Lack-of-fit Tests to Indicate Material Model Improvement or Experimental Data Noise Reduction

Charles F. Jekel* and Raphael T. Haftka†
University of Florida, Gainesville, FL, 32611, USA

Gerhard Venter‡ and Martin P. Venter §
Stellenbosch University, Stellenbosch, Western Cape, 7600, South Africa

A lack-of-fit test can be used to determine whether a finite element (FE) material model, or experimental results should be the focus to improve overall model accuracy. The lack-of-fit test compares the unbiased estimate of variance to the estimated variance from noise within data. GSJ and Hart methods from literature are provided to estimate the noise within data, and a generalized lack-of-fit test is described. The variance estimators are compared and convergence is demonstrated on a simple regression problem. The lack-of-fit test is then applied to a material calibration problem using the FE method. The lack-of-fit test indicates that the shear component of a non-linear orthotropic material models could be improved, despite already being considered an excellent fit. Additionally the lack-of-fit test is used with a load-dependent Poisson's ratio to demonstrate that the variance in the experimental data should be reduced in order to improve the model.

Nomenclature

a	Linear parameter for GSJ
a_n	Parameter for Hart
b	Linear parameter for GSJ
e	Residual value
$\tilde{e}$	Pseudo residual value
$\boldsymbol{e}$	Residual value vector
$\tilde{\boldsymbol{e}}$	Pseudo residual value vector
$f(x)$	Dependent variable value at x
$\hat{f}(x)$	Predictive model value at x
F	F-statistic value
$\boldsymbol{H}$	Tridiagonal matrix
n	Number of data points
n_β	Number of model parameters
P	P-value
x	Independent variable value
$\boldsymbol{x}$	Vector of independent variable values
$\boldsymbol{X}$	Regression matrix
β	Model parameter
σ^2	Variance
$\hat{\sigma}^2$	Unbiased estimate of variance
$\hat{\sigma}_e^2$	GSJ variance estimator

*PhD Student, Dept. of Mechanical and Aerospace Engineering University of Florida, and Dept. of Mechanical and Mechatronic Engineering Stellenbosch University.
†Distinguished Professor, Dept. of Mechanical and Aerospace Engineering University of Florida, and AIAA Fellow.
‡Professor, Dept. of Mechanical and Mechatronic Engineering Stellenbosch University.
§Lecturer, Dept. of Mechanical and Mechatronic Engineering Stellenbosch University.

$\tilde{\sigma}^2$ Hart variance estimator
$\hat{\sigma}_x^2$ Variance estimator between consecutive x
$\hat{\sigma}_f^2$ Variance estimator between consecutive $f(x)$
Subscript
i Index of data

I. Introduction

It has become common practice to calibrate material model parameters on experimental data for finite element (FE) analyses. The overall accuracy of the FE model is dependent on the model's ability to replicate the material behavior. Thus it is important to select the best material parameters to accurately describe the material behavior. In general parameter values are selected such that the FE model reproduces some experimental results. The accuracy of the FE model can be improved by either improving the model (e.g., finding a better material model), or by improving the experimental data (e.g., reducing the uncertainty in the experimental data). Potentially a lack-of-fit test may be used to decide which aspect (model or data) should be the focus for improvement.

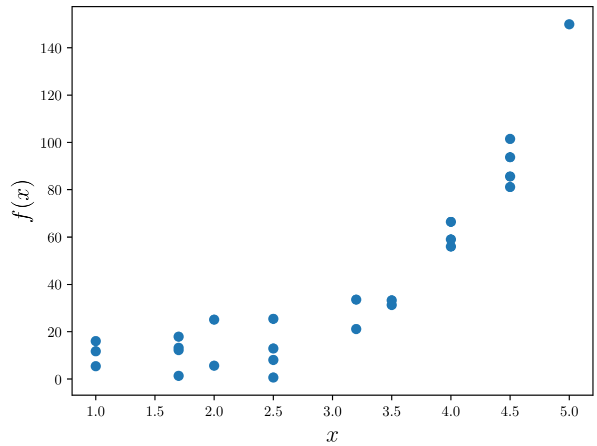
In this context a lack-of-fit investigation looks at whether the primary contribution of error is due to the inherent variance in experimental data, or the inability of the material model to replicate the experimental data. Mismatch between the FE model and experimental data is due to both the material model not being accurate enough, and noise in the experimental data. However noise doesn't just appear in experimental data, and sometimes a poor model may generate aditional noise as demonstrated by Park et al.[1] Noise leads to discrepancies between the true material model and the captured experimental data. Additionally noise in data may yield calibrated material parameters that are not accurate. A model passing the lack-of-fit test indicates that the noise in the experimental data is limiting the accuracy of the calibrated parameters. Alternatively a failed lack-of-fit test indicates that the material model is unable to replicate the response captured in the experimental data.

Analysis of variance (ANOVA) is a widely used set of statistical models aimed at comparing variation between data. ANOVA models include the partitioning of the sum of squares, lack-of-fit tests, likelihood ratio test, and the F-test. Traditionally these methods require the data to have true replicates, though practically obtaining these replicates is difficult for material tests. In terms of material response modeling, true replicates would require multiple stress values for each strain value. However material test data is typically collected in a way such that there is a dependent response as a function of some other independent variable (e.g., variation of stress values among strain values for repeated uniaxial tests).

