert whose unconscious bias influences the model output and discovering and using countermeasures to find and prevent these biases. Countermeasures include replacing point probabilities with beta distributions to incorporate uncertainty, 95% confidence levels, and using a multitude of different types of sensitivity analyses to draw attention to potential trouble spots. Finally, this paper uses a new technique named 'confidence level shifting' to optimally reduce epistemic uncertainty in the model. Taken together, the set of tools described in this paper will allow an engineer to cost effectively determine which areas of the manufacturing process are most responsible for performance variance and to determine the most effective approach to reducing that variance in order to reach a target reliability.

Keywords: Credibility; Expert elicitation; Confidence level shifting; Monte Carlo; Uncertainty quantification; Targeted testing; Unitized testing; Uncertainty reduction; Epistemic uncertainty; Reliability targets

Background

The Defense Advanced Research Projects Agency Defense Science Office (DARPA/DSO) is sponsoring Open Manufacturing (OM), an initiative to develop new technologies, new computational tools, and rapid qualification to accelerate the manufacturing innovation timeline. Certification Methodology to Transition Innovation (CMTI), one of the programs in the OM portfolio, has developed a methodology to quantify the effect of manufacturing variability on product performance to address the risk to cost and performance associated with failure to take manufacturing capability and material and fabrication/assembly variation into account early in the design process.

Motivation

The goal motivating this research is to first credibly ascertain the reliability of a product by including the effects of variability and defects in manufacturing as well as uncertainty in the environment. Note that credibility is a key requirement and is made more difficult when using expert elicitation to determine the value of model parameters used to calculate the reliability. Secondly, upon evaluation of the manufacturing process, it is very likely that product reliability will fall short of the desired target. The set of tools described in this paper will allow the engineer to cost effectively determine which areas of the manufacturing process are most responsible for performance variance and to determine the most effective approach to reducing that variance in order to reach a target reliability. These benefits will be explained in detail in the sections entitled 'Techniques to meet a target POF with a 95% confidence level' and 'Putting it all together - an example using credibility tools'.

An exemplar problem

An exemplar problem used to drive development of the framework was the manufacture of an out-of-autoclave composite panel stiffened with three hats (Figure 1 [1]). The hat-stiffened panel represents a design/manufacturing feature-based element or subcomponent in the traditional building block approach. The manufacturing steps called out in the fabrication work order serve as the initial basis for creating a Bayesian network (BN) [2] used to tailor risk with a quality control plan and to determine the probability of defects. The BN is used as a graphical representation of the contributing factors that lead to manufacturing defects. The reliability of the final product is then analyzed using the contributing factors.
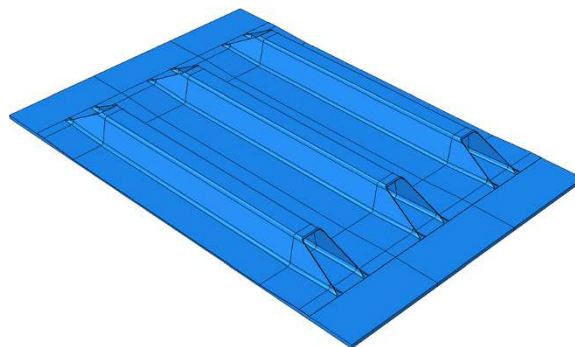

Figure 1 Isometric view of hat-stiffened panel (69 cm wide × 91 cm long).
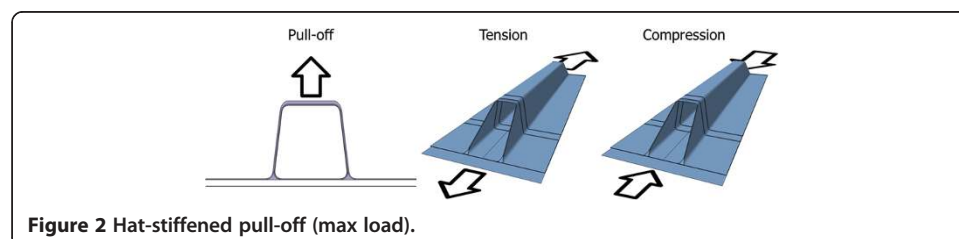
The environmental condition that this part is to be evaluated against for the purposes of this paper is out-of-plane pull-off of a hat structure (Figure 2). It should be noted that this approach used multiple load cases as shown in Figure 2, but for simplicity, only the pull-off case will be discussed. As described in detail in reference [3], given a probabilistic environmental load, geometrical and material property variation, as well as probabilistic manufacturing defects, a probability of failure (POF) of the hat-stiffened panel can be calculated.
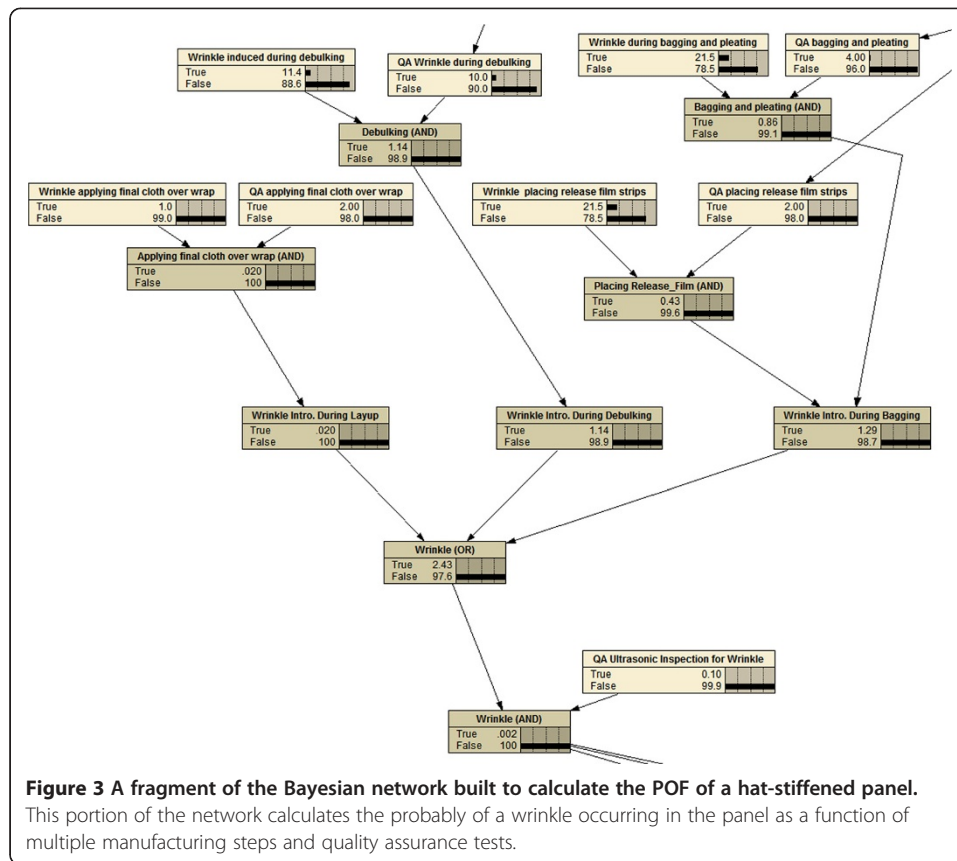
POF is partially a function of the probability of defects occurring during the manufacture of a part. A subset of all possible defects that can be introduced by the manufacturing process, tooling, etc., and that are thought to contribute to failure under acknowledged conditions, were identified. The focus of this work was on quantifying and reducing manufacturing defects. Although material and environmental variability were accounted for, examination of the costs and benefits of reducing the variability of those factors were beyond the scope of this research. They are however important factors and will be considered in future work. For the hat-stiffened panel, the defects recognized were wrinkles (nugget/noodle fiber waviness), noodle void/porosity/geometry, lower radius thickening, upper radius thinning, and top crowning. If there is any doubt as to if a defect can affect performance, it should be included in the analysis.

Once the defects of interest were identified, the step-by-step manufacturing process was analyzed to determine which steps or combinations of steps could possibly produce one or more of those defects. Additionally, options that could affect the probability of introducing defects were identified such as tooling choices, manufacturing capability levels, and manufacturing process alternatives.

At this point in the knowledge collection process, enough information was obtained to create the structure of a Bayesian network. If there is any doubt as to whether or not a factor can influence the probability of a defect, it should be included in the model and assigned a high uncertainty. Uncertainty assignment will be discussed later in this paper. A fragment of the complete BN is shown in Figure 3. The overall purpose of this Bayesian network is to calculate the probability of defects and the resulting probability of failure of the structure. For those readers familiar with process flows, it should be noted that the BN will not necessarily mimic this flow, but will instead be built to capture direct relationships between the process variables. The resultant probability of each defect (and thus the POF) is a function of any and all combinations of the choices that can go into the manufacturing process. The BN is used to determine POF with acceptable cost, or it can be used to find the cost optimal manufacturing process choices given a desired POF as described in [3].

In order for the network to calculate these total defect and POF probabilities, however, the probability that each individual manufacturing step can induce a defect

Figure 2 Hat-stiffened pull-off (max load).

Figure 3 A fragment of the Bayesian network built to calculate the POF of a hat-stiffened panel.
This portion of the network calculates the probably of a wrinkle occurring in the panel as a function of
multiple manufacturing steps and quality assurance tests.

and that each individual quality assurance test can find it, if it exists, must be
provided.

Methods

An approach to determining the credibility of models

For those programs where there is little relevant data to support the probabilities
needed to populate the BN model, expert elicitation must be used to provide them.
Even under objective 'Good Faith' opinions, the expert himself has a lot of uncertainty
in that opinion. On a given day, the expert may even choose different probabilities than
the one he had chosen earlier. Given that the output of the model has real-world con-
sequences, possibly in terms of customer acceptance of the product, there may be bias
when choosing probabilities - especially if the model has not shown that it can meet a
target POF.

The chosen way to approach this issue is to make the potential bias explicit by figur-
ing out the best way to modify model inputs to get a desired result. Once this method-
ology is known, the idea is to reverse engineer it to find countermeasures and establish
the credibility of the model.

The expert's first approach - using and adjusting point probabilities

The expert's initial approach is to make estimates of the probabilities as objectively
as possible. If the target POF is reached, then the expert is done. If not, he will find

the point probabilities that the output is most sensitive to and adjust them as little as possible such that the target is reached. Note that point probabilities are probabilities that are assumed to be known with absolute certainty and are represented by a single scalar value. The reasoning behind this strategy is that adjusting the probabilities that do not have much of an effect would require large unrealistic changes to have a significant effect.

Derivative-based approach to sensitivity

The typical way to perform a sensitivity analysis [4] is to calculate the partial derivative of the output with respect to a model parameter as shown in Equation 1.

$$S_{D_i} = \frac{\delta Y}{\delta X_i} \tag{1}$$

where S_{D_i} is the derivative-based sensitivity measure with respect to parameter i, Y is the model output of interest, and Xi is the model parameter i.

