International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 12 Issue 7, July-2023, Impact Factor: 7.957

Box–Behnken Response-Surface Modelling of AHSS Spot Welding: A Reproducible Simulation Study

Seshabattar Phaneendra[1], Dr. Dheeraj Nimawat[2]

[1]Research Scholar, Department of Mechanical Engineering, JS University, Shikohabad, UP
[2]Supervisor, Department of Mechanical Engineering, JS University, Shikohabad, UP

ABSTRACT

Process-parameter studies of resistance spot welding often move rapidly from a fitted response surface to an optimum schedule. That transition is defensible only when the provenance of the observations, the estimability of the model and the limits of prediction are explicit. This paper develops a reproducible computational demonstration using the planning domain and synthetic results of an AHSS spot-welding thesis. Two response scenarios carry the labels DP780 and DP980 for continuity, but their coefficients are assumed rather than calibrated against steel measurements. A three-factor Box–Behnken design contains twelve edge settings and five centre rows. Three synthetic replicates are generated for tensile-shear and cross-tension load at each row, and ten-term quadratic models are fitted to row means. The full design matrix has rank ten. Overall coefficients of determination exceed 0.99, yet the DP780-labelled cross-tension model produces a lack-of-fit probability of 0.0304. Withholding all centre rows makes the quadratic model unidentifiable because the sum of the squared factor coordinates is constant at every edge setting. These findings demonstrate why high apparent fit cannot replace a design audit or a prediction-domain statement. The paper provides the equations, coefficient tables, design structure, diagnostics and reproduction logic needed to inspect the calculation. Its contribution is methodological transparency, not a new welding mechanism or a validated processing recommendation. Physical interpretation requires independently characterized material, calibrated process measurements and fresh destructive tests.

Keywords: resistance spot welding; AHSS; Box–Behnken design; synthetic data; response surface; identifiability; lack of fit.

INTRODUCTION

A welded sheet assembly is not defined solely by the strength class printed on its material certificate. The joint introduces a local change in geometry and material condition, and the usefulness of the assembly depends on how that region transfers load. Resistance spot welding is important in automotive sheet joining because it creates a localized connection through a combination of electrode force, electric current and timed restraint. Kimchi and Phillips (2023) provide the directly relevant process background. The present paper does not attempt to repeat their treatment; instead, it examines the statistical reasoning used when a small process experiment is converted into a predictive model.

The motivating thesis proposes similar-grade DP780/DP780 and DP980/DP980 joints with a preferred common sheet thickness of 1.2 mm. Its experimental chapters describe a future programme, whereas its revised conclusions contain an explicitly uncalibrated numerical example. This distinction determines the scope of the present manuscript. No specimen manufacture, mechanical testing, metallographic measurement or machine qualification is asserted. The material labels identify two assumed mathematical scenarios. They do not establish that the numerical difference between the scenarios is a real difference between commercial steels.

A computational example is still useful when it exposes decisions that are frequently concealed in abbreviated reporting. A reader should be able to identify the experimental unit, distinguish replicate observations from repeated instrument readings, reconstruct the model matrix and explain why a prediction is supported at one location but not another. The

example also allows a deliberately inconvenient diagnostic outcome to remain visible. Such an outcome is easier to discuss honestly before a laboratory programme has produced an expensive dataset that investigators may feel pressure to defend.
The central research question is therefore not which current produces the strongest weld. It is whether a declared three-factor quadratic workflow can be reproduced and interpreted without overstating its evidential strength. Three subsidiary questions follow. First, which features of the Box–Behnken layout make the quadratic coefficients identifiable? Second, how do averaging and centre replication influence residual and lack-of-fit calculations? Third, what does prediction error mean when the validation exercise is generated by the same mathematical assumptions as the fitted data?

The contribution is an auditable numerical case study. It combines an explicit generator with a design-rank argument, response-specific diagnostics and a carefully restricted cross-validation exercise. It does not claim that quadratic regression, Box–Behnken designs or least squares are novel. Nor does it claim that a synthetic demonstration is a substitute for welding research. Its value lies in making the inferential boundary visible and in providing a basis on which a later experiment can be planned, challenged and reported.

This manuscript is the modelling member of a three-paper set derived from the same thesis. A companion manuscript considers preference-sensitive optimization using the same synthetic dataset. Another develops a proposed laboratory validation framework. These manuscripts are related analyses, not three independently executed experiments. Any eventual submission must disclose that relationship and assess whether separate publication is justified by genuinely distinct questions and sufficient additional evidence.

PROCESS CONTEXT AND EVIDENCE BOUNDARY

Why three process factors are retained

Current, current-application time and electrode force are retained because they are the three controllable factors specified in the source thesis. The simplified electrical expression involving current squared, resistance and time explains why current and duration belong in a welding discussion, but it does not furnish the numerical response coefficients used here. Resistance evolves, the heat path extends beyond the joining interface, and electrode force affects contact and restraint. A statistical load model should therefore not be relabelled as a thermal or metallurgical simulation.

In particular, tensile-shear load and cross-tension load are not calculated from conservation of energy in this example. They are generated directly from assumed polynomials expressed in kilonewtons. No temperature field, phase-transformation kinetics, electrical contact law or fracture criterion is solved. The units help identify the response being imitated, but they do not transform an assumed load equation into a constitutive model. This distinction is retained in every numerical table and figure.

