

# DoEIY: A User Guide

Niamh Mac Fhionnlaoich<sup>1,2</sup>, Ye Yang<sup>1</sup>, Federico Galvanin<sup>1</sup>, and Stefan Guldin\*<sup>1</sup>

<sup>1</sup>Department of Chemical Engineering, UCL, London

<sup>2</sup>APC Ltd., Cherrywood Business Park, Loughlinstown, Dublin, D18 DH50, Ireland

## 1 Introduction

Welcome to DoEIY! This tool helps plan, conduct, and analyse Design of Experiments (DoE) in a structured, accessible, and user-friendly way. It guides the user from creating a design through model fitting and analysis, and on to visualising results and optimising responses. No software installation is needed, only an internet connection. This guide provides clear explanations of key DoE concepts along with step-by-step instructions.

## 2 DoEIY Overview

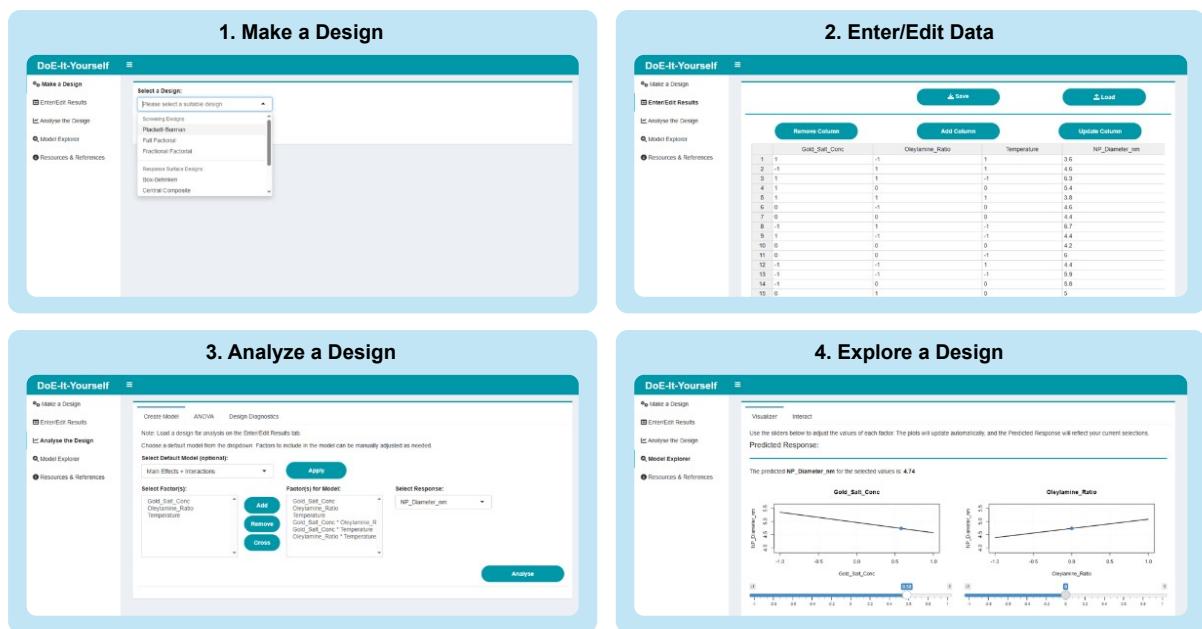


Figure 1: **Overview of the DoEIY software workflow.** 1. Make a Design: users select a design type, define factors, and generate the experimental design. 2. Enter/Edit Results: experimental responses are entered directly into the editable design table, with options to save or load designs. 3. Analyze the Design: users select or create custom models; the model summary and statistics as well as design diagnostics are presented. 4. Explore the Design: interactive visualization of factor-response relationships and optimization and prediction tools allow users to identify factor settings that maximize, minimize, or achieve a target response.

\*Corresponding author: s.guldin@ucl.ac.uk

The DoEIY software comprises 5 modules or tabs; these can be found in the navigation bar on the left (see Figure 1):

- **Make a Design**
- **Enter/Edit Results**
- **Analyze the Design**
- **Model Explorer**
- **Resources & References**

This layout mirrors the natural progression of a DoE; first you create an experimental design based on your needs, then you carry out the experiments and record your results, next you fit and assess a model, and finally, you use the model as needed (e.g. to maximize a response, understand how different experimental conditions impact the response). When you launch the webapp, you will land on the **Make a Design** tab. You can click on any of the other tabs in the navigation bar on the left to navigate to one of the other sections.

**Note:** The software will time-out after 30 minutes of inactivity. DoEIY doesn't store any information between sessions so be sure to save your design to continue working on it later.

### 3 Make a Design

The first step in the DoE process is to Make a Design. You begin by selecting a design from the dropdown menu (Figure 2):

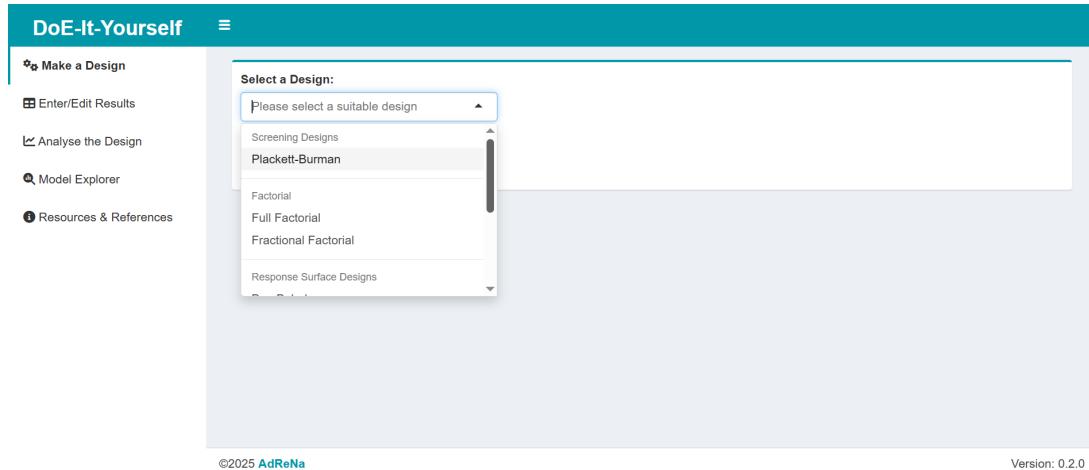


Figure 2: **Make a Design.** The *Make a Design* page in DoEIY, where you select a design type from the dropdown menu and enter the factor details to create a design.

