

A Statistical Theory of Bidirectionality

Richard DeLoach¹

NASA Langley Research Center, Hampton, Virginia, 23681

and

Norbert Ulbrich²

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

Original concepts related to the quantification and assessment of bidirectionality in strain-gage balances were introduced by Ulbrich in 2012. These concepts are extended here in three ways: 1) the metric originally proposed by Ulbrich is normalized, 2) a categorical variable is introduced in the regression analysis to account for load polarity, and 3) the uncertainty in both normalized and non-normalized bidirectionality metrics is quantified. These extensions are applied to four representative balances to assess the bidirectionality characteristics of each. The paper is tutorial in nature, featuring reviews of certain elements of regression and formal inference. Principal findings are that bidirectionality appears to be a common characteristic of most balance outputs and that unless it is taken into account, it is likely to consume the entire error budget of a typical balance calibration experiment. Data volume and correlation among calibration loads are shown to have a significant impact on the precision with which bidirectionality metrics can be assessed.

Nomenclature

Bidirectionality	=	A state in which strain-gage balance loads of opposite sign and equal magnitude generate electrical outputs of opposite sign but unequal magnitude.
Categorical	=	A type of variable only capable of assuming discrete levels, which may or may not be numerical
Coding	=	A linear transformation, typically involving centering and scaling, which maps numerical variables into some prescribed range such as [-1, +1]
Design space	=	A coordinate system in which each axis corresponds to a different factor in an experiment
Factor	=	Also called “independent variable.” A quantity for which levels are intentionally changed in an experiment
Level	=	A specific value for a given factor
Numerical	=	A type of variable capable of assuming any value from a continuum of numbers between lower and upper limits
PDF	=	Probability Density Function
Pure load	=	A load applied to one of the load channels only, with any loading of the other channels attributable to imperfections in the balance or loading alignment
Reference		
Distribution	=	A distribution of event probabilities associated with a specified hypothesis
RSM	=	Response Surface Model
Significant	=	Large enough to detect with a specified probability (statistical significance) or too large to ignore (practical significance)
Site	=	A specific combination of factor levels, represented geometrically as a point in a design space
b_{Ai}	=	Regression coefficient for term with absolute value load, index i
b_i	=	Regression coefficient without regard for bidirectionality, index i

¹ Senior Research Scientist, NASA Langley Research Center, MS 238, 4 Langley Blvd, Hampton, VA 23681, Associate Fellow, AIAA.

² Aerodynamicist, Jacobs Technology, Inc., NASA Ames Research Center.

b_{Si}	= Regression coefficient for term with signed load, index i
c_i	= Change induced by load polarity reversal in regression coefficient with index i
H	= Level of factor in physical units corresponding to +1 in coded units
H_A	= Alternative hypothesis: Bidirectionality at a great enough level to be of concern.
H_0	= Null hypothesis: No significant bidirectionality.
L	= Level of factor in physical units corresponding to -1 in coded units
n	= Number of points in calibration data sample
p	= Number of terms in polynomial response model, including the intercept
x_i	= Independent variable in coded units, index i
y	= Balance output at full calibration load, microV/V
y_{\pm}	= Balance output at full positive or negative calibration load, microV/V
z_{α}	= Standard normal deviate associated with significance level of α
z_{β}	= Standard normal deviate associated with significance level of β
α	= Maximum acceptable probability of erroneously rejecting the null hypothesis
β	= Maximum acceptable probability of erroneously rejecting the alternative hypothesis
Λ	= Bidirectionality metric
Λ_{\pm}	= Non-normalized bidirectionality metric for \pm full calibration load
σ_y	= Standard error in estimating balance output at full calibration load
σ_{Λ}	= Standard error in estimating the non-normalized bidirectionality coefficient
σ_{τ}	= Standard error in estimating the normalized bidirectionality coefficient
τ	= Normalized bidirectionality metric
τ_{\pm}	= Normalized bidirectionality metric for \pm full calibration load
ξ_i	= Independent variable in physical units, index i

I. Introduction

A full 2^{nd} -order polynomial is often used as a starting point for developing calibration equations for internal strain-gage balances. Eq. (1) represents such a full second-order polynomial in k factors, describing the i^{th} gage output.

$$y_i = b_0 + \sum_{i=1}^k b_i x_i + \sum_{i=1}^k \sum_{j>i}^k b_{ij} x_i x_j + \sum_{i=1}^k b_{ii} x_i^2 \quad (1)$$

The b_0 term is the intercept and from left to right the three summations represent first-order terms, mixed second-order (or interaction) terms, and pure quadratic terms, respectively. The x terms are loads, and the b terms are coefficients, typically quantified by the regression analysis of a suitable sample of calibration data.

A variety of calibration loading schedules have been used historically to obtain suitable regression coefficients. While differing in details, most of these loading schedules share certain common characteristics. They tend to be scaled either to the loading capacity of the balance under calibration, or to some sub-range of loads dictated by a proposed use of that balance, and they are often generally if not exactly symmetrical, with positive and negative loadings that are similar except for sign. Exceptions to the loading symmetry are made for balance elements that do not experience symmetrical loads in use.

If after calibration, a perfect balance was subjected first to a positive load of a specified magnitude, and then to a negative load of the same magnitude, the electrical outputs would be expected to differ in sign, but not in magnitude. That is, the outputs would be expected to be proportional to the loads for both positive and negative loads. Unfortunately, because of manufacturing imperfections, design characteristics, and other reasons, the electrical outputs can differ in magnitude as well as in sign, even when the loads differ only in sign. This phenomenon is known as bidirectionality.

Some degree of bidirectionality is common in balances of multi-piece construction. It can be attributed in large part to the fact that this type of construction permits different load paths through the balance for different loads.

While single-piece balances do not generally display significant bidirectionality, it may be unwise to assume that all such balances are immune to this phenomenon, absent some exonerating measurement that confirms insignificant bidirectionality. One reason is that even in a single-piece balance, slight second-order nonlinearities result in components of the output signal that are even functions of the applied load. These signal components are

represented mathematically by the quadratic terms in Eq. (1), and have the same algebraic sign whether the applied load is positive or negative. Such second-order nonlinearities ensure that the output signals are not exactly directly proportional to the loads, so that applied loads of the same magnitude but opposite sign produce output signals that differ slightly in magnitude as well as sign even in single-piece balances.

To address bidirectionality in the calibration of a balance, the response model of Eq. (1) can be augmented with a number of absolute value terms when bidirectionality is anticipated, as when multi-piece balances are calibrated. The AIAA Recommended Practice on strain-gage balance calibration, AIAA R-091-2003¹, offers the following augmented model:

$$y_i = b_0 + \left(\sum_{i=1}^k b_{Si} x_i + \sum_{i=1}^k \sum_{j>i}^k b_{Sij} x_i x_j + \sum_{i=1}^k b_{Sii} x_i^2 \right) + \left(\sum_{i=1}^k b_{Ai} |x_i| + \sum_{i=1}^k \sum_{j>i}^k b_{AAij} |x_i x_j| + \sum_{i=1}^k \sum_{j>i}^k b_{ASij} |x_i| x_j + \sum_{i=1}^k \sum_{j>i}^k b_{SAij} x_i |x_j| + \sum_{i=1}^k b_{SAii} x_i |x_i| \right) \quad (2)$$

Some modification to the nomenclature of the Recommended Practice has been introduced in Eq. (2) to highlight departures from Eq. (1). For example, subscripts *S* and *A* have been added to the regression coefficients to distinguish between terms of the original, non-augmented model that only involve (S)igned loads, and those that include (A)bsolute value terms. Terms from the model in Eq. (1), are displayed in black. For $k = 6$ primary loads, this model consists of an intercept plus 27 terms that are each functions of signed loads only. The 57 additional terms displayed in blue in Eq. (2) are the ones needed to account for bidirectionality using absolute values of the calibration loads. For $k = 6$ factors, the entire augmented 2^{nd} -order polynomial with absolute value terms consists of 85 terms, including the intercept.

There is a practical difficulty in finding the coefficients of a strain-gage balance calibration model with so many terms. The output of a strain-gage balance carries a finite amount of information about the applied load, which is distributed over the terms in the response model. The estimated output is constructed as the sum of contributions from all those terms. However, the information is not distributed uniformly among all the terms. For a pure loading, for example, the first-order primary load term typically carries more than 99% of all of the information, with all the other terms in the model accounting for the small amount of additional information needed to describe slight interactions and some non-linearity. In more complex compound loadings, most of the information is still carried by a relatively small subset of all the possible terms in a full-order polynomial, with the remaining information distributed over all the other terms.

If the calibration model is comprised of a very large number of terms as in Eq. (2), then many individual terms beyond the first-order primary load term will contribute relatively little to the output estimate. Furthermore, because the regression coefficients of each term are estimated empirically from imperfect experimental data, there will be some uncertainty in each estimate. With such a thin distribution of information over so many terms, the regression coefficient for any one term may not be large enough to significantly exceed the uncertainty in estimating it. In practice, this is often the case for many terms in a calibration response model. Such terms will carry more “noise” than “signal,” and are said to be “statistically insignificant.”

A process known as *model reduction* has been developed to cope with statistically insignificant terms in a response model. Model reduction refines the response model by rejecting or retaining terms based upon their signal to noise ratio. The details will not be described here in any depth, but they can be found in standard references on response surface modeling²⁻⁴. Specific details for strain-gage balance calibration applications can also be found in the literature⁵⁻⁹. The rejection/retention criterion for a given term is typically associated with some threshold probability that the estimated response would significantly change if that term were rejected, given the level of random error in the calibration data. Any term that can be eliminated from the model without significantly changing the response prediction is not retained.

Model reduction often eliminates half or more of the terms in a balance calibration response model. Because so many terms are eliminated that make no substantive contribution to the output, there is greater clarity in the resulting model. It is more compact, and therefore generates response predictions with a greater signal to noise ratio. Also, since only those terms responsible for true primary and secondary effects will remain, this reduced model will generally provide clearer insights into the operation of the balance.

Unfortunately, as noted above the practical mechanics of model reduction can be complicated when the starting point is a full model with as many terms as Eq. (2), especially when relatively little information is distributed over all of them. In such a case, if all of the terms were rank-ordered according to their signal to noise ratios, there would be very little difference in adjacent terms on such a list. This makes it difficult to establish a physically significant cut-off criterion for discarding terms with inadequate signal to noise ratio. Very small adjustments to what in practice will always be an arbitrary cut-off criterion can mean the difference between accepting and rejecting a relatively large number of similar, small candidate terms. Generally, no single such term will have a significant impact on the response model output. However, the cumulative effect of many terms that are either retained in the model or rejected can influence the output by an amount that, while small in absolute terms, can constitute a large portion of the error budget in a high-precision application such as an instrumentation calibration experiment.

It is only necessary to replace the relatively simple response model of Eq. (1) with the more complex model of Eq. (2) if the balance truly does display bidirectionality. It would therefore be useful if an assessment of bidirectionality could be made for any given balance, to objectively determine if tripling the number of calibration model terms is actually necessary in future calibrations. Likewise, many balances are calibrated under the assumption that they are free of bidirectionality, which it would be useful to objectively verify. It would also be useful to the balance design community to be able to rank-order balances according to the degree of their bidirectionality, in order to gain insights into which construction details, design elements, etc., are most conducive to bidirectionality.

In 2012, Ulbrich¹⁰ demonstrated how the bidirectionality of a balance could be established empirically, and proposed a metric for bidirectionality. He also proposed a standard for unacceptable levels of bidirectionality, developed at the NASA Ames Balance Calibration Laboratory. This paper presents an extension of Ulbrich's original work to include a useful normalization scheme, as well as uncertainty estimates for bidirectionality metrics generated from experimental calibration data. The uncertainty estimates facilitate a formal hypothesis test of the significance of bidirectionality detected in both multi-piece and single-piece balances.

In Section II we describe how Eq. (1) can be extended to account for bidirectionality without the added complexity of absolute value terms, by introducing a categorical variable into the analysis. The formula for a bidirectionality metric based on this categorical variable is also derived in Section II. Section III extends the formula for bidirectionality by introducing a useful normalization procedure. Uncertainty in the empirical estimation of bidirectionality is treated in Section IV.

Inference error probability and hypothesis testing is introduced in Section V and its relationship to uncertainty analysis is described. Section VI provides a detailed computational example using data from an existing calibration. Section VII presents results of applying the theory to four representative balances characterized by a range of bidirectionality conditions. Section VIII is a discussion of some of the details of the bidirectionality theory, and the paper ends with a summary and concluding remarks in Section IX.

II. Categorical Regression Variables and a Metric for Bidirectionality

In this section we will augment Eq. (1) by introducing a new variable, z , into the calibration equation. The new variable permits bidirectionality to be assessed without the additional complexity of absolute-value terms. It also facilitates the derivation of an explicit formula for bidirectionality that is easy to interpret and easy to implement. The formula is cast in terms of ordinary regression coefficients that fall out of the standard analysis of a calibration experiment; no additional measurements or analysis is required. Introducing this new variable also facilitates an assessment of the uncertainty in empirical estimates of bidirectionality. This, in turn, enables an objective test of the significance of any estimate of bidirectionality, which gives rise to a more reliable inference than simply comparing a bidirectionality estimate to some acceptability criterion without accounting for the uncertainty in that estimate. We expand on each of these points below.

The new variable that we will add to Eq. (1) is a *categorical* variable, z . Categorical variables differ from common numerical variables in one important respect. While numerical variables are generally free to assume any value from a continuum of possible values between specified limits, categorical variables are constrained to take on only certain discrete values. In a balance calibration experiment, the applied loads are numerical variables that can assume any value within the load range. The new categorical variable, z , is called the "load polarity", and has two discrete values. It assumes a value of "-1" for negative primary gage loads and "+1" for non-negative primary gage loads (including zero).

We augment Eq. (1) to include z and the interactions between z and the first- and second-order primary load terms. This leads to the following response model:

$$y_i = b_0 + \sum_{i=1}^k b_i x_i + \sum_{i=1}^k \sum_{j>i}^k b_{ij} x_i x_j + \sum_{i=1}^k b_{ii} x_i^2 + c_0 z + \sum_{i=1}^k c_i z x_i + \sum_{i=1}^k \sum_{j>i}^k c_{ij} z x_i x_j + \sum_{i=1}^k c_{ii} z x_i^2 \quad (3)$$

where b_0 is the intercept, the other b_i are regression coefficients for first-order and quadratic load terms as well as two-way load interactions, c_0 is the load polarity coefficient, and the other c 's are regression coefficients for various interactions between the polarity and load factors, as indicated. Here, y represents a particular primary gage output as before.

Since there is a unique load polarity variable for each primary gage load, to be rigorously correct we should index the z variable, and include all six such variables in the response model just as we included all six load variables. To do so, however, would introduce considerable complexity, accounting not only for bidirectionality in the primary gage output, but in how bidirectionality in secondary loads would impact the bidirectionality in primary loads, etc. Following Ulbrich¹⁰, these “second order effects of second-order effects” are ignored for purposes of quantifying bidirectionality in the primary gage output. This simplification is reflected in Eq. (3).

It is convenient to rewrite Eq. (3) by gathering terms, as follows:

$$y_i = (b_0 + c_0 z) + \sum_{i=1}^k (b_i + c_i z) x_i + \sum_{i=1}^k \sum_{j>i}^k (b_{ij} + c_{ij} z) x_i x_j + \sum_{i=1}^k (b_{ii} + c_{ii} z) x_i^2 \quad (4)$$

Again following Ulbrich¹⁰, we ignore the role of secondary loads on estimates of bidirectionality. That is, while theoretically there could be an infinite number of bidirectionality levels corresponding to an infinite number of secondary loading combinations, we define bidirectionality in a specific primary gage output only for the corresponding primary gage load, and then only for the case in which all secondary loads are zero. We apply these constraints to Eq. (4) for the case in which x is the primary gage load and y is the corresponding primary gage output for some specified gage for which bidirectionality is to be evaluated:

$$y = (b_0 + c_0 z) + (b_1 + c_1 z) x + (b_2 + c_2 z) x^2 \quad (5)$$

Here, we have dropped the “ i ” index from “ x_i ” because we are now dealing with a single specified primary gage load. We have also replaced the b_i , c_i , b_{ii} , and c_{ii} notation of Eq. (4) with b_1 , c_1 , b_2 , and c_2 , respectively. With the original notation, b_i and b_{ii} indicated first- and second-order coefficients for the i^{th} general gage load. Since we have dropped the general “ i ” subscript to simplify the notation, we now use “ b_1 ” and “ b_2 ”, respectively, for these coefficients, and we adopt analogous notational changes for the c -coefficients.

A clear physical interpretation of the c -coefficients follows when we note explicitly that in Eq. (5) the categorical variable, z , can only take on values of ± 1 :

$$y = (b_0 \pm c_0) + (b_1 \pm c_1) x + (b_2 \pm c_2) x^2 \quad (6)$$

If there were no bidirectionality, all c -coefficients in Eq. (6) would be zero and the primary gage response model when there are no secondary loads would be:

$$y = b_0 + b_1 x + b_2 x^2 \quad (7)$$

but if bidirectionality is present, c_0 , c_1 , and c_2 in Eq. (6) represent *changes* in the intercept, slope, and curvature that are induced by the bidirectionality. The “plus” signs are retained in Eq. (6) for positive primary gage loading

($z = +1$) and the “minus” signs are retained when the loading is negative ($z = -1$), although the c -coefficients themselves can be either positive or negative (or zero) in either case.

We seek a metric for bidirectionality that is in some sense a measure of how the output of a balance changes when bidirectionality is introduced. We want this metric to be zero when there is no bidirectionality and non-zero otherwise, with a magnitude that depends on the degree of bidirectionality. Again following Ulbrich¹⁰, we compute such a metric by first generating separate response models for the gage outputs under positive and negative loads. We then develop a model based on calibration loads that span the full negative/positive load range.

We use these models to estimate the primary gage output at full positive balance capacity and at full negative balance capacity. Ideally, the balance would be calibrated over its full loading range, but if the actual calibration range is less than the full load range, Ulbrich¹⁰ recommends extrapolating the response model to full load to facilitate comparisons of bidirectionality estimates made with calibration data sets that may have spanned different load ranges.

Ulbrich computes the difference between the primary gage output predicted by the model based on positive loads and the primary gage output predicted by the model based on all loads. He likewise computes the difference between the output of the negative-load model and the full-load model. The bidirectionality metric he recommends is the larger of the absolute values of these two differences. He proposes that the balance be declared bidirectional if this metric exceeds 0.5% of the balance output at full load capacity.

We follow an analogous development using the categorical variable response model introduced above. We fit a polynomial response model in six numerical load factors and six categorical load polarity factors, which reduces to Eq. (5) for the special case in which there are no secondary loads. No consider Eq. (5) for the special case of positive loading only, so $z = 1$. Call the resulting output prediction y_+ . Then

$$y_+ = (b_0 + b_1x + b_2x^2) + (c_0 + c_1x + c_2x^2) \quad (8)$$

The quantity y_+ represents the primary gage output of a bidirectional balance under positive load, x . If we subtract Eq. (7) from Eq. (8), we get the change in output attributable to bidirectionality when there is a positive load, x . We use the symbol, Λ_+ , to represent this difference:

$$\Lambda_+ = (b_0 + b_1x + b_2x^2 + c_0 + c_1x + c_2x^2) - (b_0 + b_1x + b_2x^2) \quad (9)$$

or

$$\Lambda_+ = c_0 + c_1x + c_2x^2 \quad (10)$$

By an analogous development we have for negative loading and therefore $z = -1$:

$$\Lambda_- = (b_0 + b_1x + b_2x^2 - c_0 - c_1x - c_2x^2) - (b_0 + b_1x + b_2x^2) \quad (11)$$

or

$$\Lambda_- = -c_0 - c_1x - c_2x^2 \quad (12)$$

This is the change in output attributable to bidirectionality when there is a negative load of magnitude x .

