

Optimization of Regression Models of Experimental Data using Confirmation Points

N. Ulbrich*

Jacobs Technology Inc., Moffett Field, California 94035-1000

A new search metric is discussed that may be used to better assess the predictive capability of different math term combinations during the optimization of a regression model of experimental data. The new search metric can be determined for each tested math term combination if the given experimental data set is split into two subsets. The first subset consists of data points that are only used to determine the coefficients of the regression model. The second subset consists of confirmation points that are exclusively used to test the regression model. The new search metric value is assigned after comparing two values that describe the quality of the fit of each subset. The first value is the standard deviation of the PRESS residuals of the data points. The second value is the standard deviation of the response residuals of the confirmation points. The greater of the two values is used as the new search metric value. This choice guarantees that both standard deviations are always less or equal to the value that is used during the optimization. Experimental data from the calibration of a wind tunnel strain-gage balance is used to illustrate the application of the new search metric. The new search metric ultimately generates an optimized regression model that was already tested at regression model independent confirmation points before it is ever used to predict an unknown response from a set of regressors.

Nomenclature

a_1, a_2, \dots	= regression coefficients
d	= response residual of a <i>data point</i>
d'	= PRESS residual of a <i>data point</i>
i	= <i>data point</i> index
j	= <i>confirmation point</i> index
NF	= normal force of a wind tunnel balance, [lbs]
p	= number of <i>data points</i>
q	= number of <i>confirmation points</i>
$R1, R2, \dots, R6$	= electrical outputs of the strain-gages of a wind tunnel balance, [microV/V]
δ	= response residual of a <i>confirmation point</i>
σ	= standard deviation
σ_α	= standard deviation of the PRESS residuals of the <i>data points</i>
σ_β	= standard deviation of the response residuals of the <i>confirmation points</i>

I. Introduction

During the past 5 years a candidate math model search algorithm was developed for NASA Ames' Wind Tunnel Division that optimizes regression models of multivariate experimental data (see Ref. [1] and [2] for more detail). The goal of the optimization is the automated selection of a regression model, i.e., the

* Aerodynamicist, Jacobs Technology Inc.

recommended math model, for the given data set that (i) meets strict statistical quality requirements and (ii) prevents “overfitting” of the data set’s responses. These characteristics of the algorithm make it possible for an inexperienced user to apply advanced statistical metrics during a regression analysis of data that the user may not be familiar with and that were only available to a highly skilled analyst in the past.

The algorithm was originally developed for the more efficient analysis of wind tunnel strain–gage balance calibration data (see Ref. [3], [4], and [5]). It is, however, also applicable to general global regression analysis problems. The algorithm was implemented in a regression analysis software package called BALFIT that is used on a regular basis at Ames Research Center for the analysis of wind tunnel strain–gage balance calibration data and other experimental data sets.

Figure 1 depicts key elements of the candidate math model search algorithm that BALFIT uses for the regression model optimization and the identification of the recommended math model. The algorithm is discussed in great detail in Ref. [1]. In principle, it minimizes a search metric that is a measure of the predictive capability of different regression model term combinations that are compared during the search (Ref. [1]). In addition, a primary and a secondary search constraint are enforced during the optimization that make it possible to examine only those regression models that (i) have statistically significant terms and (ii) do not contain near–linear dependencies between terms. The author’s experience has shown that math term “hierarchy” is not a necessary condition for a good regression model of experimental data (see also the author’s detailed discussion of the hierarchy rule in Ref. [1], pp.6–8). Therefore, the candidate math model search algorithm allows the user to enforce math term hierarchy only as an optional constraint that may be applied after the completion of the regression model search.

A significant improvement of the candidate math model search algorithm was made in 2009: A new search metric was selected for the algorithm that uses regression model independent *confirmation points* for the first time. Therefore, the new search metric is a better indicator of the predictive capability of math term combinations that are tested during the regression model search. The new search metric will be discussed in great detail in the next sections of the paper. First, however, two previously used search metrics are reviewed in order to illustrate the problem that the new search metric is trying to solve. Then, basic elements of the new search metric are discussed. Finally, data from the calibration of a wind tunnel strain–gage balance is used to illustrate the application of the new search metric to a realistic experimental data set.

II. Previously Used Search Metrics

In the past, two alternate search metrics were used in the candidate math model search algorithm in order to identify the recommended math model. They are reviewed in some detail in this section. This review will provide a better understanding of the advantages that the recently implemented new search metric offers.

Both previously used search metrics only analyze the subset of an experimental data set that is used to obtain the regression coefficients. Points belonging to this subset may be called “data points.” They are defined as follows:

DEFINITION 1: DATA POINTS
 THE SUBSET OF A GIVEN EXPERIMENTAL DATA SET THAT IS EXCLUSIVELY
 USED TO DETERMINE THE COEFFICIENTS OF THE REGRESSION MODEL.

The first search metric, i.e., *Search Metric 1*, was introduced in the initial version of the candidate math model search algorithm in 2004 (see Ref. [3]). The metric equals the standard deviation of the response residuals of the *data points*. It can be written as follows:

$$Search\ Metric\ 1 = \sigma(\underbrace{d_1, d_2, d_3, \dots, d_i, \dots, d_{p-2}, d_{p-1}, d_p}_{response\ residuals\ of\ data\ points}) \quad (1)$$

where d_i equals the response residual (i.e., the difference between the fitted and measured response) of a *data point*, i is the *data point* index, and p is the total number of *data points*. *Search Metric 1* has a major disadvantage: it only assesses the predictive capability of a tested regression model at points that are used to determine the regression model. In other words – the regression model is not tested at *confirmation points*, i.e., points that are independent of the regression model coefficients.

A second search metric, i.e., *Search Metric 2*, was introduced in 2007 that tries to address the shortcoming of *Search Metric 1* (see Ref. [2]). At that time the author realized that the standard deviation of the PRESS residuals of the *data points* would be a better search metric as the PRESS residual of a *data point* is computed by using the *data point* as a *confirmation point*. In addition, PRESS residuals are explicitly recommended in the literature for the comparison of the predictive capability of different regression models (see Ref. [6], p.142). *Search Metric 2* can be summarized as follows

$$\text{Search Metric 2} = \sigma(\underbrace{d'_1, d'_2, d'_3, \dots, d'_i, \dots, d'_{p-2}, d'_{p-1}, d'_p}_{\text{PRESS residuals of data points}}) \quad (2)$$

where d'_i equals the PRESS residual of the *data point*, i is the *data point* index, and p is the total number of *data points*. The calculation of the PRESS residual of a *data point* requires several steps. The PRESS residual is evaluated by removing the *data point* from the set of *data points*, fitting a regression model to the remaining *data points*, and testing the regression model at the omitted *data point*. Therefore, the standard deviation of the PRESS residuals is a metric that uses each *data point* as a *confirmation point*, i.e., as a point that is not used to generate the regression model when its predictive capability is tested.

Search Metric 2, however, also has a disadvantage. All *data points* must be used to develop the final regression model of the data set after the completion of the candidate math model search. No *data point* can be omitted. Therefore, *Search Metric 2* is really a compromise. Initially, *data points* are used as *confirmation points* in order to test the regression model. Afterwards, however, all *data points* are needed in order to compute the final values of the regression model coefficients. The author also observed that the difference between the standard deviation of the PRESS residuals of the *data points* (*Search Metric 2*) and the standard deviation of the response residuals of the *data points* (*Search Metric 1*) continues to become smaller as the number of *data points* is increased. In other words – *Search Metric 2* gradually loses its diagnostic advantage as the number of *data points* increases. An alternate search metric had to be found that would process independent sets of *data points* and *confirmation points*. Ultimately, this conclusion led to the development of *Search Metric 3* that will be discussed in the next section of the paper.

III. New Search Metric

In 2009 the author developed, implemented, and extensively tested a new search metric, i.e., *Search Metric 3*, in order to address the shortcomings of *Search Metric 1* and *Search Metric 2* that were discussed above. The new search metric was developed after concluding that the *data points* required to determine the regression model coefficients are as important as the *confirmation points* used to test the regression model.

It is necessary to define the term “*confirmation point*” more precisely before *Search Metric 3* can be explained in detail. Often, *data points* of an experimental data exist and no *confirmation points* are available. Then, in order to be able to apply *Search Metric 3*, two options exist that may be used to obtain *confirmation points*. Option 1: Additional *confirmation points* are acquired that do not match any of the *data points*. Option 2: The original experimental data set is split into two subsets; the first subset contains the *data points* that are used to determine the regression coefficients; the second subset contains the *confirmation points* that are used to test the regression model. In many situations it is difficult to apply Option 1. Then, Option 2 is the only way to obtain a suitable set of *confirmation points*. These *confirmation points* may be defined as follows:

DEFINITION 2: CONFIRMATION POINTS

THE SUBSET OF A GIVEN EXPERIMENTAL DATA SET THAT IS EXCLUSIVELY USED TO TEST THE PREDICTIVE CAPABILITY OF A REGRESSION MODEL. THE COEFFICIENTS OF THE REGRESSION MODEL ARE INDEPENDENT OF THE CHOSEN CONFIRMATION POINT SET.

