Choosing Between Several Designed Experiments based on Multiple Design Performance Criteria

Christine Anderson-Cook, Los Alamos National Laboratory

Abstract

When selecting a best design for an experiment, it is beneficial to consider what the objectives of the experiment are and to use several criteria that match these objectives to quantify differences between the potential design choices. JMP software provides multiple options for creating designs to be compared, and a rich set of design summaries that allow an experimenter to understand the trade-offs between candidate designs and make a judicious choice appropriate for that experiment.

This paper demonstrates the process of constructing and evaluating designs for a specific screening design example using the Pareto front and Data Filter approaches, and describes how this process could be generalized to other design of experiment and response optimization settings.

1. Introduction

In design of experiments, the quality of a design can be measured by several criteria, each representing different aspects of performance. For example, the D- criterion is useful for evaluating the overall precision of the estimated model parameters whereas the I- and G- criteria focus on minimizing prediction variance in the design region. Other criteria reflect a design's robustness to model mis-specification in terms of the degree of bias in the estimated coefficients or experimental error variance. Good designs should balance precision in parameter estimation or prediction when the model is correct with adequate bias protection when the specified model is incorrect (Anderson-Cook, 2010).

Consider a real case (Lu, Anderson-Cook and Robinson, 2011) in which researchers wish to select a designed experiment for studying the relationship between the response and five potentially active factors (A-E), each at two levels (+1 and -1). Due to tight cost restrictions, the experimenters can only collect 14 observations. Prior scientific knowledge indicates that the two-factor interactions AB, AC, BD, and CE are most likely to exist, with the remaining two-way interactions less likely to be active or important. Hence the researchers want to select an optimal 14-run design for estimating a model with all five main effects and interactions AB, AC, BD, and CE, written here as

$$Y = \beta_0 + \beta_A A + \beta_B B + \beta_C C + \beta_D D + \beta_E E + \beta_{AB} AB + \beta_{AC} AC + \beta_{BD} BD + \beta_{CE} CE + \varepsilon. \quad (1)$$

The objective is to estimate the ten model parameters $(\beta_0, \beta_A, \beta_B, \cdots, \beta_{CE})$ as precisely as possible, while limiting the impact of potential model misspecification (from some of the other interaction terms being present) on the estimation of coefficients and experimental error variance.

2. Metrics to Evaluate the Designs

We now present some details about the three metrics considered. When the focus is on precise parameter estimation, D-optimality is the most popular design optimality criterion due to its ease of implementation and computation. The D-criterion is defined as $|\mathbf{M}(\xi)| = |\mathbf{X}'(\xi)\mathbf{X}(\xi)| / N^p$, where $\mathbf{X}(\xi)$ is the $N \times p$ model matrix for design ξ and $\boldsymbol{\beta}$ is the associated $p \times 1$ vector of regression coefficients. The D-optimal design maximizes $|\mathbf{M}(\xi)|$ over the design space spanned by all possible designs, $\xi \in \boldsymbol{\Omega}$. The D-efficiency of a particular design, ξ^*, is given by

$$D_{\text{eff}}(\xi^*) = \left(|\mathbf{M}(\xi^*)| / \underset{\xi}{\text{Max}} |\mathbf{M}(\xi)| \right)^{1/p}.$$

To address the potential impact of model mis-specification, consider the general situation in which the user specifies an inadequate model of the form $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$ where $\mathbf{X}_1$ is an $N \times p_1$ matrix, while the true model is $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$. For our example, the additional set of terms $\mathbf{X}_2\boldsymbol{\beta}_2$ contains the two-factor interactions (AD, AE, BC, BE, CD, DE). The bias vector of the specified model parameter estimates is given by

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}_1) - \boldsymbol{\beta}_1 &= [\boldsymbol{\beta}_1 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\boldsymbol{\beta}_2] - \boldsymbol{\beta}_1 \\ &= \mathbf{A}\boldsymbol{\beta}_2, \end{aligned}$$

where $\mathbf{A} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$ is the *alias matrix* (see Myers, Montgomery, and Anderson-Cook, 2009, p. 418). A measure of transmitted bias to the estimated coefficients is the sum of squared transmitted bias (SSB) given by

$$SSB = \boldsymbol{\beta}_2'\mathbf{A}'\mathbf{A}\boldsymbol{\beta}_2.$$

While $\boldsymbol{\beta}_2$ is unknown, it is often assumed to follow the multivariate normal distribution $\boldsymbol{\beta}_2 \sim \mathrm{N}\left(\mathbf{0}, \sigma^2_{\boldsymbol{\beta}_2}\mathbf{I}\right)$ (Draper and Guttman, 1992). Anderson-Cook, Borror and Jones (2009) consider the impact of changing this assumption. It is straightforward to show that the expectation of SSB is given by

$$\mathrm{E(SSB)} = \sigma^2_{\boldsymbol{\beta}_2}\,\mathrm{tr}(\mathbf{AA}').$$

An optimal design for minimizing the impact of aliasing on the estimated coefficients is one that minimizes $\mathrm{tr}(\mathbf{AA}')$, as proposed by Bursztyn and Steinberg (2006).

Model mis-specification can also transmit bias to the estimate of σ^2 since

$$\begin{aligned} E(\mathrm{MSE}_{\text{user}}) - \sigma^2 &= \boldsymbol{\beta}_2'[\mathbf{X}_1\mathbf{A} - \mathbf{X}_2]'[\mathbf{X}_1\mathbf{A} - \mathbf{X}_2]\boldsymbol{\beta}_2 / p_1 \\ &= \boldsymbol{\beta}_2'\mathbf{R}'\mathbf{R}\boldsymbol{\beta}_2 / p_1, \end{aligned}$$

where MSE_{user} denotes the residual mean squared error from the mis-specified model and $\mathbf{R} = \mathbf{X}_1\mathbf{A} - \mathbf{X}_2$. If positive bias is transmitted to the estimate of error, power to detect effects can be reduced due to the inflated error variance estimate. An optimal design for this aspect is one that minimizes $\mathrm{tr}\left(\mathbf{R}'\mathbf{R}\right)$ (Myers et al., 2009).