Welding metallurgy and joining performance supply the physical context for a future experiment, not evidence for the chosen coefficients. Kou (2020) addresses metallurgical aspects of welding, while Dwivedi (2022) addresses joining mechanisms and performance. Their inclusion does not imply that either book reports the generator used in this paper. The generator is attributed only to the supplied thesis calculation and its disclosed assumptions.

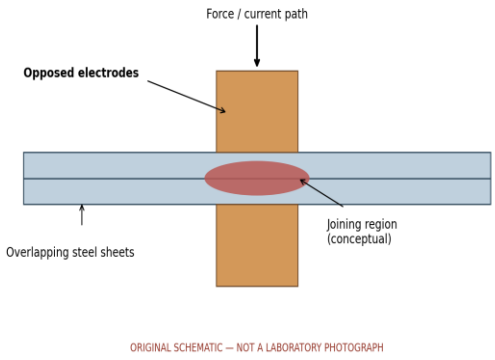


Figure 1. Resistance spot-welding arrangement. Original schematic, not to scale; no laboratory installation or measured nugget is depicted.

Observation classes

Four classes of information are separated. Planning information includes grade labels, preferred thickness and factor ranges. Assumed information includes all generating coefficients and noise scales. Computed information includes fitted

coefficients, residuals, analysis-of-variance quantities and prediction errors. Missing physical information includes chemistry, coating condition, instrument calibration, measured force, fracture appearance and actual nugget dimensions. Keeping these classes separate prevents a table of computed values from acquiring an experimental meaning through its placement in a manuscript.

The distinction also applies to pictures. Figure 1 is an original explanatory schematic, not a photograph of a welding installation. The response plots are generated from numerical arrays, not from imaging instruments. No micrograph is synthesized to suggest that a ferrite–martensite structure or a heat-affected-zone boundary was observed. A future paper can replace or complement these illustrations with real images, provided their specimen identifiers, preparation procedures and measurement scales are retained.

Domain of the calculation

The physical-coordinate ranges are 6–10 kA for current, 150–350 ms for time and 2.5–4.5 kN for force. These are planning bounds inherited from the thesis, not demonstrated safe operating limits. Coding maps them to a common interval so that polynomial terms are numerically interpretable. The centre is 8 kA, 250 ms and 3.5 kN. A coded unit corresponds to 2 kA, 100 ms and 1 kN respectively.

The statistical domain is narrower than an unrestricted three-dimensional box when predictions are interpreted conservatively. The design does not observe the cube corners where all three coded factors are simultaneously extreme. This paper therefore distinguishes the rectangular coordinate bounds from the convex hull of the actual design settings. The missing corners remain visible in the graphical treatment rather than being filled by a smooth surface and then described as equally supported.

DESIGN AND SYNTHETIC-DATA GENERATION

Design construction

The design contains the four sign combinations for each pair of active factors while the third factor is zero. This produces twelve edge settings. Five additional rows repeat the centre, yielding seventeen rows but only thirteen distinct settings. Each response receives three independently drawn synthetic replicates at each row. Consequently, there are fifty-one synthetic values per response and scenario, but the regression reported here uses seventeen row means.

The centre rows are not redundant in the same sense as repeated digits in a spreadsheet. They serve both an estimation role and a residual-comparison role. In the fitted polynomial, the centre directly informs the intercept while the edge settings inform combinations of intercept, linear effects, interactions and curvature. In the lack-of-fit partition, the variation between centre row means provides four pure-error degrees of freedom at the analysis scale used here.

The layout is not a Taguchi L9 or L27 array. Some columns are orthogonal because of sign balance, but the complete quadratic matrix contains relationships that cannot be summarized by calling the whole experiment orthogonal. The distinction matters because a factor-level average from a screening array and a coefficient from a quadratic surface answer different questions. A label should identify the actual design, not merely signal that a statistical package was used.

Table 1. Coded and physical design; planning/computational settings, not an executed experiment.

Row	x1	x2	x3	I (kA)	t (ms)	F (kN)
1	-1	-1	0	6.0	150.0	3.5
2	-1	1	0	6.0	350.0	3.5
3	1	-1	0	10.0	150.0	3.5
4	1	1	0	10.0	350.0	3.5
5	-1	0	-1	6.0	250.0	2.5
6	-1	0	1	6.0	250.0	4.5
7	1	0	-1	10.0	250.0	2.5

Row	x1	x2	x3	I (kA)	t (ms)	F (kN)
8	1	0	1	10.0	250.0	4.5
9	0	-1	-1	8.0	150.0	2.5
10	0	-1	1	8.0	150.0	4.5
11	0	1	-1	8.0	350.0	2.5
12	0	1	1	8.0	350.0	4.5
13	0	0	0	8.0	250.0	3.5
14	0	0	0	8.0	250.0	3.5
15	0	0	0	8.0	250.0	3.5
16	0	0	0	8.0	250.0	3.5
17	0	0	0	8.0	250.0	3.5

Coding and model basis

Let x1 = (I − 8)/2, x2 = (t − 250)/100 and x3 = F − 3.5, with the physical variables expressed in the units stated above. The row vector used for both generation and fitting is f(x) = [1, x1, x2, x3, $x1^2$, $x2^2$, $x3^2$, x1x2, x1x3, x2x3]. The mean response is μ(x) = f(x)b. The order of these ten columns is part of the reproduction specification and must remain unchanged when coefficients are transferred between software environments.