In practice for a Bayesian network model, this sensitivity analysis is performed numerically. The output of interest for the exemplar problem is the POF of the three-hat-stiffened panel under a pull-off load. To calculate the sensitivity of the POF to each node's point probability of either inducing a defect or for QA nodes, the probability of missing a defect if it exists, Equation 2 is used:

$$S_{D_i} = \frac{\Delta \text{POF}}{\Delta P_i} \tag{2}$$

where S_{D_i} is the derivative-based sensitivity measure with respect to probability i, POF is the probability of failure of the three-hat panel under pull-off load, and P_i is the probability of node i inducing a defect or failing to detect a defect.

With this technique in hand, the expert would calculate the sensitivity of POF to changing the probabilities within every manufacturing step or QA test node of the network. After sorting from most sensitive to least sensitive, he could calculate exactly how much to change the top few nodes to reach his target POF.

Quantifying uncertainty in the model probability parameters - the beta distribution

Although one possible countermeasure to this would be to perform a similar sensitivity analysis and use the findings to focus scrutiny on the highest sensitivity nodes, a bigger issue is that the point probabilities express complete certainty in the model values. The first countermeasure is to note that it is not possible for an expert to have absolute certainty in the values used as model parameters. Even with hard data, there is some uncertainty as to the exact value. The conclusion to be drawn, then, is that each model parameter, each of which is a probability, should be represented by a probability density function, ideally one that has zero as a lower bound and one as a higher bound due to the fact that probabilities by definition are always between 0 and 1. The beta distribution is such a distribution and is useful for representing binary success/failure problems, specifically the proportion of successes or failures that would be expected over time. Mathematically, the beta distribution is represented by Equation 3 [5].

$$beta(a,b) = \frac{1}{\beta(a,b)} p^{a-1} (1-p)^{b-1} \tag{3}$$

where p = proportion of flaws $0 <= p <= 1$, a = number of flawed examples (positive number), b = number of flawless examples (positive number), and β = the beta function (not to be confused with the beta distribution).

The beta distribution is quite flexible and can represent uniform ($a = 1$, $b = 1$), ramp, symmetrical, and asymmetrical distributions. It is simple to implement Bayesian learning using newly introduced data using this distribution. For each step in the manufacturing process, if a flaw occurs during that step, the parameter 'a' merely needs to be incremented by one. Likewise, if no flaw is introduced during that step, the parameter 'b' should be incremented by 1. For quality assurance (QA) tests, if a QA test does not miss a flaw that exists, then 'b' should be incremented. If the test misses a flaw that exists, 'a' should be incremented.

Using expert opinion elicitation to determine the parameters of the beta distribution

The goal of expert opinion elicitation is to determine the parameters of beta distribution such that it accurately represents the expert's opinion about the most likely value of the probability to be specified (the mode) as well as his uncertainty in that opinion. In this methodology, this is accomplished by having the expert express his uncertainty in terms of the number of samples he has seen. The following basic example will build the readers intuition about this process.

Upon purchasing an unlabeled trick coin received at a magic shop, it is desirable to determine the characteristics of that coin. The bin the coin was stored in noted only that the coin could be weighted to always come up heads (a two-headed coin), always come up tails (a two-tailed coin), or anything in between. This information from the bin represents prior knowledge that the coin can be weighted to any degree. This can be represented by the uniform distribution denoted beta (1, 1). Beta (1, 1) is also known as the non-informative prior distribution. Figure 4A shows a beta distribution in which every weighting is equally likely. To gather more data, the coin is flipped three times and this results in three heads. Clearly, while this coin is definitely not weighted to always come up tails, it is certainly possible that this is a fair coin that has been weighted to come up heads or tails an equal number of times, but it seems more likely that this coin is weighted towards heads. Figure 4B plots that beta distribution after adding in this new data as beta (1, 1 + 3 heads).

Finally after 30 flips, and obtaining 30 heads in a row, it is clear that this is not a fair coin but is very highly weighted towards coming up heads. The updated beta distribution is shown in Figure 4C. Note that even after 30 flips, it is not a sure thing that a heads result will always be obtained. Also note that the distribution is getting narrower and narrower representing an increase in the certainty of the coin's weighting value.

Using the above example as an intuitive example of the meaning of 'samples seen', the expert can be asked to provide a level of uncertainty in terms of samples seen. For the exemplar problem, what proportion of wrinkles has been observed during the debulking process (i.e., the probability of a wrinkle occurring during the debulking process)? Note that this is represented by the node named 'Wrinkle induced during debulking' in Figure 3.

Given these two pieces of information, the most likely value (mode) and the uncertainty in terms of samples seen, the parameters of the beta distribution that meets these requirements are as follows [6]:

$$a = \text{mode} \times (k-2) + 1 \tag{4}$$

$$b = (1-\text{mode}) \times k-2) + 1 \tag{5}$$
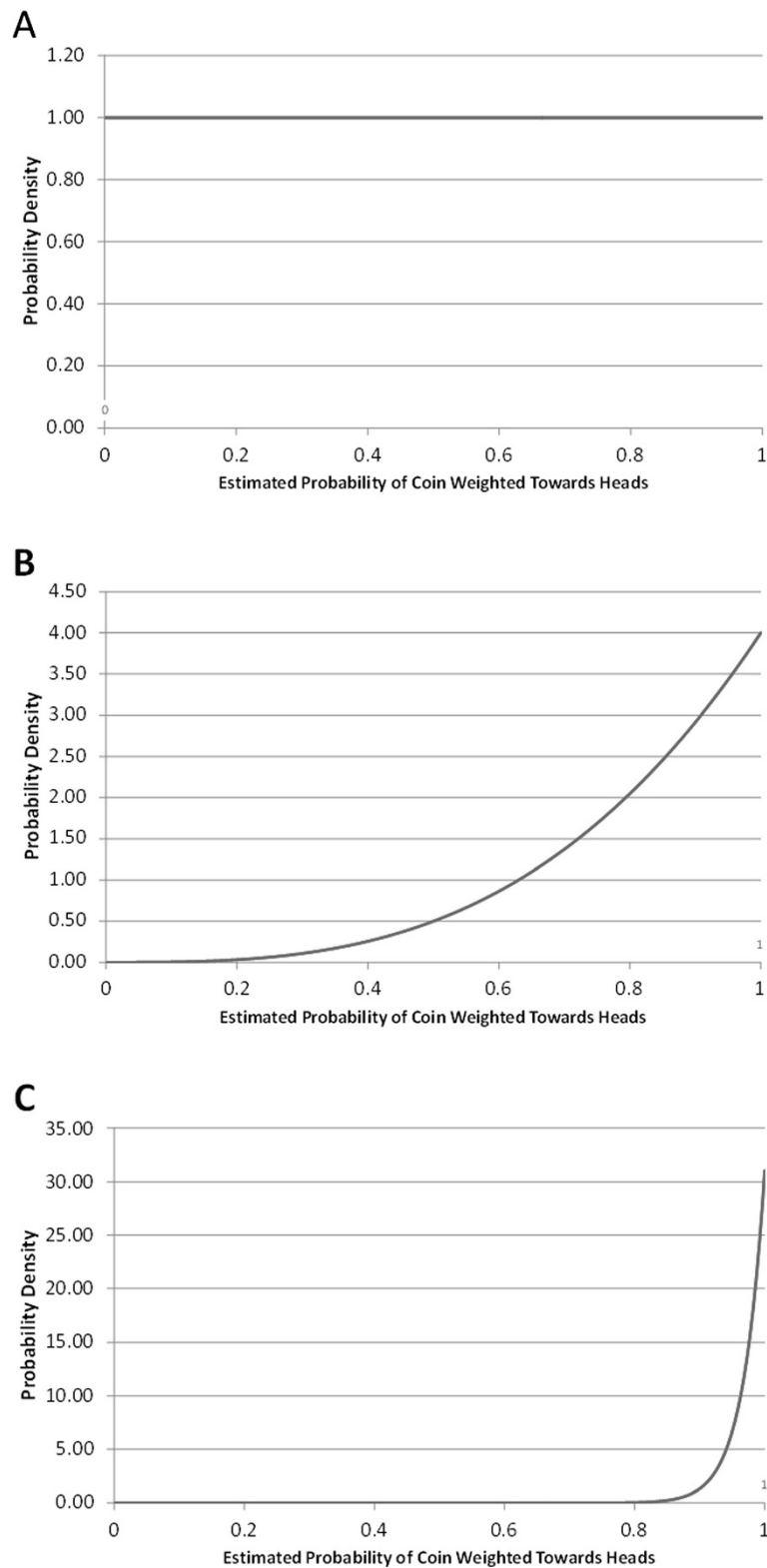
Figure 4 The non-informative beta distribution updated with additional data. (A) The beta distribution beta (1,1), which represents the non-informative state also known as the uniform. **(B)** The beta distribution beta (1,4), which represents uniform distribution updated by getting three heads in three flips. **(C)** The beta distribution beta (1,31), which represents uniform distribution updated by getting 30 heads in thirty flips.

