Propagating Skewness and Kurtosis Through Engineering Models for Low-Cost, Meaningful, Nondeterministic Design

Travis V. Anderson
Graduate Research Assistant
e-mail: travisanderson@byu.edu

Christopher A. Mattson[1]
Assistant Professor
e-mail: mattson@byu.edu

Design Exploration Research Group,
Department of Mechanical Engineering,
Brigham Young University,
Provo, UT 84602

System models help designers predict actual system output. Generally, variation in system inputs creates variation in system outputs. Designers often propagate variance through a system model by taking a derivative-based weighted sum of each input's variance. This method is based on a Taylor-series expansion. Having an output mean and variance, designers typically assume the outputs are Gaussian. This paper demonstrates that outputs are rarely Gaussian for nonlinear functions, even with Gaussian inputs. This paper also presents a solution for system designers to more meaningfully describe the system output distribution. This solution consists of using equations derived from a second-order Taylor series that propagate skewness and kurtosis through a system model. If a second-order Taylor series is used to propagate variance, these higher-order statistics can also be propagated with minimal additional computational cost. These higher-order statistics allow the system designer to more accurately describe the distribution of possible outputs. The benefits of including higher-order statistics in error propagation are clearly illustrated in the example of a flat-rolling metalworking process used to manufacture metal plates. [DOI: 10.1115/1.4007389]

1 Introduction

A system model uses known system inputs to predict system outputs. Almost always, variation in system inputs is present, which also produces variation in system outputs. The system designer is interested in whether or not a system will accomplish the design objectives even in the presence of this variation. Consequently, a system model that takes a known input distribution and produces an output distribution may be more helpful than a deterministic model [1].

A statistical error distribution is often obtained by propagating variance from system inputs to system outputs [2] using Eq. (1), where the partial derivatives are evaluated at the input mean, $x = \bar{x}$. This equation is based on a first-order Taylor-series approximation expanded about the input mean, $\bar{x}$. Its derivation and a more detailed discussion is presented later in this paper

$$\sigma_y^2 \approx \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma_{x_i}^2 \quad (1)$$

It is interesting to note that this formula predicts an output variance σ_y^2 only. Since all higher-order statistics (e.g., skewness, kurtosis, etc.) are ignored, outputs are usually assumed to be Gaussian. This assumption is often wrong and does not accurately reflect reality.

To illustrate this point, consider the simple quadratic function, $y = x^2$. Assume the input x is a Gaussian distribution with a mean $\bar{x}$ and a standard deviation σ_x both equal to 1. Equation (1) can be used to propagate this input distribution through the system model and predict the Gaussian output distribution shown in Fig. 1(*a*). This predicted output is very different from the actual system output distribution, shown in Fig. 1(*c*). However, if skewness and kurtosis are also propagated through the system model, the predicted output distribution (shown in Fig. 1(*b*)) resembles actual system output much more closely [3].

This paper discusses statistical error propagation through engineering models. The assumption of Gaussian outputs is dismissed and a method to more accurately describe an output distribution is presented. This method relies on second-order Taylor series to propagate higher-order statistics, such as skewness and kurtosis, in addition to a mean and variance. By way of example, this method is applied to a flat-rolling metalworking process used to manufacture metal plates.

2 Error Propagation Techniques

Many methods are currently in use or being researched that can propagate error through a system, including (1) nondeterministic analysis via brute force (such as Monte Carlo (MC)), (2) univariate dimension reduction, (3) deterministic model composition, (4) error budgets, (5) interval analysis, (6) Bayesian inference, (7) anti-optimizations, and (8) error propagation via Taylor-series expansion. A brief review of these methods and their benefits and drawbacks is discussed below. As indicated, some of these methods have significant computational cost, some require independent variables, some have complex implementations, and some are unable to predict an entire output distribution. Sections 3–5 of this paper present a Taylor-series-based method to overcome these grievances with minimal or no loss in accuracy.

2.1 Monte Carlo and Sampling Techniques. Due to the complexities of nondeterministic modeling, most nondeterministic error analysis techniques represent uncertainty with probabilities, which are then propagated through a deterministic model [4]. This is commonly achieved with a brute-force or sampling approach

[1]Corresponding author.

Contributed by the Design Automation Committee of ASME for publication in the JOURNAL OF MECHANICAL DESIGN. Manuscript received July 29, 2011; final manuscript received March 19, 2012; published online September 21, 2012. Assoc. Editor: Irem Y. Tumer.

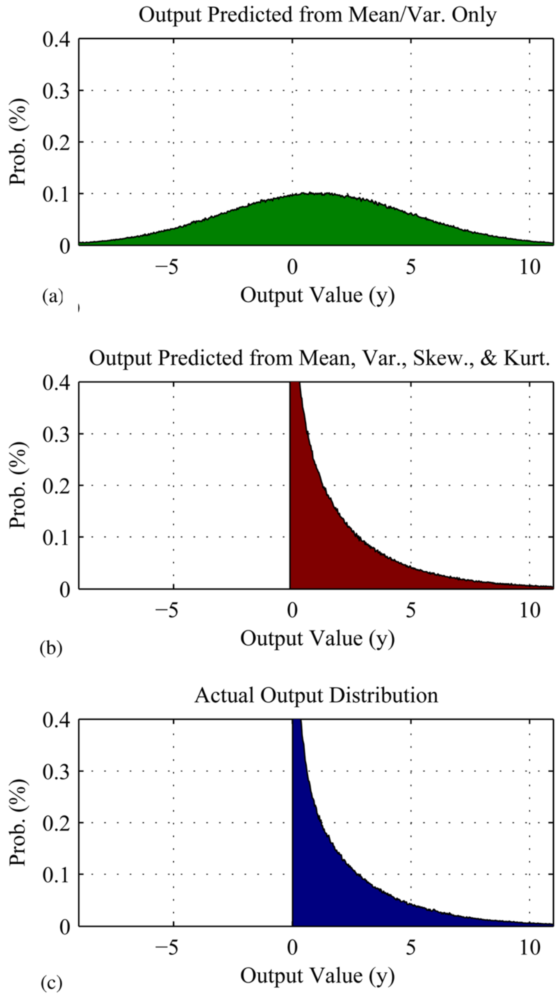

Fig. 1 Predicted output distributions obtained from propagating (*a*) mean and variance only, and (*b*) mean, variance, skewness, and kurtosis. Actual system output distribution is shown in (*c*).

that uses Monte Carlo, quasi Monte Carlo [5,6], Latin hypercube, Latin supercube [7], a hybrid [8], or some other technique. These approaches do not need to assume a Gaussian output distribution. Consequently, an estimate of the fully described output distribution can be obtained.