The model includes all linear, squared and pairwise interaction terms. It does not include a three-factor interaction, cubic terms or a discontinuity representing expulsion. This is a deliberate restriction. The example is designed to inspect a familiar local response-surface model, not to claim that every important feature of spot welding is smooth and quadratic. A real transition between failure mechanisms may demand a different representation or a restricted fitting region.

Coefficient interpretation depends on location. The slope with respect to x1 is b1 + 2b11x1 + b12x2 + b13x3. Accordingly, b1 is the local slope at the coded centre, not a universal effect of current across the domain. Physical-coordinate slopes additionally require the scale transformation: the current derivative is the coded derivative divided by two. Forgetting that conversion would produce a numerical error even if the fitted surface itself were correct.

Generator assumptions

The four coefficient vectors are listed in Table 2. The two scenarios intentionally have different intercepts and curvatures. Negative squared coefficients were selected to create a region of diminishing response and an interior trade-off problem. These features are imposed before any synthetic values are drawn. Recovering them through regression is therefore an internal consistency result, not a discovery of actual material behaviour.

Table 2. Assumed generator coefficients; all load coefficients in kN.

Term	780 TS	780 CT	980 TS	980 CT
Intercept	8.400	3.800	9.100	4.000
x1	0.800	0.350	0.950	0.300
x2	0.450	0.350	0.350	0.420
x3	0.200	0.280	0.300	0.320
$x1^2$	-0.700	-0.420	-0.850	-0.450
$x2^2$	-0.450	-0.380	-0.550	-0.480

Term	780 TS	780 CT	980 TS	980 CT
$x3^2$	-0.300	-0.320	-0.350	-0.360
x1x2	0.180	0.120	0.150	0.100
x1x3	0.100	0.080	0.100	0.060
x2x3	-0.080	0.050	-0.080	0.040

Independent Gaussian perturbations are added with standard deviations of 0.18 kN for tensile-shear load and 0.10 kN for cross-tension load. These standard deviations are assumptions, not estimated laboratory repeatability. Within the example, a row mean has a noise standard deviation reduced by the square root of three because it averages three independent draws. That mathematical consequence must not be confused with a demonstrated benefit from taking three readings on one physical specimen.

The independence assumption is particularly important. Real welds produced on the same day or with the same electrodes may share disturbances. The generator contains no day effect, material-batch effect, common current drift or correlation between response channels. It therefore cannot test how well the model would behave under those sources of dependence. Such extensions would constitute new simulation scenarios and should be identified as such rather than silently incorporated into this reproduction.

Random-number sequence

The calculation uses NumPy default_rng with seed 260228. DP780 is processed before DP980. Within each scenario, the tensile-shear noise array is drawn before the cross-tension array, each with three rows and seventeen columns. The resulting data are rearranged into design-row, replicate and response dimensions. The ordering is stated because a fixed seed alone does not specify which random values are assigned to which observations.

The source calculation also draws later numerical holdout values for the first scenario before beginning the second. These draws advance the random generator. A reproduction that skips them will recover DP780 but not the reported DP980 realization. The manuscript therefore retains the complete sequence in its reproduction description. This is a software-provenance issue rather than a scientific effect, but it materially changes printed coefficients if ignored.

ESTIMATION AND DESIGN IDENTIFIABILITY

Least-squares fitting

Ordinary least squares minimizes the sum of squared differences between the seventeen synthetic row means and their fitted values. Each response is fitted separately, using the same ten-column design matrix. No stepwise deletion, selective exclusion or search for a favourable subset of terms is performed. The calculation follows the general linear-model framework discussed by Beder (2022), while all numerical results arise from the disclosed generator.

Fitting row means is a transparent convention for this example, but it is not the only possible analysis. A model fitted to all fifty-one draws would retain within-row variation and would have a different residual partition. Both analyses can be coherent if their assumptions and experimental units are clearly specified. The present manuscript does not combine the coefficient estimates from one convention with the degrees of freedom from the other.

Equal replication and equal assumed variance make unweighted fitting of row means appropriate within this generator. If a later experiment has unequal replication or materially different variance across settings, this justification will no longer hold automatically. Weighted least squares, variance modelling or a hierarchical analysis may then be needed. The decision should follow the measurement structure rather than a desire to preserve the appearance of the original table.

Rank of the full design

The complete design matrix has rank ten, matching the number of fitted coefficients. This establishes that a unique least-squares coefficient vector is identifiable for each response under the declared basis. It does not establish low uncertainty, good extrapolation or physical correctness. Rank answers whether the columns can be separated algebraically; it does not answer whether the resulting estimates are sufficiently precise for an engineering decision.

Sign balance creates useful cancellations. The linear columns sum to zero and their pairwise inner products are zero. However, squared columns are nonnegative and overlap at edge settings. Their relationships with the intercept and with one

another differ from those of the signed linear columns. A complete audit therefore inspects the actual matrix rather than inferring all its properties from the symmetry of a picture.

