https://ntrs.nasa.gov/search.jsp?R=20170005257 2020-03-07T04:04:53+00:00Z

Wind Tunnel Strain–Gage Balance Calibration Data Analysis using a Weighted Least Squares Approach

N. Ulbrich† and T. Volden‡
Jacobs Technology Inc., Moffett Field, California 94035

A new approach is presented that uses a weighted least squares fit to analyze wind tunnel strain–gage balance calibration data. The weighted least squares fit is specifically designed to increase the influence of single–component loadings during the regression analysis. The weighted least squares fit also reduces the impact of calibration load schedule asymmetries on the predicted primary sensitivities of the balance gages. A weighting factor between zero and one is assigned to each calibration data point that depends on a simple count of its intentionally loaded load components or gages. The greater the number of a data point's intentionally loaded load components or gages is, the smaller its weighting factor becomes. The proposed approach is applicable to both the Iterative and Non–Iterative Methods that are used for the analysis of strain–gage balance calibration data in the aerospace testing community. The Iterative Method uses a reasonable estimate of the tare corrected load set as input for the determination of the weighting factors. The Non–Iterative Method, on the other hand, uses gage output differences relative to the natural zeros as input for the determination of the weighting factors. Machine calibration data of a six–component force balance is used to illustrate benefits of the proposed weighted least squares fit. In addition, a detailed derivation of the PRESS residuals associated with a weighted least squares fit is given in the appendices of the paper as this information could not be found in the literature. These PRESS residuals may be needed to evaluate the predictive capabilities of the final regression models that result from a weighted least squares fit of the balance calibration data.

Nomenclature

$\mathbf{A}$	= matrix that has the regressors of an ordinary least squares fit
$\overline{\mathbf{A}}$	= subset of matrix $\mathbf{A}$ that is used for the calculation of PRESS residuals
AF	= axial force of a strain–gage balance
$\mathbf{B}$	= vector that has the responses of an ordinary least squares fit
$\overline{\mathbf{B}}$	= subset of vector $\mathbf{B}$ that is used for the calculation of PRESS residuals
$c_1, c_2, \ldots, c_m$	= regression coefficients used in connection with an ordinary or weighted least squares fit
$\overline{c_1}, \overline{c_2}, \ldots, \overline{c_m}$	= regression coefficients associated with the calculation of PRESS residuals
$\mathbf{e}$	= unit basis vector
e	= component of unit basis vector
i	= index of a data point
$\mathbf{I}$	= identity matrix
j	= index of a data point
k	= index of a data point
l	= index of a regressor
m	= total number of regressors or regression coefficients
n_k	= total number of intentionally loaded components of a data point with index k

† Aerodynamicist, Jacobs Technology Inc.
‡ Computer Engineer, Jacobs Technology Inc.

n'_μ	= total number of intentionally loaded components of a load series with index μ
$N1, N2$	= normal force components of a force balance
p	= total number of data points
q	= total number of load series of an OFAT type balance calibration data set
$\mathbf{Q}$	= vector that has the regressor values of a specific data point
r	= response variable
rAF	= electrical output of the axial force gage of a strain–gage balance
RM	= rolling moment of a strain–gage balance
$S1, S2$	= side force components of a force balance
$\mathbf{U}$	= matrix that has the "weighted" regressors of a "weighted" least squares fit
$\overline{\mathbf{U}}$	= subset of matrix $\mathbf{U}$ that is used for the calculation of PRESS residuals
$\mathbf{V}$	= vector that has the "weighted" responses of a "weighted" least squares fit
$\overline{\mathbf{V}}$	= subset of vector $\mathbf{V}$ that is used for the calculation of PRESS residuals
w_k	= weighting factor of data point with point index k
$w_{k,\mu}$	= weighting factor of data point with point index k and load series index μ
$\mathbf{W}$	= diagonal matrix that has the weighting factors of each data point
$\overline{\mathbf{W}}$	= subset of matrix $\mathbf{W}$ that is used for the calculation of PRESS residuals
$\mathcal{W}$	= diagonal matrix that has the square root of the weighting factors of each data point
$\overline{\mathcal{W}}$	= subset of matrix $\mathcal{W}$ that is used for the calculation of PRESS residuals
$\mathbf{X}$	= vector that has the regression coefficients of a least squares fit
$\overline{\mathbf{X}}$	= subset of vector $\mathbf{X}$ that is used for the calculation of PRESS residuals
ΔAF	= axial force residual $\equiv$ difference between the tare corrected and fitted load value
Δr	= response residual that is obtained after performing a weighted least squares fit
δr	= PRESS residual of response
μ	= index of a load series
$\xi_{l,k}$	= value of a regressor variable (regressor index $\equiv l$, data point index $\equiv k$)
ρ	= fitted response that is obtained after performing a weighted least squares fit
$\overline{\rho}$	= fitted response that is needed for the calculation of PRESS residuals
σ_{PRESS}	= standard deviation of PRESS residuals
ψ	= exponent used in authors' definition of a weighting factor

I. Introduction

Calibration data of a multi–component wind tunnel strain–gage balance is traditionally analyzed using a multivariate least squares fit so that balance loads can be predicted from measured gage outputs during a wind tunnel test (see Ref. [1] for more detail). Analysts often do not realize that the traditionally applied ordinary least squares fit treats every balance calibration point equally. In other words, every data point has a weighting factor of "one" if an ordinary least squares fit is performed. Consequently, calibration load schedule design and, in particular, the number of repeat data points, "implicitly" determine the "weight" of a data point during the regression analysis of the calibration data.

The authors observed over the years that many balance calibration data sets have a relatively small number of single–component loadings. This calibration load schedule characteristic often results from the fact that a balance calibration laboratory wants to include as many combined loadings as possible in the calibration data so that related coefficients of the regression model's cross–product terms can be estimated with confidence. Therefore, the true influence of the existing single–component loadings on the final regression coefficient estimates may be too small for some balance calibration data sets considering the fact that single–component loadings significantly influence the magnitude of the load predictions. In addition, single–component loadings have a strong impact on the least squares estimates of the gage sensitivities as the electrical outputs of a strain–gage balance show highly linear behavior when (i) the balance is analyzed in its design format and (ii) a primary output is plotted versus the related primary load.

Hidden asymmetries in the chosen load schedule design themselves can also impose an unwanted "implicit" weighting on the data points during the regression analysis of the calibration data. A typical example of this situation is a load schedule design that first uses very few single–component loadings in order to

apply the axial force over its entire positive and negative load range. Afterwards, only positive axial force loadings are applied as constant auxiliary loads in order to obtain load combinations that are needed for the characterization of certain cross–product terms.

The examples discussed above clearly indicate that the ability to assign an "explicit" weighting factor to each individual data point of a balance calibration data set could be beneficial during the regression analysis of certain types of calibration data. Therefore, the authors developed a new approach that uses some very general data point specific information for the definition of a weighted least squares fit in order (i) to give more weight to single–component loadings and (ii) to reduce the negative impact of load schedule asymmetries on the regression coefficient estimates.

In the next section, relationships are reviewed that are needed for a basic understanding of a weighted least squares fit of wind tunnel strain–gage balance calibration data. Then, the authors' weighting factor definition is discussed. Afterwards, data from the calibration of a wind tunnel balance is used to illustrate benefits of the application of a weighted least squares fit to balance calibration data. In addition, it is shown in the appendices how a weighted least squares fit influences the calculation of PRESS residuals of the outputs and loads of a balance (the appendices were included in the paper because this important information could not be found in the existing literature).

II. Weighted Least Squares Approach

In general, wind tunnel strain–gage balance data is analyzed using a multivariate least squares fit that is described in matrix format (see Refs. [1], [2], and [3] for more details). This approach assembles the regressors (including the regressor associated with the intercept term) in a rectangular matrix $\mathbf{A}$. The number of rows of matrix $\mathbf{A}$ equals the total number "p" of calibration data points. The number of columns of matrix $\mathbf{A}$, on the other hand, equals the total number "m" of regressors (i.e., math terms). The responses are also needed for a complete description of the multivariate least squares problem of the balance calibration data. They are assembled in a column vector $\mathbf{B}$. The number of rows of vector $\mathbf{B}$ equals the total number "p" of calibration data points. Then, assuming that the *Iterative Method* described in Ref. [1] is used for the balance calibration data analysis, the "ordinary" least squares problem of balance calibration data can be described by the following matrix equation

$$\textit{ordinary least squares problem} \quad \Longrightarrow \quad \mathbf{A}_{p\times m} \cdot \mathbf{X}_{m\times 1} \;=\; \mathbf{B}_{p\times 1} \tag{1}$$

where vector $\mathbf{X}$ describes a column vector with m rows that has the regression coefficients of the math model of the calibration data. The general solution of the ordinary least squares problem defined in Eq. (1), i.e., the solution of vector $\mathbf{X}$, is known from the literature (see Refs. [2] and [3]). It can be summarized as follows:

$$\textit{solution of ordinary least squares problem} \quad \Longrightarrow \quad \mathbf{X}_{m\times 1} \;=\; \left[\, \mathbf{A}^{\mathbf{T}} \cdot \mathbf{A} \,\right]^{-1}_{m\times m} \cdot \left[\, \mathbf{A}^{\mathbf{T}} \cdot \mathbf{B} \,\right]_{m\times 1} \tag{2}$$

The above solution of the ordinary least squares problem assumes that the weighting factors $w_1, \ldots, w_p$ of all data points are one. This conclusion can be expressed as follows:

$$\textit{ordinary least squares problem} \quad \Longrightarrow \quad w_1 \;=\; w_2 \;=\; \ldots \;=\; w_p \;=\; 1.0 \tag{3}$$

Sometimes, an analyst of balance calibration data wants to use subject matter knowledge in order to increase the influence of certain groups of calibration points during the regression analysis. Then, a "weighted" least squares fit has to be applied to the data that assigns an individual weighting factor to each data point (see Ref. [3] for a general description of the weighted least squares approach). The matrix equation of the corresponding weighted least squares problem can easily be obtained from the equation of an ordinary least squares problem if the diagonal matrix $\mathbf{W}$ is introduced. This matrix stores the weighting factors of the individual data points on its principle diagonal. The matrix $\mathbf{W}$ looks as follows:

$$\mathbf{W}_{p\times p} \;=\; \begin{bmatrix} w_1 & 0 & \cdots & 0 & 0 \\ 0 & w_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & w_{p-1} & 0 \\ 0 & 0 & \cdots & 0 & w_p \end{bmatrix} \tag{4}$$

The weighting factors are introduced into the equations of the weighted least squares fit by simply multiplying matrix $\mathbf{A}$ and vector $\mathbf{B}$ of Eq. (1) with matrix $\mathbf{W}$. Then, the matrix equation of a weighted least squares problem can be expressed as follows (see also Ref. [3])

$$\textit{weighted least squares problem} \quad \Longrightarrow \quad \mathbf{W}_{p\times p} \cdot \mathbf{A}_{p\times m} \cdot \mathbf{X}_{m\times 1} \;=\; \mathbf{W}_{p\times p} \cdot \mathbf{B}_{p\times 1} \tag{5}$$

where vector $\mathbf{X}$ has the regression coefficients that result from the weighted least squares fit. The normal equations of the weighted least squares problem are obtained by simply multiplying both sides of Eq. (5) with the transpose of matrix $\mathbf{A}$ (see again Ref. [3] for more details). Then, we get:

$$\textit{normal equations} \quad \Longrightarrow \quad \mathbf{A}^{\mathbf{T}}_{m\times p} \cdot \mathbf{W}_{p\times p} \cdot \mathbf{A}_{p\times m} \cdot \mathbf{X}_{m\times 1} \;=\; \mathbf{A}^{\mathbf{T}}_{m\times p} \cdot \mathbf{W}_{p\times p} \cdot \mathbf{B}_{p\times 1} \tag{6}$$

It is useful to transform the normal equations of the "weighted" least squares problem to the normal equations of an equivalent ordinary least squares fit. Then, existing computer code can be used to solve the weighted least squares problem if the code was originally developed to perform an ordinary least squares fit. Therefore, an auxiliary matrix $\mathcal{W}$ is introduced. This matrix is defined as follows:

$$\mathcal{W}_{\mathbf{p}\times\mathbf{p}} \;=\; \mathbf{W}^{\mathbf{1/2}}_{\mathbf{p}\times\mathbf{p}} \;=\; \begin{bmatrix} \sqrt{w_1} & 0 & \cdots & 0 & 0 \\ 0 & \sqrt{w_2} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sqrt{w_{p-1}} & 0 \\ 0 & 0 & \cdots & 0 & \sqrt{w_p} \end{bmatrix} \tag{7a}$$

where

$$\mathbf{W}_{p\times p} \;=\; \mathbf{W}^{1/2}_{p\times p} \cdot \mathbf{W}^{1/2}_{p\times p} \;=\; \mathcal{W}_{p\times p} \cdot \mathcal{W}_{p\times p} \tag{7b}$$