where a = beta distribution parameter expressing the number of flawed examples, b = beta distribution parameter expressing the number of flawless examples, mode = the most likely probability of defect, and k = the expert's confidence in the estimate expressed in terms of equivalent prior sample size (minimum 2).
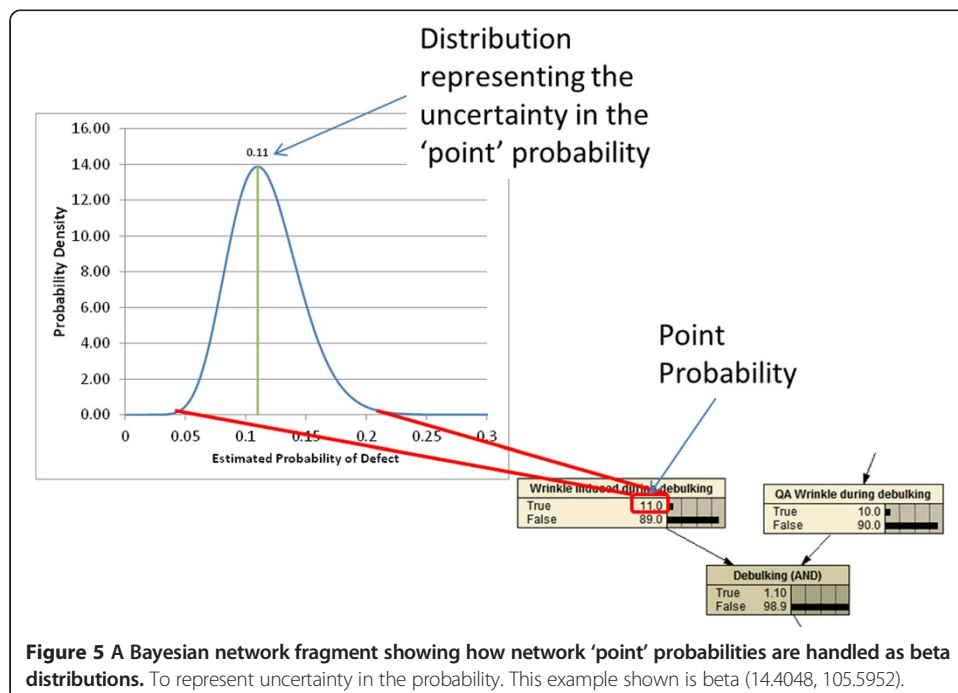
In the debulking example above, the expert provided a most likely value of 0.11 with a sample size of 120. Figure 5 shows how a point probability of 0.11 is now represented by a beta distribution with parameters of beta (13.98, 106.02). Note that in this example, the expert is implying that the range of likely values for that probability is approximately between 0.04 and 0.22.
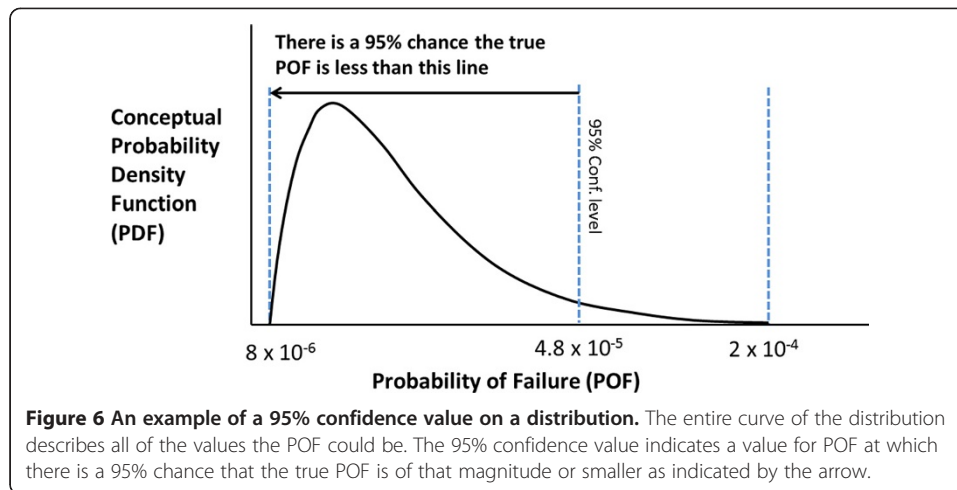
This process is continued for every node in the network until every parameter of the network is represented by a beta distribution.

Determining a 95% confidence value on the model output using Monte Carlo methods

Now that all of the model parameters have been replaced by beta distributions, the distribution of the model output can be computed using Monte Carlo methods [7]. Specifically, a 95% confidence value [8] can be calculated on the model output, which for the exemplar case is POF. Figure 6 provides intuition about the 95% confidence level (CL). As shown in the figure, it is the value at which there is a 95% chance that the true value is less than or equal to the CL. As can be seen, this is a quite conservative estimate since the most likely value (peak or mode) of the shown distribution is much smaller.

A Monte Carlo analysis entails pulling a single sample from each distribution within the model, populating the model with these new samples, and then running the model to get a single answer. This process is then repeated thousands of times to collect enough data to establish the distribution of the output parameter of interest such as POF and enables calculating its CL.

Figure 5 A Bayesian network fragment showing how network 'point' probabilities are handled as beta distributions. To represent uncertainty in the probability. This example shown is beta (14.4048, 105.5952).

Figure 6 An example of a 95% confidence value on a distribution. The entire curve of the distribution describes all of the values the POF could be. The 95% confidence value indicates a value for POF at which there is a 95% chance that the true POF is of that magnitude or smaller as indicated by the arrow.

Finding the 95% CL using histogram data is a straightforward process involving sorting the POF data from lowest value to highest value and then selecting the value for which 95% of the data is that value or smaller.

The expert's second approach - adjusting mode and certainties of PDFs

Using probability distributions instead of point probabilities and instituting a 95% confidence level to introduce conservatism is a good first step for establishing credibility. The expert's bias, if it exists, may now show up as a lower mode value or as a higher level of certainty in that mode value. This type of bias must be detected if it is to be countered.
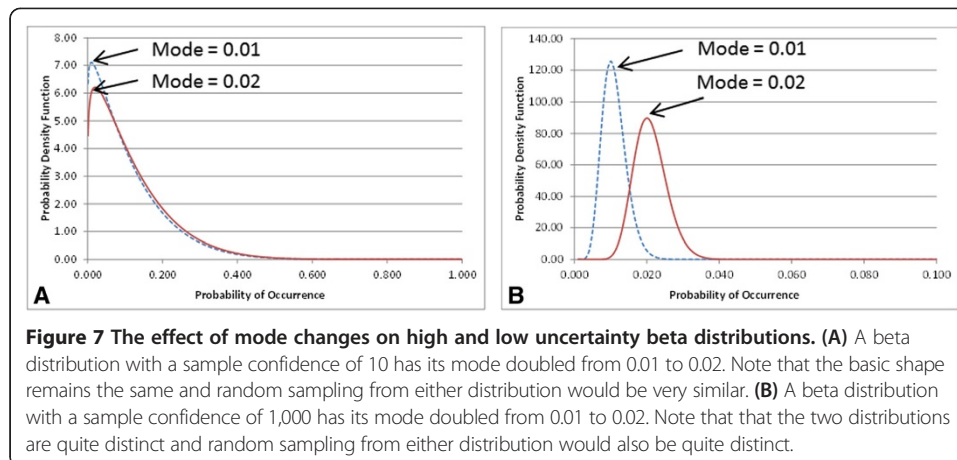
Countermeasure to the expert's second approach - sensitivity analysis on the mode with respect to POF

The countermeasure to the impact of the expert's mode and certainty selection is to perform a sensitivity analysis of the 95% confidence level to changes in the mode.

There are two conditions that have to be met before the 95% confidence level shows significant sensitivity to the mode of a node:

- The node has to have already been shown to be important through the use of sensitivity analysis. If the output has no sensitivity to the node, then the mode of the node is inconsequential.
- The sensitivity of 95% CL to a mode increases as the certainty in the value of probabilities increases. As discussed above, for beta distributions, certainty is a function of the number of samples expressed. A sample size of 2 will result in no sensitivity to mode with the sensitivity increasing as the number of samples increases.

Figure 7 illustrates these concepts. In this figure, a beta distribution's mode is doubled from 0.01 to 0.02 under 7A, a condition of high uncertainty (samples = 10) and 7B, a condition of low uncertainty (samples = 1,000). Under the condition of high uncertainty, a mode change has very little effect on the basic shape of the distribution. Under the condition of low uncertainty, the two distributions are much more distinct.

Figure 7 The effect of mode changes on high and low uncertainty beta distributions. (A) A beta distribution with a sample confidence of 10 has its mode doubled from 0.01 to 0.02. Note that the basic shape remains the same and random sampling from either distribution would be very similar. **(B)** A beta distribution with a sample confidence of 1,000 has its mode doubled from 0.01 to 0.02. Note that that the two distributions are quite distinct and random sampling from either distribution would also be quite distinct.

Mathematically, the sensitivity of 95% CL to Δmode is measured as:

$$S_{95\%\text{CL}_i} = \frac{\Delta 95\%\text{CL}}{\Delta \text{Mode}_i} \tag{6}$$

Where:

where $S_{95\%\text{CL}_i}$ = the sensitivity of the 95% confidence level of the model output of interest to a change in mode of node i, $\Delta 95\%$ CL = the change in the 95% confidence level due to a change in mode of node i, and ΔMode_i = the change in mode of node i.

The following process can be used to calculate $S_{95\%\text{CL}_i}$

- Calculate the baseline 95% CL by running a Monte Carlo analysis on the baseline model.
- Choose a node i.
- Increase the mode of node i by a delta value.
- Calculate the new 95% confidence level by running a Monte Carlo analysis.
- Calculate the sensitivity as per Equation 6.
- Repeat this process for each node i of interest.

Table 1 shows the results of this process on three nodes for an illustrative exemplar.

As shown in Table 1, the credibility of the mode value in the first two nodes listed in the table is very high as even a very large change in the mode value would have very little effect on the 95% CL POF. In fact, the mode value of node 'QA test finds debulking wrinkle if it exists' has absolutely no effect on the value of 95% CL POF due to having the maximum uncertainty in its value. It should be noted, however, that the uncertainty in the mode value has a large impact on the variance of the output, as will be discussed in more detail below. The final node, 'UltraSonic inspection finds wrinkle,' with a sensitivity of 0.01453, indicates that it would have been possible for the expert to significantly change the 95% CL POF by changing the mode. More specifically, the 95% CL changes by 1.45×10^{-5} for every 0.001 change in the mode. This means, for example, that if the mode was originally 0.008 and the expert lowered it to 0.001, the 95% CL would have been $7 \times 1.45 \times 10^{-5}$ higher or 1.5×10^{-4} instead of the reported 4.8×10^{-5}

Table 1 Sensitivity calculations of 95% CL to a change in mode for an illustrative exemplar

Node i	NumSamples	Old_mode	New_mode	Delta_mode	95POF	New95POF	Delta_POF	Sen
Wrinkle introduced during debulking	120	0.114	0.227	0.114	4.80×10^{-5}	8.07×10^{-5}	3.26×10^{-5}	0.00029
QA test finds debulking wrinkle if it exists	2	0.100	0.200	0.100	4.80×10^{-5}	4.80×10^{-5}	$0.00E + 00$	_0.00000_[a]
UltraSonic inspection finds wrinkle	1,000	0.001	0.002	0.001	4.80×10^{-5}	6.25×10^{-5}	1.45×10^{-5}	_0.01453_[b]

[a]Note that the sensitivity of 95% CL POF is zero for the mode of node 'QA test finds debulking wrinkle if it exists'. This is due to the mode having maximum uncertainty (samples = 2) as discussed previously; [b]the sensitivity of the mode of node 'UltraSonic inspection finds wrinkle' is fairly high at 0.01453. Thus, for every .001 change in mode, 95% CL POF changes by 1.45×10^{-5}. This effectively means that if the expert started with a mode of .008 and reduced it down to 0.001, a magnitude change in 95% CL POF would have occurred.

from Table 1. The consequence of this observation is that the expert should be required to provide documented proof of the 1,000 sample size or else he should be required to reduce his reported sample size.

Techniques to meet a target POF with a 95% confidence level

With these procedures in place, the expert may find that it is not possible to meet the target value of product reliability (i.e., a low enough probability of failure). What guidance can be provided to the expert to cost effectively increase reliability?

The goal is to raise the 'certain' reliability cost-effectively. The word 'certain' here is used to indicate that a reported low reliability may be due in part to a lack of process knowledge, while the other portion is due to variability in the manufacturing process coupled with a lack of suitable quality assurance tests. These ideas are captured by the following two types of uncertainty [9]:

- *Aleatory* variability is the natural randomness in a process. Aleatory uncertainty cannot be reduced thru data collection. For example, the knowledge of what number will turn up on a six-sided die. This type of uncertainty can be reduced through better process control and through quality assurance testing. In the die analogy, this is equivalent to reducing the number of sides on the die or weighting the die to come up favorably.
- *Epistemic* uncertainty is the scientific uncertainty in the model of the process. It is due to limited data and knowledge. This uncertainty can be reduced through more data collection, better expert knowledge, or through analytical means.

