Rice University, Department of Mechanical Engineering

AOX_BalCal User Guide

Meade Research Group

John Potthoff
4-28-2020

Table of Contents

Introduction

AOX_BalCal was developed as a tool for use in internal balance calibration. AOX_BalCal utilizes a non-iterative approach to model component loads directly from the strain gage (aka, gage) outputs. The program was designed to be efficient in processing large datasets and flexible in accepting data inputs. Rather than relying on single data points and an iterative process to estimate tare loads, AOX_BalCal estimates tare loads simultaneously as it computes the model's coefficients using global least squares regression. This procedure bases the tare estimates on every point in a series rather than a single measurement. The loads may be modeled using the polynomial combinations of the gage outputs typically employed by the balance community. Additionally, AOX_BalCal is capable of employing GRBFs to fully model the gage output to load relationship or to improve a polynomial model. The GRBFs are added sequentially by a greedy algorithm that is designed to minimize the need for trial-and-error attempts by the user. This includes automatically selecting the locations for GRBF centers, determining the GRBF widths and coefficients, and automatically terminating the GRBF addition algorithm prior to overfitting the calibration data. AOX_BalCal can be applied to generate the mathematical model (calibration), test the model's ability to generalize with new data (validation), and approximate unknown component loads (approximation).

The program was written and tested in Matlab 2019b. The choice of Matlab as a programming language provides transparency for the user in the software's approach and can be run with common operating systems (e.g., Windows, OS X, Linux). The program can also be run as a compiled standalone application on computers without the full Matlab software package, for wider distribution to the balance community. If the program is run in the full Matlab environment, the results can be easily examined or further processed as desired by the user through the major variables stored in the local workspace. Lastly, the program format enables further improvements or modifications to be made with relative ease.

Though the program was originally developed for use with wind tunnel strain-gage balance calibration datasets, it can also be applied for regression analysis of more general multivariate experimental datasets. An option in the graphical user interface (GUI) switches between the "Balance Calibration" and "General Function Approximation" modes. This document will generally use the terminology for the balance calibration mode. "Component Loads" are the response variables that the generated mathematical models will predict. "Gage Outputs" are the predictor variables used in these mathematical models.

A detailed description of the approach taken by AOX_BalCal and necessary background information is provided in the April 2020 thesis written by John Potthoff. This document is intended to provide a brief overview of getting started with the program. It focuses on the program settings, input files, and possible outputs.

Required Software

If running and editing AOX_BalCal in the MATLAB environment, MATLAB version 9.7 (R2019b) and the Statistics and Machine Learning Toolbox (version 11.6) are required. Using AOX_BalCal as a standalone application requires version 9.7 (R2019b) of the MATLAB Runtime. If not already installed, the MATLAB runtime will be added during the standalone application installation.

GUI Overview

Inputs to AOX_BalCal are made though a Graphical User Interface (GUI). To start, run the script:
AOX_BalCal.m. This will launch the GUI. The GUI is shown here in Balance Calibration mode. A button in the top left corner of the GUI switches to General Function Approximation mode.

Application Mode
- ◉ Balance Calibration
- ○ General Function Approximation
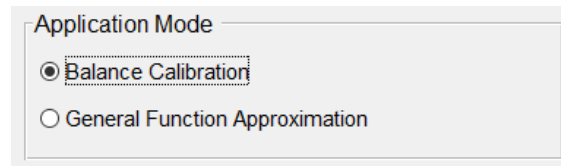
The following section explains each of the GUI options for Balance Calibration mode.
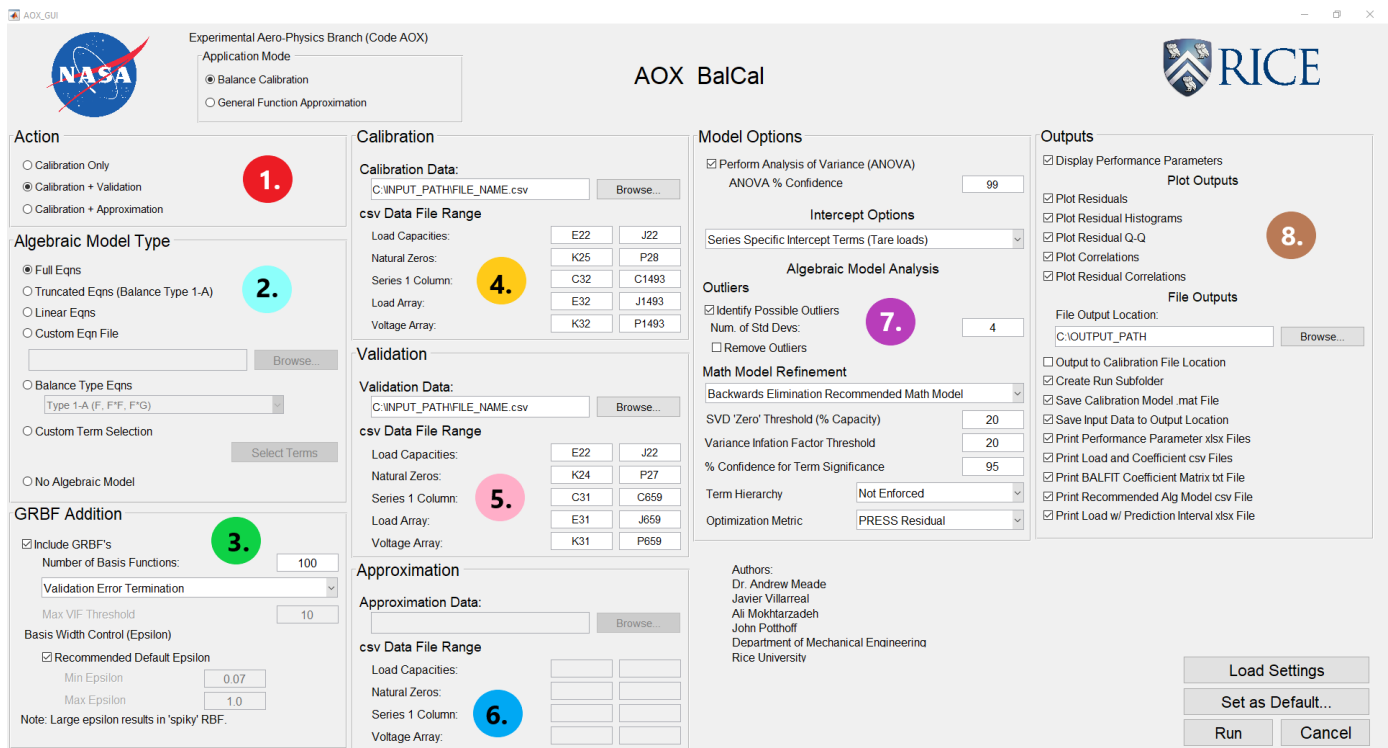
GUI Sections

The main GUI is split into the following eight panels (numbered on the figure above), based on the program option categories:

1. Action
2. Algebraic Model Type
3. GRBF Addition
4. Calibration
5. Validation
6. Approximation
7. Model Options
8. Outputs

The following sections detail the options within each of these panels.

Action Panel

Within the action panel, the user may select between the following options:

1. Calibration Only
2. Calibration + Validation
3. Calibration + Approximation

Calibration develops the mathematical model for the response variables from predictor variables. Validation applies this model to new, unseen data to test its ability to generalize. Approximation applies the model to a set of predictor variables to approximate the unknown response variables. Within AOX_BalCal, calibration must always be performed. If the user desires apply a model developed in AOX_BalCal to approximate new data without performing calibration again, they may use the program **AOX_Approx**.

Algebraic Model Type Panel

The Algebraic Model Type panel selects the polynomial terms which may be included in the mathematical model. The options are as follows:

1. **Full Eqns:** Include the full set of possible polynomial terms as shown below [1]. *F* is the component load (response variable) and Δ*R* is the gage output difference (predictor variable).

$$F_i = \sum_{j=1}^{n} b1_{i,j}\Delta R_j + \sum_{j=1}^{n} b2_{i,j}\left|\Delta R_j\right| + \sum_{j=1}^{n} c1_{i,j}\Delta R_j^2 + \sum_{j=1}^{n} c2_{i,j}\Delta R_j\left|\Delta R_j\right| +$$

$$\sum_{j=1}^{n}\sum_{k=j+1}^{n} c3_{i,j,k}\Delta R_j\Delta R_k + \sum_{j=1}^{n}\sum_{k=j+1}^{n} c4_{i,j,k}\left|\Delta R_j\Delta R_k\right| +$$

$$\sum_{j=1}^{n}\sum_{k=j+1}^{n} c5_{i,j,k}\Delta R_j\left|\Delta R_k\right| + \sum_{j=1}^{n}\sum_{k=j+1}^{n} c6_{i,j,k}\left|\Delta R_j\right|\Delta R_k +$$

$$\sum_{j=1}^{n} d1_{i,j}\Delta R_j^3 + \sum_{j=1}^{n} d2_{i,j}\left|\Delta R_j^3\right| + \sum_{j=1}^{n}\sum_{\substack{k=1\\k\neq j}}^{n} c7_{i,j,k}\Delta R_j\Delta R_k^2 +$$

$$\sum_{j=1}^{n}\sum_{k=j+1}^{n}\sum_{q=k+1}^{n} c8_{i,j,k,q}\Delta R_j\Delta R_k\Delta R_q$$

2. **Truncated Eqns (Balance Type 1-A):** Basic model for modeling component loads without bi-directional behavior. The **Truncated Eqn** is the same as the **Balance Type 1-A** option under option 5. The following terms are included:

$$F_i = \sum_{j=1}^{n} b1_{i,j}\Delta R_j + \sum_{j=1}^{n} c1_{i,j}\Delta R_j^2 + \sum_{j=1}^{n}\sum_{k=j+1}^{n} c3_{i,j,k}\Delta R_j\Delta R_k$$

3. **Linear Eqns:** Only include linear gage output differences as shown:

$$F_i = \sum_{j=1}^{n} b1_{i,j}\Delta R_j$$

4. **Custom Eqn File:** Input a custom set of terms using a .csv file. The polynomial terms included may differ between component loads. The .csv file should include the first row labeling the target component load and the first column labeling the polynomial terms. Beyond these labels, the file should contain an array to denote which terms to include. Non-zero entries in the array denote the term should be included. Zero entries indicate the term should not be included. The gage outputs and component loads do not need to match between the current calibration and the .csv file. **The dimensions of predictor and response variables must match, however.** Therefore, the recommended procedure is to first run the dataset once using a different selection for polynomial terms. After this run, the files **AOX_ALG_MODEL_COEFFICIENT_MATRIX.csv** and **DIRECT_RECOMM_CustomEquationMatrix.csv** will both have the proper labels and dimensionality for the dataset. The user can then modify these files, if desired, by denoting which terms should be included using 1's and 0's. The file **AOX_ALG_MODEL_COEFFICIENT_MATRIX.csv** may also be used in its unedited form to select the polynomial terms. Any terms with non-zero coefficients will be included. When using a coefficient matrix as the **Custom Eqn File**, AOX_BalCal will recalculate the coefficients, not using those from the previous run.

