Quality and Productivity Research Conference, June 10 -12, 2015
College of Textiles, North Carolina State University, Raleigh, NC, USA

Modeling Strategies for Definitive and Bayesian D Optimal Screening Designs

Maria Weese, Ph.D.
Department of Information Systems and Analytics
Miami University
Oxford, OH, USA
weeseml@miamioh.edu

Philip J. Ramsey, Ph.D.
University of New Hampshire & North Haven Group
Durham, NH, USA
philip.ramsey@unh.edu

Douglas Montgomery, Ph.D.
Department of Industrial Engineering
Tempe, AZ, USA
doug.montgomery@asu.edu

Outline

- Introduction
- Simulation Protocol
- Model Selection Strategies
- The Dantzig Selector
- Simulation Results for DSDs (Preliminary)
- Simulation Results DSD vs. BDD
- Fermentation Case Study

Introduction

"The best way to predict the future is to create it." Peter Drucker.

Definitive Screening Designs (DSD), are receiving a lot of interest due in part due their high efficiency, which makes them attractive for experimenters where experimental resources are limited.

Although popular there has only recently been much interest in the question of how to best analyze these designs.

The question of analysis strategies also applies to other types of screening designs such as **Bayesian D-optimal Designs (BDD)** where again the amount of research on analysis is limited.

In this talk we will investigate, through simulations, model selection strategies for both DSDs and BDDs and make some comparisons between the two types of screening designs.

Simulation Details

Four separate simulation scenarios were employed.

	Main Effects		Quadratic Effects		Interaction Effects		Heredity
Scenario	a	τ	a	τ	a	τ	
1	n/3	6	1	3	1	3	Strong
2	3	6	3	6	3	6	Strong
3	n/4	3	2	9	2	9	Strong
4	3	6	3	6	3	6	Weak

In the table *a* is the number of effects of each type randomly selected from a full quadratic model for *k* experimental factors and τ represents the magnitude of the selected effects.

The signs of the coefficients were randomly assigned.

Inactive effect coefficients were assigned from a value of 0 – a topic for consideration in future simulations .

Random error was added to the response with $\varepsilon \sim N(0;1)$.

Simulation Details

Each Scenario required that 14 designs be generated: 7 DSDs and 7 BDDs in order to cover the range of k.

As an example, for Scenario 1, a DSD with $k = 5$ has $n = 13$ trials, a randomly selected model would have 4 main effects of size 6, 1 quadratic effect of size 3, and 1 interaction of size 3.

Bayes D optimal designs were generated in JMP Pro 11 by specifying, for a full quadratic model, all main and quadratic effects as *Necessary* and all two-way interactions as *If Possible*.

The number of trials n was specified to match the required number of trials for the corresponding DSD.

For each scenario, and number of factors $k = 5, 6, \ldots, 11$ combination 1,000 iterations were performed where the appropriate model was randomly generated for each iteration.