Reducing aleatory uncertainty through improved process control to lower randomness is application dependent and will not be discussed in this paper other than to note that the identification of the processes that drive uncertainty in the output is invaluable.

This section will discuss improving reported reliability by reducing epistemic uncertainty through targeted testing.

The most direct way to reduce uncertainty in the output (thus reducing 95% CL) is by reducing the variance in the output. Thus, the goal at this stage is to discover which nodes are most responsible for variance in the output. Once that is known, the focus should be on reducing the variance of those nodes. This may involve breaking a single mode into multiple subnodes to increase the level of detail of a particular process.

Saltelli et al. [4] have developed a technique to efficiently determine which variables in a probabilistic model contribute the most to variance in the output. This technique is called variance-based global sensitivity analysis and herein will also be referred to as the Saltelli method.

It is illuminating to compare point or derivative-based sensitivity analysis (previously discussed) and to which Equation 1 refers, with Saltelli global sensitivity analysis.

Conventional derivative (point)-based sensitivities

- Do not take into account uncertainty in the parameters.
- Do provide good information about a parameter at its most likely value.

Global sensitivity analysis (GSA) (the Saltelli method)

- Does take into account uncertainty in the parameters
- Is capable of determining which factors have a major effect on the variance of the POF calculation
- Is capable of determining which factors interact with others in an important way (synergistic effects)
- Is especially useful for determining the small subset of parameters that are important
- Is essentially a variance decomposition algorithm - it determines to some degree what portion of the output variance is due to variance in a particular parameter

The Saltelli process produces two sensitivity measures for each variable. S_i indicates the main effect of variable i, and S_{Ti} indicates the total effect of variable i. There are a few characteristics of these two types of sensitivities that are important to know. S_i indicates by how much one could reduce (on average) the output variance if variable i could be fixed. It is a measure of the main effect. S_{Ti} is useful in determining two important aspects of a variable. This first is if it has interactions with other variables. This can be measured by $(S_{Ti} - S_i)$. The second is if the variable is non-influential and can safely be ignored by setting it to a fixed value when performing time consuming analyses. This is indicated by $S_{Ti} = 0$.

Table 2 shows the results of a Saltelli global sensitivity analysis of the exemplar problem. The variables are sorted from greatest total effect to least total effect. For the example BN configuration, 8 out of 81 nodes have been identified as contributing significantly to variance in the BN output as shown by a total effect of over 0.013. Note that this algorithm tends to 'bottom out' at a non-zero number which in this case is approximately 0.013. Observe that all eight significant nodes are related to inducing or detecting wrinkles in the part.

Reducing epistemic uncertainty using confidence level shifting (CLS)

Now that the nodes causing variance in the output have been identified, the next step is to determine which, how much, and in what order testing should be done to most

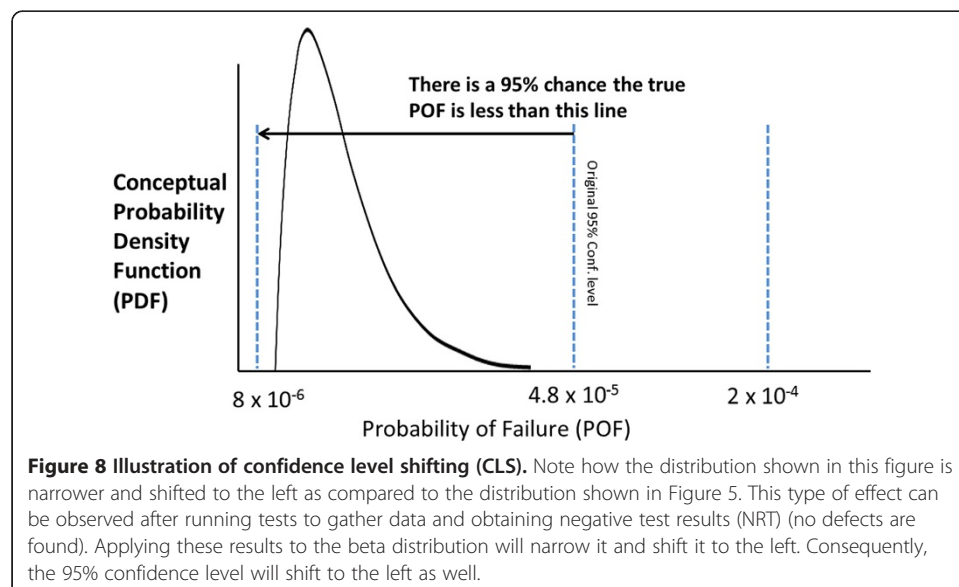**Table 2 Saltelli global sensitivity analysis of the exemplar problem**

Factor name	Hat_max_load_FE	Hat_max_load_TE	Mode	NumSamples
QA test finds debulking wrinkle if it exists	*0.407223*	*0.633979*	*0.010*	*20*
Wrinkle introduced during debulking	*0.210948*	*0.525119*	*0.010*	*120*
Wrinkle intro. during final cloth overwrap	*0.056164*	*0.187127*	*0.010*	*120*
QA test finds final cloth overwrap wrinkle	*0.033471*	*0.172301*	*0.020*	*120*
QA test finds bagging wrinkle	*0.024144*	*0.154588*	*0.010*	*120*
Wrinkle introduced during release film	*0.029566*	*0.152148*	*0.010*	*120*
QA test finds release film wrinkle	*0.025807*	*0.15214*	*0.010*	*120*
Wrinkle introduced during bagging	*0.033863*	*0.142558*	*0.010*	*120*
Radius thickening intro. during cloth overwrap	0.002573	0.126621	0.020	35
QA test finds debulking radius thickening	0.002591	0.12661	0.010	2

Hat_Max_Load_FE corresponds to S_i whereas Hat_Max_Load_TE corresponds to S_{Ti}. The variables are sorted from greatest total effect to least total effect. For the example model configuration, 8 out of 81 nodes have been identified as contributing significantly to variance in the model output as shown by the italicized entries. Note that all eight nodes are related to inducing or detecting wrinkles in the part. Mode refers to the mode of the beta distribution and NumSamples refers to the parameter 'k' in Equations 4 and 5.

effectively reduce 95% CL. This is known as 'targeted testing' using confidence level shifting (CLS). Confidence levels were explained in Figure 6. Examine Figure 8 in comparison to Figure 6 to understand the idea behind confidence level shifting. To get to Figure 8 from Figure 6, testing would take place to understand if a particular manufacturing step introduces a flaw or not. If not, the step was performed flawlessly and the parameter 'b' should be incremented as described above. A flawless test is also known as a negative result test or NRT. As NRTs accumulate, the beta distribution will narrow and shift to the left. Likewise, its associated 95% confidence level will also shift to the left. This is what is known as confidence level shifting or CLS.

To begin the CLS process, a Monte Carlo procedure will be run for the baseline network and the 95% confidence level of POF will be calculated before any NRTs have been applied to a beta distribution. Note that each beta distribution represents a probability (a factor). Next, a NRT is applied to a single factor and the Monte Carlo analysis is rerun and a new 95% confidence level of POF will be calculated. This provides enough information to calculate a Δ 95% confidence level for POF which is calculated as the original 95% confidence level for POF minus the newly calculated original 95% confidence level for POF. This term can be expressed more compactly as Δ 95% POF or even more simply as ΔPOF. If the cost of performing the test is known, another term, ΔPOF/$, can be defined which is the amount of change in 95% POF per dollar spent. The metric can be used to determine what data should be collected to most cost effectively drive the 95% POF value to the left.

With the previous information as background and referring to Figure 9, the CLS process can be explained. First determine ΔPOF/$ for each factor of the set. Add 1 to the 'b' value of the factor with the highest ΔPOF/$. If the target 95% POF has not been reached continue the process while keeping track of which factors received the 'b' increment and in what order. When the target 95% POF has been reached, the process provides a list of what data should be taken and in what order to most cost-effectively drive the 95% POF value to its target value.
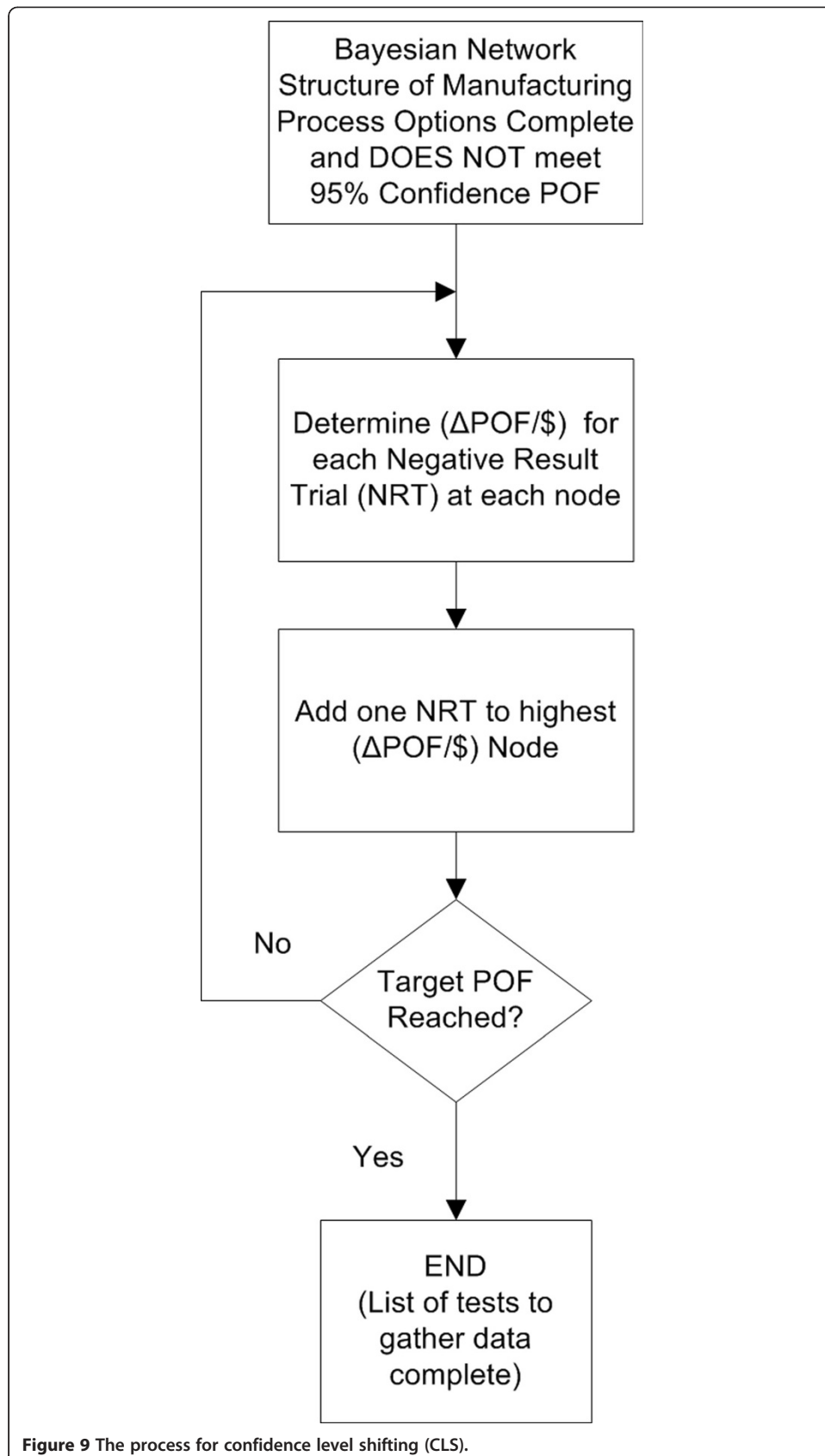

Figure 8 Illustration of confidence level shifting (CLS). Note how the distribution shown in this figure is narrower and shifted to the left as compared to the distribution shown in Figure 5. This type of effect can be observed after running tests to gather data and obtaining negative test results (NRT) (no defects are found). Applying these results to the beta distribution will narrow it and shift it to the left. Consequently, the 95% confidence level will shift to the left as well.

