DESIGN CREATOR FRONT-END USER MANUAL

Note: This user manual is a modified excerpt from the appendix in MacCalman (2013) and refers to chapters in this dissertation:
MacCalman, A. D. 2013. Flexible Space-Filling Designs for Complex System Simulations. Doctoral dissertation. Monterey, CA, Naval Postgraduate School.

This document is a user manual of the Front-End Tool in the DesignCreatorv2.xlsm file used to run our genetic algorithm. The purpose of the tool is to allow the user to create a custom design, with a specified number of design points and number of factors, by type, number of levels with the desired balance, and the model terms included in the regression matrix. In addition, the user can start the algorithm with an existing design and add columns to it; this allows us to leverage the cataloged 2nd Order NOLH designs that are included in the workbook by adding columns to them. Once the algorithm creates the design, there are some utilities available that will create a spreadsheet to translate a design, create higher-order terms, calculate the maximum absolute pairwise correlation, and create dummy variables for categorical factors.

The algorithm was written in Java™ 2 and requires the user to ensure that the Java Platform (JDK) is downloaded on their computer; visit the Oracle website at http://www.oracle.com/technetwork/java/javase/downloads/index.html to download. You can download the tool from the SEED Center website at http://harvest.nps.edu/software.html. Once downloaded, there will be two files: DesignCreator.xlsm (containing the Front-End Tool with utilities) and DOE.jar (the executable .jar file written in Java). Ensure that these files are saved to the same folder. If you are on a shared network computer we do not recommend that you save the files to the desktop. When opening the DesignCreator.xlsm file, the user must enable the macros in order to utilize the buttons throughout the workbook. The Front-End Tool will create an input.csv file and a runit.bat file (for Windows computers) or runit.txt file (for Macintosh computers) and save them to the same folder; these are the files the DOE.jar file needs to execute the algorithm from the Windows computer Command line or the Macintosh computer Terminal window.

Once the algorithm is complete, the output design will be saved as a .csv file in the same folder the DOE.jar file is in. The output file title name will have the number of rows, columns, the ρ_{map}, ML_2, and the initial seed used for the random number generator (see Chapter II for the definition of ρ_{map} and ML_2). In the .csv file, the first four rows will contain the following, respectively: the factor type, the number of levels, the model terms included in the regression matrix, and the factor name, x_i, where i is the column number. If there are discrete or categorical factors in the design, the last row, separated by the word "balance," will have the factor's balance metric indicating the spread of the levels across the design points; see Chapter VI for the definition of balance. As a general rule, the user should never delete or change any of the worksheet names in the DesignCreator.xlsm file. Each section in this appendix describes the worksheets in the DesignCreator.xlsm file and provides instructions where appropriate.

readme

The *readme* worksheet provides the purpose of the tool, explains how to create designs and use the utilities. In addition, it references literature that pertains to the designs created by the genetic algorithm.

glpl

The worksheet describes the terms of the GNU Lesser General Public License as published by the Free Software Foundation, either version 2.1 of the License or (at your option) any later version. This license ensures that the algorithm is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY, without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.

Front End

Input Parameter Settings

The *Front End* worksheet allows the user to enter the genetic algorithm input parameters. The blue-colored cells are the factor entry area used to specify the number of factors, by type, number of levels, and the model terms included in the regression matrix for the ρ_{map} calculation. The four types of factors are: continuous, discrete, categorical, and binary. For continuous factors, the number of levels must be equal to whatever is set as the "Number of Design Points" parameter in the green-colored entry area. For categorical and binary factors, only the main (linear) terms can be added to the regression matrix (model terms must be set to "M.") Binary factors can only have the number of levels set to 2. Generally, the user should set the highest-order model terms in the first set of rows. The model term designations are the following: M for main effects; MQ for main and quadratic effects; MI for main and two-way interactions; and MQI for main, quadratics, and two-way interactions. The model terms order, from highest to lowest, are MQI, MI, MQ, and M. The model term designations significantly impact the algorithm run time. The Column labeled "Minimum Feasible Imbalance" shows the minimum analytically achievable imbalance for all discrete and categorical factors. To see these imbalance values, press the "Calculate Balance Feasibility" button. See the Balance Check worksheet for guidance on the number of design points needed to achieve a desired imbalance amount for a given number of levels. Figure A1 shows a snapshot of the factor entry area in the *Front End* worksheet.