3. Process for Selecting a Best Design

The process for finding a best design for our purposes can be summarized by a multi-stage algorithm, with the following steps:

1. Create a collection of designs, and measure D-efficiency, $\mathrm{tr}(\mathbf{AA}')$ and $\mathrm{tr}\left(\mathbf{R}'\mathbf{R}\right)$ for all designs.
2. Construct the Pareto front, which consists of all designs which are not dominated by any other designs (more formal definition below in Section 3.2)
3. Select a best design from the Pareto front which best suits the needs of the experimenter.

We now present some of the details about each of the following steps for the example described in the Introduction.

3.1 Creating Collection of Design for Consideration

Since we are looking at a screening experiment where we wish to estimate a model with main effects and selected two-way interactions, while protecting against alternate two-way interactions, a natural choice of candidate locations from which to choose design locations are any combination of high and low levels for each factor. Hence, we consider a candidate set comprised of 32= 2^5 points, where each has the form $(\pm1, \pm1, \pm1, \pm1, \pm1)$. Since we are looking for a design with 14 runs, we consider any design made up of 14 runs from our candidate set, where we allow replicates and hence can sample from the 32 candidate points with replacement. Not all designs selected as a random set of runs will be able to estimate the model in (1), and hence we discard any such choices.

There are multiple approaches for generating designs, from randomly generating a large set of randomly constructed designs, to using a standard exchange algorithm, to using the PAPE algorithm (Lu, Anderson-Cook and Robinson, 2011). For the illustration used in the talk, we generated a large number of designs (2 million designs generated in under an hour on a standard desktop computer) and proceeded from there to evaluate which of those should be considered further with the Pareto front approach. The information needed for the remainder of the algorithm is a table with all criteria to be optimized in individual columns. Once a “best” design has been selected, the user needs to be able to map this to the original design matrix of 14 rows and 5 columns (A-E).

Figure 1 shows the pairwise scatterplot of all of the potential designs. Using the surface plot in JMP (Graph > Surface Plot) is an excellent way to dynamically visualize the Pareto front, which generally looks like an uneven outer surface of portion of a sphere. We can see that some trade-offs between criteria are needed, as improving one design criterion typically comes at the expense of worsening another.

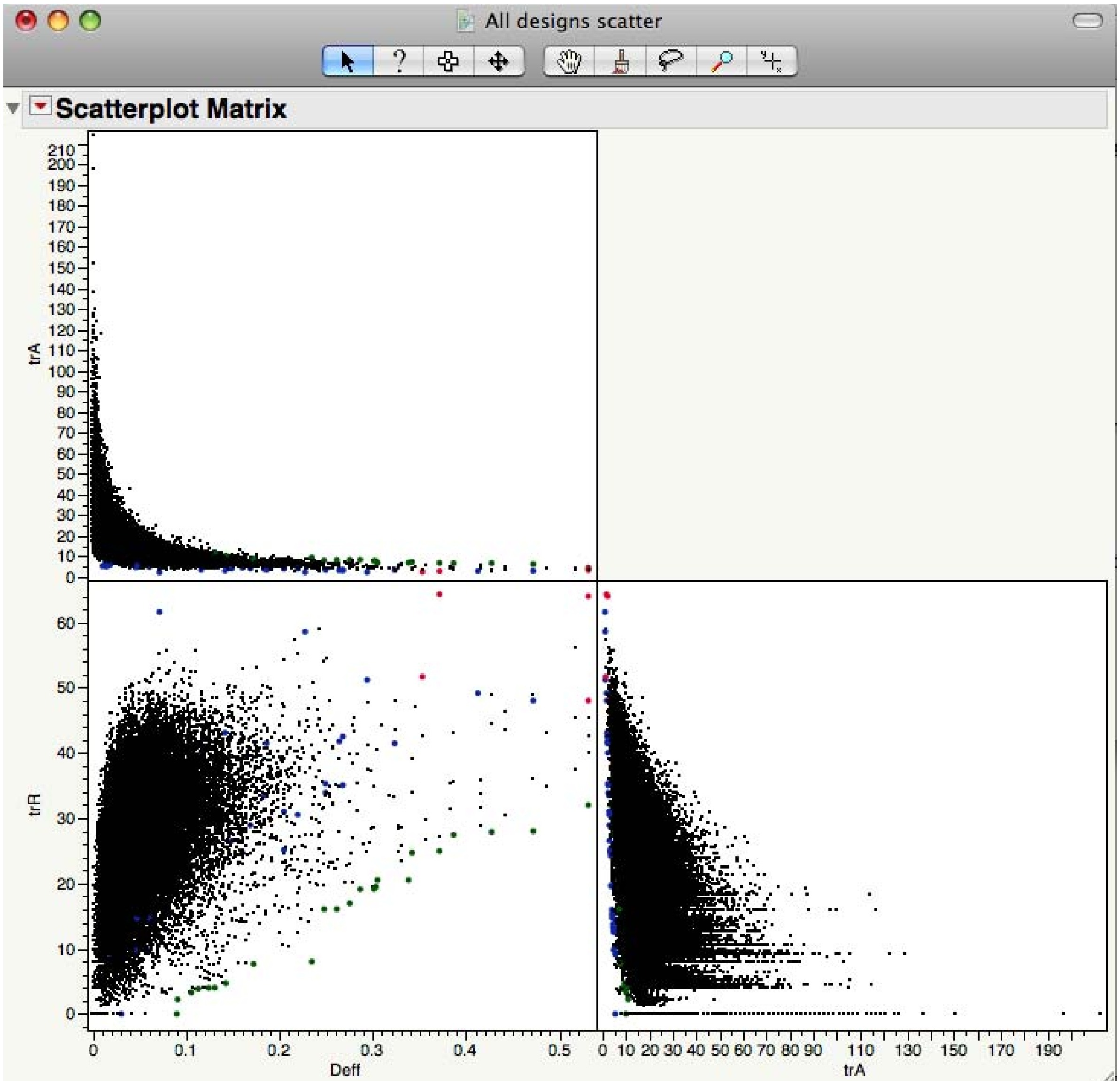

Figure 1: Pairwise Scatterplot of All Generated Designs for D-efficiency, $\mathrm{tr}(\mathbf{AA}')$ and $\mathrm{tr}\left(\mathbf{R}'\mathbf{R}\right)$.