Simulation Details

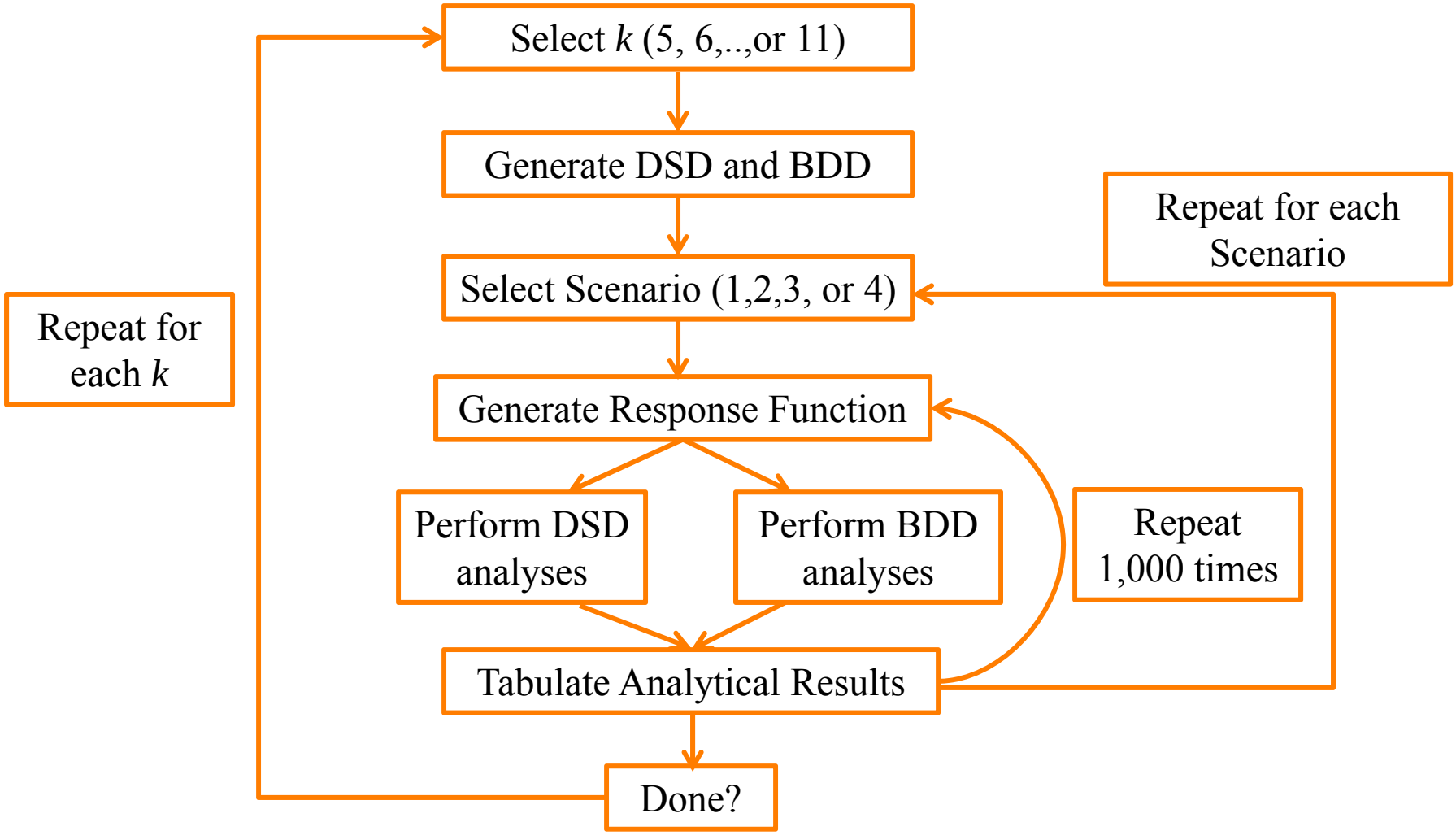
Each simulation iteration was analyzed with various methods and each iteration used the same response function for both design types.

The following were measured for each design type, scenario, and *k*:

- Power (overall number of times the correct active factors were identified) and the associated Type II error rates.
- Type I error rates (number of times inactive factors were identified as active).
- Average total number of effects identified as active.
- The power to detect the different types of effect, i.e. power to correctly identify the quadratic effects, the interaction effects and the main effects.
- PRESS statistic for the selected models and their responses.

Simulation Details

Below is a flowchart of the simulation path:



Simulation Details

The following analysis strategies were employed:

- **Forward Selection** using AICc and BIC as a stopping criteria.
- **The Dantzig Selector** using AICc and BIC to choose δ, γ=1.5.
- **All Subsets Regression** (DSDs only) using AICc and BIC for the selection criteria with max model size set to 10 effects.

All Subsets Regression is computationally very expensive, so a limit of 10 was chosen.

Also, under the assumption of Effect Sparsity it would be unusual to find more than 10 active effects in a screening design.

Effect Heredity restrictions are not enforced in any of the analyses – this is a topic for future research.

Model Selection

AICc is the **Akaike Information Criterion** with a bias correction factor "c" for small data sets.

BIC is the **Bayesian Information Criterion**.

The AICc, assuming the response is normally distributed, has the following formula:

$$AICc = nLn\left(\frac{SSE(p)}{n}\right) + \frac{2p(p+1)}{n-p-2}.$$

Overfitting penalty term

Where *SSE* is sum of squares error, n is the number of observations and p is the number of model terms **including** the intercept and the estimate of σ.

Model Selection

The BIC is similar in form to the AICc and assuming the response is normally distributed has the form:

$$BIC = nLn\left(\frac{SSE(p)}{n}\right) + pLn(n).$$

Overfitting penalty term

The model with the smallest BIC value is interpreted as the one with the largest **posterior probability given the data**.

The basis for the BIC lies in Bayesian probability theory while the basis for AICc lies in thermodynamics and is related to Boltzmann's entropy.

Generally the two popular criteria for model selection do not agree on a best model due to the different manner in which over fitting is penalized.

There is no consensus as to which may be the preferred criterion.

Model Selection

The DSD and BDD are supersaturated for the full quadratic model, so modeling strategies must sequentially determine important effects to add to the model from the set of all possible effects.

Forward Selection starts with no terms in the model and adds terms sequentially that provide the largest gain in the selection criteria – AICc and BIC are used in the simulation study.