These methods can require a higher computational cost than some of the other techniques discussed below, and the entire simulation must be executed again each time the model or any input value changes. This can be prohibitive in an iterative design process.

2.2 Univariate Dimension Reduction. The goal of univariate dimension reduction is primarily to reduce the complexities of dimension explosion by reducing multivariate problems into a set of univariate problems. Collectively, a set of univariate statistical integrals is much easier to solve than a single multivariate statistical integral. In some situations, data analysis may even be more accurate in the reduced space than in the original space, such as with regressions [9].

Univariate dimension reduction is capable of predicting an entire output distribution. However, even the univariate expectation integrals can be difficult to solve and all inputs must be independent of each other. When inputs are correlated, they can be transformed into independent variables with a Rosenblatt transformation [10], which adds an additional level of complexity to the model.

2.3 Deterministic Error by Model Composition. Uncertainty can also be propagated deterministically through a compositional system model [11]. This technique produces max/min error bounds by augmenting the system model with component error models and comparing the resulting output with the original system model's output [12]. Errors do not need to be independent and component models do not have to be mathematical or closed-form functions.

However, deterministic error analysis requires known error models for every component, and the max/min error bounds obtained from this method are often much too large to offer practical assistance to the system designer. No information is obtained regarding the statistical probabilities of outputs within the max/min error envelop [1].

2.4 Error Budgets. The method of error budgets involves propagating the error of each component through the system separately, and resolving each component's error to the contribution it makes on the total system error [13,14]. This is done by perturbing one error source at a time and observing the effect this has on the total system error. Consequently, this method requires either that component errors be independent or that a separate model showing component error interactions be developed, which typically is not done [15]. If the error sources are not actually independent, this method will not necessarily describe the full range of possible model error.

2.5 Interval Analysis Methods. Interval analysis methods bound rounding and measurement errors in mathematical computation. Arithmetic can then be performed using intervals instead of a single nominal value [16]. These techniques can be used to propagate error envelopes, or intervals, through a system model. These methods, however, are typically limited to the basic operations of addition, subtraction, multiplication, and division.

2.6 Bayesian Inference. Bayesian inference is a method of statistical inference whereby the probability that a hypothesis is true is inferred based on both observed evidence and the prior probability that the hypothesis was true [17]. It combines common-sense knowledge with observational evidence in an attempt to eliminate needless complexity in a model by declaring only meaningful relationships [18] and disregarding the influences of all other variables on system outputs.

2.7 Anti-Optimization Techniques. Anti-optimization techniques allow the designer to find the worst-case scenario for a given problem. This results in a two-level optimization problem, where the uncertainty is anti-optimized on the lower level and the overall design is optimized on a higher level [19].

2.8 Taylor-Series and Central Moments. A derivatives-based technique can be much simpler to implement than most of the methods previously discussed, does not require independent inputs, is efficient in achieving high levels of accuracy, and can predict an entire output distribution. While lower-order statistical error propagation via Taylor-series expansion is common practice, system designers typically only propagate variance and do not consider the higher-order statistics of skewness and kurtosis [20–22].

This paper shows that these higher-order statistics can be easily propagated along with variance using a Taylor series. This generates a significantly more accurate and fully described output distribution with little additional effort or cost. This technique is easy to implement and works well with correlated variables.

These higher-order statistics are determined by propagating central moments. Central moments are commonly used in statistical analysis [21,23], particularly in the field of tolerance analysis. The kth central moment is given by Eq. (2)

$$
\begin{aligned}
\mu_k &= \int_{-\infty}^{\infty} (x-\bar{x})^k f(x)dx \\
&= E\left[(x-\bar{x})^k\right] \\
&= \frac{1}{N}\sum_{j=1}^{N}(x_j-\bar{x})^k
\end{aligned}
\tag{2}
$$

where x represents some distribution of N values, $\bar{x}$ represents the input mean, and E is the expectation operator. The central moments of a population can easily be estimated using any appropriate population sampling technique.

The zeroth central moment is always equal to one, the first central moment is always equal to zero [24], and the second central moment is equivalent to the variance. The third and fourth moments are used in the calculation of the higher-order statistics skewness and kurtosis. This paper presents a discussion of these higher-order statistics and addresses the benefits, limitations, and underlying assumptions associated with using a Taylor-series and central moments to propagate higher-order statistics.

3 Propagation of Variance

This section derives the first- and second-order Taylor-series formulas typically cited in literature for variance propagation.

3.1 Derivation of Variance Propagation Formula Using First-Order Taylor Series. While other sources of error exist (such as unmodeled behavior, emergent behavior, and measurement error), this paper focuses only on the variation in system outputs caused by variation in system inputs.

Let y be some function of n inputs x, where each input x_i is some distribution of possible values. The first-order Taylor-series approximation expanded about the input means $\bar{x}$ is shown in Eq. (3)

$$
y \approx f(\bar{x}_1,.,\bar{x}_n) + \sum_{i=1}^{n}\frac{\partial f}{\partial x_i}(x_i-\bar{x}_i)
\tag{3}
$$

where the partial derivatives are evaluated at the mean $x_i = \bar{x}_i$. An approximation of the output mean $\bar{y}$ is given in Eq. (4)

$$
\begin{aligned}
\bar{y} &= E[y] \\
&\approx E\left[f(\bar{x}_1,.,\bar{x}_n) + \sum_{i=1}^{n}\frac{\partial f}{\partial x_i}(x_i-\bar{x}_i)\right] \\
&\approx f(\bar{x}_1,.,\bar{x}_n) + \sum_{i=1}^{n}\frac{\partial f}{\partial x_i}\mu_{1,i} \\
&\approx f(\bar{x}_1,.,\bar{x}_n)
\end{aligned}
\tag{4}
$$

where E is the expectation operator and $\mu_{k,i}$ is the kth central moment for the ith input, as given previously in Eq. (2). (Recall that the first central moment is equal to zero.) Subtracting Eq. (3) from Eqs. (4) produces (5)

$$
y-\bar{y} \approx \sum_{i=1}^{n}\frac{\partial f}{\partial x_i}(x_i-\bar{x}_i)
\tag{5}
$$