Following Ulbrich¹⁰, we would evaluate Λ_+ and Λ_- for x corresponding to the maximum positive and negative load capacities, respectively, and then select as the bidirectionality metric whichever one had the largest absolute value.

A. Coded Variables

One must use a particular primary gage load, x , to compute numerical values for the bidirectionality metrics of Eqs. (10) and (12). This computation can be simplified by specifying that the regression models be represented in terms of *coded variables*.

Variables are coded by means of a linear transformation performed prior to the regression analysis, in which the physical calibration load range is mapped linearly into a range that might span $[-1, +1]$ in coded units, for example. Eq. (13) shows such a transformation.

$$x = \frac{\xi - \frac{1}{2}(H + L)}{\frac{1}{2}(H - L)} \quad (13)$$

If ξ represents some primary gage load variable in a strain-gage balance calibration experiment expressed in physical units such as pounds or inch-pounds, and if H and L represents the upper and lower limits of the calibration load range—say, +2500 pounds and -2500 pounds for example—then x would be the load variable in coded units. Note that for $\xi = H$, $x = +1$, and for $\xi = L$, $x = -1$, so the coded variable does span $[-1, +1]$ in Eq. (13). If ξ is in the middle of the range; that is, if $\xi = L + \frac{1}{2}(H - L)$, then $x = 0$. The coding transformation of Eq. (13) thus centers the variable as well as scaling it. The coded variable assumes a value of zero in the center of the range, which may coincidentally correspond to a value of zero in physical units for a load range that is symmetrical about zero. For load ranges that are not symmetrical about zero, the coding transformation still centers the variable at the mid-range value, even if it is not zero in physical units.

Scaling and centering variables prior to a regression analysis achieve a number of benefits in the regression itself, quite apart from considerations of bidirectionality. A detailed treatment of these benefits is beyond the scope of this paper, but two are mentioned here for illustration. First, the regression coefficients become independent of the units in which physical variables are expressed. This simplifies comparisons of the relative contribution of linear effects, quadratic effects, and interactions. Secondly, the intercept is decoupled from the remaining regressor terms in the response model. That is, errors in estimating the other regressor terms in the model are not strongly influenced by errors in estimating the intercept. This decoupling leads to what Marquardt and Snee¹¹ describe as a reduction in “nonessential ill-conditioning” that is manifested in a reduction in the inflation of variance in individual regression coefficients than could otherwise be induced if there is correlation between the intercept and other model regressors. Model prediction is therefore improved by coding the variables, especially at off-design points within the design space. Montgomery, Peck, and Vining¹² show how this decoupling simplifies calculations of the confidence interval associated with model predictions. Further details about the advantages of coding in general regression analysis can be found elsewhere¹³ but in this paper we highlight a simplification in the bidirectionality metric that ensues if the regression model is based on coded variables rather than variables in physical units. Note that any regression model cast originally in coded units can always be recast in physical units at the end of the analysis by simply invoking this inverse of the transformation of Eq. (13):

$$\xi = \frac{1}{2}[H(x + 1) - L(x - 1)] \quad (14)$$

The resulting response model in physical units will display the same functional dependence as the model in coded variables displayed if the model is hierarchical; that is, if all sub-elements of each term in the model are also represented in the model (For example, a model featuring an AB^2 term is hierarchical if it also has a first-order “A” term, a quadratic “B²” term, and a first-order “B” term). A detailed discussion of hierarchy is beyond the scope of this paper, but it is not uncommon for response modeling software to impose it as an option⁷ or as a strongly recommended default¹⁴. Valid transitions between coded and physical units are therefore practical even for fairly complex models. References 15 and 16 provide further information on hierarchy.

B. Bidirectionality Metric Using Coded Variables

When the bidirectionality metrics in Eqs. (10) and (12) are expressed in terms of regression coefficients for a coded-variable model, it is necessary to express the primary gage load, x , as a coded variable also. Quantifying bidirectionality at the maximum calibration load generates a more accurate indicator variable than using some intermediate load. This also results in a substantial simplification of the bidirectionality metric if the coding variables are scaled to the full calibration range of the balance data sample (which is not necessarily the full load range of the balance). This is because in that case, $x = 1$ in Eqs. (10) and (12) and explicit references to the load for which bidirectionality is estimated can be dropped. The load is implicitly incorporated into the metric by virtue of the load coding and the convention that bidirectionality is evaluated at the highest positive and negative calibration loads of the primary strain gage.

Equation (10) represents bidirectionality when the balance is under a positive load, in which case $x = +1$. Likewise, Eq. (12) represents bidirectionality when the balance is under a negative load. In coded units, $x = -1$ in that case. The two bidirectionality metrics can then be expressed as follows:

$$\begin{aligned}\Lambda_+ &= c_1 + (c_0 + c_2) \\ \Lambda_- &= c_1 - (c_0 + c_2)\end{aligned}\tag{15}$$

or

$$\Lambda_{\pm} = c_1 \pm (c_0 + c_2)\tag{16}$$

Following Ulbrich¹⁰, we will characterize bidirectionality by the one with the largest absolute value:

$$\Lambda = \text{MAX} \left[\text{ABS}(\Lambda_+), \text{ABS}(\Lambda_-) \right]\tag{17}$$

Equation (16) makes it clear that the bidirectionality indicator variable reduces to the algebraic sum of three regression coefficients. These coefficients, by Eq. (3), correspond to the categorical variable, z , introduced to characterize bidirectionality, and the interaction of that categorical variable with the mean gage load slope and curvature terms. These coefficients directly quantify changes in intercept, slope, and curvature that are attributable to bidirectionality. It is therefore not surprising that the bidirectionality matrix is a simple linear combination of them.

Note also that insofar as the regression coefficients depend on the dynamic range of the calibration loads, the bidirectionality metric may also be load-dependent. That is, to such an extent that the calibration equation might be load-dependent so that the calibration over a restricted load range differs from the calibration over the full load range of the balance, say, the bidirectionality metric may also be load-dependent. It is recommended that the calibration load range be specified whenever a value is reported for the bidirectionality metric, Λ .

III. Normalizing the Bidirectionality Indicator Variable

The first step in assessing bidirectionality is to quantify it, but it is not sufficient to simply discover that the bidirectionality of a balance is non-zero; it must also exceed some consensus threshold to be regarded as great enough to be of concern. As of this writing there is no industry-wide consensus as to what such a threshold should be, but Ulbrich's original proposal for a threshold of 0.5% of the primary gage output at full load will be used in the numerical examples that follow.

To determine whether the bidirectionality computed by Eqs. (16) and (17) exceeds the acceptable limit, one must first compute that limit by using the response model to estimate the primary gage output at either the maximum positive primary gage load or the maximum negative primary gage (depending on which version of Λ in Eq. (16) is used, based on Eq. (17)). The absolute value of this quantity is then multiplied by 0.005 and compared with the value of Λ computed by Eq. (17). Whether or not the balance is subject to being declared "bidirectional" will depend on the relative magnitude of Λ and the computed 0.5% of output at peak load.

This procedure has some minor drawbacks that can be easily eliminated. First, a separate bidirectionality threshold must be calculated for every combination of primary gage and maximum calibration load for that gage. This entails some additional calculation that is not actually needed, as we will soon show. Secondly, the approach as outlined thus far casts bidirectionality in absolute physical units, which complicates direct comparisons among different balances. For example, if for one balance, $\Lambda = 5$ millivolts/volt, and for another, $\Lambda = 10$ millivolts/volt, which displays the greater multicollinearity? It is not possible to say without additional information about the maximum calibration load and the corresponding balance output. Only if both balances have the same output sensitivity and are each calibrated over the same load range can a direct comparison be made of bidirectionality metrics computed in this way. Even then, one would need to know the actual outputs at maximum calibration load to say if either level of bidirectionality is of concern.

If one balance is calibrated over different load ranges at different times, comparisons of absolute bidirectionality measures would require some normalization of calibration loads. As a practical matter, this would entail

extrapolating to a convenient comparison point, such as the physical load capacity of the balance. However, the comparison of extrapolated non-linear results is always problematical.

Comparisons of bidirectionality metrics computed for the same balance using data sets that span different load ranges, and comparisons among balances of different output sensitivity, can be more readily made if the bidirectionality metric is normalized by the primary gage output at maximum calibration load. Since this load never exceeds the calibration load range, difficulties associated with extrapolation are avoided. Such normalization also permits an immediate assessment as to whether the level of bidirectionality is great enough to be of concern.

Equation (5) permits us to normalize by an estimate of the primary gage output at maximum calibration load, x_{\max} , when there is no bidirectionality, by setting the three c -coefficients in Eq. (5) to zero. We will call this output $Y_{A=0}$. This results in the following:

$$Y_{A=0} = b_0 + b_1 x_{\max} + b_2 x_{\max}^2 \quad (18)$$

The load in Eq. (18), x_{\max} , is expressed as a coded variable. By the convention proposed earlier, bidirectionality will be evaluated at maximum calibration load, which in coded units is either at $x_{\max} = -1$ or at $x_{\max} = +1$. We will normalize the A of Eq. (15) by the $Y_{A=0}$ of Eq. (18) evaluated at $x_{\max} = -1$, and we will normalize the A_+ of Eq. (15) by $Y_{A=0}$ of Eq. (18) evaluated at $x_{\max} = +1$. We will use τ and τ_+ , respectively, to designate these normalized bidirectionality metrics.

Note that the coded numerical variable representing maximum calibration load, x , and the categorical load variable, z , are correlated in Eq. (5). Each of these two variables can assume one of the two values, -1 or +1, but the primary gage load, x , is only at $x_{\max} = +1$ when the gage loading is positive ($z = +1$). Likewise, x is only at $x_{\max} = -1$ for negative primary gage loading, for which z is also at -1. So τ is only defined for x and z both at -1 and τ_+ is only defined for x and z both at +1. Thus we have:

$$\begin{aligned} \tau_+ &= \frac{c_1 + (c_0 + c_2)}{(b_0 + b_2) + b_1} \\ \tau_- &= \frac{c_1 - (c_0 + c_2)}{(b_0 + b_2) - b_1} \end{aligned} \quad (19)$$

or

$$\tau_{\pm} = \frac{c_1 \pm (c_0 + c_2)}{(b_0 + b_2) \pm b_1} \quad (20)$$

Eq. (20) is the normalized analogy of Eq. (16). As in the case of the non-normalized bidirectionality metric, we follow Ulbrich¹⁰ and select that variation of the metric with the largest absolute value:

$$\tau = \text{MAX} \left[\text{ABS}(\tau_+), \text{ABS}(\tau_-) \right] \quad (21)$$

It will be convenient later to let y_+ and y_- represent the primary gage output evaluated at $x = -1$ and $x = +1$, respectively, with $z = 0$, as estimated by Eq. (5):

$$\begin{aligned} y_+ &= (b_0 + b_2) + b_1 \\ y_- &= (b_0 + b_2) - b_1 \end{aligned} \quad (22)$$

or

$$y_{\pm} = (b_0 + b_2) \pm b_1 \quad (23)$$

That is, y_+ and y_- in Eqs. (22) and (23) represent the primary gage outputs at the positive and negative extremes of the primary gage calibration load range (not necessarily the load capacity of the balance) that would be estimated from a fit to the full calibration data set (positive and negative loads) without regard to bidirectionality.

Figure 1 illustrates this normalization graphically. The tangent of the angle θ in this figure is $\Lambda/Y_{\Lambda=0}$, which represents the normalized bidirectionality metric. That is, $\tan(\theta) = \Lambda/Y_{\Lambda=0} = \tau$. Note that this angle will display a mild dependence on the maximum calibration load, because Λ and Y are both non-linear functions of x . However, because the nonlinearities are relatively small, the normalized bidirectionality metric should be fairly stable over practical ranges of maximum calibration load.

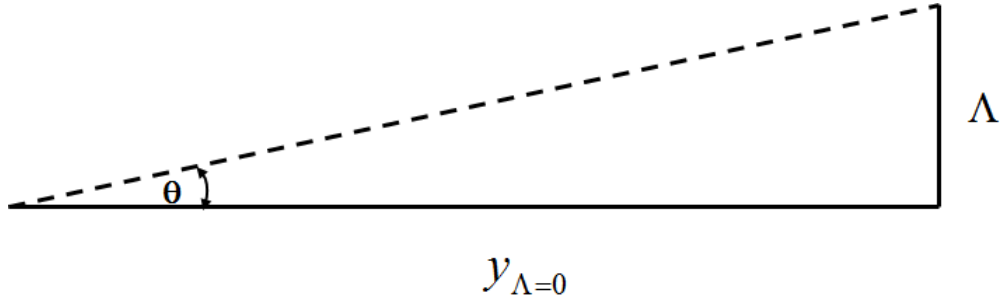


Figure 1. Bidirectionality is normalized by dividing it by the principal gage output at full calibration load. Slope of angle is normalized metric.

As in the case of the non-normalized bidirectionality metric of Eq. (16), the bidirectionality metric will be load-dependent if the regression coefficients depend on the dynamic range of the calibration loads. That is, if the calibration over a restricted load range differs from the calibration over the full load range of the balance, say, the bidirectionality metric may also be load-dependent. It is recommended that the calibration load range be specified whenever a value is reported for the normalized bidirectionality metric, τ , just as was recommended above for the non-normalized bidirectionality metric.

IV. Uncertainty in the Bidirectionality Indicator Variable

The normalized and non-normalized bidirectionality metrics are both computed using regression coefficients that are estimated from imperfect experimental data. These experimental errors in the balance calibration data translate into uncertainty in the bidirectionality estimates. Therefore, we cannot simply compare an estimate of bidirectionality to some fixed criterion such as 0.5% of primary gage output at maximum calibration load, and infer with complete confidence that the balance is or is not bidirectional. It will depend on how much uncertainty there is in the bidirectionality estimate. If experimental uncertainty caused an underestimation of bidirectionality, one might fail to detect a truly bidirectional balance. Likewise, experimental error could cause a balance with negligible bidirectionality to be improperly characterized as bidirectional.

Extreme bidirectionality is easy to detect, so this case is not very interesting from a balance characterization perspective. Levels of bidirectionality large enough to be troubling but too small to be obvious are more interesting, because the more subtle the bidirectionality, the easier it is to make an inference error. The probability of making an improper inference in these more challenging cases depends on how much uncertainty there is in estimating bidirectionality. We will therefore consider that uncertainty in some detail here.

A. Uncertainty in the Non-Normalized Bidirectionality Indicator Variable

Consider the two Eqs. (16), which represent interim non-normalized bidirectionality metrics for positive and negative loads. (Recall that the final metric is the one of these with the largest absolute value). Each is a function of three regression coefficients, c_0 , c_1 , and c_2 . There is some variance associated with the estimate of each of the three coefficients, due to experimental error in the data used to compute them. This variance can be estimated from the regression covariance matrix.

We can propagate these regression coefficient variances into an estimate of the variance in the bidirectionality metric by exploiting a well-known general error propagation formula. For the special case of a function of three variables, $y = f(x_1, x_2, x_3)$ with all x_i independent of each other (with uncorrelated errors), the following relationship exists between the variance in y and the variances in the x_i :

$$\sigma_y^2 = \left(\frac{\partial y}{\partial x_1} \right)^2 \sigma_{x_1}^2 + \left(\frac{\partial y}{\partial x_2} \right)^2 \sigma_{x_2}^2 + \left(\frac{\partial y}{\partial x_3} \right)^2 \sigma_{x_3}^2 \quad (24)$$

We note in passing that straight-forward variations of this formula will accommodate more than three independent variables, or fewer than three. We will use this formula for the case of two independent variables later.

Let $y = A_+ = A_- = A$, and let $x_1 = \pm c_0$, $x_2 = c_1$, and $x_3 = \pm c_2$. Then by applying Eq. (24) to Eqs. (16):

$$\left(\frac{\partial y}{\partial x_i} \right)^2 = 1 \text{ for } i = [1..3] \rightarrow \sigma_{A_+}^2 = \sigma_{A_-}^2 = \sigma_A^2 = (\sigma_{c_0}^2 + \sigma_{c_1}^2 + \sigma_{c_2}^2) \quad (25)$$

Therefore,

$$\sigma_A = \sqrt{\sigma_{c_0}^2 + \sigma_{c_1}^2 + \sigma_{c_2}^2} \quad (26)$$

Eq. (26) provides a useful geometrical interpretation of the standard error in A , the non-normalized bidirectionality metric. It is simply the diagonal of a three-dimensional rectangle with sides of length σ_{c_0} , σ_{c_1} , and σ_{c_2} . That is, it is a vector with the standard errors of the three regression coefficients as components that determine its overall magnitude. As with the bidirectionality metric itself, since the calibration regression coefficients may be load dependent, so may the standard error in estimating bidirectionality, Eq. (26).

B. Uncertainty in the Normalized Bidirectionality Indicator Variable

The normalized bidirectionality metric of Eq. (20) is somewhat more complicated than the non-normalized metric of Eq. (16) in that it is a function of six regression coefficients rather than three. There is uncertainty in each of the six coefficients, which must be propagated into an uncertainty estimate for the normalized bidirectionality indicator variable.

Extending Eq. (24) to a function of six variables, the general uncertainty propagation formula for $\tau_{\pm} = f(c_0, c_1, c_2, b_0, b_1, b_2)$ becomes:

$$\sigma_{\tau_{\pm}}^2 = \left(\frac{\partial \tau_{\pm}}{\partial c_0} \right)^2 \sigma_{c_0}^2 + \left(\frac{\partial \tau_{\pm}}{\partial c_1} \right)^2 \sigma_{c_1}^2 + \left(\frac{\partial \tau_{\pm}}{\partial c_2} \right)^2 \sigma_{c_2}^2 + \left(\frac{\partial \tau_{\pm}}{\partial b_0} \right)^2 \sigma_{b_0}^2 + \left(\frac{\partial \tau_{\pm}}{\partial b_1} \right)^2 \sigma_{b_1}^2 + \left(\frac{\partial \tau_{\pm}}{\partial b_2} \right)^2 \sigma_{b_2}^2 \quad (27)$$

The variance for each of the regression coefficients is available from the covariance matrix, as before. The partial derivatives for τ_{\pm} are computed from Eq. (20):

$$\begin{aligned} \left(\frac{\partial \tau_{\pm}}{\partial c_i} \right)^2 &= \frac{1}{[(b_0 + b_2) \pm b_1]^2} & i = 0..2 \\ \left(\frac{\partial \tau_{\pm}}{\partial b_i} \right)^2 &= \left\{ \frac{c_1 \pm (c_0 + c_2)}{[(b_0 + b_2) \pm b_1]^2} \right\}^2 & i = 0..2 \end{aligned} \quad (28)$$

Inserting Eqs. (28) into Eq. (27) and rearranging terms:

$$\sigma_{\tau_{\pm}}^2 = \frac{\left[(b_0 + b_2) \pm b_1 \right]^2 (\sigma_{c_0}^2 + \sigma_{c_1}^2 + \sigma_{c_2}^2) + \left[c_1 \pm (c_0 + c_2) \right]^2 (\sigma_{b_0}^2 + \sigma_{b_1}^2 + \sigma_{b_2}^2)}{\left[(b_0 + b_2) \pm b_1 \right]^4} \quad (29)$$

Equation (29), the equation for the variance in τ_{\pm} , can be simplified by recognizing that it contains as one of its elements the variance in the non-normalized bidirectionality metric, Λ , first encountered above in Eq. (25) and repeated here for convenience:

$$\sigma_{\Lambda}^2 = \sigma_{c_0}^2 + \sigma_{c_1}^2 + \sigma_{c_2}^2 \quad (25)$$

Equation (29) also contains the equation for the non-normalized bidirectionality metric itself, first introduced in Eq. (16) and reproduced here for convenience:

$$\Lambda_{\pm} = c_1 \pm (c_0 + c_2) \quad (16)$$

Eq. (23), reproduced here for convenience, is also embedded in Eq. (29).

$$y_{\pm} = (b_0 + b_2) \pm b_1 \quad (23)$$

Finally, applying the general variance propagation formula of Eq. (24) to Eq. (23), we obtain this result:

$$\sigma_{y_{\pm}}^2 = \sigma_{b_0}^2 + \sigma_{b_1}^2 + \sigma_{b_2}^2 = \sigma_y^2 \quad (30)$$

where for clarity, we have dropped the “ \pm ” subscript from σ_y , since the signed terms on the right of Eq. (30) are all positive whether $y = y_+$ or $y = y_-$.