Then, after using the right–hand side of Eq. (7b) in order to substitute the matrix $\mathbf{W}$ in Eq. (6), we get the following transformed normal equations of the weighted least squares problem:

$$\mathbf{A}^{\mathbf{T}}_{m\times p} \cdot \mathcal{W}_{p\times p} \cdot \mathcal{W}_{p\times p} \cdot \mathbf{A}_{p\times m} \cdot \mathbf{X}_{m\times 1} \;=\; \mathbf{A}^{\mathbf{T}}_{m\times p} \cdot \mathcal{W}_{p\times p} \cdot \mathcal{W}_{p\times p} \cdot \mathbf{B}_{p\times 1} \tag{8}$$

We also know from the operator rules of matrix algebra that $\mathcal{W}^{\mathbf{T}} = \mathcal{W}$ because $\mathcal{W}$ is a diagonal matrix. In addition, we know that $\mathbf{A}^{\mathbf{T}}\mathcal{W} = [\mathcal{W}\mathbf{A}]^{\mathbf{T}}$. Consequently, Eq. (8) can also be expressed as follows:

$$[\, \mathcal{W}_{p\times p} \cdot \mathbf{A}_{p\times m} \,]^{\mathbf{T}} \cdot [\, \mathcal{W}_{p\times p} \cdot \mathbf{A}_{p\times m} \,] \cdot \mathbf{X}_{m\times 1} \;=\; [\, \mathcal{W}_{p\times p} \cdot \mathbf{A}_{p\times m} \,]^{\mathbf{T}} \cdot [\, \mathcal{W}_{p\times p} \cdot \mathbf{B}_{p\times 1} \,] \tag{9}$$

Now, the auxiliary matrix $\mathbf{U}$ and the auxiliary vector $\mathbf{V}$ are introduced so that Eq. (9) can be expressed in the format of an ordinary least squares fit. We get:

$$\mathbf{U}_{p\times m} \;=\; \mathcal{W}_{p\times p} \cdot \mathbf{A}_{p\times m} \tag{10a}$$

$$\mathbf{V}_{p\times 1} \;=\; \mathcal{W}_{p\times p} \cdot \mathbf{B}_{p\times 1} \tag{10b}$$

Then, after using the left–hand sides of Eqs. (10a) and (10b) to replace the contents of the brackets in Eq. (9), we get the following relationship for the normal equations of the weighted least squares fit:

$$\mathbf{U}^{\mathbf{T}}_{m\times p} \cdot \mathbf{U}_{p\times m} \cdot \mathbf{X}_{m\times 1} \;=\; \mathbf{U}^{\mathbf{T}}_{m\times p} \cdot \mathbf{V}_{p\times 1} \tag{11}$$

Finally, the solution of the weighted least squares problem, i.e., the solution for the regression coefficients of the balance calibration data, can be expressed by the following matrix equation:

$$\textit{solution of weighted least squares problem} \quad \Longrightarrow \quad \mathbf{X}_{m\times 1} \;=\; \left[\, \mathbf{U}^{\mathbf{T}}\, \mathbf{U} \,\right]^{-1}_{m\times m} \cdot \left[\, \mathbf{U}^{\mathbf{T}}\, \mathbf{V} \,\right]_{m\times 1} \tag{12}$$

It remains to define values for the weighting factors of the data points of a strain–gage balance calibration data set. The authors' proposed weighting factor definition is discussed in the next section of the paper.

III. Weighting Factor Definition for Balance Calibration Data

The successful use of a weighted least squares fit during the regression analysis of multivariate experimental data highly depends on a suitable choice of a weighting factor for each data point. An analyst will quickly find out that the weighting factor selection is often more difficult than initially anticipated. This problem is explicitly mentioned in the literature. For example, *Miller* makes the following two statements in his textbook: "*... If God or the Devil were willing to tell us the values for ... [the weights] ..., the solution would be ...*" and "*... Since most of us cannot get help from above or below, we are faced with having to estimate the unknown weights. ...*" (taken from Ref. [4], p. 210). Consequently, the authors realized from the start that the definition of a reliable set of weighting factors for the analysis of strain–gage balance calibration data had to be based on some very general subject matter knowledge about the behavior of a balance and its calibration process. This approach helps avoid the situation that the chosen weighting factor definition would only work successfully for one class of balance calibration load schedule designs.

The authors decided to define a weighting factor for each calibration data point such that the overall influence of points with more complex load combinations is reduced during the regression analysis. Then, points of single–component loadings would have greater influence on the final regression coefficient set. The authors' weighting approach is also supported by two observations that can be made whenever loads are analyzed in the "design format" of the balance (e.g., a force balance is analyzed in force balance format): (i) a typical strain–gage balance shows highly linear behavior when it is analyzed in its "design format" and the primary gage outputs are plotted versus related primary gage loads; (ii) more than 90% of the magnitude of a regression model's prediction of an output is determined by regressors that are constructed as a purely linear function of a single balance load assuming the balance is analyzed using the *Iterative Method*.

The authors' weighting factor definition is based on a simple "count" of the number of "intentionally loaded" load components (or gages) of a data point. The weighting factor is constructed such that an increase of the number of "intentionally loaded" load components (or gages) decreases the influence of the data point on the regression coefficient estimates. The authors consider a balance load component of a data point to be "intentionally loaded" if the magnitude of the load component exceeds the recommended threshold of 20 % of the load capacity. This conservative threshold choice guarantees that data points with small to moderate load levels always get the highest weighting factor. In addition, in order to improve clarity, the authors decided to map the weighting factor's numerical value to the fixed interval between zero and one. In conclusion, the authors' weighting factor definition can be summarized as follows ...

WEIGHTING FACTOR DEFINITION

$\Longrightarrow$ applicable to all load schedules $\Longleftarrow$

$$w_k = \left[\frac{\lambda_k}{MAX\ \{\ \lambda_1\ ,\ \ldots\ ,\ \lambda_p\ \}} \right]^{\psi} \quad ; \quad 1 \leq k \leq p \qquad (13a)$$

where

$$\lambda_k = \frac{MAX\ \{\ n_1\ ,\ \ldots\ ,\ n_p\ \}}{n_k} \qquad (13b)$$

$$\psi = 2 \qquad (13c)$$

$$n_k = 0 \quad \Longrightarrow \quad w_k = 1.0 \qquad (13d)$$

where (i) w_k is the weighting factor of the data point with index k and (ii) n_k is the number of intentionally loaded load components or gages of the data point. The definition above uses the power of a fraction to

control the strength of the weighting. The exponent choice of $\psi = 2$ suggested itself because the square root $\sqrt{w_k}$ of the individual weighting factors of each data point is needed to describe the "weighted" least squares problem in terms of the normal equations of an ordinary least squares problem (see Eq. (7a)). In theory, an exponent choice of $\psi > 2$ could also be used for the weighting factor definition. It would simply increase the strength of the weighting that is applied during the least squares fit of the calibration data. However, the authors recommend to use an exponent of $\psi = 2$ for balance calibration data because they did not observe a noticeable improvement in the quality of the least squares fit when an exponent of $\psi > 2$ was used.

The definition above, i.e, Eq. (13a), uses information for the calculation of a data point's weighting factor that is obtained from its unique loading characteristics. In other words, the weighting factor of a data point is completely independent of load combinations seen by other data points. Consequently, the authors' definition works with all load schedule designs. Sometimes, however, a balance is calibrated using the so–called *One–Factor–At–A–Time* (OFAT) approach. This approach uses multiple load series that each represent one specific load configuration. In addition, the OFAT approach varies only a single load component within each load series while keeping all other load components at a constant value during the execution of the load series. Therefore, it is reasonable to assume that all data points of an OFAT load series have a single common error source. Consequently, the authors developed an alternate weighting factor definition for OFAT type balance calibration data sets that (i) first determines the maximum number of intentionally loaded load components or gages of an OFAT type load series, (ii) uses this value in combination with the basic idea expressed in Eq. (13a) to compute the load series' weighting factor, and (iii) assigns this weighting factor to all data points of the OFAT load series. This alternate weighting factor definition can be summarized as follows ...

ALTERNATE WEIGHTING FACTOR DEFINITION

$\Longrightarrow$ only applicable to "One–Factor–At–A–Time" type load schedules $\Longleftarrow$

$$w_{k,\mu} = \left[\frac{\lambda'_\mu}{MAX\ \{\ \lambda'_1\ ,\ \ldots\ ,\ \lambda'_q\ \}} \right]^\psi \quad ; \quad 1 \leq \mu \leq q \qquad (14a)$$

where

$$\lambda'_\mu = \frac{MAX\ \{\ n'_1\ ,\ \ldots\ ,\ n'_q\ \}}{n'_\mu} \qquad (14b)$$

$$\psi = 2 \qquad (14c)$$

$$n'_\mu = 0 \quad \Longrightarrow \quad w_{k,\mu} = 1.0 \qquad (14d)$$

where $w_{k,\mu}$ is the weighting factor of the data point with index k that belongs to the load series with index μ, and, n'_μ is the maximum number of "intentionally loaded" load components or gages of the load series. It has to be emphasized that the alternate weighting factor definition given in Eq. (14a) is "optional" because the original weighting factor definition given in Eq. (13a) can also be applied to OFAT type calibration data sets. Therefore, the authors recommend to apply the two weighting factor definitions separately to a given OFAT type calibration data set and compare the resulting least squares fits afterwards in order to decide which approach works best for the specific characteristics of the given calibration data.

It is useful to discuss a simple example to better illustrate the calculation of the weighting factor that is defined in Eqs. (13a) and (13b). It is assumed, that a 6–component force balance is calibrated in force balance format. It is also assumed that the final calibration data set consists of 500 data points and that the maximum number of "intentionally loaded" load components of a data point equals four. Then, we can summarize the given calibration load schedule characteristics as follows:

$$\textit{total number of data points} \quad \Longrightarrow \quad p = 500$$

$$\textit{maximum number of intentionally loaded load components} \implies MAX\{n_1, \ldots, n_k, \ldots, n_{500}\} = 4$$

$$\textit{possible number of intentionally loaded load components} \implies 0 \leq n_k \leq 4$$

Now, the possible weighting factors of the balance calibration data set can be computed by applying the weighting factor definition given in Eqs. (13a) and (13b). Table 1 below shows the corresponding values for each possible number of "intentionally loaded" load components of a data point.

Table 1: Possible weighting factors w_k of the balance calibration data example.

n_k	λ_k	$\lambda_k/MAX\{\lambda_k\}$	w_k
0	–	–	Eq. (13d) $\Longrightarrow$ 1.0
1	$4/1 = 4.00$	$4.00/4.00 = 1.00$	$(1.00)^2 = 1.0000$
2	$4/2 = 2.00$	$2.00/4.00 = 0.50$	$(0.50)^2 = 0.2500$
3	$4/3 = 1.3\overline{3}$	$1.3\overline{3}/4.00 = 0.3\overline{3}$	$(0.3\overline{3})^2 = 0.111\overline{1}$
4	$4/4 = 1.00$	$1.00/4.00 = 0.25$	$(0.25)^2 = 0.0625$

The example above illustrates another property of the weighting factor definition: its application depends on an analyst's ability to objectively "count" the number n_k of "intentionally loaded" load components (or gages) of each calibration data point. This task is, in reality, more complicated than it initially appears because (i) balance loads have to be described relative to the absolute load datum of the balance before the regression analysis can take place and (ii) fundamental characteristics of the chosen load prediction methods have to be taken into account. These issues are discussed in more detail in the next section of the paper.

IV. Number of Intentionally Loaded Load Components

A. General Remarks

It was mentioned in the previous section that the determination of the number of applied load components of a data point is more complicated than it may initially appear. The successful completion of this task depends on a good understanding of the importance of the absolute load datum of the balance. In addition, basic characteristics of the chosen load prediction method need to be taken into account. Two different methods are used for this purpose in the aerospace testing community. They are called the *Iterative Method* and the *Non–Iterative Method* (see Ref. [5] for more detail). The methods differ in the choice of independent variables, i.e., factors, that are used to construct the regressors for the least squares fit of the balance calibration data. First, the determination of the number of "intentionally loaded" load components is described assuming that an analyst applies the *Iterative Method* to the balance calibration data. Afterwards, the same task is discussed assuming that an analyst applies the *Non–Iterative Method*.