5. **Balance Type Eqns:** The user may select common subsets of the possible polynomial terms from a dropdown menu. Within the dropdown, *F* corresponds to Δ*R_j* and *G* corresponds to Δ*R_k* in the equation provided for **Full Eqns.**

6. **Custom Term Selection:** The user may select their own custom set of terms to include in the model using a window accessed from the **Select Terms** button. The window is shown below. The polynomial terms selected will be included in the models for each component load. In the term selection window, F corresponds to ΔR_j, G corresponds to ΔR_k, and H corresponds to ΔR_q in the equation provided for **Full Eqns.**
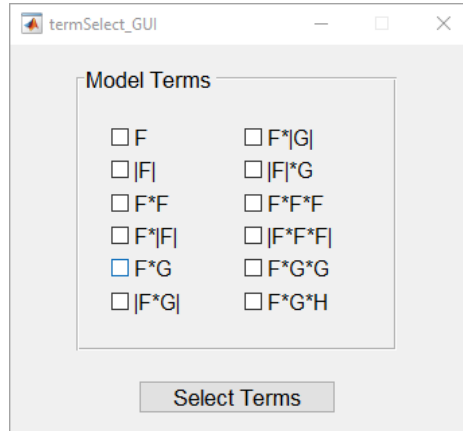
7. **No Algebraic Model:** Select this option to include no polynomial terms if you would like to generate a GRBF-only model

GRBF Addition Panel

This panel provides the options related to including Gaussian Radial Basis Functions (GRBFs) in the model. The following options are provided:
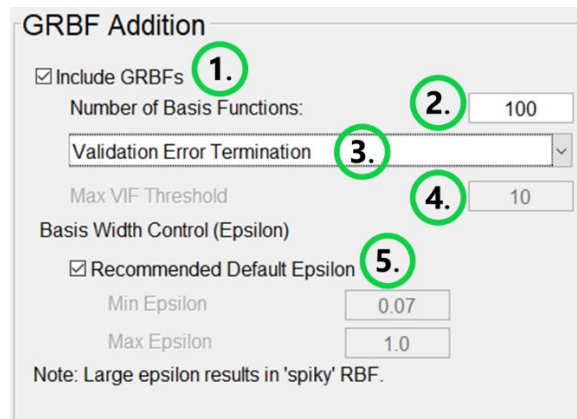
1. **Include GRBFs:** Selection for if GRBFs should be included in the model
2. **Number of Basis Functions:** Input the maximum number of GRBFs to include in the model for each component load as a positive integer. If this input is greater than the number of calibration datapoints, it will be set within AOX_BalCal to the number of calibration datapoints.
3. **Early Termination Dropdown:** The addition of GRBFs may be terminated automatically to prevent overfitting. **Validation Error Termination** is only available as an option if **Calibration + Validation** is selected in the Action Panel. The remaining three early termination options

(Options c, d, and e) are only available if **Perform Analysis of Variance** is selected in the Model Options panel. GRBF addition will terminate independently for each component load model. The following options may be selected from a dropdown:

    a. **No Early Termination:** Place the number of GRBFs specified in option 2.

    b. **Validation Error Termination:** Halt GRBF addition based on increasing residual RMS for validation data. Option only available if **Calibration + Validation** is selected in the Action Panel. Trim final number of GRBFs in each channel for the minimum Validation residual RMS.

    c. **PRESS Termination:** Halt GRBF addition based on increasing PRESS statistic [2]. Option only available if **Perform Analysis of Variance** is selected in the Model Options panel. Trim final number of GRBFs in each channel for minimum PRESS statistic.

    d. **Prediction Interval Termination:** Halt GRBF addition based on increasing mean Prediction Interval for calibration datapoints [2]. Option only available if **Perform Analysis of Variance** is selected in the Model Options panel. Trim final number of GRBFs in each channel for minimum mean prediction interval.

    e. **VIF + Prediction Interval Termination:** Halt GRBF addition based on increasing mean Prediction Interval for calibration datapoints or when no GRBFs can be added without exceeding the VIF threshold set in option 4 [2]. Option only available if **Perform Analysis of Variance** is selected in the Model Options panel. Trim final number of GRBFs in each channel for minimum mean prediction interval.

4. **Max VIF Threshold:** Limit for maximum VIF for predictor terms in model if **VIF + Prediction Interval Termination** is selected. If adding GRBFs to a polynomial model and the polynomial model already exceeds the VIF threshold, no GRBFs will be added.

5. **Recommended Default Epsilon:** Option to use the recommended default range for epsilon or input a custom range. Epsilon is the shape parameter which controls the GRBF widths [1]. It is optimized for each GRBF as it is placed. A larger value of epsilon results in a narrow or sharp GRBF while a small value of epsilon results in a wide or a gradual GRBF.

Calibration/Validation/Approximation Panel

The calibration, validation, and approximation panels specify data inputs to AOX_BalCal. The first field in each panel specifies the path do the input file. The data may be input via .csv files or .cal, .val, or .app files for calibration, validation, and approximation, respectively. During a run, AOX_BalCal produces .cal, .val, and .app files with the calibration, validation, and approximation input data as appropriate. These files may be used to perform additional runs on the same dataset without having to specify the .csv file ranges. If inputting data from a .csv file, the following ranges must be specified using spreadsheet notation (e.g., A1). The following figures show the panels for calibration, validation, and approximation completed to input data from a .csv file. A figure is also provided showing the example .csv file. In this figure, the appropriate data file ranges are highlighted and labeled.

1. **Load Capacities:** Capacities for the component loads. Range must match the dimensionality of component loads. Capacities should be in the same order as within the **Load Array** field.
2. **Natural Zeros:** Range of raw gage outputs from the balance used to approximate natural zeros. Each row in the range will be treated as the output from a roll orientation. Columns in range must match the order and number of gage output channels in the **Voltage Array** field.

AOX_BalCal calculates the natural zeros as the average of rows provided. If gage output differences are provided in the data file, the **Natural Zeros** field may be selected as an array of zeros where the number of columns matches the number of gage outputs.

3. **Series 1 Column:** Column for integer series labels. AOX_BalCal calculates different tare loads for each series.
4. **Load Array:** Array of component loads (response variables). Columns in the array denote the load component and rows contain the individual datapoints. Any number of component loads may be provided.
5. **Voltage Array:** Array of gage outputs (predictor variables). Columns in the array denote the gage output channels and rows contain the individual datapoints. Number of columns may differ between the **Load Array** and **Voltage Array.** The number of predictor variables may be greater than or equal to the number of component loads in balance calibration mode.

Note: Loads are not known when performing approximation, so the **Load Array** field is not included in the Approximation Panel. The input data file requirements are consistent between the Approximation action in AOX_BalCal and the standalone program AOX_Approx.

Model Options

The Model Options panel contains various options for developing and evaluating the calibration model. They options are as follows:

1. **Perform Analysis of Variance (ANOVA):** Selecting this option evaluates the mathematical model generated during calibration using an assortment of statistical metrics.
2. **ANOVA % Confidence:** If ANOVA is performed, this setting determines the level of confidence used where performing hypothesis tests.

Algebraic Model Analysis:

3. **Intercept Options Dropdown:** The following options are provided:
   a. **Series Specific Intercept Terms (Tare loads):** Select this option to estimate tare loads. Indicator variables are used to generate a unique intercept term for each series. These intercepts are the offset between the applied loads in the series and the total load on the balance. Therefore, they provide the tare loads for each series.
   b. **Global Intercept Term:** A single intercept term will be included for all series. This intercept will also be included when applying the model to new data. As component loads are approximated as a function of gage output differences, this intercept should not be necessary, especially if only using polynomial terms. It is included as an option for unique scenarios, however.

c. **No Intercept Term:** No intercept terms will be included in the model. This is the recommended selection if the dataset has been tare-corrected so the supplied loads are the total loads.

4. **Identify Possible Outliers:** Selecting this option returns a list of data points from the calibration dataset which are outliers based on the residual magnitude compared to residual standard deviation.

5. **Num. of Std Devs:** Input the threshold number of standard deviations from the mean residual for a residual to be identified as an outlier.

6. **Remove Outliers:** If selected, AOX_BalCal will remove the datapoints from calibration data identified as outliers based on the definition in Option 5. Then it will recalculate the calibration model without these outlier terms. A datapoint which is an outlier based on its residual in one or more component load channels is removed when recalculating the calibration model for all channels.

7. **Math Model Refinement Dropdown:** The set of polynomial terms selected in the **Algebraic Model Type** panel can be refined automatically based on the following options:

   a. **No Model Refinement:** All polynomial terms selected in the **Algebraic Model Type** panel will be included in the model

   b. **SVD for Non-Singularity (Permitted Math Model):** Singular Value Decomposition (SVD) is applied to ensure no exact linear dependencies exist between the polynomial terms. This ensures a unique solution exists for the coefficients. This is referred to as the Permitted Math Model in BALFIT. Generating this model is the first set in all the following refinement options.

   c. **BALFIT Legacy Constrained (Suggested) Math Model:** Selecting this option applies constraints on maximum VIF and term significance to the permitted math model. This option applies the constraints using BALFIT's routine for generating its Suggested Math Model [3].

   d. **Updated Constrained (Suggested) Math Model:** Apply constraints on maximum VIF and term significance to the permitted math model using an updated routine. This routine is more sequential to ensure the final model will meet both constraints but is more computationally expensive than the Legacy Constrained algorithm.

   e. **Forward Selection Recommended Math Model:** The recommended model takes the set of polynomial terms in the permitted math model, applies the constraints on maximum VIF and term significance, and optimizes the set of included terms based on a selected metric. Forward selection starts by including the minimum set of required polynomial terms and builds the model by sequentially adding more terms.

   f. **Backwards Elimination Recommended Math Model:** The recommended model takes the set of polynomial terms in the permitted math model, applies the constraints on maximum VIF and term significance, and optimizes the set of included terms based on a selected metric. Backwards elimination starts with the permitted model and builds the model by sequentially removing terms.

8. **SVD 'Zero' Threshold (% Capacity):** Input is used when generating the Permitted Model. This input sets a threshold as a percentage of gage output capacity. If gage output capacities are not found in the input data file, the maximum magnitudes from raw gage outputs are considered the capacity. When generating the permitted model, gage output differences below this

threshold are temporarily set to zero. The selected polynomial terms are generated from the set of gage output differences with the threshold applied. SVD is applied to these polynomial terms to test for linear independence. Lowering the threshold will allow more terms to be included in the permitted model. If SVD should be applied to the polynomial terms generated with no threshold applied, set this input to 0.

9. **Variance Inflation Factor Threshold:** This input sets the maximum VIF allowed when generating the constrained or recommended math models.

10. **% Confidence for Term Significance:** This input sets the confidence level used to determine if polynomial terms are statistically significant. All included terms must be statistically significant at this confidence level for the constrained or recommended math models.

11. **Term Hierarchy:** Dropdown with the following options for enforcing term hierarchy [4] when generating the constrained or recommended model:
    a. **Not Enforced:** Included polynomial terms do not need to be hierarchically supported
    b. **Enforced After Search:** Check is performed after generating the model to ensure the polynomial terms are hierarchically supported. For the Constrained and Forward Selection recommended model, this option will add terms required for hierarchical support but may result in a final model that does not meet the maximum VIF and term significance constraints
    c. **Enforced During Search:** Term hierarchy is enforced during the search algorithm for the Constrained or Recommended Model. This is the recommended option if desiring a model which obeys the hierarchy rule. This model ensures the final set of polynomial terms are hierarchically supported and meet the maximum VIF and term significance constraints.

12. **Optimization Metric:** When generating the recommended math model, AOX_BalCal will optimize the included terms based on the metric selected from the following options:
    a. **PRESS Residual:** Minimize the PRESS statistic
    b. **Sqrt of Residual Mean Squared:** Minimize the square root of residual mean squared value
    c. **Model F-Value:** Maximize the model's F-Statistic

Outputs

The options in this panel specify the plots which should be generated and files which should be saved.

1. **Display Performance Parameters:** Performance parameters and results for all performed actions will be output to the Command Window in Matlab.