3.2 Pareto Front Approach to Reduce Designs Considered

When balancing multiple objectives in design optimization, it is important to be able to identify the tradeoffs between competing design objectives. The Pareto approach (Kasprzak and Lewis, 2001, Gronwald et al., 2008, and Trautmann and Mehnen, 2009) enables the user to not only identify existing tradeoffs among criteria, but it also finds a suite of potential solutions for diverse weighting and scaling schemes associated with the criteria of interest.

Without loss of generality, assume the goal of a general multiple criteria design optimization problem is to maximize $C(\geq 2)$ criteria simultaneously given constraints on the input factors. Let $\boldsymbol{\xi} = (\mathbf{d}_1', \mathbf{d}_2', \ldots, \mathbf{d}_N')' \in \boldsymbol{\Omega}$ denote a design matrix of dimension *N*x*k* where *N* is the number of design points and *k* is the number of design variables. The set of all possible *N*x*k* design matrices for a given candidate set is denoted by $\boldsymbol{\Omega}$. Let $\mathbf{y} = \mathbf{F}(\boldsymbol{\xi}) = (f_1(\boldsymbol{\xi}), f_2(\boldsymbol{\xi}), \ldots, f_C(\boldsymbol{\xi}))'$ denote the vector of criteria values corresponding to $\boldsymbol{\xi}$. The space containing all obtainable criteria vectors is called the criterion space.

A solution $\boldsymbol{\xi}_1$ is said to *Pareto dominate* another solution $\boldsymbol{\xi}_2$ if $f_j(\boldsymbol{\xi}_1) \geq f_j(\boldsymbol{\xi}_2)$ for all $j \in \{1, 2, \ldots, C\}$ and there exists at least one $j \in \{1, 2, \ldots, C\}$ such that $f_j(\boldsymbol{\xi}_1) > f_j(\boldsymbol{\xi}_2)$. In this case, the criteria vector $\mathbf{F}(\boldsymbol{\xi}_2)$ is said to be *dominated* by $\mathbf{F}(\boldsymbol{\xi}_1)$. In layman's terms, this means that the design $\boldsymbol{\xi}_1$ is at least as good as $\boldsymbol{\xi}_2$ for all criteria, and better for at least one. In the remainder of the paper, the criteria vector corresponding to a particular solution is referred to a *point* in the criterion space. A solution is *Pareto optimal* if and only if no other solution dominates it and its corresponding criteria vector is a *non-dominated* vector. Henceforth, the *Pareto optimal set* will refer to the set of Pareto optimal solutions and the corresponding set of criteria vectors will be referred to as the *Pareto front*. A good overview of the Pareto-related concepts is available in Marler and Arora (2004).

Methods for generating the Pareto optimal set continue to be a popular research topic. The classical approach (Ngatchou et al., 2005) involves repeatedly finding optimal solutions using desirability functions or constrained optimization with modified weightings or restrictions. This method is generally inefficient and sensitive to the shape of the Pareto front as it often fails to capture the non-convex points on the front. Recent techniques focus on direct generation of the Pareto front based on Pareto ranking of the solutions. Among these approaches, genetic algorithms (Fonseca and Fleming 1993, Srinivas and Deb 1994, and Horn et al. 1994) and evolutionary algorithms (Zitzler and Thiele 1998) are most common. Lu et al. (2011) describe the PAPE algorithm which can populate the Pareto front with an adaptation of a point exchange algorithm based on a specified candidate set of design locations.

In JMP, the Pareto front can be easily extracted with the Row Selection option (Rows > Row Selection > Select Dominant), where the columns of interest as well as the direction to optimize (here maximize D-efficiency, and minimize both $\operatorname{tr}(\mathbf{A}\mathbf{A}')$ and $\operatorname{tr}\left(\mathbf{R}'\mathbf{R}\right)$) are specified. Figure 2 shows how to select the direction of the optimization.

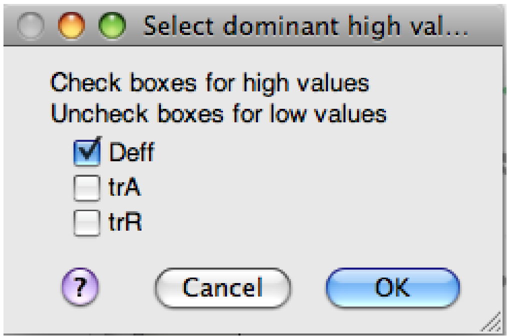

Figure 2: Selection menu to specify maximization or minimization for different criteria.