B. Iterative Method

The *Iterative Method* requires two steps in order to predict balance loads from measured gage outputs. First, the method fits the outputs as a function of the balance loads. In other words, it treats the loads as the independent variables of the calibration data set and uses them to construct the regressors for the least squares fit. The balance loads themselves should have been tare corrected, i.e., corrected for loads associated with the weight of the calibration hardware, before the regression analysis takes place so that all loads are described relative to the common load datum of zero absolute load. Then, after completion of the least squares fit of the outputs, a load iteration scheme is constructed from the resulting regression coefficients of the outputs so that loads can be predicted from the measured outputs during the wind tunnel test.

The description of the *Iterative Method* above makes one remark that is important when it comes to the determination of the number "n_k" of the "intentionally loaded" load components of a data point: the tare corrected loads, i.e., loads relative to the load datum of zero absolute load, have to be used to determine n_k. Therefore, an analyst must be able to estimate the magnitude of the tare corrected calibration load set <u>before</u> the final regression analysis of the calibration data takes place. In addition, the authors recommend to use a threshold of ± 20 % of the capacity of a load component in order to decide whether or not a load component is considered to be "intentionally loaded." This recommendation, of course, will assign the value of $n_k = 0$ to a data point if the magnitude of none of its load components exceeds 20 % of the related load

capacity. In this case, the condition defined in Eq. (13d) is applied. Consequently, those data points will always get the highest possible weighting factor of one.

C. Non–Iterative Method

The *Non–Iterative Method* may also used to predict balance loads from measured strain–gage outputs during a wind tunnel test. The *Non–Iterative Method* directly fits the tare corrected balance loads as a function of the gage outputs that were measured during the calibration. Ideally, the difference between a gage output and its natural zero, i.e., its output at zero absolute load, should be used to construct the regressors for the least squares fit of a balance load. Then, no confusion about the output datum associated with the regression model of the load exists when the regression model is used during a wind tunnel test as, by design, the output datum associated with output differences is zero.

The description of the *Non–Iterative Method* above contains one statement that is relevant as far as the determination of n_k, i.e., of the number of "intentionally loaded" load components is concerned: the difference between a gage output and its natural zero should be used to construct the regressors for the least squares fit of the balance load. In other words, output differences and <u>not</u> the tare corrected loads are the independent variables of the least squares fit that the *Non–Iterative Method* performs. Therefore, an analyst must "count" the number of "intentionally loaded" gages instead of the number of "intentionally loaded" load components as the output differences are the electrical representation of the load state of the balance. In addition, as only "intentionally loaded" gages of a data point are to be used for the determination of n_k, the authors recommend to use a threshold of ± 20 % of the maximum output difference of a given gage in order to decide whether or not a gage is considered to be "intentionally loaded." This recommendation, of course, will assign the value of $n_k = 0$ to a data point if the magnitude of none of its output differences exceeds 20 % of the related maximum output difference. Then, the condition defined in Eq. (13d) is again applied. Consequently, those data points will always get the highest possible weighting factor of one.

A balance calibration data example is discussed in the next section to better illustrate the application of the authors' weighted least squares fit approach to real–world balance calibration data.

V. Discussion of Example

A. General Remarks

Data from a machine calibration of NASA's MC60E balance is used in this section to illustrate both application and benefits of a weighted least squares fit of strain–gage balance calibration data. The MC60E balance is a six–component force balance that was manufactured and originally calibrated by *Force Measurement Systems (FMS)* in San Diego. It has a diameter of 2.0 inches and measures five forces and one moment ($N1$, $N2$, $S1$, $S2$, AF, RM). Table 2 below lists load capacities of the balance:

Table 2: Load capacities of the MC60E balance.

	$N1$, lbs	$N2$, lbs	$S1$, lbs	$S2$, lbs	RM, in–lbs	AF, lbs
Capacity	2500	2500	1250	1250	5000	700

The chosen machine calibration data set was obtained in 2016 in *FMS' Automatic Balance Calibration System (ABCS)*. The data set consisted of 2091 points that were distributed across sixteen load series. The calibration load set was defined by using a load schedule design for force balances that was originally developed at NASA Ames Research Center (see Ref. [6], Fig. 3, Fig. 4). This design consists of six single–component load series, three two–component load series, four three–component load series, one four–component load series, and two five–component load series. Loadings of the multi–component series are structured such that all combinations of the components are applied within the limits of their load ranges. The load ranges are gradually reduced from $\pm 100\%$ to $\pm 33\%$ of capacity as the number of applied load components of a series increases. This constraint was introduced to decrease the potential peak stress that the balance may experience during calibration.

The Ames' load schedule design uses an OFAT calibration approach during six of its sixteen load series. Consequently, it was only possible to compare results for two different data point weighting methods during the data analysis (i.e., equally weighted data points and the weighting factor definition that is described in Eq. (13a)). Characteristics of these two weighting methods are summarized in Table 3 below.

Table 3: Description of weighting methods that were applied during the calibration data analysis.

Weighting Method	Description
Method 1	equal weight of one is applied to all data points (default weighting approach)
Method 2	weight depends on a "count" of the intentionally loaded gages $\Longrightarrow$ Eq. (13a)

Method 1 above can be described as an **"implicit"** weighting approach as (i) the weighting factors of all data points are identical and (ii) the calibration load schedule structure alone determines the true "weight" of each data point. *Method 2*, on the other hand, can be described as an **"explicit"** weighting approach as the weighting factor of each individual data point is determined analytically and assigned on a point–by–point basis by either an analyst or existing analysis software logic.

B. Results for Implicit Weighting (Method 1)

First, the *Iterative Method* was applied to the 2091–point machine calibration data set. The six outputs of the original calibration data set were each fitted using a 28–term math model (28 terms $\equiv$ 27 terms plus one intercept; no absolute value terms were needed; no near–linear dependencies between the regressors were introduced by this choice). An ordinary least squares fit was performed during the regression analysis that assigned a weighting factor of one to all data points (Weighting Method $\equiv$ *Method 1*). Afterwards, the load iteration scheme was constructed from the regression models. Finally, the load residuals, i.e., the differences between tare corrected and fitted loads, were computed for the six load components of the balance.

The authors decided to only discuss the accuracy of the axial force predictions in more detail in the paper. First, characteristics of the axial force load schedule of the 2091–point data set are reviewed. The axial force load schedule is shown in *Fig. 1a* below. It depicts the tare corrected axial forces of the calibration data when plotted versus the data point index (the alternating gray scale highlights the load series structure).

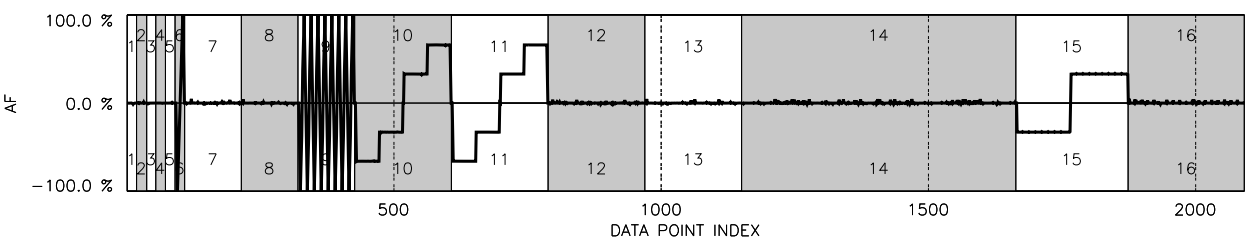

Fig. 1a - Axial force load schedule used for the 2091–point machine calibration of the MC60E balance.

Figure 1b below shows the axial gage outputs plotted versus the tare corrected axial force for all 2091 calibration points. It is observed that the outputs are highly linear in nature when plotted versus the tare corrected axial force. *Figure 1b* also indicates that the calibration points are well distributed across the axial force range of the balance from -700 [lbs] to $+700$ [lbs].

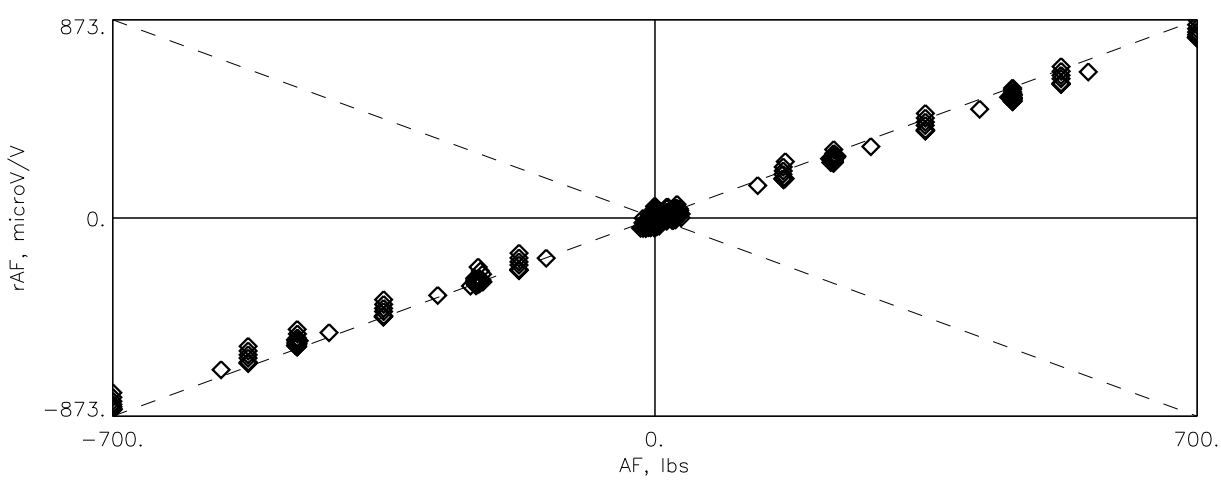

Fig. 1b - Axial gage output plotted versus the axial force for the 2091–point machine calibration.

The axial force residuals of the 2091–point data set were computed after applying the load iteration scheme of the *Iterative Method* to the measured outputs. *Figure 2* below shows the resulting residuals of the calibration data when plotted versus the tare corrected axial force. It is observed that the residuals stay within the 0.25 % threshold that is traditionally used to evaluate load residuals of strain–gage balances.

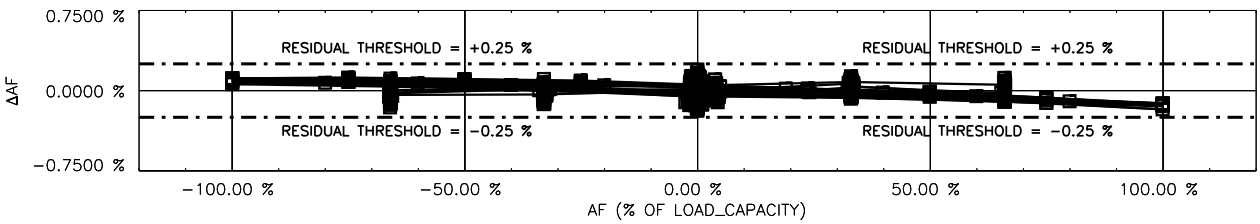

Fig. 2 - Axial force residuals plotted versus the axial force for the 2091–point machine calibration data set after weighting Method 1 was applied during the regression analysis of the data.

Now, in order to illustrate the benefit of the weighted least squares fit, an asymmetric load schedule was generated from the original 2091–point calibration data set. This asymmetric data set was obtained by simply removing almost all data points from the 2091–point data set that have a negative axial force. Only axial loadings of load series 6 were left unchanged. *Figure 3a* below shows the modified axial force load schedule of the resulting 1210–point subset of the original 2091–point calibration data set.

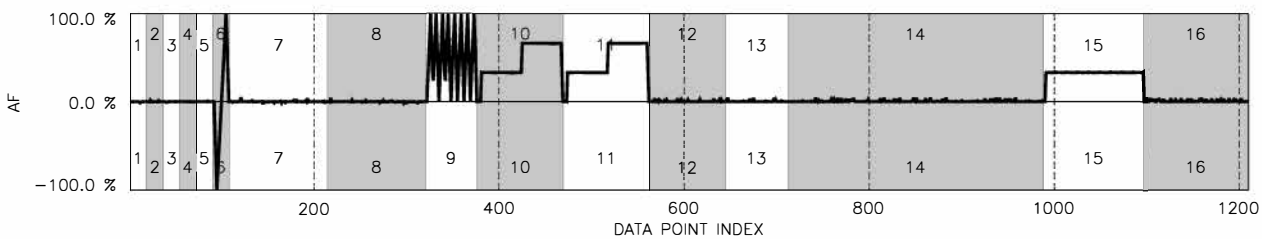

Fig. 3a - Axial force load schedule used for the 1210–point subset of the 2091–point machine calibration.

Figure 3b below shows the axial gage outputs of the 1210–point subset when plotted versus the tare corrected axial forces. It can clearly be seen in *Fig. 3b* that the 1210–point subset has substantially fewer negative axial force loadings. Consequently, an ordinary least squares fit is expected to apply too much "implicit" weight on the positive axial force loadings which may negatively influence the resulting least squares estimate of the axial gage sensitivity.