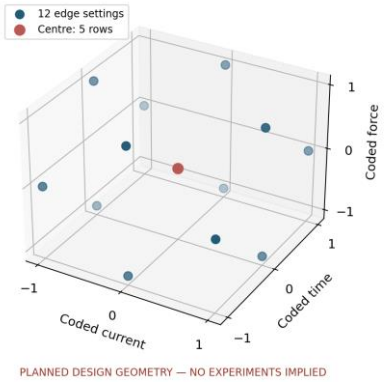


Figure 2. Three-factor Box–Behnken layout. Twelve distinct edge settings and five repeated centre rows; planned/computational geometry only.

Why deleting the centre is different

At every edge setting, $x1^2 + x2^2 + x3^2 = 2$. If all centre rows are removed, the intercept column becomes linearly dependent on the three squared columns. The ten-term model cannot then identify the intercept and the common component of curvature separately. The loss is structural; more numerical precision or a different least-squares routine cannot recreate information absent from the design.

This observation explains a limitation of leave-one-setting-out validation. Withholding an edge setting leaves sufficient information for the declared model, whereas withholding the unique centre setting removes all five centre rows and creates rank deficiency. Reporting an apparently successful prediction for that fold by relying on an arbitrary generalized inverse would conceal the fact that the full coefficient vector is no longer uniquely determined.

The centre fold is therefore excluded explicitly from the reported cross-validation summary. The resulting error statistic is named leave-one-edge-setting-out error. It covers twelve held-out locations and must not be promoted to full-domain validation. The distinction is not cosmetic: changing the label changes the reader's understanding of what has actually been challenged.

RESULTS OF THE REPRODUCIBLE CALCULATION

Coefficient recovery

Table 3 reports the fitted coefficients in the declared order. Their proximity to the generating coefficients is expected because the fitting basis matches the generating basis and the imposed perturbations are relatively small. Differences reflect the particular random realization. They do not estimate a discrepancy between a physical welding mechanism and a statistical approximation because no physical mechanism was used to generate the data.

Table 3. Fitted coefficients from synthetic row means; use coded coordinates.

Term	780 TS	780 CT	980 TS	980 CT
Intercept	8.4888	3.7924	9.1588	3.9989
x1	0.7897	0.3581	0.9920	0.3373
x2	0.4882	0.3569	0.4073	0.4338
x3	0.1626	0.2970	0.3537	0.2955

Term	780 TS	780 CT	980 TS	980 CT
$x1^2$	-0.7085	-0.4231	-0.8430	-0.4813
$x2^2$	-0.5230	-0.3560	-0.5360	-0.4864
$x3^2$	-0.3547	-0.2818	-0.4067	-0.3146
x1x2	0.1770	0.1039	0.1313	0.0892
x1x3	0.0758	0.0992	0.0819	0.0323
x2x3	-0.0827	0.0672	-0.1945	0.0411

The fitted intercepts are close to the assumed centre responses. Linear terms retain positive signs in all four models, while squared terms remain negative. These signs describe the chosen numerical example. The interpretation should stop there unless laboratory observations independently establish the same pattern. A statement that current increases real weld strength would require consideration of the relevant location, range, response, failure mode and uncertainty.

The coefficients also illustrate why comparing raw magnitudes across terms can mislead. A squared coefficient multiplies a different function of the factor than a linear coefficient, and an interaction modifies the slope conditionally. Moreover, the physical scaling differs by factor. Any ranking of importance must specify whether it refers to a derivative at the centre, a response change across a specified interval, an incremental sum of squares or some other defined quantity.

Overall fit

The DP780-labelled tensile-shear and cross-tension models have R^2 values of approximately 0.9925 and 0.9950. The corresponding DP980-labelled values are approximately 0.9952 and 0.9928. These values indicate that the fitted polynomials explain most of the variation in the synthetic row means. They are not probabilities that the models are true, and they provide no direct evidence of external predictive accuracy.

The residual standard errors are about 0.1112 and 0.0577 kN for the first scenario and 0.1046 and 0.0775 kN for the second. These statistics use seven residual degrees of freedom. They should be compared with the scale of row-mean noise rather than with an assumed distribution of individual weld strengths. Averaging and the number of fitted coefficients both influence their interpretation.

Table 4. Synthetic model diagnostics; model/residual/total df = 9/7/16; lack-of-fit/pure-error df = 3/4.

Scenario / response	R^2	Adj. R^2	RSE (kN)	Model F	Model p	LOF p
DP780 TS	0.9925	0.9829	0.1112	103.11	1.33e-06	0.1482
DP780 CT	0.9950	0.9886	0.0577	155.32	3.21e-07	0.0304
DP980 TS	0.9952	0.9890	0.1046	160.87	2.84e-07	0.1176
DP980 CT	0.9928	0.9836	0.0775	107.56	1.15e-06	0.8285

Lack-of-fit finding

The DP780-labelled cross-tension model has a lack-of-fit F statistic of approximately 8.9026 and a probability of 0.0304 using three and four degrees of freedom. At a nominal five-percent threshold, this is an adverse diagnostic. It remains in the manuscript. No seed is changed, no centre row is deleted and no alternative analysis is selected solely to make that probability exceed the threshold.