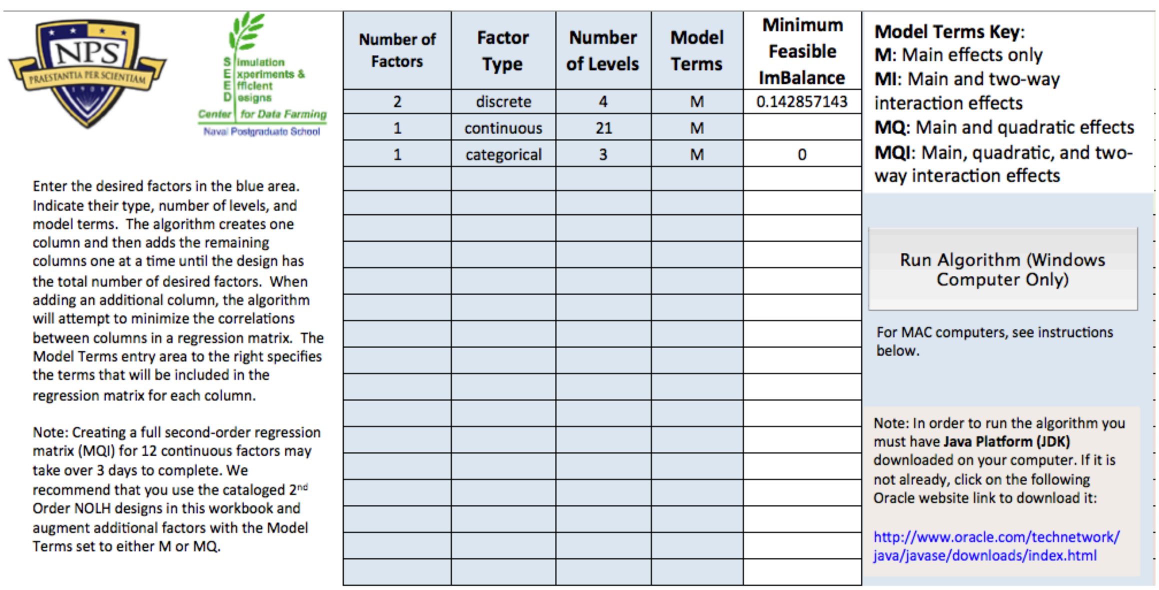

Figure A1. Factor entry area in the *Front End* worksheet.

The red-colored cells are the algorithm's internal input parameters that will not be of interest to the general user of the design creator. Chapter IV discusses the experimental designs we performed to determine the appropriate input parameter settings for design searches. The user can change these input settings, if desired (see Chapter III for the algorithm steps and definitions of input parameters), and can restore the default settings by pressing the macro button underneath the red-colored cell area. Changing these internal input parameter settings will impact the algorithm's performance and run-time length; see Chapter IV for guidance on the performance and run-time length for different number of design points and columns with the default internal parameter settings. For designs that are not difficult to minimize the ρ_{map}, we recommend setting the number of trials (*numTrials*) equal to 1 in order to speed up the algorithm's run time.