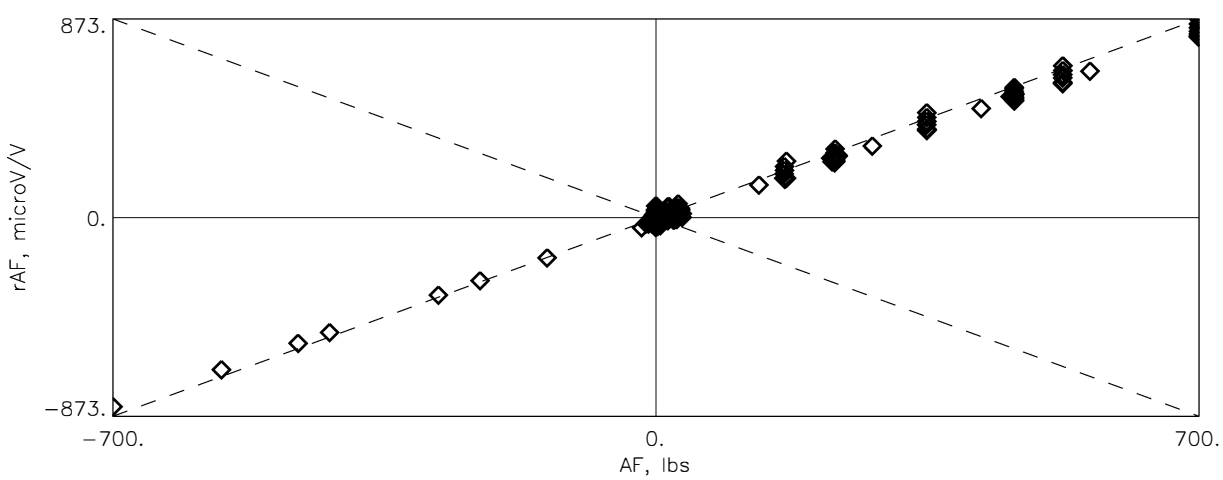

Fig. 3b - Axial gage output plotted versus the axial force for the 1210–point subset.

In the next step, *Method 1*, i.e., equal weighting of all data points, was chosen for the regression analysis

of the 1210–point subset of the original calibration data. Again, it was decided to apply the *Iterative Method* during the analysis and use the same math model that was selected to process the original 2091–point data set. The resulting axial force residuals for the 1210–point subset are shown in *Fig. 4a* below.

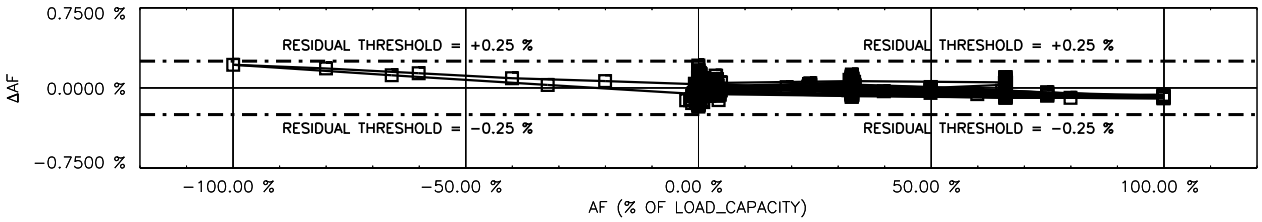

Fig. 4a - Axial force residuals plotted versus the axial force for the 1210–point subset after weighting Method 1 was applied during the regression analysis of the data.

The residuals plotted for the negative axial force range in *Fig. 4a* above belong to data points of load series 6 as only load series 6 has negative axial force loadings of significant magnitude (see also *Fig. 3a*). The residuals exhibit a distinct "slope" that did not exist when the original 2091–point data set was processed using *Method 1* (compare residuals plotted in *Fig. 2* with residuals plotted in *Fig. 4a*).

C. Results for Explicit Weighting (Method 2)

In the next step, *Method 2*, i.e., "explicit weighting," was used during the regression analysis of the 1210–point subset. Therefore, the weighting factor of each data point was computed by using Eq. (13*a*). The 1210–point subset also consists of the 16 load series that the original 2091–point calibration data set has (the chosen load series structure is shown in Ref. [6], Fig. 3, Fig. 4). The first six load series of the subset have a maximum number of <u>one</u> "intentionally loaded" load component. Therefore, the weighting factor of all data points from series 1 to series 6 is one. In all other cases, the weighting factors of the data points of each series will range from the weighting factor that is obtained for the maximum number of "intentionally loaded" load components of the given load series to the maximum value of one. The final list of the chosen weighting factor set of each load series of the 1210–point subset is given in Table 4 below.

Table 4: Weighting factors used by *Method 2* for the analysis of the MC60E calibration data.

Load Series	Maximum Number of Loaded Gages (see also Ref. [6], Fig. 3, Fig. 4)	Weighting Factor Set
1, 2, 3, 4, 5, 6	1	1.0000
7, 8, 9	2	0.2500, 1.000
10, 11, 12, 13	3	$0.111\overline{1}$, 0.2500, 1.000
14	4	0.0625, $0.111\overline{1}$, 0.2500, 1.000
15, 16	5	0.0400, 0.0625, $0.111\overline{1}$, 0.2500, 1.000

Again, it was decided to apply the *Iterative Method* to the calibration data using the same 28–term math model that was used to analyze the original 2091–point calibration data set. This time, however, the weighting factor sets listed in Table 4 above were used during the regression analysis of the data. The resulting axial force residuals for the 1210–point subset are shown in *Fig. 4b* below.

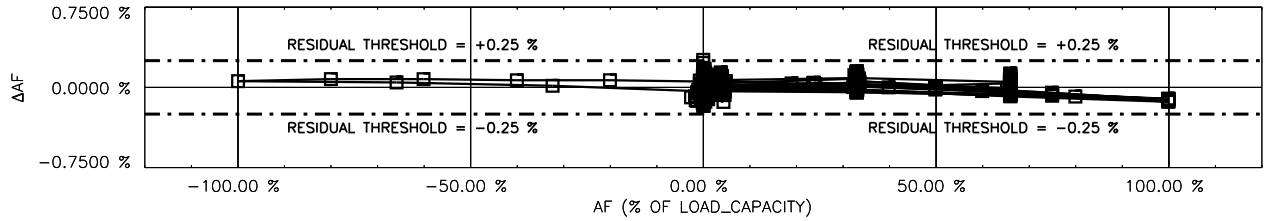

Fig. 4b - Axial force residuals plotted versus the axial force for the 1210–point subset after weighting Method 2 was applied during the regression analysis of the data.

The shown residuals no longer have the "slope" for the negative axial force range that was observed when the 1210–point subset was analyzed using *Method 1* (compare *Fig. 4b* with *Fig. 4a*). Therefore, it appears that the use of *Method 2* has made the axial force residuals of the "asymmetric" 1210–point subset look more like the axial force residuals of the original "symmetric" 2091–point calibration data set (compare *Fig. 4b* with *Fig. 2*). This observation can be quantified by comparing least squares estimates of the axial gage sensitivity for the three cases that were investigated so far (Case 1 $\equiv$ 2091–point data set & *Method 1*; Case 2 $\equiv$ 1210–point subset & *Method 1*; Case 3 $\equiv$ 1210–point subset & *Method 2*). The least squares estimate of the axial gage sensitivity is simply the coefficient of the axial force AF in the regression model of the axial gage output rAF if the *Iterative Method* is used to analyze balance calibration data. Table 5 below lists the least squares estimate of the axial gage sensitivity for the three cases that were investigated.

Table 5: Least squares estimates of the axial gage sensitivity of the MC60E.

Case Number	Number Of Data Points	Weighting Method	Axial Force Residual Plot	Axial Gage Sensitivity $\partial rAF/\partial AF$, $(\mu V/V)/lbs$
1	*2091*	*Method 1*	*Fig. 2*	1.1700
2	*1210*	*Method 1*	*Fig. 4a*	1.1692
3	*1210*	*Method 2*	*Fig. 4b*	1.1700

The result for Case 3, i.e., the axial gage sensitivity obtained after analyzing the 1210–point subset with *Method 2*, shows good agreement with the corresponding value that was obtained during the analysis of the original 2091–point data set using *Method 1* (see Case 1). Therefore, it is confirmed that the use of *Method 2* was beneficial during the analysis of the "asymmetric" 1210–point subset. *Method 2* was able to compensate for the negative influence of the partially missing negative axial force loadings on the least squares estimate of the axial gage sensitivity.

It is important to mention that the intentionally introduced "asymmetries" in the calibration load schedule of the axial force in the MC60E data set caused a residual shift of about 0.20 % of capacity if (i) the axial gage was loaded near −100 % of its capacity and (ii) *Method 1* instead of *Method 2* was applied during the regression analysis of the calibration data (compare *Fig. 4a* with *Fig. 4b*). The authors observed that load schedule "asymmetries" hidden in "real–world" data sets often resulted in residual shifts that were much higher than 0.20 % of capacity. Shifts between 0.50 % and 0.75 % near maximum load capacity were sometimes seen if *Method 1* (implicit weighting) instead of *Method 2* (explicit weighting) was applied during the regression analysis of those data sets. In all those cases, a simple switch from *Method 1* to *Method 2* reduced load residuals at load capacity to values that were equal or less than 0.10 % of capacity.

D. Implicit Weighting with Multiple Repeats

In theory, it is possible to reduce the negative impact of the "asymmetries" of the 1210–point subset on the least squares estimate of the axial gage sensitivity by adding "repeats" of load series 6 to the 1210–point subset. This load schedule change is expected to increase the "implicit" weight of points that influence the estimate of the axial gage sensitivity. Consequently, load residuals of negative axial forces should decrease and the least squares estimate of the axial gage sensitivity should improve. These conclusions can be demonstrated by adding, for example, nine "perfect" repeats of load series 6 to the 1210–point subset. The axial force load schedule of the resulting 1363–point data set is shown in *Fig. 5a* below.

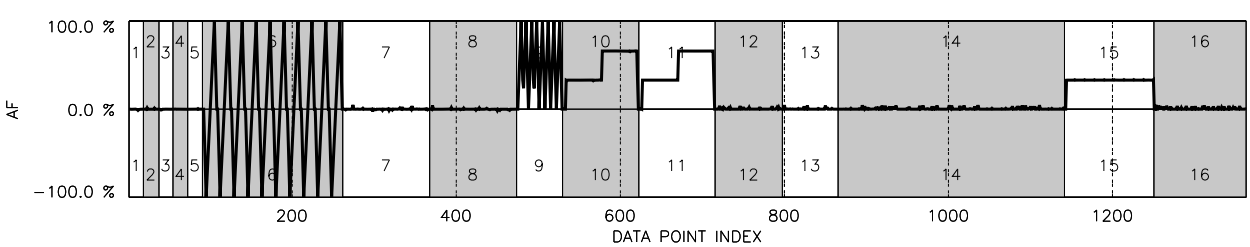

Fig. 5a - Axial force load schedule used for the 1363–point data set, i.e., the expanded 1210–point subset.

Differences between the 1210–point and the 1363–point load schedule can be seen by comparing *Fig. 3a* with *Fig. 5a*. In the next step, the 1363–point data set was analyzed by apply the *Iterative Method* to the data. The chosen math model was identical to the math model that was used to analyze the 1210–point subset. In addition, weighting *Method 1* was applied during the regression analysis of the data. The final axial force residuals for the 1363–point data set are shown in *Fig. 5b* below.

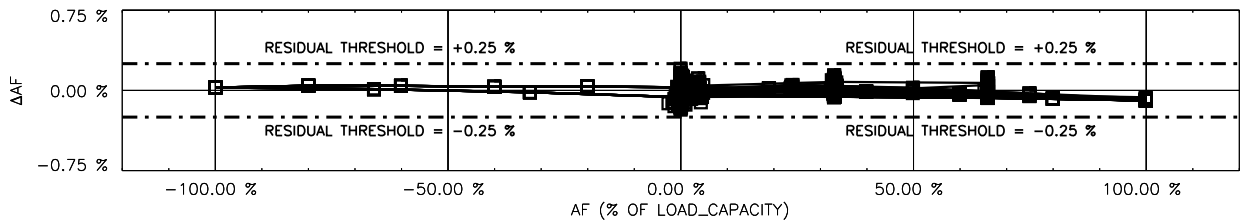

Fig. 5b - Axial force residuals plotted versus the axial force for the 1363–point data set after weighting Method 1 was applied during the regression analysis of the data.