Figure 9 The process for confidence level shifting (CLS).

Note that a single complete fabrication of a part with appropriate inspection will simultaneously provide a data point for every node in the network. Partial part constructions can be accomplished to gather data for just the most important nodes. The ratio of decrease in POF to the cost of running a trial is the measure by which it is decided which trials to run. Note that a test that simultaneously provides data for multiple nodes is called a 'unitized test'. Unitized tests are a time- and cost-efficient technique for generating data to reduce epistemic uncertainty.

Figure 10 shows a plot of decreasing 95% CL POF as a function of (non-unitized) targeted testing. Note that in this example, 700 discrete targeted tests must be run to reduce the 95% CL probability of failure from 4.8×10^{-4} to 1×10^{-4}.

Table 3 is a detailed look at exactly what tests were performed and in what order. Table 3 shows the results of targeted testing analysis using confidence level shifting. Each row of this table provides a breakdown of how much data should be collected for each factor for a maximum reduction in 95% CL POF. For example of 75 collected data points, 25 data points should be collected to check for inducing wrinkles during the final cloth overwrap, and 50 QA tests should be performed to see if wrinkles can be detected during the debulking manufacturing step.

Unitized test for efficient collection of testing data

The example shown in Figure 10 and Table 3 shows the most efficient possible data collection to reduce epistemic uncertainty using individual tests. In this case, however, the burden of testing is high, requiring 700 individual tests to reach a target POF. One technique to reduce this burden is to create a unitized test structure that can test all eight significant features per test. As per Figure 11, only 95 of these unitized tests would have to be performed to reach the target 95% CL POF. This is a sevenfold reduction in the
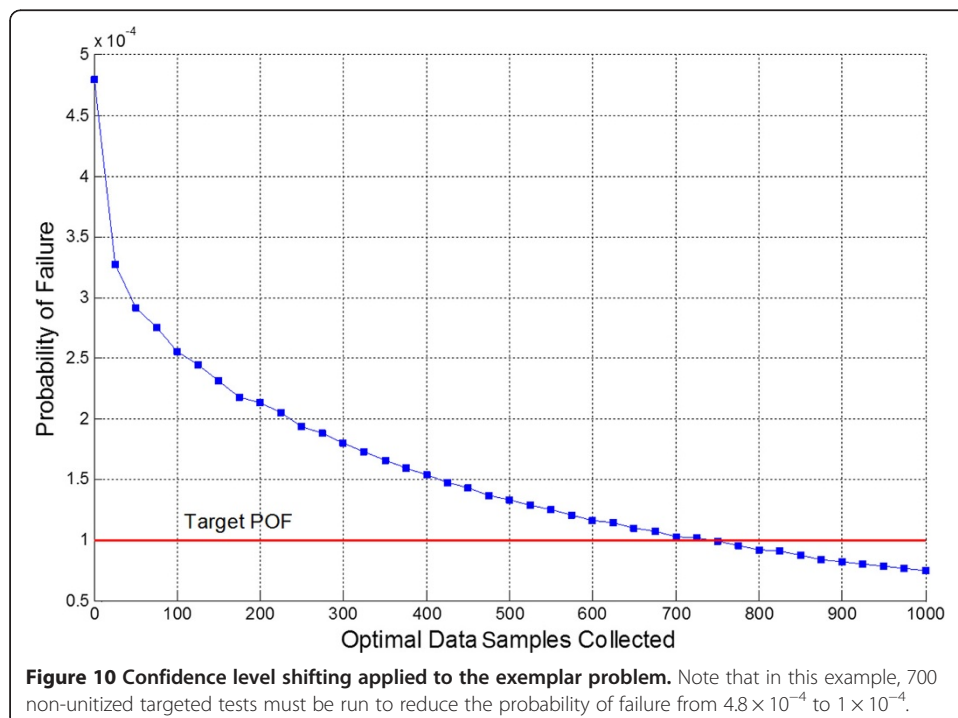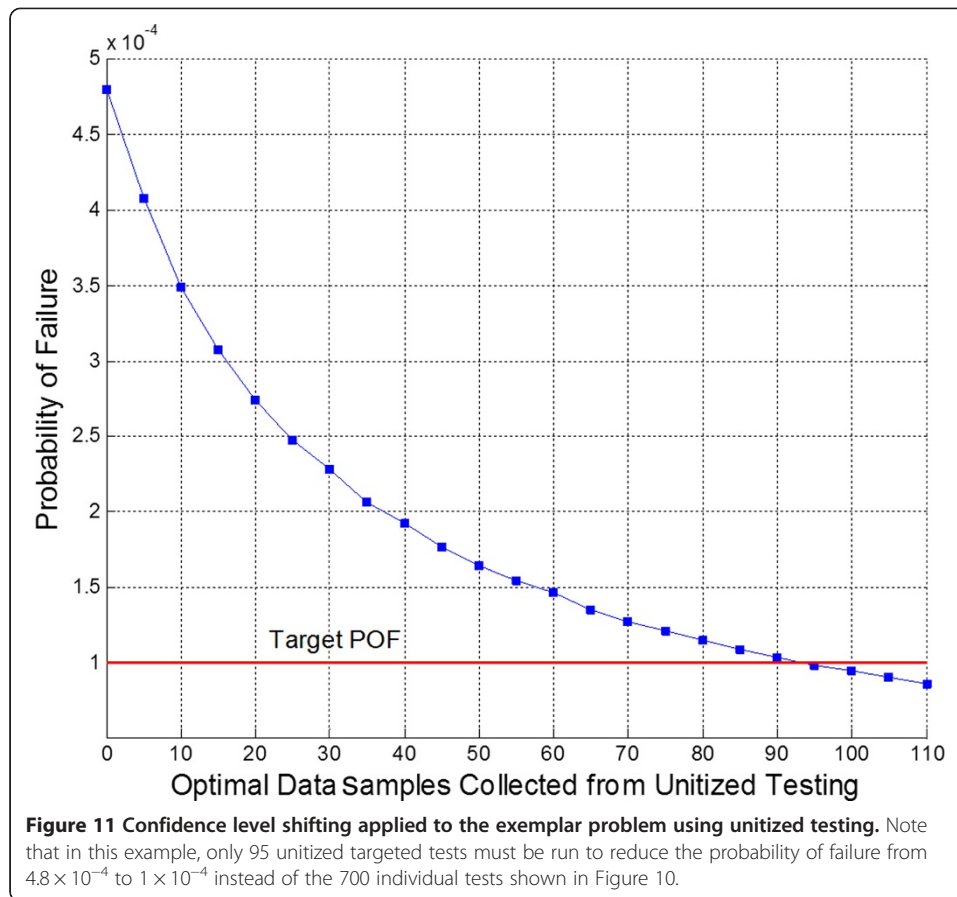

Figure 10 Confidence level shifting applied to the exemplar problem. Note that in this example, 700 non-unitized targeted tests must be run to reduce the probability of failure from 4.8×10^{-4} to 1×10^{-4}.

Table 3 Results of targeted testing analysis using confidence level shifting

Data	MS final cloth overwrap wrinkle	QA final cloth overwrap wrinkle	MS during debulking wrinkle	Q during debulking wrinkle	MS during bagging wrinkle	Q during bagging wrinkle	MS placing release film wrinkle	Q placing release film wrinkle	Lowest POF
0	0	0	0	0	0	0	0	0	4.80×10^{-4}
25	0	0	0	25	0	0	0	0	3.28×10^{-4}
50	0	0	0	50	0	0	0	0	2.92×10^{-4}
75	*0*	*25*	*0*	*50*	*0*	*0*	*0*	*0*	*$2.75 \times 10-4$*
100	0	25	0	75	0	0	0	0	2.56×10^{-4}
125	25	25	0	75	0	0	0	0	2.45×10^{-4}
150	25	25	0	75	0	25	0	0	2.31×10^{-4}
175	25	25	0	75	0	25	25	0	2.18×10^{-4}
200	25	25	0	75	0	25	25	25	2.13×10^{-4}
225	25	50	0	75	0	25	25	25	2.05×10^{-4}
250	25	50	0	100	0	25	25	25	1.94×10^{-4}
275	25	50	0	100	0	25	25	50	1.88×10^{-4}
300	25	75	0	100	0	25	25	50	1.80×10^{-4}

Each row of this table provides a breakdown of how much data should be collected for each factor for a maximum reduction in 95% CL POF. For the example of 75 collected data points (shown in italics), 25 data points should be collected to check for inducing wrinkles during the final cloth overwrap, and 50 QA tests should be performed to see if wrinkles can be detected during the debulking manufacturing step.