All Subsets Regression fits all possible models of every size up to some specified maximum size.

For p effects there are approximately 2^p possible models (not all models for a DSD and BDD are going to be estimable).

Penalized Regression techniques also sequentially add terms to the model, which provide the largest gain in the selection criterion.

Many forms of penalized regression exist.

The Dantzig Selector

The **Dantzig Selector** is a shrinkage method often used when the number of possible effects $p > n$. The solution is found by solving the linear programming problem:

$\hat{\beta}_{DS}$ is the solution to

$$\min_{\hat{\beta} \in B} \left\| \hat{\beta} \right\|_1 \text{ subject to } \left\| X^T \left(Y - X\hat{\beta} \right) \right\|_\infty \leq \delta$$

The tuning parameter δ is determined empirically with AICc or BIC.

A two stage procedure is often used where the initial effects are estimated then any effect < γ is removed and the remaining effects estimated by OLS; γ = 1.5 was used (Marley and Woods, 2010).

Several others Marley and Woods (2010), Draguljić et al. (2014), Weese et al (2015) have shown success in terms of identifying the correct active factors in supersaturated designs using the Dantzig selector.

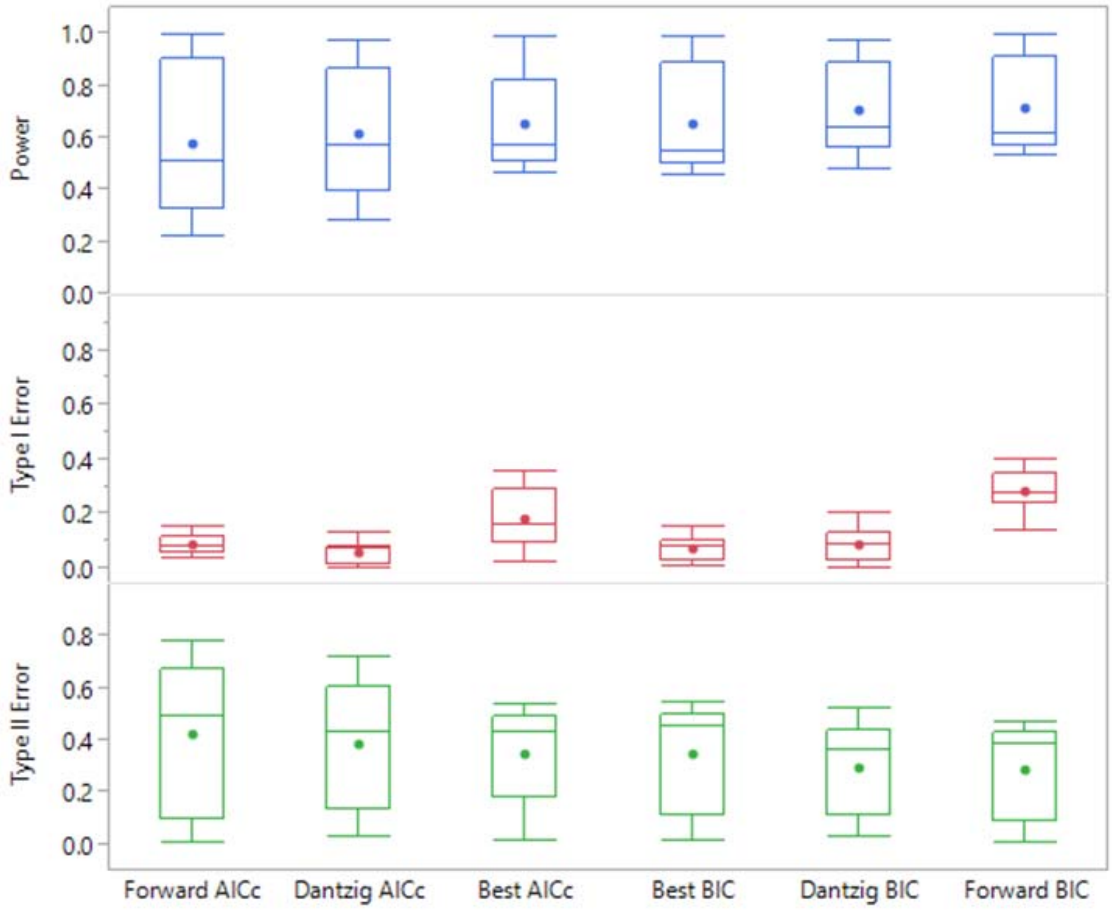
Preliminary Simulation Results for DSDs

To the right is a graph of the overall results across the 4 Scenarios and number of factors *k*.

The differences in overall Power are relatively small (9% spread).