The residuals of the negative axial forces no longer show the unwanted "slope" that was observed for the 1210–point subset even though *Method 1* was applied (compare *Fig. 5b* with *Fig. 4a*). In addition, as predicted, the residuals of the negative axial forces of the 1363–point data set are smaller than values that were observed for the 1210–point subset. The computed least squares estimate of the axial gage sensitivity of the 1363–point data set is 1.1704 $(\mu V/V)/lbs$. It is larger than the value of 1.1692 $(\mu V/V)/lbs$ that was computed for the 1210–point subset (see Table 5, Case 2). However, as expected, the axial gage sensitivity value of the 1363–point subset shows slightly better agreement with the reference sensitivity value of 1.1700 $(\mu V/V)/lbs$ that was computed for the original 2091–point machine calibration data set (see Table 5, Case 1).

VI. Summary and Conclusions

A new approach was presented that uses a weighted least squares fit for the analysis of wind tunnel strain–gage balance calibration data. The weighted least squares fit is specifically designed to increase the influence of single–component loadings during the regression analysis. A weighting factor between zero and one is assigned to each calibration data point that depends on a "count" of its "intentionally loaded" load components or gages. A variation of the authors' original weighting factor definition was also discussed that can only be applied to OFAT type balance calibration data set.

The authors' approach works well with both the *Iterative Method* and the *Non–Iterative Method* that are used for the analysis of strain–gage balance calibration data in the aerospace testing community. The *Iterative Method* uses a good estimate of the tare corrected load set as input for the determination of the weighting factors. The *Non–Iterative Method*, on the other hand, uses gage output differences relative to the natural zeros as input for the determination of the weighting factors. Machine calibration data of a six–component force balance was used to illustrate the application and potential benefits of the proposed weighted least squares fit. It was shown that the use of a weighted least squares fit during the analysis of balance calibration data can improve the least squares estimates of the sensitivity of a balance gage whenever a given calibration load schedule has a highly "asymmetric" distribution of the applied calibration loads.

It was also demonstrated that a load schedule structure change, e.g., the use of multiple repeats of a single–component load series, can have a positive impact on the "implicit" weighting that is acting during the regression analysis of balance calibration data. However, it is important to point out that beneficial load schedule structure changes are often difficult to execute during a "real–world" balance calibration considering (i) the multivariate nature of strain–gage balance behavior, (ii) calibration hardware and process limitations, and (iii) calibration schedule constraints and cost. The authors' weighted least squares fit, on the other hand, can easily be applied to an existing balance calibration data set as (i) no changes of the original load schedule need to be made, and, (ii) the "explicit" weighting factors are derived from very general characteristics of the balance calibration data.

Acknowledgement

The authors want to thank Bob Gisler and Jon Bader of NASA for their critical and constructive review of the final manuscript. The work reported in this paper was supported by the Wind Tunnel Division at NASA Ames Research Center under contract NNA16BD26C.

References

[1]AIAA/GTTC Internal Balance Technology Working Group, *Recommended Practice: Calibration and Use of Internal Strain–Gage Balances with Application to Wind Tunnel Testing*, *AIAA R–091–2003*, sponsored and published by the American Institute of Aeronautics and Astronautics, Reston, Virginia, 2003.

[2]Strang, G., *Introduction to Applied Mathematics*, Wellesley–Cambridge Press, Wellesley, Massachusetts, 1986, pp. 35–39 .

[3]Montgomery, D. C., Peck, E. A., and Vining, G. G., *Introduction to Linear Regression Analysis*, 4th ed., John Wiley & Sons, Inc., New York, 2006; pp. 179–183.

[4]Miller, R. G., Jr., *Beyond ANOVA: Basics of Applied Statistics*, reprint of 1st ed., Chapman & Hall – CRC Press LLC, Boca Raton (Florida), 1998; pp. 210–214.

[5]Ulbrich, N., "Comparison of Iterative and Non–Iterative Strain–Gage Balance Load Calculation Methods," *AIAA 2010–4202*, paper presented at the 27th AIAA Aerodynamic Measurement Technology and Ground Testing Conference, Chicago, Illinois, June/July 2010.

[6]Ulbrich, N., Gisler, R., and Kew, R., "Assessment of New Load Schedules for the Machine Calibration of a Force Balance," *AIAA 2015–2023*, paper presented at the 53rd AIAA Aerospace Sciences Meeting, Kissimmee, Florida, January 2015, Fig. 3, Fig. 4.

[7]Ulbrich, N., "Regression Model Optimization for the Analysis of Experimental Data," *AIAA 2009–1344*, paper presented at the 47th AIAA Aerospace Sciences Meeting, Orlando, Florida, January 2009.

[8]Burden, R. L. and Faires, J. D., *Numerical Analysis*, 3rd edition, PWS–Kent Publishing Company, Boston, Massachusetts, 1985, p. 334.

Appendix 1: Weighted PRESS Residual of a Response

1.1 General Remarks

PRESS residuals of the solution of a least squares fit for a given set of experiments data are important metrics that may be used for comparison and assessment of the predictive capabilities of different regression models. In principle, the calculation of PRESS residuals of a given data point requires only a few steps. At first, the data point is temporarily removed from the original data set. Then, the regression model is fitted using the remaining data points. In the next step, the regression model is used to predict the response of the withheld data point. Finally, the difference between the predicted and measured response of the withheld data point is computed. This difference is the PRESS residual of the response of the data point. A very detailed derivation of the PRESS residuals of the responses was presented in the appendices of Ref. [7]. However, this derivation assumes that an ordinary least squares fit of experimental data is performed and that all data points of the experimental data set have a weighting factor of one. Therefore, it became necessary to develop a more general derivation of the PRESS residuals of the responses that can also be used in connection with a weighted least squares fit. This derivation is presented for the first time in the three appendices of this paper.

The present derivation of the PRESS residual associated with a weighted least squares fit is complex. In principle, it follows a derivation of the PRESS residuals of an ordinary least squares fit that is outlined in Ref. [3] and discussed in more detail in Ref. [7]. Many intermediate steps and missing proofs of important auxiliary relationships were added to the present derivation of the PRESS residuals of a weighted least squares fit. These changes and additions have made the derivation more complete and easier to understand.

The derivation will show that a close connection between the classical response residual of a weighted least squares fit and the related PRESS residual of a response exists. Therefore, the derivation has two parts. At first, an equation for the classical weighted response residual is derived. Afterwards, it is shown how the equation of the weighted response residual is connected to the equation of the weighted PRESS residual.

1.2 Weighted Response Residual

The derivation of the classical weighted response residual starts by defining the regression model of a set of responses. The regression model is given by the following equation:

REGRESSION MODEL OF A RESPONSE

$$r_k \;=\; \underbrace{c_1 \cdot 1 \;+\; c_2 \cdot \xi_{1,k} \;+\; c_3 \cdot \xi_{2,k} \;+\; \ldots}_{m\ regression\ coefficients\ (c_1,\ c_2,\ \ldots,\ c_m)} \quad ; \quad 1 \le k \le p \qquad (15)$$

The regression model of each point of a given experimental data set can be written in a more compact format using a matrix and two vectors. In addition, weighting factors can be introduced for each data point (see Ref. [3], pp.179–183 for more details). Then, we get the following general equation for the weighted least squares fit (see also Eq. (5) in the body of the text):

$$\underbrace{\begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & w_p \end{bmatrix}}_{\mathbf{W}_{p\times p}} \cdot \underbrace{\begin{bmatrix} 1 & \xi_{1,1} & \xi_{2,1} & \cdots \\ 1 & \xi_{1,2} & \xi_{2,2} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \xi_{1,k} & \xi_{2,k} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \xi_{1,p} & \xi_{2,p} & \cdots \end{bmatrix}}_{\mathbf{A}_{p\times m}} \cdot \underbrace{\begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_m \end{bmatrix}}_{\mathbf{X}_{m\times 1}} = \underbrace{\begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & w_p \end{bmatrix}}_{\mathbf{W}_{p\times p}} \cdot \underbrace{\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_k \\ \vdots \\ r_p \end{bmatrix}}_{\mathbf{B}_{p\times 1}} \qquad (16a)$$

Equation (16a) may be written in abbreviated form using the matrix and vector symbols that were

introduced above. Then, we get for the matrix equation description of the weighted least squares problem the following expression

$$\textit{weighted least squares problem} \quad \Longrightarrow \quad \mathbf{W}_{p\times p} \cdot \mathbf{A}_{p\times m} \cdot \mathbf{X}_{m\times 1} \;=\; \mathbf{W}_{p\times p} \cdot \mathbf{B}_{p\times 1} \tag{16b}$$

The solution of the weighted least squares problem above was derived in great detail in the body of the text of the paper (see Eq. (12)). It can be summarized by the following matrix equations:

SOLUTION OF WEIGHTED LEAST SQUARES PROBLEM
(original experimental data set)

$$\mathbf{X}_{m\times 1} \;=\; \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1}_{m\times m} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{V}\right)_{m\times 1} \tag{17a}$$

$$\mathbf{U}_{p\times m} \;=\; \mathcal{W}_{p\times p} \cdot \mathbf{A}_{p\times m} \tag{17b}$$

$$\mathbf{V}_{p\times 1} \;=\; \mathcal{W}_{p\times p} \cdot \mathbf{B}_{p\times 1} \tag{17c}$$

where

$$\mathcal{W}_{\mathbf{p}\times\mathbf{p}} \;=\; \begin{bmatrix} \sqrt{w_1} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \sqrt{w_2} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \sqrt{w_3} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \sqrt{w_{p-1}} & 0 \\ 0 & 0 & 0 & \cdots & 0 & \sqrt{w_p} \end{bmatrix} \tag{17d}$$

It is convenient to define a column vector $\mathbf{Q}$ that can be used to write matrix $\mathbf{A}$ and its transpose $\mathbf{A}^{\mathbf{T}}$ in a format that is useful for the derivation of the residuals of a weighted least squares fit. This column vector and its transpose can be expressed as follows:

$$\left[\mathbf{Q}_k\right]_{m\times 1} \;=\; \begin{bmatrix} 1 \\ \xi_{1,k} \\ \xi_{2,k} \\ \vdots \end{bmatrix} \tag{18a}$$

$$\left[\mathbf{Q}_k^{\mathbf{T}}\right]_{1\times m} \;=\; \begin{bmatrix} 1 & \xi_{1,k} & \xi_{2,k} & \cdots \end{bmatrix} \tag{18b}$$

Now, after comparing the contents of matrix $\mathbf{A}$ shown in Eq. (16a) with the definition of vectors $\mathbf{Q}$ and $\mathbf{Q}^{\mathbf{T}}$ above, we conclude that matrix $\mathbf{A}$ and its transpose $\mathbf{A}^{\mathbf{T}}$ can be expressed as follows:

$$\mathbf{A}_{p\times m} \;=\; \begin{bmatrix} \mathbf{Q}_1^{\mathbf{T}} \\ \vdots \\ \mathbf{Q}_k^{\mathbf{T}} \\ \vdots \\ \mathbf{Q}_p^{\mathbf{T}} \end{bmatrix} \tag{18c}$$

$$\mathbf{A}^{\mathbf{T}}{}_{m\times p} \;=\; \begin{bmatrix} \mathbf{Q}_1 & \cdots & \mathbf{Q}_k & \cdots & \mathbf{Q}_p \end{bmatrix} \tag{18d}$$

Then, after applying Eq. (18c) to Eq. (17b) and knowing that $\mathcal{W}^{\mathbf{T}} = \mathcal{W}$ because $\mathcal{W}$ is a diagonal matrix, the matrix $\mathbf{U}$ of the weighted least squares problem and its transpose $\mathbf{U}^{\mathbf{T}}$ can be expressed as follows:

$$\mathbf{U}_{p\times m} \quad = \quad \mathcal{W}_{p\times p} \cdot \mathbf{A}_{p\times m} \quad = \quad \mathcal{W}_{p\times p} \cdot \begin{bmatrix} \mathbf{Q}_1^{\mathbf{T}} \\ \vdots \\ \mathbf{Q}_k^{\mathbf{T}} \\ \vdots \\ \mathbf{Q}_p^{\mathbf{T}} \end{bmatrix} \tag{19a}$$

$$\mathbf{U}_{m\times p}^{\mathbf{T}} \quad = \quad [\,\mathbf{Q}_1 \;\cdots\; \mathbf{Q}_k \;\cdots\; \mathbf{Q}_p\,] \cdot \mathcal{W}_{p\times p}^{\mathbf{T}} \quad = \quad [\,\mathbf{Q}_1 \;\cdots\; \mathbf{Q}_k \;\cdots\; \mathbf{Q}_p\,] \cdot \mathcal{W}_{p\times p} \tag{19b}$$

Finally, the weighted response residual of a data point with index k can be expressed as the difference between the original response r_k and the fitted response ρ_k. We get:

WEIGHTED RESPONSE RESIDUAL
(original experimental data set)

$$\Delta r_k \quad = \quad r_k \;-\; \rho_k \tag{20a}$$

$$\rho_k \quad = \quad \mathbf{Q}_k^{\mathbf{T}} \cdot \mathbf{X} \tag{20b}$$

The calculation of response residuals of all data points requires a <u>single</u> matrix inversion as $\mathbf{X}$, i.e., the solution of the global regression problem given in Eq. (17a) above, exclusively depends on the product of inverse matrix $(\mathbf{U}^{\mathbf{T}}\mathbf{U})^{-\mathbf{1}}$ with vector $\mathbf{U}^{\mathbf{T}}\mathbf{V}$.