Figure 11 Confidence level shifting applied to the exemplar problem using unitized testing. Note that in this example, only 95 unitized targeted tests must be run to reduce the probability of failure from 4.8×10^{-4} to 1×10^{-4} instead of the 700 individual tests shown in Figure 10.
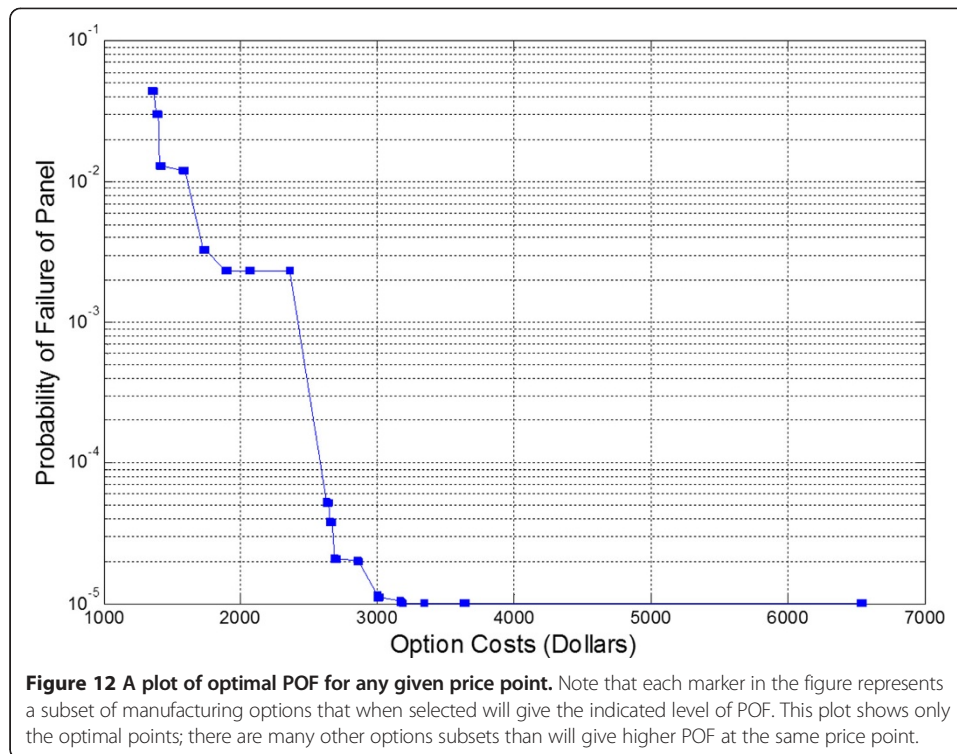
number of tests. Despite this large reduction in the number of required tests, in some cases, this will still be too expensive or time-consuming. A few points should be noted however. The first is that every problem will be different. For example, in some cases, the number of tests may be reduced from 70 down to 10. In addition, CLS is only one of the tools used to improve the 95% CL. As discussed above, reducing aleatory variability through the use of QA tests and better process control are alternative options that can be used in addition to or in place of CLS depending in the problem at hand.

Results and discussion

Putting it all together - an example using credibility tools

This section will provide an example of using the credibility tools discussed in this paper to reach a 95% confidence level probability of failure of 1×10^{-4} when starting with a manufacturing process for a three-hat-stiffened panel that has a 4.4% probability of failure under certain environmental conditions when no quality assurance testing is done.

As discussed in detail in reference [3], a Bayesian network is constructed for the three-hat-stiffened panel that includes all possible manufacturing options, including many potential quality assurance (QA) tests to catch defects both during the manufacturing process and as a final check. Using the network's point probabilities, it is possible to quickly evaluate all possible combinations of options to find the highest reliability part at any given price point or conversely the lowest probability of failure. Figure 12 is plot of optimal POF for any given price point. A few conclusions can be

Figure 12 A plot of optimal POF for any given price point. Note that each marker in the figure represents a subset of manufacturing options that when selected will give the indicated level of POF. This plot shows only the optimal points; there are many other options subsets than will give higher POF at the same price point.

drawn from Figure 12. This first is that the manufacturing process is highly dependent on QA checks for reliability. With no quality checks, the POF is .044 or 4.4% per part (the leftmost marker in the plot). The POF can range as low as 1.01×10^{-5} with a full set of QA checks in place. It should be noted that this plot does not include the price of rework or scrap due to faulty manufacturing. Future work will address this issue.

To reach the stated goal of 1×10^{-4} POF or 0.9999 reliability, the $2,626 option is the most cost-effective (not including scrap or rework). This option represents the case that all QA tests are off except for the ultrasonic inspection for wrinkle QA test.

The Bayesian network with this configuration is evaluated using Monte Carlo analysis to include the effects of uncertainty in the expert elicited opinions. The result is shown in Figure 13.

While the mode of this analysis meets the goal of 1×10^{-4} (being 5.6×10^{-5}), the 95% confidence value in POF is slightly too large at 2.3×10^{-4}. Another issue with using only the ultrasonic inspection for wrinkle QA test is that it only catches the wrinkle after the part is complete, leading to a very high rejection rate of a finished part. According to the model, there is a 46% chance of a wrinkle defect. This means that nearly half of the completed parts would have to be rejected. This is unacceptable. To better understand what is causing the wrinkles, an analysis of manufacturing steps as modeled by the Bayesian network is undertaken.

By removing all QA tests and running a Saltelli global sensitivity analysis, the manufacturing steps most responsible for output variation can be found.

Looking at Table 4, it is clear that wrinkles are the primary cause of increased failure and that there are four manufacturing steps that contribute to wrinkles. Given the high mode values (these are used as the point probabilities) it is also clear that the probability of incurring wrinkles during the manufacturing process is quite high.

Figure 13 Monte Carlo result of a three-hat panel network using only ultrasonic Inspection as QA.
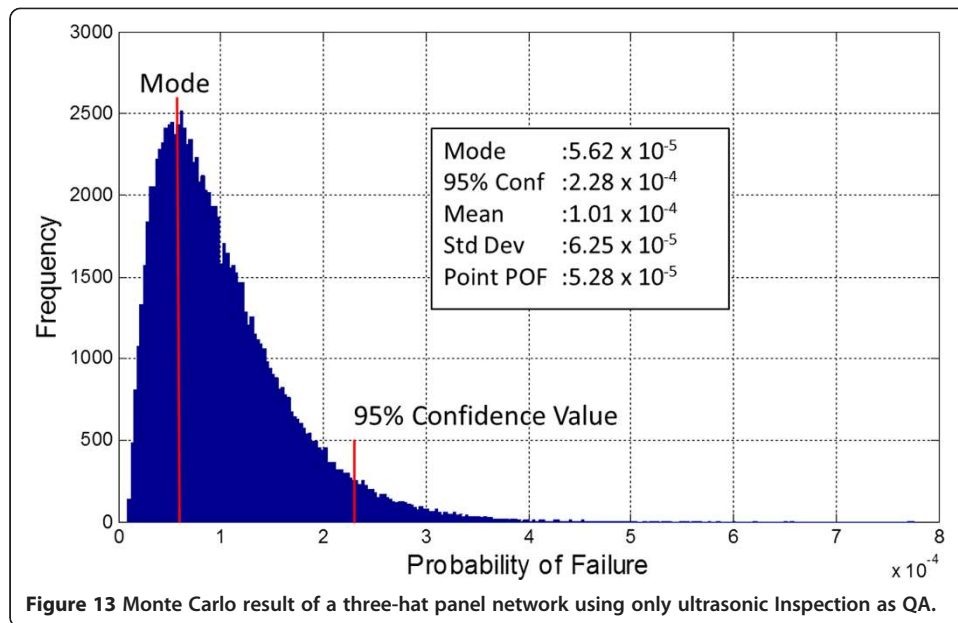
The relationship between p (wrinkle) and POF

Now that it is clear the wrinkles are the primary driver behind increased POF for this particular example, it is informative to note the relationship between the two quantities. Figure 14 shows a plot of POF vs. p (wrinkle). Note that by Figure 14, POF is linearly related to p (wrinkle) and POF is generally 0.096 of p (wrinkle). Thus, to get a POF of 1×10^{-4} or better, p (wrinkle) must be 1×10^{-3}. Occasionally, it is simpler to use p (wrinkle) as a proxy for POF, as this is directly displayed in the Bayesian network rather than needing to be calculated outside of it.

Examining the effect of four in-process QA checks

Observing the mode values of Table 4, the major issue in reaching a POF of 1×10^{-4} is not the uncertainty around the manufacturing process but the high mode value. The high mode value indicates that the manufacturing steps have a very high probability of inducing a wrinkle. Improving the actual manufacturing process reduces or eliminates rework or scrap and for that reason is normally considered the best way to improve quality and will be considered in the next section. The most expeditious way, however, to attack this problem without expending effort improving the manufacturing process is to apply quality assurance tests (QA tests) at the four manufacturing steps that can induce wrinkling in order to find and correct the wrinkles at the stage where they are introduced. The four tests, their costs, and their relative effectiveness are shown in Table 5. These tests are included in the Bayesian model to view their effect in p (wrinkle).

Table 4 Saltelli global sensitivity analysis results

Title	Hat_max_load_FE	Hat_max_load_TE	Mode	NumSamples
Wrinkle placing release film strips	0.435196	1.101102	0.2153	120
Wrinkle induced during debulking	0.192996	0.950746	0.1136	120
Wrinkle during bagging and pleating	0.713702	0.824299	0.2153	120
Wrinkle applying final cloth over wrap	0.039608	0.776457	0.0100	120

Top four variables contributing to variance identified. Hat_Max_Load_FE corresponds to S_i whereas Hat_Max_Load_TE corresponds to S_Ti. The variables are sorted from greatest total effect to least total effect. For the example model configuration, 4 out of 81 nodes have been identified as contributing significantly to variance in the model output. Note that all four nodes are related to inducing or detecting wrinkles in the part. Mode refers to the mode of the beta distribution and NumSamples refers to the parameter 'k' (expert confidence) in Equations 4 and 5.

Figure 14 Probability of failure as a function of probability of wrinkle.
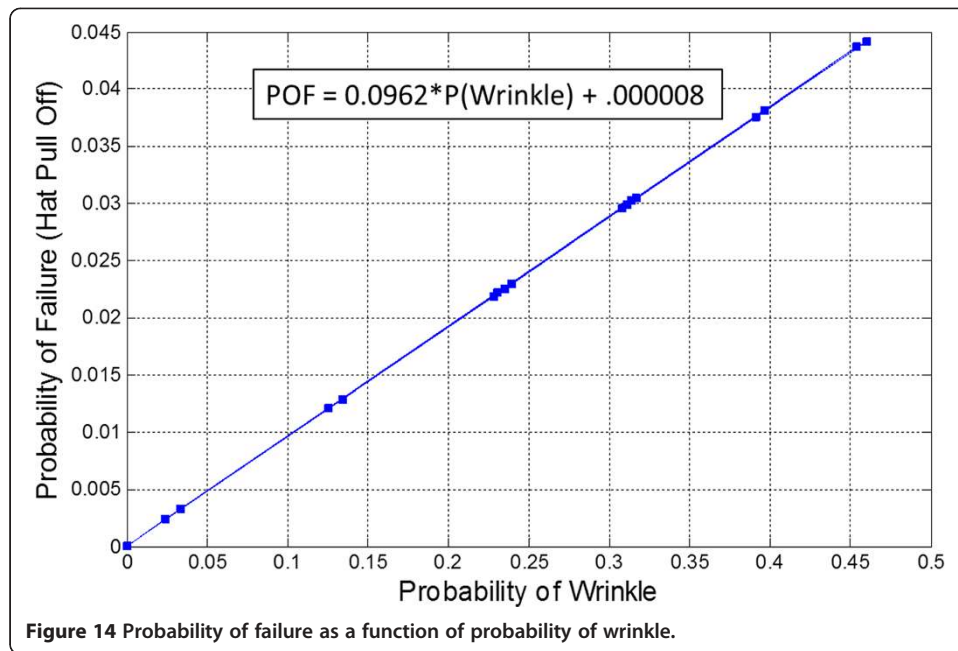
As Figure 15 shows, the p (wrinkles) is reduced from 46% to 2.43% by applying all four of these QA tests. Although that is a significant drop, p (wrinkle) must be lowered to 0.1% (1×10^{-3}) in order to reach the reliability target.