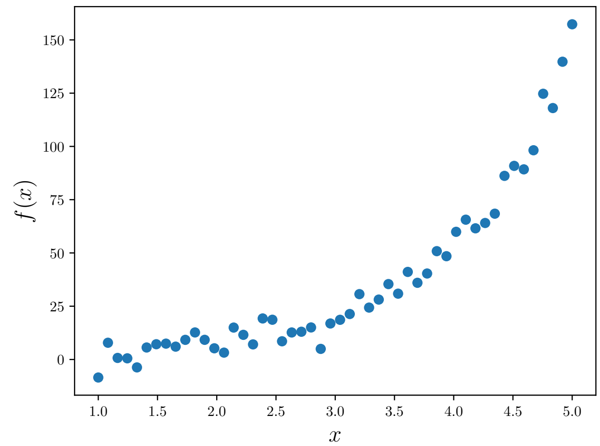
An example of the difference between true replicates and a typical material response can be seen in figure 1. There are multiple $f(x)$ values for a given x when looking at a response containing true replicates. True replicates of experimental responses can be difficult to capture. An experiment is generally performed to captured a response $f(x)$ which varies with respect to x. The experiment itself is repeated, however the captured responses may not contain replicates for the exact same independent variable x. There are classical statistical methods that can be applied when true replicates exist.[2]

This paper goes on to describe the lack-of-fit test which is based on comparing a ratio of variances. First the unbiased variance is calculated which represents the variance between the model and data. Then the variance from the noise (or scatter) within the data is estimated. Two different variance estimators are used from literature. First the popular GSJ variance estimator is described.[3] Then a generalized variance estimator proposed by Hart is described.[4] These variance estimators are used to estimate the variance due to the noise (or scatter) within experimental data. A classical lack-of-fit test is then described from the literature by comparing these variances.

The GSJ and Hart variance estimators are compared against each other. Additionally the convergence behavior of the variance estimators and the lack-of-fit tests is demonstrated on a simple linear regression example. The lack-of-fit tests are then used on a material parameter identification problem. An example shows a case where it is suggested that the shear component of a non-linear orthotropic material model be improved in order to improve the accuracy of the FE model. Another example uses the lack-of-fit test to show that the variation within experimental data be reduced in order to improve a load-dependent Poisson's ratio model.



(a) True replicates



(b) Typical response

Figure 1. True replicates shown in (a) where there are multiple $f(x)$ for a unique x. A more typical response is shown in (b) of collected experimental data where $f(x)$ appears to demonstrate inherent variation amid different x values.

II. Statistical Methods

The variances investigated either represent the variance from the *error* of the fit, or the variance from *noise* (or scatter) within the data. Mathematical definitions of the variances are proposed from literature. The unbiased variance represents the variance of the error between the model and data. The GSJ and Hart methods are presented to approximate the inherent variance from noise within the data. Popular alternative variance estimates are also mentioned, though this work emphasizes the GSJ and Hart methods. Lastly the F-statistic is proposed for a generalized lack-of-fit test.

A. Unbiased Variance

The residual $\boldsymbol{e}$ is described as

$$\boldsymbol{e} = f(\boldsymbol{x}) - \hat{f}(\boldsymbol{x}) \tag{1}$$

where the predictive model is represented by $\hat{f}(\boldsymbol{x})$, and the data is represented by $f(\boldsymbol{x})$. The unbiased estimate of the variance is

$$\hat{\sigma}^2 = \frac{\boldsymbol{e} \cdot \boldsymbol{e}^{\mathrm{T}}}{n - n_\beta} \tag{2}$$

for a linear regression problem, with n number of data points, and n_β number of parameters in the predictive model.[2a] The unbiased estimate of the variance assumes that the model accurately describes the data. If this unbiased variance is significantly larger than the variance due to noise in the data, then the model is not accurate.

The measure of the lack-of-fit can be described by the difference between the unbiased variance and the variance from noise in data. It is common to use the unbiased variance as a goodness-of-fit metric for a predictive model. It can be noted that $\boldsymbol{e} \cdot \boldsymbol{e}^{\mathrm{T}}$ represents the sum of the square of the residuals, another metric that is often used to represent goodness-of-fit for regression problems.

B. GSJ Variance Estimator

Gasser et al. proposed a variance estimator based on local linear fitting between three consecutive points.[3] This method for estimating variance is commonly referred to as the GSJ method and was originally intended for non-linear regression. However, the GSJ method isn't just limited to applications in non-linear regression. For instance the GSJ method can be used to approximate the variance due to noise (or scatter) within data.

[a]Myers et al. in Ref. 2 refers to $\hat{\sigma}^2$ as the unbiased estimator. It was chosen to refer to this term as the unbiased variance in order to avoid confusion from the later described variance estimators.