Because the mean generator is exactly quadratic, this particular result demonstrates that a nominal rejection can occur through the selected random realization and the limited pure-error denominator. That explanation does not justify ignoring the diagnostic in an actual experiment. In laboratory data, the generating mean is unknown, and the same signal could arise from model inadequacy, dependence, variance changes, measurement problems or a combination of causes. The appropriate response is to distinguish interpretation under known simulation truth from interpretation under unknown physical truth. Here, the analyst can trace the observation to a declared generator. In a real study, the analyst would inspect residual structure, replicate consistency and process records, then consider preplanned additional observations or a revised model. Repeating an experiment only until a preferred probability appears would not constitute validation.

Residual and held-out prediction patterns

Residual plots show the component of the synthetic row means not absorbed by the fitted quadratic. The horizontal index is design-row order, not physical production time. Consequently, a sequence of residuals cannot be interpreted as electrode wear, operator learning or machine heating in this example. Those interpretations would require a real execution order and associated process records.

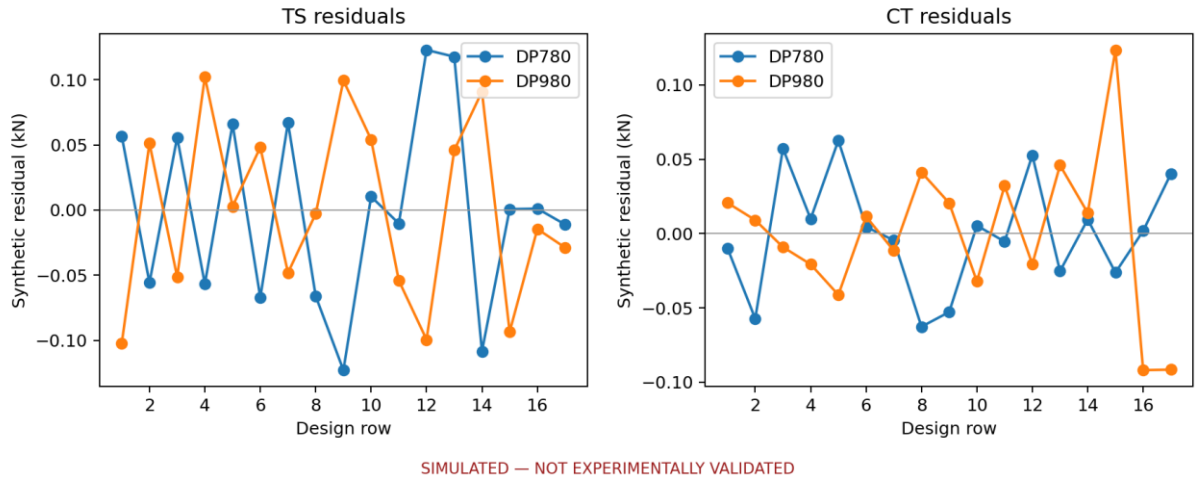


Figure 3. Residuals of synthetic row means against computational row order. The horizontal axis is not laboratory execution time.

The leave-one-edge-setting-out root-mean-square errors are approximately 0.2849 and 0.1645 kN for DP780-labelled tensile-shear and cross-tension responses. For the DP980-labelled scenario, they are approximately 0.2744 and 0.1007 kN. These exceed the corresponding in-sample residual errors, which is unsurprising because each prediction is made without the withheld setting and with fewer observations available to estimate ten coefficients.

Table 5. Leave-one-edge-setting-out error on synthetic row means; centre fold excluded for rank deficiency.

Scenario	TS RMSE (kN)	CT RMSE (kN)	Supported folds
DP780	0.2849	0.1645	12 edge settings
DP980	0.2744	0.1007	12 edge settings

The disparity between fitted and held-out errors is informative even though neither is an experimental validation statistic. It shows how a visually impressive in-sample fit can coexist with greater sensitivity to the removal of a design location. The

limited number of unique settings means that one withheld row can materially change the local curvature estimate. This should encourage scrutiny of prediction support rather than a search for a single favourable goodness-of-fit number.

DISCUSSION

The meaning of a successful simulation

A successful reproduction establishes that a reader can recover the stated arrays and statistics under the same assumptions. It does not establish that the assumptions describe a welding machine. Reproducibility and validity are related but different requirements. A computational result can be perfectly reproducible and physically irrelevant, while an apparently plausible physical result can be impossible to audit because its processing choices were not retained.

The present exercise deliberately prioritizes the first requirement while identifying the evidence needed for the second. It is therefore most appropriately read as a methods case study or teaching example. Its numerical load values should not be used to design an automotive connection, qualify a welding procedure or claim improvement over an existing production schedule. Such uses would require evidence that is absent from the source thesis.

What is and is not learned about model complexity

The quadratic fit is deliberately well matched to the generator. This provides a controlled setting for studying rank and residual partitioning, but it is a favourable test of functional form. The example does not show that quadratic regression outperforms another model family on independent physical data. Nor does it demonstrate that additional complexity would be beneficial. A flexible model could fit the same observations more closely while becoming less identifiable or less interpretable.