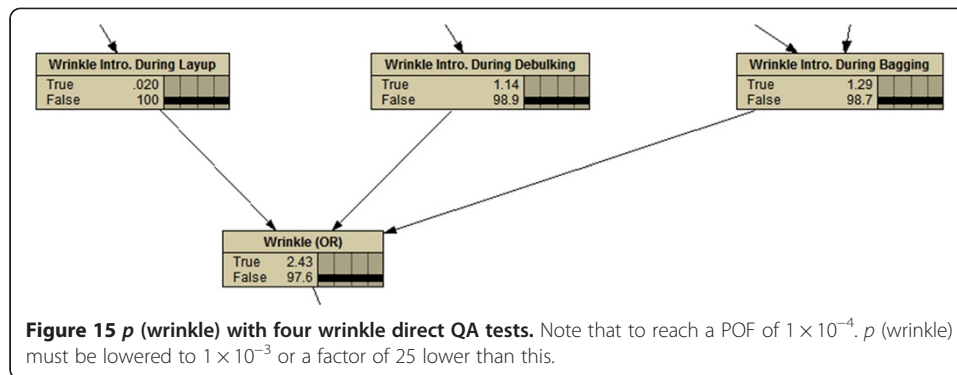### Identifying manufacturing process areas to target for improvements

Based on the point probability of a wrinkle being 0.0243 and needing to be 0.001, it is clear that process improvements must be made. Both global and point sensitivity analysis are performed on the four QA test model to determine the focus areas. Tables 6 and 7 show the results of this analysis.

Table 6 shows that most of the variance is due to the debulking and bagging steps and also that there is synergy between those nodes and other nodes. The point sensitivity analysis of Table 7 shows that those nodes are also prominent in affecting POF. Based on these two tables, it appears that the most efficient means of decreasing POF is by improving the manufacturing debulking related steps and the QA of the debulking simultaneously (top two globally sensitive) to take advantage of the synergy between them as well as their high point sensitivity. If this is not enough, then the nodes related to bagging and the release film should be worked on next.

To simulate working on the debulking process, it is assumed for the purposes of this example that engineers improve the process by examining and improving such elements as the bulk factor of the product form, tack or lack of tack, and the debulk process itself.

Table 5 Efficacy and cost of QA tests for wrinkles

Node	Mode	NumSamples	Cost
QA test finds bagging wrinkle	0.0400	120	$30
QA test finds debulking wrinkle	0.1000	2	$315
QA test finds release film wrinkle	0.0200	120	$30
QA test finds final cloth overwrap wrinkle	0.0200	120	$165

Figure 15 *p* (wrinkle) with four wrinkle direct QA tests. Note that to reach a POF of 1×10^{-4}. *p* (wrinkle) must be lowered to 1×10^{-3} or a factor of 25 lower than this.

After those improvements, there is only a .01 chance of inducing a wrinkle during the debulking process (improved from .11) and a .99 chance of finding a wrinkle at that point with a 20 sample size. With these improvements in place, the point probability network results are shown in Figure 16. This figure shows that the probability of wrinkle is still much higher than the 0.1% needed and that the major source of the high probability appears to be due to the steps making up bagging.

To help verify this conclusion, another global sensitivity analysis is performed with the results shown in Table 8. This table shows that the top four nodes contributing to output variance are now all related to bagging. Note that placing release film is part of the bagging process. Since Table 8 verifies what was seen in Figure 16, it is clear that improving the nodes related to inducing wrinkles during the bagging process would most benefit POF. This time, it is assumed that engineers improve the bagging process steps such that there is a 0.01 chance of wrinkle and improve the QA tests such that there is a 0.99 chance of detection.

As shown in Figure 17, these changes improve the point model probability of wrinkle to 0.05% which is better than the goal of 0.1%.

At this point, a Monte Carlo analysis must be run to take into account the uncertainty in the model parameters, and a 95% confidence level must be calculated to add the necessary amount of conservatism to the estimate. Figure 18 shows the results of the Monte Carlo process. Note that the mode of the POF distribution at 1.2×10^{-4} is close to the target value of 1×10^{-4}, but the 95% confidence level is 4.93×10^{-4}, which is about a factor of five from the target level.

At this stage, it may be cost-effective to perform targeted testing to reduce epistemic uncertainty by using the confidence level shifting (CLS) analysis technique. The first

Table 6 Global sensitivity analysis of the manufacturing network that includes the four wrinkle direct QA tests

Factor name	Hat_max_load_FE	Hat_max_load_TE	Mode	NumSamples
QA test finds debulking wrinkle if it exists	*0.817142*	*0.88832*	*0.1000*	*2*
Wrinkle introduced during debulking	*0.13929*	*0.279974*	*0.1136*	*120*
QA test finds bagging wrinkle	*0.003301*	*0.096139*	*0.0400*	*120*
QA test finds release film wrinkle	0.00201	0.087771	0.0200	120
Wrinkle introduced during bagging	0.000163	0.080021	0.2153	120
Wrinkle introduced during release film	−0.003777	0.082574	0.2153	120
QA test finds final cloth overwrap wrinkle	−0.005413	0.082446	0.0200	120
Wrinkle intro. during final cloth overwrap	−0.005268	0.082371	0.0100	120

The italicized entries indicate the significant variables identified through the sensitivity analysis.

Table 7 Point sensitivity analysis of the manufacturing network that includes the four wrinkle direct QA tests

Node	POFHat_max_load_sens	Mode	NumSamples
QA test finds bagging wrinkle	*0.02038*	*0.040*	*120*
Wrinkle introduced during release film	*0.02029*	*0.020*	*120*
QA test finds debulking wrinkle if t exists	*0.01079*	*0.100*	*2*
Wrinkle introduced during debulking	*0.00949*	*0.114*	*120*
Wrinkle introduced during bagging	*0.00379*	*0.215*	*120*
Wrinkle introduced during release film	0.00189	0.215	120
Wrinkle intro. during final cloth overwrap	0.00188	0.010	120
QA test finds final cloth overwrap wrinkle	0.00094	0.020	120

The italicized entries indicate the significant variables identified through the sensitivity analysis.

step in the CLS process is to identify which nodes are causing the most variance on the output POF using global sensitivity analysis. Table 2 from the main body of this paper shows the results of this. With the steps and QA tests related to wrinkles balanced in terms of performance by engineers, all eight nodes related to wrinkles are found to be important. Figure 9 from this paper's CLS section shows that 95 successful unitized targeted tests can be run to drive the 95% CL POF to 1×10^{-4}.

Now that the manufacturing process has been modified and additional testing performed to drive the 95% CL POF down to 1×10^{-4}, the final step is to perform a sensitivity analysis of the 95% confidence level POF to a change in mode, which is a good indication of influential distributions that must be justified by documentation. Table 9 shows the results of this analysis. The table is sorted by node with those nodes that have the most influence starting at the top. The top eight influential nodes turn out to be the eight nodes that had 95 unitized tests performed to make their estimates more certain. Due to the collection of this extra confirmatory data, these nodes can be considered credible. The ninth node in the list - radial thickening (RT) during debulking (not shown in Figure 3) has a sensitivity of 0.28×10^{-5} per 0.01 of mode change. This means that the given mode (0.01) could be as high as .04, or four times higher than the given mode, before the 95% CL POF reaches the target value. This value is thus judged to be credible.

Summary of using credibility tools to reach a credible target 95% CL POF

In summary, an example project consisting of manufacturing a three-hat-stiffened panel was used as a case study for exercising the credibility tools detailed in this paper. The goal of the example was to analyze a manufacturing process in terms of the factors that
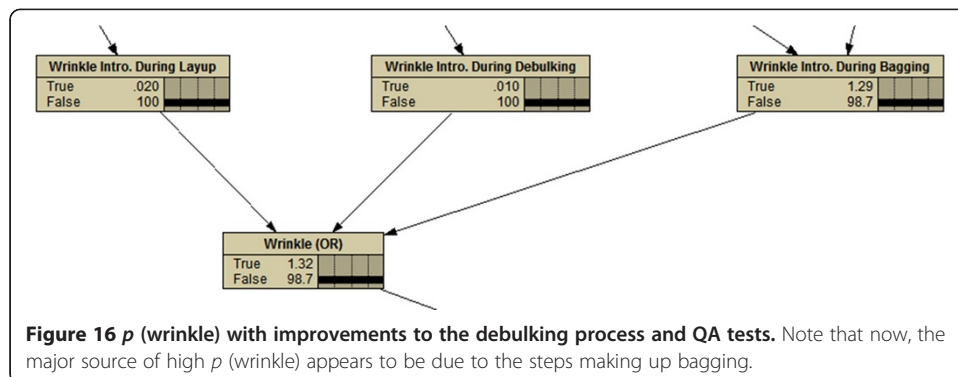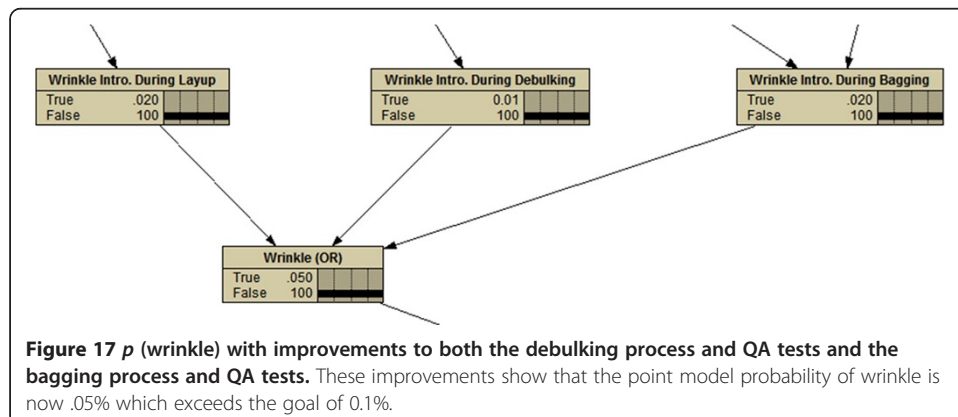

Figure 16 p (wrinkle) with improvements to the debulking process and QA tests. Note that now, the major source of high p (wrinkle) appears to be due to the steps making up bagging.