Let's considered a data set $f(\boldsymbol{x})$, where $\boldsymbol{x}$ is ordered as $x_1 < x_2 < \cdots < x_n$. A vector of pseudo residuals $\tilde{\boldsymbol{e}}$ can be constructed by considering every three consecutive data points. A single pseudo residual $\tilde{e}_i$ is described by

$$\tilde{e}_i = \frac{x_{i+1} - x_i}{x_{i+1} - x_{i-1}} f(x_{i-1}) + \frac{x_i - x_{i-1}}{x_{i+1} - x_{i-1}} f(x_{i+1}) - f(x_i) \tag{3}$$

$$= a_i f(x_{i-1}) + b_i f(x_{i+1}) - f(x_i) \tag{4}$$

for $i = 2, 3, \cdots, n-1$. The variance $\hat{\sigma}_e^2$ of the estimated noise within data is obtained by taking

$$\hat{\sigma}_e^2 = \frac{1}{n-2} \sum_{i=2}^{n-1} \frac{e_i^2}{a_i^2 + b_i^2 + 1} \tag{5}$$

where n represents the number of data points. It's worthwhile to note that $\hat{\sigma}_e^2$ can be calculated for all one dimensional data containing at least three points. The GSJ variance estimator becomes more accurate as the number of data points increases, and this convergence behavior is demonstrated in Section III.

C. Hart Variance Estimator

Hart proposed a generalized variance estimator useful for lack-of-fit tests.[4] The estimator considers the noise within data by taking differences between every two consecutive residuals. The derivation is meant for linear regression problems, however it can be extended to non-linear regression problems by using the linearized regression matrix described by Coppe et al.[5] An example regression matrix $\boldsymbol{X}$ for fitting a quadratic polynomial to data is represented bellow.

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \\ 1 & x_n & x_n^2 \end{bmatrix} \tag{6}$$

The data should be ordered consecutively as $x_1 < x_2 \cdots < x_n$. Thus the first residual is defined as the difference between the data and predictive model at the first location $e_1 = f(x_1) - \hat{f}(x_1)$, the second residual $e_2 = f(x_2) - \hat{f}(x_2)$, and so forth as $e_n = f(x_n) - \hat{f}(x_n)$. The estimate of the variance from the noise in the data $\tilde{\sigma}^2$ is described as

$$\tilde{\sigma}^2 = \frac{\boldsymbol{e}^{\mathrm{T}} \cdot \boldsymbol{H} \cdot \boldsymbol{e}}{a_n} \tag{7}$$

$$= \frac{1}{a_n} \sum_{i=2}^{n} \left(e_i - e_{i-1}\right)^2 \tag{8}$$

where $a_n = 2(n-1) - \mathrm{trace}(\boldsymbol{H}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}})$ and $\boldsymbol{H}$ is the $n \times n$ tridiagonal matrix.

$$H = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{bmatrix} \tag{9}$$

Hart uses this generalized variance estimate to perform non-parametric lack-of-fit tests, stating that it is more conservative than GSJ. However unlike GSJ, Hart's method is dependent upon the chosen regression model. The convergence behavior of the Hart variance estimator (with respect to the accuracy of the regression model and the number of data points) is demonstrated in Section III.

D. Alternative Variance Estimators

Various other methods have been used to estimate the variance from noise in data. Notable simple methods include estimating the noise by the variation between consecutive x points as $\hat{\sigma}_x^2$, or by the variation between consecutive $f(x)$ points as $\hat{\sigma}_f^2$.

$$\hat{\sigma}_x^2 = \frac{1}{2(n-1)} \sum_{i=2}^{n} \left(x_i - x_{i-1}\right)^2 \tag{10}$$

$$\hat{\sigma}_f^2 = \frac{1}{2(n-1)} \sum_{i=2}^{n} \left(f(x_i) - f(x_{i-1})\right)^2 \tag{11}$$

However Gasser et al. demonstrated that the GSJ method was more accurate than $\hat{\sigma}_f^2$.[3] Additionally $\hat{\sigma}_x^2$ will not provide useful insight when x has been systematically spaced. It is worthwhile to note that Altman and Paulson provided an alternative derivation for the GSJ method when the spacing between x points is consistent.[6] Hart stated in Ref. 4 that his generalized method is closely related to the cubic spline method created by Munson and Jernigan.[7] Expansions of the GSJ method could be created using higher order (quadratic, cubic) interpolations between consecutive data points.

E. F-test and lack-of-fit

The F-test represents a statistical test for investigating lack-of-fit. Essentially F represents our test statistic, which will be used to determine whether or not a model is adequate with the provided data. A classical lack-of-fit test categorizes the error resulting from either *lack-of-fit* or *pure error*.[2] Lack-of-fit error refers to the error between the model and true function. While pure error represents the inherent error from the noise (or scatter) within the data. Using these principles, the F-statistic can be generalized as the the ratio of variances

$$F \approx \frac{\hat{\sigma}^2}{\hat{\sigma}_e^2} \approx \frac{\hat{\sigma}^2}{\tilde{\sigma}^2} \tag{12}$$

where the unbiased estimate of variance is compared to the estimated variance of noise within the data.

The F-statistic follows the $\mathrm{F}(X; n-n_\beta, n-1)$ distribution where n represents the number of data points and n_β represents the number of model parameters. The null hypothesis H_0 is constructed assuming that there is no lack-of-fit, and the model is adequate given the scatter in the data. For this hypothesis, the P-value is obtained by taking the complementary cumulative distribution of $\mathrm{F}(F; n-n_\beta, n-1)$. The null hypothesis is accepted when

$$H_0 : \text{P-value} > 0.05 \tag{13}$$

and the P-value is greater than some level of confidence, which conventionally is 0.05 for significance.