The difference in Type I error is more striking.

The Dantzig Selector is performing as well as All Subsets Regression and is computationally more efficient



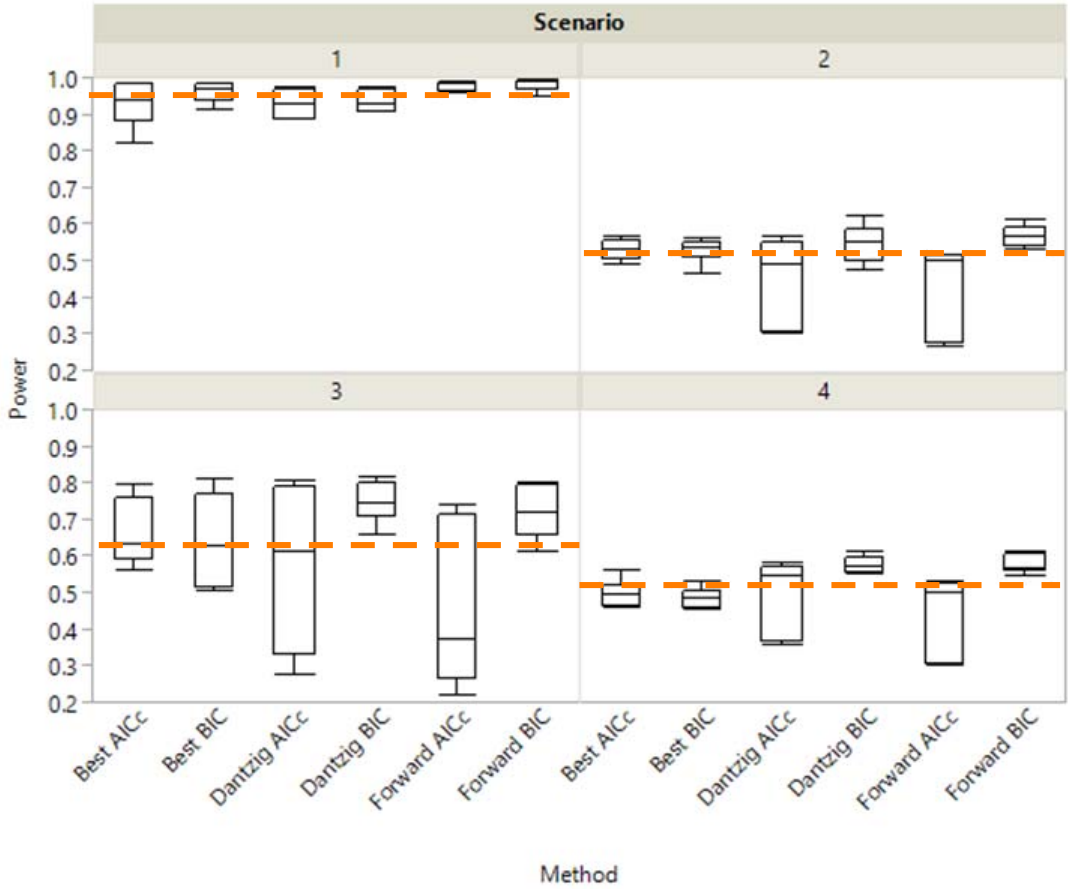
Preliminary Simulation Results for DSDs

To the right is a graph of results by Scenario.

Scenario 1 with dominant main effects had high power.

Scenario 3 with dominant quadratic and interaction effects had relatively high power.

Scenarios 2 and 4 with equal magnitude of effects did relatively poorly in power.



	Main Effects		Quadratic Effects		Interaction Effects		Heredity
Scenario	a	τ	a	τ	a	τ	
1	n/3	6	1	3	1	3	Strong
2	3	6	3	6	3	6	Strong
3	n/4	3	2	9	2	9	Strong
4	3	6	3	6	3	6	Weak