Plot Outputs

2. **Plot Residuals:** Plot residuals normalized by load capacity vs datapoint number.
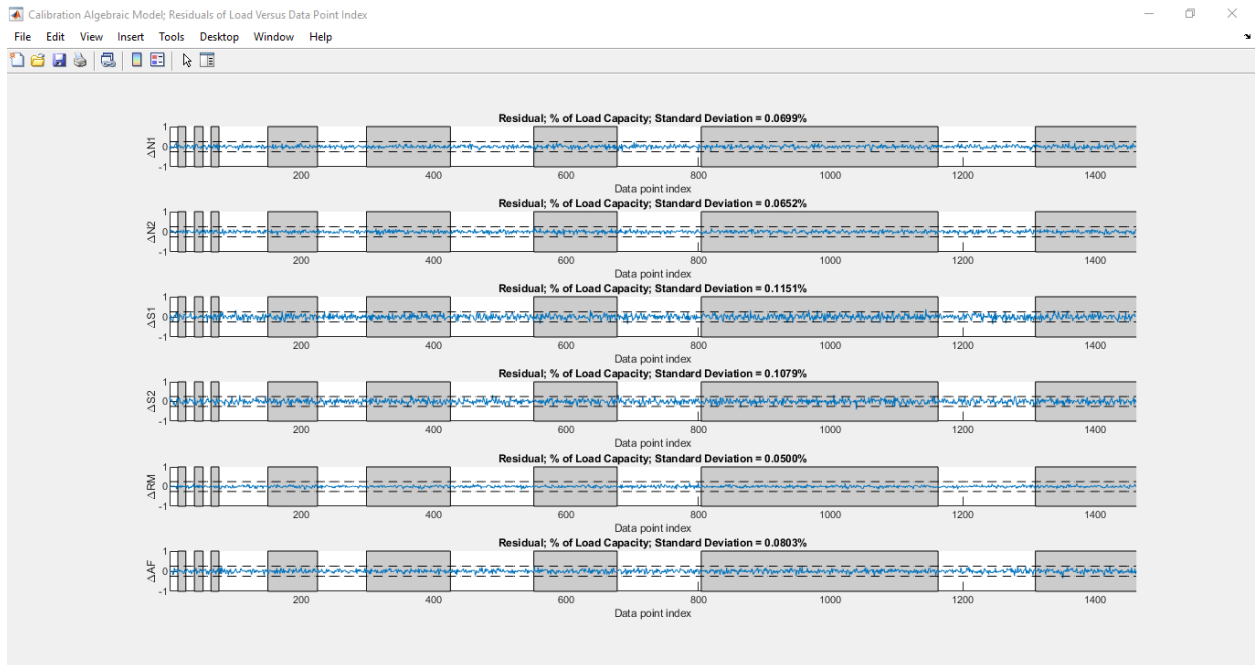
3. **Plot Residual Histograms:** Generate histograms for residual distribution. The histogram can be used to evaluate the assumption that the residuals are normally distributed. An overlay on the histogram provides the standard normal distribution. For calibration residuals, titles above each histogram provide the results for a Shapiro-Wilk test as a confidence level that the residuals are non-normally distributed [5]. A level of confidence above 90% indicates that the normality assumption may be faulty.
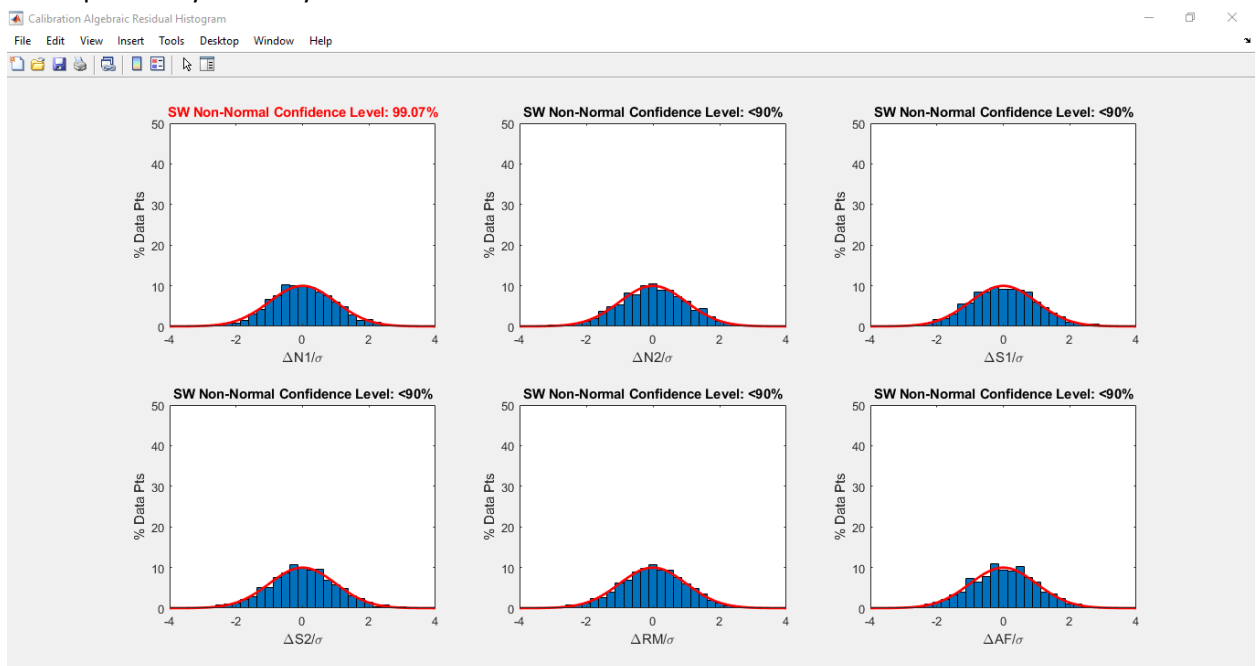
4. **Plot Residual Q-Q Plot:** Generate a Quartile-Quartile (Q-Q) plot for the residuals. This plots the actual residual quartiles vs the standard normal quartiles. If the residuals are normally distributed, the Q-Q plot should be linear. Deviations from linearity indicate the residuals may not be normally distributed. For calibration residuals, titles above each histogram provide the results for a Shapiro-Wilk test as a confidence level that the residuals are non-normally distributed [5]. A level of confidence above 90% indicates that the assumption may be faulty.
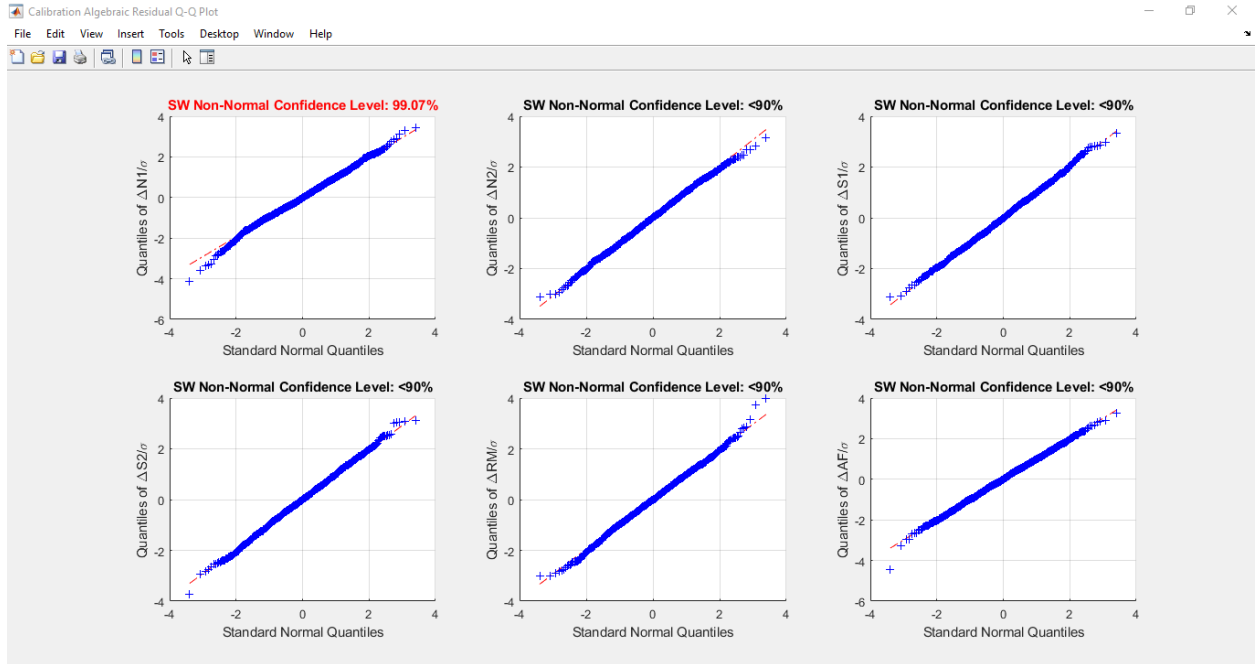
5. **Plot Correlations:** Generate a scatter plot for each of the load components vs each of the gage outputs. This is useful for evaluating if there is obvious correlation between any of the predictor and response variables.
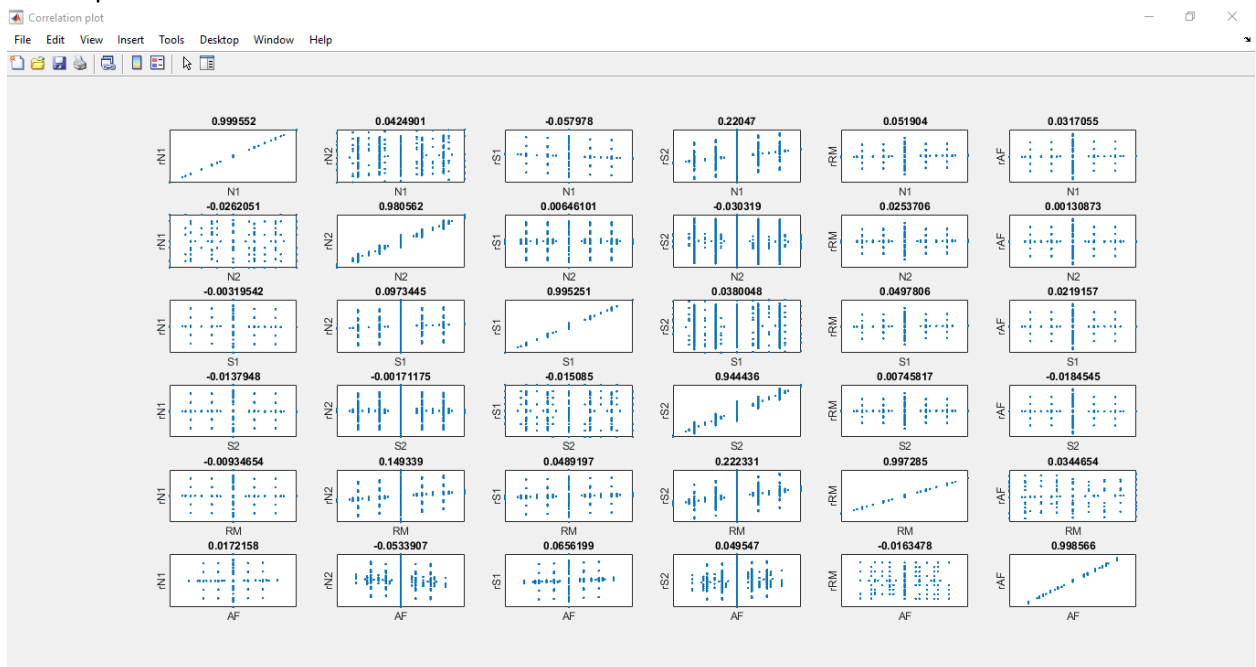
6. **Plot Residual Correlations:** Generate a scatter plot for the load residuals vs gage outputs. The residual distributions should generally be uniform and random. Noticeable patterns in the distribution could indicate a deficiency in the model.
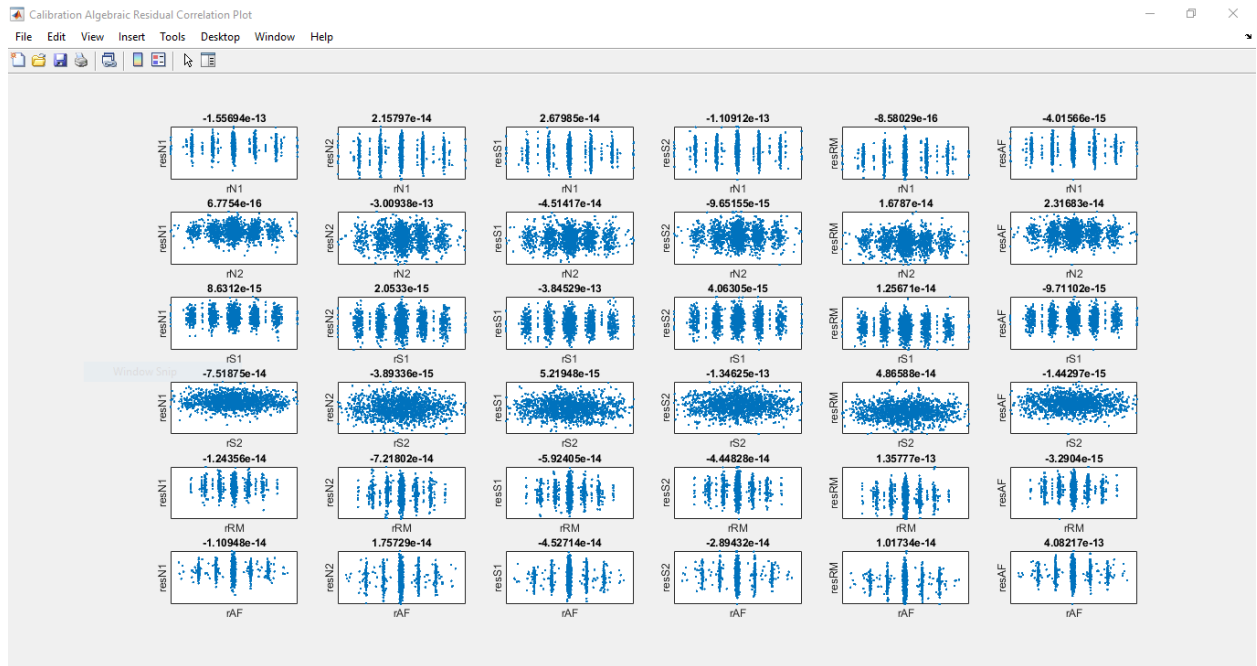
File Outputs:

7. **File Output Location:** Specify the path to save output files in the computer.
8. **Output to Calibration File Location:** Selecting this option deactivates option 7 and saves output files to the location of the calibration input file.
9. **Create Run Subfolder:** Selecting this option creates a new folder labeled with the run time in the output location and saves all files in this folder.
10. **Save Calibration Model .mat File:** Selecting this model generates a .mat file containing the calibration model. It may be input into the program AOX_Approx so as to apply the model to new data.
11. **Save Input Data to Output Location:** Selecting this option will save .cal, .val, and .app files which can be input in AOX_BalCal to quickly perform additional runs using the same calibration, validation, and approximation data, respectively.
12. **Print Performance Parameter xlsx Files:** Selecting this option will save performance parameters for calibration and validation to .xlsx files.
13. **Print Load and Coefficient csv Files:** Selecting this option saves the load approximations and coefficients to the mathematical model in .csv files
14. **Print BALFIT Coefficient Matrix txt File:** Selecting this option generates .txt files for the model coefficients for input into BALFIT's non-iterative mode. Based on BALFIT's input requirements for the non-iterative approach, each target component load requires an individual .txt file.
15. **Print Recommended Alg Model csv File:** Selecting this option outputs a .csv file with an array of 1's and 0's for which polynomial terms are statistically significant, based on the ANOVA calculations. It may be used as the **Custom Eqn File** input in future runs. ANOVA must be performed for this option.

16. **Print Load w/ Prediction Interval xlsx File:** Selecting this option applies the calibration results from ANOVA to calculate prediction intervals for validation and approximation loads. The results are saved to a .xlsx file. The prediction intervals are based on the level of confidence specified for **ANOVA % Confidence** in the Model Options panel.

Other Buttons

The following buttons are provided in the bottom right corner the GUI:

1. **Load Settings:** Load the GUI settings for AOX_BalCal from a .ini file. If the input data is saved using the **Save Input Data to Output Location** option in the Outputs panel, the settings for the run are also saved in a .ini file. Loading this file allows the setting used in the run to be quickly reused.
2. **Set as Default:** Save the current GUI settings to the file default.ini. This file will be loaded as the default settings during the next run of AOX_BalCal.
3. **Run:** Begin run. The GUI will close automatically after selecting **Run.** Progress of AOX_BalCal can be monitored in MATLAB's Command Window.
4. **Cancel:** Close the GUI without performing analysis.

General Function Approximation GUI Differences:

Largely, the GUI options remain the same when switching from Balance Calibration to General Function Approximation mode. Differences are as follows and labeled in the GUI figure:
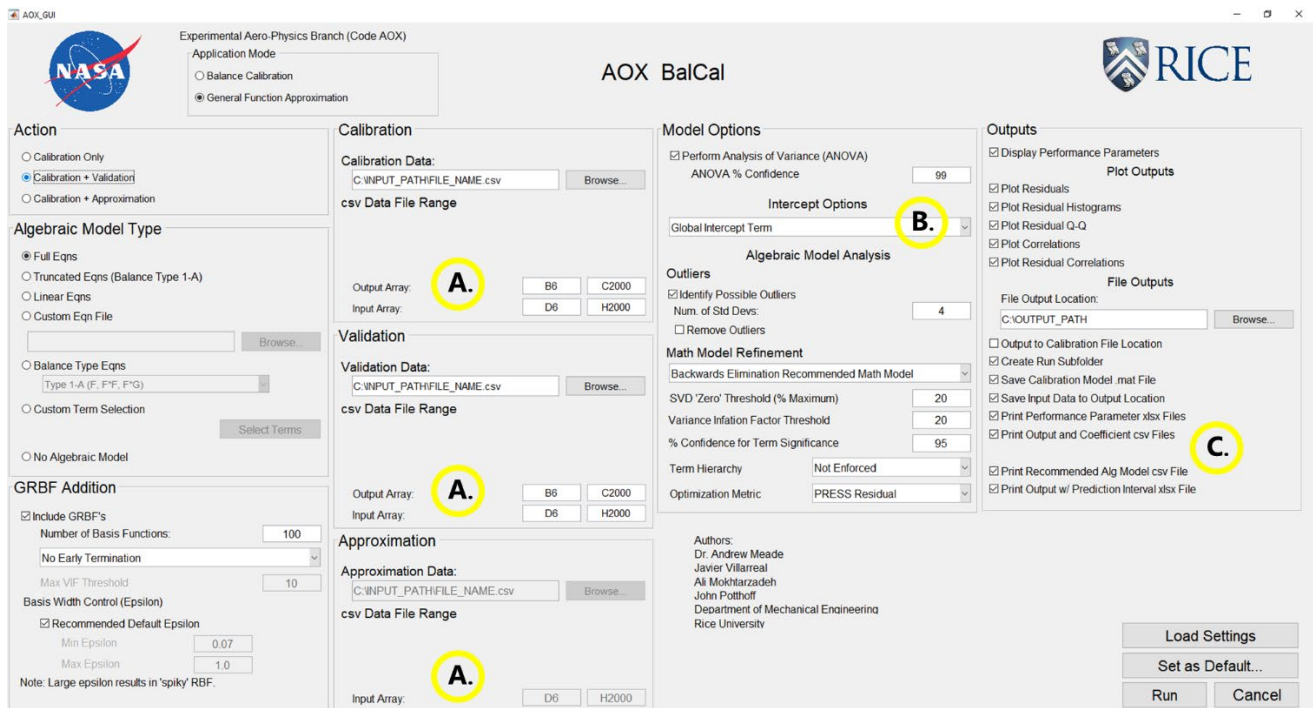
A. .csv File Input Ranges: When inputting Calibration, Validation, and Approximation data, the ranges which must be specified are reduced to **Output Array** and **Input Array**. **Output Array** is the range for output (response) variables and **Input Array** is the range for input (predictor) variables. There are no constraints on the number of input and output variables. In order to generate all possible polynomial interaction terms, however, at least 3 input variables must be

provided. The following figures show the Calibration, Validation, and Approximation panels in general function approximation mode. The corresponding example input .csv file is also provided with the ranges highlighted and labeled.

1. **Output Array:** Array of target output variables. Columns in the array denote the output variable and rows contain the individual datapoints. Any number of output variables may be provided.
2. **Input Array:** Array of input variables. Columns in the array denote the input variables and rows contain the individual datapoints. Number of columns may differ between the **Output Array** and **Input Array.** Any number of input variables may be provided. In order to generate all possible polynomial interaction terms, however, at least 3 input variables must be provided.

   Note: Output variables are not known when performing approximation, so the **Output Array** field is not included in the Approximation Panel. The input data file requirements are consistent between the Approximation action in AOX_BalCal and the standalone program AOX_Approx.

B. Intercept Options: Options for intercepts are reduced to **Global Intercept Term** and **No Intercept Term.**

C. Outputs: **Print BALFIT Coefficient Matrix .txt File** is removed

Balance Calibration Input File Considerations

Initial inputs for calibration, validation, and approximation data in AOX_BalCal are made using .csv files. Following the initial test on a dataset, the generated .cal, .val, and .app files may be input to quickly run additional tests on the data without entering the .csv file ranges. AOX_BalCal was designed to accept inputs from .csv files without stringent formatting requirements. Special attention was paid to reading data from BALFIT input files of calibration data. Data can also be input from .csv files with differing formats. However, several important considerations are provided in the following section when inputting balance calibration data from .csv files. For each of the following considerations, images are provided from an example .csv file. The area of consideration is outlined in red.

1. **Semicolons in Column A:** Any rows with a semicolon in Column A will be ignored when reading the datapoint sets.

2. **Point ID Labels:** Point ID labels are expected in the column immediately left of the Series 1 range. They may be a combination of characters and digits. Point IDs are used to label data points in program outputs. If they cannot be read from this location, AOX_BalCal generates labels automatically in the format "P-#".

; data for analysis (series 1 = load series number, series 2 = natural zero series number)

Point ID	Series1	Series2	N1	N2	S1	S2	RM	AF	rN1	rN2	rS1	rS2	rRM	rAF
P-0001	1	A	0	0	0	0	0	0	8.639711	33.24823	-43.1855	-6.32391	2.464813	-11.4089
P-0002	1	A	-833.333	0	0	0	0	0	-344.776	7.293228	-27.7253	-59.7949	-21.6158	-15.8787
P-0003	1	A	-1666.67	0	0	0	0	0	-695.528	-18.6692	-11.7947	-114.249	-45.1115	-21.4025
P-0004	1	A	-2500	0	0	0	0	0	-1046.99	-45.1973	5.008331	-168.529	-68.0345	-26.9587
P-0005	1	A	-2000	0	0	0	0	0	-836.528	-28.8294	-4.12248	-135.635	-53.6827	-23.5238
P-0006	1	A	-1500	0	0	0	0	0	-624.886	-13.8094	-14.6101	-102.033	-40.0521	-19.3097
P-0007	1	A	-1000	0	0	0	0	0	-413.842	1.084941	-24.3649	-70.2968	-27.0888	-17.7143
P-0008	1	A	-500	0	0	0	0	0	-203.303	17.3659	-32.4484	-37.4387	-11.4749	-12.4635
P-0009	1	A	0	0	0	0	0	0	8.375373	32.68348	-43.6015	-4.06358	1.326037	-9.81751
P-0010	1	A	500	0	0	0	0	0	220.449	47.87595	-54.061	28.54437	15.28988	-6.43136
P-0011	1	A	1000	0	0	0	0	0	434.502	63.33111	-64.3322	61.41572	25.31389	-4.87485
P-0012	1	A	1500	0	0	0	0	0	645.9571	77.93485	-74.6006	95.34536	36.94624	-0.0721

3. **Series 1:** The Series 1 entries must be sequential positive integers.
4. **Number of points in Series 1:** AOX_BalCal uses series specific intercept terms in the set of predictor variables to estimate tare loads. Using this technique, it can provide estimates for the tare loads even if only one data point is provided in a series. If a single datapoint is provided in a series, however, the PRESS Statistic results will be poor as the series specific intercept is calculated from an individual datapoint.