The green-colored cells are the input parameters the general users will need to set each time they run the algorithm. Because the algorithm is run as a batch file from the Command or Terminal window, the user may decide to increase the number of algorithm instantiations that will be executed. Setting the "Number of Batch Replications" parameter to greater than 1 will allow the user to send a batch file to a computer cluster to perform multiple replications of the algorithm. Because of the stochastic nature of the algorithm, we recommend performing multiple replications when searching for efficient designs and then selecting the design with the smallest ρ_{map}. If the user does not intend to send a batch file to a computer cluster, he/she can run the algorithm multiple times in separate Command/Terminal windows. The "Number of Design Points" parameter is the number of experiments or rows in the desired output design matrix. The "Start With Design" boolean parameter lets the algorithm know whether to add the desired factors entered in the blue-colored cell area to an existing design located in the *Start Design* worksheet. When the "Start With Design" parameter is set to TRUE, ensure that the "Number of Design Points" parameter is set to the same number of rows in the design that is pasted into the *Start Design* worksheet. The "Jiggle Operations" boolean parameter lets the algorithm know whether to perform the jiggle operations on the continuous factors (see Chapter III for a description of the jiggle operations). If the algorithm starts with an existing design, the jiggle operation will only be performed on the newly added continuous columns. The "Show Comments" boolean parameter lets the algorithm know whether to show the comments in the Command/Terminal window during the algorithm's execution. When sending a batch file to a computer cluster, the "Show Comments" parameter should be set to FALSE. Figure A2 shows a snapshot of the input parameter entry area in the *Front End* worksheet.

Input Parameter	Setting	Description
Number of Batch Replications	1	The number of command line batch replications written to the batch file.
Number of Design Points	20	The number of rows in the design matrix. Each row designates the factor settings for each experiment.
Start With Design	FALSE	TRUE means that the algorithm will add columns to the design that is pasted into the Start Design worksheet. FALSE means that the algorithm will create a new design.
Perform Jiggle Operations	TRUE	TRUE means that the algorithm will perform the jiggle operation, FALSE means that it will not. The jiggle operation will not be performed on columns in the Start Design worksheet.
Show Comments	TRUE	TRUE means that the algorithm comments will be displayed in the command/termainal window. Set to FALSE when sending batch files to a high performance computer cluster.
numExploreGen	100	Number of exploration generations.
numExploitGen	200	Number of exploitation generations.
popSize	100	Size of the population of candidate columns.
copyPortion	0.1	Portion of candidate columns copy into the next generation.
halfWidth	0.5	The bounded distance that prevents the jiggle operator for perturbing outside a range.
numJigGen	100	Number of jiggle generations.
numTrials	3	Number of exploration trials each consisting of a set of exploration generations with its own initial population of candidate columns.
swapPortion	0.2	Portion of design points swapped during a swap operation.
poolSize	100	Size of the pool that contains a set of candidate columns.
genExitCriteria	20	Number of generations performed without improvement of the fitness function.
jigglePortion	0.2	Portion of design point jiggled during a jiggle operation.
colAttempts	3	Number of attempts to find a column with a new initial population of solutions if an attempt did not meet the maximum correlation threshold.
jigglePasses	3	Number of times the jiggle operator is performed on the columns.
corrThreshold	0.05	The maximum correlation a column threshold must be before added to the design. The algorithm will continue to find a column to add to the design for a set number of attempts (*colAttempts*).

Figure A2. Input parameter entry area in the *Front End* worksheet.

Algorithm Execution

Once the input parameters are set, the steps to execute the algorithm will depend on the type of operating system on your computer (Windows or Macintosh). For Windows computers, simply press the "Run Algorithm" macro button; each time you press this button, a new Command line window will open and run a different instantiation of the algorithm. Macintosh computers must run the algorithm from the Terminal window, with the current directory set to the file location where the DesignCreator.xlsm and DOE.jar files are saved. The first step is to press the "Create Flat Files" macros button. Then, open the Terminal window and change the directory to where the algorithm is saved. At the Terminal Command prompt, type the following:

. ./runit.txt

To run additional algorithm instantiations simultaneously, open a new Terminal window and repeat the above steps. To open the Terminal window from the Finder, the user can go to System Preferences and click on "Keyboard," select the "Keyboard Shortcuts" tab and click "Services" from the left menu; scroll down on the right and check the box next to "New Terminal at Folder." Setting this preference will allow the user to right click on a folder in the Finder and click "New Terminal at Folder" to open the Terminal at the desired folder. This preference setting will save the user from having to change the directory manually to where the algorithm is located each time you open the Terminal window.