The basic idea behind *Search Metric 3* can be understood if an “ideal” and “overfitted” regression analysis result are compared. Let us assume, for example, that the “dependent variable,” i.e., the response, of some experimental data is described using a polynomial of a single “independent variable”. Then, an “ideal” least squares fit result may look like the one depicted in Fig. 2. The response residuals of the *data points* should have a magnitude that is very similar to the magnitude of the response residuals of the *confirmation points*. This definition of an “ideal” regression analysis result can be expressed using the standard deviations of the corresponding residuals. We can write

DEFINITION 3: “IDEAL” REGRESSION ANALYSIS RESULT

$$\underbrace{\sigma(d_1, d_2, d_3, \dots, d_i, \dots, d_{p-2}, d_{p-1}, d_p)}_{\text{response residuals of data points}} \approx \underbrace{\sigma(\delta_1, \delta_2, \delta_3, \delta_4, \dots, \delta_j, \dots, \delta_{q-2}, \delta_{q-1}, \delta_q)}_{\text{response residuals of confirmation points}} \quad (3a)$$

where d_i is a response residual of a *data point* with index i and δ_j is a response residual of a *confirmation point* with index j . Figure 3 shows an “overfitted” regression analysis result of the same experimental data set. Now, the response residuals of the *data points* have a magnitude that is significantly smaller than the magnitude of the response residuals of the *confirmation points*. This definition of an “overfitted” result can be expressed using the standard deviations of the corresponding residuals. Then, we can write:

DEFINITION 4: “OVERFITTED” REGRESSION ANALYSIS RESULT

$$\underbrace{\sigma(d_1, d_2, d_3, \dots, d_i, \dots, d_{p-2}, d_{p-1}, d_p)}_{\text{response residuals of data points}} \ll \underbrace{\sigma(\delta_1, \delta_2, \delta_3, \delta_4, \dots, \delta_j, \dots, \delta_{q-2}, \delta_{q-1}, \delta_q)}_{\text{response residuals of confirmation points}} \quad (3b)$$

The standard deviation of the response residuals of the *data points* is often very close to the standard deviation of the PRESS residuals of the *data points*. In addition, both values depend on the *data points*. Therefore, the standard deviation of the response residuals of the *data points* may be substituted by the standard deviation of the PRESS residuals of the *data points*. We get the relationship:

$$\underbrace{\sigma(d_1, d_2, d_3, \dots, d_i, \dots, d_{p-2}, d_{p-1}, d_p)}_{\text{response residuals of data points}} \approx \underbrace{\sigma(d'_1, d'_2, d'_3, \dots, d'_i, \dots, d'_{p-2}, d'_{p-1}, d'_p)}_{\text{PRESS residuals of data points}} \quad (4)$$

The suggested substitution will also make it possible to define *Search Metric 3* such that it agrees with *Search Metric 2* if no *confirmation points* are available for analysis.

At this point a question emerges: How can the results presented in Fig. 2 and Fig. 3 be used to define a new search metric for the candidate math model search? It is obvious, after comparing the two figures, that “overfitting” can be avoided if (i) the standard deviation of the PRESS residuals of the *data points* and (ii) the standard deviation of the response residuals of the *confirmation points* are of similar magnitude. The first value, i.e., the standard deviation of the PRESS residuals of the *data points*, assesses the quality of the tested regression model at the *data points*. The second value, i.e., the standard deviation of the response residuals of the *confirmation points*, assess the quality of the tested regression model at the *confirmation points*. The “ideal” regression model of the *data points* should neither favor the quality of the fit at the *data*

points nor the quality of the fit at the *confirmation points*. This behavior can be guaranteed if the maximum of the two values is selected as the search metric. Then, the new metric can be defined as follows:

$$\text{Search Metric 3} = \mathcal{MAX}\{\sigma_\alpha, \sigma_\beta\} \quad (5a)$$

where

$$\sigma_\alpha = \sigma(\underbrace{d'_1, d'_2, d'_3, \dots, d'_i, \dots, d'_{p-2}, d'_{p-1}, d'_p}_{\text{PRESS residuals of data points}}) \quad (5b)$$

$$\sigma_\beta = \sigma(\underbrace{\delta_1, \delta_2, \delta_3, \delta_4, \dots, \delta_j, \dots, \delta_{q-2}, \delta_{q-1}, \delta_q}_{\text{response residuals of confirmation points}}) \quad (5c)$$

where σ_α equals the standard deviation of the PRESS residuals of the *data points* and σ_β equals the standard deviation of the response residuals of the *confirmation points*. Figure 4 summarizes the differences between *Search Metric 1*, *Search Metric 2*, and *Search Metric 3*. In addition, Fig. 5 lists the individual steps that are needed in order to determine *Search Metric 3*.

Search Metric 3, similar to *Search Metric 1* and *Search Metric 2*, uses the standard deviation of a given set of residuals in order to define a search metric for the candidate math model search. Therefore, a minimum number of *data points* and *confirmation points* is required in order to compute the search metric. The number of *data points* is usually sufficiently large (e.g., > 20) so that the standard deviation of the PRESS residuals of the *data points* can be determined. The number of available *confirmation points*, on the other hand, may be limited. Therefore, *Search Metric 3* was implemented in the BALFIT software package such that it is only used if the number of *confirmation points* is at least 20. This threshold choice follows recommendations that are made in the literature regarding the minimum number of *confirmation points* (see discussion in Ref. [6], p.308).

In the next section of the paper experimental data from a calibration of NASA's MC60D wind tunnel strain-gage balance will be used (i) to demonstrate the application of the candidate math model search algorithm's new search metric to a realistic data set and (ii) to compare characteristics of the new search metric with those of a previously applied search metric.

IV. Application of New Search Metric to Wind Tunnel Balance Data

A. Data Description

A calibration data set of NASA's MC60D strain-gage balance was selected for the present study (i) to demonstrate the application of *Search Metric 3* to a realistic data set and (ii) to compare the optimization results for the previously used *Search Metric 2* with the results for *Search Metric 3*. The use of a strain-gage balance needs to be explained in more detail in order to better understand the regression analysis results that are discussed in this section. A strain-gage balance is a measuring device that is used in wind tunnel testing. It makes the accurate measurement of a wind tunnel model's aerodynamic loads possible. These aerodynamic forces and moments are obtained after processing the measured electrical outputs of the strain-gages of the balance using a multivariate regression model of the balance characteristics. This regression model is the result of the regression analysis of high-precision calibration data that relates a set of known balance loads to corresponding measured electrical outputs of the gages. The MC60D balance is a modern high capacity force type balance that was manufactured by Triumph (Force Measurement Systems) in 2008. The balance calibration data was obtained in Triumph's ABCS calibration machine. This calibration machine makes it possible to obtain data for a complete description of the physical behavior of the balance. Six balance loads and electrical outputs of the gages were recorded for each point that was taken during the calibration. The final calibration data set consisted of 1906 individual loadings of the balance. An initial analysis of the data showed that the selected balance calibration data set is of excellent quality.

For the present study it was decided to process the data in “direct read format.” Consequently, the original calibration loads were converted from “force balance format” (forward normal force, aft normal force, forward side force, aft side force, rolling moment, axial force) to “direct read format” (normal force, pitching moment, side force, yawing moment, rolling moment, and axial force) using the load transformation equations that are listed in Ref. [6]. The electrical outputs of the gages were already corrected for the tare weight of the calibration machine hardware and the balance shell. Therefore, they could be used for the regression analysis without further modifications.

B. Iterative versus Non-Iterative Methods

In the aerospace testing community both iterative and non-iterative methods are used for the regression analysis of strain-gage balance calibration data. The iterative method uses the balance loads as “independent variables” and the electrical outputs of the gages as “dependent variables” in order to define the regressors and responses for the regression analysis. Consequently, the iterative method fits the electrical outputs of the gages as a function of the balance loads and then uses an iteration scheme in order to relate the measured electrical outputs to balance loads during a wind tunnel test (see Refs. [3], [4], and [7] for more detail). The non-iterative method, on the other hand, uses the electrical outputs of the gages as “independent variables” and the balance loads as “dependent variables” in order to define the regressors and responses for the regression analysis. This method is identical with a classical regression analysis of balance calibration data that directly fits each balance load as a function of the electrical outputs of the gages.

Studies of the author showed that the iterative and the non-iterative method have the same balance load prediction accuracy as long as (i) the balance calibration experiment is well designed and (ii) the regression models of the calibration data meet strict statistical quality requirements. Therefore, in order to simplify the discussion of the application of *Search Metric 3* to the chosen experimental data set, it was decided (i) to select the non-iterative method for the regression analysis and (ii) to optimize the regression model of only the normal force component. Consequently, the normal force (NF) was treated as the “dependent variable,” i.e., as the “response,” and the electrical outputs of the gages ($R1, R2, \dots, R6$) were treated as “independent variables,” i.e., as variables that define the “regressors” for the regression analysis.

C. Regression Analysis of Normal Force Data

The regression analysis and regression model optimization of the normal force data of the balance was performed for two different data set examples. This choice made it possible to compare the predictive capability of two optimized regression models that are the result of using either *Search Metric 2* or *Search Metric 3*. Table 1 below lists key features of the two selected examples that were processed using the regression model optimization process. The threshold values for the primary and secondary search constraints as listed as well. The optional math term hierarchy constraint was not applied.

Table 1: Description of Data Analysis Examples.

	EXAMPLE 1	EXAMPLE 2
DEPENDENT VARIABLE (RESPONSE)	NF	NF
INDEPENDENT VARIABLES (DEFINE REGRESSORS)	$R1, R2, \dots, R6$	$R1, R2, \dots, R6$
NUMBER OF DATA POINTS (p)	1906	201
NUMBER OF CONFIRMATION POINTS (q)	0	1705
SEARCH METRIC USED FOR OPTIMIZATION	<i>Search Metric 2</i>	<i>Search Metric 3</i>
PRIMARY CONSTRAINT (P-VALUE OF T-STATISTIC)	< 0.0001	< 0.0001
SECONDARY CONSTRAINT (VARIANCE INFLATION FACTOR)	< 5	< 5
HIERARCHY CONSTRAINT (OPTIONAL)	<i>not applied</i>	<i>not applied</i>

The first example, i.e., *Example 1*, was processed assuming that no *confirmation points* were available for the regression model optimization. Therefore, all 1906 calibration points of the experimental data set were considered to be *data points* and only *Search Metric 2* could be used during the optimization.

The second example, i.e., *Example 2* was constructed so that the new *Search Metric 3* could be applied.