Inserting Eqs. (23), (25), (16), and (30) into Eq. (29):

$$\sigma_{\tau_{\pm}}^2 = \frac{y_{\pm}^2 \sigma_{\Lambda}^2 + \Lambda_{\pm}^2 \sigma_y^2}{y_{\pm}^4} \quad (31)$$

Eq. (31) is thus the variance of the normalized bidirectionality metric. Note also, by inserting Eqs. (16) and (23) into Eq. (20) we get the following intuitively clear formula for the normalized bidirectionality indicator variable under positive or negative load:

$$\tau_{\pm} = \frac{\Lambda_{\pm}}{y_{\pm}} \quad (32)$$

We could have arrived at Eq. (31) by simply applying the general variance propagation formula to Eq. (32). That is, we could have begun with Eq. (32) as a rational definition of normalized bidirectionality, and derived the same result as the more rigorous development produced, so it is possible to arrive at Eq. (31) by two different routes.

Note in Eq. (31) that the non-normalized indicator variables for positive and negative load, Λ_+ and Λ_- , are not necessarily equal, nor are the primary gage outputs at positive and negative load, y_+ and y_- . The variance in these two quantities is the same, however, for positive and negative load, so there is no polarity subscript for the variance of either variable. The physical interpretation of this result is that the uncertainty with which we can estimate the primary gage output at maximum calibration load, and the uncertainty with which we can estimate the *non-normalized* bidirectionality metric, are independent of load polarity. The uncertainty in estimating the *normalized*

bidirectionality metric *does* depend on polarity, however, because it is a function of the primary gage output at maximum calibration load, which can be different for positive and negative loads. The Appendix demonstrates that while the uncertainty in empirical estimates of the normalized bidirectionality metric is different for positive and negative loading, for symmetrical load schedules the difference is typically too small to have any practical significance. That is, for all practical purposes the precision with which normalized bidirectionality can be quantified is virtually independent of load polarity, as it was for the case of the non-normalized metric derived above.

V. Inference Error Probability and Hypothesis Testing

The only reason we quantify the bidirectionality of a balance is to determine if it is significantly bidirectional. If the bidirectionality is not zero, we still say the balance is not *significantly* bidirectional if it is below some threshold deemed to be worrisome. Only if we can say with some prescribed minimum level of confidence that the bidirectionality exceeds such a threshold are we entitled to declare the balance bidirectional.

Unfortunately, because of experimental uncertainty there is always some possibility of error in any assessment of bidirectionality. There are four possible outcomes of any such assessment. The balance may actually display significant bidirectionality (enough to be of concern) or it may not, but in either case we must *infer* whether the balance is significantly bidirectional or not. If the balance *is* bidirectional and we infer that it is, or if the balance is *not* bidirectional and we infer that it is not, the inference will have been valid. However, if the balance is *not* bidirectional and because of experimental error we infer that it *is*, or if the balance *is* bidirectional and experimental error leads us to infer that it is *not*, we will have made an inference error. We can compute the probability of making either of these two types of error as part of a thorough bidirectionality assessment. However, it is more common to compute the minimum level of empirically estimated bidirectionality beyond which it is “unlikely” (probability below some specified threshold) that the bidirectionality truly is zero, given the amount of experimental error in the data.

Let us assume that we have used Eqs. (20) and (29) to compute the normalized bidirectionality metric, τ , and its variance. We appeal to the Central Limit Theorem in order to claim that the error distribution for τ is Gaussian, with a mean of zero and a variance that can be computed by Eq. (29). For the sake of illustration, let us say that Eq. (29) has been used to estimate a variance of 1E-06. The standard error (“one sigma”) is the square root of this, 0.001 or 0.1%. If the error distribution can indeed be assumed to be Gaussian, then neglecting bias errors in the data we would say with 95% confidence that the true normalized bidirectionality metric lies roughly within $\pm 0.2\%$ (± 2 sigma) of our computed estimate.

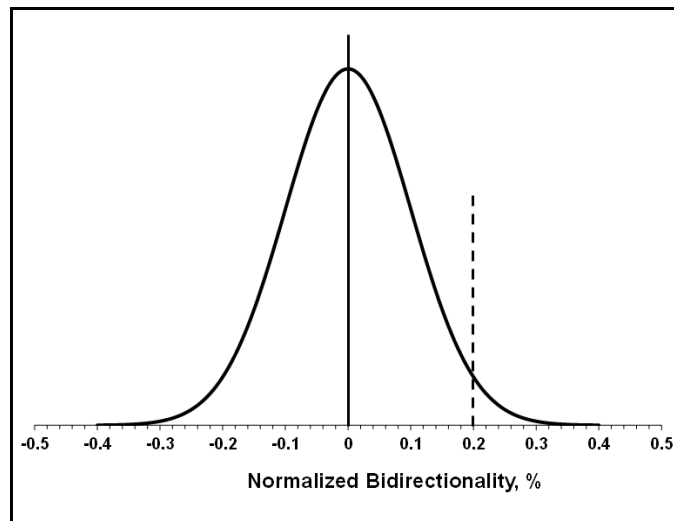


Figure 2. H_0 Reference Distribution for Normalized Bidirectionality, Mean = 0.0%, Standard Error = 0.1%.

We establish a null hypothesis, H_0 , that the true normalized bidirectionality metric is 0%. Figure 2 is a normal probability density function for the case in which H_0 is true. Its mean is 0% and its standard deviation is 0.1%. The PDF in Fig. 2 is a reference distribution, which describes the likelihood of observing a given value of τ when the

null hypothesis is in fact true (that is, when τ is actually 0%). Experimental error ensures that any given empirical estimate of τ is likely to be non-zero even if H_0 is true, but smaller departures from zero are more likely than larger ones in that case, and an empirical estimate of τ is more likely to lie near the mean of the reference distribution than far from it.

The dashed line in Fig. 2 is at 0.2%, or two standard deviations away from zero. The actual value of τ can be either positive or negative but by convention we represent the bidirectionality metric as an absolute value term per Eqs. (17) and (21). If it lies to the right of the dashed line, we are entitled to reject the null hypothesis and to claim that there is no more than a 5% probability of an inference error. That is, we are entitled to claim that there is at least a 95% probability that even given the limitations on precision induced by experimental error; bidirectionality has been detected in this case. If τ is to the *left* of the dashed line in Fig. 2, we cannot reject the null hypothesis with less than a 5% probability of an inference error; therefore we cannot conclude with 95% confidence that the balance is bidirectional.

If, as in this example, we establish 95% as the minimum level of confidence with which we are willing to make an inference; that is to say, if we are unwilling to accept more than a 5% probability of an inference error, then we must report any estimate of bidirectionality within $\pm 0.2\%$ as indistinguishable from zero. In that case we cannot claim that the bidirectionality *actually is* zero. We claim only that given the quality of the data from which we are making an inference, we are unable to say with the requisite level of confidence whether the bidirectionality is non-zero or not. With this precision it would be meaningless to report a level of bidirectionality of 0.15%, say, claiming that the balance is indeed bidirectional, but only to a relatively small degree. We cannot resolve any difference between 0.15% and zero in this example, and so we should simply report “no detectable bidirectionality.”

The reference distribution in Fig. 2 is associated with a null hypothesis, but there is also a reference distribution that is associated with the alternative to this null hypothesis. In this case, the alternative hypothesis is that bidirectionality has been detected at a sufficient level to be of concern.

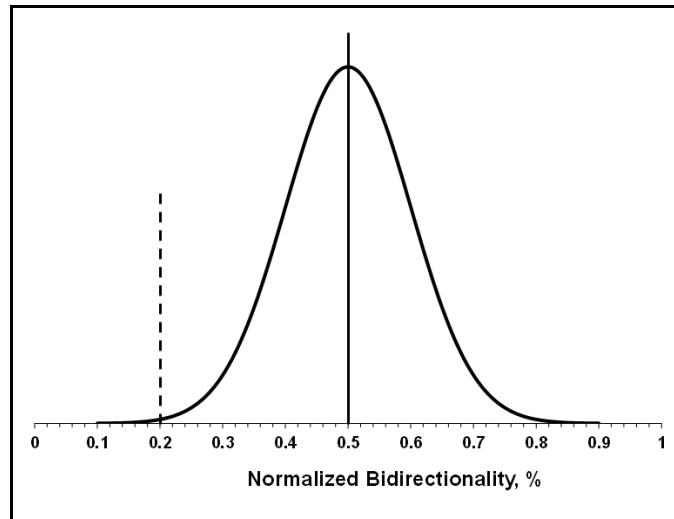


Figure 3. H_A Reference Distribution for Normalized Bidirectionality, Mean = 0.5%, Standard Error = 0.1%.

We must also decide whether or not to reject the *alternative* to the null hypothesis. Again, we use a reference distribution for this purpose. The null hypothesis was that there is no significant bidirectionality and its alternative is that there is in fact significant bidirectionality. Figure 3 is the reference distribution for the alternative hypothesis, where for this example we are using “0.5%” as the tolerance level for bidirectionality. We assert that bidirectionality of 0.5% or greater is unacceptable.

The dashed line in Fig. 3 is a criterion that we use in this example to accept or reject the alternative hypothesis. Any value of τ computed by Eq. (21) that lies to the left of this dashed line is sufficiently small to reject the alternative hypothesis with an inference error probability no greater than the area under the reference distribution to the left of the dashed line. The alternative hypothesis is always single-sided and the dashed line in this example is three standard deviations to the left of the mean. Therefore the probability of erroneously rejecting the alternative

hypothesis by declaring a truly bidirectional balance to be free of significant bidirectionality is no greater than 0.0014 in this case if we require that τ be less than 0.2% as a condition for rejecting the alternative hypothesis.

Figures 2 and 3 are combined in Fig. 4. We compute τ using Eq. (21). If it is less than the criterion marked with the dashed line, 0.2%; that is, if it is more likely to have been drawn from the distribution on the left represented in green than the red one on the right, we reject the alternative hypothesis and declare the balance free of significant bidirectionality. If $\tau \geq 0.2\%$ and so more likely to have been drawn from the red distribution on the right than the green one on the left, we reject the null hypothesis and declare that the balance is bidirectional to a sufficient degree that we are unable to declare it free of significant bidirectionality.

The bidirectionality criterion level of 0.2% was selected so that the area under the H_0 reference distribution to the right of this, and the area under the H_A reference distribution to the left of this, were low enough to satisfy our inference error risk requirements. In this example, we are willing to accept a probability of no greater than 0.05 (one chance in 20) of erroneously rejecting the null hypothesis by declaring that a non-bidirectional balance is in fact bidirectional, and we are willing to accept a probability of no greater than 0.0014 (one chance in 714) of erroneously rejecting the alternative hypothesis by declaring that a bidirectional balance is not in fact bidirectional.

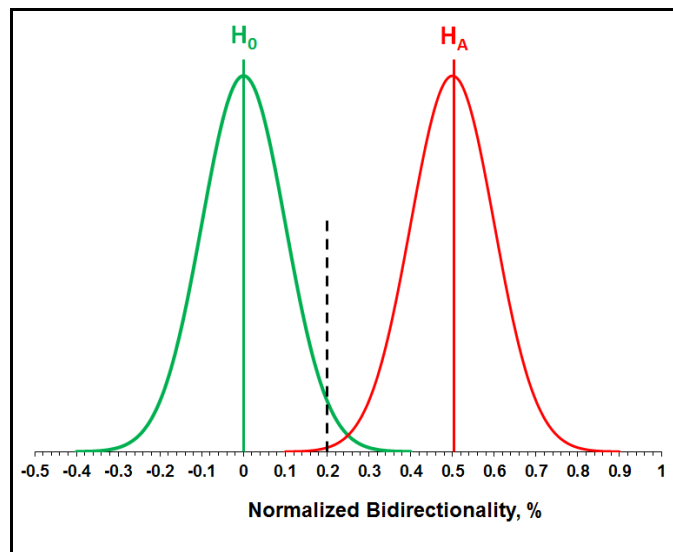


Figure 4. H_0 and H_A Reference Distributions for Normalized Bidirectionality, Means = 0.0% and 0.5%, Standard Error = 0.1%.

In this example we required greater inference error insurance against erroneously rejecting the alternative hypothesis than erroneously rejecting the null hypothesis because the consequences of committing this error are more severe than erroneously rejecting the null hypothesis. If we erroneously reject the null hypothesis, the result is that we prescribe more care than is necessary in analyzing the balance, because we claim it is bidirectional when it is not. This is not a desirable outcome, but it would be much worse to declare a significantly bidirectional balance to be free of bidirectionality. In that case, we could fail to account for the bidirectionality of the balance and produce an erroneous calibration.

The example cited in this section was artificial, and intended simply to illustrate the process of making an inference about bidirectionality under uncertainty. In the next section, we show how this process is applied to actual balance calibration data.

VI. Assessing Bidirectionality in a Force Balance: An Example Using Existing Calibration Data

This paper proposes that the bidirectionality of a force balance can be assessed through a series of steps that will be illustrated in this section, using existing data from a multi-piece Task balance that we have designated as “balance1.” The steps are grouped into four phases. The first phase consists of specifying tolerance levels that will determine when we are prepared to claim that a balance is bidirectional. The second phase consists of specifying an initial balance calibration response model. Starting with that model we perform a standard term-reduction regression analysis⁸ on a set of calibration data, eliminating terms from the model with regression coefficients that are too small

to resolve with a specified level of confidence given the unexplained variance (experimental error) of the data. The third phase is to use the regression coefficients quantified in Phase Two to compute a bidirectionality metric and its variance, and the fourth phase is to make some inference about the bidirectionality of the balance based on the bidirectionality metric estimated in Phase Three (plus its variance) and the tolerance levels established in Phase One.

A. Phase One

The bidirectionality assessment will turn on three tolerance levels that are established before any calibration data are analyzed. The first is a tolerance level for bidirectionality; the greatest degree of bidirectionality we are willing to accept without taking it into account through a more complex calibration model, for example. The remaining two tolerances are in the form of inference error probabilities. One of these probabilities represents the greatest acceptable risk of a false alarm (claiming a balance is significantly bidirectional when it is not), and the other represents the greatest acceptable risk of failing to detect that a balance is significantly bidirectional when in fact it is.

It is important to quantify inference error risk tolerance because experimental error ensures that we can never be absolutely certain that our estimate of bidirectionality exceeds our tolerance for it. There will always be some non-zero probability that we get this wrong by either validating a balance that actually is bidirectional, or by falsely indicting a balance that in truth does not exceed our tolerance for bidirectionality. So it is not sufficient to simply declare the balance to be “bidirectional” or “not bidirectional.” For either case we need to be able to say how confident we are in our conclusion.

No matter the inference, if our confidence level does not exceed some prescribed minimum, we are not justified in making it. For example, if we can say with 80% confidence that a balance displays bidirectionality that exceeds our tolerance for it, but our standards demand 95% confidence, then all we can say is that we are unable to determine with sufficient confidence that the balance is bidirectional. Note that this is not the same thing as declaring that the balance is NOT bidirectional; it simply acknowledges limits on the quality and/or volume of calibration data that preclude making an acceptable inference.

We begin in this example by arbitrarily specifying 0.5% as our tolerance for normalized bidirectionality. If the principal gage output based on a calibration that accounts for bidirectionality differs from the principal gage output based on a calibration that does not, that difference is considered unacceptable if it exceeds a half percent of the principal gage output when bidirectionality is not considered. This is consistent with the tolerance level initially proposed by Ulbrich¹⁰. We will also specify in this example probabilities of 0.05 and 0.01 as, respectively, the maximum acceptable risk of a false alarm (erroneously rejecting the null hypothesis) and the maximum acceptable risk of a missed detection of significant bidirectionality (erroneously rejecting the alternative hypothesis). These specific inference error probabilities are arbitrary and can be individually specified to fit the analyst’s tolerance for inference error risk, reflecting the fact that the consequences of each of the two inference errors might be different.

In this example, the different inference error risk tolerance specifications reflect a sense that failing to detect significantly bidirectionality would have greater consequences than erroneously inferring that a balance is bidirectional. The former error could result in an invalid calibration, while the latter is only likely to result in a more elaborate calibration than is necessary—one possibly featuring a number of unnecessary absolute-value terms, for example. The inference error risk tolerances specified in this example reflect a willingness to accept one chance in 20 of a false alarm, but only one chance in 100 of failing to detect a truly bidirectional balance. That is, in this illustration we will require 95% confidence in an inference that the balance is bidirectional, but we will require 99% confidence to infer that a balance is free of significant bidirectionality.

B. Phase Two

We adopt Eq. (3) as the initial calibration response model to use in this phase of the analysis. A commercial software package¹⁷ was used to determine the statistically significant b_i and c_i coefficients from this model, applying the backward elimination method of term reduction to Eq. (3).

The reader is referred to standard references for a detailed description of backward elimination, but briefly, we exploit the fact that every calibration data set features variance, most of it intentionally induced by applying a variety of different calibration loads. We refer to such variance as “explained” (by the response model). Any part of the variance that cannot be attributed to known load changes using the calibration response model is “unexplained,” and therefore contributes to uncertainty in calibration response predictions. We seek a response model that explains as much of the total variance as possible.

We begin by proposing the full response model of Eq. (3), and as part of the regression analysis we make a statistical assessment of how the unexplained variance changes as each term in the model is eliminated, one at a

time. Equivalently, we assess how the *explained* variance changes, since the explained and unexplained variance sum to the total variance in the calibration data, which is a constant for a given data set.

Each term is declared statistically significant and thus retained in the model if rejecting it would increase the unexplained variance by an amount that can be detected with a significance of 0.001. In such a case, even given the ambiguity induced by ordinary experimental error, we can be at least 99.9% sure that the model coefficient is non-zero. The analyst can choose other values besides 0.001 for the significance level in backward elimination; larger values are less restrictive, while smaller values are more so. It is largely a matter of the analyst's experience in similar applications.

The effect of this analysis is to eliminate candidate model terms unless we are quite sure that they belong in the model, leaving a relatively lean model in which we are confident that "every term counts." In practice, most of the terms eliminated in this process are unambiguously negligible, although there can be "borderline terms" that might be either rejected or retained. For these borderline terms, the experience, judgment, and subject matter expertise of the analyst all come into play in making the final rejection/retention decision. The backward elimination cycle is repeated until no further modifications to the model significantly increase the explained variance (or equivalently, decrease the unexplained variance).

Eq. (3) is a second-order polynomial in k factors. For this analysis, $k = 7$ since the response of each primary gage output is modeled as a function of six loads and one categorical variable. It would have been possible to model each response as a function of 12 factors, the six loads plus categorical variables associated with each load, but this could have resulted in models for one response featuring categorical variables associated with other responses, or even interactions among categorical variables. The first case would imply that the degree of bidirectionality in one response is a function of whether there is bidirectionality in another gage. The second would imply that the degree of bidirectionality in one response is a function of the degree of bidirectionality in another gage.