Table 8 Global sensitivity analysis of the manufacturing network that includes four wrinkle direct QA tests and improved debulking steps
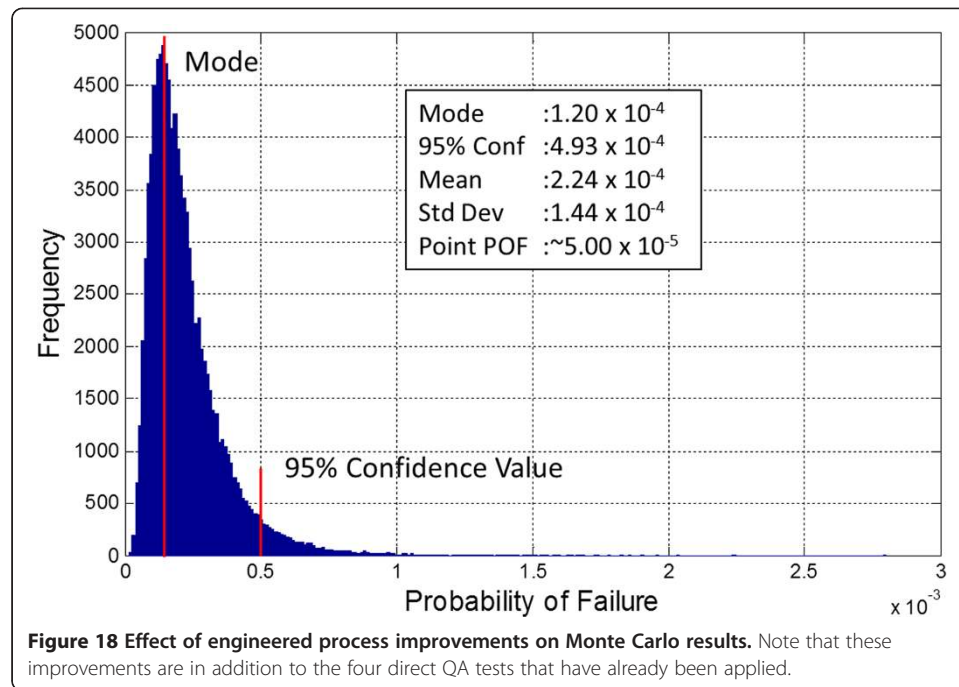
Factor name	Hat_max_load_FE	Hat_max_load_TE	Mode	NumSamples
QA test finds bagging wrinkle	*0.488152*	*0.625328*	*0.0400*	*120*
QA test finds release film wrinkle	*0.298514*	*0.388214*	*0.0200*	*120*
Wrinkle introduced during bagging	*0.135016*	*0.145427*	*0.2153*	*120*
Wrinkle introduced during release film	*0.039348*	*0.123047*	*0.2153*	*120*
Wrinkle introduced during debulking	0.00898	0.122452	0.0100	120
QA test finds debulking wrinkle if it exists	0.036094	0.115424	0.0100	20
Wrinkle intro. during final cloth overwrap	0.006912	0.096334	0.0100	120
QA test finds final cloth overwrap wrinkle	0.003504	0.095578	0.0200	120
Radius thickening intro. during final cloth overwarp	0.003001	0.093529	0.0200	35

This table shows that the top four nodes (in italics) contributing to output variance are now all related to bagging. Note that placing release film is part of the bagging process.

contribute to its unreliability, to ensure that the expert opinion using to furnish the parameters of that model were credible, and then use a number of tools to determine the optimal way to create a more reliable product that met a target reliability number. Note that this was accomplished conceptually for illustrative purposes.

The examination started with noting the effect of any and all possible combinations of manufacturing options on the point reliability of the part. It was found that quality assurance tests had an extremely high impact on part reliability. After noting that a post-manufacturing QA test was effective but that it resulted in many costly part rejections, another analysis was undertaken which found four manufacturing steps that induced wrinkles also contributed the most to variance in the output. QA tests were directly applied within the model to address these four steps but it was found that they were not effective enough to reach the target 95% CL POF. An iterative process was then undertaken which involved identifying and then improving aspects of the manufacturing process until the process nearly reached the target POF. At this point, an effort to reduce epistemic uncertainty in the model was undertaken using confidence level shifting to identify target testing. This testing optimally reduced uncertainty in the model. Finally, a sensitivity analysis of the 95% confidence level POF to a change in mode was performed for each node of the model to indicate influential distributions that must be justified by

Figure 17 *p* (wrinkle) with improvements to both the debulking process and QA tests and the bagging process and QA tests. These improvements show that the point model probability of wrinkle is now .05% which exceeds the goal of 0.1%.

Figure 18 Effect of engineered process improvements on Monte Carlo results. Note that these improvements are in addition to the four direct QA tests that have already been applied.

documentation. At this point, it was found that with the targeted testing that had already occurred that the model was credible.

Conclusions

This paper details an approach to obtaining credible model output when model parameters are based on expert opinion. Although the model used as an example in this paper is a Bayesian network model, the approach and techniques described in this paper are completely transferrable to any model using uncertain parameters. This paper details an approach to obtaining credible model output based on the idea of having a hypothetical expert whose unconscious bias influences the model output and discovering and using countermeasures to find and prevent these biases. Countermeasures include replacing point probabilities with beta distributions to incorporate uncertainty and requiring 95% confidence levels to add conservatism. Multiple types of sensitivity analyses are used to identify parameters in the model that have the most influence over the model's output. This includes a derivative point probability-based sensitivity analysis that is a good indicator of relevance when all parameters are at their most likely values, a sensitivity analysis of 95% confidence level to a change in mode which is a good indication of influential distributions that must be justified by documentation and a variance-based global sensitivity analysis which is useful for identifying which model parameters contribute the most to output variance and which model parameters have synergy with other model parameters. Finally, this paper uses a new technique named 'confidence level shifting' to cost and time optimally reduce epistemic uncertainty in the model. This is useful when uncertainty in model parameters is inflating the 95% confidence level of a reported target output (such as probability of failure or probability of a defect) and needs to be brought down as cost effectively as possible.

Table 9 Results of sensitivity calculations of 95% CL to a change in mode

Nodes	numSamples	Old_mode	New_mode	Delta_mode	95POF	New95POF	Delta_POF	Sen
Wrinkle introduced during debulking	*215*	*0.0055*	*0.0155*	*0.0100*	9.26×10^{-5}	1.21×10^{-4}	2.81×10^{-5}	2.81×10^{-3}
QA test finds debulking wrinkle if it exists	*115*	*0.0055*	*0.0155*	*0.0100*	9.26×10^{-5}	1.14×10^{-4}	2.11×10^{-5}	$2.11 \times 10-3$
QA test finds bagging wrinkle	*215*	*0.0055*	*0.0155*	*0.0100*	9.26×10^{-5}	1.11×10^{-4}	1.85×10^{-5}	1.85×10^{-3}
QA test finds final cloth overwrap wrinkle	*215*	*0.0032*	*0.0132*	*0.0100*	9.26×10^{-5}	1.11×10^{-4}	1.81×10^{-5}	1.81×10^{-3}
Wrinkle introduced during bagging	*215*	*0.0055*	*0.0155*	*0.0100*	9.26×10^{-5}	1.10×10^{-4}	1.75×10^{-5}	1.75×10^{-3}
QA test finds release film wrinkle	*215*	*0.0055*	*0.0155*	*0.0100*	9.26×10^{-5}	1.10×10^{-4}	1.75×10^{-5}	1.75×10^{-3}
Wrinkle introduced during release film	*215*	*0.0055*	*0.0155*	*0.0100*	9.26×10^{-5}	1.09×10^{-4}	1.60×10^{-5}	1.60×10^{-3}
Wrinkle intro. during final cloth overwrap	*215*	*0.0055*	*0.0155*	*0.0100*	9.26×10^{-5}	1.06×10^{-4}	1.37×10^{-5}	1.37×10^{-3}
Radius thickening intro. during debulking	30	0.0100	0.0200	0.0100	9.26×10^{-5}	9.55×10^{-5}	2.88×10^{-6}	2.88×10^{-4}
QA test finds release film up. radius thickening	10	0.0100	0.0200	0.0100	9.26×10^{-5}	9.51×10^{-5}	2.50×10^{-6}	2.50×10^{-4}
QA test finds bagging radius thickening	20	0.0050	0.0150	0.0100	9.26×10^{-5}	9.45×10^{-5}	1.93×10^{-6}	1.93×10^{-4}

The top eight nodes are italicized. The sensitivity of the mode of for an average one of these nodes is roughly 2×10^{-3} or 2×10^{-5} for every .01 change in mode. This effectively means that if the expert started with a mode of .008 and reduced it down to 0.001, a magnitude change in 95% CL POF would have occurred.

Abbreviations

Δ: delta, change in value; A: beta distribution parameter expressing the number of flawed examples; B: beta distribution parameter expressing the number of flawless examples; BN: Bayesian networks; CL: confidence level; CLS: confidence level shifting; CMTI: Certification Methodology to Transition Innovation; DARPA/DSO: Defense Advanced Research Projects Agency Defense Science Office; GSA: global sensitivity analysis; K: expert confidence in estimate in terms of equivalent prior sample size; Mode: the most likely probability of a flaw; NRT: negative result test; OM: Open Manufacturing; P: proportion of flaws in the beta distribution; P_i: probability of node i inducing or failing to detect a defect; POF: probability of failure; QA: quality assurance; RT: radial thickening; $S_{95\%CLi}$: sensitivity of the 95% CL of model output due to change in mode of node i; $S_D i$: derivative-based sensitivity measure; Si: effect due to variable i; S_{Ti}: total effect due to variable i; X_i: model parameter; Y: model output.

Competing interests

The author declares that he has no competing interests.

References

1. Renieri G (2013) High performance stiffened panel design concept. Annual Report. Office of Naval Research (ONR), Arlington, VA. Contract N00014-12-C0417
2. Koller D, Friedman N (2009) Probabilistic graphical models: principles and techniques. Press, MIT, Cambridge
3. Hahn GL, Pado LE, Thomas MJ, Liguore SL (2014) Application of risk quantification approach to aerospace manufacturing using Bayesian networks. Paper presented at AIAA SciTech 2014. National Harbor, Maryland, pp 13–17. Report AIAA-2014-0463
4. Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S (2008) Global sensitivity analysis. The Primer, Wiley, John & Sons, Incorporated.
5. Kruschke J (2011) Doing Bayesian data analysis. Press, Academic, New York
6. Kruschke J (2012) Beta distribution parameterized by mode instead of mean. http://doingbayesiandataanalysis.blogspot.com/2012/06/beta-distribution-parameterized-by-mode.html
7. Haldar A, Mahadevan S (2000) Probability, reliability, and statistical methods in engineering design. Wiley, John & Sons, Incorporated, New York
8. Milton JS, Arnold JC (1986) Probability and statistics in the engineering and computing sciences. Elsevier Academic Press, Burlington, MA
9. Abrahamson N (2006) Seismic hazard assessment: problems with current practice and future developments Proceedings. First European Conference on Earthquake Engineering and Seismology, Geneva, Switzerland. September 2006