Statistical learning references provide a broader vocabulary for prediction and validation. James et al. (2023) discuss modelling and prediction tools, and Suzuki (2021) develops related mathematical exercises. In the present case, however, the small number of unique settings is the controlling consideration. Introducing a complex learner would not manufacture information about missing corners, material variability or fracture transitions. Model choice must remain proportionate to the information actually available.

Uncertainty and the scale of replication

Three synthetic draws per row reduce variance of the row mean under independence. They do not create three independently varied process settings. Likewise, five centre rows provide information about repeatability at one setting, not five additional locations spanning the design domain. Reporting both row count and unique-setting count is therefore necessary. The distinction becomes even more important when experimental specimens are nested within production batches or test sessions.

A future uncertainty analysis should separate measurement uncertainty, within-setting weld variation, between-session variation and model-form discrepancy. These sources cannot be reconstructed from the imposed Gaussian noise alone. Treating them as one residual term may be convenient, but convenience does not establish that they share a common distribution or respond similarly to additional sampling. A well-designed experiment records enough structure to identify at least the dominant components.

Limits of physical analogy

The assumed response surfaces have units and shapes that make them convenient for discussing optimization. Nevertheless, visual plausibility is not calibration. A concave response could be drawn for many processes without proving the existence or location of a physical maximum. The polynomial omits expulsion, changes in contact area, crack initiation and transitions between fracture modes. Any of those may make a nominally smooth load surface inappropriate over part of an actual domain.

The same caution applies to grade comparison. The DP980-labelled scenario has a higher generating intercept than the DP780-labelled scenario because the author selected it that way. A difference in fitted load is therefore largely predetermined. It cannot establish superiority of DP980 joints, a greater martensite fraction or improved resistance to cross-tension failure. Those are material-specific claims requiring material-specific evidence.

Reporting adverse information

The retained lack-of-fit result provides a useful test of reporting discipline. If a manuscript displays only large R^2 values and favourable optimum settings, readers cannot evaluate whether the analysis encountered contradictory evidence.

Complete reporting is not achieved merely by attaching more decimal places. It requires showing the diagnostics that could alter the conclusion, explaining their scope and acknowledging when an inferential step remains unresolved.

The present paper also avoids interpreting partial term tests as additive percentage contributions. Removing one term from a full model while retaining others produces a conditional comparison. Such comparisons can overlap in the information they explain, particularly when the columns are correlated. Percentages obtained by normalizing these increments need not partition a meaningful total. A later physical paper should define its importance metric before displaying a ranked factor chart.

Reproduction and audit protocol

The minimum reproduction package consists of the factor coding, the ordered design, the four generating coefficient vectors, the two noise scales, the random generator and its draw order. The analysis then computes row means, fits the full basis and obtains residuals. The manuscript provides these ingredients in equations and tables. The synthetic values are not proprietary instrument output, and their provenance should remain inseparable from any subsequent reuse.

An independent implementation should begin with dimensional and structural checks. The design must have seventeen rows, three coded factor columns and thirteen unique settings. The model matrix must have ten columns and rank ten. The centre should contain five rows. Each raw response array should contain three draws per design row. Failing any of these checks means that later agreement in a rounded optimum could be accidental.

The residual calculation should be performed using unrounded values. The total sum of squares is calculated about the response mean; the residual sum of squares is calculated about the fitted response; their difference gives the regression sum of squares. Model, residual and total degrees of freedom are nine, seven and sixteen. The five centre row means produce four pure-error degrees of freedom, leaving three for lack of fit.

The held-out calculation should remove all observations associated with the selected design setting. Splitting replicates from the same setting between training and validation would assess a different and easier question. Before each refit, the matrix rank is checked. The centre fold is recorded as structurally unsupported rather than assigned a numerical score that conceals non-identifiability. The twelve supported edge folds then form the stated error summary.

Numerical tolerances should be declared when comparing independent implementations. Different linear-algebra libraries can produce negligible floating-point differences, whereas changes in draw ordering or printed-coefficient rounding can produce larger differences. The reproduction should compare coefficients, residual sums of squares and selected response values rather than rely only on a matching figure. A plot can look identical even when a transcription error affects a consequential digit.

Finally, the audit should inspect language as well as arithmetic. Words such as measured, observed, confirmed and validated must refer to the kind of evidence actually available. In this manuscript, observed numerical patterns are observations of a simulated dataset. They are not observations of a laboratory process. A technically accurate table becomes misleading if its surrounding prose erases that distinction.

ADDITIONAL ANALYTICAL CHECKS

Information carried by different design locations

The matrix structure provides a useful explanation of prediction uncertainty without requiring another random experiment. Under the declared ordering, each linear column has a squared length of eight, and each interaction column has a squared length of four. The intercept column has a squared length of seventeen. Each squared column sums to eight, has a squared length of eight, and has an inner product of four with each other squared column. These quantities follow directly from counting the nonzero coordinates in the twelve edge settings and five centre rows.

The signed linear and interaction columns are orthogonal to the intercept and squared columns in this balanced layout. The intercept and squared block is not diagonal. Its nonzero off-diagonal entries explain why a coefficient uncertainty calculation must use the full information matrix rather than divide a residual variance by the squared length of each column independently. Ignoring those relationships would give incorrect uncertainty for the intercept and curvature terms even though the linear coefficient calculations might appear reasonable.