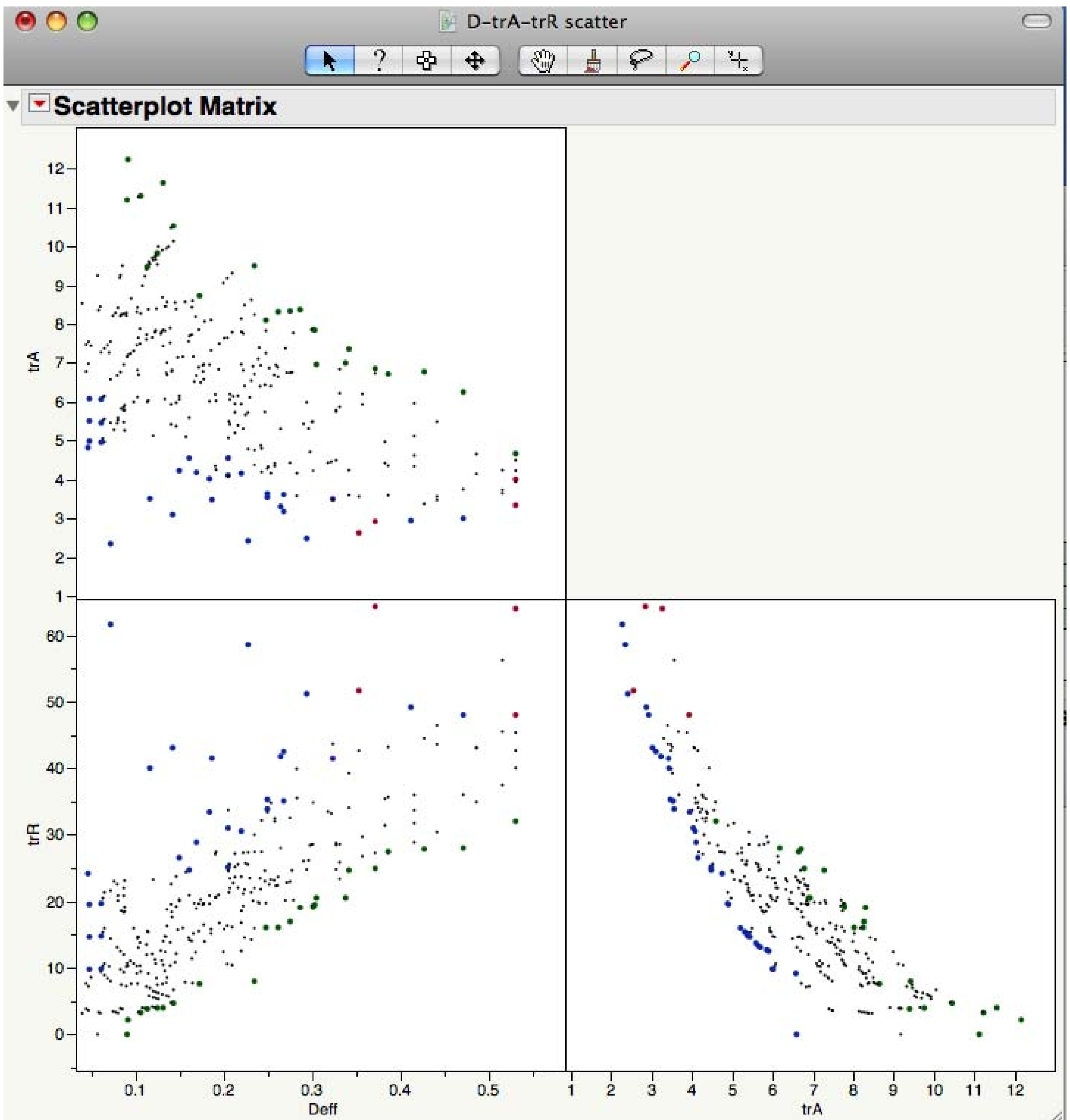

Figure 3: Pairwise Scatterplot of Pareto Front based D-efficiency, $\mathrm{tr}(\mathbf{AA}')$ and $\mathrm{tr}\left(\mathbf{R}'\mathbf{R}\right)$

These rows which comprise the Pareto front can then be extracted and explored further. Figure 3 shows the Pareto front comprised of 333 values for the three-criteria optimization, where different colors have been used to denote designs which are on the Pareto front found by optimizing the criteria in pairs (red for D-efficiency and $\mathrm{tr}(\mathbf{AA}')$, green for D-efficiency and $\mathrm{tr}\left(\mathbf{R}'\mathbf{R}\right)$),blue for $\mathrm{tr}(\mathbf{AA}')$ and $\mathrm{tr}\left(\mathbf{R}'\mathbf{R}\right)$). Points which are only on the three-criteria front are shown in black. Note that if a design is included in more than one two-criteria front, then it cannot be simultaneously shown in both colors for which it would be chosen. Figure 4 shows the scatterplot of the Pareto front based on just two criteria, D-efficiency and $\mathrm{tr}(\mathbf{AA}')$,where improvements in one criteria must necessarily lead to a worsening of the other criteria in order for neither of the two points to dominate each other. This unique ordering of points is not possible when three or more criteria are used.

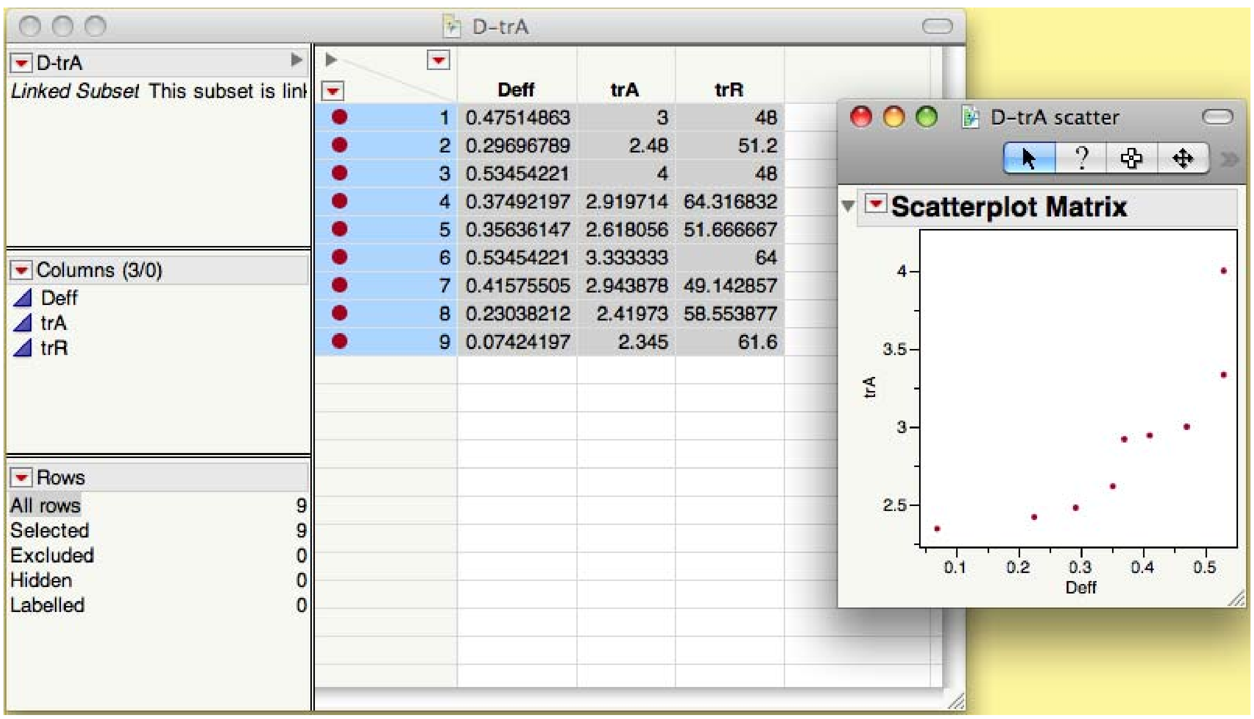

Figure 4: Pareto front based on two criteria: D-efficiency and $\mathrm{tr}(\mathbf{AA}')$

Note that the two criteria front shown in Figure 4 is comprised of 9 designs, and improving the D-efficiency (increasing), must necessarily worsen $\mathrm{tr}(\mathbf{AA}')$ (increase), and vice versa. From this summary, we see that the best possible value for D-efficiency is 0.5345, and the best possible for $\mathrm{tr}(\mathbf{AA}')$ is 2.345 (the best value for $\mathrm{tr}\left(\mathbf{R}'\mathbf{R}\right)$ is 0).