Squaring and taking the expectation of Eqs. (5) produces (6)

$$
\sigma_y^2 \approx \sum_{i=1}^{n}\left[\left(\frac{\partial f}{\partial x_i}\right)^2\sigma_{x_i}^2 + 2\sum_{j=i+1}^{n}\frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}\sigma_{x_i x_j}^2\right]
\tag{6}
$$

where σ_y^2 and σ_x^2 are the variances in y and x, respectively. Recall that variance σ^2 is the second central moment, which is given by Eq. (7)

$$
\begin{aligned}
\sigma_x^2 &= \mu_2 \\
&= \int_{-\infty}^{\infty}(x-\bar{x})^2 f(x)dx \\
&= E[(x-\bar{x})^2] \\
&= \frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2
\end{aligned}
\tag{7}
$$

The second term in Eq. (6) is the covariance term, where $\sigma_{x_i x_j}^2$ is the covariance between inputs x_i and x_j. Covariance is defined in Eq. (8)

$$
\sigma_{x_i x_j}^2 = E\left[(x_i-\bar{x}_i)(x_j-\bar{x}_j)\right]
\tag{8}
$$

When inputs are independent, the covariance term is equal to zero, and Eqs. (6) reduces to (9)

$$
\sigma_y^2 \approx \sum_{i=1}^{n}\left(\frac{\partial f}{\partial x_i}\right)^2\sigma_{x_i}^2
\tag{9}
$$

This simplifying assumption of independence is typically made, both in literature and in practice. Consequently, Eq. (9) is the formula typically given for statistical error propagation through an engineering model [22,25–27]. This is identical to Eq. (1), which was presented in the Introduction of this paper, and consequently carries with it the same limitations discussed therein.

3.2 Derivation of Variance Propagation Formula Using Second-Order Taylor Series. As shown in the preceding derivation, Eq. (9) is based on a first-order Taylor series. For nonlinear and higher-order polynomial functions, Taylor-series truncation error becomes significant and Eq. (9) can become extremely inaccurate (i.e., wrong by one or more orders of magnitude [22]).

In situations where increased accuracy is required, a second-order Taylor series is sometimes used to propagate statistical error. For the sake of brevity, the second-order derivation is presented below for a monovariable function, $y = f(x)$. Extending this derivation to multivariate functions is trivial, as it follows the same derivation steps.

The second-order Taylor series taken about the input mean $\bar{x}$ is given in Eq. (10), where the partial derivatives are again evaluated at the mean, $x = \bar{x}$.

$$
y \approx f(\bar{x}) + \frac{df}{dx}(x-\bar{x}) + \frac{1}{2}\frac{d^2f}{dx^2}(x-\bar{x})^2
\tag{10}
$$

The second-order approximation of the output mean $\bar{y}$ is given in Eq. (11).

$$
\begin{aligned}
\bar{y} &= E[y] \\
&\approx E\left[f(\bar{x}) + \frac{df}{dx}(x-\bar{x}) + \frac{1}{2}\frac{d^2f}{dx^2}(x-\bar{x})^2\right] \\
&\approx f(\bar{x}) + \frac{1}{2}\frac{d^2f}{dx^2}\mu_2
\end{aligned}
\tag{11}
$$

Subtracting Eqs. (11) from (10) gives (12).

$$
y-\bar{y} \approx \frac{df}{dx}(x-\bar{x}) + \frac{1}{2}\frac{d^2f}{dx^2}(x-\bar{x})^2 - \frac{1}{2}\frac{d^2f}{dx^2}\mu_2
\tag{12}
$$

Squaring and taking the expectation of Eqs. (12) produces (13).

$$
E\left[(y-\bar{y})^2\right] \approx \mu_2\left(\frac{df}{dx}\right)^2 + \mu_3\frac{df}{dx}\frac{d^2f}{dx^2} + \frac{1}{4}\left(\mu_4-\mu_2^2\right)\left(\frac{d^2f}{dx^2}\right)^2 = \sigma_y^2
\tag{13}
$$

If x is Gaussian, all odd moments (μ_k where k is odd) are zero and Eqs. (13) reduces to (14).

$$\sigma_y^2 \approx \left(\frac{df}{dx}\right)^2 \mu_2 + \frac{1}{4}\left(\frac{d^2 f}{dx^2}\right)^2 (\mu_4 - \mu_2^2) \approx \left(\frac{df}{dx}\right)^2 \sigma_x^2 + \frac{1}{4}\left(\frac{d^2 f}{dx^2}\right)^2 (\mu_4 - \sigma_x^4) \tag{14}$$

Furthermore, if x is Gaussian, the substitution $\mu_4 \approx 3\sigma^4$ can be made [28]. This substitution is made in Eq. (15).

$$\sigma_y^2 \approx \left(\frac{df}{dx}\right)^2 \sigma_x^2 + \frac{1}{2}\left(\frac{d^2 f}{dx^2}\right)^2 \sigma_x^4 \tag{15}$$

If y is a function of multiple independent inputs, the generalized form of Eqs. (15) is given in (16).

$$\sigma_y^2 \approx \sum_{i=1}^{n}\left(\frac{\partial y}{\partial x_i}\right)^2 \sigma_{x_i}^2 + \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{n}\left(\frac{\partial^2 y}{\partial x_i \partial x_j}\right)^2 \sigma_{x_i}^2 \sigma_{x_j}^2 \tag{16}$$

Equation (16) is the second-order formula most often cited in literature [25,29] for statistical error propagation.

3.3 Key Assumptions and Limitations.

Though common in engineering literature and academia, Eqs. (9) and (16) have many significant limitations and make many important assumptions [30]. These assumptions and limitations include the following:

(1) Taking the Taylor-series expansion about a single point ($\bar{x}$) causes the approximation to be of local validity only [8,25]. Consequently, the accuracy of the approximation generally decreases with an increase in the deviation from the input mean.
(2) The approximation is generally more accurate for linear and polynomial-type models.
(3) All inputs x_i are assumed be Gaussian. When inputs are not Gaussian, the non-Gaussian terms in Eq. (13) cannot be neglected.
(4) All inputs x_i are assumed to be independent. When inputs are not independent, the covariance terms in Eq. (13) cannot be neglected [22,26,31].
(5) The input means and variances must be known.
(6) The number of terms in Eq. (13) increases exponentially as the number of model inputs increases [30].
(7) The output error distribution is assumed to be Gaussian, described by only a mean and standard deviation.

4 Propagation of Skewness

Non-Gaussian distributions cannot be fully described with a only mean and standard deviation. Consequently, higher-order statistics, such as skewness and kurtosis, must also be used to describe non-Gaussian distributions. This section considers the definition of skewness and derives a formula for propagating skewness through an engineering system model.