; data for analysis (series 1 = load series number, series 2 = natural zero series number)

Point ID	Series1	Series2	N1	N2	S1	S2	RM	AF	rN1	rN2	rS1	rS2	rRM	rAF
P-0001	1	A	0	0	0	0	0	0	8.639711	33.24823	-43.1855	-6.32391	2.464813	-11.4089
P-0002	1	A	-833.333	0	0	0	0	0	-344.776	7.293228	-27.7253	-59.7949	-21.6158	-15.8787
P-0003	1	A	-1666.67	0	0	0	0	0	-695.528	-18.6692	-11.7947	-114.249	-45.1115	-21.4025
P-0004	1	A	-2500	0	0	0	0	0	-1046.99	-45.1973	5.008331	-168.529	-68.0345	-26.9587
P-0005	1	A	-2000	0	0	0	0	0	-836.528	-28.8294	-4.12248	-135.635	-53.6827	-23.5238
P-0006	1	A	-1500	0	0	0	0	0	-624.886	-13.8094	-14.6101	-102.033	-40.0521	-19.3097
P-0007	1	A	-1000	0	0	0	0	0	-413.842	1.084941	-24.3649	-70.2968	-27.0888	-17.7143
P-0008	1	A	-500	0	0	0	0	0	-203.303	17.3659	-32.4484	-37.4387	-11.4749	-12.4635
P-0009	1	A	0	0	0	0	0	0	8.375373	32.68348	-43.6015	-4.06358	1.326037	-9.81751
P-0010	1	A	500	0	0	0	0	0	220.449	47.87595	-54.061	28.54437	15.28988	-6.43136
P-0011	1	A	1000	0	0	0	0	0	434.502	63.33111	-64.3322	61.41572	25.31389	-4.87485
P-0012	1	A	1500	0	0	0	0	0	645.9571	77.93485	-74.6006	95.34536	36.94624	-0.0721
P-0013	1	A	2000	0	0	0	0	0	858.8919	93.34053	-84.1985	129.4061	48.98434	3.221293
P-0014	1	A	2500	0	0	0	0	0	1071.953	106.811	-95.0282	160.7433	60.87924	8.104821
P-0015	1	A	1666.667	0	0	0	0	0	715.6591	81.89128	-78.7765	106.2657	41.44099	0.855796
P-0016	1	A	833.3333	0	0	0	0	0	363.927	57.60503	-61.4251	49.66868	23.22982	-5.07797
P-0017	1	A	0	0	0	0	0	0	8.397296	33.79851	-43.2887	-6.04736	0.301228	-10.0835
P-0018	2	A	0	0	0	0	0	0	7.24799	33.63431	-42.646	-5.75415	0.156717	-12.0042
P-0019	2	A	0	-833.333	0	0	0	0	12.3771	-333.774	-43.4993	0.836882	-11.1848	-13.8314
P-0020	2	A	0	-1666.67	0	0	0	0	13.90316	-700.529	-43.6638	5.698616	-22.0654	-14.4621
P-0021	2	A	0	-2500	0	0	0	0	18.71299	-1067.6	-46.9833	12.13614	-34.1189	-15.5182

5. **Series 2:** The Series 2 entries are currently not used in AOX_BalCal. They may be numerical or character entries without issue.

; data for analysis (series 1 = load series number, series 2 = natural zero series number)

Point ID	Series1	Series2	N1	N2	S1	S2	RM	AF	rN1	rN2	rS1	rS2	rRM	rAF
P-0001	1	A	0	0	0	0	0	0	8.639711	33.24823	-43.1855	-6.32391	2.464813	-11.4089
P-0002	1	A	-833.333	0	0	0	0	0	-344.776	7.293228	-27.7253	-59.7949	-21.6158	-15.8787
P-0003	1	A	-1666.67	0	0	0	0	0	-695.528	-18.6692	-11.7947	-114.249	-45.1115	-21.4025
P-0004	1	A	-2500	0	0	0	0	0	-1046.99	-45.1973	5.008331	-168.529	-68.0345	-26.9587
P-0005	1	A	-2000	0	0	0	0	0	-836.528	-28.8294	-4.12248	-135.635	-53.6827	-23.5238
P-0006	1	A	-1500	0	0	0	0	0	-624.886	-13.8094	-14.6101	-102.033	-40.0521	-19.3097
P-0007	1	A	-1000	0	0	0	0	0	-413.842	1.084941	-24.3649	-70.2968	-27.0888	-17.7143
P-0008	1	A	-500	0	0	0	0	0	-203.303	17.3659	-32.4484	-37.4387	-11.4749	-12.4635
P-0009	1	A	0	0	0	0	0	0	8.375373	32.68348	-43.6015	-4.06358	1.326037	-9.81751
P-0010	1	A	500	0	0	0	0	0	220.449	47.87595	-54.061	28.54437	15.28988	-6.43136
P-0011	1	A	1000	0	0	0	0	0	434.502	63.33111	-64.3322	61.41572	25.31389	-4.87485
P-0012	1	A	1500	0	0	0	0	0	645.9571	77.93485	-74.6006	95.34536	36.94624	-0.0721

6. **Natural Zeros:** The range entered for natural zeros should contain columns for each of the gage outputs provided in the Voltage Array and rows for each roll orientation recorded. AOX_BalCal averages across the rows to estimate the true natural zero voltages. If an uneven number of readings are recorded at different orientations, the gage outputs should be averaged at each orientation and those averages should be provided in the natural zero range to avoid weighting some orientations too heavily in AOX_BalCal. If the gage outputs in the data file are already gage output differences, select an array of zeros with columns for each column in the Voltage Array and one row as the Natural Zeros range.

9	; natural zeros											
10		NZ-01	A	0					4.325866	26.3029	-45.4492	-6.93579	-0.93436	-11.7178
11		NZ-01	A	90					14.49802	32.04156	-38.5671	-4.54863	-0.19286	-11.0523
12		NZ-01	A	180					29.90688	31.50237	-51.7487	-2.20118	0.991151	-10.8812
13		NZ-01	A	270					9.669238	31.75316	-39.435	-3.51441	-0.66394	-11.1486

7. **Load and Voltage Arrays:** The ranges for the load and voltage arrays must be uninterrupted ranges of numerical entries. Each column provides a variable and each row contains an individual datapoint. The number of gage outputs (predictor variables) may be greater than or equal to the number of component loads (target variables). For n component loads, the first n gage output columns are expected to be the gage outputs for each component load channel. Additional predictor variables such as temperatures and bellows pressures for air balances may be included to the right of these first n columns.

15	; data for analysis (series 1 = load series number, series 2 = natural zero series number)														
16	;	Point ID	Series1	Series2	N1	N2	S1	S2	RM	AF	rN1	rN2	rS1	rS2	rRM	rAF
17		P-0001	1	A	0	0	0	0	0	0	8.639711	33.24823	-43.1855	-6.32391	2.464813	-11.4089
18		P-0002	1	A	-833.333	0	0	0	0	0	-344.776	7.293228	-27.7253	-59.7949	-21.6158	-15.8787
19		P-0003	1	A	-1666.67	0	0	0	0	0	-695.528	-18.6692	-11.7947	-114.249	-45.1115	-21.4025
20		P-0004	1	A	-2500	0	0	0	0	0	-1046.99	-45.1973	5.008331	-168.529	-68.0345	-26.9587
21		P-0005	1	A	-2000	0	0	0	0	0	-836.528	-28.8294	-4.12248	-135.635	-53.6827	-23.5238
22	;	P-0006	1	A	-1500	0	0	0	0	0	-624.886	-13.8094	-14.6101	-102.033	-40.0521	-19.3097
23	;	P-0007	1	A	-1000	0	0	0	0	0	-413.842	1.084941	-24.3649	-70.2968	-27.0888	-17.7143
24		P-0008	1	A	-500	0	0	0	0	0	-203.303	17.3659	-32.4484	-37.4387	-11.4749	-12.4635
25		P-0009	1	A	0	0	0	0	0	0	8.375373	32.68348	-43.6015	-4.06358	1.326037	-9.81751
26		P-0010	1	A	500	0	0	0	0	0	220.449	47.87595	-54.061	28.54437	15.28988	-6.43136
27		P-0011	1	A	1000	0	0	0	0	0	434.502	63.33111	-64.3322	61.41572	25.31389	-4.87485
28		P-0012	1	A	1500	0	0	0	0	0	645.9571	77.93485	-74.6006	95.34536	36.94624	-0.0721
29		P-0013	1	A	2000	0	0	0	0	0	858.8919	93.34053	-84.1985	129.4061	48.98434	3.221293

8. **Load and Gage Output Labels:** Labels for the component loads and gage are expected in the row above the specified ranges for the Load Array and Voltage Array, respectively. If labels cannot be read from this location, default labels are used.

15	; data for analysis (series 1 = load series number, series 2 = natural zero series number)														
16	;	Point ID	Series1	Series2	N1	N2	S1	S2	RM	AF	rN1	rN2	rS1	rS2	rRM	rAF
17		P-0001	1	A	0	0	0	0	0	0	8.639711	33.24823	-43.1855	-6.32391	2.464813	-11.4089
18		P-0002	1	A	-833.333	0	0	0	0	0	-344.776	7.293228	-27.7253	-59.7949	-21.6158	-15.8787
19		P-0003	1	A	-1666.67	0	0	0	0	0	-695.528	-18.6692	-11.7947	-114.249	-45.1115	-21.4025
20		P-0004	1	A	-2500	0	0	0	0	0	-1046.99	-45.1973	5.008331	-168.529	-68.0345	-26.9587
21		P-0005	1	A	-2000	0	0	0	0	0	-836.528	-28.8294	-4.12248	-135.635	-53.6827	-23.5238
22	;	P-0006	1	A	-1500	0	0	0	0	0	-624.886	-13.8094	-14.6101	-102.033	-40.0521	-19.3097
23	;	P-0007	1	A	-1000	0	0	0	0	0	-413.842	1.084941	-24.3649	-70.2968	-27.0888	-17.7143
24		P-0008	1	A	-500	0	0	0	0	0	-203.303	17.3659	-32.4484	-37.4387	-11.4749	-12.4635
25		P-0009	1	A	0	0	0	0	0	0	8.375373	32.68348	-43.6015	-4.06358	1.326037	-9.81751
26		P-0010	1	A	500	0	0	0	0	0	220.449	47.87595	-54.061	28.54437	15.28988	-6.43136
27		P-0011	1	A	1000	0	0	0	0	0	434.502	63.33111	-64.3322	61.41572	25.31389	-4.87485

9. **Load and Gage Output Units:** AOX_BalCal searches for row containing the string "units". In the following row, the load and gage output units are read from the columns provided for the load and voltage arrays. If units cannot be read from this location, default units are used.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	BALANCE_NAME	MC60-E (5F1M)														
2	DESCRIPTION	Artificial data set for tare calculation validation														
3	; load and output symbols															
4					N1	N2	S1	S2	RM	AF	rN1	rN2	rS1	rS2	rRM	rAF
5	; units															
6					lbs	lbs	lbs	lbs	in-lbs	lbs	microV/V	microV/V	microV/V	microV/V	microV/V	microV/V
7	; capacities															
8					2500	2500	1250	1250	5000	700	785	850	1128	1160	1300	1655

10. **Load Capacities:** The load capacity range should contain one column for each of the columns in the Load Array and a single row.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	BALANCE_NAME	MC60-E (5F1M)														
2	DESCRIPTION	Artificial data set for tare calculation validation														
3	; load and output symbols															
4					N1	N2	S1	S2	RM	AF	rN1	rN2	rS1	rS2	rRM	rAF
5	; units															
6					lbs	lbs	lbs	lbs	in-lbs	lbs	microV/V	microV/V	microV/V	microV/V	microV/V	microV/V
7	; capacities															
8					2500	2500	1250	1250	5000	700	785	850	1128	1160	1300	1655

11. **Gage Output Capacities:** Gage output capacities are expected in the same row as load capacities in the columns provided for the Voltage Array. If the gage outputs cannot be read from this location, the maximum magnitude of raw gage outputs will be set as the capacity. Gage capacities are needed in determining the zero threshold when calculating the permitted equation.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	BALANCE_NAME	MC60-E (5F1M)														
2	DESCRIPTION	Artificial data set for tare calculation validation														
3	; load and output symbols															
4					N1	N2	S1	S2	RM	AF	rN1	rN2	rS1	rS2	rRM	rAF
5	; units															
6					lbs	lbs	lbs	lbs	in-lbs	lbs	microV/V	microV/V	microV/V	microV/V	microV/V	microV/V
7	; capacities															
8					2500	2500	1250	1250	5000	700	785	850	1128	1160	1300	1655

12. **File Description:** A description of the dataset may be included in input .csv file for inclusion in output files from AOX_BalCal. AOX_BalCal searches for a cell containing exactly "DESCRIPTION". The cell immediately to the right of this cell is read and saved as the file description.