There are a selection of design types and designs within them; you should choose what design type you need based on what you want to accomplish. The available designs are outlined in this section below.

Once you have chosen your design and click **Make Design**, a pop-up window will appear (see Figure 3a), where you will enter the factor details. The necessary factor details will depend on the chosen design. Factors can be added or removed by right-clicking on the table and selecting

the relevant option. After clicking **Next**, you will be taken to the design details (Figure 3b). If there are issues with the factor entries, the software will highlight the problems before allowing you to proceed. Once valid factor details are provided, the design details page will appear, summarising the factors, the number of runs, and the blocking options if available. At this stage, you may also choose whether to randomise the run order (the default is randomised). When you are satisfied, click **Create** and the design will be generated. The completed design will then be available under the **Enter/Edit Data** tab.

**Note:** In v1.0.0 of DoEIY, factor names may only contain letters, numbers, and underscores, i.e. syntactically valid names in R. As such, a valid factor name must:

- contain only letters, digits, underscores, and periods,
- begin with a letter or a period (but not a period followed by a digit),
- not be a reserved word (like `if`, `else`, or `function`).

### 3.1 Screening Designs

Screening designs are used at the early stage of experimentation when you have many possible active factors, i.e., factors that have an impact on the response(s), but do not yet know which ones are actually important. The goal is to quickly and efficiently identify the few factors that have the biggest influence on the response with the fewest experiments. Once the key factors are identified, you can focus later studies on those factors using more in-depth designs such as factorial or response surface methods. Currently the software offers Plackett-Burman designs for screening.

**Plackett-Burman Designs** are efficient screening tools for identifying which factors are likely to influence while keeping the number of experiments to a minimum. This makes these designs very effective for narrowing down a large set of possible factors to just those that actually affect the response. It is important to note that main effects may be mixed with higher-order interactions, meaning that what looks like a strong factor effect may in fact reflect an underlying interaction. Even so, these designs remain powerful tools for quickly focusing experimental effort on the most influential factors. This software offers Plackett-Burman designs for 4 to 23 factors.

For a Plackett–Burman design, select *Plackett–Burman* from the dropdown menu. In the factor details window (illustrated in Figure 3a), you must specify each factor with a name, its type (continuous, discrete, or categorical), and exactly two level values (e.g. `-1`, `1`). Once created, the design details will summarise the factors and total number of runs. You may then choose whether to keep the default randomised run order or disable randomisation before finalising the design. Blocking is not available for this design.

### 3.2 Factorial Designs

Factorial designs provide a structured way to study how multiple factors influence a response, both individually and in combination allowing you to estimate not only the main effects of each factor but also their interactions. This is important as in many real systems, the effect of one factor can depend on the level of another. Importantly, 2-level factorial designs are limited to estimating linear effects and interactions; they do not capture curvature in the response (e.g., quadratic effects). The software supports both full and fractional factorial designs.

**Full Factorial Designs** investigate every possible combination of factor levels, using two or more levels for each factor. This comprehensive approach enables the independent estimation

**a.**

**b.**

Figure 3: **Make a Plackett-Burman Design.** **a.** Factor details window for entering names, types, and two level values in a Plackett–Burman design. **b.** Design details window showing a summary of factors and runs, with the option to randomise the run order.

of all main effects and higher-order interactions, providing a complete understanding of factor relationships. This makes these designs ideal when the number of factors is small or resources allow for a larger number or runs. DoEIY provides full factorial designs for up to 25 factors with a maximum of 10 levels per factor.

**a. Provide Factor Details**

Full Factorial Designs are supported for 1-25 factors, which can be continuous, discrete, categorical or blocking. For each factor, 2-10 levels can be considered, select the appropriate number of levels for the factor from the corresponding dropdown. Please provide the values for the levels as comma separated values, e.g., '5, 10, 15'.

All entries in the table below must be completed. Factor Names can only contain letters, numbers, and underscores. Do NOT use spaces or special characters in the factor names.

Note: Add or remove rows by right clicking on a row.

	Factor Names	Factor Type	Num Levels	Level Values
1	Temperature	Continuous	3	5, 15, 25
2	Solvent	Categorical	4	Ethanol, Acetone, Octane, IPA
3	Number_Additions	Discrete	3	2, 4, 8

**b. Design Details**

Selected Design: Full Factorial

Factor Names	Factor Type	Num Levels	Level Values
Temperature	Continuous	3	5, 15, 25
Solvent	Categorical	4	Ethanol, Acetone, Octane, IPA
Number_Additions	Discrete	3	2, 4, 8

Number of Runs: 36

Blocking is not available for this design

Randomize Runs:

Back Create Cancel

Figure 4: **Make a Full Factorial Design.** a. Factor details window for specifying names, types, number of levels, and level values in a full factorial design. b. Design details window summarising the factors and the total number of runs, with the option to randomise the run order.

For a full factorial design, select *Full Factorial* from the dropdown menu. In the factor details window (illustrated in Figure 4a), you must provide a name, type, the number of levels, and the level values for each factor. These should be entered as comma-separated values (e.g. 0, 1, 2). Once valid factor details have been supplied, the design details will summarise the factors and the total number of runs required by the design. You may then choose whether to keep the default randomised run order or disable randomisation before finalising the design. Blocking is not currently supported for this design.

**Fractional Factorial Designs** are efficient alternatives to full factorial designs, utilizing only a fraction of the total possible runs. They allow for the estimation of main effects and, depending on the design's resolution, certain interactions. The resolution of a fractional factorial design indicates which types of effects are aliased with which. This software provides 2-level fractional factorial designs with the following resolutions:

- Resolution III - main effects are not aliased with each other but may be aliased with two-factor interactions.
- Resolution IV - main effects are not aliased with other main effects or two-factor interactions; two-factor interactions may still be aliased with each other.