A rejected null hypotheses implies that the model poorly describes the data. The null hypothesis is rejected for large F values, where the majority of variance is characterized from the inability of the model to fit the data. Alternatively all of the modeling variance could be equal to the variance from noise in the data, or $F = 1$. When $F = 1$ the model is accepted, because it is assumed that the only error in the model originates from the inherent noise in the data. It is possible for the F-statistic to be less than 1, because the estimated variance from noise may be greater than the unbiased variance (particularly when there is error in the estimated variance). The models are usually accepted for a small test statistic where $F < 1$ because the error is less than the approximated scatter within the data.

III. Demonstrated Convergence of Statistical Methods

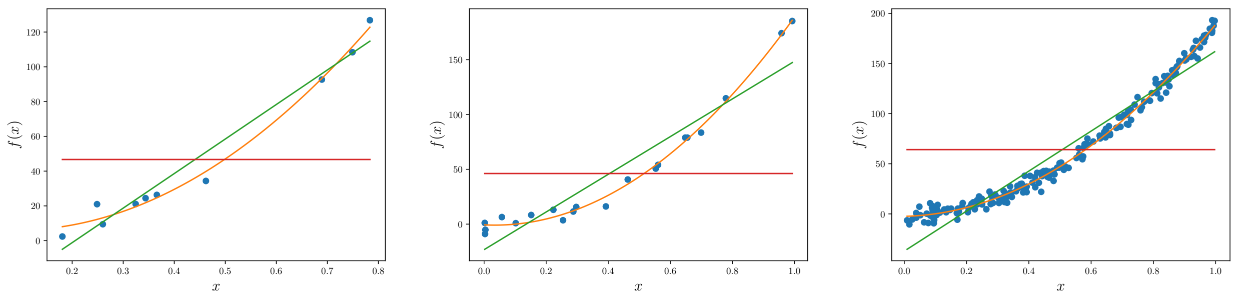
A simple example was created to demonstrate the convergence behavior of the variance estimators and lack-of-fit test. Data is sampled from the equation

$$f(x) = -1.4 + 1.5x + 190.3x^2 \tag{14}$$

with the addition of random Gaussian noise that follows the normal distribution of $N(\mu = 0, \sigma^2 = 25)$. The number of data points was increased to demonstrate the convergence behavior. The data points x_i are selected at random from $0 \leq x_i \leq 1$. The least squares method is used to fit polynomials with degrees ranging from zero to five to the data. With the Hart method, it is also important to understand the effect of the chosen regression model on the estimated variance. Polynomials with degrees three to five were chosen to provide insight on overfitting the true function. A k degree polynomial is expressed as

$$\hat{f}(\boldsymbol{x}) = \sum_{j=0}^{k} \beta_j \boldsymbol{x}^j \tag{15}$$

where the β_j parameters are determined from the least squares fit. Examples of polynomial fits for 10, 20, and 200 data points can be seen in figure 2.



(a) 10 data points (b) 20 data points (c) 200 data points

Figure 2. Constant, linear, and quadratic polynomial fits for 10, 20, and 200 data points.

A. Variance Estimator Convergence

The estimated variance from noise for the GSJ and Hart methods are compared to the variance of the added noise randomly drawn from the normal distribution. It is the intention of this comparison to demonstrate the accuracy of the estimated variance from noise for a simple data set. The estimated variances are presented in table 1. For a small number of points the GSJ method does a poor job at estimating the variance of the noise in the data. However as the number of data points increases, the accuracy of the GSJ method improves. The same trend is observed with the Hart method, in which the number of data points improved the accuracy of the estimated variance of noise. The Hart method was able to estimate the noise with reasonable accuracy when using higher degree polynomials and only 10 data points. For this example the Hart method using the 4^{th} degree polynomial produced the most accurate estimation of the variance from noise in the signal. Lastly it is worthwhile to note that even with 20,000 data points, the estimated variance from all methods converged to 25.3 despite having a true variance of 25.0. This discrepancy looks worse as a variance than as a standard deviation because of the squared term. If standard deviations were considered, the estimated standard deviation would be 5.03 and the true standard deviation would be 5.0.

It was possible the random spacing of the $\boldsymbol{x}$ data points influenced the performance of the GSJ variance. To demonstrate this affect, the GSJ variance was recalculated for linearly spaced x data points chosen from $0 \leq x_i \leq 1$. The noise from the normal distribution was preserved from the previous example. The results of the linearly spaced x values on the GSJ variance is presented in table 2. The GSJ method estimated the variance from noise more accurately when using linearly spaced data points as opposed to the randomly selected data points. This likely occurs because the linear approximation between three *far* points is much worse than three *close* data points. Despite improving, the GSJ variance estimate did a poor job at estimating the noise with few data points.

B. Lack-of-fit Convergence

The F-statistic and P-values were calculated for the previous example of fitting various degree polynomials to quadratic data with normally distributed noise. The unbiased estimate of variance $\hat{\sigma}^2$ for the model fits are seen in table 3. The F-statistic and P-value were calculated for each degree polynomial fit. The results using the GSJ method are shown in table 4 and the Hart method in table 5. In general the GSJ and Hart methods are in agreement with accepting and rejecting models based on the lack-of-fit test. Additionally the GSJ and Hart methods converge to the same F-statistic values for this example when $n = 2000$ data points are used. The only disagreement between the GSJ and Hart methods occur for the 1 degree polynomial fit

Table 1. Demonstrated convergence of the GSJ and Hart variance from noise estimation methods for the quadratic example with 10 through 20,000 data points.