These 333 designs represent the set of all non-dominated designs, and form the collection from which it is rational and defensible to select the single final "best" design, since they dominate all other designs considered.

3.3 Selecting a "Best" Design

As with the step described in Section 3.1, there are multiple approaches for paring down the multiple design alternatives found using the Pareto Front approach in Section 3.2 to identify the single final design to run in the experiment.

Lu, Anderson-Cook and Robinson (2011) describe an approach to finding the best design using the Utopia Point approach, which seeks to minimize the distance between points on the front with the ideal location which optimizing all the criteria simultaneously (called the Utopia Point). Typically this ideal does not exist for any possible design, and so this approach seeks to find the best alternative subject to some scaling of the different criteria which matches the experimenter's priorities. It has some similarities to the desirability function approach (Derringer-Suich, 1980), but with an easy method for exploring the robustness of the choice of design to different weightings of the criteria and forms for combining them. Here we consider an alternative approach, which is appropriate if the experimenter has clear ideas about what values of the different criteria are acceptable.

This approach uses the Data filter (found in Rows > Data Filter), and allows the user to specify constraints on each of the different criteria, and rows satisfying the constraints are selected. The approach starts with all rows selected, and the Data Filter showing the complete ranges of the data, as shown in Figure 5. We see the data ranges for the three criteria are D-efficiency $\in[0.013, 0.535]$,

$\mathrm{tr}(\mathbf{AA}') \in [2.345, 12.242]$ and $\mathrm{tr}\left(\mathbf{R}'\mathbf{R}\right) \in [0, 64.317]$

The user should adjust the ranges of each of the criteria in the order of importance that they feel reflects their importance as experiment goals. With each additional constraint, the available range for the remaining criteria shrinks. For example, in this case, we might feel that performing well for estimating the model parameters for our chosen model (D-efficiency) is most important, and that we would like to achieve at least 65% relative efficiency compared to the best available design (with D-efficiency value of 0.5345). Hence the lower bound of the Data Filter could be shifted to give a value of approximately 0.347. This selects 40 of the 333 possible designs as shown in Figure 6. Note how in the scatterplots, the points highlighted fall to the right of the imaginary vertical line at 0.3418.