1.3 Weighted PRESS Residual of Response

The weighted PRESS residual of a data point is closely related to its weighted response residual. Only one key difference exists. The weighted PRESS residual of a data point has to be computed using the global regression solution of the <u>modified</u> original experimental data set that has one fewer data point (i.e., the data point itself). Therefore, vectors and matrices of the original weighted least squares problem of the experimental data set change. They become a function of the omitted data point.

It is assumed that the omitted data point has the index k. Then, the weighted least squares problem of the <u>modified</u>, i.e., reduced, original data set can be described using the following equation:

WEIGHTED LEAST SQUARES PROBLEM
(modified experimental data set)

$$\left[\overline{\mathbf{U}}_k\right]_{(p-1)\times m} \cdot \left[\overline{\mathbf{X}}_k\right]_{m\times 1} \quad = \quad \left[\overline{\mathbf{V}}_k\right]_{(p-1)\times 1} \tag{21a}$$

where

$$\left[\overline{\mathbf{U}}_k\right]_{(p-1)\times m} = \overline{\mathcal{W}}_{(p-1)\times(p-1)} \cdot \overline{\mathbf{A}}_{(p-1)\times m} = \overline{\mathcal{W}}_{(p-1)\times(p-1)} \cdot \begin{bmatrix} \mathbf{Q}_1^{\mathbf{T}} \\ \vdots \\ \mathbf{Q}_{k-1}^{\mathbf{T}} \\ \mathbf{Q}_{k+1}^{\mathbf{T}} \\ \vdots \\ \mathbf{Q}_p^{\mathbf{T}} \end{bmatrix}_{(p-1)\times 1} \tag{21b}$$

$$\left[\overline{\mathbf{X}}_k\right]_{m\times 1} = \begin{bmatrix} \overline{c_1} \\ \overline{c_2} \\ \overline{c_3} \\ \vdots \\ \overline{c_m} \end{bmatrix}_{m\times 1} \tag{21c}$$

$$\left[\overline{\mathbf{V}}_k\right]_{(p-1)\times 1} = \overline{\mathcal{W}}_{(p-1)\times(p-1)} \cdot \overline{\mathbf{B}}_{(p-1)\times 1} = \overline{\mathcal{W}}_{(p-1)\times(p-1)} \cdot \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_{k-1} \\ r_{k+1} \\ \vdots \\ r_p \end{bmatrix}_{(p-1)\times 1} \tag{21d}$$

The weighting factors of the reduced data set are stored in a diagonal matrix $\overline{\mathcal{W}}$ that has one fewer row and column than the corresponding diagonal matrix $\mathcal{W}$ of the original data set. It looks as follows:

$$\overline{\mathcal{W}}_{(p-1)\times(p-1)} = \begin{bmatrix} \sqrt{w_1} & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ 0 & \sqrt{w_2} & \cdots & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sqrt{w_{k-1}} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & \sqrt{w_{k+1}} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & \sqrt{w_{p-1}} & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \sqrt{w_p} \end{bmatrix} \tag{21e}$$

The solution of the weighted least squares fit of the modified data set can easily be written down using the normal equations. In this case, similar to Eq. (17a) for the original experimental data set, we get:

SOLUTION OF WEIGHTED LEAST SQUARES PROBLEM
(modified experimental data set)

$$\overline{\mathbf{X}}_k = \left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{U}}_k\right)^{-1} \cdot \left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{V}}_k\right) \tag{22}$$

The weighted PRESS residual of a response can be computed using an equation that is very similar to the equation of the classical weighted response residual. Again, the weighted PRESS residual of a response of a data point with index k can be expressed as the difference between the original weighted response r_k and the fitted response $\overline{\rho}_k$ that was computed using the solution of the modified original data set. We get:

WEIGHTED PRESS RESIDUAL OF A RESPONSE
(modified experimental data set)

$$\delta r_k = r_k - \overline{\rho}_k \qquad (23a)$$

$$\overline{\rho}_k = \mathbf{Q}_k^{\mathbf{T}} \cdot \overline{\mathbf{X}}_k \qquad (23b)$$

It is interesting to mention that two metrics may be derived from the weighted PRESS residuals of a given experimental data set. These metrics may be used to assess the predictive capability of different regression models. The first metric is called the PRESS statistic. It is defined by the following relationship:

$$PRESS = \sum_{k=1}^{p} [\delta r_k]^2 \qquad (24)$$

The PRESS statistic is recommended in the literature for the comparison of different regression models (see, e.g., Ref. [3], p. 142). A related metric is the standard deviation of the PRESS residuals of all data points. It may also be used to compare regression models. This metric is defined as:

$$\sigma_{PRESS} = \sqrt{\frac{1}{p-1} \sum_{k=1}^{p} [\delta r_k]^2} \qquad (25)$$

It seems, superficially viewed, that the calculation of the weighted PRESS residuals of all data points of a data set is time consuming as a matrix inversion, i.e.,

$$\left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{U}}_k\right)^{-1}$$

has to be performed for each data point with index k of the data set (see also the right–hand side of Eq. (22)). However, it can be shown that the weighted PRESS residual of a data point can directly be computed using only (1) the original response residual, (2) the inverse matrix used in the weighted least squares fit of the original data set, and (3) two matrix multiplications. The proof of this surprising fact takes advantage of an important relationship between the inverse matrix used in Eq. (17a) and the inverse matrix used in Eq. (22). The relationship is derived in App. 2 of the present paper. It is given by the equation:

RELATIONSHIP BETWEEN INVERSE MATRICES
(see App. 2 for a derivation of the relationship)

$$\left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{U}}_k\right)^{-1} = \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} + \frac{\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \mathbf{Q}_k \cdot w_k \cdot \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1}}{1 - h_k} \qquad (26a)$$

$$h_k = \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \mathbf{Q}_k \cdot w_k \qquad (26b)$$

Consequently, after inserting Eq. (26a) into Eq. (22), i.e., the solution of the weighted least squares fit of the modified experimental data set, and after inserting the result into Eq. (23b), i.e., the equation of the fitted response $\overline{\rho}_k$, we get the following expression:

$$\overline{\rho}_k \quad = \quad \mathbf{Q}_k^{\mathbf{T}} \cdot \left[\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} + \frac{\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \mathbf{Q}_k \cdot w_k \cdot \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1}}{1 \; - \; h_k} \right] \cdot \left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{V}}_k\right) \tag{27}$$

Equation (27) can be simplified if (i) brackets on the right–hand side of Eq. (27) are expanded and (ii) a common denominator is used. Then, we get for the fitted response of the modified experimental data set the following equation:

$$\begin{aligned} \overline{\rho}_k \quad = \quad & \frac{[\; 1 \; - \; h_k \;] \cdot \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{V}}_k\right)}{1 \; - \; h_k} \\ + \quad & \frac{\left[\; \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \mathbf{Q}_k \cdot w_k \;\right] \cdot \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{V}}_k\right)}{1 \; - \; h_k} \end{aligned} \tag{28a}$$

Now, after using Eq. (26b) in order to replace the square bracket of the second fraction on the right–hand side of Eq. (28a), we get:

$$\begin{aligned} \overline{\rho}_k \quad = \quad & \frac{[\; 1 \; - \; h_k \;] \cdot \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{V}}_k\right)}{1 \; - \; h_k} \\ + \quad & \frac{h_k \cdot \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{V}}_k\right)}{1 \; - \; h_k} \end{aligned} \tag{28b}$$

Equation (28b) can be simplified further if the first fraction is expanded. Then, we can write:

$$\begin{aligned} \overline{\rho}_k \quad = \quad & \frac{\mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{U}}_k\right)}{1 \; - \; h_k} - \frac{h_k \cdot \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{V}}_k\right)}{1 \; - \; h_k} \\ + \quad & \frac{h_k \cdot \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{V}}_k\right)}{1 \; - \; h_k} \end{aligned} \tag{28c}$$

The second and third fraction on the right–hand side of Eq. (28c) cancel each other. Then, we get:

$$\overline{\rho}_k \quad = \quad \frac{\mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{V}}_k\right)}{1 \; - \; h_k} \tag{28d}$$

In the next step, after using Eq. (28d) to replace the fitted response on the right–hand side of Eq. (23a), we get for the weighted PRESS residual of the response the following expression:

$$\begin{aligned} \delta r_k \quad = \quad & r_k \; - \; \frac{\mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{V}}_k\right)}{1 \; - \; h_k} \\ = \quad & \frac{r_k \; - \; h_k \cdot r_k \; - \; \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{V}}_k\right)}{1 \; - \; h_k} \end{aligned} \tag{29}$$

Now, after using Eq. (26b) in order to replace h_k in the numerator of Eq. (29), we get for the weighted PRESS residual of the response:

$$\delta r_k \quad = \quad \frac{r_k \; - \; \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot [\; \mathbf{Q}_k \cdot w_k \cdot r_k \;] \; - \; \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{V}}_k\right)}{1 \; - \; h_k} \tag{30}$$

Equation (30) can be simplified significantly. From Eqs. (16b), (17c), and (17d) we know that

$$\mathcal{W}_{p\times p} \cdot \mathbf{B}_{p\times 1} = \begin{bmatrix} \sqrt{w_1}\cdot r_1 \\ \sqrt{w_2}\cdot r_2 \\ \vdots \\ \sqrt{w_k}\cdot r_k \\ \vdots \\ \sqrt{w_p}\cdot r_p \end{bmatrix} \tag{31a}$$

Consequently, it is concluded that

$$\mathbf{W}_{p\times p} \cdot \mathbf{B}_{p\times 1} = \mathcal{W}_{p\times p} \cdot \mathcal{W}_{p\times p} \cdot \mathbf{B}_{p\times 1} = \begin{bmatrix} w_1\cdot r_1 \\ w_2\cdot r_2 \\ \vdots \\ w_k\cdot r_k \\ \vdots \\ w_p\cdot r_p \end{bmatrix} \tag{31b}$$

Then, after combining Eq. (19b) with (17c), we can write the product $\mathbf{U}^{\mathbf{T}}\mathbf{V}$ in the following form:

$$\begin{aligned} \mathbf{U}^{\mathbf{T}}\ \mathbf{V} &= \underbrace{[\,\mathbf{Q}_1\ \cdots\ \mathbf{Q}_k\ \cdots\ \mathbf{Q}_p\,]\ \cdot\ \mathcal{W}_{p\times p}}_{from\ Eq.\ (19b)}\ \cdot\ \underbrace{\mathcal{W}_{p\times p}\ \cdot\ \mathbf{B}_{p\times 1}}_{from\ Eq.\ (17c)} \\ &= [\,\mathbf{Q}_1\ \cdots\ \mathbf{Q}_k\ \cdots\ \mathbf{Q}_p\,]\ \cdot\ \underbrace{\mathbf{W}_{p\times p}\ \cdot\ \mathbf{B}_{p\times 1}}_{from\ Eq.\ (31b)} \\ &= \sum_{i=1}^{p} \mathbf{Q}_i \cdot w_i \cdot r_i \end{aligned} \tag{32a}$$

Similarly, using Eq. (21b) and (21d), we can write

$$\begin{aligned} \overline{\mathbf{U}}_k^{\mathbf{T}}\ \cdot\ \overline{\mathbf{V}}_k &= \underbrace{[\,\mathbf{Q}_1\ \cdots\ \mathbf{Q}_{k-1}\ \ \mathbf{Q}_{k+1}\ \ \cdots\ \mathbf{Q}_p\,]\ \cdot\ \overline{\mathcal{W}}_{(p-1)\times(p-1)}}_{from\ Eq.\ (21b)}\ \cdot\ \underbrace{\overline{\mathcal{W}}_{(p-1)\times(p-1)}\ \cdot\ \overline{\mathbf{B}}_{(p-1)\times 1}}_{from\ Eq.\ (21d)} \\ &= [\,\mathbf{Q}_1\ \cdots\ \mathbf{Q}_{k-1}\ \ \mathbf{Q}_{k+1}\ \ \cdots\ \mathbf{Q}_p\,]\ \cdot\ \overline{\mathbf{W}}_{(p-1)\times(p-1)}\ \cdot\ \overline{\mathbf{B}}_{(p-1)\times 1} \\ &= \left[\sum_{i=1}^{p} \mathbf{Q}_i \cdot w_i \cdot r_i\right]\ -\ \mathbf{Q}_k \cdot w_k \cdot r_k \end{aligned} \tag{32b}$$