Variance metric	$n = 10$	$n = 20$	$n = 200$	$n = 2000$	$n = 20,000$
GSJ $\hat{\sigma}_e^2$	271.4	177.5	27.1	23.8	25.3
0 degree Hart $\tilde{\sigma}^2$	260.1	171.4	27.0	23.8	25.3
1 degree Hart $\tilde{\sigma}^2$	37.6	60.4	25.6	23.8	25.3
2 degree Hart $\tilde{\sigma}^2$	30.4	27.1	25.3	23.9	25.3
3 degree Hart $\tilde{\sigma}^2$	25.7	27.2	25.3	23.9	25.3
4 degree Hart $\tilde{\sigma}^2$	23.2	24.8	25.3	23.9	25.3
5 degree Hart $\tilde{\sigma}^2$	25.9	25.6	25.3	23.9	25.3
True σ^2 from noise*	23.58	24.1	24.7	24.7	25.0

* True noise refers to the variance calculated from the samples pulled from the normal distribution. The normal distribution used had a $\sigma^2 = 25.0$, however the actual variance of these samples will differ when n is small.

Table 2. Demonstrated convergence of the GSJ when data points are linearly spaced from $0 \leq x_i \leq 1$.

Variance metric	$n = 10$	$n = 20$	$n = 200$	$n = 2000$	$n = 20,000$
GSJ $\hat{\sigma}_e^2$	255.6	93.8	26.0	24.6	25.2
True σ^2 from noise*	23.58	24.1	24.7	24.7	25.0

* True noise refers to the variance calculated from the samples pulled from the normal distribution. The normal distribution used had a $\sigma^2 = 25.0$, however the actual variance of these samples will differ when n is small.

when $n = 20$, in which the GSJ method accepts the model while the Hart method rejects the model. It is noted that the GSJ method in this case is on the boundary of the 95% confidence hypothesis test, and that the model would have been rejected if 99% confidence hypothesis test was used.

Table 3. Unbiased estimate of variance $\hat{\sigma}^2$ for various degree polynomial fits to the true quadratic function which included normally distributed noise. The bottom row of the table shows the variance of the normally distributed noise in the data.

Degree	$n = 10$	$n = 20$	$n = 200$	$n = 2000$	$n = 20,000$
0	2008	3307	3571	3321	3291
1	85.9	384	208	237	227
2	27.3	25.5	24.9	24.7	25.0
3	21.3	26.2	25.0	24.7	25.0
4	21.8	22.7	25.1	24.7	25.0
5	26.3	22.4	24.7	24.7	25.0
True σ^2 from noise*	23.58	24.1	24.7	24.7	25.0

* True noise refers to the variance calculated from the samples pulled from the normal distribution. The normal distribution used had a $\sigma^2 = 25.0$, however the actual variance of these samples will differ when n is small.

Table 4. GSJ estimated variance used to calculate the F-statistic and P-value for the example quadratic with 10 through 20,000 data points. P-values greater than 0.05 indicated that the data is adequately described by the degree of polynomial fit.

Degree	$n = 10$	$n = 20$	$n = 200$	$n = 2000$	$n = 20,000$
0	$F = 7.40, \mathrm{P} = 0.03$	$F = 18.6, \mathrm{P} = 0.00$	$F = 132, \mathrm{P} = 0.00$	$F = 139, \mathrm{P} = 0.00$	$F = 130, \mathrm{P} = 0.00$
1	$F = 0.30, \mathrm{P} = 0.94$	$F = 2.16, \mathrm{P} = 0.05$	$F = 7.67, \mathrm{P} = 0.00$	$F = 9.94, \mathrm{P} = 0.00$	$F = 8.96, \mathrm{P} = 0.00$
2	$F = 0.10, \mathrm{P} = 1.00$	$F = 0.14, \mathrm{P} = 1.00$	$F = 0.92, \mathrm{P} = 0.73$	$F = 1.04, \mathrm{P} = 0.22$	$F = 0.99, \mathrm{P} = 0.79$
3	$F = 0.08, \mathrm{P} = 1.00$	$F = 0.14, \mathrm{P} = 1.00$	$F = 0.92, \mathrm{P} = 0.72$	$F = 1.04, \mathrm{P} = 0.22$	$F = 0.99, \mathrm{P} = 0.79$
4	$F = 0.08, \mathrm{P} = 1.00$	$F = 0.13, \mathrm{P} = 1.00$	$F = 0.93, \mathrm{P} = 0.71$	$F = 1.04, \mathrm{P} = 0.21$	$F = 0.99, \mathrm{P} = 0.79$
5	$F = 0.10, \mathrm{P} = 0.98$	$F = 0.13, \mathrm{P} = 1.00$	$F = 0.91, \mathrm{P} = 0.74$	$F = 1.04, \mathrm{P} = 0.21$	$F = 0.99, \mathrm{P} = 0.79$

Table 5. Hart estimated variance used to calculate the F-statistic and P-value for the example quadratic with 10 through 20,000 data points. P-values greater than 0.05 indicated that the data is adequately described by the degree of polynomial fit.