Therefore, the original experimental data set of 1906 calibration points was split into two subsets. The split was performed by randomly selecting about 10% of the calibration points, i.e., 201 points, and assigning them to be *data points*. Sufficient information about the normal force characteristics was contained in the randomly chosen *data points* so that a meaningful regression analysis of the normal force data was possible. The remaining 90% of the calibration points, i.e., 1705 points, were assigned to be *confirmation points*.

A multivariate quadratic was chosen to be the upper bound of all math models that were investigated during the optimization of the two regression models. This upper bound has the following form:

$$NF = a_1 + a_2 \cdot R1 + a_3 \cdot R2 + \cdots + a_7 \cdot R6 + a_8 \cdot (R1)^2 + a_9 \cdot (R2)^2 + \cdots + a_{13} \cdot (R6)^2 + a_{14} \cdot (R1 \cdot R2) + a_{15} \cdot (R1 \cdot R3) + \cdots + a_{28} \cdot (R5 \cdot R6) \quad (6)$$

Figure 7a shows the independent variables, i.e., the electrical outputs of the strain-gages, that were used for the regression model optimization of *Example 1*. A total number of 1906 *data points* was used for the calculation of the regression coefficients. Figure 7b shows the dependent variables (responses), i.e., the normal forces, that were fitted during the regression model optimization of *Example 1*. Figure 7c depicts *Search Metric 2* for *Example 1* as a function of the number of regression model terms. This metric was minimized during the optimization. Figures 7d and 7e show the p-value of t-statistic maximum (primary search constraint) and the variance inflation factor maximum (secondary search constraint) as a function of the regression model terms. The most conservative thresholds (< 0.0001 and < 5) were chosen for the constraints during the optimization. Table 2 below shows the terms of the recommended math model that was obtained for *Example 1*. This optimized regression model consists of a total of 24 terms.

Table 2: Recommended Math Model Terms for Example 1 and Example 2.

INDEX	MATH TERM	EXAMPLE 1	EXAMPLE 2
1	intercept (constant)	×	×
2	$R1$	×	×
3	$R2$	×	×
4	$R3$	×	×
5	$R4$	×	×
6	$R5$	×	×
7	$R6$	×	×
8	$(R1)^2$	×	×
9	$(R2)^2$	×	×
10	$(R3)^2$	×	–
11	$(R4)^2$	–	–
12	$(R5)^2$	×	×
13	$(R6)^2$	×	–
14	$(R1 \cdot R2)$	×	–
15	$(R1 \cdot R3)$	–	–
16	$(R1 \cdot R4)$	×	–
17	$(R1 \cdot R5)$	×	×
18	$(R1 \cdot R6)$	–	–
19	$(R2 \cdot R3)$	×	–
20	$(R2 \cdot R4)$	×	–
21	$(R2 \cdot R5)$	×	×
22	$(R2 \cdot R6)$	×	–
23	$(R3 \cdot R4)$	×	–
24	$(R3 \cdot R5)$	×	×
25	$(R3 \cdot R6)$	×	–
26	$(R4 \cdot R5)$	×	×
27	$(R4 \cdot R6)$	–	–
28	$(R5 \cdot R6)$	×	–

Figure 7f shows the analysis of variance results for the recommended math model of *Example 1*. The analysis of variance results and other metrics depicted in Fig. 7f indicate that the recommended math model met all statistical quality requirements that were imposed during the optimization process. Figure 7g shows the response (normal force) residuals of the *data points*, i.e., the difference between the fitted and applied normal force, as a function of the applied normal force. The standard deviation of the response residuals is only a very small percentage (0.0646%) of the largest normal force magnitude that was applied during the balance calibration. Therefore, the 24 term recommended math model of *Example 1* is a good characterization of the expected behavior of normal force of the balance during a wind tunnel test.

Figure 8a shows the independent variables, i.e., the electrical outputs of the strain-gages, that were used for the regression model optimization of *Example 2*. This time, only the 201 randomly selected *data points* were used to calculate the regression coefficients. Figure 8b shows the dependent variables (responses), i.e., the normal forces, that were fitted during the regression model optimization of *Example 2*. Figure 8c depicts *Search Metric 3* for *Example 2* as a function of the number of math terms. This metric was minimized during the search. Figures 8d and 8e show the p-value of t-statistic maximum (primary search constraint) and the variance inflation factor maximum (secondary search constraint) as a function of the number of math terms for *Example 2*. Table 2 above lists terms of the recommended math model that was obtained for *Example 2*. This time, the optimized regression model consists of only 14 terms, i.e., 10 fewer terms than the optimized model of *Example 1*.

Figure 8f shows the analysis of variance results for the recommended math model of *Example 2*. The analysis of variance results and other metrics shown in Fig. 8f indicate that the recommended math model of *Example 2* met all statistical quality requirements that were imposed during the optimization process. Figure 8g shows the response (normal force) residuals of the *data points*, i.e., the difference between the fitted and applied normal force, as a function of the applied normal force. The standard deviation of the response residuals of the *data points* is only a very small percentage (0.0718%) of the largest normal force magnitude that was applied during the balance calibration. Figure 8h shows the response (normal force) residuals of the *confirmation points* as a function of the applied normal force. Again, as the 1705 *confirmation points* were used during the optimization to independently test the predictive capability of the processed math term combinations, the standard deviation of the response residuals of the *confirmation points* is only a very small percentage (0.0765%) of the largest normal force magnitude that was applied during the balance calibration. Both the standard deviation of the response residuals of the *data points* and the standard deviation of the response residuals of the *confirmation points* are of about the same magnitude and very small. Therefore, the 14 term recommended math model of *Example 2* is also an excellent characterization of the expected behavior of normal force of the balance during a wind tunnel test.

The hierarchy rule was not explicitly applied during the optimization of the two regression models (see Table 1). However, it is interesting to note that the optimized models listed in Table 2, Fig. 7f, and Fig. 8f turned out to be both hierarchical. Therefore, it can be concluded that the original calibration data set of the MC60D balance contains data that supports a hierarchical regression model.

A question remains: Which one of the two recommended math models of the calibration data is expected to have better predictive capabilities? Several observations can be made after comparing the two recommended math models. – *Observation 1*: Comparing the standard deviation of the response residuals of the 1906 *data points* that were used to determine the recommended math model of *Example 1* (0.0646%, see Fig. 7g) with the standard deviation of the response residuals of the 201 *data points* that were used to determine the recommended math model of *Example 2* (0.0718%, see Fig. 8g) we see that the difference of the values is 0.0072%. – *Observation 2*: Comparing the standard deviation of the response residuals of the 1906 *data points* that were used to determine the recommended math model of *Example 1* (0.0646%, see Fig. 7g) with the standard deviation of the response residuals of the 1706 *confirmation points* that were used to test the recommended math model of *Example 2* (0.0765%, see Fig. 8h) we see that the difference of the values is 0.0119%. – *Observation 3*: The recommended math model of *Example 2* was obtained using only about 10% of the *data points* that were required to obtain the recommended math model of *Example 1*. – *Observation 4*: The recommended math model of *Example 2* uses only 14 of the 24 math terms that the recommended math model of *Example 1* uses. – *Observation 5*: The recommended math model of *Example 1* was not tested at regression coefficient independent *confirmation points*.

It is concluded, after reviewing the five observations, that the difference between the standard deviations

of the two recommended math models is relatively small ($\approx 0.01\%$ of the largest normal force magnitude). This observation is remarkable considering the significant differences in (i) the number of *data points* that were used to develop the two regression models and (ii) the number of math terms of the two models. Both math models appear to have similar predictive capabilities. The recommended math model of *Example 2*, however, has the advantage that it was already successfully tested at regression coefficient independent *confirmation points*. Therefore, it appears that the recommended math model of *Example 2* is the better and more reliable regression model for the given calibration data of the MC60D balance.

D. Tare Corrected Confirmation Point Loads

Tare corrections are used in strain-gage balance calibration data analysis in order to correct the original balance calibration data for the effects of the weight of the calibration hardware and the balance shell. The manufacturer of the MC60D balance supplied the calibration data set in “tare corrected” format. Consequently, the data could directly be used for the regression analysis example that is discussed above. Often, however, balance calibration data is provided for analysis that still needs to be tare corrected. Different approaches are used in the aerospace testing community to estimate these tare corrections. One approach, the so-called “tare load iteration,” estimates tare corrections for the given balance calibration loads (see Ref. [3], pp.13–14, or Ref. [7], p.17 for a discussion of the tare load iteration process). Now a question must be asked: What impact do these tare corrections have on *confirmation points* that may be obtained after splitting an uncorrected balance calibration data set into a subset of *data points* and a subset of *confirmation points*? In general, tare corrections have to be computed for each balance calibration load series assuming that the weight of the calibration fixtures changes from load series to load series. Some points of a load series may belong to the subset of *data points*. Other points of the same load series may belong to the subset of *confirmation points*. All points of a load series, however, have the same tare load corrections. Therefore, the tare load corrections need to be estimated using the *data points* of the load series and afterwards applied to both the loads of the *data points* and the loads of the *confirmation points* before the calibration data set can be processed using the new search metric and the regression model optimization algorithm.

V. Summary and Conclusions

A regression model optimization algorithm is currently being used at NASA Ames’ Wind Tunnel Division that identifies an optimized regression model, i.e., the so-called recommended math model, for a given multivariate experimental data set. This optimized regression model is designed (i) to meet strict statistical quality requirements and (ii) to prevent “overfitting” of the experimental data set’s responses. A new search metric was implemented in the optimization algorithm in 2009. This new search metric simultaneously uses *data points* and *confirmation points* in order to better assess and compare the predictive capability of different math term combinations that are tested during the optimization.

Experimental data from a calibration of NASA’s MC60D wind tunnel strain-gage balance was used in the present study to illustrate the application of the new search metric during the regression model optimization of a realistic data set. The optimization results for the new search metric were compared with results that were obtained for a previously used search metric. This comparison showed that the new search metric has the ability to generate a math model that is expected to have better predictive capabilities than an optimized regression model that is obtained using a *confirmation point* independent search metric.