- Resolution V - two-factor interactions are not aliased with main effects or other two-factor interactions, but may be aliased with three-factor interactions.

**a. Provide Factor Details**

Fractional Factorial designs are supported for 3 to 25 factors, which can be continuous, categorical, or discrete. Each factor is limited to two levels. Please provide the values for the levels as comma separated values, e.g., '-1, 1'. When setting up the design, select the required resolution from the dropdown menu. Note: Resolution III requires at least 3 factors. Resolution IV requires at least 4 factors, and Resolution V requires at least 5 factors. These factors can be continuous or categorical. For each factor, 2-10 levels can be considered, select the appropriate number of levels for the factor from the corresponding dropdown.

Note: Add or Remove columns by right clicking on the table.

	Factor Names	Factor Type	Level Values
1	Temperature	Continuous	5, 25
2	Solvent	Categorical	Ethanol, Octane
3	Number_Additions	Discrete	2, 6
4	Concentration_A	Continuous	5, 60
5	Reaction_Time	Continuous	30, 120

Next Cancel

**b. Design Details**

Selected Design: Fractional Factorial

Factor Names	Factor Type	Level Values
Temperature	Continuous	5, 25
Solvent	Categorical	Ethanol, Octane
Number_Additions	Discrete	2, 6
Concentration_A	Continuous	5, 60
Reaction_Time	Continuous	30, 120

Select number of runs: 16      Select number of blocks: 2

This is a **Resolution V** design.  
Randomize Runs:

Back Create Cancel

Figure 5: Make a Fractional Factorial Design. **a.** Factor details window for entering names, types, and two level values in a fractional factorial design. **b.** Design diagnostics window with dropdown menus to select the total number of runs and the number of blocks, plus the option to randomise.

Resolution III designs are generally suitable for screening studies, when the primary goal is simply to flag potentially important factors. Resolution IV is often considered the minimum acceptable for most studies, as it provides clearer estimation of main effects. Resolution V is preferred when interactions are expected to be important and you want confidence in interpreting both main and two-factor effects. Increasing the resolution provides clearer interpretation, but at the cost of a higher experimental workload. Selecting the appropriate resolution depends on balancing the level of detail you need against the number of runs that can be executed.

For a fractional factorial design, select *Fractional Factorial* from the dropdown menu. In the factor details window (illustrated in Figure 5a), you must provide a name, type, and level values for each factor. Only two levels are permitted, entered as comma-separated values (e.g. -1, 1). After completing the factor details, the design details page will appear, Figure 5b. Here, the first dropdown menu lets you select the total number of runs for the fractional factorial design, this number is always a power of 2. Based on your selection, the second dropdown menu will update to show the number of blocks the design can accommodate. You can then choose the appropriate number of blocks from the available options before finalising the design. As with the other designs, you can choose whether to randomize the order of the experimental runs.

### 3.3 Response Surface Designs

Response surface designs are used when you need to explore non-linearity in the relationship between factors and the response(s). The use more than two levels for each factor, allowing for the estimation of quadratic effects for modelling curvature in the response(s). While a 3-level full factorial design can provide this information, response surface designs are typically more efficient, needing fewer experimental runs to achieve the same insight.

**Box-Behnken Designs** are response surface designs that use 3 levels for each factor to efficiently estimate quadratic models. They allow independent estimation of main effects, two-factor interactions, and quadratic effects, without requiring the full number of runs of a three-level factorial. In this software, Box-Behnken designs are available for experiments with 3 to 7 factors.

The figure consists of two screenshots labeled 'a.' and 'b.' showing software interfaces for design creation.

**Screenshot a. Provide Factor Details:** This window is titled 'Provide Factor Details'. It contains instructions about Box-Behnken Designs and a table for entering factor names and level values. The table has columns for 'Factor Names' and 'Level Values'. Four factors are listed: Temperature (levels 5, 15, 25), Reaction\_Duration (levels 30, 75, 120), Concentration\_A (levels 50, 75, 100), and Concentration\_B (levels 50, 75, 100). Buttons for 'Next' and 'Cancel' are at the bottom.

	Factor Names	Level Values
1	Temperature	5, 15, 25
2	Reaction_Duration	30, 75, 120
3	Concentration_A	50, 75, 100
4	Concentration_B	50, 75, 100

**Screenshot b. Design Details:** This window is titled 'Design Details'. It shows the 'Selected Design: Box-Behnken' and lists the four factors with their levels. It includes a dropdown for 'Select number of blocks' set to 1, a 'Randomize Runs' toggle switch (which is on), and buttons for 'Back', 'Create', and 'Cancel'.

Figure 6: **Make a Box-Behnken Design.** **a.** Factor details window for entering names and three equally spaced levels for each continuous factor in a Box-Behnken design. **b.** Design diagnostics window where blocking can be selected for 4- or 5-factor designs, along with the option to randomise the run order.

For a Box-Behnken design, select *Box-Behnken* from the dropdown menu. In the factor details window (see Figure 6a), you must provide a name and level values for each factor. All factors must be continuous and must have exactly three levels. These levels should ideally be equally spaced (e.g. 5, 15, 25 rather than 5, 15, 45). Once the factor details are valid, the design details page will appear (Figure 6b). For designs with four or five factors, blocking is available; in these cases, you may select the number of blocks from the dropdown menu based on the available options. As with other designs, you may also choose whether to keep the default randomised run order or disable randomisation before creating the design.

**Central Composite Designs** are response surface designs used to fit quadratic models by employing 3 levels for each factor. They are composed of three parts:

- Factorial Points - a full or fractional factorial design (typically Resolution V) that estimates the main effects and two factor interactions
- Axial Points - a set of points places along each axis (between or beyond the extremes of the factors) to quadratic effects
- Center Points - located at the midpoint of all factors; helps improve precision and estimate experimental error, while also preserving orthogonality and rotatability (which determines the number of center points)

The positioning of the axial points depends on whether the design is Circumscribed, Inscribed, or Face-Centred.

- Circumscribed - the axial points lie outside the range of the factorial design.
- Inscribed - the axial points lie within the provided upper and lower bounds and the factorial design is 'shrunk' to fit within the range to preserve rotatability.
- Face Centered - the axial points lie on the center of each face of the factorial design. These axial points do not exceed the bounds of the factorial design. This option is not rotatable.