While such second-order effects involving bidirectionality interactions are not inconceivable, they are believed to be too small to warrant the substantial additional complexity required to quantify them in a general balance calibration response model. We can avoid much of this added complexity by adopting a convention that the bidirectionality associated with one primary gage load will be defined only for zero loads on the other gages. It is then not necessary to model categorical variables associated with the other gages, or the interactions of those other categorical variables with each other or with the load variables. This enormously simplifies the response model.

There is one other convention adopted in the regression analysis described in this paper: Hierarchy is imposed on all response models. A comprehensive description of hierarchy and the role it plays in producing what Kempthorne¹⁷ describes as "well formulated" models is beyond the scope of this paper, but the interested reader can consult the literature^{13, 15, 16} for additional information. The discussion around Eq. (14) provides a brief description.

The calibration data sample acquired for "balance1" featured six primary gage outputs in microV/V, labeled R_1 through R_6 . The calibration consisted of 1,751 combinations of six corresponding primary gage loads labeled N_1 , N_2 , S_1 , S_2 , RM , and AF . The RM loads were in inch-pounds and all other loads were in pounds. Table 1 lists the calibration data ranges for the six primary gage loads of balance1, and also displays the physical variable ranges into which linear transformations mapped the loads into the coded variable range of ± 1 , using Eq. (13). The constants " L " and " H " in Eq. (13) appear in the last two columns of Table 1.

Table 1. Load ranges for balance1 calibration data

Coded Variables	Physical Variables		Calibration Range		Coded Variable Range	
	Load	Units	Min	Max	L ($x=-1$)	H ($x=+1$)
x_1	N_1	lbs	-2073	2126	-2100	2100
x_2	N_2	lbs	-2033	2128	-2100	2100
x_3	S_1	lbs	-686	689	-700	700
x_4	S_2	lbs	-709	716	-700	700
x_5	RM	in-lbs	-3766	3998	-4000	4000
x_6	AF	lbs	-349	352	-350	350

Note that for coding the independent variables, the upper and lower range limits were specified as round numbers near, but not exactly equal to, the absolute values of the maximum calibration loads. For example, the N_1

load range of -2073 lbs to +2126 lbs was linearly mapped into a coded variable range for which ± 1 in coded variables corresponded to ± 2100 lbs. This simply means that the loads with the largest absolute value will not always be precisely 1.000 in coded units but might be slightly greater or smaller than 1. This has no impact on the analysis, although the customary cautions against extensive extrapolation apply.

Symmetry about zero in the coding range has a further advantage for the bidirectionality analysis performed here, in that it ensures that when the independent variable is zero in coded units, it is also zero in physical engineering units. This facilitates the simplifications that occur when we adopt the convention of defining primary gage bidirectionality only in terms of zero secondary loads. By default, the commercial software package¹⁴ used in this analysis performs all of its regression analyses in coded units, requiring only that the user specify the coded unit ranges listed in Table 1.

The calibration data set for “balance1” was delivered in the form of an Excel spreadsheet with twelve columns and 1,751 rows. The first six columns were the loads described in Table 1, and the next six columns were the corresponding gage outputs, in microV/V. The load columns were treated as independent variables in the ensuing regression analysis to which one more column was added to represent an associated categorical variable. In the spreadsheet, each element of this categorical variable column was set either to -1 or +1, depending on whether the primary gage load was negative or greater than or equal to zero. These data were then copied from the spreadsheet and pasted into the software.

The software permits the user to specify a starting response model for the backward elimination regression procedure described above. For each gage, the starting model was selected to be a full third-order polynomial in the seven independent variables (six numerical load factors plus the categorical polarity designation variable for the gage under evaluation). A third-order model was specified simply to accommodate interaction terms involving the categorical variable and conventional second-order load interaction terms. Third-order terms involving only numerical load variables were rejected from the starting model. This ensured that after backward elimination the reduced model would feature the categorical variable designating the sign of the primary gage load for each data point, a number of conventional load terms no higher than second-order, and interactions of such load terms with the categorical variable.

The calibration data sample provided for “balance1” was analyzed using the backward elimination term removal method described above. We adopt the notional convention of using x_i [$i = 1..6$] to represent the load variables in coded units, as in Table 1. With this convention, the reduced model for R_i , the output corresponding to the N_i primary gage load, is as follows for the “balance1” data summarized in Table 1:

$$\begin{aligned} R_i = & (b_0 + c_0 z) + (b_1 + c_1 z) x_1 + (b_2 + c_2 z) x_2 + b_3 x_3 + b_5 x_5 \\ & + (b_{12} + c_{12} z) x_1 x_2 + b_{13} x_1 x_3 + b_{15} x_1 x_5 + b_{35} x_3 x_5 \\ & + (b_{11} + c_{11} z) x_1^2 + b_{22} x_2^2 + b_{33} x_3^2 + b_{55} x_5^2 \end{aligned} \quad (33)$$

Here we have reverted to the general subscripted notation used in Eqs. (1) through (4). Numerical values for the coefficients, as well as the standard errors in estimating them, were generated by the regression software.

Figure 5 displays the regression coefficients of Eq. (33) graphically, as multiples of the standard error (“one sigma”) in estimating them. The horizontal red line marks the retention/rejection threshold criterion established for this study; for a candidate term to be retained in the calibration response model, we require that its coefficient exceed the standard error in estimating it by a factor of just over 3, corresponding to a significance of 0.001 or a confidence level of 99.9%. Of the 56 terms in Eq. (3), the original starting candidate response model, only 18 survived the backward elimination process to appear in Eq. (33).

As Fig. 5 reveals, the magnitudes of most of these coefficients exceed the detection threshold by a comfortable margin (note the logarithmic scale). Two exceptions are the intercept (“Int” in Fig. 5), which is just marginally below the threshold, and the quadratic N_i load term, which is well below the threshold. We force the intercept to be retained regardless of its statistical significance. We retain the quadratic N_i term to maintain hierarchy, as required by the statistical significance of the $N_i \times N_j \times z$ term representing an interaction between the quadratic N_i term and the categorical variable z . (Since $N_i \times N_j \times z$ is significant, we must retain all of its components to maintain hierarchy: N_i , z , $N_j \times z$, and $N_i \times N_j$, regardless of their statistical significance. In this case, each of the components is significant except the quadratic N_i term).

Several interesting observations and related tentative conclusions are available from Fig. 5. Note, for example, that the intercept term is marginally insignificant while the coefficient of the categorical variable term, z , is about 10 standard deviations away from zero and therefore unambiguously significant. This implies that the intercept of a

regression model fitting loads that span the full positive and negative range of the calibration data set will be just barely indistinguishable from zero at the 0.001 level of significance. However, the fact that the z coefficient is significant implies a discontinuity at the intercept when positive and negative loads are fitted separately. This is a tell-tale sign of bidirectionality.

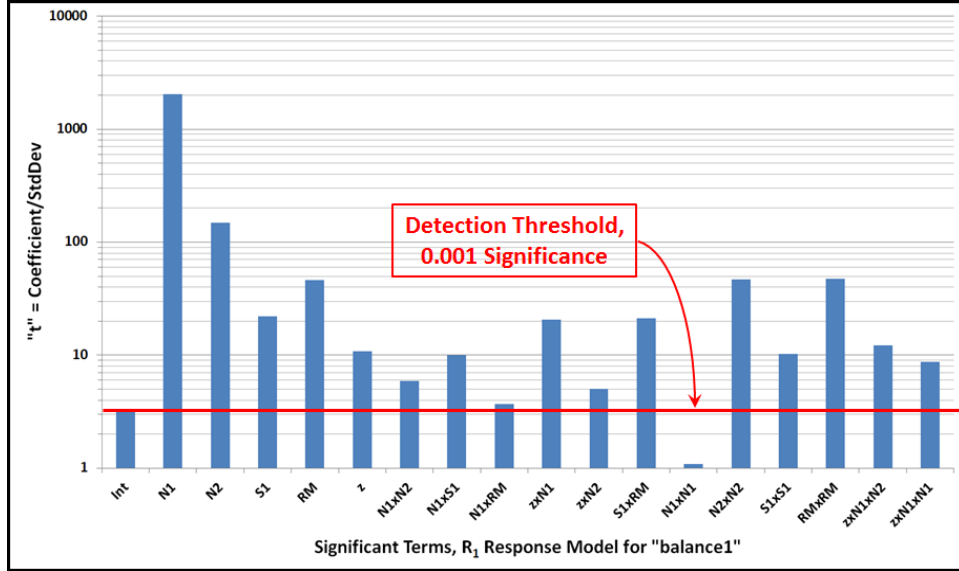


Figure 5. Significant terms in the R_1 response model for “balance1,” multiples of standard error

The first-order N_1 term dominates the regression model. At over 2,000 standard deviations away from zero, its coefficient is an order of magnitude larger than the coefficient of any other term in the model. This is expected because R_1 is the primary output corresponding to loadings of the N_1 gage, but note that the $N_1 \times z$ term is also unambiguously significant, at about 20 standard deviations away from zero. This implies a change in slope when response models are fitted to positive and negative load data separately. This change in slope with load polarity is another sign of bidirectionality.

The fact that the coefficient of the quadratic N_1 term is so far below the detection threshold in Fig. 5 indicates that no curvature would be detected in a fit of data symmetrically spanning the full positive and negative load range. However, just as the intercept and slope of R_1 as a function of N_1 are different for positive and negative loads, so is the curvature, or second-order effect. This is clear from the fact that the coefficient of the $N_1 \times N_1 \times z$ term is over eight standard deviations away from zero and thus statistically significant. This change in curvature across zero is a further indication that the balance is bidirectional.

Recall that we define bidirectionality for the primary load gage only for the case in which secondary gage loads are zero. For the R_1 response model in Eq. (33), this implies that $x_2 = x_3 = x_4 = x_5 = 0$ (the x_6 or Axial Force term was not statistically significant), so that Eq. (33) becomes:

$$R_1 = (b_0 + c_0 z) + (b_1 + c_1 z)x_1 + (b_{11} + c_{11} z)x_1^2 \quad (34)$$

This is just Eq. (4) for the special case in which all secondary loads are zero and we designate the primary gage load as x_1 . We adopt notational simplifications described after Eq. (4) that are facilitated by the fact that secondary loads are all zero so that we no longer need to index the applied loads. We also explicitly display the ± 1 values that the categorical variable, z , assumes according to whether the primary gage load is positive or negative. Eq. (34) then is of the form of Eq.(6), reproduced here for convenience:

$$y = (b_0 \pm c_0) + (b_1 \pm c_1)x + (b_2 \pm c_2)x^2 \quad (6)$$

where when y is the primary gage response to a *positive* primary gage load, x , the c_i categorical variable coefficients are all preceded with plus signs, and when y is the primary gage response to a *negative* primary gage load, x , the c_i

categorical variable coefficients are all preceded with minus signs. As before, b_0 , b_1 , and b_2 are simplified notation for the regression coefficients of the intercept, first- and second-order x coefficients, and c_0 , c_1 , and c_2 are regression coefficients for the categorical variable, z , and its interactions with the first- and second-order x terms in the calibration response model, respectively.

The coefficients b_0 , b_1 , b_2 , c_0 , c_1 , and c_2 in Eq. (6) comprise what we refer to as the “Cardinal Six” coefficients for assessing bidirectionality. The numerical values for these coefficients and their standard errors (“one sigma”) are displayed in Table 2. Note that these coefficients correspond to a response model for which the independent variables are expressed in coded variables, not in physical units such as pounds and inch-pounds. See Section II.

Table 2. “Cardinal Six” terms in the R_1 response model for “balance1.” Ratio of Coefficient to Standard Error is “t”

Model Components		Empirical Estimates		t
Term	Coefficient	Coefficient	Std Err	
Int	b_0	-0.235	0.072	3.265
N1	b_1	792.873	0.392	2025
N1xN1	b_2	0.483	0.447	1.082
z	c_0	-0.735	0.068	10.75
zxN1	c_1	8.103	0.392	20.68
zxN1xN1	c_2	3.812	0.437	8.727

With the construction of Table 2, we have completed the first two of four proposed phases in formally evaluating the bidirectionality of a force balance. We will now illustrate the third phase of bidirectionality assessment.

C. Phase Three

The third of four proposed phases in assessing the bidirectionality of a force balance is to use the regression coefficients quantified in Phase Two to compute bidirectionality metrics and their variances. We will insert numerical values for these coefficients from Table 2 into Eqs. (16), (20), (26), and (29) to quantify normalized and non-normalized bidirectionality metrics and the uncertainty in estimating them.

We begin with the formula for the non-normalized bidirectionality metric derived in Eq. (16). For the R_1 response of balance1 we have:

$$\begin{aligned}\Lambda_+ &= c_1 + (c_0 + c_2) = 8.103 + [(-0.735) + (+3.812)] = +11.180 \\ \Lambda_- &= c_1 - (c_0 + c_2) = 8.103 - [(-0.735) + (+3.812)] = +5.026\end{aligned}\tag{35}$$

Following Ulbrich¹⁰, we characterize bidirectionality by the metric in Eq. (35) with the largest absolute value:

$$\Lambda = \text{MAX} [\text{ABS}(\Lambda_+), \text{ABS}(\Lambda_-)] = \text{MAX}(11.180, 5.026) = 11.180\tag{36}$$

The standard error (“one sigma” value) for this estimate of Λ is computed from Eq. (26), which for the R_1 response of balance1 is:

$$\sigma_\Lambda = \sqrt{\sigma_{c_0}^2 + \sigma_{c_1}^2 + \sigma_{c_2}^2} = \sqrt{0.068^2 + 0.392^2 + 0.437^2} = 0.591\tag{37}$$

We estimate the normalized bidirectionality metric in a similar way starting with Eq. (20), which for the R_I response of balance1 we have:

$$\begin{aligned}\tau_+ &= \frac{c_1 + (c_0 + c_2)}{(b_0 + b_2) + b_1} = \frac{(8.103) + [(-0.735) + (3.812)]}{[(-0.235) + (0.483)] + (792.837)} = +0.01409 = +1.41\% \\ \tau_- &= \frac{c_1 - (c_0 + c_2)}{(b_0 + b_2) - b_1} = \frac{(8.103) - [(-0.735) + (3.812)]}{[(-0.235) + (0.483)] - (792.837)} = -0.00634 = -0.63\%\end{aligned}\quad (38)$$

Again we follow Ulbrich¹⁰ and select that variation of the metric with the largest absolute value:

$$\tau = \text{MAX} [\text{ABS}(\tau_+), \text{ABS}(\tau_-)] = \text{MAX} [1.41\%, 0.63\%] = 1.41\% \quad (39)$$

Equation (31), reproduced here for convenience, represents the variance in estimating τ for positive and negative load:

$$\sigma_{\tau_{\pm}}^2 = \frac{y_{\pm}^2 \sigma_{\Lambda}^2 + \Lambda_{\pm}^2 \sigma_y^2}{y_{\pm}^4} \quad (31)$$

The uncertainty in estimating the normalized bidirectionality metric is a function of the non-normalized bidirectionality metric and its standard error, and the output near maximum calibration load and its standard error. The non-normalized bidirectionality metric and its standard error have been quantified for the current case in Eqs. (36) and (37). We can quantify y_{\pm} and the associated standard error (same for both polarities) in a similar way by inserting values from Table 2 into Eqs. (22) and (30):

$$\begin{aligned}y_+ &= (b_0 + b_2) + b_1 = [(-0.235 + 0.483)] + 792.837 = +793.121 \\ y_- &= (b_0 + b_2) - b_1 = [(-0.235 + 0.483)] - 792.837 = -792.624\end{aligned}\quad (40)$$

$$\sigma_{y_{\pm}}^2 = \sigma_{b_0}^2 + \sigma_{b_1}^2 + \sigma_{b_2}^2 = \sigma_y^2 = 0.072^2 + 0.392^2 + 0.447^2 = 0.3587 \rightarrow \sigma_y = 0.599 \quad (41)$$

Inserting values from Eqs. (36), (38), (40), and (41) into Eq. (31), we obtain for the R_I response of balance1:

$$\begin{aligned}\sigma_{\tau_+}^2 &= 5.5559 \times 10^{-7} \rightarrow \sigma_{\tau_+} = 0.075\% \\ \sigma_{\tau_-}^2 &= 5.5550 \times 10^{-7} \rightarrow \sigma_{\tau_-} = 0.075\%\end{aligned}\quad (42)$$

For this case, the standard error in estimating the normalized bidirectionality metric, τ , is essentially the same for positive and negative loads. We demonstrate in the Appendix that this is a general result for calibration load schedules that generate outputs under positive and negative loading that are nominally the same magnitude. That is, the uncertainty in estimating τ is essentially independent of load polarity for symmetric balance outputs near maximum calibration load.

We summarize the key results for Phase 3 of the bidirectionality assessment analysis of balance1's R_I by gathering the results of various calculations from above into Table 3.

The quantities y , Λ , and τ were each evaluated for positive and negative loading, and following Ulbrich¹⁰, the largest absolute values are retained in Table 3. We now have the information necessary to proceed to the fourth and final phase of the bidirectionality analysis.

Table 3. Bidirectionality Metrics and Standard Errors, “balance1” Output R_I

Quantity		Standard Error	
Symbol	Value	Symbol	Value
y	793.12	σ_y	0.60
Δ	11.18	σ_Δ	0.59
τ	1.41%	σ_τ	0.08%

D. Phase Four

In this final phase, we infer whether we can conclude with acceptable confidence that bidirectionality has been detected, given the level of experimental error in the calibration data. If we infer that bidirectionality is severe enough to *detect* unambiguously, we must then infer whether we can conclude with acceptable confidence that it is great enough to be *of concern*. We apply the formal hypothesis testing methods first introduced in Section V above to make these inferences, as we will now illustrate by continuing the balance1 bidirectionality assessment.

From Table 3 we see that the non-normalized bidirectionality metric, Δ , has a value of 11.18 microV/V for the R_I output of balance1. For this case, this means that at a load of +2100 pounds (see Table 1), the R_I output forecasted from a regression model based only on positive loads is expected to differ from a regression model based on the full range of positive and negative calibration loads by 11.18 microV/V. We test this result against a null hypothesis that there is no real difference in the outputs forecasted by the two models, with any perceived difference due only to experimental error.

In Phase 1 we established an acceptable inference error probability of 0.05 for erroneously rejecting the null hypothesis. That is, we stipulated in Phase 1 that we were willing to accept one chance in 20, and no more, of a false alarm by declaring a balance to be bidirectional when it is not.

If the null hypothesis is correct and the true bidirectionality metric is in fact zero, then an estimate of the metric that is approximately two standard deviations away from zero would be large enough to reject the null hypothesis with 95% confidence; that is, with an inference error probability of no more than 0.05. To be more precise, since the variance in the non-normalized bidirectionality metric estimate recorded in Table 3 is based on 1753 residual degrees of freedom, a value that is greater than 1.961 standard deviations is sufficiently remote from zero to reject the null hypothesis with an inference error probability of no more than 0.05. (See the Discussion for distinctions between one-sided and two-sided null hypotheses, and remarks about how the analysis is impacted by the fact that bidirectionality is represented as an absolute value).

Assuming a standard error in Δ of 0.59 microV/V as in Table 3, our criterion for rejecting the null hypothesis is then $1.961 \times 0.59 = 1.16$ microV/V. An empirical estimate of Δ must be greater than this before we can say with 95% confidence that we have detected bidirectionality. If the empirical estimate of Δ is less than or equal to this we do not say there is no bidirectionality (a negative assertion can never be proven); we say instead that the bidirectionality of the balance is too small to detect with the requisite level of confidence, given the quality and the volume of the data in hand.