Table 1 Comparison of positive and negative skew

Sign	Skewed	Mean versus median [32]
Negative	Left	Mean < median (typically)
Positive	Right	Mean > median (typically)

4.1 Definition of Skewness.

The first-order statistic of a distribution is its mean, the second-order statistic is its standard deviation, and the third-order statistic is its skewness. Skewness is a measure of a distribution's asymmetry. Skewness (denoted γ_1) is defined in Eq. (17).

$$\gamma_1 = E\left[\left(\frac{x-\bar{x}}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3} = \frac{\kappa_3}{\kappa_2^{1.5}} \tag{17}$$

where E is the expectation operator, μ_3 is the third central moment, σ is the standard deviation, and κ_2 and κ_3 are the second and third cumulants, respectively.

Table 1 and Fig. 2 illustrate some characteristics and terminology of positively- and negatively-skewed distributions. A skewness of zero indicates a symmetric distribution.

Skewness is an important defining characteristic of statistical distributions. A measure of skewness is required to fully describe any asymmetric distribution. Traditional uncertainty propagation, however, only propagates a mean and variance. With no skewness information available, skewness is neglected (assumed equal to zero) and a Gaussian distribution is assumed.

4.2 Skewness Propagation Formula Derivation.

Using a first-order Taylor series to propagate skewness through a system model results in an output skewness equal to the input skewness. It has already been demonstrated that this often does not reflect reality, even for simple functions. Consequently, a second-order Taylor series will be used to derive a formula for skewness propagation.

The second central moment of output y has already been given in Eq. (13), and the third moment is given in Eq. (18).

$$E\left[(y-\bar{y})^3\right] \approx \begin{bmatrix} \mu_3 \\ \frac{3}{2}(\mu_4 - \mu_2^2) \\ \left(\frac{3}{4}\mu_5 - \frac{3}{2}\mu_2\mu_3\right) \\ \left(\frac{1}{4}\mu_2^3 - \frac{3}{8}\mu_2\mu_4 + \frac{1}{8}\mu_6\right) \end{bmatrix} \cdot \begin{bmatrix} \partial_1^3 \\ \partial_1^2\partial_2 \\ \partial_1\partial_2^2 \\ \partial_2^3 \end{bmatrix} \tag{18}$$

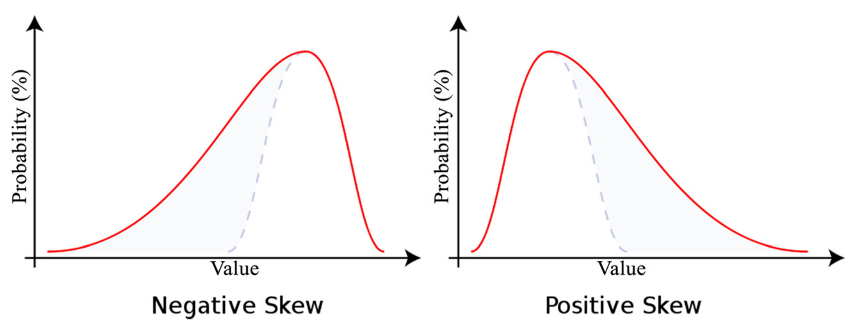

Fig. 2 Examples of negative (left) and positive (right) skewness

where μ_k is the kth central moment of input x, and ∂_1 and ∂_2, respectively, represent the partial derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial^2 f}{\partial x^2}$, evaluated at the mean $x = \bar{x}$. The third moment is a cubic function, and consequently it has four terms. Equation (18) has both first and second partial derivatives, because it is based on a second-order Taylor series. If a higher-order Taylor series were used, Eq. (18) would contain higher-order partial derivatives.

The second moment from Eq. (13) and the third moment from Eq. (18) can be used with the definition of skewness given by Eq. (17) to estimate the skewness in output y. This output skewness estimation is given in Eq. (19).

$$\gamma_1 = \frac{E\left[(y-\bar{y})^3\right]}{\left\{E\left[(y-\bar{y})^2\right]\right\}^{1.5}} \approx \frac{\begin{bmatrix} \mu_3 \\ \frac{3}{2}\left(\mu_4 - \mu_2^2\right) \\ \left(\frac{3}{4}\mu_5 - \frac{3}{2}\mu_2\mu_3\right) \\ \left(\frac{1}{4}\mu_2^3 - \frac{3}{8}\mu_2\mu_4 + \frac{1}{8}\mu_6\right) \end{bmatrix} \cdot \begin{bmatrix} \partial_1^3 \\ \partial_1^2\partial_2 \\ \partial_1\partial_2^2 \\ \partial_2^3 \end{bmatrix}}{\left[\mu_2\partial_1^2 + \mu_3\partial_1\partial_2 + \frac{1}{4}\left(\mu_4 - \mu_2^2\right)\partial_2^2\right]^{1.5}} \tag{19}$$

Equation (19) estimates output skewness using the input central moments, μ_k. If input skewness, kurtosis, and higher-order statistics are known instead of input moments, these statistics can easily be substituted into Eq. (19) in place of these moments.

Again for the sake of brevity, the skewness propagation formula has only been derived for monovariate functions. However, this derivation can easily been extended to multivariate functions as desired.

4.3 Key Assumptions and Limitations. The following four assumptions and limitations apply to the method just presented to propagate skewness:

(1) Equation (19) is based on a second-order Taylor series. Consequently, it will predict output skewness perfectly for second-order (or lower) functions. Accuracy decreases with increasing nonlinearity.
(2) System model outputs and derivatives must be obtainable from given system inputs (either analytically or numerically). This is possible for closed-form differentiable equations and many numerical models.
(3) Taking the Taylor-series expansion about a single point ($\bar{x}$) causes the approximation to be of local validity only [8,25]. Consequently, the accuracy of the approximation generally decreases with an increase in the input moments μ_k.
(4) The statistical input distribution must be known.

4.4 Skewness Propagation With Gaussian Inputs. Consider the propagation of Gaussian error. With a Gaussian distribution, the following expressions are true:

- All odd moments (μ_k, where k is odd) are equal to zero.
- The fourth moment is approximately three times the second moment squared [28] ($\mu_4 \approx 3\mu_2^2$).
- The sixth moment is approximately fifteen times the second moment cubed ($\mu_6 \approx 15\mu_2^3$).