	A	B	C	D	E	F	G	H	I
1	BALANCE_NAME	MC60-E (5F1M)							
2	DESCRIPTION	Artificial data set for tare calculation validation							
3	; load and output symbols								
4					N1	N2	S1	S2	RM

13. **Balance Name:** The name of the internal balance may be included in input .csv file for inclusion in output files from AOX_BalCal. AOX_BalCal searches for a cell containing exactly "BALANCE_NAME". The cell immediately to the right of this cell is read and saved as the balance name.

	A	B	C	D	E	F	G	H	I
1	BALANCE_NAME	MC60-E (5F1M)							
2	DESCRIPTION	Artificial data set for tare calculation validation							
3	; load and output symbols								
4					N1	N2	S1	S2	RM

Balance Calibration Mode File Outputs

The following section provides a list of all possible file outputs from AOX_BalCal and their contents. All files are saved to the **File Output Location** specified in the Outputs GUI panel.
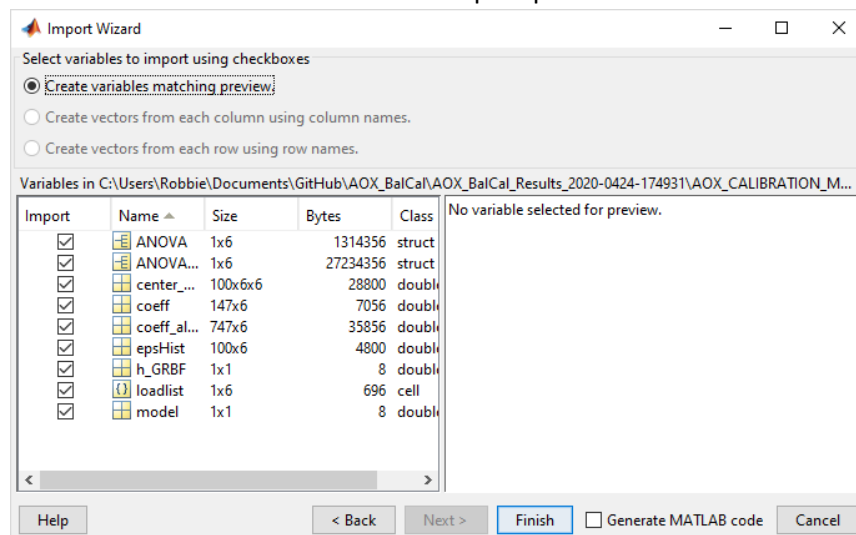
Algebraic and GRBF Calibration Outputs

1. **CALIB Report.xlsx:** Summary of calibration results and evaluation metrics for algebraic and GRBF sections (if run). Results from algebraic and GRBF sections are output in separate tabs. Output using **Print Performance Parameter xlsx Files** button in Outputs panel.

	A	B	C
1	Calibration Algebraic Results		
2	Software Mode: FORCE BALANCE CALIBRATION		
3	REPORT NO: 2020-0424-174931		
4	Calibration Input File: MC60E-LargeArtificial-2018_cal_rand.cal		
5	Calibration ALG Outliers Flagged: FALSE		
6	Calibration ALG Outliers Removed: FALSE		
7	Algebraic Model Used: FULL		
8	Number of Datapoints: 1462		
9	GRBF Addition Performed: TRUE		
10	Number GRBFs: 100		
11			
12	Percent Load Capacity of Residual Standard Deviation		
13		N1	N2
14		0.069851317	0.06516859
15			
16	Tares		
17	Series	N1	N2
18	1	-16.61289983	7.018648659
19	2	-17.34601544	6.666569595
20	3	-19.84032395	7.404288634
21	4	-13.20639482	6.203922568
22	5	-15.66717603	6.952842174
23	6	-5.793519653	1.519326209
24	7	-16.09169029	7.327652147
25	8	-16.21210997	5.986666582
26	9	-17.69304086	6.294944093
27	10	-8.862080544	4.031264168
28	11	-16.40678732	5.536978943
29	12	-17.64685271	6.198045548

Tabs: **Algebraic Results** | GRBF Results

	A	B	C
1	Calibration GRBF Results		
2	Software Mode: FORCE BALANCE CALIBRATION		
3	REPORT NO: 2020-0424-174931		
4	Calibration Input File: MC60E-LargeArtificial-2018_cal_rand.cal		
5	Calibration ALG Outliers Flagged: FALSE		
6	Calibration ALG Outliers Removed: FALSE		
7	Algebraic Model Used: FULL		
8	Number of Datapoints: 1462		
9	GRBF Addition Performed: TRUE		
10	Number GRBFs: 100		
11			
12	Percent Load Capacity of Residual Standard Deviation		
13		N1	N2
14		0.062332617	0.060759956
15			
16	Tares		
17	Series	N1	N2
18	1	-15.82558903	7.571711354
19	2	-16.77834271	6.683157526
20	3	-19.67930541	7.59291558
21	4	-13.2906874	6.321962895
22	5	-16.07768426	6.983816611
23	6	-5.504780365	1.284343585
24	7	-16.0907906	7.751827475
25	8	-16.18194432	6.144488064
26	9	-18.58590639	6.706734238
27	10	-8.254220642	3.641365821
28	11	-16.17890282	5.423880182
29	12	-18.22781053	6.582460456

Tabs: Algebraic Results | **GRBF Results**

2. **AOX_CALIBRATION_MODEL_*TIMESTAMP*.mat:** File containing everything from calibration model (including GRBFs) needed to run in the standalone approximation program. Output using **Save Calibration Model .mat File** button in the Outputs panel.

Import Wizard

Select variables to import using checkboxes
- (•) Create variables matching preview.
- ( ) Create vectors from each column using column names.
- ( ) Create vectors from each row using row names.

Variables in C:\Users\Robbie\Documents\GitHub\AOX_BalCal\AOX_BalCal_Results_2020-0424-174931\AOX_CALIBRATION_M...

Import	Name ▲	Size	Bytes	Class
☑	ANOVA	1x6	1314356	struct
☑	ANOVA...	1x6	27234356	struct
☑	center_...	100x6x6	28800	doubl
☑	coeff	147x6	7056	doubl
☑	coeff_al...	747x6	35856	doubl
☑	epsHist	100x6	4800	doubl
☑	h_GRBF	1x1	8	doubl
☑	loadlist	1x6	696	cell
☑	model	1x1	8	doubl

No variable selected for preview.

Help   < Back   Next >   Finish   ☐ Generate MATLAB code   Cancel

Calibration Algebraic Section Outputs

1. **AOX_ALG_MODEL_COEFFICIENT_MATRIX.csv:** CSV file of algebraic coefficients for model. Output using **Print Load and Coefficient csv Files** button in the Outputs panel.
2. *LOADLABEL*_**BALFIT_REGRESSION_COEFFICIENT_MATRIX_IN_AMES_FORMAT.txt:** Regression coefficient matrix for input into BALFIT non-iterative approach. A separate .txt file is generated for each component load. Output using **Print BALFIT Coefficient Matrix txt File** button in Outputs panel.
3. **CALIB ALG ANOVA STATS.xlsx:** ANOVA results for algebraic calibration model. Output if **Perform Analysis of Variance (ANOVA)** is selected in Model Options panel.
4. **CALIB ALG Tare Corrected Load Approximation.csv:** Load approximation from calibration input voltages using algebraic model with tare loads subtracted. Output using **Print Load and Coefficient csv Files** button in the Outputs panel.
5. **DIRECT_RECOMM_CustomEquationMatrix.csv:** Custom equation matrix of statistically significant predictor terms based on ANOVA. This file is a useful starting point for generating a user built custom equation matrix for input into AOX_BalCal. Output using **Print Recommended Alg Model csv File** button in Outputs panel.

Calibration GRBF Section Outputs

1. **AOX_ALG-GRBF_MODEL_COEFFICIENT_MATRIX.csv:** CSV file of algebraic and GRBF coefficients for model. Output using **Print Load and Coefficient csv Files** button in the Outputs panel.
2. **AOX_GRBF_Centers.csv:** Index of calibration datapoints where GRBFs were placed. Output using **Print Load and Coefficient csv Files** button in the Outputs panel.
3. **AOX_GRBF_Coefficients.csv:** Coefficients used for GRBFs placed. Output using **Print Load and Coefficient csv Files** button in the Outputs panel.
4. **AOX_GRBF_Epsilon.csv:** Epsilon values (width parameters) for GRBFs placed. Output using **Print Load and Coefficient csv Files** button in the Outputs panel.
5. **AOX_GRBF_h.csv:** h value for GRBFs placed (single h for all GRBFs). Output using **Print Load and Coefficient csv Files** button in the Outputs panel.
6. **CALIB GRBF ANOVA STATS.xlsx:** ANOVA results for calibration model including GRBFs. Output if **Perform Analysis of Variance (ANOVA)** is selected in Model Options panel.
7. **CALIB GRBF Tare Corrected Load Approximation.csv:** Load approximation from calibration input voltages using algebraic model with GRBFs added with tare loads subtracted. Output using **Print Load and Coefficient csv Files** button in the Outputs panel.

Algebraic and GRBF Validation Outputs

1. **VALID Report.xlsx:** Summary of validation results and evaluation metrics for algebraic and GRBF sections (if run). Results from algebraic and GRBF sections are output in separate tabs. Output using **Print Performance Parameter xlsx Files** button in Outputs panel.

Validation Algebraic Section Outputs

1. **VALID ALG Tare Corrected Load Approximation.csv:** Load approximation from validation input voltages using algebraic model with tare loads subtracted. Output using **Print Load and Coefficient csv Files** button in the Outputs panel.
2. **VALID ALG Tare Corrected Load Approximation w PI.xlsx:** Load approximation from validation input voltages with tares subtracted on Tab 1. Tab 2 contains the approximation +/- the

prediction interval. Tab 3 contains the prediction interval values. Output using **Print Load w/ Prediction Interval xlsx File** button in Outputs panel.

Validation GRBF Section Outputs

1. **VALID GRBF Tare Corrected Load Approximation.csv:** Load approximation from validation input voltages using algebraic model with GRBFs added with tare loads subtracted. Output using **Print Load and Coefficient csv Files** button in the Outputs panel.
2. **VALID GRBF Tare Corrected Load Approximation w PI.xlsx:** Load approximation from validation input voltages with tares subtracted on Tab 1.   Tab 2 contains the approximation +/- the prediction interval. Tab 3 contains the prediction interval values. Output using **Print Load w/ Prediction Interval xlsx File** button in Outputs panel.