In the current example, the empirical estimate of the bidirectionality metric is, from Table 3, 11.18 microV/V. Since this is almost 19 standard deviations away from zero, we infer that Δ is too large to attribute to random error. We therefore conclude that this balance is in fact bidirectional. Since 19 standard deviations is so much greater than the minimum 1.961 needed to reject the null hypothesis with no more than an 0.05 probability of an improper inference, the inference error probability is substantially less than 0.05 in this case, so it can be concluded with very little inference error risk that the gage output does display bidirectional behavior.

Figure 6 is a probability distribution of empirical Δ estimates for output R_I of balance1 under the null hypothesis, given that the standard error estimating Δ is 0.59 microV/V as in Table 3. It serves as a graphical reference for assessing statistical significance. Even if the true value of Δ is zero (that is, even if the null hypothesis is exactly true), experimental error could result in an empirical estimate of Δ that differs slightly from zero.

However, in that case we would still expect to find the Λ estimate somewhere near the zero mean of the reference distribution, if not precisely at zero.

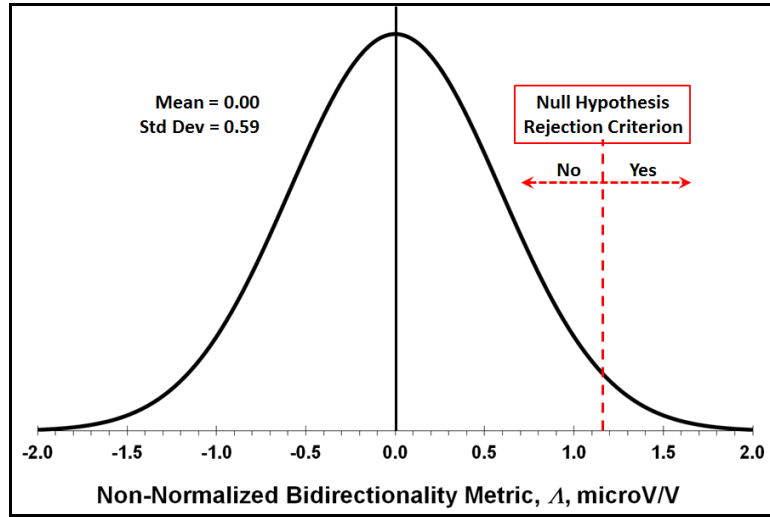


Figure 6. **Non-Normalized Bidirectionality Reference Distribution. balance1, Output R_1 : Risk ≤ 0.05**

The dashed line marks the criterion for rejecting the null hypothesis, so any empirical estimate of Λ to the right of this line is far enough away from zero to reject the null hypothesis and to conclude with at least 95% confidence that the balance is bidirectional. For output R_1 of balance1, the Λ estimate of 11.18 MicroV/V is so far to the right of the null hypothesis criterion that it is literally “off the chart,” strongly suggesting that the true value of Λ is non-zero.

We have now demonstrated that the true value of Λ for the R_1 output of balance1 is very likely to be non-zero, but we have not as yet demonstrated that it is large enough to be of concern. To do this, we begin by establishing an alternative to the null hypothesis stating that the magnitude of the bidirectionality is in fact large enough to be of concern, and then we test that hypothesis just as we tested the null hypothesis. That is, we establish a reference distribution for the alternative hypothesis and a quantitative criterion for accepting or rejecting it. We then compare the empirical estimate of Λ with this criterion, and accept or reject the alternative depending on whether Λ is greater than this criterion or not.

In Phase 1 we declared that for a given load, if the principal gage output based on a calibration that accounts for bidirectionality differs from the principal gage output based on a calibration that does not, that difference would be considered unacceptable if it exceeded a half percent of the principal gage output when bidirectionality is not considered. This is the tolerance level first proposed by Ulbrich¹⁰.

For the case of the R_1 principal gage output of balance1 that we are currently considering, Table 3 indicates that the output near maximum calibration load is 793.12 MicroV/V. This corresponds to the 2100 pound load for which the coded principal gage load has a magnitude of “1” per Table 1, which is the load for which we are evaluating bidirectionality in the current case. Our bidirectionality tolerance is a half percent of this, or $0.005 \times 793.12 = 3.966$ MicroV/V. Figure 7 is a reference distribution corresponding to the alternative hypothesis that Λ is just out of tolerance at 3.966 MicroV/V. The uncertainty in estimating the bidirectionality metric is the same whether the true bidirectionality is zero or just large enough to be of concern, so as with the reference distribution for the null hypothesis, Fig. 6, the standard error of this distribution is 0.59 MicroV/V. Again, even if the balance is just marginally bidirectional so that the true value of Λ is 3.966 MicroV/V, experimental error could result in an empirical estimate of Λ that is larger or smaller.

We specified in Phase 1 of this analysis that we would not accept a probability any greater than 0.01 of failing to detect significant bidirectionality. That is, we established this as the greatest risk we would accept of erroneously rejecting the alternative hypothesis. The dashed line in Fig. 7 is placed at the point for which the area under the reference distribution to the left of it is 0.01, which can be shown to be 2.329 standard deviations from the mean. With a standard deviation of 0.59 microV/V, this places the criterion $2.329 \times 0.59 = 1.374$ microV/V to the left of the mean, and since the mean is at 3.966 microV/V, this is at $3.966 - 1.374 = 2.592$ microV/V, as indicated in Fig. 7.

If the balance is in fact bidirectional, the probability that random experimental error would result in an empirical estimate of the bidirectionality metric that is smaller than 2.592 microV/V is no greater than 0.01. There is thus a minimum probability of 0.99 that even a marginally bidirectional balance gage output will yield an empirical

bidirectionality metric to the right of the dashed threshold line. That is, even if the gage is only marginally bidirectional, we would expect an empirical estimate of λ that exceeds 2.592 microV/V. This area under the reference distribution to the right of this criterion level for the alternative hypothesis is called the “power” or “resolving power” of the test for bidirectionality.

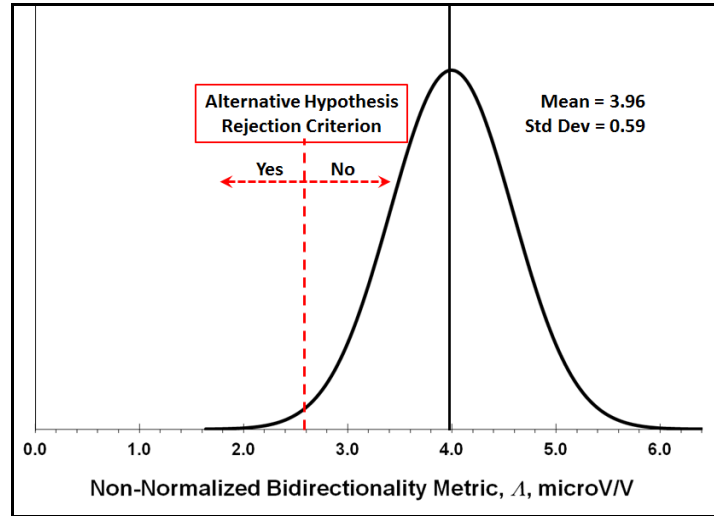


Figure 7. Non-Normalized Bidirectionality Reference Distribution. balance1, Output R_I : Risk ≤ 0.01

For output R_I of balance1, it has already been shown that the λ estimate of 11.18 microV/V is large enough to reject the null hypothesis and to conclude with at least 95% confidence that the bidirectionality of this balance is non-zero. Since λ also exceeds the alternative hypothesis criterion by a substantial margin, we are unable to reject the alternative hypothesis, and we conclude in this case with at least 99% confidence that the bidirectionality is great enough to be of concern.

Table 3 also lists the *normalized* bidirectionality metric, τ , for output R_I of balance1 and its standard error. These enable us to construct reference distributions for the null and alternative hypotheses that correspond to those for the non-normalized metric displayed in Figs. 6 and 7. Figure 8 displays both reference distributions, as well as the empirical estimate of the *normalized* bidirectionality metric for output R_I of balance1.

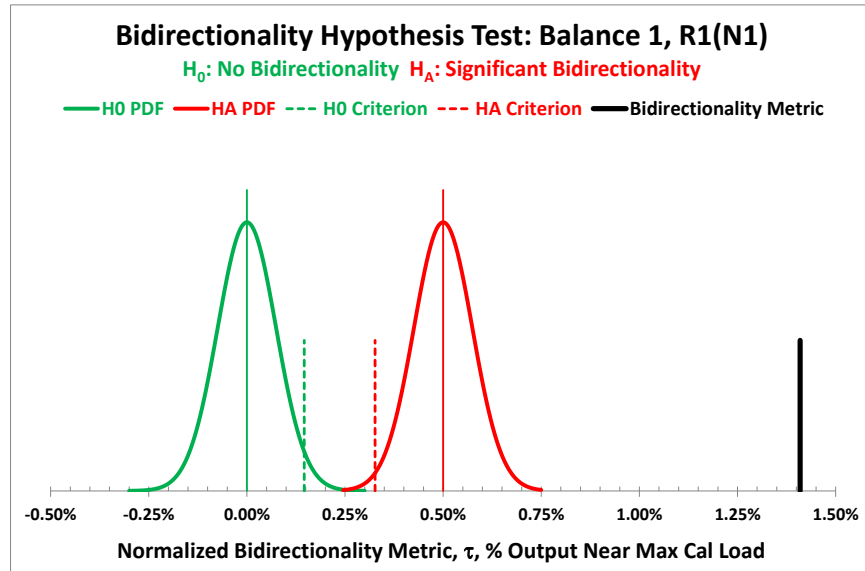


Figure 8. Normalized Bidirectionality Reference Distributions for Null and Alternative Hypotheses: balance1, Output R_I

Obviously, we make the same inference from the normalized metric, τ , in Fig. 8 as we did from the non-normalized metric, Λ ; namely, that the bidirectionality is large enough to detect unambiguously (so we reject the null hypothesis that $\tau = 0\%$ at the 0.05 significance level), and it is also large enough that we cannot reject at the 0.01 significance level the alternative hypothesis that $\tau \geq 0.5\%$ of the primary gage output near maximum calibration load. We therefore conclude, as before, that for output R_I of balance1 the observed bidirectionality is both large enough to detect and large enough to be of concern.

VII. Bidirectionality in Representative Balances

The previous section illustrated the detailed recipes for computing normalized and non-normalized bidirectionality metrics and their corresponding standard errors (“one-sigma”) for one of the primary gage outputs of a multi-piece Task balance designated as “balance1.” Criteria for rejecting formal null and alternative hypotheses about bidirectionality were also presented for this case. In this section we provide the results of similar calculations for the other five outputs of balance1, as well as for all outputs of three additional representative balances for which calibration data sets were available.

A. Balance1

Table 4 extends the information in Table 3 to include all six primary load gages for balance1, the multi-piece Task force balance for which one of the outputs has been examined in some detail already. Formal inferences for each primary gage output—whether to reject the null hypothesis and conclude that bidirectionality is greater than zero or to reject the alternative and conclude that the true bidirectionality is less than the 0.5% tolerance level—are also summarized.

Table 4. Results of bidirectionality assessment for balance1, a multi-piece Task balance

Balance1	Primary Gage Outputs (Corresponding Primary Gage Loads)					
	$R_1(N_1)$	$R_2(N_2)$	$R_3(S_1)$	$R_4(S_2)$	$R_5(RM)$	$R_6(AF)$
Λ	11.18	16.52	14.09	33.44	27.67	2.32
σ_Λ	0.591	0.597	1.768	1.699	1.495	0.937
Critical Λ for H_0	1.159	1.171	3.468	3.333	2.933	1.838
Critical Λ for H_A	2.590	2.741	-0.113	0.199	1.500	1.582
y	793.1	826.2	800.6	831.2	996.3	752.8
0.05% y	3.966	4.131	4.003	4.156	4.981	3.764
σ_y	0.598	0.609	1.758	1.696	1.495	0.938
τ	1.41%	2.00%	1.76%	4.02%	2.78%	0.31%
σ_τ	0.074%	0.072%	0.221%	0.205%	0.150%	0.124%
Critical τ for H_0	0.15%	0.14%	0.43%	0.40%	0.29%	0.24%
Critical τ for H_A	0.33%	0.33%	-0.01%	0.02%	0.15%	0.21%
Bidirectionality > 0?	YES	YES	YES	YES	YES	YES
Bidirectionality $\geq 0.5\%$ FS?	YES	YES	YES	YES	YES	YES

Table 4 reveals that for all six outputs, we are able to infer that the level of bidirectionality is great enough to detect and great enough to be of concern. This agrees with our preconceived notion that such a multi-piece balance would be bidirectional; however, there are some surprises in this table. Note that for $R_6(AF)$, while the normalized bidirectionality metric has a value that exceeds the critical value for the alternative hypothesis so that we cannot reject it, it nonetheless has a value of bidirectionality that is less than the 0.5% tolerance level. In Fig. 9 the reference distributions shown separately in Figs. 6 and 7 for the R_I output of balance1 are combined for the R_6 output, illustrating this case. This is the output corresponding to the axial force primary gage load.

Figure 9 illustrates the importance of incorporating the uncertainty of the bidirectionality metric into the assessment of bidirectionality. While the normalized bidirectionality metric’s value of 0.31% is less than the tolerance level of 0.50%, the uncertainty associated with estimating the bidirectionality is great enough that we are unable to say with high confidence that this difference is due to anything other than experimental error. It is possible that the true bidirectionality metric is indeed below the tolerance level, but we are not justified in making such an inference given the variance in our experimental data. Prudence dictates in such a case that we treat this output as bidirectional.

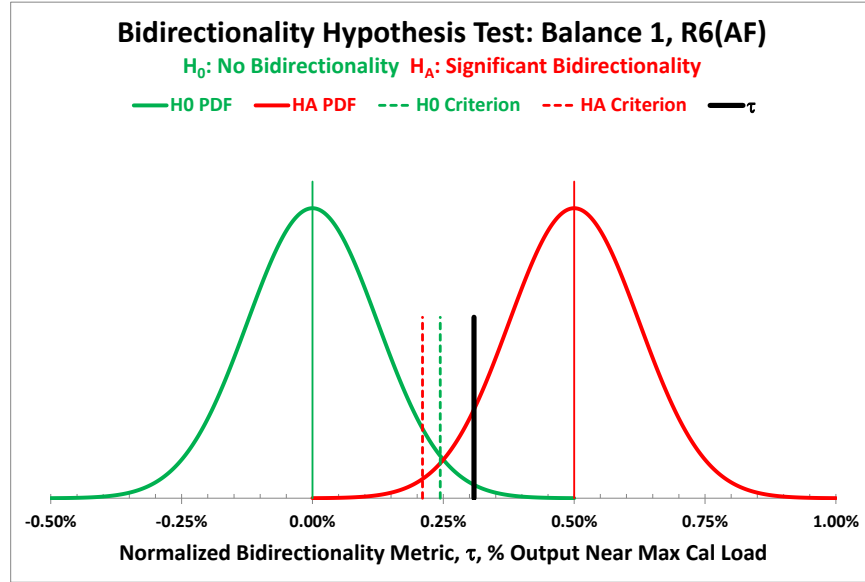


Figure 9. Normalized Bidirectionality Reference Distributions for Null and Alternative Hypotheses: balance1, Output R_6

There is another interesting feature in Fig. 9 that can be seen from Table 4 to apply also to outputs R_3 , R_4 , and R_5 as well as to R_6 . Note that the critical value for rejecting the null hypothesis (0.24%) is slightly greater than the critical value for rejecting the alternative hypothesis (0.21%). We see a somewhat more pronounced example of this behavior for output R_5 in Fig. 10. This is output corresponding to the rolling moment primary gage load.

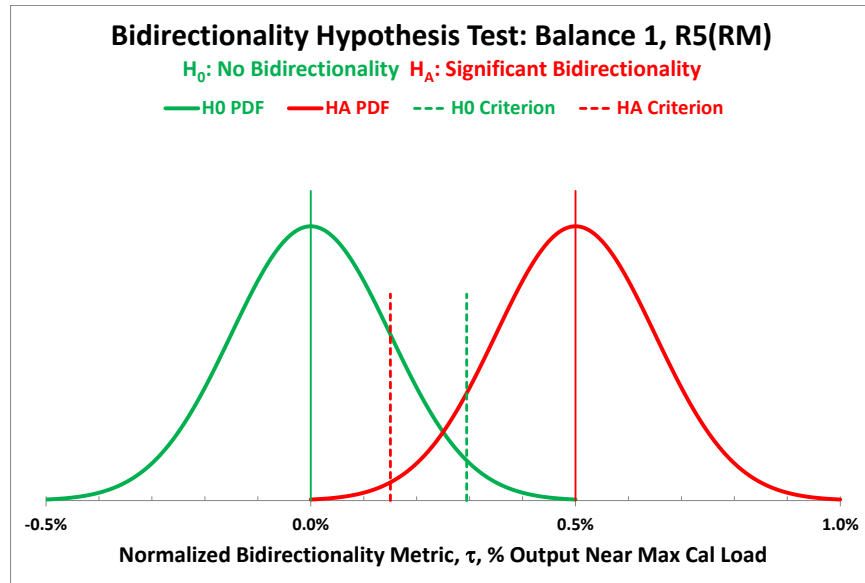


Figure 10. Normalized Bidirectionality Reference Distributions for Null and Alternative Hypotheses: balance1, Output R_5

Imagine in Figs. 9 and 10 if the bidirectionality metric were to fall between the red and green dashed lines representing the critical levels for rejecting H_0 and H_A . In such a case the bidirectionality metric would be too small to distinguish from zero with high confidence (that is, to the left of the green dashed line) and simultaneously too small to distinguish from a level of bidirectionality great enough to be of concern (that is, to the right of the red dashed line). This is a tell-tale sign of insufficient precision in the bidirectionality metric, in that we cannot clearly distinguish from zero certain levels of bidirectionality that are large enough to be of concern. For the R_5 output of

balance1 corresponding to rolling moment, the normalized bidirectionality metric was large enough that it could be unambiguously detected even given the relatively large experimental error in estimating it (2.78% so off to the right of the range illustrated in Fig. 10). Likewise, the R_6 output of balance1 corresponding to axial force can be seen in Fig. 9 to lie to the right of this criterion gap, suggesting that only the null hypothesis need be rejected. But this illustrates how calibration data quality can come into play in assessing bidirectionality.

If Figs. 9 and 10 illustrate the effects of *too little* precision in estimating bidirectionality, Fig. 8 illustrates the effect of having *too much*. There is a gap between the criteria levels in this figure as there is in Figs. 9 and 10, but the critical value for rejecting the null hypothesis (0.15%) is to the *left* of the critical value for rejecting the alternative hypothesis (0.33%). If an empirical estimate of the bidirectionality metric were to fall between the red and green dashed lines in this figure representing critical levels for rejecting H_0 and H_A , *both* hypotheses would have to be rejected. This is because the bidirectionality would be large enough to distinguish from zero but too small to be of interest.

This situation occurs when more resources have been expended than are necessary to achieve the precision necessary to make a reliable inference about the bidirectionality of the balance output. Not unlike the audiophile who has invested so much in his high fidelity stereo system that he can hear the conductor's asthma during quiet interludes of a symphony, we have invested so much in the calibration data sample that we can clearly observe effects that may be real, but are too small to be of any interest. In such a case there is some potential for resource savings, by acquiring fewer data points, for example. This would have the effect of increasing the width of the reference distributions, closing the gap between the critical levels.