Consequently, Eqs. (19) reduces to (20) when inputs are Gaussian.

$$\gamma_1 \approx \frac{3\sigma_x\partial_1^2\partial_2 + \sigma_x^3\partial_2^3}{\left(\partial_1^2 + \frac{1}{2}\partial_2^2\sigma_x^2\right)^{1.5}} \tag{20}$$

where σ_x is the standard deviation of the input distribution. Equation (20) proves that nonlinear functions (i.e., the second partial derivative is nonzero) produces a skewed non-Gaussian output, even with Gaussian inputs.

Furthermore, the most computationally expensive part to propagating skewness is calculating first and second derivatives. However, these have already been calculated in order to propagate variance if a second-order Taylor series was used, and consequently the additional cost to also propagate skewness is minimal.

Any statistical property that is propagated through a system model improves the accuracy of the predicted output distribution. For example, propagating both a mean and a variance is more accurate (and useful) than propagating a mean alone. In a similar manner, propagating skewness in addition to a mean and variance also improves the accuracy of the predicted output distribution.

Section 5 of this paper discusses the propagation of a higher-order statistic, kurtosis, which further improves the accuracy of the predicted output distribution.

5 Propagation of Kurtosis

This section defines kurtosis and excess kurtosis, and derives a formula for propagating kurtosis through an engineering system model.

5.1 Definition of Kurtosis. The fourth-order statistic is kurtosis. Kurtosis is a measure of a distribution's "peakedness," or the thickness of the distribution's tails. Kurtosis (denoted β_2) is the fourth standardized moment, and is defined in Eq. (21).

$$\beta_2 = E\left[\left(\frac{x-\bar{x}}{\sigma}\right)^4\right] = \frac{\mu_4}{\sigma^4} \tag{21}$$

The kurtosis of a Gaussian distribution is equal to 3.

5.2 Definition of Excess Kurtosis. In statistical analysis, "excess kurtosis" (denoted γ_2) is often used more than kurtosis. In practice, the term "kurtosis" more often refers to excess kurtosis instead of the fourth standardized moment. To avoid confusion, this paper uses the definition of kurtosis presented above and defines excess kurtosis as the fourth cumulant divided by the square of the second cumulant, as indicated in Eq. (22). Since a Gaussian distribution has a kurtosis of three, the "minus 3" in Eq. (22) causes a Gaussian distribution to have zero excess kurtosis

$$\gamma_2 = \frac{\kappa_4}{\kappa_2^2} = \frac{\mu_4}{\sigma^4} - 3 \tag{22}$$

The excess kurtosis of several common types of distributions are shown in Fig. 3.

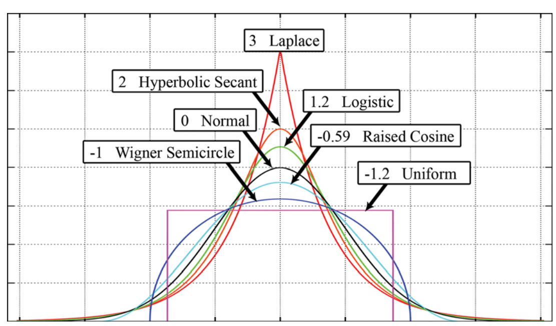

Fig. 3 The excess kurtosis of various common statistical distributions

5.3 Kurtosis Propagation Formula Derivation. A second-order Taylor series will also be used to propagate kurtosis through a system model. The third central moment of output y has already been given in Eq. (18), and the fourth moment is given in Eq. (23)

$$\mu_{4,y} = E\left[(y-\bar{y})^4\right] \tag{23}$$

$$\approx \begin{bmatrix} \mu_4 \\ 2(\mu_5 - \mu_2\mu_3) \\ \frac{3}{2}(\mu_2^3 - 2\mu_2\mu_4 + \mu_6) \\ \frac{3}{2}(\mu_2^2\mu_3 - \mu_2\mu_5 + \frac{1}{3}\mu_7) \\ \frac{1}{16}(6\mu_2^2\mu_4 - 3\mu_2^4 - 4\mu_2\mu_6 + \mu_8) \end{bmatrix} \cdot \begin{bmatrix} \partial_1^4 \\ \partial_1^3\partial_2 \\ \partial_1^2\partial_2^2 \\ \partial_1\partial_2^3 \\ \partial_2^4 \end{bmatrix} \tag{24}$$

The excess kurtosis γ_2 in the output distribution y is given by Eq. (25).

$$\gamma_2 = \beta_2 - 3 \tag{25}$$

where kurtosis β_2 is given by Eq. (26).

$$\beta_2 = \frac{E\left[(y-\bar{y})^4\right]}{\left\{E\left[(y-\bar{y})^2\right]\right\}^2} \approx \frac{\begin{bmatrix} \mu_4 \\ 2(\mu_5 - \mu_2\mu_3) \\ \frac{3}{2}(\mu_2^3 - 2\mu_2\mu_4 + \mu_6) \\ \frac{3}{2}(\mu_2^2\mu_3 - \mu_2\mu_5 + \frac{1}{3}\mu_7) \\ \frac{1}{16}(6\mu_2^2\mu_4 - 3\mu_2^4 - 4\mu_2\mu_6 + \mu_8) \end{bmatrix} \cdot \begin{bmatrix} \partial_1^4 \\ \partial_1^3\partial_2 \\ \partial_1^2\partial_2^2 \\ \partial_1\partial_2^3 \\ \partial_2^4 \end{bmatrix}}{\left[\mu_2\partial_1^2 + \mu_3\partial_1\partial_2 + \frac{1}{4}(\mu_4 - \mu_2^2)\partial_2^2\right]^2} \tag{26}$$

An estimate of the kurtosis of an output distribution can be obtained using Eq. (26) and a known input distribution. The input central moments μ_k can be estimated using any appropriate population sampling technique.

6 An Illustrative Example: Flat-Rolling Process

Consider the manufacture of steel plates or sheets via flat rolling, where material is fed between two rollers (called working rolls). This example illustrates that uncertainties in friction between rolls and rolling material—an engineering-centric concept—highly affects factory throughput and ultimately a rolling company's business plan. By using the relations derived in this paper, a more meaningful inclusion of frictional effects is made, and the rolling throughput is more effectively planned for.

In any plate-rolling mill, the gap between the working rolls is less than the thickness of the incoming material. As the working rolls rotate in opposite directions, the incoming material elongates as its thickness is reduced. This process, illustrated in Fig. 4, can be done either below the recrystallization temperature of the material (cold rolling) or above it (hot rolling).