Then, after inserting Eq. (32a) into Eq. (32b), we get the following relationship:

$$\boxed{\overline{\mathbf{U}}_k^{\mathbf{T}}\ \cdot\ \overline{\mathbf{V}}_k = \mathbf{U}^{\mathbf{T}}\ \mathbf{V}\ -\ \mathbf{Q}_k \cdot w_k \cdot r_k \qquad (33)}$$

Rearranging terms in Eq. (33), we get:

$$\mathbf{Q}_k \cdot w_k \cdot r_k = \mathbf{U}^{\mathbf{T}}\ \mathbf{V}\ -\ \overline{\mathbf{U}}_k^{\mathbf{T}}\ \cdot\ \overline{\mathbf{V}}_k \tag{34}$$

Now, after using the right–hand side of Eq. (34) to replace the contents of the square bracket in the numerator of Eq. (30) and after simplifying terms, Eq. (30) can be written as follows:

$$\delta r_k = \frac{r_k\ -\ \mathbf{Q}_k^{\mathbf{T}}\ \cdot\ \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1}\ \cdot\ \mathbf{U}^{\mathbf{T}}\ \mathbf{V}}{1\ -\ h_k} \tag{35}$$

We also know, after inserting Eq. (17a) into Eq. (20b), that the fitted response of the weighted least squares problem of the original experimental data can be written in the following form:

$$\rho_k \quad = \quad \mathbf{Q}_k^{\mathbf{T}} \cdot \mathbf{X} \quad = \quad \mathbf{Q}_k^{\mathbf{T}} \cdot \underbrace{\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \mathbf{U}^{\mathbf{T}}\ \mathbf{V}}_{from\ Eq.\ (17a)} \tag{36}$$

Inserting Eq. (36) into Eq. (35), we get:

$$\delta r_k \quad = \quad \frac{r_k \ - \ \rho_k}{1 \ - \ h_k} \tag{37}$$

Finally, after (i) using Eq. (20a) to replace the numerator of Eq. (37) and after (ii) using Eq. (26b) to replace h_k in Eq. (37), we get for the weighted PRESS residual of the response the equation:

WEIGHTED PRESS RESIDUAL OF A RESPONSE

$$\delta r_k \quad = \quad \frac{\Delta r_k}{1 \ - \ h_k} \quad = \quad \frac{\Delta r_k}{1 \ - \ \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \mathbf{Q}_k \cdot w_k} \tag{38}$$

Equation (38) shows that the weighted PRESS residual of the response of a data point is essentially a scaled value of the corresponding weighted response residual of the data point. The scaling factor depends on (i) the inverse of matrix $\mathbf{U}^{\mathbf{T}}\mathbf{U}$ that is needed to obtain the solution of the original weighted least squares problem and (ii) a vector $\mathbf{Q}_k$ defined in Eq. (18a) that is a function of the regressor values of the data point.

Appendix 2: Derivation of Relationship between Inverse Matrices

An important relationship between two inverse matrices was used in App. 1 in order to derive a formula for the weighted PRESS residual of a response. The relationship is given by the following set of equations:

RELATIONSHIP BETWEEN INVERSE MATRICES

$$\left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{U}}_k\right)^{-1} = \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} + \frac{\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \mathbf{Q}_k \cdot w_k \cdot \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1}}{1 - h_k} \tag{26a}$$

$$h_k = \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \mathbf{Q}_k \cdot w_k \tag{26b}$$

The above relationship needs to be proven rigorously. This proof is possible if another relationship between products of sets of matrices is used. This relationship is given by the following expression:

RELATIONSHIP BETWEEN PRODUCTS OF MATRICES
(see App. 3 for a derivation of the relationship)

$$\left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{U}}_k\right) = \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right) - \mathbf{Q}_k \cdot w_k \cdot \mathbf{Q}_k^{\mathbf{T}} \tag{39}$$

A proof of Eq. (39) can be found in App. 3 of the present paper. For the time being it is assumed that Eq. (39) is valid so that the proof of Eqs. (26a) and (26b) above can be completed. The proof of Eq. (26a) and (26b) starts by realizing that the following relationship between a matrix and its inverse applies:

$$\mathbf{I} = \left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{U}}_k\right)^{-1} \cdot \left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{U}}_k\right) \tag{40}$$

Therefore, it is concluded that Eq. (26a) and (26b) can be proven by simply showing that the product of the right–hand side of Eq. (26a) with the right–hand side of Eq. (39) equals the identity matrix. The corresponding matrix product can be expressed as follows:

$$\mathbf{P} = \underbrace{\left[\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} + \frac{\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \mathbf{Q}_k \cdot w_k \cdot \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1}}{1 - h_k}\right]}_{right-hand\ side\ of\ Eq.\ (26a)} \cdot \underbrace{\left[\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right) - \mathbf{Q}_k \cdot w_k \cdot \mathbf{Q}_k^{\mathbf{T}}\right]}_{right-hand\ side\ of\ Eq.\ (39)} \tag{41}$$

Brackets on the right–hand side of Eq. (41) can be expanded. Then, we get:

$$\begin{aligned} \mathbf{P} = {} & \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right) - \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \mathbf{Q}_k \cdot w_k \cdot \mathbf{Q}_k^{\mathbf{T}} \\ & + \frac{\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \mathbf{Q}_k \cdot w_k \cdot \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)}{1 - h_k} \\ & - \frac{\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \mathbf{Q}_k \cdot w_k \cdot \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \mathbf{Q}_k \cdot w_k \cdot \mathbf{Q}_k^{\mathbf{T}}}{1 - h_k} \end{aligned} \tag{42}$$

We also know that

$$\mathbf{I} \quad = \quad \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right) \tag{43}$$

Then, after using Eq. (43) in order to simplify Eq. (42), we get:

$$\begin{aligned} \mathbf{P} \quad = \quad \mathbf{I} \; - \; \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \mathbf{Q}_k \cdot w_k \cdot \mathbf{Q}_k^{\mathbf{T}} \; + \; \frac{\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \mathbf{Q}_k \cdot w_k \cdot \mathbf{Q}_k^{\mathbf{T}} \; \cdot \; \mathbf{I}}{1 \; - \; h_k} \\ - \; \frac{\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \mathbf{Q}_k \cdot w_k \cdot \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \mathbf{Q}_k \cdot w_k \cdot \mathbf{Q}_k^{\mathbf{T}}}{1 \; - \; h_k} \end{aligned} \tag{44}$$

Terms in Eq. (44) can be rearranged. Then, Eq. (44) becomes:

$$\begin{aligned} \mathbf{P} \; = \; \mathbf{I} \; - \; \frac{\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \; \mathbf{Q}_k \; \cdot \; w_k \; \cdot \; [\; 1 \; - \; h_k \;] \; \cdot \; \mathbf{Q}_k^{\mathbf{T}}}{1 \; - \; h_k} \; + \; \frac{\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \mathbf{Q}_k \cdot w_k \cdot \mathbf{Q}_k^{\mathbf{T}}}{1 \; - \; h_k} \\ - \; \frac{\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \mathbf{Q}_k \cdot w_k \cdot \left[\; \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \; \mathbf{Q}_k \; \cdot \; w_k \;\right] \cdot \mathbf{Q}_k^{\mathbf{T}}}{1 \; - \; h_k} \end{aligned} \tag{45}$$

We also know from Eq. (26*b*) that the following relationship applies:

$$h_k \quad = \quad \left[\; \mathbf{Q}_k^{\mathbf{T}} \; \cdot \; \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \; \cdot \; \mathbf{Q}_{k_{m \times 1}} \; \cdot \; w_k \;\right] \tag{46}$$

Then, after using the left–hand side of Eq. (46) to replace the contents of the square bracket of the numerator of the third fraction of Eq. (45), we get

$$\begin{aligned} \mathbf{P} \quad = \quad \mathbf{I} \; - \; \frac{\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \; \cdot \; \mathbf{Q}_k \; \cdot \; w_k \; \cdot \; [\; 1 \; - \; h_k \;] \; \cdot \; \mathbf{Q}_k^{\mathbf{T}}}{1 \; - \; h_k} \\ + \; \frac{\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \; \cdot \; \mathbf{Q}_k \; \cdot \; w_k \; \cdot \; \mathbf{Q}_k^{\mathbf{T}}}{1 \; - \; h_k} \; - \; \frac{\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \; \mathbf{Q}_k \; \cdot \; w_k \; \cdot \; h_k \; \cdot \; \mathbf{Q}_k^{\mathbf{T}}}{1 \; - \; h_k} \end{aligned} \tag{47}$$

Finally, after combining the last two fractions of the right–hand side of Eq. (47), we get:

$$\begin{aligned} \mathbf{P} \quad = \quad \mathbf{I} \; - \; \frac{\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \; \cdot \; \mathbf{Q}_k \; \cdot \; w_k \; \cdot \; [\; 1 \; - \; h_k \;] \cdot \mathbf{Q}_k^{\mathbf{T}}}{1 \; - \; h_k} \\ + \; \frac{\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \; \cdot \; \mathbf{Q}_k \; \cdot \; w_k \; \cdot \; [1 \; - \; h_k] \; \cdot \; \mathbf{Q}_k^{\mathbf{T}}}{1 \; - \; h_k} \end{aligned} \tag{48}$$

The two fractions on the right–hand side of Eq. (48) cancel each other. Then, we get the final result:

$$\mathbf{P} \quad = \quad \mathbf{I} \tag{49}$$

Therefore, it is proven that Eq. (26*a*) and (26*b*) are valid as the product of the right–hand side of Eq. (26*a*) with the right–hand side of Eq. (39) equals the identity matrix.

Appendix 3: Derivation of Relationship between Products of Matrices

An auxiliary relationship between two sets of matrices is proven in this appendix that is urgently needed for the proof of Eq. (26a) in App. 2. The relationship has the following form:

RELATIONSHIP BETWEEN PRODUCTS OF MATRICES

$$\left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{U}}_k\right)_{m\times m} = \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)_{m\times m} - \mathbf{Q}_k \cdot w_k \cdot \mathbf{Q}_k^{\mathbf{T}} \qquad (39)$$

The proof of Eq. (39) can be accomplished by first showing that the following relationship is valid:

$$\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right) = \sum_{k=1}^{p} \mathbf{Q}_k \cdot w_k \cdot \mathbf{Q}_k^{\mathbf{T}} \qquad (50)$$

Matrix $\mathbf{U}$ above was defined in Eq. (19a). We know, by inspection, that the following relationship is valid:

$$\mathbf{U}_{p\times m} = \mathcal{W}_{p\times p} \cdot \mathbf{A}_{p\times m} = \mathcal{W}_{p\times p} \cdot \begin{bmatrix} \mathbf{Q}_1^{\mathbf{T}} \\ \mathbf{Q}_2^{\mathbf{T}} \\ \vdots \\ \mathbf{Q}_k^{\mathbf{T}} \\ \vdots \\ \mathbf{Q}_p^{\mathbf{T}} \end{bmatrix}_{p\times m} = \mathcal{W}_{p\times p} \cdot \left[\sum_{k=1}^{p} [\mathbf{e}_k]_{p\times 1} \cdot \left[\mathbf{Q}_k^{\mathbf{T}}\right]_{1\times m} \right] \qquad (51)$$

where the unit basis vector $\mathbf{e}_k$ can be expressed as

$$[\mathbf{e}_k]_{p\times 1} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_\mu \\ \vdots \\ e_p \end{bmatrix}_{p\times 1} \quad ; \quad e_\mu = \begin{cases} 0 & \text{if } \mu \neq k \\ 1 & \text{if } \mu = k \end{cases} \qquad (52)$$

and where vector $\mathbf{Q}_k^{\mathbf{T}}$ is given by Eq. (18b) as

$$\left[\mathbf{Q}_k^{\mathbf{T}}\right]_{1\times m} = [\ 1 \quad \xi_{1,k} \quad \xi_{2,k} \quad \cdots\ ]_{1\times m} \qquad (18b)$$

We also know, that the transpose of the sum of two matrices equals the sum of the transpose of each matrix (from Ref. [8], p. 334):

$$[\ \mathbf{C} + \mathbf{D}\ ]^{\mathbf{T}} = \mathbf{C}^{\mathbf{T}} + \mathbf{D}^{\mathbf{T}} \qquad (53a)$$