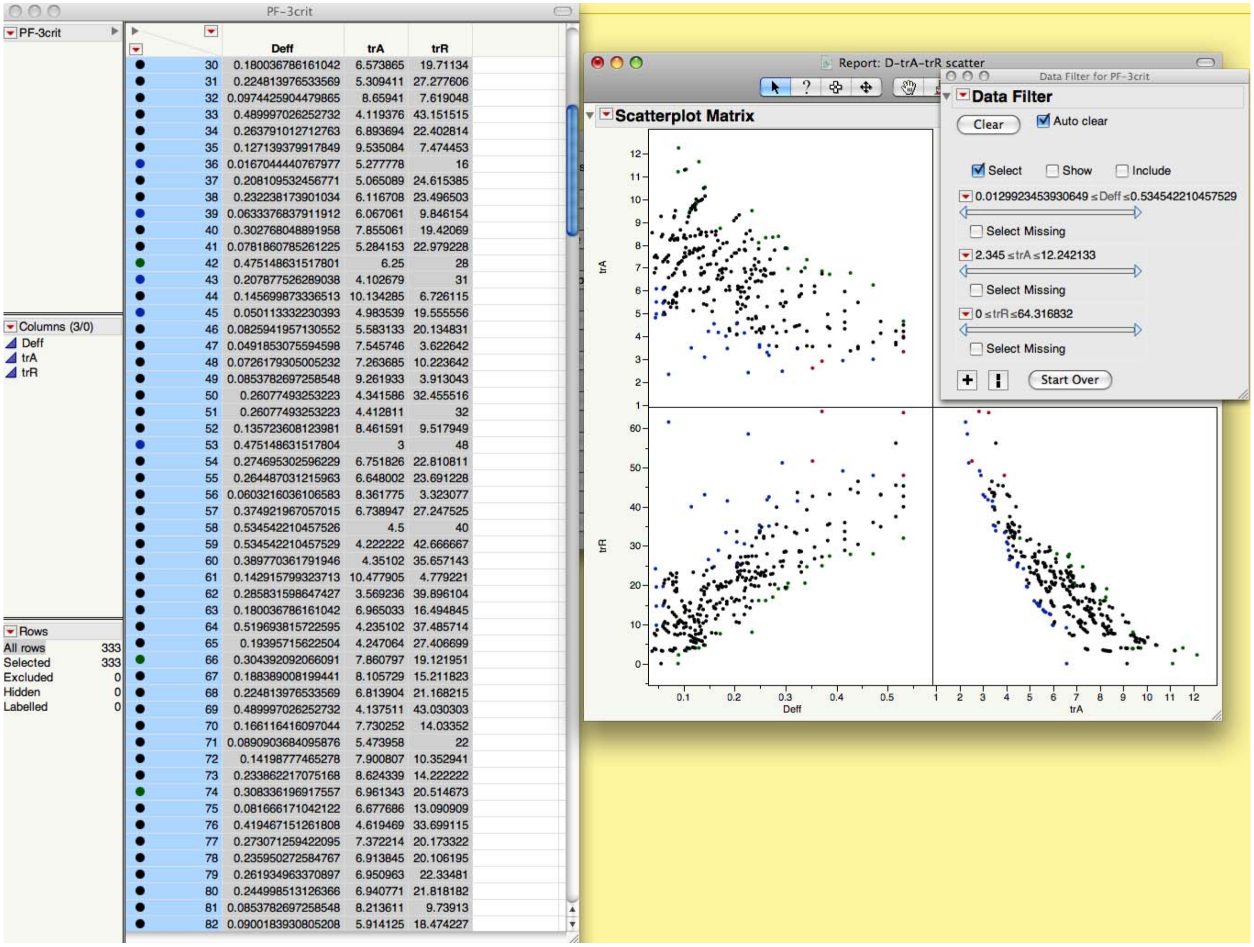

Figure 5: Starting point for Data Filter approach for selecting a best design

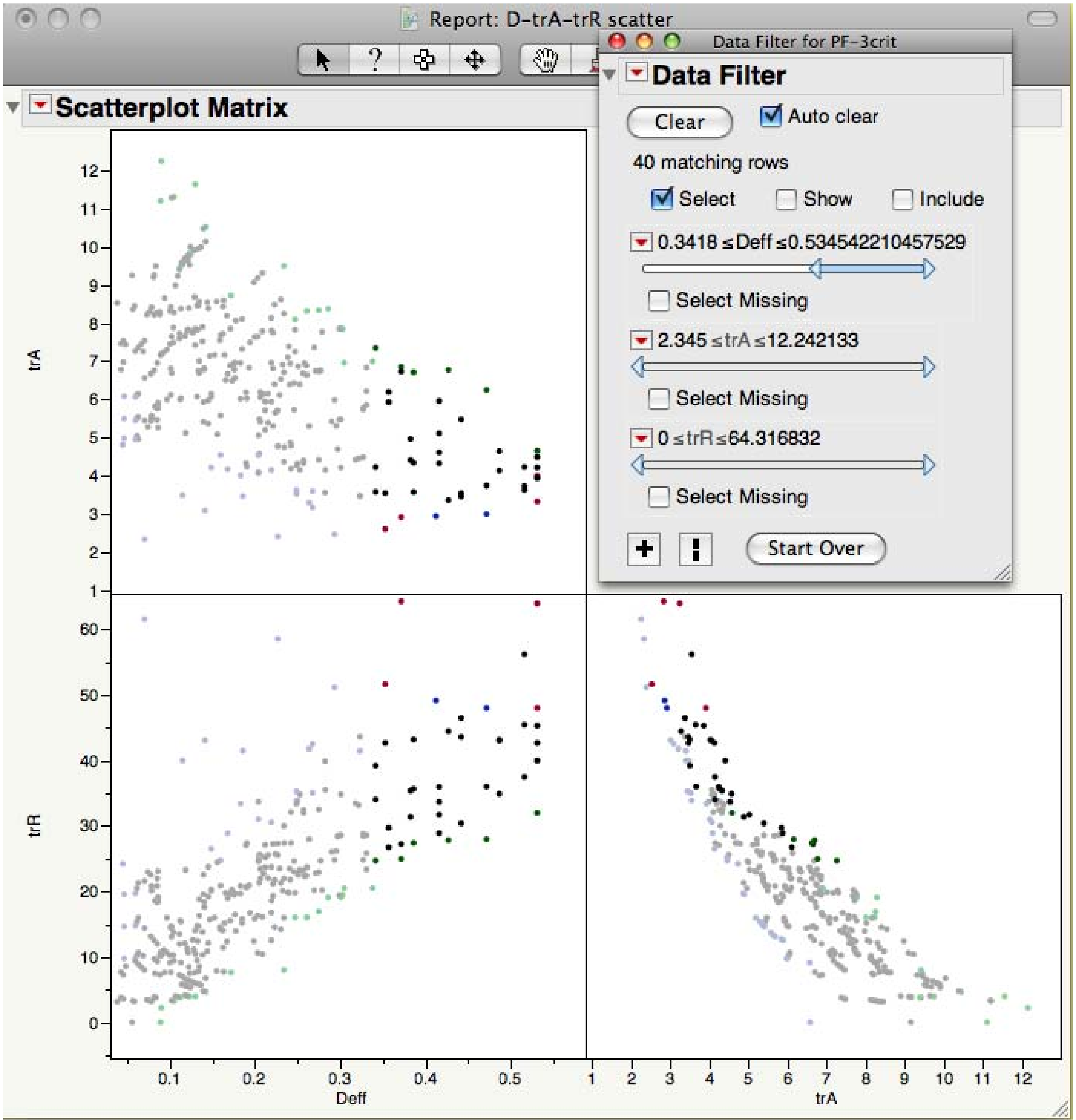

Figure 6: Using the Data Filter to Select Design with at least 65% relative D-efficiency

Next we adjust the remaining criteria (here since we are seeking to minimize $\mathrm{tr}(\mathbf{AA}')$ and $\mathrm{tr}\left(\mathbf{R}'\mathbf{R}\right)$, we adjust the upper bound for each criterion) until we obtain a best design. Note that in this case, we adjusted $\mathrm{tr}(\mathbf{AA}')$ first, and so this allowed us very little flexibility to further restrict the range of $\mathrm{tr}\left(\mathbf{R}'\mathbf{R}\right)$. Figure 7 shows the final highlighted design based on this set of criteria.