The ideal scenario is one in which the critical levels for the null and alternative hypotheses coincide, so there is only one criterion for bidirectionality. In that case, metrics smaller than this criterion justify rejecting the alternative hypothesis, validating the output as free of bidirectionality, while larger metrics justify rejecting the null hypothesis, indicating that the output as bidirectional. The “scale” of the calibration experiment (that is, the volume of data acquired) can be optimized during the design of the calibration experiment to close the gap between the H_0 and H_A criteria levels. However, the current study utilized existing calibration data samples that were not scaled for bidirectionality. As a result, gaps exist between the H_0 and H_A criteria levels for all the balance outputs examined, indicating either a surplus or a deficit of precision depending on whether the H_0 criterion was less than (to the left of) the H_A criterion, or greater than (to the right) of it. Unlike the case for balance1, the estimated bidirectionality metric did fall within this gap for some of the outputs of other balances examined in this study.

B. Balance2

Calibration data for a second balance, identified as “balance2” in this study, were also analyzed to assess the bidirectionality of each of its outputs. This was a single-piece “hybrid” balance with outputs and load ranges as displayed in Table 5.

Table 5. Load ranges for balance2 calibration data

Coded Variables	Physical Variables		Calibration Range		Coded Variable Range	
	Load	Units	Min	Max	L ($x=-1$)	H ($x=+1$)
x_1	N_1	lbs	-2508	2500	-2500	2500
x_2	N_2	lbs	-2477	2481	-2500	2500
x_3	S_1	lbs	-1249	1248	-1250	1250
x_4	S_2	lbs	-1245	1249	-1250	1250
x_5	RM	in-lbs	-5034	4965	-5000	5000
x_6	AF	lbs	-696	695	-700	700

Table 6 displays the same kind of bidirectionality assessment results for balance2 as are displayed in Table 4 for balance1. The normalized and non-normalized bidirectionality metrics for all six primary gage outputs of balance2 are listed, as well as rejection criteria for the null hypothesis (no bidirectionality) and its alternative (bidirectionality great enough to be of concern). Formal inferences drawn by comparing the empirical estimates of bidirectionality with these criteria are summarized at the bottom of the table.

Table 6. Results of bidirectionality assessment for balance2, a single-piece “hybrid” balance

Balance2	Primary Gage Outputs (Corresponding Primary Gage Loads)					
	$R_1(N_1)$	$R_2(N_2)$	$R_3(S_1)$	$R_4(S_2)$	$R_5(RM)$	$R_6(AF)$
Λ	2.49	3.25	0.80	4.15	7.57	1.94
σ_Λ	0.314	0.505	0.470	0.441	1.408	0.270
Critical Λ for H_0	0.615	0.991	0.922	0.864	2.762	0.529
Critical Λ for H_A	4.011	4.598	1.782	1.902	3.445	3.569
y	948.4	1154.9	575.2	585.5	1344.8	839.5
0.05% y	4.742	5.774	2.876	2.927	6.724	4.197
σ_y	0.307	0.500	0.470	0.442	1.410	0.270
τ	0.26%	0.28%	0.14%	0.71%	0.56%	0.23%
σ_τ	0.033%	0.044%	0.082%	0.075%	0.105%	0.032%
Critical τ for H_0	0.06%	0.09%	0.16%	0.15%	0.21%	0.06%
Critical τ for H_A	0.42%	0.40%	0.31%	0.32%	0.26%	0.43%
Bidirectionality > 0?	YES	YES	NO	YES	YES	YES
Bidirectionality $\geq 0.5\%$ FS?	NO	NO	NO	YES	YES	NO

There are a number of interesting observations to be made in Table 6. This balance has been regarded historically as “non-bidirectional,” with no special provisions for bidirectionality normally made during its calibration. However, bidirectionality at some level was detected in five of the six primary gage outputs; all except the R_3 output corresponding to the forward side-force gage loading, S_1 .

At 0.14%, the normalized bidirectionality metric for the S_1 gage is indeed the lowest of all the outputs for balance2, but our inability to reject the null hypothesis for this gage is due at least as much to the relatively large standard error in estimating the metric for that gage as it is to the small size of the metric itself. At 0.082% the standard error (“one sigma” value) has a value of 58.7% of reading, over three times larger than the next-largest uncertainty expressed as a percent of reading. The 95% confidence interval (“two sigma”) encompasses both $\tau = 0$ and $\tau = 0.14\%$, rendering them indistinguishable with that level of confidence. This illustrates the importance of accounting not only for the magnitude of the bidirectionality metric, but the uncertainty in estimating it.

Table 6 indicates that levels of bidirectionality large enough to be of concern were not observed in every gage for which non-zero bidirectionality was detected, but such levels were observed in two of those five gages. Specifically, the R_4 and R_5 outputs corresponding to primary gage loads for the aft side-force gage (S_2) and for rolling moment, respectively, did display levels of bidirectionality large enough to be of concern, as Fig. 12 shows.

Figures 11 and 12 graphically display the bidirectionality of each of the six primary gage outputs for balance2, in multiples of the standard deviation in estimating each. Normalizing the bidirectionality metric by its standard error in this way permits all six gage outputs to be compared in one figure to a single reference distribution, even though the standard errors of all gages differ. Note that the peak of the normalized reference distribution is 0 in both Fig. 11 and Fig. 12 because in these normalized bidirectionality displays, the x-axis simply represents a displacement—a number of standard deviations away from where the reference distribution peaks.

The reference distribution in Fig. 11 corresponds to the null hypothesis, for which the peak occurs at $\tau = 0.0\%$. If the bidirectionality is greater than the criterion marked with a dashed line in this figure, we are entitled to reject the null hypothesis and conclude that the bidirectionality is non-zero, with no greater probability than 0.05 in this case of being in error. Figure 11 reveals that this was the case for five of the six primary gage outputs as noted earlier, with only the output corresponding to the forward side-force load (S_1) displaying bidirectionality at too low a level to be distinguished from zero with 95% confidence.

Figure 12 displays the reference distribution for the alternative hypothesis, for which the peak corresponds to $\tau = 0.5\%$. The alternative hypothesis states that bidirectionality exists at levels large enough to be of practical concern ($\geq 0.5\%$). If we reject this hypothesis, we are concluding that any non-zero bidirectionality there may be is at such a low level that we can afford to ignore it. If the empirical estimate of bidirectionality is to the left of the dashed line in Fig. 12, we can reject the alternative hypothesis with a probability no greater than 0.01 of an improper inference. This value of 0.01 represents the inference error risk we are willing to assume, which, for a specified volume of data, is dictated by how far the criterion is placed from the mean of the reference distribution. In Fig. 12 it is nominally 2.3 standard deviations from the mean, where the area under the reference distribution to the left of the criterion is 0.01.

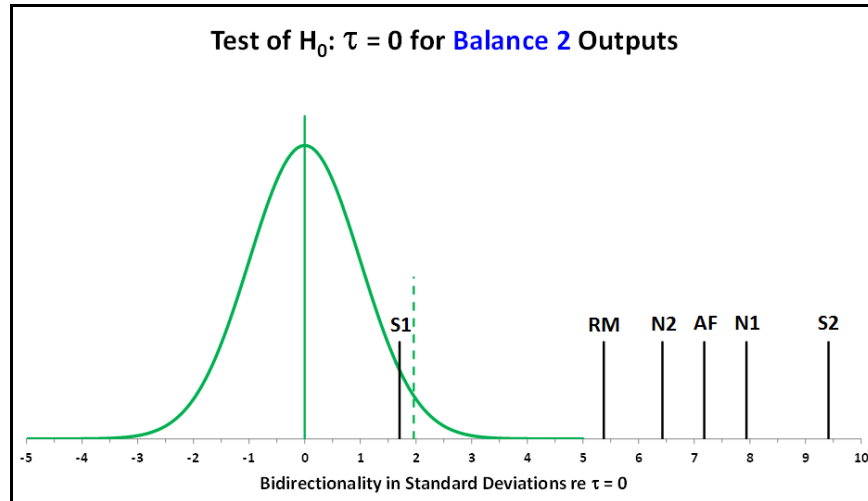


Figure 11. Balance2 Bidirectionality Reference Distribution for Null Hypothesis. Bidirectionality for each output expressed in multiples of its standard deviation.

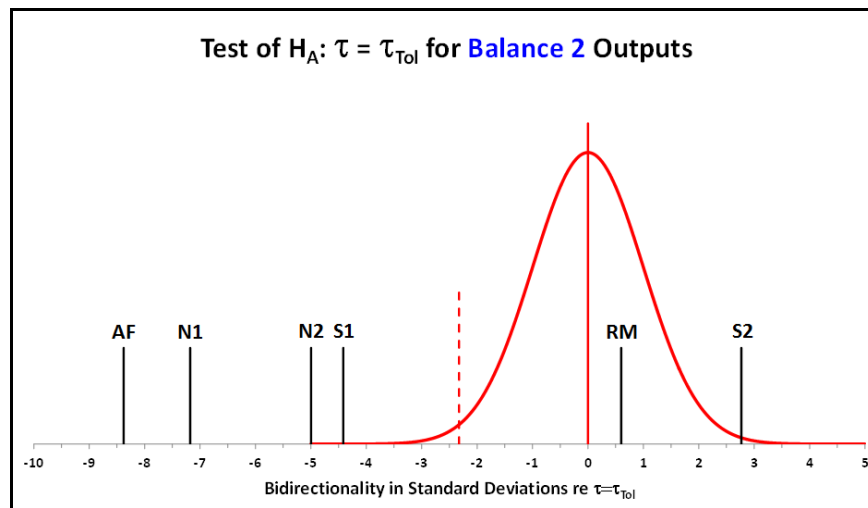


Figure 12. Balance2 Bidirectionality Reference Distribution for Alternative Hypothesis. Bidirectionality for each output expressed in multiples of its standard deviation.

We set the criterion somewhat further from the mean for the alternative hypothesis than for the null (2.3 sigma vs. 2.0 sigma) to make it less likely ($p = 0.01$ vs. $p = 0.05$) that we will erroneously reject the alternative hypothesis. The reason for this difference is that while erroneously rejecting either hypothesis is undesirable, the consequences are not the same. If we erroneously reject the null hypothesis, we attribute significant bidirectionality to a balance that is in fact free of significant bidirectionality. This may result in unnecessary precautions to account for the non-existent bidirectionality, but except for the additional resources wasted in that effort, there is relatively little harm done.

On the other hand, if we erroneously reject the alternative hypothesis, we declare a bidirectional balance to be free of significant bidirectionality, and validate the decision not to take bidirectionality into account in the calibration. This will result in an improper calibration, which we judge to be a more serious consequence than falsely indicting a non-bidirectional balance. For this reason, while we accept one chance in 20 of erroneously rejecting the null hypothesis, we only accept one chance in 100 of erroneously rejecting the alternative hypothesis.

In Figs. 8 through 10 above, we displayed the reference distributions for both the null hypothesis and its alternative in a single figure for one gage output, but for multiple gages we are unable to display both reference distributions in a single figure without clutter. This is because the displacement between peaks of the two reference distributions for each gage would be different if expressed in multiples of their unique standard deviations. For the

purpose of making an inference about the bidirectionality of a given primary gage output, however, it is sufficient to consider the null and alternative hypotheses separately as in Figs. 11 and 12.

C. Balance3

A single-piece moment balance identified as “balance3” was included in this study. It featured outputs and load ranges as displayed in Table 7.

Table 7. Load ranges for balance3 calibration data

Coded Variables	Physical Variables		Calibration Range		Coded Variable Range	
	Load	Units	Min	Max	L ($x=-1$)	H ($x=+1$)
x_1	PM1	in-lbs	-32053	32098	-32000	32000
x_2	PM2	in-lbs	-32264	32077	-32000	32000
x_3	YM1	in-lbs	-16708	16755	-18000	18000
x_4	YM2	in-lbs	-18128	18140	-18000	18000
x_5	RM	in-lbs	-8804	8837	-9000	9000
x_6	AF	lbs	-398	397	-400	400

Table 8 displays the bidirectionality assessment results for balance3. As in Tables 4 and 6, formal inferences were made by comparing empirical estimates of bidirectionality with criteria levels for the corresponding null and alternative hypotheses. These are summarized at the bottom of the Table 8.

Table 8. Results of bidirectionality assessment for balance3, a single-piece moment balance

Balance3	Primary Gage Outputs (Corresponding Primary Gage Loads)					
	negR ₂ (PM ₁)	R ₁ (PM ₂)	negR ₃ (YM ₁)	R ₄ (YM ₂)	R ₅ (RM)	R ₆ (AF)
Λ	6.51	3.26	4.53	2.92	5.59	3.89
σ_Λ	0.440	0.494	0.851	0.619	0.605	0.954
Critical Λ for H ₀	0.864	0.969	1.670	1.214	1.187	1.872
Critical Λ for H _A	7.486	7.408	5.169	6.109	1.664	3.269
y	1702.3	1711.8	1430.4	1510.1	614.8	1098.3
0.05% y	8.511	8.559	7.152	7.550	3.074	5.491
σ_y	0.439	0.494	0.851	0.737	0.605	0.955
τ	0.38%	0.19%	0.32%	0.19%	0.91%	0.35%
σ_τ	0.026%	0.029%	0.059%	0.041%	0.098%	0.087%
Critical τ for H ₀	0.05%	0.06%	0.12%	0.08%	0.19%	0.17%
Critical τ for H _A	0.44%	0.43%	0.36%	0.40%	0.27%	0.30%
Bidirectionality > 0?	YES	YES	YES	YES	YES	YES
Bidirectionality \geq 0.5% FS?	NO	NO	NO	NO	YES	YES

Even though balance3 is a single-piece balance historically assumed to be non-bidirectional, some level of bidirectionality was quantified for all six of its outputs. The bidirectionality of four of the six outputs was small enough to be of no practical significance, but the rolling moment output displayed rather substantial bidirectionality. The axial force output exhibited a level of bidirectionality that was less than the 0.5% tolerance threshold, but by so small a margin as to be indistinguishable from 0.5% within experimental error. Prudence would dictate in such a case that precautions should be taken in analyzing the axial force output to account for its bidirectionality, which we are unable to dismiss as too low to be of practical significance because of the uncertainty in estimating it.

The bidirectionality of each of the six primary gage outputs for balance3 is displayed in Figures 13 and 14 as multiples of the standard deviation in estimating each. Taken together, the results presented in these two figures are rather surprising. They suggest that we may not be justified in assuming that a balance is non-bidirectional, simply because it has historically been regarded as such because of its single-piece construction.

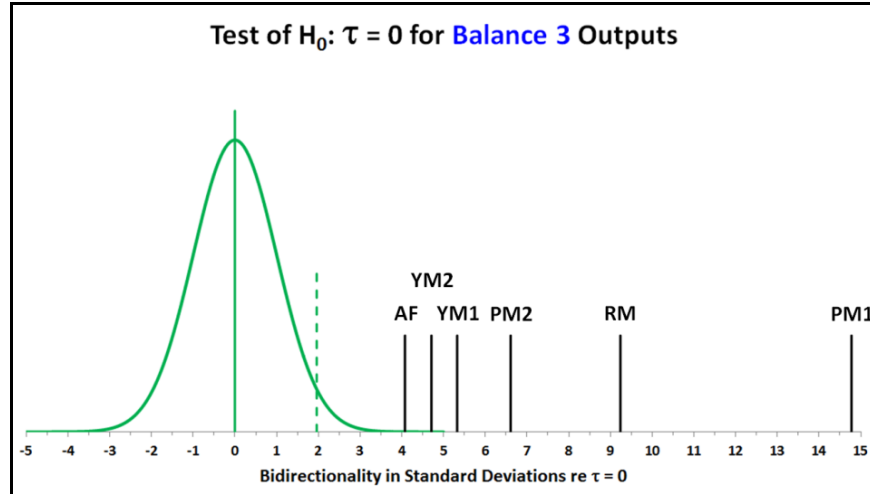


Figure 13. Balance3 Bidirectionality Reference Distribution for Null Hypothesis. Bidirectionality for each output expressed in multiples of its standard deviation.

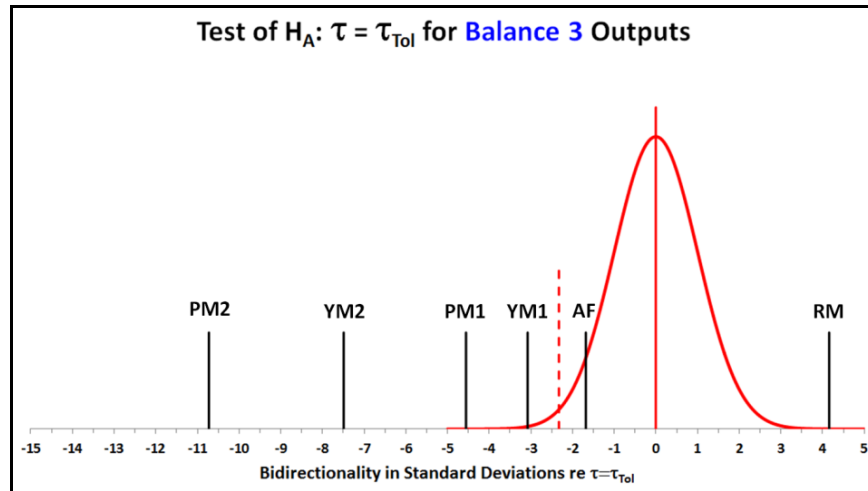


Figure 14. Balance3 Bidirectionality Reference Distribution for Alternative Hypothesis. Bidirectionality for each output expressed in multiples of its standard deviation.

D. Balance 4

The fourth and final balance examined in this study is a single-piece direct-read semispan balance with five outputs, identified as “balance4.” The outputs and load ranges of this balance are displayed in Table 9.

Table 9. Load ranges for balance4 calibration data

Coded Variables	Physical Variables		Calibration Range		Coded Variable Range	
	Load	Units	Min	Max	L ($x=-1$)	H ($x=+1$)
x_1	NF	lbs	-40000	40000	-40000	40000
x_2	PM	ft-lbs	-20000	20000	-20000	20000
x_3	YM	ft-lbs	-40041	40041	-40000	40000
x_4	RM	ft-lbs	-200233	200233	-190000	190000
x_5	AF	lbs	-8000	8000	-8000	8000

Table 10 displays the bidirectionality assessment results for balance4 in what by now the reader will recognize as a standard format used for all four balances of this study. As with the previous three balances, formal inferences

made by comparing empirical estimates of bidirectionality with criteria levels for the corresponding null and alternative hypotheses are summarized at the bottom of this table.

Table 10. Results of bidirectionality assessment for balance10, a single-piece semi-span balance

Balance4	Primary Gage Outputs (Corresponding Primary Gage Loads)				
	R ₁ (NF)	R ₂ (PM)	R ₃ (YM)	R ₄ (RM)	R ₅ (AF)
Λ	0.99	1.63	6.38	2.52	2.27
σ_{Λ}	1.775	0.896	1.023	1.396	1.147
Critical Λ for H_0	3.487	1.760	2.011	2.743	2.255
Critical Λ for H_A	0.225	0.874	1.769	0.621	0.435
γ	873.6	592.9	831.5	775.9	622.7
0.05% γ	4.368	2.964	4.158	3.880	3.113
σ_{γ}	1.775	0.898	1.290	1.413	1.119
τ	0.11%	0.27%	0.77%	0.33%	0.37%
σ_{τ}	0.203%	0.151%	0.123%	0.180%	0.184%
Critical τ for H_0	0.40%	0.30%	0.24%	0.35%	0.36%
Critical τ for H_A	0.03%	0.15%	0.21%	0.08%	0.07%
Bidirectionality > 0?	NO	NO	YES	NO	YES
Bidirectionality \geq 0.5% FS?	YES	YES	YES	YES	YES

A glance at the summary of bidirectionality inferences at the bottom of Table 10 reveals a curious result. For three of the five outputs, the bidirectionality metric could not be distinguished from zero with 95% confidence, given the uncertainty in estimating it. These are the normal force, pitching moment, and rolling moment outputs. However, for these three outputs as well as the remaining two outputs of the balance, we are able to infer bidirectionality levels large enough to be of concern!