**a. Provide Factor Details**

Central Composite designs are supported for 2 more factors. These factors must be continuous. For each factor, three levels are considered. Please provide the upper and lower limits for these levels as comma separated values, e.g., '-1, 1'. The mid level will be automatically calculated as the midpoint of the provided values.

All entries in the table below must be completed. Factor Names can only contain letters, numbers, and underscores. Do NOT use spaces or special characters in the factor names.

Note: Add or remove rows by right clicking on a row.

	Factor Names	Level Values
1	Temperature	5, 25
2	Reaction_Duration	30, 120
3	Concentration_A	50, 100
4	Concentration_B	50, 100

Next Cancel

**b. Design Details**

Selected Design: Central Composite

Factor Names	Level Values
Temperature	5, 25
Reaction_Duration	30, 120
Concentration_A	50, 100
Concentration_B	50, 100

Select Central Composite Design Type:

- Circumscribed
- Inscribed**
- Face Centered

Back Create Cancel

Figure 7: Make a Central-Composite Design. **a.** Factor details window for specifying names and upper/lower level values for each continuous factor in a Central Composite design. **b.** Design details window where the design type (circumscribed, inscribed, or face-centred) is selected, with the option to randomise the run order.

For a Central Composite design, select *Central Composite* from the dropdown menu. In the factor details window (Figure 7a), you must provide a name and the upper and lower level values for each factor. All factors must be continuous. The software will then calculate the additional levels automatically, based on the type of central composite design selected. In the design details

pane (Figure 7b), a dropdown menu allows you to choose between the Circumscribed, Inscribed, or Face-Centred variants. Blocking is not available for this design. As with other designs, you may also choose whether to keep the default randomised run order or disable randomisation before finalising the design.

### 3.4 Space Filling Designs

Space-filling designs are used to explore a wide region of the design space efficiently. The goal is to maximize coverage while ensuring the experimental runs are spread as evenly as possible across the design space. They can be particularly useful when little is known about the system or when machine learning models will be applied. In this software, space-filling designs are implemented through Latin Hypercube sampling.

**a. Provide Factor Details**

Latin Hypercube Designs are supported for any number of factors or runs. These factors can be continuous discrete, categorical, or blocks. For continuous factors, provide the upper and lower limits as comma separated values, e.g., -1, 1. For discrete, categorical, or blocking factors, provide the allowable values as comma separated values. Note: if blocking factors are included, it is recommended to choose a number of runs that is divisible by the number of blocks.

All entries in the table below must be completed. Factor Names can only contain letters, numbers, and underscores. Do NOT use spaces or special characters in the factor names.

Note: Add or remove rows by right clicking on a row.

	Factor Names	Factor Type	Level Values
1	Temperature	Continuous	5, 25
1	Reaction_Duration	Continuous	30, 120
1	Solvent	Categorical	Ethanol, Acetone, Octane, DCM
1	Number_Additions	Discrete	2, 4, 8
1	Concentration_A	Continuous	50, 100

**b. Design Details**

Selected Design: Latin Hypercube Sampling

Factor Names	Factor Type	Level Values
Temperature	Continuous	5, 25
Reaction_Duration	Continuous	30, 120
Solvent	Categorical	Ethanol, Acetone, Octane, DCM
Number_Additions	Discrete	2, 4, 8
Concentration_A	Continuous	50, 100

Number of Experimental Runs:  ⓘ

Blocking is not available for this design

Randomize Runs:

Back Create Cancel

Figure 8: Make a Latin Hypercube Design. **a.** Factor details window for entering names, types, and level values in a Latin Hypercube design. Continuous factors require upper and lower limits, while discrete and categorical factors require all allowable values. **b.** Design details window for selecting the number of runs (at least one more than the number of factors), with the option to randomise the run order.

**Latin Hypercube Sampling** is a statistical method for creating designs that efficiently explore the design space with a set number of runs. Each factor is sampled uniformly across its range, ensuring all portions of the design space are represented. This allows for broad and balanced coverage of the design space with fewer runs than classical designs. This make Latin Hypercube designs very useful for exploring complex, high-dimensional systems when other designs would require an infeasible number of runs. Note: Because this method incorporates an element of randomness, two designs generated from the same inputs may differ; however, each design is still optimised for balance and coverage.

For a Latin Hypercube design, select *Latin Hypercube* from the dropdown menu. In the factor details window (Figure 8a), you must provide a name, type, and level values for each factor. For continuous factors, you must supply the upper and lower limits; for discrete or categorical factors, you must list all allowable values. Once valid factor details are provided, the design details page will appear (see Figure 8b). Here, you select the total number of runs for the design. The number of runs should be at least one greater than the number of factors included. You may also choose whether to keep the default randomised run order or disable randomisation before finalising the design.

### 3.5 Custom Designs

Custom designs are flexible designs that tailor the experimental runs to the specific model you want to fit. Instead of following a fixed template, you define the terms in the model and a custom design is constructed to maximise the statistical information for it. They are particularly useful when you have a mix of factor types, a limited number of runs that can be performed, or other restrictions that render classical designs unsuitable. DoEIY offers custom designs using the D-Optimal methodology.

**D-Optimal Designs** are built around a specified model that includes the effects and interactions you wish to estimate. You specify which model terms to include (main effects, interactions, quadratic terms, or custom combinations), and the D-Optimal algorithm will generate a set of runs optimised to estimate those terms as efficiently as possible. This software supports D-Optimal designs for any number of factors, levels, and factor types, including continuous, discrete, categorical, and blocking factors.

For a D-optimal design, select *D-optimal* from the dropdown menu. In the factor details window (Figure 9a), you must provide a name, type, and level values for each factor. For continuous factors, this means specifying the upper and lower limits, for discrete and categorical factors, you must list all allowable values. Once factor details are complete, the design details page will appear (see 9b). Here, you select the model components to include, such as main effects, two-factor interactions, and quadratic terms. You may also define custom interactions between specific factors if required. Based on these choices, the software will display the minimum and maximum number of runs allowed. You must then specify a number of runs that falls within this range before creating the design.