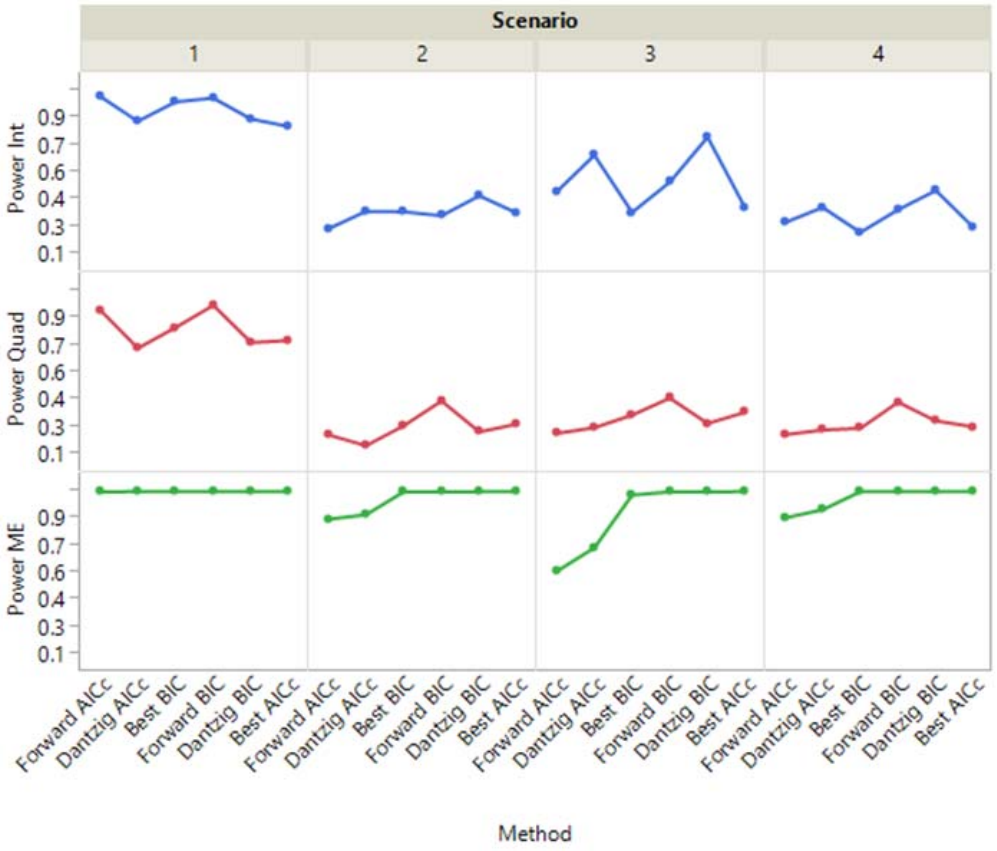
Preliminary Simulation Results for DSDs

To the right is a graph of results by Scenario and type of effect.

Scenario 1 with sparse interaction and quadratic effects has high power for all effects.

The power for main effects was high in all 4 Scenarios.

But Forward and Dantzig Selector using AICc did the worse in 2 and 3.



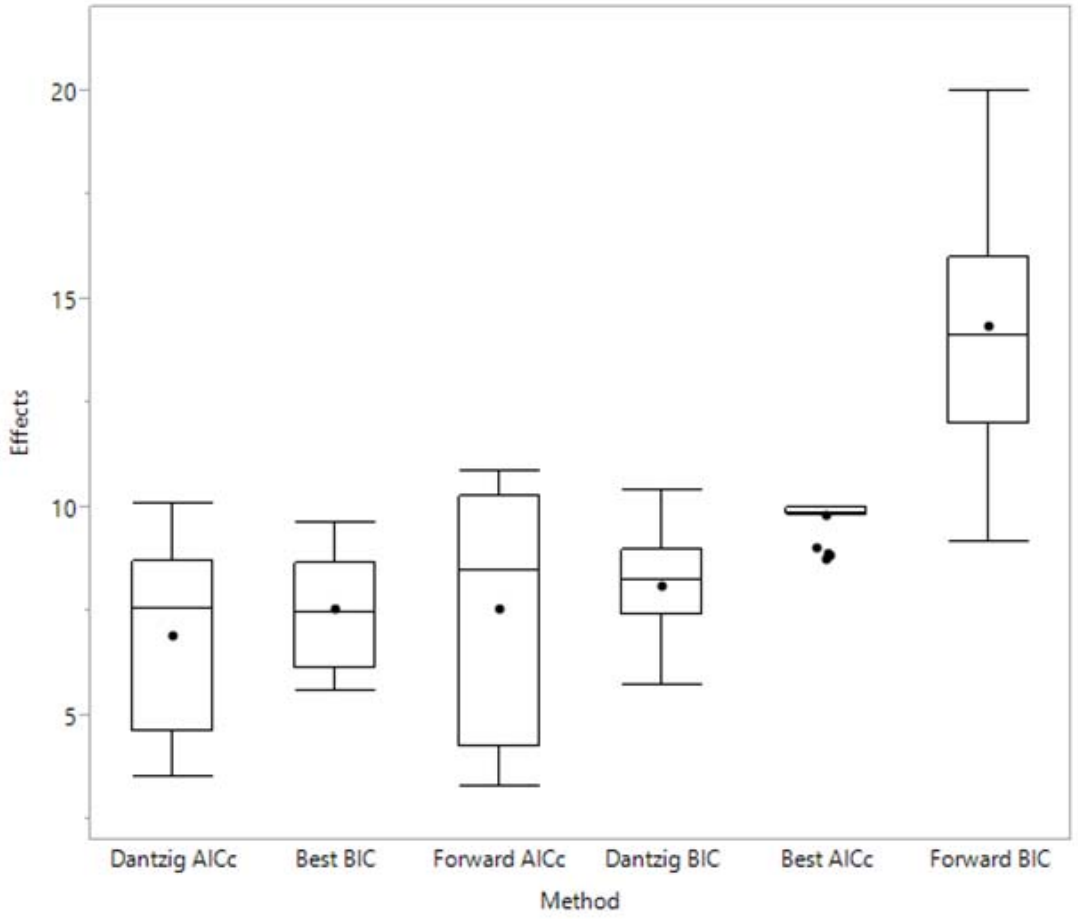
	Main Effects		Quadratic Effects		Interaction Effects		Heredity
Scenario	a	τ	a	τ	a	τ	
1	n/3	6	1	3	1	3	Strong
2	3	6	3	6	3	6	Strong
3	n/4	3	2	9	2	9	Strong
4	3	6	3	6	3	6	Weak