It is interesting to note that all original points of a given experimental data set are used during the optimization of the regression model if the new search metric is applied. None of the information is discarded that may be contained in the experimental data set. Only a portion of the experimental data set, however, is needed to determine the coefficients of the regression model. The new search metric ultimately generates an optimized regression model of the experimental data set that was already tested at regression coefficient independent *confirmation points* before it is ever used to predict an unknown response from a set of regressors.

VI. Acknowledgements

The author would like to thank Tom Volden of Jacobs Technology and Jon Bader of NASA Ames Research Center for their critical and constructive review of the final manuscript. Thanks also go to Ray

Rhew of NASA Langley Research Center for providing the balance calibration data that is used in the examples. The work reported in this paper was partially supported by NASA's Aeronautic Test Program and the Wind Tunnel Division at Ames Research Center under contract NNA09DB39C.

VII. References

¹Ulbrich, N., "Regression Model Optimization for the Analysis of Experimental Data," AIAA 2009-1344, paper presented at the 47th AIAA Aerospace Sciences Meeting and Exhibit, Orlando, Florida, January 2009.

²Ulbrich, N. and Volden, T., "Regression Analysis of Experimental Data Using an Improved Math Model Search Algorithm," AIAA 2008-0833, paper presented at the 46th AIAA Aerospace Sciences Meeting and Exhibit, Reno, Nevada, January 2008.

³Ulbrich, N. and Volden, T., "Strain-Gage Balance Calibration Analysis Using Automatically Selected Math Models," AIAA 2005-4084, paper presented at the 41st AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit, Tucson, Arizona, July 2005.

⁴Ulbrich, N. and Volden, T., "Application of a New Calibration Analysis Process to the MK-III-C Balance," AIAA 2006-0517, paper presented at the 44th AIAA Aerospace Sciences Meeting and Exhibit, Reno, Nevada, January 2006.

⁵Ulbrich, N., and Bader, J., "Analysis of Sting Balance Calibration Data Using Optimized Regression Models," AIAA 2009-5372, paper presented at the 45th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit, Denver, Colorado, August 2009, pp.18-21.

⁶Montgomery, D. C., Peck, E. A., and Vining, G. G., *Introduction to Linear Regression Analysis*, 4th ed., John Wiley & Sons, Inc., New York, 2006, p. 142, p. 308.

⁷American Institute of Aeronautics and Astronautics, "Recommended Practice - Calibration and Use of Internal Strain Gage Balances with Application to Wind Tunnel Testing," AIAA R-091-2003, American Institute of Aeronautics and Astronautics, Reston, Virginia, 2003.

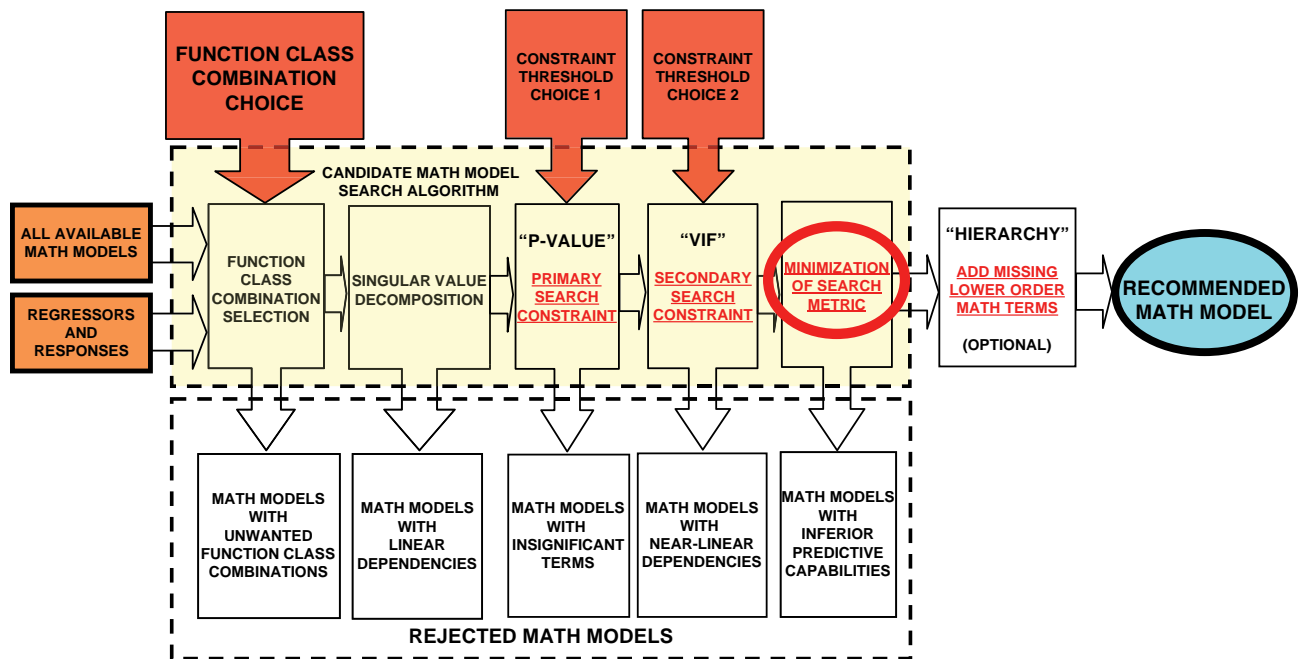


Fig. 1 Key elements of candidate math model search algorithm.

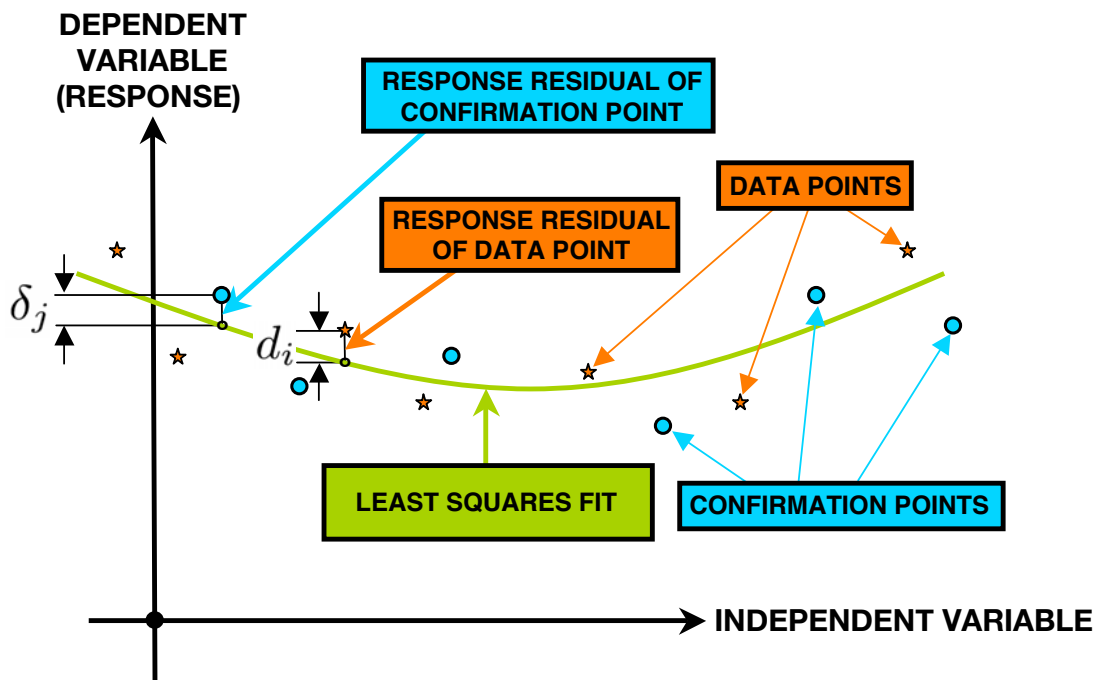


Fig. 2 “Ideal” regression analysis result of experimental data.

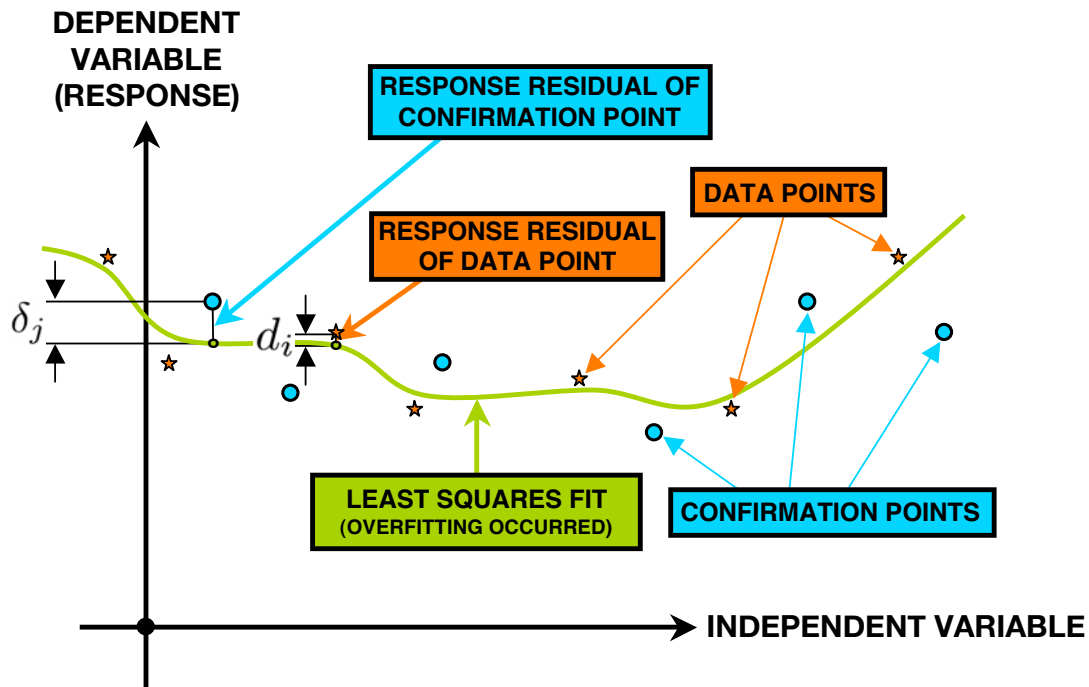
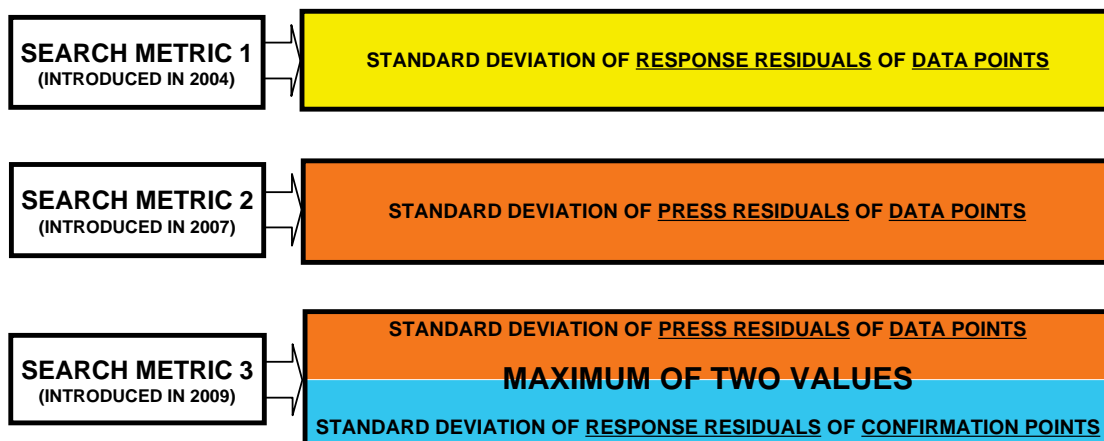