## 4 Entering and Editing Results

Once you have finalised your experimental design, the **Enter/Edit Results** tab will display the design matrix in an editable table (see Figure 10). This is where you enter the responses (experimental outcomes) for each run of the design. Each row corresponds to a run from your chosen design, and each column corresponds to either a factor or a response.

You can type your measured responses directly into the table. For example, if your response is yield, particle size, or another experimental measurement, simply enter the values in the appropriate response column next to the corresponding run.

### 4.1 Saving and Loading Designs

- The **Save** button stores your design and its factor metadata in a `.csv` file. This file can later be reloaded to continue your work.

**a.** **Provide Factor Details**

D-Optimal Designs are supported for any number of factors, which can be continuous discrete, categorical, or blocks. For continuous factors, provide the upper and lower limits as comma separated values, e.g., -1, 1. For discrete, categorical, or blocking factors, provide the allowable values as comma separated values. Note: if blocking factors are included, it is recommended to choose a number of runs that is divisible by the number of blocks.

All entries in the table below must be completed. Factor Names can only contain letters, numbers, and underscores. Do NOT use spaces or special characters in the factor names.

Note: Add or remove rows by right clicking on a row.

	Factor Names	Factor Type	Level Values
1	Temperature	Continuous	5, 25
1	Reaction_Duration	Continuous	30, 120
1	Solvent	Categorical	Ethanol, Acetone, Octane
1	Number_Additions	Discrete	2, 4, 8
1	Concentration_A	Continuous	50, 100

Next Cancel

**b.** **Design Details**

Selected Design: D-Optimal

Factor Names	Factor Type	Level Values
Temperature	Continuous	5, 25
Reaction_Duration	Continuous	30, 120
Solvent	Categorical	Ethanol, Acetone, Octane
Number_Additions	Discrete	2, 4, 8
Concentration_A	Continuous	50, 100

For a D-Optimal design, it's essential to identify the effects you wish to estimate in advance. Please use the interface below to select all relevant effects, including main effects, interactions, quadratic terms, and any other higher-order effects.

Main Effects    2-Way Interactions    Quadratics

Add    Create Interaction    Remove

Factor: Temperature, Reaction\_Duration, Solvent, Number\_Additions, Concentration\_A

Selected Effects: Concentration\_A, Number\_Additions, Reaction\_Duration, Solvent, Temperature, Number\_Additions:Concentration\_A, Reaction\_Duration:Concentration\_A, Reaction\_Duration:Number\_Additions

Number of Runs: 40    Min Runs: 32    Max Runs: 72

Randomize Runs:

Back Create Cancel

Figure 9: **Make a D-Optimal Design.** **a.** Factor details window for entering names, types, and level values in a D-optimal design. Continuous factors require upper and lower limits, while discrete and categorical factors require all allowable values. **b.** Design details window for selecting model components (e.g. main effects, two-factor interactions, quadratics) and any custom interactions. The software then displays the minimum and maximum number of runs allowed, and you must choose a run size within this range.

	Reducing\_Agent\_Ratio	Reaction\_Time	Entropy
1	0	0	1.37
2	1	-1	3.54
3	-1	-1	3.05
4	-1	0	2.76
5	0	-1	2.46
6	1	0	1.75
7	1	1	1.46
8	0	0	1.66
9	-1	1	2.37
10	0	1	1.81
11	0	0	1.68

 ©2025 AdReNa Version: 0.2.0
 

Figure 10: **Enter/Edit Results.** The *Enter/Edit Results* tab, where you can enter responses, save or load designs, and edit rows or columns in the design matrix.

- The **Load** button restores a previously saved design into the table for further editing or analysis.

**Important:** The DoEIY webapp does not auto-save and will time out after 30 minutes of inactivity. To avoid losing your work, save your progress frequently. Editing the `.csv` file outside the webapp is discouraged; the format of the file must be maintained for compatibility and external changes may make it impossible to reload correctly.

## 4.2 Editing the Table

When a design is loaded, additional editing options appear:

- **Remove Column** – deletes a column from the table.
- **Add Column** – creates a new column (e.g. if you want to add an additional response).
- **Update Column** – allows you to change a column’s name, type (numeric or categorical), or its position in the table.

For editing rows, right-click on any row to open the context menu:

- **Insert above** – add a new row directly above the selected one.
- **Insert below** – add a new row directly below the selected one.
- **Remove** – delete the selected row.

This editing functionality allows you to flexibly manage your dataset as you collect experimental results.

## 5 Analyze the Design

Once responses have been entered, you can move to the **Analyze the Design** tab. This tab contains three subtabs: **Create Model**, **ANOVA**, and **Design Diagnostics**. These tools allow you to specify a model, evaluate its significance, and check for design limitations.

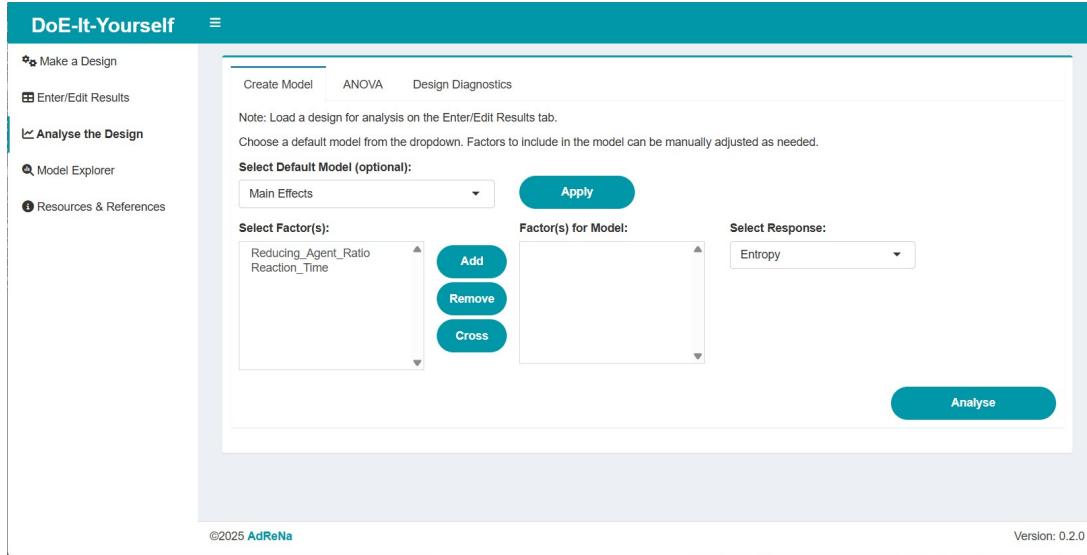


Figure 11: **Analyze the Design**. The *Analyze the Design* tab, which contains the *Create Model*, *ANOVA*, and *Design Diagnostics* subtabs used to specify, fit, and evaluate the model.