Degree	$n = 10$	$n = 20$	$n = 200$	$n = 2000$	$n = 20,000$
0	$F = 7.70, \mathrm{P} = 0.02$	$F = 19.3, \mathrm{P} = 0.00$	$F = 132, \mathrm{P} = 0.00$	$F = 139, \mathrm{P} = 0.00$	$F = 130, \mathrm{P} = 0.00$
1	$F = 2.27, \mathrm{P} = 0.12$	$F = 6.37, \mathrm{P} = 0.00$	$F = 8.12, \mathrm{P} = 0.00$	$F = 9.94, \mathrm{P} = 0.00$	$F = 8.96, \mathrm{P} = 0.00$
2	$F = 0.90, \mathrm{P} = 0.55$	$F = 0.94, \mathrm{P} = 0.55$	$F = 0.98, \mathrm{P} = 0.55$	$F = 1.04, \mathrm{P} = 0.22$	$F = 0.99, \mathrm{P} = 0.79$
3	$F = 0.83, \mathrm{P} = 0.58$	$F = 0.96, \mathrm{P} = 0.53$	$F = 0.99, \mathrm{P} = 0.54$	$F = 1.04, \mathrm{P} = 0.22$	$F = 0.99, \mathrm{P} = 0.79$
4	$F = 0.94, \mathrm{P} = 0.50$	$F = 0.92, \mathrm{P} = 0.56$	$F = 0.99, \mathrm{P} = 0.52$	$F = 1.04, \mathrm{P} = 0.22$	$F = 0.99, \mathrm{P} = 0.79$
5	$F = 1.02, \mathrm{P} = 0.44$	$F = 0.87, \mathrm{P} = 0.60$	$F = 0.98, \mathrm{P} = 0.56$	$F = 1.04, \mathrm{P} = 0.21$	$F = 0.99, \mathrm{P} = 0.79$

It is important to remember that both the GSJ and Hart methods are estimations of the variance from the scatter in the data. When there is a limited number of data points, the methods may yield different results. However the example problem demonstrated that both the GSJ and Hart methods converge to the same values given a large number of data.

The F-statistic is useful in determining whether a model adequately describes data. In the quadratic example with added normally distributed noise, the F-statistic demonstrated that constant (degree 0) and linear (degree 1) models were inadequate to represent the quadratic data. The models were inadequate because a significant portion of the unbiased variance $\hat{\sigma}^2$ originated from the model's inability to fit the data and not from the inherent variance (from noise) in the data. It is worthwhile to note that the F-statistic does not provide insight about over fitting, since the polynomials with degrees 3-5 were always accepted with the F-statistic.

IV. Application to Material Parameter Identification

The F-statistic is applied to a material calibration example of Jekel et al., in which a non-linear orthotropic material model is fit to uniaxial test data using non-linear regression.[8] There are three distinctive FE models, one for the warp, fill, and 45° bias uniaxial tests. Each FE model is compared to the corresponding uniaxial experimental data. The intention is to apply the F-statistic to investigate the lack-of-fit for a practical material calibration problem. In terms of material calibration, the F-statistic and lack-of-fit tests represent a useful tool on demonstrating whether the model would improve the most from improvements to the FE model, or from reducing the variance in the experimental data.

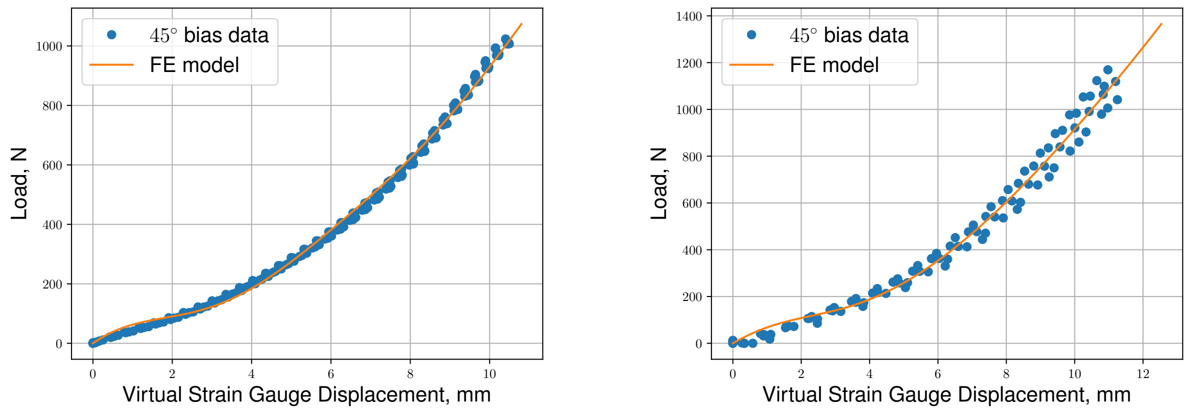
Both the GSJ and Hart methods were used to investigate the lack-of-fit. It's worthwhile to note that it is significantly simpler to apply the GSJ method, as the Hart method requires a linearized regression matrix. Since this is a non-linear regression problem, a linearized regression matrix was approximated using finite differences following Coppe et al.[5]

The lack-of-fit results are presented in table 6, where the F-statistic and P-values were calculated for the non-linear orthotropic model on two different PVC-coated polyesters. The Hart and GSJ methods were in agreement for all cases, where only the CF0700T 45° bias model was deemed inadequate with the provided data. It is interesting to note that this particular model failed the lack-of-fit test, despite being a relatively good fit to the data. For all of the other 5 examples, the lack-of-fit tests demonstrated that the models adequately described the test data. An accepted lack-of-fit test (P-value > 0.05) would indicate that the variance within the experimental data should be reduced in order to improve the overall FE model. Alternatively a rejected lack-of-fit test (P-value ≤ 0.05) would suggest improvements like using a material model that more accurately represented the experimental response.