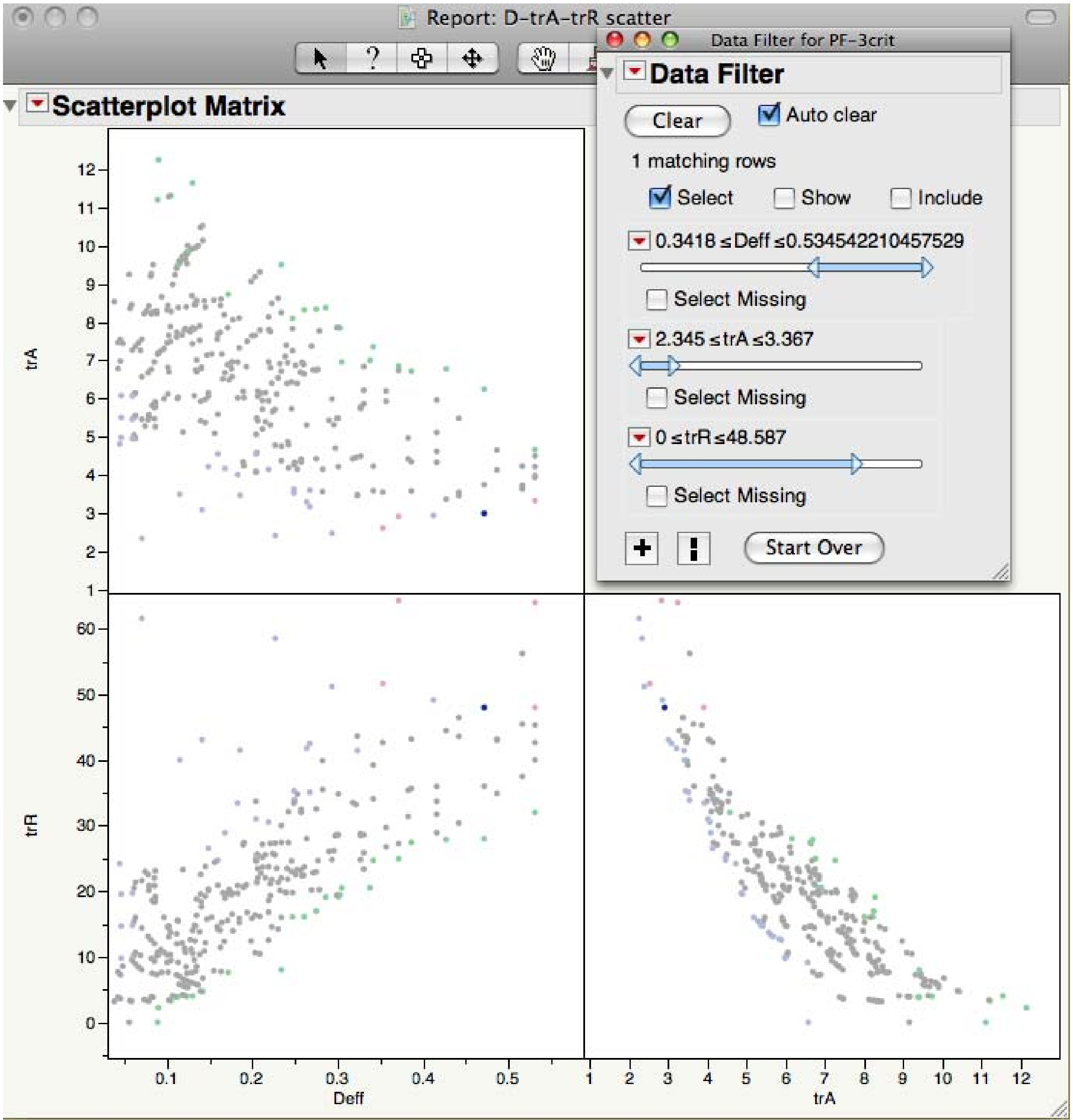

Figure 7: Final selected design using Data Filter approach

Figure 8 shows the final criterion values for the selected design based on this set of constraints. In general, we recommend experimenting with the different thresholds to identify several different candidate “best” designs, and then carefully comparing their criterion values to determine which suits the goals of the experiment best.

PF−3crit

PF-3crit

Columns (3/0)

Deff

trA

trR

	Deff	trA	trR
30	0.180036786161042	6.573865	19.71134
31	0.224813976533569	5.309411	27.277606
32	0.0974425904479865	8.65941	7.619048
33	0.489997026252732	4.119376	43.151515
34	0.263791012712763	6.893694	22.402814
35	0.127139379917849	9.535084	7.474453
36	0.0167044440767977	5.277778	16
37	0.208109532456771	5.065089	24.615385
38	0.232238173901034	6.116708	23.496503
39	0.0633376837911912	6.067061	9.846154
40	0.302768048891958	7.855061	19.42069
41	0.0781860785261225	5.284153	22.979228
42	0.475148631517801	6.25	28
43	0.207877526289038	4.102679	31
44	0.145699873336513	10.134285	6.726115
45	0.050113332230393	4.983539	19.555556
46	0.0825941957130552	5.583133	20.134831
47	0.0491853075594598	7.545746	3.622642
48	0.0726179305005232	7.263685	10.223642
49	0.0853782697258548	9.261933	3.913043
50	0.26077493253223	4.341586	32.455516
51	0.26077493253223	4.412811	32
52	0.135723608123981	8.461591	9.517949
53	0.475148631517804	3	48
54	0.274695302596229	6.751826	22.810811
55	0.264487031215963	6.648002	23.691228
56	0.0603216036106583	8.361775	3.323077
57	0.374921967057015	6.738947	27.247525
58	0.534542210457526	4.5	40
59	0.534542210457529	4.222222	42.666667
60	0.389770361791946	4.35102	35.657143
61	0.142915799323713	10.477905	4.779221

Figure 8: Data Table with Selected Final design

Once a best design has been identified, then the details of the design can be examined to see if the geometry of the design in the factor space is appealing. Figure 9 shows the design with gray circles denoting where observations exist. Note that for any of the factors, the levels of -1 and +1 can be reversed. However, because of the non-symmetric model in (1), the factors cannot be relabeled (for example, changing the roles of factor A and E) and still preserve the characteristics of D-optimality, $\mathrm{tr}(\mathbf{A}\mathbf{A}')$ and $\mathrm{tr}(\mathbf{R}'\mathbf{R})$.

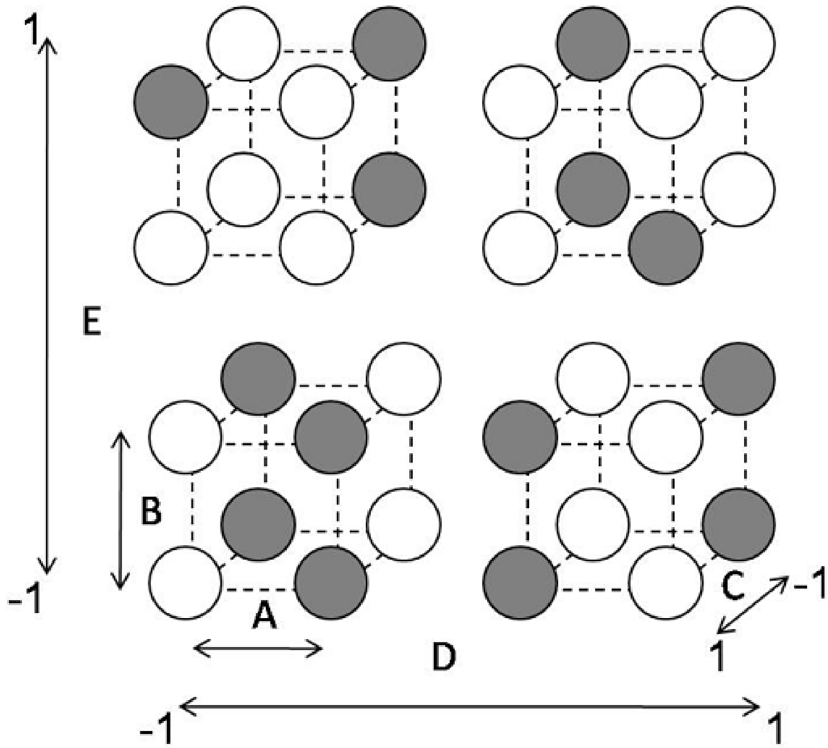

Figure 9: Final design selected using Data Filter approach.