When the "Show Comments" parameter is set to TRUE, the comments shown in the Command or Terminal window reveal the progress of the algorithm. Figure A3 shows a Command line window that searched for a three continuous factor 2^{nd} order design with 20 design points. The algorithm performed three exploration trials (*numExploreGen* = 3) and three jiggle generation passes (*jigglePasses* = 3). The final time shown at the bottom of Figure A3 is in hours.

```
Y:\trunk\Dissertation\FrontEnd>java -jar DOE.jar 20 False True True 100 200 100
0.1 0.5 100 3 0.2 100 20 0.2 3 3 0.05
design Points: 20 seed: 2149891

1 column, continuous factor type, 20 discreteLevels, mode: MQI, 1 columnAttempt,
 designSize: 1
1 exploration trial correlation: 0.0030075187969924814
2 exploration trial correlation: 0.0035120253120068706
3 exploration trial correlation: 0.0038094149848396284
best from exploration generations: 0.0030075187969924814 time elapsed: 0.0661632
0333333334 minutes
best from exploitation generations: 0.0030075187969924814


2 column, continuous factor type, 20 discreteLevels, mode: MQI, 1 columnAttempt,
 designSize: 2
1 exploration trial correlation: 0.04602415128730918
2 exploration trial correlation: 0.02857142857142857
3 exploration trial correlation: 0.0298522151520584
best from exploration generations: 0.02857142857142857 time elapsed: 0.209332816
66666667 minutes
best from exploitation generations: 0.02857142857142857

1 jiggle pass:
2 jiggled column: 0.013533527345797927 original column: 0.02857142857142857
1 jiggled column: 0.00662377093987147 original column: 0.013533527345797927
0 jiggled column: 0.005520766537884698 original column: 0.013516121050120719

2 jiggle pass:
2 jiggled column: 0.005520766537884698 original column: 0.005520766537884698
1 jiggled column: 0.003323758202155466 original column: 0.006572470707733691
0 jiggled column: 0.003943008496000514 original column: 0.005520766537884698

3 jiggle pass:
2 jiggled column: 0.00394300849600049 original column: 0.00394300849600049
1 jiggled column: 0.002658994167946492 original column: 0.00394300849600049
0 jiggled column: 0.0034261573867559723 original column: 0.005472171446045812

design size: 3, maximum correlation: 0.003426157386755961, time: 0.0191719396666
6667,ML2: 0.006221993446700047,seed: 2149891,

Y:\trunk\Dissertation\FrontEnd>
```

Figure A3. Command line window during the algorithm execution.

Balance Check

This worksheet calculates the minimum analytically achievable imbalance for a given number of discrete or categorical factor levels. Use this worksheet to help decide how many design points you need to ensure the design's imbalance is minimized. The algorithms will attempt to find discrete or categorical factors with an imbalance < 0.1. After 50 attempts, if the algorithm did not find a column, then it will return the column with the lowest imbalance. There are some design point and level combinations that cannot achieve a 0.1 balance. This worksheet will help guide which design points are feasible for a given number of factor levels.

Cataloged Designs

This worksheet has hyperlinks that will navigate the user to other worksheets that contain the cataloged 2^{nd} Order NOLH design. Once there, the user can press the macro button to automatically copy the design into the *Start Design* worksheet. We recommend using these cataloged designs for up to 12 continuous factors when you can afford to perform the number of experiments needed for each design. When the user desires to add discrete factors to a set of continuous factors (up to 12), with the model terms set to "MQI" (for a full second-order model), we recommend copying a cataloged design to the *Start Design* worksheet and then deleting two continuous columns for every one discrete factor (this is only a rule of thumb). Adding additional columns (of any type) to the cataloged designs, with the model terms set to "M" or "MQ" do not require that you delete continuous columns.

Start Design

If the user desires to add additional columns to an existing design, paste the design into this worksheet and set the "Start Design" parameter to TRUE in the *Front End* worksheet. The first row designates the factor type. Ensure one of the following text entries is in each column in the first row: continuous, discrete, binary, or categorical. Specify the number of levels for the factor in the second row. For continuous factors, the algorithm does not care what is entered because the number of levels for a continuous factor is always the number of design points. The third row contains the model terms (M, MI, MQ, and MQI). These entries have no impact to the algorithm. The fourth row is reserved for the factor name. Ensure that the design (with the first four rows) is pasted into cell B1.