Approximation Algebraic Section Outputs

1. **APPROX ALG Global Load Approximation.csv:** Global Load approximation from approximation input voltages using algebraic model (No tare correction). Output using **Print Load and Coefficient csv Files** button in the Outputs panel.
2. **APPROX ALG Global Load Approximation w PI.xlsx:** Global Load approximation from approximation input voltages on Tab 1 (No tare correction).   Tab 2 contains the approximation +/- the prediction interval. Tab 3 contains the prediction interval values. Output using **Print Load w/ Prediction Interval xlsx File** button in Outputs panel.

Approximation GRBF Section Outputs

1. **APPROX GRBF Global Load Approximation.csv:** Global Load approximation from approximation input voltages using algebraic model with GRBFs added (No tare correction). Output using **Print Load and Coefficient csv Files** button in the Outputs panel.
2. **APPROX GRBF Global Load Approximation w PI.xlsx:** Global Load approximation from approximation input voltages on Tab 1 (No tare correction).   Tab 2 contains the approximation +/- the prediction interval. Tab 3 contains the prediction interval values. Output using **Print Load w/ Prediction Interval xlsx File** button in Outputs panel.

General Function Input File Considerations

Initial inputs for calibration, validation, and approximation data in AOX_BalCal are made using .csv files. Following the initial test on a dataset, the generated .cal, .val, and .app files may be input to quickly run additional tests on the data without entering the .csv file ranges. AOX_BalCal was designed to accept inputs from .csv files without stringent formatting requirements. However, several important considerations are provided in the following section when inputting data from .csv files for general function approximation using AOX_BalCal. If the scales of measurement are drastically different between input variables, it is recommended that the input array be normalized to a consistent range between inputs. This will prevent a single input variable from dominating the distance measurements during GRBF addition. For each of the following considerations, images are provided from the example .csv file provided above. The area of consideration is outlined in red.

1. **Semicolons in Column A:** Any rows with a semicolon in Column A will be ignored when reading the datapoint sets.

	A	B	C	D	E	F	G	H	I
1	DESCRIPTION	Example dataset for general function approximation							
2	; units								
3		lbs	lbs	microV/V	microV/V	microV/V	microV/V	microV/V	
4	; data for analysis								
5	;	Out1	Out2	In1	In2	In3	In4	In5	
6		0	0	8.639711	33.24823	-43.1855	-6.32391	2.464813	
7		-833.333	0	-344.776	7.293228	-27.7253	-59.7949	-21.6158	
8		-1666.67	0	-695.528	-18.6692	-11.7947	-114.249	-45.1115	
9		-2500	0	-1046.99	-45.1973	5.008331	-168.529	-68.0345	
10		-2000	0	-836.528	-28.8294	-4.12248	-135.635	-53.6827	
11	;	-1500	0	-624.886	-13.8094	-14.6101	-102.033	-40.0521	
12	;	-1000	0	-413.842	1.084941	-24.3649	-70.2968	-27.0888	
13		-500	0	-203.303	17.3659	-32.4484	-37.4387	-11.4749	
14		0	0	8.375373	32.68348	-43.6015	-4.06358	1.326037	

2. **Output and Input Arrays:** The ranges for the output and input arrays must be uninterrupted ranges of numerical entries. Each column provides a variable and each row contains an individual datapoint. Any number of input and output variables may be provided. In order to generate all possible polynomial interaction terms, however, at least 3 input variables must be provided. As mentioned above, if the input variables have measurements on different orders of magnitude, it is recommended that the input array is normalized to a consistent range between variables.

	A	B	C	D	E	F	G	H	I
1	DESCRIPTION	Example dataset for general function approximation							
2	; units								
3		lbs	lbs	microV/V	microV/V	microV/V	microV/V	microV/V	
4	; data for analysis								
5	;	Out1	Out2	In1	In2	In3	In4	In5	
6		0	0	8.639711	33.24823	-43.1855	-6.32391	2.464813	
7		-833.333	0	-344.776	7.293228	-27.7253	-59.7949	-21.6158	
8		-1666.67	0	-695.528	-18.6692	-11.7947	-114.249	-45.1115	
9		-2500	0	-1046.99	-45.1973	5.008331	-168.529	-68.0345	
10		-2000	0	-836.528	-28.8294	-4.12248	-135.635	-53.6827	
11	;	-1500	0	-624.886	-13.8094	-14.6101	-102.033	-40.0521	
12	;	-1000	0	-413.842	1.084941	-24.3649	-70.2968	-27.0888	
13		-500	0	-203.303	17.3659	-32.4484	-37.4387	-11.4749	
14		0	0	8.375373	32.68348	-43.6015	-4.06358	1.326037	

3. **Output and Input Variable Labels:** Labels for the output and input are expected in the row above the specified ranges for the output array and input array, respectively. If labels cannot be read from this location, default labels are used.

	A	B	C	D	E	F	G	H	I
1	DESCRIPTION	Example dataset for general function approximation							
2	; units								
3		lbs	lbs	microV/V	microV/V	microV/V	microV/V	microV/V	
4	; data for analysis								
5	;	Out1	Out2	In1	In2	In3	In4	In5	
6		0	0	8.639711	33.24823	-43.1855	-6.32391	2.464813	
7		-833.333	0	-344.776	7.293228	-27.7253	-59.7949	-21.6158	
8		-1666.67	0	-695.528	-18.6692	-11.7947	-114.249	-45.1115	
9		-2500	0	-1046.99	-45.1973	5.008331	-168.529	-68.0345	

4. **Output and Input Variable Units:** AOX_BalCal searches for row containing the string "units". In the following row, the output and input variable units are read from the columns provided for the output and input arrays. If units cannot be read from this location, no units are included in outputs.

	A	B	C	D	E	F	G	H	I
1	DESCRIPTION	Example dataset for general function approximation							
2	; units								
3		lbs	lbs	microV/V	microV/V	microV/V	microV/V	microV/V	
4	; data for analysis								
5	;	Out1	Out2	In1	In2	In3	In4	In5	
6		0	0	8.639711	33.24823	-43.1855	-6.32391	2.464813	
7		-833.333	0	-344.776	7.293228	-27.7253	-59.7949	-21.6158	
8		-1666.67	0	-695.528	-18.6692	-11.7947	-114.249	-45.1115	
9		-2500	0	-1046.99	-45.1973	5.008331	-168.529	-68.0345	

5. **File Description:** A description of the dataset may be included in input .csv file for inclusion in output files from AOX_BalCal. AOX_BalCal searches for a cell containing exactly "DESCRIPTION". The cell immediately to the right of this cell is read and saved as the file description.

	A	B	C	D	E	F	G	H	I
1	DESCRIPTION	Example dataset for general function approximation							
2	; units								
3		lbs	lbs	microV/V	microV/V	microV/V	microV/V	microV/V	
4	; data for analysis								
5	;	Out1	Out2	In1	In2	In3	In4	In5	
6		0	0	8.639711	33.24823	-43.1855	-6.32391	2.464813	
7		-833.333	0	-344.776	7.293228	-27.7253	-59.7949	-21.6158	

General Function Approximation Mode File Outputs

The following section provides a list of all possible file outputs from AOX_BalCal and their contents. All files are saved to the **File Output Location** specified in the Outputs GUI panel.

Algebraic and GRBF Calibration Outputs

1. **CALIB Report.xlsx:** Summary of calibration results and evaluation metrics for algebraic and GRBF sections (if run). Results from algebraic and GRBF sections are output in separate tabs. Output using **Print Performance Parameter xlsx Files** button in Outputs panel.
2. **AOX_CALIBRATION_MODEL_*TIMESTAMP*.mat:** File containing everything from calibration model needed to run in standalone approximation program (including GRBFs if run). Output using **Save Calibration Model .mat File** button in the Outputs panel.

Calibration Algebraic Section Outputs

1. **AOX_ALG_MODEL_COEFFICIENT_MATRIX.csv:** CSV file of algebraic coefficients for model. Output using **Print Output and Coefficient csv Files** button in the Outputs panel.
2. **CALIB ALG ANOVA STATS.xlsx:** ANOVA results for algebraic calibration model. Output if **Perform Analysis of Variance (ANOVA)** is selected in Model Options panel.
3. **CALIB ALG Output Approximation.csv:** Output variable approximation from calibration input variables using algebraic model. Output using **Print Output and Coefficient csv Files** button in the Outputs panel.
4. **DIRECT_RECOMM_CustomEquationMatrix.csv:** Custom equation matrix of statistically significant predictor terms based on ANOVA. This file is a useful starting point for generating a user built custom equation matrix for input into AOX_BalCal. Output using **Print Recommended Alg Model csv File** button in Outputs panel.

Calibration GRBF Section Outputs

1. **AOX_ALG-GRBF_MODEL_COEFFICIENT_MATRIX.csv:** CSV file of algebraic and GRBF coefficients for model. Output using **Print Output and Coefficient csv Files** button in the Outputs panel.
2. **AOX_GRBF_Centers.csv:** Index of calibration datapoints where GRBFs were placed. Output using **Print Output and Coefficient csv Files** button in the Outputs panel.
3. **AOX_GRBF_Coefficients.csv:** Coefficients used for GRBFs placed. Output using **Print Output and Coefficient csv Files** button in the Outputs panel.
4. **AOX_GRBF_Epsilon.csv:** Epsilon values (width parameters) for GRBFs placed. Output using **Print Output and Coefficient csv Files** button in the Outputs panel.
5. **AOX_GRBF_h.csv:** h value for GRBFs placed (single h for all GRBFs). Output using **Print Output and Coefficient csv Files** button in the Outputs panel.
6. **CALIB GRBF ANOVA STATS.xlsx:** ANOVA results for calibration model including GRBFs. Output if **Perform Analysis of Variance (ANOVA)** is selected in Model Options panel.
7. **CALIB GRBF Output Approximation.csv:** Output variable approximation from calibration input variables using algebraic model with GRBFs added. Output using **Print Output and Coefficient csv Files** button in the Outputs panel.

Algebraic and GRBF Validation Outputs

1. **VALID Report.xlsx:** Summary of validation results and evaluation metrics for algebraic and GRBF sections (if run). Results from algebraic and GRBF sections are output in separate tabs. Output using **Print Performance Parameter xlsx Files** button in Outputs panel.

Validation Algebraic Section Outputs

1. **VALID ALG Output Approximation.csv:** Output variable approximation from validation input variables using algebraic model. Output using **Print Output and Coefficient csv Files** button in the Outputs panel.
2. **VALID ALG Output Approximation w PI.xlsx:** Output variable approximation from validation input variables on Tab 1.   Tab 2 contains the approximation +/- the prediction interval. Tab 3 contains the prediction interval values. Output using **Print Output w/ Prediction Interval xlsx File** button in Outputs panel.

Validation GRBF Section Outputs

1. **VALID GRBF Output Approximation.csv:** Output variable approximation from validation input variables using algebraic model with GRBFs added. Output using **Print Output and Coefficient csv Files** button in the Outputs panel.
2. **VALID GRBF Output Approximation w PI.xlsx:** Output variable approximation from validation input variables on Tab 1.   Tab 2 contains the approximation +/- the prediction interval. Tab 3 contains the prediction interval values. Output using **Print Output w/ Prediction Interval xlsx File** button in Outputs panel.

Approximation Algebraic Section Outputs

1. **APPROX ALG Output Approximation.csv:** Output variable approximation from approximation input variables using algebraic model. Output using **Print Output and Coefficient csv Files** button in the Outputs panel.
2. **APPROX ALG Output Approximation w PI.xlsx:** Output variable approximation from approximation input variables on Tab 1.   Tab 2 contains the approximation +/- the prediction interval. Tab 3 contains the prediction interval values. Output using **Print Output w/ Prediction Interval xlsx File** button in Outputs panel.