The transpose of matrix $\mathbf{U}$ may be obtained by simply transposing both sides of Eq. (51). Then, as $\mathcal{W}^{\mathbf{T}} = \mathcal{W}$, we get:

$$\mathbf{U}^{\mathbf{T}} = \left[\mathcal{W} \cdot \sum_{k=1}^{p} \mathbf{e}_k \cdot \mathbf{Q}_k^{\mathbf{T}} \right]^{\mathbf{T}} = \left[\sum_{k=1}^{p} \left[\ \mathbf{e}_k \cdot \mathbf{Q}_k^{\mathbf{T}}\ \right]^{\mathbf{T}} \right] \cdot \mathcal{W} \qquad (53b)$$

The following two theorems for (i) the transpose of a matrix and for (ii) the transpose of the product of two matrices apply (from Ref. [8], p. 334):

$$\mathbf{C} \quad = \quad \left(\mathbf{C}^{\mathbf{T}}\right)^{\mathbf{T}} \tag{54a}$$

$$(\ \mathbf{C}\ \mathbf{D}\ )^{\mathbf{T}} \quad = \quad \mathbf{D}^{\mathbf{T}}\ \mathbf{C}^{\mathbf{T}} \tag{54b}$$

Therefore, Eq. (53b) can also be written as:

$$\mathbf{U}^{\mathbf{T}} \quad = \quad \left[\ \sum_{k=1}^{p}\ \mathbf{Q}_k \cdot\ \mathbf{e}_k^{\mathbf{T}}\right] \cdot\ \mathcal{W} \tag{55}$$

Now, after multiplying the right–hand side of Eq. (55) with the right–hand side of Eq. (51), we get:

$$\begin{aligned}\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right) \quad &= \quad \left[\ \sum_{i=1}^{p}\ \mathbf{Q}_i \cdot\ \mathbf{e}_i^{\mathbf{T}}\right] \cdot\ \mathcal{W}\ \cdot\ \mathcal{W}\ \cdot\ \left[\ \sum_{j=1}^{p}\ \mathbf{e}_j \cdot\ \mathbf{Q}_j^{\mathbf{T}}\right] \\ &= \quad \left[\ \sum_{i=1}^{p}\ \mathbf{Q}_i \cdot\ \mathbf{e}_i^{\mathbf{T}}\right] \cdot\ \ \mathbf{W}\ \cdot\ \left[\ \sum_{j=1}^{p}\ \mathbf{e}_j \cdot\ \mathbf{Q}_j^{\mathbf{T}}\right]\end{aligned} \tag{56a}$$

The summation sign associated with index j may be moved in front of the term $\mathbf{Q}_i$ because the three terms $\mathbf{W}$, $\mathbf{Q}_i$, and $\mathbf{e}_i^{\mathbf{T}}$ are independent of index j. Then, we get:

$$\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right) \quad = \quad \sum_{i=1}^{p}\ \sum_{j=1}^{p}\ \left[\ \mathbf{Q}_i\ \cdot\ \mathbf{e}_i^{\mathbf{T}}\ \cdot\ \mathbf{W}\ \cdot\ \mathbf{e}_j\ \cdot\ \mathbf{Q}_j^{\mathbf{T}}\ \right] \tag{56b}$$

It is known, by inspection, that the following relationship between the product of two unit basis vectors with a diagonal matrix is valid:

$$\left[\mathbf{e}_i^{\mathbf{T}}\right]_{1\times p}\ \cdot\ \mathbf{W}_{p\times p}\ \cdot\ \ [\mathbf{e}_j]_{p\times 1} \quad = \quad \begin{cases} 0 & \text{if}\ \ i \neq j \\ w_i & \text{if}\ \ i = j \end{cases} \tag{57}$$

Consequently, a summation term used on the right–hand side of Eq. (56b) only differs from zero if the summation indices i and j are identical. Therefore, Eq. (56b) becomes:

$$\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right) \quad = \quad \sum_{i=1}^{p}\ \left[\ \mathbf{Q}_i\ \cdot\ w_i\ \cdot\ \mathbf{Q}_i^{\mathbf{T}}\ \right] \tag{58}$$

Now, after expanding the summation on the right–hand side of Eq. (58) and after moving the summation terms with index k to the left–hand side of the equation, we get:

$$\begin{aligned}\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)\ -\ \mathbf{Q}_k\ \cdot\ w_k\ \cdot\ \mathbf{Q}_k^{\mathbf{T}}\ \ &= \ \ \mathbf{Q}_1\ \cdot\ w_1\ \cdot\ \mathbf{Q}_1^{\mathbf{T}}\ +\ \mathbf{Q}_2\ \cdot\ w_2\ \cdot\ \mathbf{Q}_2^{\mathbf{T}}\ +\ \ldots \\ &\ldots\ +\ \mathbf{Q}_{k-1}\ \cdot\ w_{k-1}\ \cdot\ \mathbf{Q}_{k-1}^{\mathbf{T}}\ +\ \mathbf{Q}_{k+1}\ \cdot\ w_{k+1}\ \cdot\ \mathbf{Q}_{k+1}^{\mathbf{T}}\ +\ \ldots \\ &\ldots\ +\ \mathbf{Q}_{p-1}\ \cdot\ w_{p-1}\ \cdot\ \mathbf{Q}_{p-1}^{\mathbf{T}}\ +\ \mathbf{Q}_p\ \cdot\ w_p\ \cdot\ \mathbf{Q}_p^{\mathbf{T}}\end{aligned} \tag{59}$$

We also know, after applying Eq. (58) to a matrix $\overline{\mathbf{U}}$ that equals matrix $\mathbf{U}$ with the k–th row removed, that the following relationship is valid:

$$\begin{aligned}\left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{U}}_k\right) \quad &= \quad \mathbf{Q}_1\ \cdot\ w_1\ \cdot\ \mathbf{Q}_1^{\mathbf{T}}\ +\ \mathbf{Q}_2\ \cdot\ w_2\ \cdot\ \mathbf{Q}_2^{\mathbf{T}}\ +\ \ldots \\ &\ldots\ +\ \mathbf{Q}_{k-1}\ \cdot\ w_{k-1}\ \cdot\ \mathbf{Q}_{k-1}^{\mathbf{T}}\ +\ \mathbf{Q}_{k+1}\ \cdot\ w_{k+1}\ \cdot\ \mathbf{Q}_{k+1}^{\mathbf{T}}\ +\ \ldots \\ &\ldots\ +\ \mathbf{Q}_{p-1} \cdot w_{p-1} \cdot \mathbf{Q}_{p-1}^{\mathbf{T}}\ +\ \mathbf{Q}_p \cdot w_p \cdot \mathbf{Q}_p^{\mathbf{T}}\end{aligned} \tag{60}$$

It can be seen that the right–hand side of Eq. (59) equals the right–hand side of Eq. (60). Therefore, it is concluded that the left–hand side of Eq. (59) must equal the left–hand side of Eq. (60). This observation can be summarized as follows:

$$\boxed{\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{U}}_k \quad = \quad \mathbf{U}^{\mathbf{T}}\mathbf{U} \quad - \quad \mathbf{Q}_k \cdot w_k \cdot \mathbf{Q}_k^{\mathbf{T}} \qquad (61)}$$

Consequently, knowing that vector $\mathbf{Q}_k^{\mathbf{T}}$ is the k–th row of matrix $\mathbf{U}$, it is concluded that the right–hand side of Eq. (61) equals the matrix product <u>after</u> the k–th row is removed in matrix $\mathbf{U}$.

Appendix 4: Weighted PRESS Residual of a Balance Load

Strain–gage balances are used in wind tunnel testing to measure forces and moments that act on a wind tunnel model during a test. A strain–gage balance has to be calibrated so that measured electrical outputs of the strain–gages can be related to aerodynamic loads that the wind tunnel model experiences during the test. Typical load combinations are applied during the calibration of the balance and corresponding strain–gage outputs are recorded. Then, a regression analysis of the calibration data is performed that fits the measured electrical strain–gage outputs (responses) as a function of the applied calibration loads that are used to construct the regressors of the multivariate least squares fit. In a final step, the result of the regression analysis is used to define an iteration equation that makes it possible to directly compute loads, i.e., the value of the independent calibration variable, from the measured electrical outputs during a wind tunnel test. This iteration process is unique to the analysis and use of wind tunnel strain–gage balance calibration data. A more detailed explanation of the strain–gage balance calibration analysis process and the derivation of an iteration equation that allows for a direct calculation of balance loads from a given set of responses (measured electrical strain–gage output) is given in Ref. [1].

A calculation of the weighted PRESS residual of a balance load is of great interest to the wind tunnel testing community as balance loads are often the most important result of wind tunnel tests. Unfortunately, the weighted PRESS residual is computed for a response value (see App. 1). However, the author realized that weighted PRESS residuals of the balance loads can be computed in a two step process. This two step process combines (i) the regression analysis of the modified experimental data set with (ii) an iteration process for each data point.

The modified experimental data set needed for the calculation of the weighted PRESS residual of a balance load is identical with the modified data set that is discussed in App. 1. In principle, the calculation of the weighted PRESS residuals of the balance load needs the solution of the weighted least squares fit and the load iteration result for the data point if the data point (index k) itself is withheld from the original data set. The modified data set equals the original data set with the data point itself withheld. Consequently, matrix $\overline{\mathbf{U}}$ of the regression problem has one fewer row. The regression solution now becomes a function of the data point (index k) that is omitted for the calculation of the PRESS residuals.

The weighted least squares fit solution for the reduced calibration data set is needed to start the load iteration for a given data point where the weighted PRESS residual of the balance load is to be computed. It is assumed, similar to Eq. (21a) in App. 1, that the matrix and vector describing the weighted least squares problem of the modified data set are given as:

$$\overline{\mathbf{U}}_k \cdot \overline{\mathbf{X}}_k = \overline{\mathbf{V}}_k \tag{62}$$

The solution of the weighted least squares problem is simply given by the following:

$$\overline{\mathbf{X}}_k = \left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{U}}_k\right)^{-1} \cdot \left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{V}}_k\right) \tag{63}$$

It can be seen from Eq. (63) that the solution of the global regression problem of the modified data set is a function of the following matrices and vectors:

$$\overline{\mathbf{X}}_k \implies \mathcal{F}\left\{ \left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{U}}_k\right)^{-1} \; ; \; \overline{\mathbf{U}}_k \; ; \; \overline{\mathbf{V}}_k \right\}$$

It is obvious that the calculation of the inverse matrix on the right–hand side of Eq. (63) is a very time consuming operation if the solution of the weighted least squares problem has to be found. It appears, superficially viewed, that the calculation of the solution requires a complete calculation of the inverse matrix for each data point if the weighted PRESS residuals of all data points are to be determined. Fortunately, we can take again advantage of Eq. (26a) and (26b) that are derived in App. 2. These equations have the following form:

RELATIONSHIP BETWEEN INVERSE MATRICES

$$\left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{U}}_k\right)^{-1} = \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} + \frac{\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \mathbf{Q}_k \cdot w_k \cdot \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1}}{1 - h_k} \qquad (26a)$$

$$h_k = \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \mathbf{Q}_k \cdot w_k \qquad (26b)$$

We get, after inserting Eq. (26a) and (26b) into Eq. (63), the following solution of the global regression problem of the modified data set:

$$\overline{\mathbf{X}}_k = \left[\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} + \frac{\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \mathbf{Q}_k \cdot w_k \cdot \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1}}{1 - \mathbf{Q}_k^{\mathbf{T}} \cdot \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \cdot \mathbf{Q}_k \cdot w_k} \right] \cdot \left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{V}}_k\right) \qquad (64)$$

Now, the weighted least squares solution of the modified data set depends on the following matrices, vectors, and scalars:

$$\overline{\mathbf{X}}_k \implies \mathcal{F}\left\{ \left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1} \; ; \; \mathbf{Q}_k \; ; \; \overline{\mathbf{U}}_k \; ; \; \overline{\mathbf{V}}_k \; ; \; w_k \right\}$$

The originally required matrix inversion of the modified data set for each data point ($1 \leq k \leq p$), i.e.,

$$\left(\overline{\mathbf{U}}_k^{\mathbf{T}} \cdot \overline{\mathbf{U}}_k\right)^{-1}$$

has been replaced by simple matrix additions and multiplications that use the fixed inverse matrix of the original regression problem, i.e.,

$$\left(\mathbf{U}^{\mathbf{T}}\mathbf{U}\right)^{-1}$$

The inverse matrix of the original regression problem only has to be computed <u>once</u> for all data points. Therefore, the time consuming part of the calculation of the weighted PRESS residuals of the balance loads is <u>not</u> the calculation of the global regression solution of the reduced data set. Instead, it is the iteration that has to be performed for each data point after the solution of the weighted least squares problem of the modified data set is obtained. All matrices used in this iteration equation are directly derived for each data point with index k from the global regression solution $\overline{\mathbf{X}}_k$ of the modified, i.e., reduced, original balance calibration data set.