Coded Design

Paste a design with the first four row entries as indicated in the *Start Design* worksheet instructions into cell B1. If there are discrete or categorical factors in the original .csv output file, be sure not to paste the word "balance" and the balance metric into this worksheet. Also, avoid pasting empty cells that may get highlighted after selecting the current region in the .csv output file. Press the "Create Translation Worksheet" macro button to create a formula worksheet that will allow the user to translate the coded design point levels to the factors range desired for the experiments. To calculate the ML_2 and ρ_{map} metrics, press the "Insert Design into Design Tools Worksheet" macro button. If the design has categorical factors and the user wants to examine the first-order correlations of the design with the categorical dummy variables, press the "Insert Design into Categorical Design Worksheet."

Translated Design

After pressing the "Create Translation Design" macro button in the *Coded Design* worksheet, the macro will insert the formulas into the cells that will allow the user to translate the design to the desire factor ranges. The blue-colored cells are copies of the first three rows from the *Coded Design* worksheet (factor type, number of levels, and model terms). For continuous factors, enter the low and high setting for each factor. Users have the option to round the continuous factor to a discrete factor; however, we do not recommend doing this. Rounding a continuous factor is an old technique to create discrete factors but can severely impact the ρ_{map} of the original design (especially the 2^{nd} Order ρ_{map}). We should not have to round a continuous factor anymore because our algorithm is capable of creating designs with discrete factors for a specified number of levels. If the factor column is discrete, the sixth row allows the user to scale the column instead of rounding. Scaling a discrete factor to a number greater than 1 will spread the discrete levels over a wider range of values. If the factor type is either discrete or categorical, the high level will be protected and will add the number of levels to the low-level setting. The yellow-colored cells are protected to ensure the user does not change the translation formulas. After establishing the low and high levels and naming the factors, the user can copy and paste special values the translated design into another spreadsheet for their experiment.

Design Tools

After pressing the "Insert Design into Design Tools Worksheet" macro button in the *Coded Design* worksheet, the design will appear (with the factor names only in the first row) in cell B1. The available macro buttons allow the user to calculate the ML_2 space-filling metric; center the design by subtracting the mean; create the quadratic terms; the second-, third-, and fourth-order terms; calculate the ρ_{map}, and calculate the distribution of all absolute pairwise correlations. Before you create the higher-order terms, you must ensure that you center the design first; otherwise, the main factors will be highly correlated with its own quadratic. Be sure to only press the higher-order macros button once; otherwise, the macro will expand out the terms with whatever is currently in the worksheet. Delete the high-order terms in the worksheet if you desire to recreate a different set of higher-order terms. When the user presses the "Collect and Sort Abs Corr Distribution" macro button, the distribution of all absolute pairwise correlations of whatever design is currently in the worksheet will get pasted and sorted into the *Abs Corr Distro* worksheet.

Abs Corr Distro

After pressing the "Collect and Sort Abs Corr Distribution" macro button, the absolute pairwise correlation distribution will get pasted and sorted into this worksheet.

Categorical Design

After pressing the "Insert Design into Categorical Design Worksheet" macro button in the *Coded Design* worksheet, the design will appear in cell B1. From here, the user can designate the dummy variable convention before creating the dummy variables (see Chapter VI for a description of the different dummy variable conventions). After pressing the "Create Dummies" macro button, a new design will get pasted into the *Dummy Variables* worksheet with all the categorical factors converted into the set of dummy variables determined by the number of levels.

Dummy Variables

This worksheet will contain the design with dummy variables after pressing the "Create Dummies" macro button in the *Categorical Design* worksheet. Pressing the "Find First-Order Correlation Distro with Dummy Variables" macro button will paste and sort the absolute pairwise correlation distribution into the *Abs Dummy Corr Distro* worksheet.