Preliminary Simulation Results for DSDs

To the right is a graph of the total number of effects found by method.

Forward Selection BIC is noticeably higher and is consistent with a known tendency of BIC to over fit models when *n* is small relative to *p*.

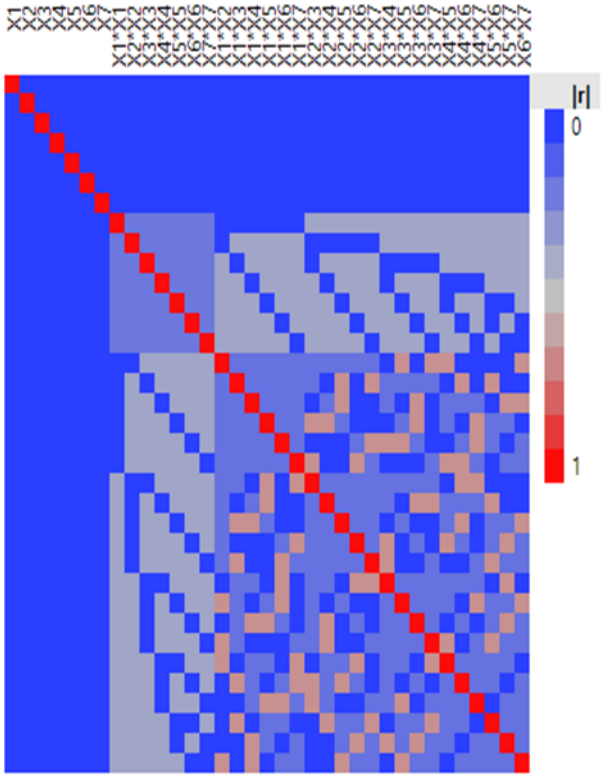
In fact the largest number of active effects across all cases would be **10**.

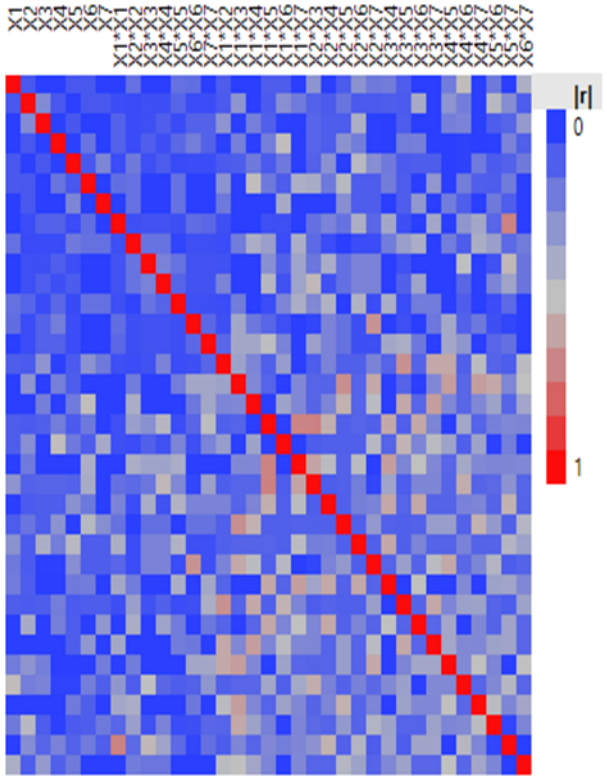


Preliminary Simulation Results DSD vs. BDD

One advantage of the DSD is that the main effects are orthogonal and this may not be the case for a BDD.