Table 6. F-statistics and P-values (model is accepted when P > 0.05) for the non-linear orthotropic material model on two types of PVC-coated polyester.

Material	Method	Warp	Fill	45° bias
CF0700T	GSJ	$F = 0.40, \mathrm{P} = 1.00$	$F = 0.85, \mathrm{P} = 0.73$	$F = 1.59, \mathrm{P} = 0.00$
CF0700T	Hart	$F = 1.35, \mathrm{P} = 0.14$	$F = 1.20, \mathrm{P} = 0.25$	$F = 2.79, \mathrm{P} = 0.00$
VALMEX® 7318	GSJ	$F = 0.46, \mathrm{P} = 1.00$	$F = 0.98, \mathrm{P} = 0.53$	$F = 0.84, \mathrm{P} = 0.79$
VALMEX® 7318	Hart	$F = 1.28, \mathrm{P} = 0.18$	$F = 1.04, \mathrm{P} = 0.45$	$F = 0.86, \mathrm{P} = 0.77$

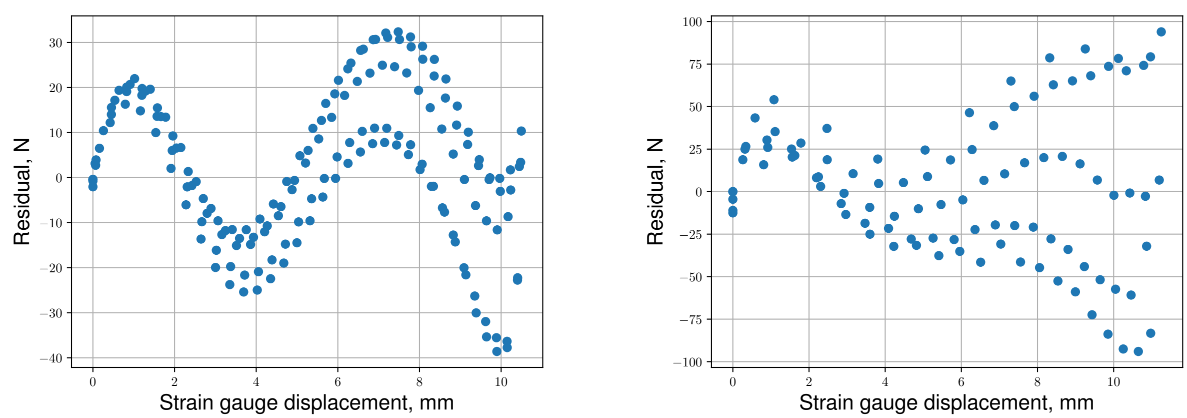
The fit of the CF0700T 45° bias test data and FE model is shown in (a) of figure 3, while (b) shows the VALMEX® 7318 45° bias fit. In both cases the FE model appears to capture the experimental response well. One way to qualitatively investigate lack-of-fit is to plot the residuals $\boldsymbol{e}$ of the model. If a model has lack-of-fit, there will be a systematic departure in the residual plot. Alternatively if a model has no lack-of-fit, the residual plot will appear as random noise. To visualize this description, the residual plots from the 45° bias test direction is seen in figure 4 for the two types of PVC-coated polyester. The CF0700T 45° bias test is shown in (a) where the systematic departure indicates that lack-of-fit is present. The VALMEX® 7318 45° bias test is shown in (b), where the inherent randomness (especially at high displacements) indicates inherent variability in the experimental data. Patterns in the residual plot can be an indication of a case where it is best to improve the model. While a zero centered residual plot with random scatter could indicate lack-of-fit.



(a) CF0700T 45° bias test (b) VALMEX® 7318 45° bias test

Figure 3. The experimental data and final FE model results. The accuracy of the FE model depends on the material model and the uncertainty in the experimental data.

The CF0700T 45° bias test was described as an exceptionally good fit because the average residual was small relative to the data values. The residuals were on the order of 20 N, while the data had a maximum value around 1200 N. This demonstrates an example where the lack-of-fit test indicates that an existing model could be improved, despite being already considered a good fit. In this case a second order polynomial was used for the shear moduli and potentially a higher order polynomial would allow the model to better match the experimental response.



(a) CF0700T 45° bias test (b) VALMEX® 7318 45° bias test

Figure 4. Residual plots of 45° bias uniaxial test for two different PVC-coated polyesters. In (a) the lack-of-fit test was failed, while (b) passed the lack-of-fit test.