### 5.1 Create Model

In this subtab you specify the structure of the statistical model. You can choose one of the default model structures:

- **Main Effects** — includes only the main factors.
- **Main Effects + Interactions (2-way)** — includes main factors and all two-factor interactions.
- **Response Surface** — includes main factors, two-factor interactions, and quadratic terms.

Click **Apply** to populate the model term list. You may also build a custom model by using **Add**, **Remove**, or **Cross** (to generate interactions).

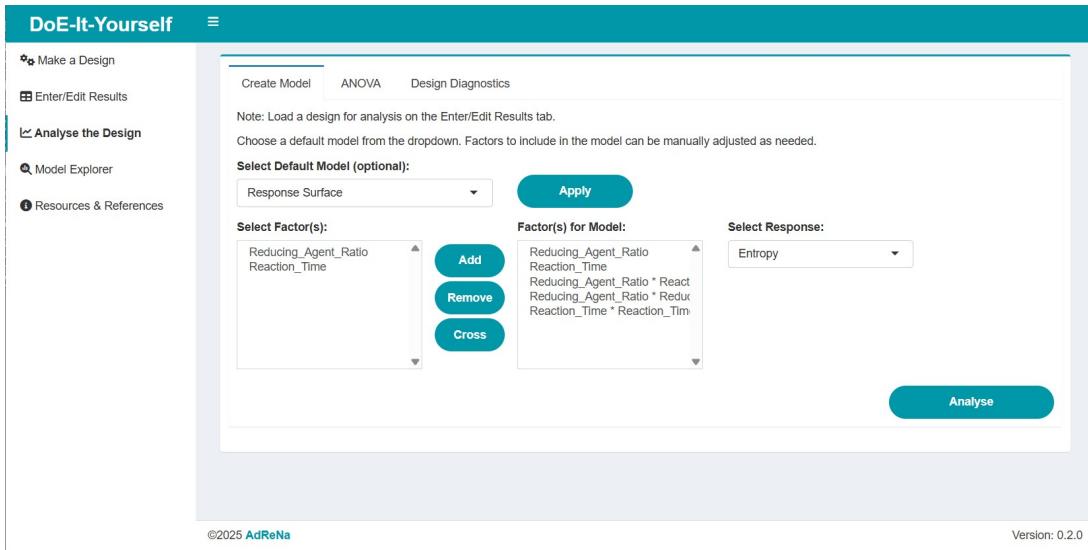
Next, select the response variable you wish to model and click **Analyze**. This fits the model and moves you automatically to the ANOVA subtab.

*How the model is fitted:* Models are fitted by ordinary least squares regression using R's `aov()` function. Analysis of variance (ANOVA) is then applied to partition variance and test the significance of the terms.

### 5.2 ANOVA

The **ANOVA** subtab reports model fit and term significance. Outputs include:

- **A parity plot** (predicted vs. observed values).



**Figure 12: Model Specification.** The *Create Model* subtab, showing the default-model dropdown, model term lists, and controls for adding, removing, or crossing terms.

- Summary metrics:  $R^2$ , adjusted  $R^2$ , the  $F$ -statistic with degrees of freedom, and the overall  $p$ -value.
- An **ANOVA table** listing degrees of freedom, sum of squares, mean square,  $F$ -statistic, and  $p$ -values. Significance is indicated with asterisks.
- An **Estimates table** showing coefficients, standard errors,  $t$ -statistics, and  $p$ -values, again with asterisks marking significance.

### 5.3 Design Diagnostics

The **Design Diagnostics** subtab provides tools to evaluate the quality of the fitted model.

- **Alias structure:** identifies terms that cannot be distinguished due to the design (aliased terms).
- **Correlation heatmap:** visualises collinearity among model terms to assess whether effects can be estimated independently.

## 6 Model Explorer

The **Model Explorer** tab allows you to interact with your fitted model. It contains two subtabs: **Visualizer** and **Interact**. These tools help you explore predicted responses, factor relationships, and optimisation settings.

### 6.1 Visualizer

The **Visualizer** subtab displays plots of the response against each factor. Sliders allow you to adjust factor values interactively. If the model includes interactions, adjusting one factor updates the plots for interacting factors. The predicted response at the current slider settings is displayed, helping you understand how factor changes affect the outcome.