Below is a color map of a DSD aliasing (left) and the corresponding BDD design, $k = 7$ and $n = 17$.



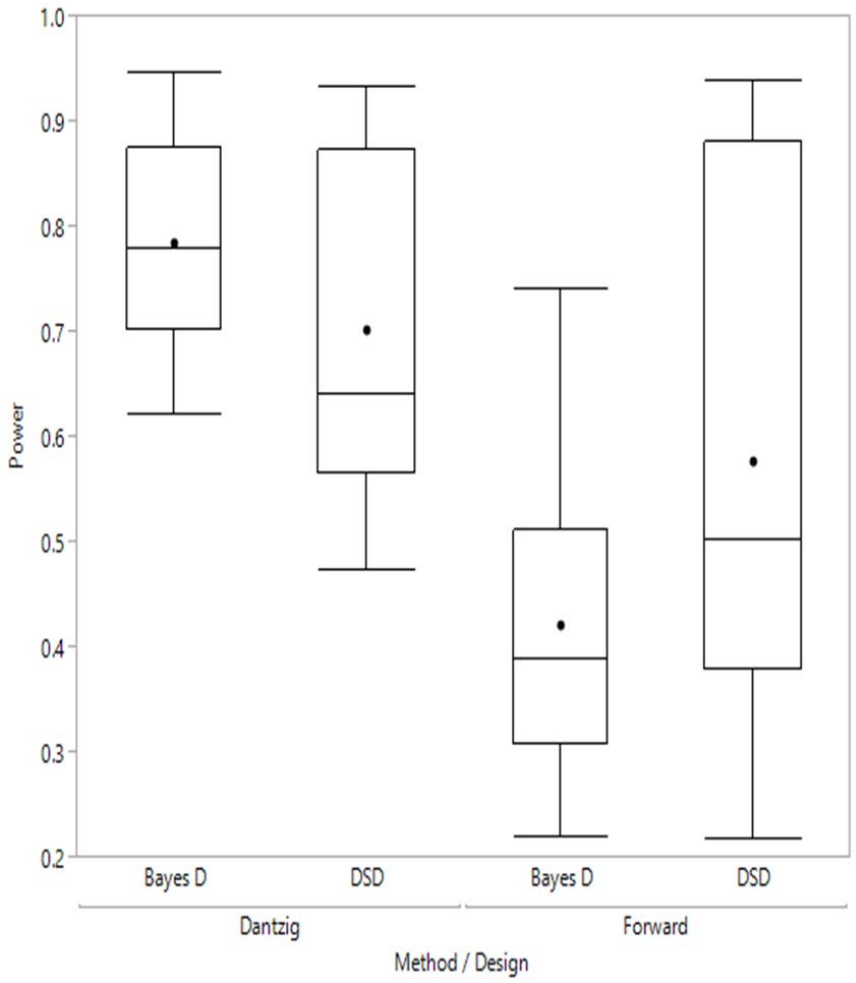


*Preliminary Simulation Results DSD vs. BDD

In these simulations the Dantzig Selector with BIC and Forward Selection with AICc were used; Subsets Regression was not performed.

To the right is a graph of power by Design Type, and Method.

Interestingly the Forward Selection performed very poorly for the BDDs and generally lower in power for both designs.



*In these simulations the inactive coefficients were assigned values from N(0; 0.2) and the active effects were assigned value from a N(3; 0.2) or N(6; 0.2) as appropriate.