The design can be further evaluated in JMP with the DOE > Augment Design feature, which has a large number of Design Evaluation tools available automatically. Figure 10 shows a snapshot of the list of tools available to assess a design.

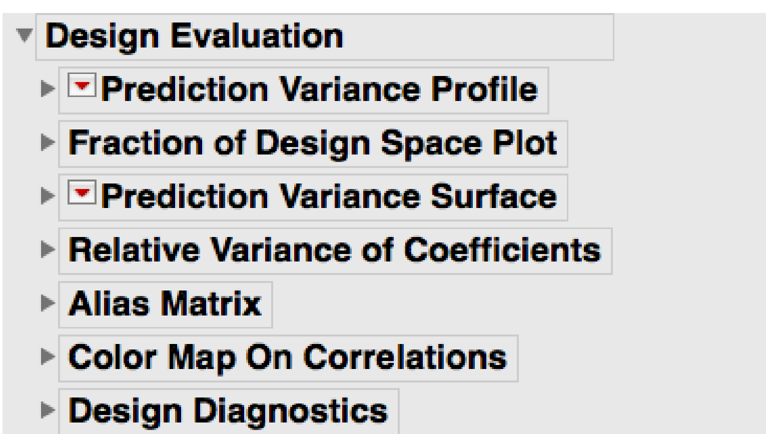

Figure 10: Snapshot of JMP list of Design Evaluation tools (accessible from Custom Design and Augment Design)

4. Conclusions

In design of experiments, a "best" design should be selected based on several criteria which highlight varied aspects of a "good design", such as the precision of parameter estimation and protection against model misspecification. Quite often, these criteria cannot be simultaneously optimized and trade-offs must be incorporated. We illustrated the Pareto approach for design construction and selection using an example to find an optimal 14-run design involving five two-level factors and the specified model in (1) using D-optimality, $\mathrm{tr}(\mathbf{AA}')$ and $\mathrm{tr}(\mathbf{R}'\mathbf{R})$.

This Pareto optimization approach combined with the Data Filter selection method allows the user to identify a collection of best designs from which to select a final experiment to be run. The approach is flexible enough to be applied to other design of experiment situations where other criteria might be of primary interest. For creating the Pareto front and using the Data Filter to select a final design, all that is required is a table (like the one shown in Figure 8) with a column for each criterion and rows for all potential choices. It is also appropriate for response optimization approaches as well. Once a set of responses with a corresponding table of candidate values has been identified, the same selection procedure applies.

Many decisions involve complex trade-offs between competing objectives. Having a formal method to eliminate dominated designs from consideration, and then selecting from among the Pareto front candidates can lead to better understanding of available options and hence improved decision-making.

References

Anderson-Cook, C. M. (2010), "A Matter of Trust: Balance Confidence in Your Model While Avoiding Pitfalls", *Quality Progress*, March, 56-58.

Anderson-Cook, C.M., Borror, C.M., Jones, B. (2009) "Graphical Tools for Assessing the Sensitivity of Response Surface Designs to Model Misspecification" *Technometrics* 51 75-87.

Bursztyn, D. and Steinberg, D.M. (2006), "Comparisons of Designs for Computer Experiments", *Journal of Statistical Planning and Inference*, Vol. 136, 1103-1119.

Draper, N.R. and Guttman, I. (1992), "Treating Bias as Variance for Experimental Design Purposes", *Annals of the Institute of Statistical Mathematics*, Vol. 44, 659-671.

Fonseca, C., and Fleming, P.J. (1993), "Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion, Generalization", proceedings of the Fifth International conference on Genetic Algorithms, San Mateo, CA, 416-423.

Gronwald, W., Hohm, T., and Hoffmann, D. (2008), "Evolutionary Pareto-optimization of Stably Folding Peptides", *BMC-Bioinformatics*, Vol. 9, 109.

Horn, J., Nafpliotis, N. , and Goldberg, D.E. (1994), "A Niched Pareto Genetic Algorithm for Multiobjective Optimization", proceedings of the First IEEE Conference on Evolutionary Computation, IEEE World Congress on Computational Intelligence, Piscataway, NJ. Vol. 1, 82-87.

Kasprzak, E.M. and Lewis, K.E. (2001), "Pareto Analysis in Multiobjective Optimization Using the Collinearity Theorem and Scaling Method", *Structural Multidisciplinary Optimization*, Vol. 22, 208-218.

Lu, L., Anderson-Cook, C.M., Robinson, T.J. (2011) “Optimization of Designed Experiments Based on Multiple Criteria Utilizing a Pareto Frontier” *Technometrics* (in press).

Marler, R.T., and Arora, J.S. (2004), “Survey of Multi-Objective Optimization Methods for Engineering”, *Structural Multidisciplinary Optimization*, Vol. 26, 369-395.

Myers, R.H., Montgomery, D.C., and Anderson-Cook, C.M. (2009), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, Wiley: New York, 3rd ed.

Ngatchou, P., Zarei, A., and El-Sharkawi, M.A. (2005), “Pareto Multi Objective Optimization”, proceedings of IEEE International Conference on Intelligent Systems Application to Power Systems, Washington DC, 84-91.

Srinivas, N., and Deb, K. (1994), “Multiobjective Function Optimization Using Nondominated Sorting Genetic Algorithms”, *Evolutionary Computation*, Vol. 2(3), 221-248.

Trautmann, H., and Mehnen, J. (2009), “Preference-based Pareto Optimization in Certain and Noisy Environments”, *Engineering Optimization*, Vol. 41, 23-38.

Zitzler, E., and Thiele, L. (1998), “An Evolutionary Algorithm for Multiobjective Optimization: The Strength Pareto Approach”, TIK Tech. Report No. 43, Swiss Federal Institute of Technology (ETH).