Fig. 3 “Overfitted” regression analysis result of experimental data.



DATA POINTS ==> POINTS THAT ARE USED TO DETERMINE THE REGRESSION MODEL COEFFICIENTS
 CONFIRMATION POINTS =====> POINTS THAT ARE USED TO TEST THE REGRESSION MODEL
 REQUIREMENT ==> A CONFIRMATION POINT CANNOT BELONG TO THE CHOSEN SET OF DATA POINTS

Fig. 4 Search metric options for a candidate math model search.

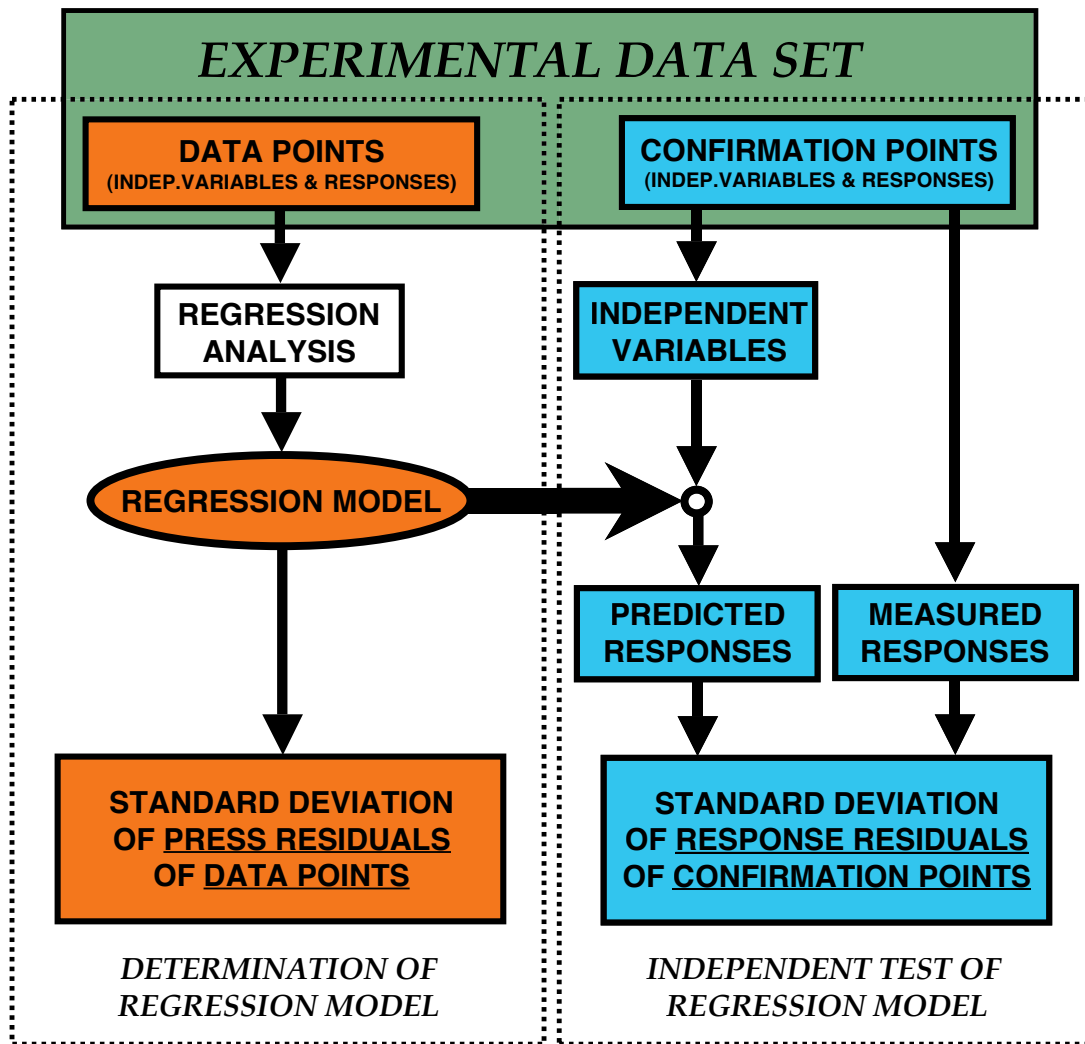


Fig. 5 Determination of *Search Metric 3*.

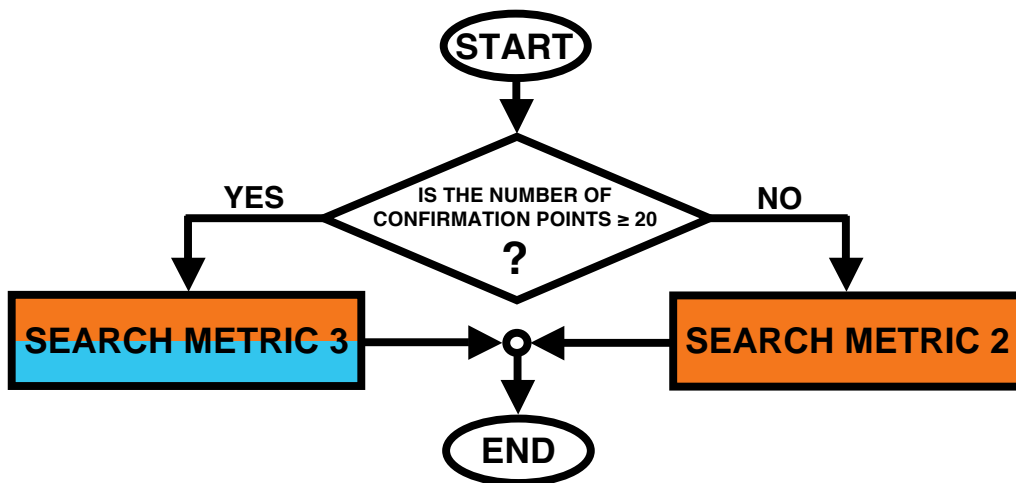


Fig. 6 Implementation of *Search Metric 3* in candidate math model search algorithm.