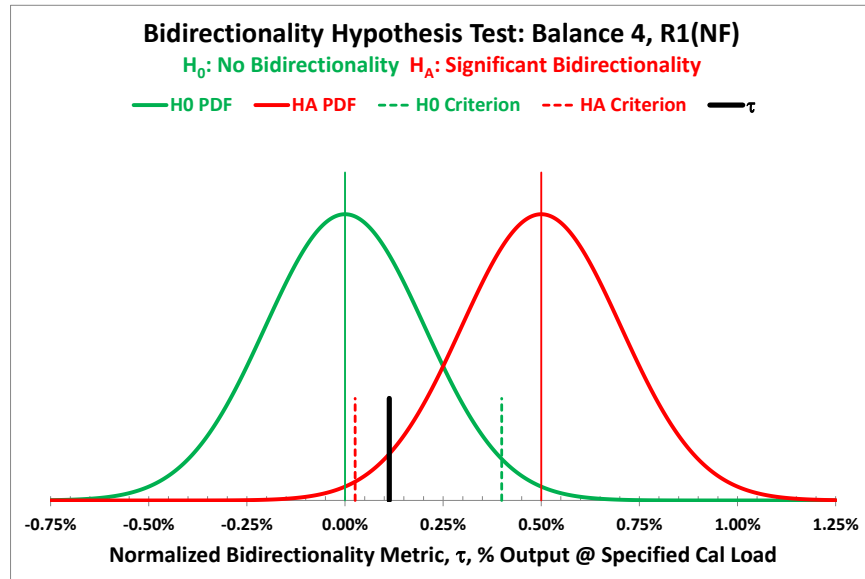


Figure 15. Normalized Bidirectionality Reference Distributions for Null and Alternative Hypotheses: balance4, Output R₁

How is it possible for the bidirectionality to be large enough to be troubling and at the same time small enough to be indistinguishable from zero, as with the *NF*, *PM*, and *RM* outputs of balance4? Figure 15 reveals how this mystery can be explained for the normal force output in terms of the unusually large experimental error in the bidirectionality metric for this output. The same explanation applies to the other balance outputs for which

bidirectionality levels are observed that are large enough to be of concern at the same time they are indistinguishable from zero within experimental error.

In the case of the normal force output illustrated in Fig. 15, the uncertainty in estimating bidirectionality is so large that the reference distributions for the null and alternative hypotheses overlap substantially. In fact, they overlap so much that the criteria for rejecting each hypothesis “switch sides.” That is, the criterion for rejecting the null hypothesis appears to the *right* of the criterion for rejecting the alternative hypothesis. If, as in Fig. 15, the bidirectionality metric falls in the gap between these two criteria, then it is to the *left* of the criterion for the null hypothesis (rendering it inappropriate to reject the null hypothesis) at the same time it is to the *right* of the criterion for the alternative hypothesis, rendering it likewise inappropriate to reject the alternative hypothesis. So for bidirectionality metrics falling in this gap, we can reject neither hypothesis; *both* must be embraced simultaneously!

We might suppose that this lack of precision reflects an inadequate number of points in the calibration data sample. That is not likely to be the explanation for balance4, however. It can be shown¹⁸ that the volume of data necessary to generate a polynomial response model with a prediction error tolerance of δ and inference error tolerances for erroneously rejecting the null and alternative hypotheses of α and β , respectively, given a standard error in the measurements of σ , can be computed as follows, where p is the number of terms in the polynomial response model, including the intercept:

$$n = \left[\left(z_\alpha + z_\beta \right)^2 \left(\frac{\sigma}{\delta} \right)^2 \right] p \quad (43)$$

As an absolute minimum one must have at least one data point for every term in a polynomial response model, so $n = p$ is the fewest points that can be specified. The term in brackets is a multiplier that describes how much this minimum must be extended when data are acquired in a measurement environment characterized by a standard data error of σ , to account for specified levels of prediction tolerance, δ , and inference error risk tolerance, α and β .

The volume of data prescribed by Eq. (43) is sufficient to ensure that there will be a probability no greater than α of erroneously declaring a residual to be an outlier, and a probability no greater than β of erroneously validating a residual as within the tolerance of δ when it is not. The tolerance value is subject to the experimenter’s discretion but absent any other special preference, a reasonable value might be the “95% Least Significant Difference (LSD).” By definition, a response model prediction that differs by no more than the 95% LSD from a confirmation measurement at the same point is close enough that no difference between measurement and model prediction can be resolved at the 95% confidence level. A model that can be said with 95% confidence to predict responses that are indistinguishable from measurements would be regarded as adequate in many practical aerospace applications.

The 95% LSD has a value of two times the square root of sigma. Establishing this as the prediction tolerance, we have

$$\delta = 2\sqrt{2}\sigma \quad (44)$$

Inserting this into Eq. (43) yields this result:

$$n = \left[\frac{1}{2} \left(\frac{z_\alpha + z_\beta}{2} \right)^2 \right] p \quad (45)$$

That is, for a given measurement environment, once the prediction tolerance is established the data volume specification is simply a matter of how much inference error risk one is willing to tolerate. In the present study, α and β had values of 0.05 and 0.01, respectively, for which the corresponding z-statistics (unit normal deviates) z_α and z_β are 1.96 and 2.33, respectively. Inserting these values into Eq. (45) produces this simple data volume specification:

$$n = 2.3p \quad (46)$$

That is, we need a minimum of 2.3 data points for each term in the response model to ensure that a well-fitted model produces residuals within the 95% Least Significant Difference such that true outliers are mislabeled as within tolerance no more than 1% of the time, and points that are within tolerance are mislabeled as outliers no more than 5% of the time.

The number of terms, p , in the response model depends on the order of the model, the number of independent variables, and the fraction of model terms that survive the term reduction process by which terms are rejected from the model unless they are large enough compared to the uncertainty in estimating them. For balance4, the largest response model has a total of 26 terms. This happens to be the response model for the normal force primary load that is represented in Fig. 15. We would therefore expect that a calibration load schedule with at least $2.3 \times 26 = 59.8 = 60$ data points would provide adequate precision for the calibration model. There were in fact 498 points in the calibration data sample. This is over eight times more data than necessary to produce an adequate calibration model by the quality specifications for δ , α , and β that have been specified. While Eq. (45) describes data volume requirements for an adequate calibration response model and not an adequate bidirectionality assessment per se, it is implausible that eight times more data than necessary to generate adequate calibration modeling precision would still be insufficient to ensure adequate precision for the evaluation of bidirectionality.

A more likely explanation for the poor precision in estimating bidirectionality for balance4 is that the bidirectionality metric is a function of regression coefficients for which the uncertainty in estimating them is inflated by the presence of multicollinearity. A Variance Inflation Factor (VIF) quantifies the impact of multicollinearity on the uncertainty of each regression coefficient, and is typically computed automatically by standard regression analysis software packages.

The VIF has a value of “1” when there is no inflation and thus no adverse impact of multicollinearity. For the normal force response model of balance4, VIF values as high as 7000 were associated with the regression coefficients upon which the bidirectionality assessment depends! This implies a high degree of correlation among the individual regressors in the response model.

The multicollinearity exhibited in the calibration load schedule for balance4 was so pervasive that a full second-order model could not be generated because the design matrix was of inferior rank. That is, there were fewer independent regressors than terms in the full second-order model. Only after the term reduction process had eliminated numerous model terms was the design matrix of sufficient rank to evaluate.

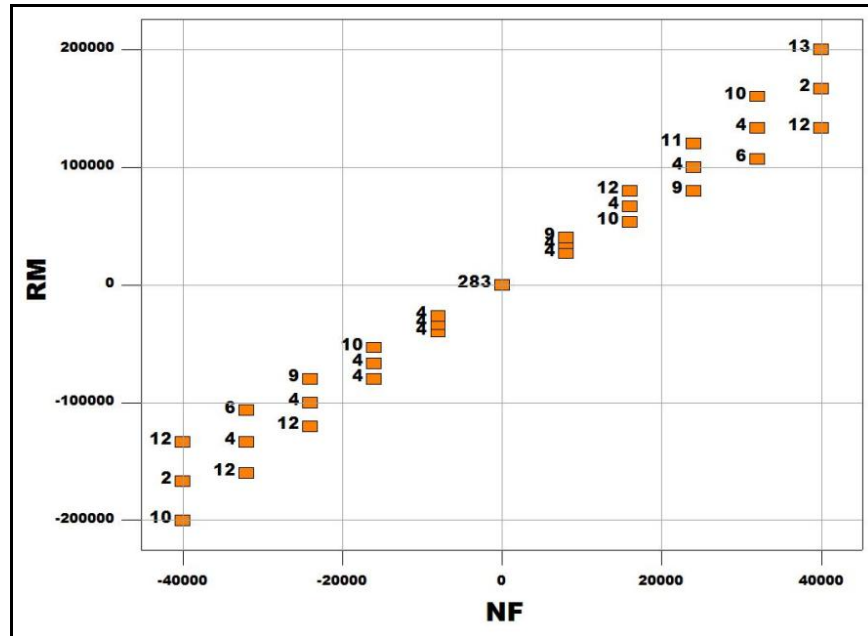


Figure 16. Correlation between two balance4 regressors. Numbers indicate replicated combinations of rolling moment and normal force.

Figure 16 illustrates the correlation among two of the balance4 regressors. The number of load combinations with identical levels of the two plotted loads (rolling moment and normal force) is displayed next to each point. The

clear correlation between these two regressors is evident. High levels of correlation were also observed among other balance4 regressors.

The resulting inflation of variance in key regression coefficients no doubt contributed to the imprecision with which bidirectionality could be evaluated for this balance. See Section IV, *Uncertainty in the Bidirectionality Indicator Variable*, and especially Eqs. (26) and (29). These two equations reveal the relationship between the variance in normalized and non-normalized bidirectionality metrics and the variances of the regression coefficients upon which they depend (substantially inflated for balance4). It is the resulting inflation in the variance of the bidirectionality metric that is revealed graphically in Fig. 15.

E. Summary of Bidirectionality Assessments

Bidirectionality assessments were performed for all outputs of four representative balances to illustrate the statistical theory of bidirectionality developed earlier in this paper. Each assessment was based on the rejection of one of two formal hypotheses: a null hypothesis that no significant bidirectionality is detected, or an alternative hypothesis that bidirectionality, τ , is out of tolerance with levels that equal or exceed 0.5% of the output at full calibration load. Figure 17 displays a summary of all possible combinations of these two inferences.

		Detectable?	
		YES	NO
Unacceptable?	NO	<p><i>More Precision than Needed</i></p> <p>Enough to Detect Insignificant Bidirectionality</p>	<p>Not Enough Bidirectionality to Detect</p>
	YES	<p>Significant Bidirectionality</p>	<p><i>Not Enough Precision</i></p> <p>Cannot Detect Significant Bidirectionality</p>

Figure 17. Possible outcomes of inferences made during bidirectionality assessment.

One normally expects the null and alternative hypotheses to be mutually exclusive, so that rejecting the null hypothesis implies that the alternative hypothesis is not rejected, and conversely. However, for the balance outputs assessed in this study, the variance in the bidirectionality metric tended to one of two extremes that each caused this convention to be violated, as in the upper left and lower right quadrants of Fig. 17.

In some cases, the bidirectionality metric could be assessed with such high precision that levels too small to be of concern could be easily resolved. The *null* hypothesis could be rejected in such cases because the metric was large enough to be easily distinguished from zero, and at the same time the *alternative* hypothesis could be rejected because the metric was so small as to be clearly below the 0.5% tolerance level. Thus, whenever the bidirectionality metric was quantified with significantly greater precision than necessary to ensure acceptable inference error risk, it became necessary to reject *both* hypotheses. This happened in about a third of the cases examined (7 cases in 23, or 31%).

There were also cases in which the uncertainty in estimating bidirectionality was so great that *neither* the null hypothesis *nor* the alternative hypothesis could be rejected. These were cases in which the bidirectionality metric could not be clearly distinguished *either* from zero *or* from the specified tolerance level of 0.5%. The uncertainty in estimating the bidirectionality metric was so great that not even certain levels large enough to be of concern could not be distinguished from zero with acceptable levels of confidence. This imprecision, which has been attributed to multicollinearity in the load schedule, was observed in 3 of the 23 cases examined, or 13%.

Figure 18 displays the frequency with which each of the outcomes displayed in Fig. 17 occurred in the present study. As noted, a total 23 balance outputs were assessed, from three 6-channel balances and one balance with only five outputs.

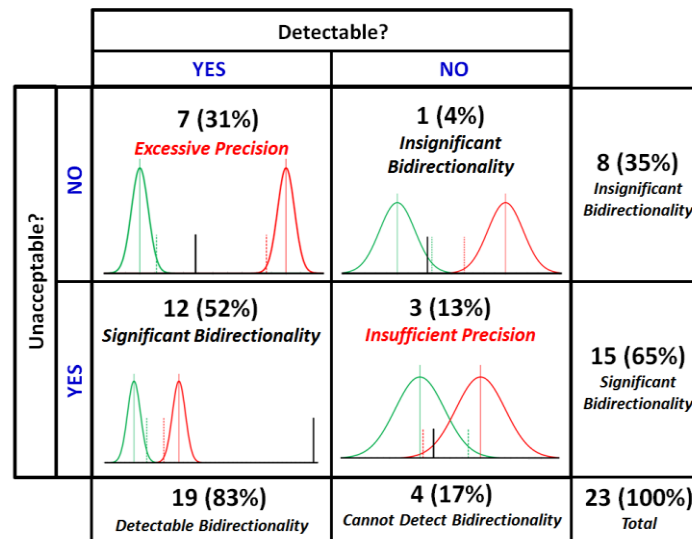


Figure 18. Frequency of inference combinations made during bidirectionality assessment.

Each quadrant of Fig. 18 displays a sketch of typical reference distributions for the null hypothesis (green, on the left, with a mean of 0%) and the alternative hypothesis (red, on the right, with a mean of 0.5%). The black vertical line in each sketch marks where the bidirectionality metric would typically lie for the scenario being illustrated, and the green and red dashed lines are critical values of the null and alternative hypotheses, respectively. The number of balance outputs represented by each quadrant is displayed, and in parentheses this number is expressed as a percentage of the 23 outputs that were examined. A brief comment describes the circumstances represented in each quadrant.

The left column of Fig. 18 represents cases in which the bidirectionality was large enough to be distinguished from zero with 95% confidence, given the existing experimental error in the data. For scenarios depicted on the right, the signal-to-noise ratio was poor enough that bidirectionality could not be distinguished from zero with at least 95% confidence. This could be the result of a relatively small level of bidirectionality in the balance output, or relatively large experimental error in estimating it. Levels of bidirectionality great enough to distinguish from zero with at least 95% confidence were observed in a total of 19 of the 23 outputs examined, or 83% of the time.

In at least two of the remaining four outputs, the bidirectionality metric was great enough to be indistinguishable from the 0.5% tolerance level within experimental error, but it still could not be distinguished from zero simply because the experimental error was so large. With more representative levels of bidirectional uncertainty, those two cases would also have been distinguishable from zero with at least 95% confidence, raising the total from 19 cases to 21 out of 23, or 91%. We conclude that bidirectionality is as ubiquitous a characteristic of balance outputs as non-linearity and channel interactions. If the sample of balances examined in this study is representative, as we believe it to be, then situations in which no bidirectionality can be detected are expected to be rare. It appears not to be so much a question of *whether* a balance output displays bidirectionality, but *how much* bidirectionality it displays.

The upper row in Fig. 18 depicts cases in which the alternative hypothesis is rejected because the empirical estimate of bidirectionality can be said with 99% confidence to lie below the 0.5% tolerance threshold at or beyond which the bidirectionality is judged to be great enough to warrant special precautions in the balance calibration. For the lower row, bidirectionality levels were detected that are either indistinguishable within experimental error from the 0.5% threshold, or unambiguously greater, so that special precautions to account for bidirectionality would be recommended when developing the calibration response models.

In 15 of the 23 cases examined (65%), we were unable to say with 99% confidence that the true bidirectionality was small enough to validate it as within the 0.5% tolerance level. In only eight cases (35%) were we able to declare with 99% confidence that the true bidirectionality level was less than the 0.5% tolerance threshold.

Three overarching characteristics of bidirectionality emerge from this study. The first is that it is apparently ubiquitous; prudence requires a default assumption of some level of bidirectionality in every balance output. The second is that the bidirectionality of any given balance output is *likely* to be great enough to warrant special precautions when the calibration equation is developed. In this study the odds were 2:1 that bidirectionality would be large enough to be of concern in any given balance output. Such levels were observed across all balance types, including single-piece balances that may have been thought to be free of bidirectionality because of their construction details and other design characteristics. The third broad characteristic of bidirectionality observed in this study is that bidirectionality is more properly associated with individual balance outputs than with the balance as a whole. In general, a given balance is likely to feature some outputs with levels of bidirectionality large enough to be of concern, and others with bidirectionality levels well below the 0.5% tolerance threshold.

One other generality can be offered based on the results of this study. Rolling moment seems to be especially vulnerable to bidirectionality. For all four balances, the bidirectionality metric for rolling moment either exceeded the 0.5% tolerance level or was indistinguishable from it within experimental error.

VIII. Discussion

Miscellaneous remarks on a number of topics are collected in this section. These topics include a clarification of why a two-sided null hypothesis is adopted when the bidirectionality metric is defined in terms of its absolute value, and a brief summary of how bidirectionality uncertainty assessments can provide insights into the nature of the calibration load schedule. We also outline ways that examining bidirectionality components can provide additional information about the balance and the calibration process. Finally, we discuss two topics that impact the mechanics of computing the bidirectionality metrics; namely the symmetry of the load schedules and the non-orthogonality of the calibration response models.

A. Polarity of the Null Hypothesis

There may seem to be a complication in the analysis due to the absolute value of the bidirectionality metric. It is not uncommon for a null hypothesis to be “two-sided,” meaning that it can be said to be true if the estimated metric of interest resides within a specified range on *either side* of the mean of some reference distribution.

Because the bidirectionality metric is an absolute value, it may seem that the null hypothesis should be “one-sided.” That is, it may seem that if we have associated with the null hypothesis an inference error probability of 0.05, say, and we do not reject the null hypothesis, then we must mean that the probability is less than 0.05 that the estimated metric is more than a specified distance from zero, and the probability is 0 that it is negative. However, it is merely a labeling convention to express the bidirectionality metric as an absolute value, and such an arbitrary convention can have no actual impact on the probability of an inference error.

The empirically estimated bidirectionality metric can be less than the true value of the metric or greater, since it is just as likely that experimental error will cause one to understate the true bidirectionality metric as overstate it. The null hypothesis is therefore properly regarded as two-sided. Nonetheless, because of the absolute value convention, we represent the bidirectionality metric as positive in various graphical depictions, with no loss of generality. We note in passing that no such ambiguity attaches to the alternative hypothesis, which is always one-sided in hypothesis testing; we ask only if a potential effect is non-zero, not whether it is positive or negative.

B. Insights into the Load Schedule

The discussion surrounding Fig. 17 centers on the consequences of too much or too little precision in the empirical estimate of bidirectionality. The widths of the reference distributions associated with the null and alternative hypotheses provide a graphical indication of how adequate the precision is. If the widths of the two distributions are narrow compared to the 0.5% tolerance level so that a relatively broad range of bidirectionality levels lies between them, this is an indication that more loading combinations were included in the load schedule than necessary to assess the bidirectionality of the balance output. In that case, it is possible to detect bidirectionality levels much too low to be of interest. Since there is no need to pay for levels of precision so great that they permit uninteresting effects to be observed, a broad valley between two narrow peaks suggests that there is potential to save time and direct operating costs by acquiring fewer points in subsequent calibrations.