The Poisson's ratio for the CF0700T PVC-coated polyester is now considered as another material parameter calibration example.[8] A linear trend was fitted to the Poisson's ratio that varied with respect to the load, as seen in figure 5. A lack-of-fit test ($F \approx 0.9$) indicates that the linear trend is an accepted model. The lack-of-fit test indicated that the major source of the variance was from the experimental data. In this case the only way to improve a trend to the linear Poisson's ratio model would be to reduce the variance in the experimental data. Attempting to use a better model for the trend, such as quadratic or cubic polynomials, will not gain much improvement in the overall accuracy. For instance the unbiased variance for the linear trend is $\hat{\sigma}^2 = 5.7 \times 10^{-}4$, quadratic trend is $\hat{\sigma}^2 = 5.7 \times 10^{-}4$, and for a cubic trend is $\hat{\sigma}^2 = 5.6 \times 10^{-}4$.

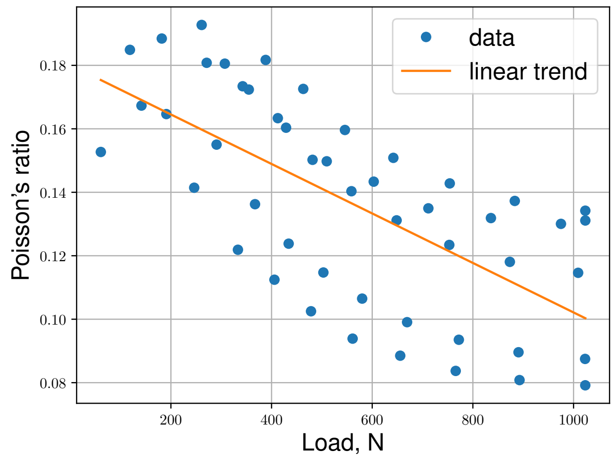


Figure 5. Poisson's ratio as a function of the applied load for CF0700T PVC-coated polyester, with a linear trend fitted to the data.

It is worthwhile to consider that for these examples, conclusions as to whether to improve the model or data could have been drawn without a lack-of-fit test using visual methods. Though the lack-of-fit test serves as a basis to aid such decisions. Additonally it may be difficult to visualize high dimensional data, and thus it is anticipated that extensions of the lack-of-fit in high dimmensions may be useful when dealing with data that is difficult to visualize. Unfortunately the GSJ and Hart methods as proposed will require modification to be extended to higher dimensions. If true replicates exist, a classical lack-of-fit test can be applied following Myers et al. regardless of the dimension of data.[2]

V. Conclusion

A generalized lack-of-fit test was described that could be applied to experimental data when calibrating a material model. The F-statistic represented a comparison of the unbiased estimate of variance to the estimated noise within the data. Two methods, GSJ and Hart, were used to approximate the variance from the noise within experimental data for cases without true replicates. The GSJ method can be directly applied to data, while the Hart method requires a linear regression model. This gives the GSJ an advantage in simplicity. Though it was shown that the Hart method can be more accurate with fewer data points, provided that the regression model used closely resembles the true function. The lack-of-fit test can be used to indicate if the model, or the data has the largest potential for improvement.

The lack-of-fit tests were applied to a material calibration problem, in which a non-linear orthotropic model was used to describe PVC-coated polyester. It is worthwhile to note that the existing model was previously considered an excellent fit. Despite this, the lack-of-fit test indicated that one of the experimental responses was inadequately described by the model. This lack-of-fit could be visualized as a systematic departure in a residual plot. An interesting take away is that even if a material model is of exceptional fit, the material calibration parameters may fail the lack-of-fit test. In this case it would be recommenced to consider improving the model. Alternatively with the Poisson's ratio example, the lack-of-fit test indicates that the uncertainty in the experimental data should be reduced in order to predict a better trend.

References

[1]Park, C., Fernandez-Godino, M. G., Kim, N. H., and Haftka, R. T., "Validation, Uncertainty Quantification and Uncertainty Reduction for a Shock Tube Simulation," *18th AIAA Non-Deterministic Approaches Conference*, AIAA SciTech Forum, American Institute of Aeronautics and Astronautics, jan 2016.

[2]Myers, R. H., Montgomery, D. C., and Anderson-Cook, C. M., *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, Wiley Series in Probability and Statistics, Wiley, 2016.

[3]Gasser, T., Sroka, L., and Jennen-Steinmetz, C., "Residual Variance and Residual Pattern in Nonlinear Regression," *Biometrika*, Vol. 73, No. 3, 1986, pp. 625–633.

[4]Hart, J. D., *Classical Lack-of-Fit Tests*, Springer New York, New York, NY, 1997, pp. 117–143.

[5]Coppe, A., Haftka, R. T., and Kim, N. H., "Uncertainty Identification of Damage Growth Parameters Using Nonlinear Regression," *AIAA Journal*, Vol. 49, No. 12, dec 2011, pp. 2818–2821.

[6]Altman, N. S. and Paulson, C. P., "Some remarks about the gasser-sroka-jennen-steinmetz variance estimator," *Communications in Statistics - Theory and Methods*, Vol. 22, No. 4, jan 1993, pp. 1045–1051.

[7]Munson, P. J. and Jernigan, R. W., "A Cubic Spline Extension of the Durbin-Watson Test," *Biometrika*, Vol. 76, No. 1, 1989, pp. 39–47.

[8]Jekel, C. F., Venter, G., and Venter, M. P., "Modeling PVC-coated polyester as a hypoelastic non-linear orthotropic material," *Composite Structures*, Vol. 161, feb 2017, pp. 51–64.