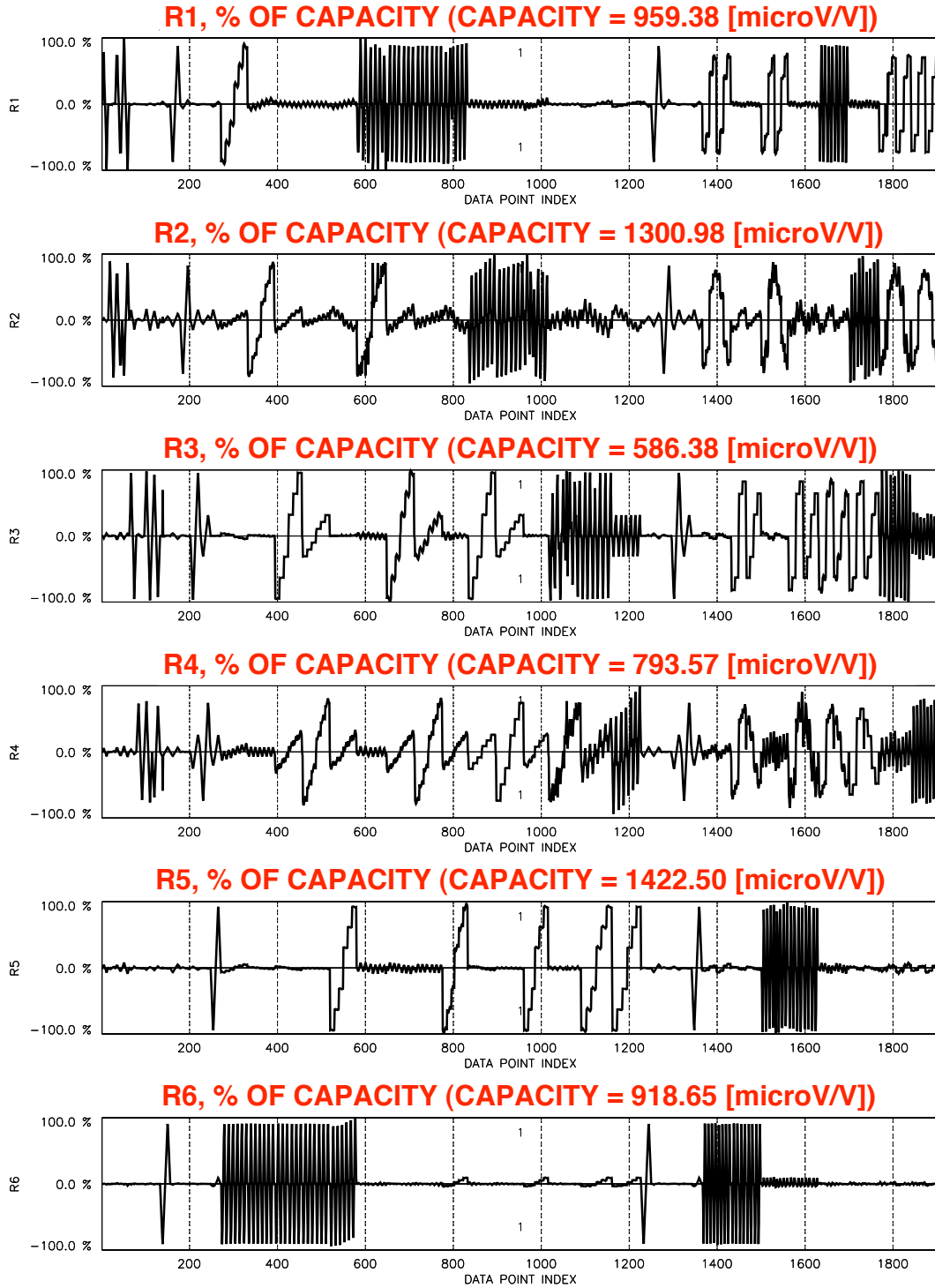


Fig. 7a Example 1: Electrical outputs of strain-gages versus *data point index* (1906 *data points*).

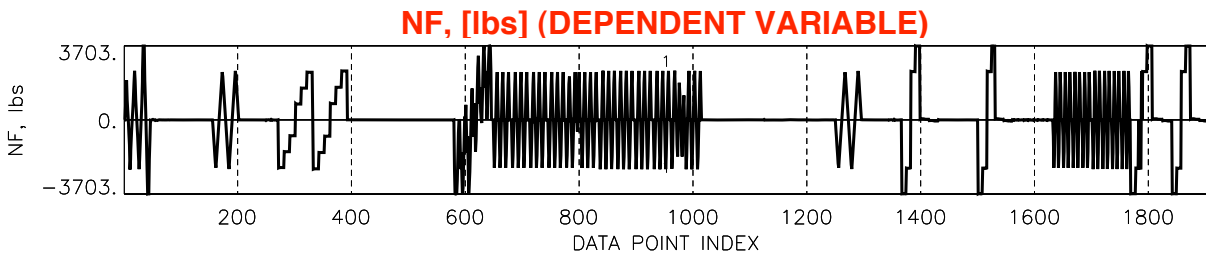


Fig. 7b Example 1: Normal forces versus *data point* index (1906 *data points*).

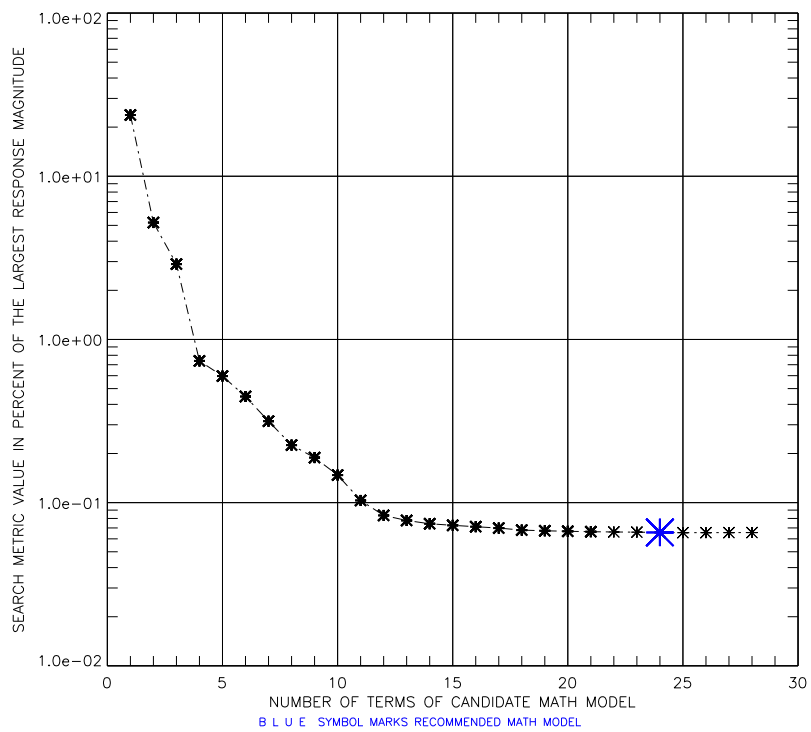


Fig. 7c Example 1: *Search Metric 2* versus number of regression model terms.

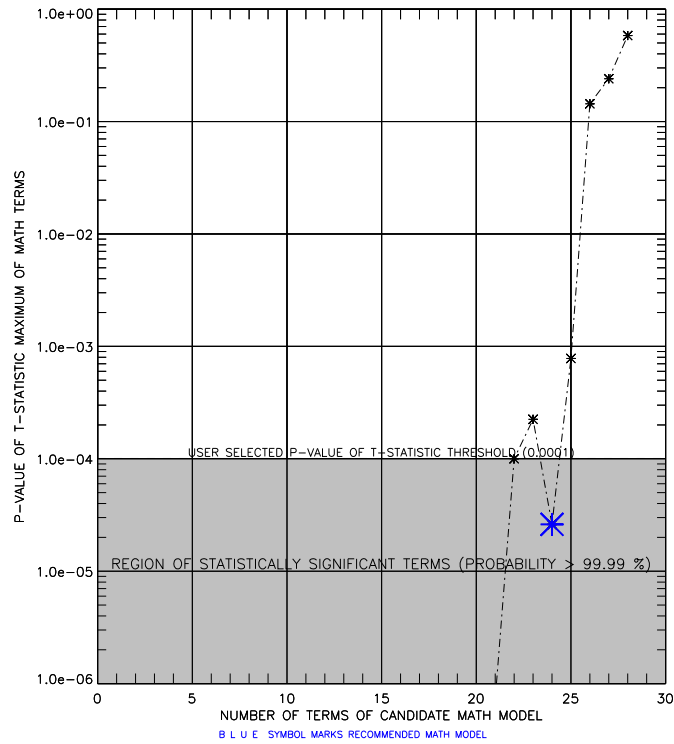


Fig. 7d Example 1: Primary search constraint versus the number of regression model terms.

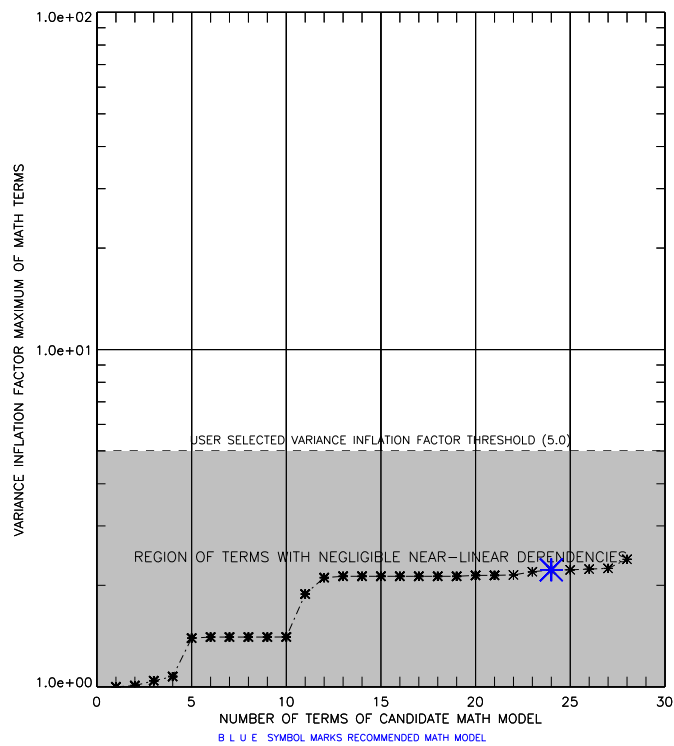


Fig. 7e Example 1: Secondary search constraint versus the number of regression model terms.