For independent row-mean errors with common variance s^2, the fitted coefficient covariance is s^2 multiplied by the inverse of X-transpose-X. In the present design, the intercept diagonal entry of that inverse is 0.20. Each linear diagonal entry is

0.125, and each interaction diagonal entry is 0.25. Each squared-term diagonal entry is 0.2375. The covariance multiplier between two different squared terms is −0.0125, and that between the intercept and a squared term is −0.10. These values are analytical properties of this matrix, not estimates of a material property.

The intercept estimate is therefore closely tied to the centre information. Indeed, with this layout and full quadratic basis, the fitted intercept equals the mean of the five centre row means. The edge settings alone cannot determine it separately from the common curvature component. This offers a direct interpretation of the rank result: the centre observations anchor the vertical level that distinguishes an overall offset from a common increase or decrease in all squared terms.

Leverage and supported prediction

The fitted-mean variance at a coded point is $s^2\, f(x)(X\text{-transpose-}X)^{-1}f(x)\text{-transpose}$. At the design centre, its variance multiplier is 0.20. At each edge setting, the multiplier is 0.75. The trace check is twelve times 0.75 plus five times 0.20, which equals ten, the number of fitted coefficients. This provides an additional reproducibility check that is independent of the chosen response values and random seed.

These leverage values explain why equal geometric distances on a diagram do not automatically mean equal estimation precision. The repeated centre supplies relatively strong information about its fitted mean, while each individual edge setting carries greater leverage. A point with high leverage can strongly influence its own fitted value and the nearby surface. Removing it for prediction assessment can therefore change the estimated curvature more than the small in-sample residual at that point might suggest.

Leverage is not a defect flag by itself. Designed experiments deliberately place observations where they carry information about model terms. The relevant question is whether the design provides enough independent support for the intended predictions and whether influential locations have reliable measurements. In a physical experiment, a high-leverage point affected by an unnoticed fixture problem could materially distort the surface. Replication and process traceability are therefore especially valuable when an observation has substantial influence.

The calculation also clarifies two different uncertainty intervals. An interval for the fitted mean uses the leverage multiplier. A prediction interval for a new independent row mean adds the new observation’s error variance. If the target is an individual weld rather than a three-weld mean, the additional variance must be defined at that individual scale. Using the mean-response interval as though it described individual joint performance would understate variation even if the regression assumptions were otherwise correct.

Rounding and coefficient transport

A printed coefficient table is a useful audit aid, but it is not always a lossless data format. Suppose each of the ten coefficients is rounded to four decimal places. Within the coded cube, every basis function has absolute value at most one. The worst-case bound on the resulting prediction difference is then ten times half of one unit in the fourth decimal place, or 0.0005 kN. This conservative bound applies to coefficient rounding alone, not to refitting from rounded raw values or to changing the model basis.

The bound helps distinguish harmless presentation differences from consequential implementation errors. A disagreement of several tenths of a kilonewton cannot plausibly be explained only by rounding these coefficients to four decimals. More likely causes would include using physical coordinates with coded coefficients, swapping interaction columns, omitting the intercept or changing the random-number sequence. A reproduction report should investigate such structural errors before attributing differences to floating-point arithmetic.

Units deserve the same discipline. Entering time in seconds while using a coding expression intended for milliseconds produces a large coordinate error. Entering force in newtons rather than kilonewtons does the same. A program may still return a numerical answer because the polynomial accepts any real input. The existence of that answer does not establish that the input was meaningful. Explicit unit checks are therefore part of model implementation rather than an optional presentation detail.

Distinguishing planned contrasts from exploratory stories

A physical study might ask whether increasing current from one declared level to another changes a response at a fixed time and force. That is a planned contrast between two model predictions and can be accompanied by an uncertainty calculation using the difference between their basis vectors. It is a more precise question than asking whether current is important in general. The contrast states both the locations and the response whose change is being assessed.

The present synthetic example can calculate such contrasts, but their interpretation remains conditional on its generator. A useful classroom exercise is to compare a centre-based current contrast with a contrast near another time level. The interaction terms make the answers different even though the same current increment is used. This exposes why a single main-effect coefficient is insufficient to describe the whole response surface.

Exploratory interpretation is not prohibited, but it should be identified. If an investigator inspects many plots and then selects one striking pattern, that pattern has a different evidential status from a prespecified contrast. In a later laboratory paper, the distinction should guide which results are described as confirmatory and which motivate another experiment. The synthetic exercise provides a low-cost setting in which to practise that discipline before physical resources are committed.

What an expanded simulation would need to test

A stronger methodological investigation could systematically vary noise magnitude, replicate correlation, heteroscedasticity, omitted cubic terms and block drift. Each extension would answer a different robustness question. Increasing noise tests statistical precision under the same mean structure; adding a cubic term tests functional-form mismatch; adding a common session effect tests dependence. Combining all disturbances at once would make it harder to identify which assumption caused a change in performance.