Approximation GRBF Section Outputs

1. **APPROX GRBF Output Approximation.csv:** Output variable approximation from approximation input variables using algebraic model with GRBFs added. Output using **Print Output and Coefficient csv Files** button in the Outputs panel.
2. **APPROX GRBF Output Approximation w PI.xlsx:** Output variable approximation from approximation input variables on Tab 1.   Tab 2 contains the approximation +/- the prediction interval. Tab 3 contains the prediction interval values. Output using **Print Output w/ Prediction Interval xlsx File** button in Outputs panel.

Warning Messages

The following warning messages may be output to the command window while running AOX_BalCal to provide alerts to the user:

1. **Unable to read gage capacities for calibration data. Using maximum absolute value of voltage as gage capacity:** Returned if gage capacities were not found in the calibration .csv file. If calculating the permitted equation, the maximum gage output magnitudes will be used to set the zero threshold. See **SVD 'Zero' Threshold (% Capacity)** in Model Options GUI panel.

2. **Unable to read load capacities for calibration data. Using maximum absolute value of applied loads as load capacity:** Returned if load capacities could not be read in the .csv file. Load capacities are used to normalize residuals. The maximum load magnitudes will be used to normalize the residuals.

3. **Less than 2 input data dimensions. Unable to create interaction terms: F*G, |F*G|, F*|G|, |F|*G, F*G*G, F*G*H:** Returned if the voltage array (Input Array) only contains one variable. Polynomials terms for interactions between predictor variables will not be generated.

4. **Less than 3 input data dimensions. Unable to create interaction term: F*G*H:** Returned if the voltage array (Input Array) only contains two variables. Polynomials terms for interactions between three predictor variables will not be generated.

5. **Linear dependence found between predictor variable terms. Terms eliminated to ensure non-singularity:** Returned while refining the included polynomial terms to the form the permitted equation if linear dependence between included polynomial terms causes terms to be discarded for the permitted equation.

6. **Permitted Math Model is not hierarchically supported. Removing unsupported terms:** May be returned if enforcing term hierarchy while refining for a constrained or recommended equation. If the linearly independent set of polynomial terms (permitted equation) is not hierarchically supported [4], unsupported terms are removed before continuing the model refinement.

7. **VIF calculation indicates strong multicollinearity. Analysis of Variance results are unreliable:** High multicollinearity between predictor variables makes the statistical metrics calculated in the ANOVA function unreliable [2]. This warning is returned if the maximum Variance Inflation Factor (VIF) among predictor variables is greater than 10.

8. **VIF calculation indicates some multicollinearity. Analysis of Variance results may be unreliable:** High multicollinearity between predictor variables makes some of the statistical metrics calculated in the ANOVA function unreliable. This warning is returned if the maximum Variance Inflation Factor (VIF) among predictor variables is greater than 5.

9. **Input Max # GRBF > # Calibration datapoints. Setting Max # GRBF = # Calibration datapoints:** A maximum of one GRBF may be centered at each calibration datapoint. If the number of GRBFs input by the user in the GUI is greater than the number of calibration datapoints, the maximum number of GRBFs is set to the number of calibration datapoints.

10. **Ill-Conditioned matrix for load channel *LABEL,* Terminating RBF addition. Final # RBF=*NUMBER:*** As GRBFs are added in the self-assembling algorithm, all coefficients (for polynomial terms and GRBFS) in the model are resolved at each iteration. If a large number of GRBFs is set, eventually the matrix of predictor terms including GRBFs may become ill-conditioned, based on the uncertainty principle [6]. If this occurs the newest GRBF placed is

discarded and this warning message is returned. GRBF addition is halted in the affected channel to maintain numerical stability in the coefficients.

11. **Disagreement between calculated series intercept tare loads and tare load datapoints. Check results:** AOX_BalCal calculates tare loads based on all datapoints in a series using series specific intercepts, rather than single points where the gage outputs are recorded with no measured calibration loads applied to the balance (tare load datapoints). The total loads approximated by the model for these tare load datapoints should agree with the tare load estimates. This warning is returned if there is disagreement between the two estimates. If ANOVA calculations were performed, the difference between the estimates is acceptable if the tare estimate is within the prediction interval for the tare load datapoints. If ANOVA was not performed, the acceptable difference is twice the residual standard deviation. A difference between the load approximation at tare load datapoints and the tare load approximated using series specific intercepts may indicate an issue with the recorded tare load gage outputs or an issue with the model.

12. **Shapiro-Wilk test found non-normally distributed residuals for channels: *LABELS*. Hypothesis testing results and calculated intervals may be inaccurate. Recommend inspection of residual histograms and Q-Q plots:** To perform hypothesis testing and calculate intervals within the ANOVA function, a normal distribution is assumed for the residuals from the model. The Shapiro-Wilk test is applied to test this assumption [5]. If the Shapiro-Wilk test rejects the hypothesis of a normal distribution at the confidence level used in the ANOVA function, this warning is returned. It is recommended that the user also inspects the residual histograms and Q-Q plots to confirm that an issue exists with the normality assumption. Based on reviewing all these results, the user should use their judgement with assessing the reliability of the ANOVA results.

13. **Shapiro-Wilk test might be inaccurate due to large sample size (> 5000):** The Shapiro-Wilk test may be applied to datasets with between 3 and 5000 points [7]. For datasets larger than this size, the user should recognize that Shapiro-Wilk results may be inaccurate and instead rely on the residual histograms and Q-Q plots to evaluate the assumption of normality.

Standalone Approximation Program

A standalone program named 'AOX_approx' was created to apply a calibration model developed in AOX_BalCal to approximate new component loads from gage outputs. Like AOX_BalCal, AOX_approx may be run in the MATLAB environment or as a standalone program. In MATLAB, the program is run using the script 'AOX_approx.m'. The calibration model is input to AOX_approx using the file **AOX_CALIBRATION_MODEL_*TIMESTAMP*.mat,** output from AOX_BalCal using the **Save Calibration Model .mat File** button in the Outputs panel. If the .mat file provided to AOX_approx does not contain the necessary information, an error message will be displayed and the 'Run' button will be disabled. In a manner identical to the approximation section of the AOX_BalCal GUI, new gage outputs are provided through an .app file or .csv file with specified data ranges. If GRBFs were included in the calibration model, the user may specify if they should be included when performing approximation. If ANOVA was performed during calibration, prediction intervals for the new load approximations may be calculated at a specified confidence level. The possible file outputs from AOX_approx are identical to those for the approximation action in AOX_BalCal. The following figure shows the GUI for AOX_approx.
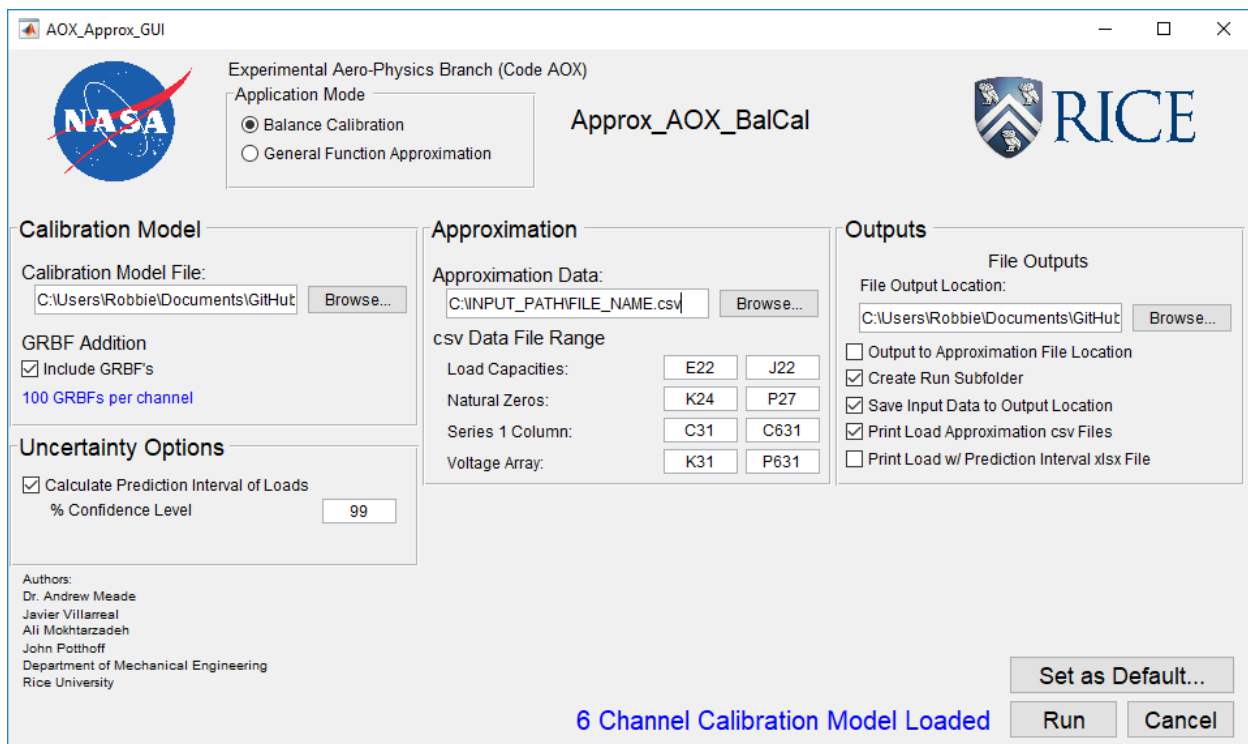
Calibration and Validation Data Splitter Program

A MATLAB script named 'Cal_Val_dataSplitter_GUI.m' was also made to split a dataset from experimental calibration testing into calibration and validation sets. The program accepts the original full dataset as a .csv file with ranges specified for the Series 1 Column, Load Array, and Voltage Array. The calibration and validation sets are saved from the program in two generated .csv files. The user specifies the output location and names for these files. Additionally, the user specifies the percentage of the original data that will go to the calibration set. The data may be split using Latin Hypercube Sampling (LHS) or Randomly.

If splitting using LHS, first, datapoints with voltages in the bottom or top 5% of the full range are automatically put into the calibration dataset. Next, the remaining points are split based on the set percentages using LHS. This may result in not every series being represented in the calibration or validation datasets. The first datapoints in each series are included in both the calibration and validation datasets.

When performing a random split, the set percentage of points are randomly assigned to calibration, and the remainder go to validation in each series. This guarantees that each series will be represented in both calibration and validation. The first datapoints in each series are included in both the calibration and validation datasets.

The output .csv files for calibration and validation also include the point IDs, series labels, and top header section above the load and voltage arrays from the original data file. Additionally, a description is added to the top row of the .csv files containing a description of the method used to split the data. The data splitter GUI is shown below.

References

[1] Potthoff, J. Development of Wind Tunnel Internal Strain-Gage Balance Calibration Software with Self-Assembling Gaussian Radial Basis Function Algorithm. Rice University, Houston, Texas, 2020.

[2] Montgomery, D. C. Introduction to Linear Regression Analysis / Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining. Wiley, Hoboken, N.J, 2012.

[3] Ulbrich, N. Analysis of Multivariate Experimental Data Using A Simplified Regression Model Search Algorithm. In AIAA Ground Testing Conference, No. AIAA 2013-2996, San Diego, CA, 2013.

[4] Ulbrich, N. Regression Model Optimization for the Analysis of Experimental Data. In 47th AIAA Aerospace Sciences Meeting and Exhibit, No. 2009–1344, Orlando, Florida, 2009.

[5] Mohd Razali, N., and Yap, B. "Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests." J. Stat. Model. Analytics, Vol. 2, 2011.

[6] Fasshauer, G. E. Meshfree Approximation Methods with MATLAB. World Scientific, Singapore ; Hackensack, N.J, 2007.

[7] Royston, P. "Remark AS R94: A Remark on Algorithm AS 181: The W-Test for Normality." Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 44, No. 4, 1995, pp. 547–551.