It can be argued by those who use automated calibration machines that as a practical matter, once the purchasing costs have been incurred, data volume is an insignificant contributor to automated calibration costs because of the per-point efficiency of such machines. There is thus little incentive to control costs by acquiring less data. A comprehensive cost/benefit comparison of automated calibration machines versus manual dead-weight loading transcends the scope of the current paper, except to say that a specific volume of data suffices to perform an

adequate calibration and bidirectionality assessment (see Eqs. (43) through (46) and surrounding discussion). While acquiring substantially more than this does no real harm, the benefits are likewise limited.

It is seldom the case in practice that calibration data sets are too small to provide adequate precision; the converse is much more likely. However, there are situations in which there can be insufficient precision to adequately calibrate a balance and assess the bidirectionality of its outputs even when an ample volume of data is available. This was discussed at some length above when the results of the balance4 bidirectionality assessment were presented. That discussion described how poor precision can result from multicollinearity inflating the variance of regression coefficients that comprise the calibration response model.

Multicollinearity occurs when calibration loadings are correlated as in Fig. 16, which represents a two-dimensional design space in which each data point is a site in that space. The design space site distribution can be optimized to reduce multicollinearity, however. One way, which can be applied to the full multidimensional design space that includes all regressors in the calibration model, is to use a D-optimal design. In a D-optimal design, the loading combinations are chosen to minimize the volume of the joint confidence ellipsoid for the response model regression coefficients. This can reduce the Variance Inflation Factor for these coefficients, which will reduce the variance in the normalized and non-normalized bidirectionality metrics, per Eqs. (26) and (29), as previously noted. D-optimal designs are readily available from various experiment design software packages^{14, 19, 20}.

In summary, while the level of the bidirectionality metric conveys important information about the calibration of the balance output, the uncertainty in estimating it reveals useful information about the calibration load schedule. If the reference distributions used to assess bidirectionality are very narrow, more data may have been acquired than are needed to make reliable inferences. If the data volume is ample but the reference distributions are still very wide, and especially if they are so wide that they overlap substantially, correlated loads may be producing levels of multicollinearity that result in excessive inflation of the regression coefficient variance.

C. Insights into “Process Discontinuities”

The bidirectionality metric is a function of coefficients that quantify the change in intercept, slope, and curvature that are induced by a change in load polarity. See Eqs. (16) and (20). The c_0 coefficient in these equations quantifies the change in intercept that accompanies a load polarity change, and may be of special practical interest.

While the c_1 and c_2 coefficients that quantify changes in sensitivity and non-linearity may provide some insight into the effects of certain balance design characteristics, the c_0 coefficient could reveal something about the calibration process itself. If this coefficient is frequently found to be significant, it is possible that it might be attributable to subtle differences in the way that positive and negative loads are physically applied during the calibration, especially in dead-weight manual calibrations. This could lead to procedural improvements that minimize offsets associated with load polarity changes. The contrary is also true. If the c_0 coefficient is consistently shown to be insignificant, this would tend to validate the “process methodology.” Either result would be useful.

D. Asymmetry in the Calibration Loading Schedule

The bidirectionality metrics proposed in this paper for a given primary gage output are only defined when the other independent variables in the calibration response model are zero. If the regression were performed with independent variables expressed in engineering units (typically pounds for forces and inch-pounds or foot-pounds for moments), this would imply zero physical loads on the other balance inputs. However, the bidirectionality metrics presented here were developed in terms of coded variables.

Coded variables have a range of ± 1 , into which the range of a corresponding physical variable is linearly mapped. “Zero” in coded units corresponds to the center of the range of the corresponding physical variable, which is only zero in physical units if the range is symmetrical about zero. This is typically the case for balance calibration loads, so it is not usually an issue.

Furthermore, the loading does not have to be perfectly symmetrical as long as the physical loads corresponding to coded variables ± 1 differ only in sign and span enough of the load range to avoid excessive extrapolation in representing the largest physical loads. For example, the actual N_1 load range for balance1 was -2073 lbs to +2126 lbs. This was linearly mapped into a coded variable range for which ± 1 corresponded to ± 2100 lbs, which has the requisite symmetry. In this case, loads with the largest absolute values will not always be precisely 1.000 in coded units but might be slightly greater or smaller than 1. This has no impact on the analysis as long as customary cautions against extensive extrapolation are observed.

If the load schedule included, say, axial force loads that were only of one polarity, then the coded variables would imply that the bidirectionality metric could only be defined for an axial force of half the maximum load. It would be possible to derive a more general metric that is dependent on all secondary loads instead of assuming them to be zero, but the complexity of that effort renders it beyond the scope of this initial effort. Therefore at the current

stage, symmetry in the loading schedule is a prerequisite for a valid bidirectionality assessment using the metrics presented in the paper.

E. Non-orthogonality Effects

An orthogonal polynomial response model has a property that would be especially convenient for assessing bidirectionality. The numerical value of each coefficient in such a model is independent of whether any other term in the model is retained or rejected. Extending a first-order model by adding second-order terms would of course change model predictions because the second-order effects would be included, but the coefficient of each first-order term would remain unchanged by the addition of the higher-order terms. If the model is not orthogonal, then adding terms to the model or deleting them will not only alter model predictions, it can also change the numerical value of coefficients for retained terms that were previously in the model.

Unfortunately, polynomial response models typically encountered in balance calibrations are not orthogonal. This is a complication, because it is common practice to improve the precision of calibration model predictions by rejecting as many terms from the full response model as possible. This is because the prediction variance averaged over all terms in the regression is directly proportional to p , the number of terms retained in the fitted polynomial. Terms are therefore only retained when the regression coefficient is sufficiently large compared to the standard error in estimating it.

The assessment of “sufficiently large” entails some judgment, however. The situation is further exacerbated by the fact that each time a term is rejected from a non-orthogonal model, the coefficients of all other terms can shift by an amount that depends on which term was rejected. The *order* in which terms with small coefficients are rejected from the model is therefore also a factor in determining what the final ensemble of retained terms will be. Different analysts seldom generate identical reduced models.

Equations (20) and (29) reveal that the normalized bidirectionality metric and its variance depend on only six regression coefficients, dubbed for the purposes of this analysis “the Cardinal Six” coefficients. These are b_i and c_i , for $i = 0, 1$, and 2 . In the present analysis, the Cardinal Six terms were always retained during term reduction. Unfortunately, because of the non-orthogonality of each polynomial response model, the precise numerical value of these six coefficients will always depend to some degree on which additional terms are retained in the response model. That in turn can depend on how the final response model is constructed. This means that it is possible to develop somewhat different values for the bidirectionality metric and also the criteria for rejecting null and alternative hypotheses in the bidirectionality assessment process, depending on which terms appear in the final response model besides the Cardinal Six.

The possible variance in bidirectionality assessment results due to non-orthogonality in the calibration response models is acknowledged, but there is good reason to believe that it will be small. Terms that are very nearly insignificant are very small indeed, so even if different subsets of them are retained in the final model causing different non-orthogonality effects in the Cardinal Six terms that determine the bidirectionality metric and its variance, those effects should likewise be small. Nonetheless, we should be sensitive to the fact that slightly different bidirectionality metrics might be produced by different analysts from the same calibration data sample. In certain borderline cases, it may even be possible for such differences to influence the assessment of bidirectionality, although this is expected to occur only rarely.

F. Additional Caveats

The current study utilized balances believed to be representative, and included a moment balance, a force balance, a hybrid balance, and a direct-read balance. However, the selection of balances to study was dictated largely by the availability of existing calibration data sets. A comprehensive assessment of the methodology will depend on its application to balances with a variety of construction/design details.

IX. Summary and Concluding Remarks

This paper extends the theory of bidirectionality introduced by Ulbrich¹⁰ at the 28th AIAA Aerodynamic Measurement Technology, Ground Testing, and Flight Testing Conference in 2012. That theory is extended in the following three ways: 1) the metric originally proposed by Ulbrich is normalized to permit easy comparisons with the tolerance level he proposed, 2) a categorical variable is introduced in the regression analysis to account for load polarity, and 3) the uncertainty in both the normalized and non-normalized bidirectionality metrics is quantified. These extensions are applied to four representative balances to quantify bidirectionality metrics for each, as well as the uncertainty in each metric. This information was used to make formal inferences regarding the bidirectionality of each balance. The following results were obtained:

- 1) Bidirectionality levels large enough to be distinguished from zero with 95% confidence were detected in 83% of the balance outputs that were examined. This is roughly five times out of six.
- 2) In 65% of the cases examined, bidirectionality levels were large enough to either exceed the Ulbrich threshold of 0.5% of output at full calibration load, or to be indistinguishable from it due to experimental error. These results were not limited to multi-piece balances traditionally assumed to be bidirectional, but applied also to single-piece balances.
- 3) Results 1 and 2 suggest that bidirectionality is common, and likely to be too large to prudently ignore in any given balance output.
- 4) The categorical variable introduced to account for load polarity differences in the calibration model regression analysis facilitates an intuitively satisfying physical interpretation of bidirectionality as the net effect of changes in offset, sensitivity, and linearity induced when the load polarity changes.
- 5) Absolute levels of bidirectionality are relatively small, typically ranging from a few tenths of one percent of the balance output at full calibration load to a percent or two. While small in absolute terms, these levels constitute a large fraction—and in some cases a large multiple—of the total error budget for a representative balance calibration, in which the standard error of residuals is typically expected to be less than 0.25% of the output at full calibration load.
- 6) The precision with which bidirectionality is estimated can be great enough to resolve levels that are much too small to be of interest. These situations are common, and are attributed to calibration data samples with substantially more data than necessary to make adequate inferences.
- 7) The precision with which bidirectionality is estimated can be so low that levels of bidirectionality great enough to be of concern nonetheless cannot be distinguished from zero within experimental error. This imprecision is attributed to the multicollinearity that characterizes calibration data samples with correlated loads.

Appendix

Dependence of Bidirectionality Uncertainty on Maximum Calibration Loads

Equation (31), reproduced here for convenience as Eq. (A-1), represents the variance in estimating the normalized bidirectionality metric for positive and negative load:

$$\sigma_{\tau_{\pm}}^2 = \frac{y_{\pm}^2 \sigma_{\Lambda}^2 + \Lambda_{\pm}^2 \sigma_y^2}{y_{\pm}^4} \quad (\text{A-1})$$

In this appendix we derive the following useful result: For the commonly occurring situation in which the magnitudes of the maximum positive and negative calibration loads are nominally the same, the variance in the empirical estimate of the normalized bidirectionality metric is essentially the same for positive and negative loading.

From Eq. (A-1) we have:

$$\frac{\sigma_{\tau_{+}}^2}{\sigma_{\tau_{-}}^2} = \left(\frac{y_{-}}{y_{+}} \right)^4 \times \frac{y_{+}^2 \sigma_{\Lambda}^2 + \Lambda_{+}^2 \sigma_y^2}{y_{-}^2 \sigma_{\Lambda}^2 + \Lambda_{-}^2 \sigma_y^2} = \left(\frac{y_{-}}{y_{+}} \right)^4 \times \frac{y_{+}^2 \sigma_{\Lambda}^2 \left[1 + \left(\frac{\sigma_y}{y_{+}} \right)^2 \left(\frac{\Lambda_{+}}{\sigma_{\Lambda}} \right)^2 \right]}{y_{-}^2 \sigma_{\Lambda}^2 \left[1 + \left(\frac{\sigma_y}{y_{-}} \right)^2 \left(\frac{\Lambda_{-}}{\sigma_{\Lambda}} \right)^2 \right]} \quad (\text{A-2})$$

We introduce some notational simplification by defining t in a standard way as the ratio of some empirical estimate and the uncertainty in estimating it. Specifically, we have

$$t_{y_{+}} = \frac{y_{+}}{\sigma_y}, \quad t_{y_{-}} = \frac{y_{-}}{\sigma_y}, \quad t_{\Lambda_{+}} = \frac{\Lambda_{+}}{\sigma_{\Lambda}}, \quad t_{\Lambda_{-}} = \frac{\Lambda_{-}}{\sigma_{\Lambda}}$$

which, when inserted into Eq. (A-2), yields the following:

$$\frac{\sigma_{\tau_+}^2}{\sigma_{\tau_-}^2} = \left(\frac{y_-}{y_+} \right)^2 \times \frac{\left[1 + \left(\frac{t_{\Lambda_+}}{t_{y_+}} \right)^2 \right]}{\left[1 + \left(\frac{t_{\Lambda_-}}{t_{y_-}} \right)^2 \right]} \quad (\text{A-3})$$

The t -values for y_{\pm} represent the electrical gage outputs near maximum positive and negative calibration load as predicted by the calibration response model, in multiples of the standard error in those predictions. Because the maximum calibration loads can be relatively large, the corresponding response prediction estimates can also be relatively large, especially in comparison with the small standard errors in those predictions typically attributable to the high precision of a modern balance calibration. For example, y_+ and y_- values for the R_I output of balance1 were 793.12 and -792.62 mV/Volt, respectively, while the standard error (identical for both of them) was only 0.60 mV/Volt. The corresponding t -values were thus 1321.9 and 1321.0, respectively.

The t -values for Λ_{\pm} represent empirical estimates of the non-normalized bidirectionality metric for positive and negative load, expressed as multiples of the standard error in those estimates. Again for the R_I output of balance1 we have (see Eq. (35)) $\Lambda_+ = 11.18$ mV/Volt and $\Lambda_- = 5.03$ mV/Volt. The standard error is the same for both of them: 0.349 mV/Volt. The corresponding t -values are thus 32.0 and 14.4, respectively. For the purpose of testing a null hypothesis that the bidirectionality is zero, these Λ t -values are large, in that a little more than 3 is sufficient to infer that the bidirectionality metric is non-zero with less than one chance in a thousand of being wrong (99.9% confidence). Nonetheless, they are two orders of magnitude smaller than the t -values for y_{\pm} , so the squared t_{Λ}/t_y ratios in the numerator and denominator of the square-bracket term in Eq. (A-3) are quite small. They are about four orders of magnitude less than one ($5.9 \text{ E-}04$ and $1.2 \text{ E-}04$, respectively, in the case of the positive and negative R_I outputs of balance1). The bracketed term in Eq. (A-3) is then, to an excellent approximation, “1” (actual numerical value for R_I of balance1: 0.99953). We therefore drop the square-bracketed term from Eq. (A-3) and arrive at this interesting result:

$$\frac{\sigma_{\tau_+}}{\sigma_{\tau_-}} \approx \left| \frac{y_-}{y_+} \right| \quad (\text{A-4})$$

Equation (A-4) suggests that for loading schedules resulting in balance outputs with nominally equal magnitudes near maximum positive and negative calibration load, the uncertainty in estimating the normalized bidirectionality metric, τ , is essentially independent of the polarity of the load. It also suggests an inverse relationship between the magnitude of the response near maximum calibration load of one polarity, and the uncertainty in estimating bidirectionality at the other polarity. That is, if the loading schedule is such that the magnitude of the output near maximum positive calibration load is greater than the magnitude of the output near maximum negative calibration load, then the uncertainty in estimating the normalized bidirectionality at negative load will be greater than the uncertainty in estimating the normalized bidirectionality at positive load, and conversely.

For R_I of balance1, the magnitude of y_-/y_+ is $793.12/792.62 = 1.0006$. By Eq. (A-4) this would mean that the two bidirectionality standard errors were identical to within about 6 parts in ten thousand (0.06%) but when one accounts for the square-bracketed term of Eq. (A-3), computed for this case to be 0.99953, the variances agree even more closely—within about 1 part in ten thousand or 0.01%.

Acknowledgements

The support of the Engineering Directorate at NASA Langley Research Center and the Wind Tunnel Division at NASA Ames Research Center is gratefully acknowledged.

References

- ¹AIAA/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, American Institute of Aeronautics and Astronautics, Reston, Virginia, 2003.
- ²Box, G. E. P., and N. R. Draper: *Empirical Model Building and Response Surfaces*. New York: John Wiley & Sons. 1987
- ³Myers, R. H., and D. C. Montgomery: *Response Surface Methodology. Process and Product Optimization Using Designed Experiments*. New York: John Wiley & Sons. 1995
- ⁴Draper, N. R., and H. Smith: *Applied Regression Analysis*, 2nd ed. New York: John Wiley & Sons. 1981
- ⁵Parker, P. and DeLoach, R. “Response Surface Methods for Force Balance Calibration Modeling” 19th International Congress on Instrumentation in Aerospace Simulation Facilities, Cleveland Ohio, August 2001.
- ⁶Ulbrich, N. and Volden, T., “Strain–Gage Balance Calibration Analysis Using Automatically Selected Math Models,” AIAA 2005–4084, paper presented at the 41st AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit, Tucson, Arizona, July 2005.
- ⁷Ulbrich, N. and Volden, T., “Development of a New Software Tool for Balance Calibration Analysis,” AIAA 2006–3434, paper presented at the 25th AIAA Aerodynamic Measurement Technology and Ground Testing Conference, San Francisco, California, June 2006.
- ⁸DeLoach, R., and Ulbrich, N., “A Comparison of Two Balance Calibration Model Building Methods (Invited)”. AIAA 2007–0147, 45th AIAA Aerospace Sciences Meeting and Exhibit, Reno, Nevada. Jan 8–11, 2007.
- ⁹DeLoach, R., and Philipsen, I., “Stepwise Regression Analysis of MDOE Balance Calibration Data Acquired at DNW (Invited)”, AIAA 2007–0144, 45th AIAA Aerospace Sciences Meeting and Exhibit, Reno, Nevada Jan 8–11, 2007
- ¹⁰Ulbrich, N., Detection of Bi-Directionality in Strain Gage Balance Calibration Data, AIAA-2012-3320, 28th AIAA Aerodynamic Measurement Technology, Ground Testing, and Flight Testing Conference, 25–28 June 2012, New Orleans, Louisiana
- ¹¹Marquardt, D.W. and R. D. Snee [1975], “Ridge Regression in Practice,” *Am. Statist.*, 29(1),3–20.
- ¹²Montgomery, D.C., E.A. Peck., and C.G. Vining, *Introduction to Linear Regression Analysis*, Wiley Series in Probability and Statistics, 3rd ed., John Wiley and Sons, New York, 2001.
- ¹³DeLoach, R., “The Role of Hierarchy in Response Surface Modeling of Wind Tunnel Data” AIAA 2010-931, 48th AIAA Aerospace Sciences Meeting and Exhibit, Orlando, Florida, January 4–7, 2010.
- ¹⁴Design Expert, Software Package, Ver. 8.07.1, StatEase, Inc., Minneapolis, Minnesota, 2011.
- ¹⁵Piexoto, J.L., “A Property of Well-Formulated Polynomial Regression Models,” *Amer. Statist.*, Feb. 1990, Vol. 44, No. 1.
- ¹⁶Peixoto, J.L., “Hierarchical Variable Selection in Polynomial Regression Models,” *Amer. Statist.*, 1987, 41, pp 311–313.
- ¹⁷Kempthorne, O., “Classificatory Data Structures and Associated Linear Models,” *Statistics and Probability: Essays in Honor of C. R. Rao*, G. Kallianpur, P. R. Krishnaiah, and J. K. Ghosh, Eds., Amsterdam, North Holland, pp. 397–410.
- ¹⁸DeLoach, R., “Response Surface Modeling Tolerance and Inference Error Risk Specifications: Proposed Industry Standards (Invited)” AIAA 2012-2859, 28th AIAA Aerodynamic Measurement Technology and Ground Testing Conference. New Orleans, Louisiana. June 25–28, 2012.
- ¹⁹Minitab, Software Package, Ver. 14, Minitab, Inc., State College, Pennsylvania, 2003.
- ²⁰JMP, Software Package, Ver. 6.0.3, SAS Institute, Cary, North Carolina, 2006.