SOURCE OF VARIATION	SUM OF SQUARES	PRESS STATISTIC	DEGREES OF FREEDOM	MEAN SQUARE	F-VALUE OF REGRESSION	P-VALUE OF REGRESSION	
REGRESSION	3.3283e+09	—	23	1.4471e+08	2.4996e+07	< 0.0001	
RESIDUAL	10895.5575	11261.8144	1882	5.7894	—	—	
TOTAL	3.3283e+09	—	1905	—	—	—	
R-SQUARE		ADJ. R-SQUARE		PRESS R-SQUARE			
0.999997		0.999997		0.999997			
REGRESSION COEFFICIENT ESTIMATES AND STATISTICAL METRICS (NF)							
REGRESSION MODEL HIERARCHY CHARACTERISTICS = HIERARCHICAL							
MATH TERM INDEX	MATH TERM NAME	COEFFICIENT VALUE	STANDARD ERROR	T-STATISTIC OF COEFFICIENT	P-VALUE OF COEFFICIENT	VIF (PRIMARY)	VIF (ALTERNATE)
1	INTERCEPT	+0.9155	+0.0889	+10.3006	—	—	—
2	R1	+2.4968	+0.0002	+14853.3833	< 0.0001	+1.0533	+1.0420
3	R2	+2.1666	+0.0001	+16749.6969	< 0.0001	+1.1087	+1.0778
4	R3	-0.6736	+0.0003	-2430.4900	< 0.0001	+1.3807	+1.3489
5	R4	-0.0668	+0.0003	-261.9374	< 0.0001	+1.4132	+1.4071
6	R5	+0.2393	+0.0001	+1861.5538	< 0.0001	+1.1389	+1.1551
7	R6	-0.0564	+0.0002	-260.1130	< 0.0001	+1.0256	+1.0178
8	R1*R1	-2.4196e-05	+2.9331e-07	-82.4904	< 0.0001	+1.1562	+1.1556
9	R2*R2	-1.3850e-05	+1.7885e-07	-77.4409	< 0.0001	+1.1648	+1.1624
10	R3*R3	+5.2345e-06	+7.7884e-07	+6.7209	< 0.0001	+1.5608	+1.5606
12	R5*R5	-1.4277e-05	+1.6390e-07	-87.1061	< 0.0001	+2.1499	+2.1513
13	R6*R6	+2.1375e-06	+3.2657e-07	+6.5453	< 0.0001	+1.0379	+1.0318
14	R1*R2	+2.6057e-06	+2.8847e-07	+9.0328	< 0.0001	+1.0899	+1.0875
16	R1*R4	-3.5266e-06	+8.2062e-07	-4.2975	< 0.0001	+1.1426	+1.1312
17	R1*R5	-7.8094e-06	+4.7904e-07	-16.3023	< 0.0001	+1.1738	+1.1736
19	R2*R3	-8.9789e-06	+7.9822e-07	-11.2487	< 0.0001	+1.6778	+1.6735
20	R2*R4	+3.1959e-06	+7.5812e-07	+4.2156	< 0.0001	+1.7507	+1.7273
21	R2*R5	+4.1914e-05	+3.8570e-07	+108.6703	< 0.0001	+1.4370	+1.4386
22	R2*R6	-6.3335e-06	+5.6316e-07	-11.2463	< 0.0001	+1.1012	+1.0851
23	R3*R4	-4.4691e-06	+9.0191e-07	-4.9551	< 0.0001	+1.5095	+1.5030
24	R3*R5	+2.8673e-05	+7.9996e-07	+35.8426	< 0.0001	+1.4654	+1.4658
25	R3*R6	+1.2271e-05	+1.0648e-06	+11.5240	< 0.0001	+1.0578	+1.0434
26	R4*R5	+3.1523e-05	+7.4113e-07	+42.5332	< 0.0001	+2.2224	+2.2242
28	R5*R6	+3.3574e-06	+6.4214e-07	+5.2284	< 0.0001	+1.0944	+1.1028

Fig. 7f Example 1: Analysis of variance results for the recommended math model with 24 terms.

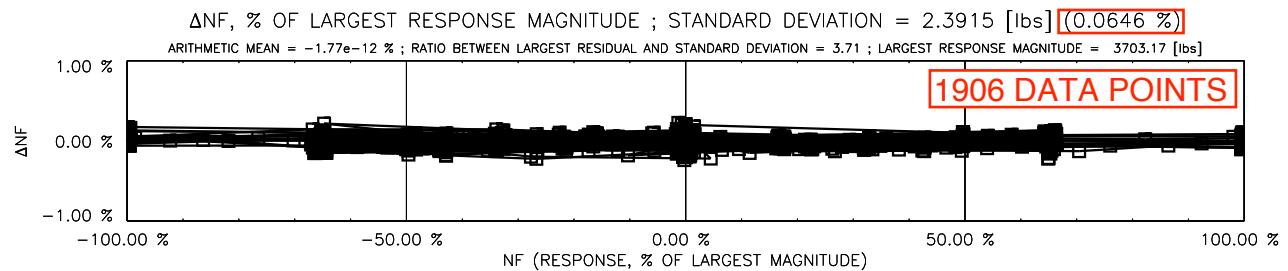


Fig. 7g Example 1: Response residuals of *data points* that were used to determine the recommended math model.

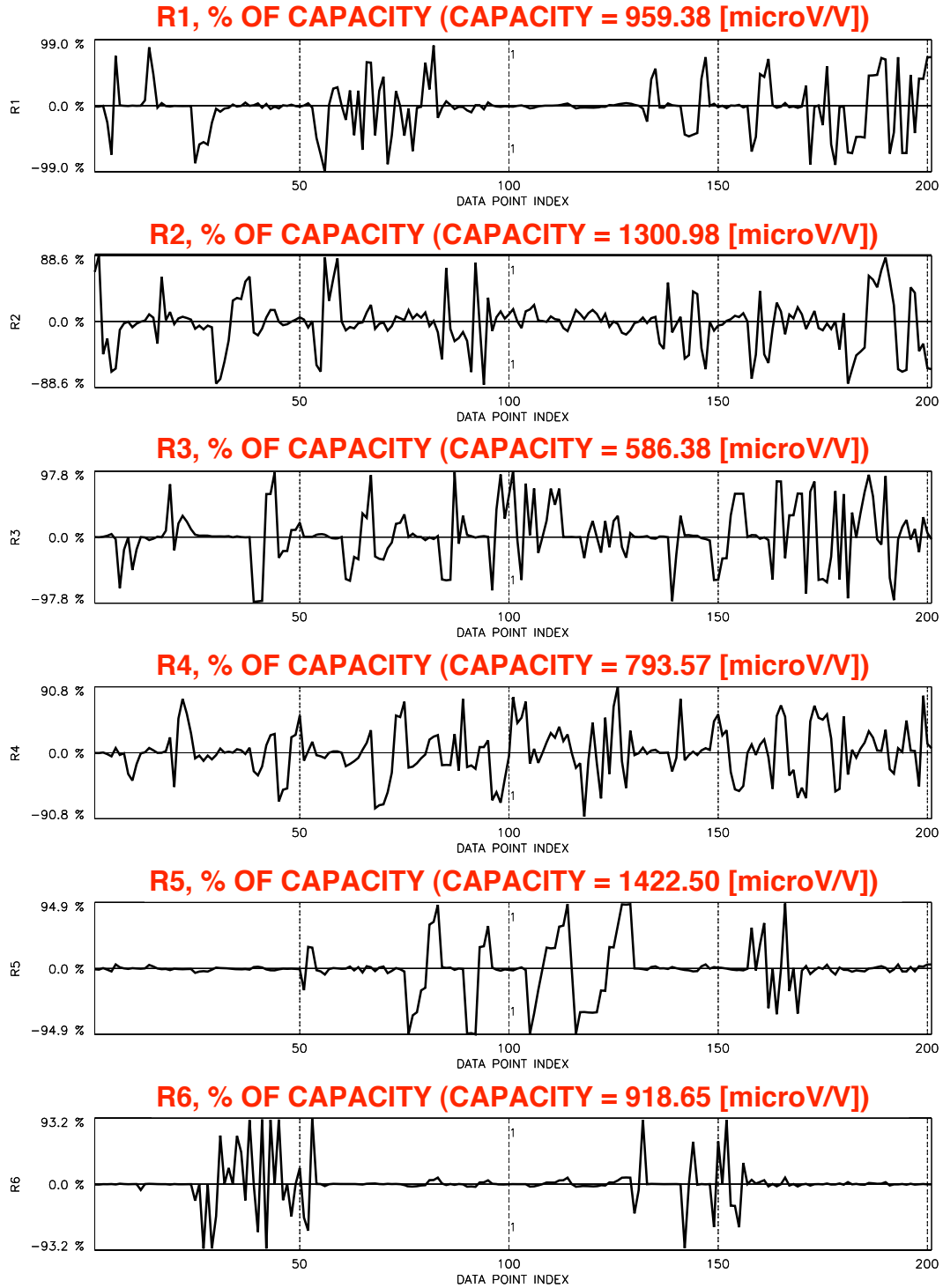


Fig. 8a Example 2: Electrical outputs of strain-gages versus *data point index* (201 *data points*).

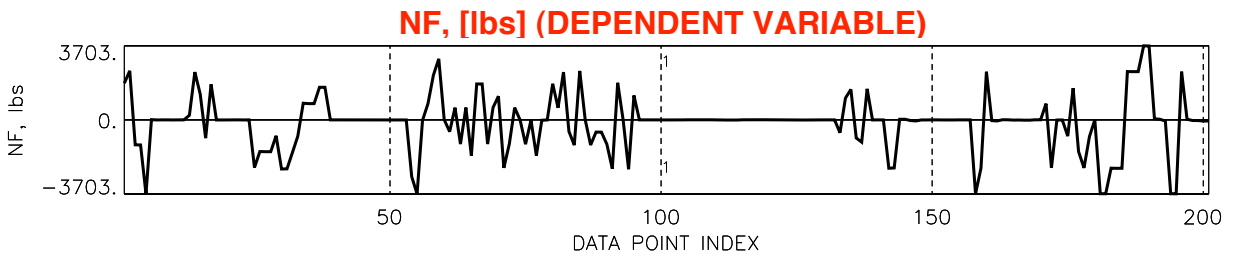


Fig. 8b Example 2: Normal forces versus *data point* index (201 *data points*).