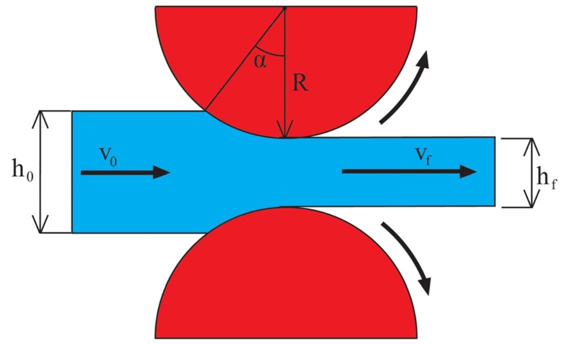

Fig. 4 The flat-rolling manufacturing process whereby plates or sheets of metal are made. Material is drawn between two rollers, which reduces the material's thickness.

6.1 The Model. The manufacturer desires to use its flat-rolling equipment more efficiently by reducing overall rolling time for each plate. Consequently, the manufacturer desires to minimize the number of passes required to achieve final plate thickness. The maximum amount of deformation that can be achieved in a single pass is a function of the friction at the interface between the rolls and the material. If the intended change in thickness is too great, the rolls will merely slip along the material without drawing it in [33]. The maximum change in thickness attainable in a single pass ($\Delta H_{\max}$) is given in Eq. (27) [34].

$$\Delta H_{\max} = \mu_f^2 R \tag{27}$$

where μ_f is the coefficient of friction between the rolls and the plate, and R is the radius of the rolls. In this example, the radius of the rolls is measured to be 0.406 m. Determining the coefficient of friction in a metalworking process is more difficult, however. The conditions surrounding friction in a metalworking process are very different from those in a mechanical device [33], as shown in Table 2.

Furthermore, lubrication is often used both to reduce friction and consequent tool wear, and to act as a thermal barrier to help regulate tool temperature [35]. All these factors and others (e.g., rolling speed, material properties, surface finishes, etc.) combine to create variation in the friction experienced in the flat-rolling metalworking process. This variation can inhibit the manufacturer's ability to specify an optimal gap width (and the resulting change in material thickness, ΔH) for each pass.

While many empirical and mathematical formulas have been presented as methods to predict the coefficient of friction in flat-rolling processes, these will not be addressed in this paper. For the purposes of this example, it is sufficient to assume that some appropriate technique has been employed to determine the distribution of friction coefficients. This distribution is described in Table 3 and shown in Fig. 5.

Table 2 Comparison of the conditions of friction found in typical mechanical devices and metalworking processes

Typical mechanical devices	Metalworking processes
Two surfaces of similar material and strength	One very hard tool and one softer material
Elastic loads and no change in shape	Plastic deformation occurs in material
Wear-in cycles produce surface compatibility	Each set of rollers makes a single pass Contact area constantly changes under deformation
Low/moderate temperature	Often elevated temperature
Friction force depends on contact pressure	Friction force depends on material strength

Table 3 Statistical properties of the distribution of the friction coefficient in a flat-rolling metalworking process

Statistical property	Value
Mean ($\bar{\mu}$)	0.35
Variance (σ^2)	9×10^{-4}
Skewness (γ_1)	0.7
Excess kurtosis (γ_2)	0.2

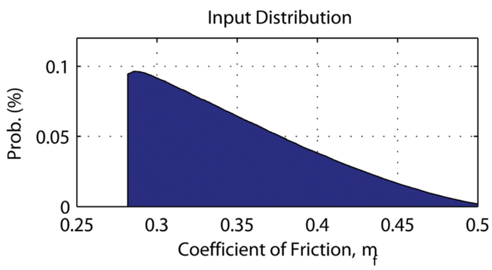

Fig. 5 Distribution of the coefficient of friction in a flat-rolling metalworking process

Table 4 Central moments of the distribution of the coefficient of friction, μ_f

Central moment	Value
First moment (μ_1)	0
Second moment (μ_2)	8.99×10^{-4}
Third moment (μ_3)	1.88×10^{-5}
Fourth moment (μ_4)	2.58×10^{-6}
Fifth moment (μ_5)	1.45×10^{-7}
Sixth moment (μ_6)	1.45×10^{-8}
Seventh moment (μ_7)	1.21×10^{-9}
Eighth moment (μ_8)	1.2×10^{-10}

6.2 Statistical System Model Output Prediction. Based on this distribution of μ_f, the first eight central moments were calculated using Eq. (2). These moments, given in Table 4, are used to propagate statistical properties from the friction coefficient distribution to predict a distribution for the maximum change in thickness attainable with a single pass.

The second-order prediction of the mean maximum reduction in thickness $\overline{\Delta H_{\max}}$ was calculated to be 5.0 cm using Eq. (11). Equation (13) predicts a variance of 2.16 cm^2. Typically, higher-order statistics are not propagated and a Gaussian output distribution is assumed, which is shown in Fig. 6(*a*). This prediction only accounts for 52.9% of the actual system output distribution (i.e., 52.9% overlap in the area under the predicted and actual probability density functions).

However, the method presented in this paper to propagate skewness and kurtosis results in a more accurate prediction of the system output. In this example, Eq. (19) estimates an output skewness of 0.954, and Eq. (26) estimates an output kurtosis of 3.286. This predicted output, shown in Fig. 6(*b*), accounts for 94.4% of actual system outputs—a large improvement over propagating a mean and variance alone.

Note once a mean, variance, skewness, and kurtosis are known, a probability density function can be determined using an empirical distribution system, such as the Pearson system or the Johnson system [10]. In this example, a Pearson system was used to generate the probability density functions shown in Fig. 6.

6.3 Ramifications of Neglecting Higher-Order Statistics. Using the probability density function obtained from the traditional approach—where only a mean and variance are propagated and a Gaussian distribution is assumed—the manufacturer would have concluded that with a 99.5% chance of success, the material thickness could only be reduced by a maximum of 1.21 cm per pass. However, using the method presented in this paper to also propagate higher-order statistics, the manufacturer can conclude that the material thickness could be reduced by 3.22 cm per pass with the same likelihood of success. This reduces the number of passes required to achieve the desired plate thickness by over two and a half times, which is a fundamental consideration to any business plan.

As this example clearly indicates, the benefits of propagating higher-order statistics through a system model can be substantial. Fortunately, estimates of output skewness and kurtosis are easy to obtain using the formulas derived in this paper. Designers can use these same formulas in a wide range of both simple and complex engineering system models.