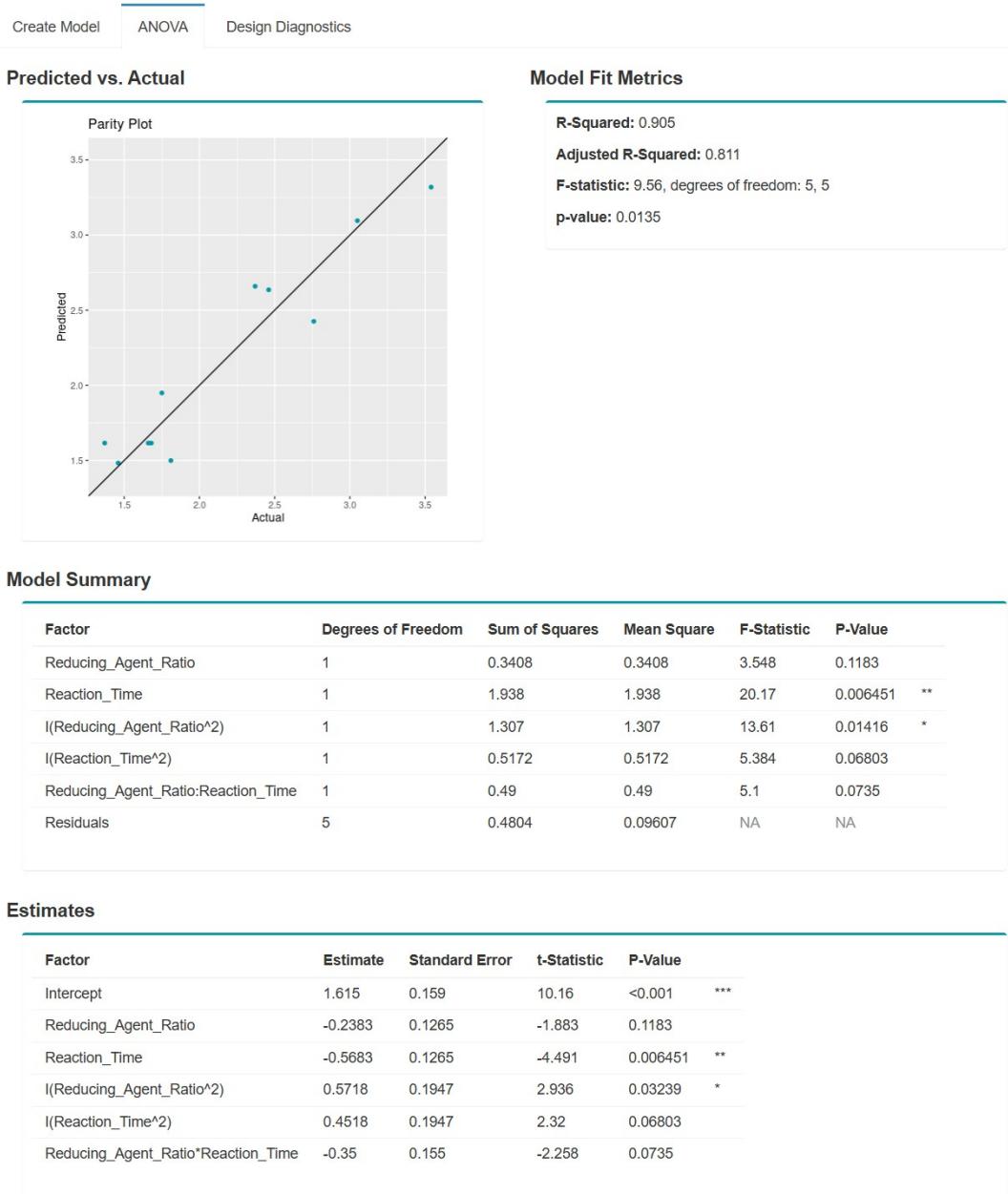


Figure 13: **ANOVA Results.** The ANOVA subtab, showing the parity plot, model fit metrics, and excerpts from the ANOVA and Estimates tables.

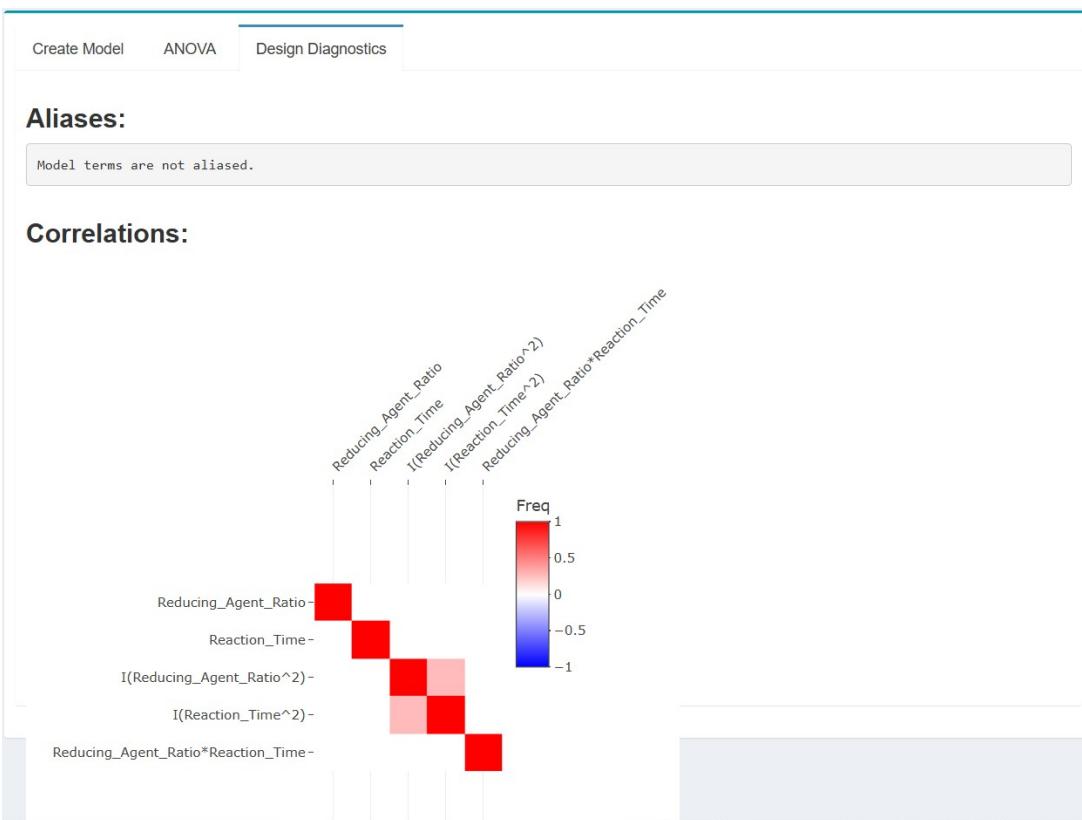


Figure 14: **Diagnostics.** The *Design Diagnostics* subtab, showing an alias report and a correlation heatmap.