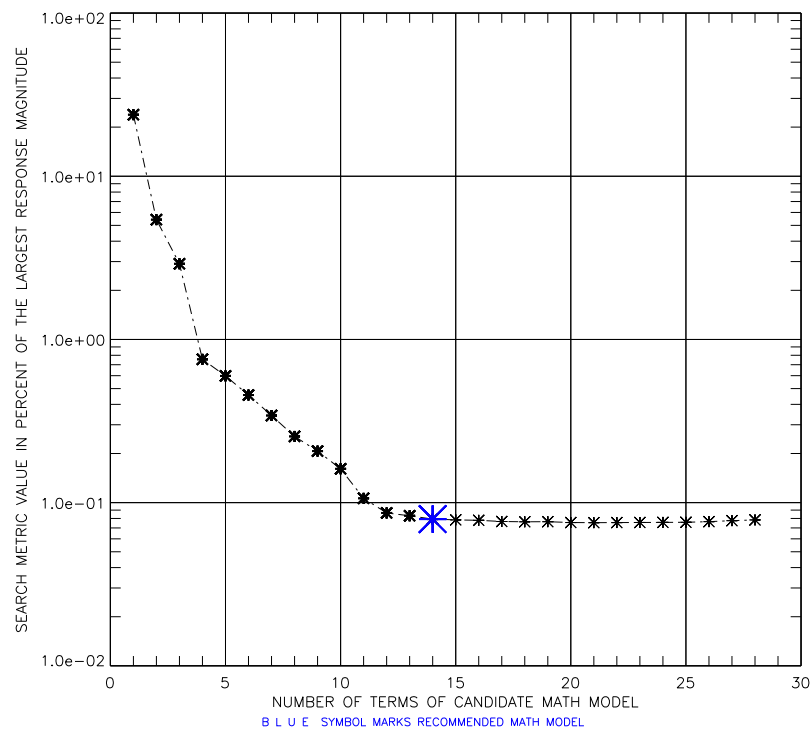


Fig. 8c Example 2: *Search Metric 3* versus number of regression model terms.

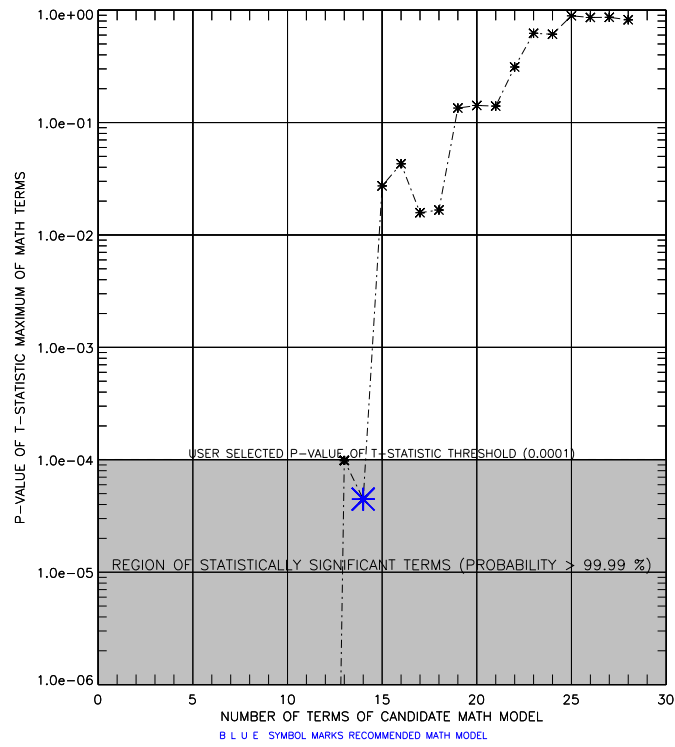


Fig. 8d Example 2: Primary search constraint versus the number of regression model terms.

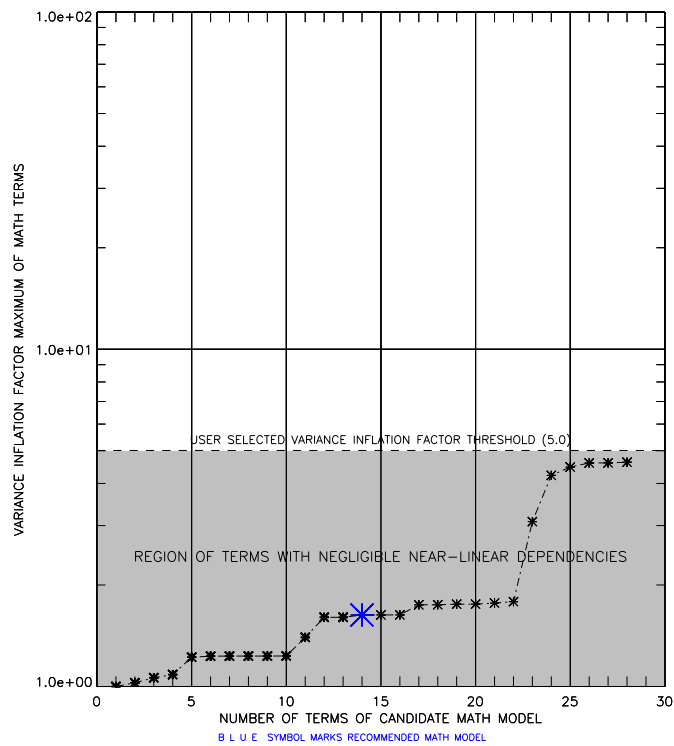


Fig. 8e Example 2: Secondary search constraint versus the number of regression model terms.

SOURCE OF VARIATION	SUM OF SQUARES	PRESS STATISTIC	DEGREES OF FREEDOM	MEAN SQUARE	F-VALUE OF REGRESSION	P-VALUE OF REGRESSION
REGRESSION	3.5825e+08	—	13	2.7557e+07	3.6429e+06	< 0.0001
RESIDUAL	1414.5894	1723.8715	187	7.5646	—	—
TOTAL	3.5825e+08	—	200	—	—	—
R-SQUARE		ADJ. R-SQUARE		PRESS R-SQUARE		
0.999996		0.999996		0.999995		

REGRESSION COEFFICIENT ESTIMATES AND STATISTICAL METRICS (NF)							
REGRESSION MODEL HIERARCHY CHARACTERISTICS = HIERARCHICAL							
MATH TERM INDEX	MATH TERM NAME	COEFFICIENT VALUE	STANDARD ERROR	T-STATISTIC OF COEFFICIENT	P-VALUE OF COEFFICIENT	VIF (PRIMARY)	VIF (ALTERNATE)
1	INTERCEPT	+1.0848	+0.2592	+4.1843	—	—	—
2	R1	+2.4966	+0.0006	+4118.2949	< 0.0001	+1.0664	+1.0804
3	R2	+2.1654	+0.0005	+4461.3122	< 0.0001	+1.1641	+1.1657
4	R3	−0.6734	+0.0009	−752.1106	< 0.0001	+1.2123	+1.2190
5	R4	−0.0667	+0.0008	−81.9190	< 0.0001	+1.2329	+1.2354
6	R5	+0.2395	+0.0005	+502.2758	< 0.0001	+1.2834	+1.1981
7	R6	−0.0558	+0.0008	−68.5852	< 0.0001	+1.0381	+1.0381
8	R1*R1	−2.4689e−05	+1.1832e−06	−20.8662	< 0.0001	+1.3263	+1.3517
9	R2*R2	−1.4095e−05	+6.7951e−07	−20.7427	< 0.0001	+1.2331	+1.2186
12	R5*R5	−1.3981e−05	+5.4117e−07	−25.8337	< 0.0001	+1.6293	+1.6325
17	R1*R5	−8.5372e−06	+2.0422e−06	−4.1805	< 0.0001	+1.2246	+1.1701
21	R2*R5	+4.3279e−05	+1.2142e−06	+35.6450	< 0.0001	+1.2184	+1.2068
24	R3*R5	+3.3467e−05	+3.5087e−06	+9.5383	< 0.0001	+1.4202	+1.4442
26	R4*R5	+3.2677e−05	+2.2004e−06	+14.8506	< 0.0001	+1.6098	+1.6098

Fig. 8f Example 2: Analysis of variance results for the recommended math model with 14 terms.

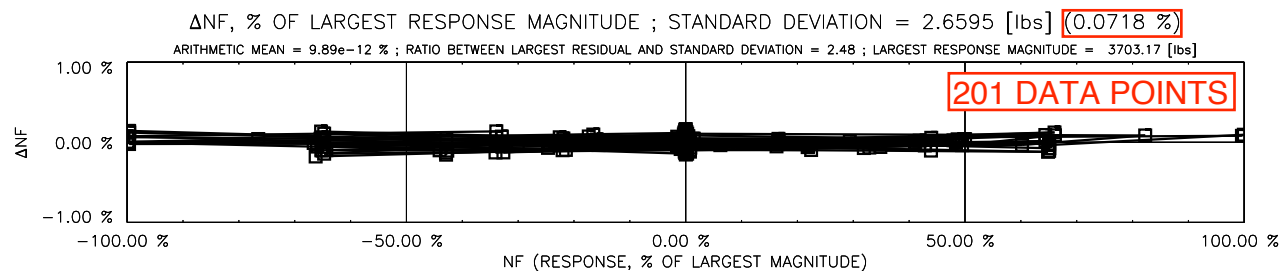


Fig. 8g Example 2: Response residuals of *data points* that were used to determine the recommended math model.

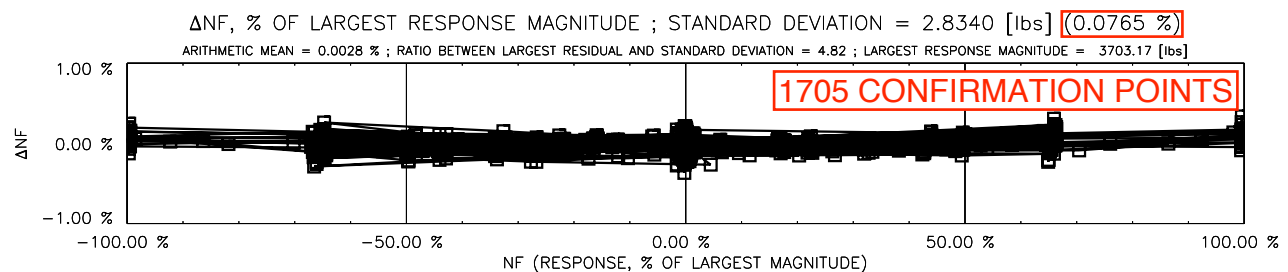


Fig. 8h Example 2: Response residuals of *confirmation points* that were used to test the recommended math model.