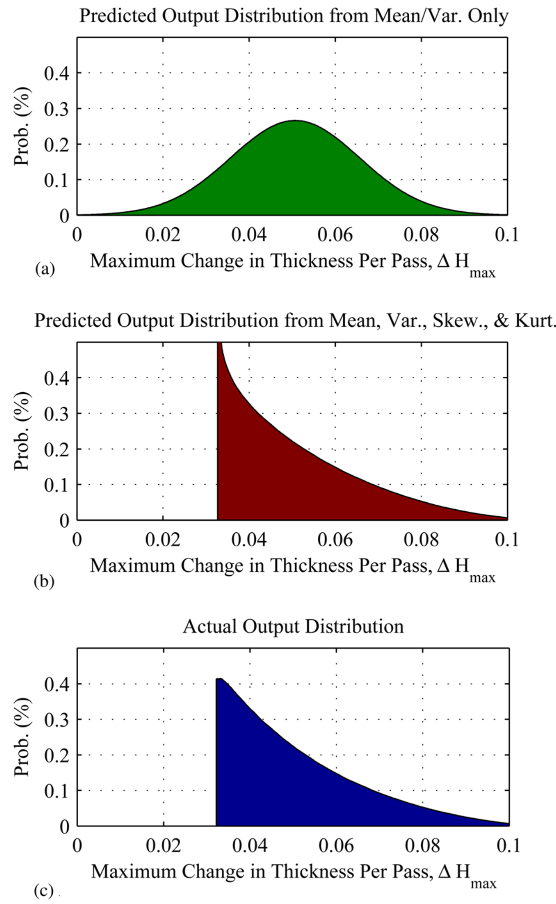

Fig. 6 Predicted output distributions obtained from propagating (*a*) mean and variance only, and (*b*) mean, variance, skewness, and kurtosis. Actual system output distribution is shown in (*c*).

6.4 Accuracy and Cost Comparisons. This example is now used to compare the accuracy and computational cost of the Taylor-series-based methods presented in this paper with other error propagation techniques. Specifically, these other techniques are a MC simulation with 1 million executions and a Latin hypercube sampling (LHC).

Accuracy was compared using the same metric introduced in Sec. 6.2 (percent of the predicted distribution that overlaps with the actual system output distribution). The accuracy of the MC simulation was 100%, the Taylor-series method presented in this paper was 94.4%, and the LHC was 94.6%. Figure 7 shows these results.

Computational cost was measured by MATLAB execution time. The MC simulation took 5.951 s, the Taylor-based method from this paper took 0.007 s, and the LHC took 0.851 s. This is illustrated in Fig. 8.

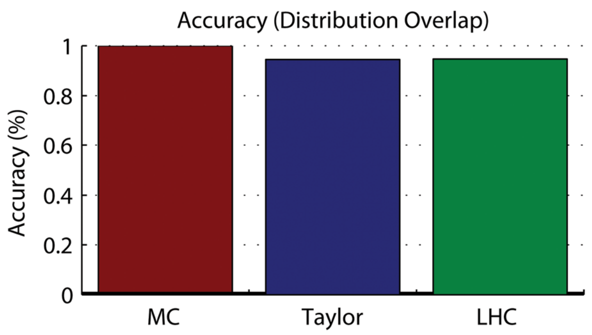

Fig. 7 Accuracy (distribution overlap) of different error propagation methods

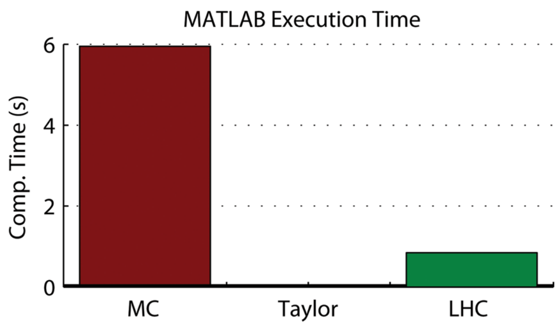

Fig. 8 Computational cost of different error propagation methods, as measured by MATLAB execution time

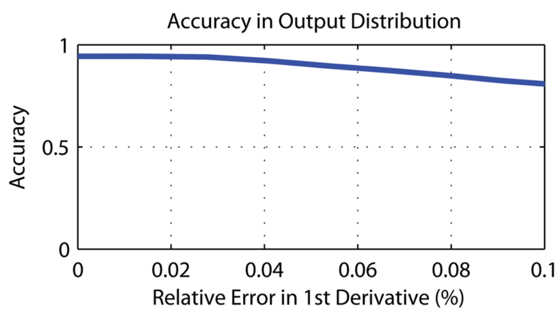

Fig. 9 Sensitivity of the accuracy of the predicted output distribution to error in the derivative approximations

While the accuracy of the Taylor method presented in this paper is similar to that achieved with LHC, the Taylor approach has significantly less computation time. Furthermore, the Taylor model also has an added advantage in that it is simple and can be worked out by hand, where the LHC requires a computer and a programmed algorithm.

6.5 Sensitivity to Derivative Approximation Errors. The method presented in this paper requires first- and second-order derivatives. Often engineering models are "black-box" functions and a finite difference method must be used to calculate these derivatives. To demonstrate the sensitivity of the output distribution to errors resulting from finite difference derivative approximations, this same example was solved repeatly using a forward-difference derivative approximation with varying step sizes. Approximately, a 3% relative error in the derivative approximation can be absorbed with little impact on overall accuracy. A relative error larger than 3% begins to linearly decrease the accuracy of the predicted output. Figure 9 shows this relationship.

As seen in Fig. 9, the method presented in this paper is not particularly sensitive to small errors in derivative approximations, and errors as large as 10% only reduced the accuracy from 94% to 80%—still significantly better than propagating a mean and variance alone with perfect derivative values.

7 Conclusions

The variance in a system's output can easily be predicted using a Taylor series and knowledge of the input variance. However, having only a mean and variance and lacking any additional information about the output distribution, system designers often make the erroneous assumption that the output is Gaussian. This paper has shown how inaccurate that assumption can be, even for very simple functions. By following the methods shown in this paper, system designers can more fully describe an output distribution by also propagating higher-order statistics, such as skewness and kurtosis, though a system model.

While sufficient for many physical systems, the approach to higher-order statistical error propagation presented in this paper may not work for all types of system models, such as state-space models, Laplace transforms, and differential equations. Additional work is required to adapt the method presented for use with these types of models.