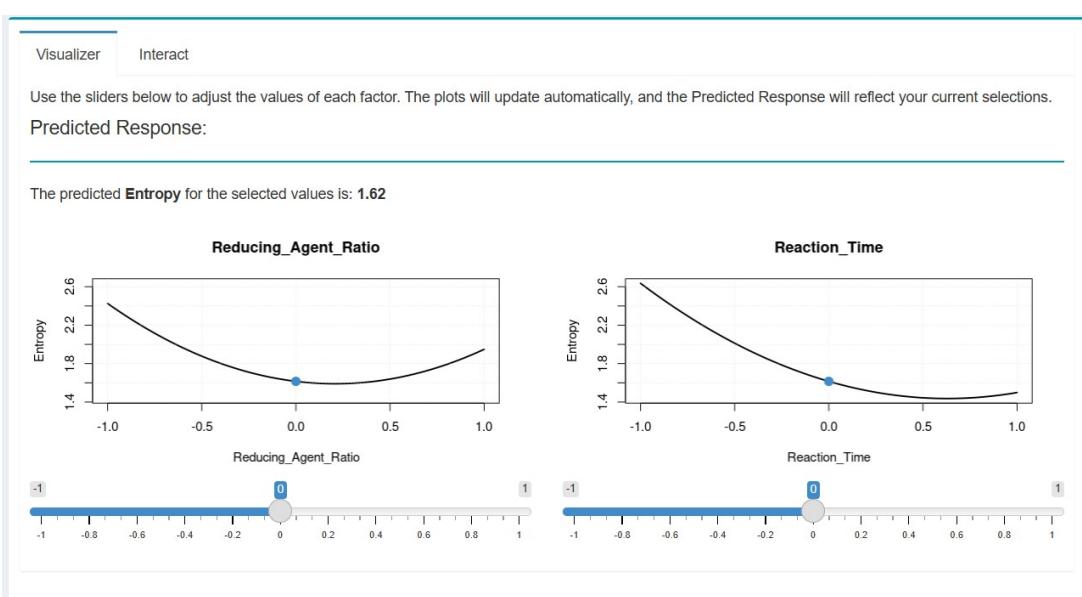


Figure 15: **Factor-Response Visualizer.** The *Visualizer* subtab, showing factor-response plots with sliders and a predicted response read-out.

## 6.2 Interact (Optimisation and Targeting)

The **Interact** subtab lets you optimise or target a response value.

- Choose a goal: **Maximise**, **Minimise**, or **Find Target**.
- For **Find Target**, enter the target response value.
- Optionally provide starting values for the factors, which can help when multiple solutions exist.
- Click **Run** to obtain factor settings and the predicted response.

*Optimisation engine:* The search uses R's `optim()` function. No additional optimisation parameters are exposed in the interface.

The screenshot shows the 'Interact' subtab of a software application. At the top, there are two tabs: 'Visualizer' and 'Interact'. The 'Interact' tab is selected. Below the tabs, the title 'Select Optimization Objective' is displayed. Underneath it, the 'Goal:' section contains three radio buttons: 'Maximize' (unchecked), 'Minimize' (checked), and 'Find Target' (unchecked). The next section is 'Starting Values:', which includes two input fields: 'Start Reducing\_Agent\_Ratio' with the value '1.25' and 'Start Reaction\_Time' with the value '75'. A large blue 'Run' button is located below these fields. At the bottom of the subtab, under 'Resulting Conditions:', the following values are listed: 'Reducing\_Agent\_Ratio: 1.59107', 'Reaction\_Time: 111.227', 'Predicted Entropy:', and 'Predicted Response: 1.33'.

Figure 16: **Interact Optimisation.** The *Interact* subtab, showing goal selection, optional starting values, and the returned solution.

## 7 DoE Glossary

Before using DoEIY, it is helpful to review some core DoE terminology:

- **Factors** are the variables you choose to study and deliberately control, such as temperature, solvent choice, or concentration.
- **Levels** are the values for each factor that are investigated. For example, for temperature, its levels could be low = 5°C, medium = 15°C, and high = 25°C.
- The **Design Space** is the multi-dimensional region defined by the factors in your experiment and the ranges over which they are studied. Each factor contributes one dimension, and together they form the boundaries within which all experimental runs are planned.

- A **Response** is the measured output(s) of interest, e.g., yield or particle size
- **Blocking** is a way to group experimental runs to account for known sources of variability, such as different days, batches, or equipment.
- An **Interaction** is when the effect of one factor on the response changes depending on the level of another factor.
- A **Run** or experimental run is a single execution of the experiment at a specific combination of factor levels.
- **Aliasing** occurs when the effects of two or more factors (or their interactions) cannot be distinguished from each other because of the structure of the experimental design. In other words, the influence of one factor is "mixed up" with another, making it impossible to separate their individual contributions to the response. For example, in a fractional factorial design, a main effect might be aliased with a two-factor interaction. This means any observed change in the response could be due to either the factor itself or the interaction, but the design does not provide enough information to tell them apart. The **Alias Structure** of a design is the complete map of which effects are aliased with each other. It shows, for every factor and interaction in the model, whether it is uniquely estimable or combined with another effect.
- **Correlation** describes how strongly two model terms vary together. A correlation close to +1 means the terms move in the same direction, while a correlation close to -1 means they move in opposite directions. A value near 0 indicates little to no linear relationship. High correlations signal collinearity and the model may have difficulty distinguishing the individual contribution of each term to the response. Tip: While low correlations are generally preferred, not all high correlations indicate aliasing. For example, if you include both a factor A and its quadratic term  $A^2$ , these terms are mathematically related and will be highly correlated, even though they are not aliased. In this case, the correlation map is highlighting a natural dependency, not a flaw in the design.
- **ANOVA** is a statistical method used to separate the total variability in your data into two parts: the variation explained by the model (due to factors and their interactions), and the residual variation (unexplained noise). By comparing these sources of variation, ANOVA tests whether each model term contributes significantly to the response. The results are presented in a table showing, for each term, its degrees of freedom, sum of squares, mean square, F-statistic, and p-value. ANOVA helps you determine which factors and interactions have a meaningful effect, and whether your overall model provides a good fit to the data.
- **Rotatability** is a property of a response surface design that ensures the precision of predictions depends only on the distance from the design centre, not on the direction. In other words, all points the same distance from the centre have the same variance in the prediction.
- **Orthogonality** is a property of an experimental design that ensures the estimates of different model terms (such as main effects and interactions) are uncorrelated with each other; this means each factor's effect can be assessed independently.