Such an expanded simulation should define performance metrics before execution. Relevant metrics might include coefficient bias, interval coverage, frequency of lack-of-fit rejection and error at independently selected supported locations. The number of simulation repetitions should be chosen to control uncertainty in those metrics. One favourable realization is not a substitute for a repeated-simulation study, just as one favourable weld is not a substitute for a replicated physical experiment. This manuscript does not claim to have completed that larger programme. It reports one fully disclosed realization and several exact matrix properties. The distinction matters because the phrase simulation study can describe anything from a single worked example to a large controlled numerical experiment. Here, the intended contribution is a reproducible case study whose limitations are sufficiently explicit to support a more ambitious follow-up.

Interpreting a surrogate before extending it

Before adding another factor, the investigator should ask whether the current model answers its intended question at the present observation scale. Adding coating as a categorical variable, for example, would not simply require another column in the existing table. It would require independent material conditions and a design that separates coating from grade, supplier and batch. Without that structure, the additional coefficient could acquire a precise-looking estimate while retaining an ambiguous physical interpretation.

Similarly, replacing programmed current with measured current changes the role of the predictor. Programmed current belongs to the assigned treatment, whereas measured current can contain delivery error and may be affected by the evolving process. Either variable can be useful, but the model should state whether its purpose is to predict the effect of setting the controller or to describe outcomes conditional on a measured waveform. Those are related but not identical scientific questions. The same reasoning applies to intermediate responses such as nugget width. Including width as a predictor of load may improve an explanatory association, yet width is itself an outcome of welding and may not be available before a schedule is selected. A model intended for prospective process selection cannot assume access to a destructive measurement that only becomes available afterward. The information available at decision time should determine the predictor set.

These distinctions connect the numerical exercise to experimental planning without pretending that the exercise already contains physical evidence. Chaturvedi and Vendan (2021) place design considerations within the broader welding context. Here, the practical lesson is to define the intended decision, the observable inputs and the unit of prediction before increasing model complexity. A richer equation is not necessarily a more useful model if it requires unavailable inputs or estimates effects that the sampling design cannot separate. Finally, model extension should preserve a clear baseline implementation. The present ten-term calculation provides such a baseline because its assumptions and matrix properties are explicit. A later alternative can be compared against it using a declared criterion, supported prediction locations and an independent dataset. That comparison would create additional evidence; merely displaying a more elaborate fitted surface would not.

CONCLUSIONS

The declared Box–Behnken layout supports a ten-term quadratic model when all centre and edge rows are retained. Its algebraic structure is not equivalent to a Taguchi orthogonal-array analysis. Removing every centre row creates a specific

dependence between the intercept and squared terms, so the full quadratic cannot be uniquely identified in that validation fold.

High overall fit does not eliminate diagnostic concerns. All four synthetic fits produce R^2 values above 0.99, while the DP780-labelled cross-tension response retains a nominal lack-of-fit signal at $p = 0.0304$. The supported edge-setting prediction errors are larger than in-sample residual errors. These findings illustrate the need to report the observation scale, validation domain and adverse diagnostics together.

The results establish a reproducible numerical workflow, not a validated description of DP780 or DP980 welding. Physical use requires independently characterized sheets, calibrated welding records, appropriate destructive tests and an experiment capable of separating process effects from nuisance variation. The companion optimization analysis must inherit these limitations rather than treating fitted surfaces as unquestioned physical truth.

Declarations and relationship to the thesis

This manuscript is a computational draft derived from the supplied thesis, "Optimization of Process Parameters for Spot Welding of Advanced High-Strength Steels." Thesis author, institution, submission year and repository details must be supplied before formal submission. No human author or affiliation has been invented. The synthetic data are shared with the companion optimization manuscript; the validation manuscript is a proposed protocol. No independent experimental campaign is claimed. AI assistance was used in drafting and numerical presentation, and the named authors must verify the content and follow the target journal's disclosure policy. Funding, conflicts of interest and author contributions remain to be completed by the actual authors. All diagrams are explanatory or computational figures; none is a laboratory photograph.

REFERENCES

[1]. Beder, J. H. (2022). Linear models and design. Springer. https://doi.org/10.1007/978-3-031-08176-7
[2]. Chaturvedi, M., & Vendan, S. A. (2021). Advanced welding techniques: Holistic view with design perspectives. Springer. https://doi.org/10.1007/978-981-33-6621-3
[3]. Dwivedi, D. K. (2022). Fundamentals of metal joining: Processes, mechanism and performance. Springer. https://doi.org/10.1007/978-981-16-4819-9
[4]. James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). An introduction to statistical learning: With applications in Python. Springer. https://doi.org/10.1007/978-3-031-38747-0
[5]. Kimchi, M., & Phillips, D. H. (2023). Resistance spot welding: Fundamentals and applications for the automotive industry (2nd ed.). Springer. https://doi.org/10.1007/978-3-031-25783-4
[6]. Kou, S. (2020). Welding metallurgy (3rd ed.). Wiley. ISBN 9781119524816. https://www.wiley.com/en-us/Welding%2BMetallurgy%2C%2B3rd%2BEdition-p-9781119524816
[7]. Suzuki, J. (2021). Statistical learning with math and Python: 100 exercises for building logic. Springer. https://doi.org/10.1007/978-981-15-7877-9
[8]. Source thesis (unpublished document supplied for this work): Optimization of Process Parameters for Spot Welding of Advanced High-Strength Steels. Author, institution and year were not supplied and must be completed before submission. The source thesis is separate from the published-book references above.