References

[1] Hamaker, H. C., 1995, "Relative Merits of Using Maximum Error Versus 3(Sigma) in Describing the Performance of Laser-Exposure Reticle Writing Systems," Proc. SPIE, **2440**, p. 550.
[2] Hamel, J., Li, M., and Azarm, S., 2010, "Design Improvement by Sensitivity Analysis Under Interval Uncertainty Using Multi-Objective Optimization," J. Mech. Des, **132**(8), p. 081010.
[3] Mekid, S., and Vaja, D., 2008, "Propagation of Uncertainty: Expressions of Second and Third Order Uncertainty With Third and Fourth Moments," Measurement, **41**(6), pp. 600–609.
[4] Oberkampf, W. L., DeLand, S. M., Rutherford, B. M., Diegert, K. V., and Alvin, K. F., 2002, "Error and Uncertainty in Modeling and Simulation," Reliab. Eng. Syst. Saf., **75**, pp. 333–357.
[5] Halton, J. H., 1960, "On the Efficiency of Certain Quasi-Random Sequences of Points in Evaluating Multi-Dimensional Integrals," Numer. Math., **2**, pp. 84–90.
[6] Hammersley, J. M., 1960, "Monte Carlo Methods for Solving Multivariate Problems," Ann. N.Y. Acad. Sci., **86**, pp. 844–874.
[7] Owen, A. B., 1998, "Latin Supercube Sampling for Very High-Dimensional Simulations," ACM Trans. Model. Comput. Simul., **8**(1), pp. 71–102.
[8] Hutcheson, R. S., and McAdams, D. A., 2010, "A Hybrid Sensitivity Analysis for Use in Early Design," J. Mech. Des., **132**(11), p. 111007.
[9] Samet, H., 2005, *Foundations of Multidimensional and Metric Data Structures (The Morgan Kaufmann Series in Computer Graphics and Geometric Modeling)*, Morgan Kaufmann Publishers, Inc., San Francisco, CA.
[10] Lee, S., and Chen, W., 2009, "A Comparative Study of Uncertainty Propagation Methods for Black-Box-Type Problems," Struct. Multidiscip. Optimiz., **37**, pp. 239–253.
[11] Larson, B., Anderson, T. V., and Mattson, C. A., 2010, "System Behavioral Model Verification for Concurrent Design and Modeling," 13th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference Proceedings, Paper No. 9104 in AIAA-2010, AIAA/ISSMO.
[12] Poolla, K., Khargonekar, P., Tikku, A., Krause, J., and Nagpal, K., 1994, "A Time-Domain Approach to Model Validation," IEEE Trans. Autom. Control, **39**(5), pp. 951–959.
[13] Evans, J. W., Zawadzki, R. J., Jones, S. M., Olivier, S. S., and Werner, J. S., 2009, "Error Budget Analysis for an Adaptive Optics Optical Coherence Tomography System," Opt. Express, **17**(16), pp. 13768–13784.
[14] Hamaker, H. C., 1995, "Improved Estimates of the Range of Errors on Photomasks Using Measured Values of Skewness and Kurtosis," Proc. SPIE, **2621**, pp. 198–207.
[15] Oschmann, J., 1997, Gemini System Error Budget Plan, January. Available at http://www.gemini.edu/documentation/webdocs/spe/spe-s-g0041.pdf
[16] Hayes, B., 2003, "A Lucid Interval," Am. Sci., **91**(6), pp. 484–488.
[17] Box, G. E., and Tiao, G. C., 1992, *Bayesian Inference in Statistical Analysis*, Wiley, Hoboken, NJ.
[18] 2011, "Basics of Bayesian Inference and Belief Networks," www.research.microsoft.com, http://goo.gl/rSJCD, July.
[19] Lombardi, M., and Haftka, R. T., 1998, "Anti-Optimization Technique for Structural Design Under Load Uncertainties," Comput. Methods Appl. Mech. Eng., **157**(1-2), pp. 19–31.
[20] Koch, P. N., 2002, "Probabilistic Design: Optimizaing for Six Sigma Quality," 43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Paper No. AIAA-2002-1471.
[21] Glancy, C., 1999, "A Second-Order Method for Assembly Tolerance Analysis," Proceedings of the 1999 ASME Design Engineering Technical Conferences, Paper No. DAC-8707 in DETC99, ASME/DETC.
[22] Vardeman, S. B., 1994, *Statistics for Engineering Problem Solving*, PWS Publishing Company, Boston, MA.

[23] Jackson, P. S., 1982, "A Second-Order Moments Method for Uncertainty Analysis," IEEE Trans. Reliab., **R-31**(4), pp. 382–384.
[24] Walwyn, R., 2005, *Moments - Encyclopedia of Statistics in Behavioral Science*, Vol. 3, John Wiley and Sons, Ltd., New Jersey.
[25] Mattson, C. A., and Messac, A., 2002, "A Non-Deterministic Approach to Concept Selection Using S-Pareto Frontiers," Proceedings of ASME DETC, Vol. 2, Paper No. DETC2002/DAC-34125, pp. 859–870.
[26] Tellinghuisen, J., 2001, "Statistical Error Propagation," J. Phys. Chem. A, **105b**(15), pp. 3917–3921.
[27] Lindberg, V., 2000, Uncertainties and Error Propagation—Part I of a Manual on Uncertainties, Graphing, and the Vernier Caliper, Internet, July. Available at http://www.rit.edu/cos/uphysics/uncertainties/Uncertaintiespart2.html
[28] Julier, S., Uhlmann, J., and Durrant-Whyte, H. F., 2000, "A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators," IEEE Trans. Autom. Control, **45**(3), pp. 477–482.
[29] Putko, M. M., Arthur C. T., III, Newman, P. A., and Green, L. L., 2002, "Approach for Uncertainty Propagation and Robust Design in CFD Using Sensitivity Derivatives," J. Fluids Eng., **124**(1), pp. 60–69.
[30] Anderson, T. V., Mattson, C. A., Larson, B. J., and Fullwood, D. T., 2011, "Efficient Propagation of Error Through System Models for Functions Common in Engineering," J. Mech. Des., **134**(1), p. 014501.
[31] Goodman, L. A., 1960, "On the Exact Variance of Products," J. Am. Stat. Assoc., **55**(292), pp. 708–713.
[32] von Hippel, P. T., 2005, "Mean, Median, and Skew: Correcting a Textbook Rule," J. Stat. Educ., **13**(2). Available at http://www.amstat.org/publications/jse/v13n2/vonhippel.html
[33] Degarmo, E. P., Black, J. T., and Kohser, R. A., 2003, *Materials and Processes in Manufacturing*, 9 ed., Wiley, New York.
[34] Tyfour, W. R., "Rolling," http://www.freewebs.com/tyfour/Rolling.doc
[35] Lenard, J. G., 2007, *Primer on Flat Rolling*, Elsevier Science, Hoboken, NJ, pp. 868–876.