Preliminary Simulation Results DSD vs. BDD

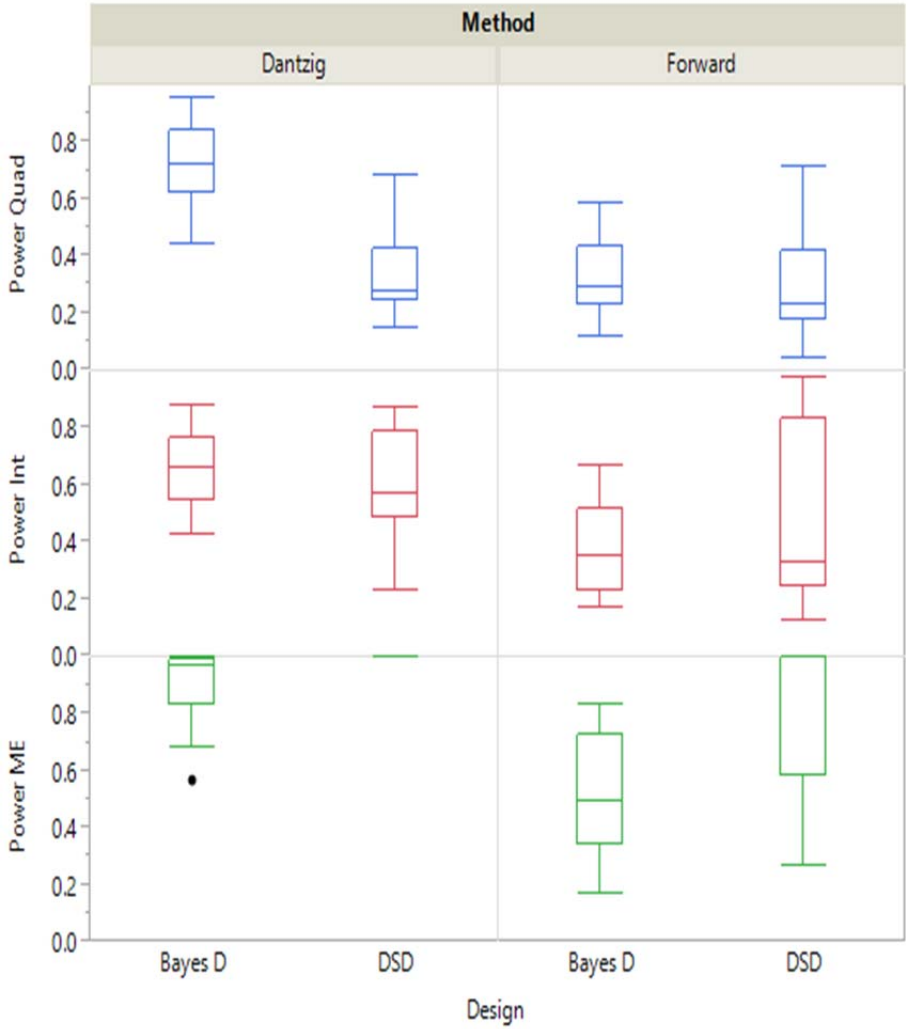
To the right is graph of power for each type of effect by Design Type and Method.

For Quadratic effects and the Dantzig Selector the BDD seems to do best.

For Interaction effects and the Dantzig Selector power is about the same.

For Main effects the DSD and Dantzig Selector is best.

Forward Selection was lower in virtually all cases.



Should you use one or the other?

There doesn't seem to be any appreciable difference in design choice.

Interestingly BDD designs seem to perform better with the Dantzig Selector in terms of power for the quadratic and interaction effects.

Overall, The DSD designs seem slightly better in terms of power for larger k (results not shown).

For DSDs the level of partial aliasing between quadratic and interaction effects decreases as k increases.

This may explain the change in performance as k increases.

For main effects the DSDs do quite well in terms of power, which is expected given the main effects are orthogonal.

The Fermentation Experiments

We now focus on an experiment to optimize the biomolecule **Yield** of the fermentation step in a bio-process.

For the experiment $K = 5$ factors were identified:

1. **pH** (6.8, 7.2) = fermentation solution pH;
2. **Dissolved Oxygen** (%DO) (target values 20%, 40%);
3. **Induction Tempe** (39.5 C, 42.5 C) = Temperature at which the biomolecule production is induced in the E. Coli cells.
4. **Induction OD_{600}** (20, 40) = biomass at which the induction is initiated as measured by optical density at 600 nm.
5. **Feed Rate** (1.9, 3.5 mL/hr) = feed rate of a growth media containing 50% glycerol added to the fermentation solution when induction is initiated.

The Fermentation Experiments

The two goals of the experiment were to **characterize the fermentation** step and to **maximize the Yield** of a biomolecule (X) produced by E. Coli cells.

Note: The goal was not necessarily to maximize the mass of the E. Coli community. It is possible to substantially increase the mass of a microbial community without maximizing biomolecule production.

The three responses of interest are:

1. **Yield** = biomolecule titer measured in units of mg/L;
2. **OD600** = measure of biomass by optical density at 600 nm;
3. **WCW** = wet cell weight in units of g/L.

Yield was the primary response and the focus of our analysis.

The Fermentation Experiments

To evaluate the DSD and a traditional Central Composite Design both designs were run in parallel.

Several analysis scenarios can be considered.

Analyze each separately using All Subsets Regression and the Dantzig Selector and compare the solutions.

Use multiple solutions for All Subset Regression and average the predictions from the subset of models – this is consistent with the current movement toward multi-model inference.

Due to the nature of the Dantzig Selector in **R** it was used separately on the DSD and CCD with δ estimated using the automatic procedure suggested by Candes and Tao (2007) and $\gamma = 1.5$ and 20.

The Fermentation Experiments

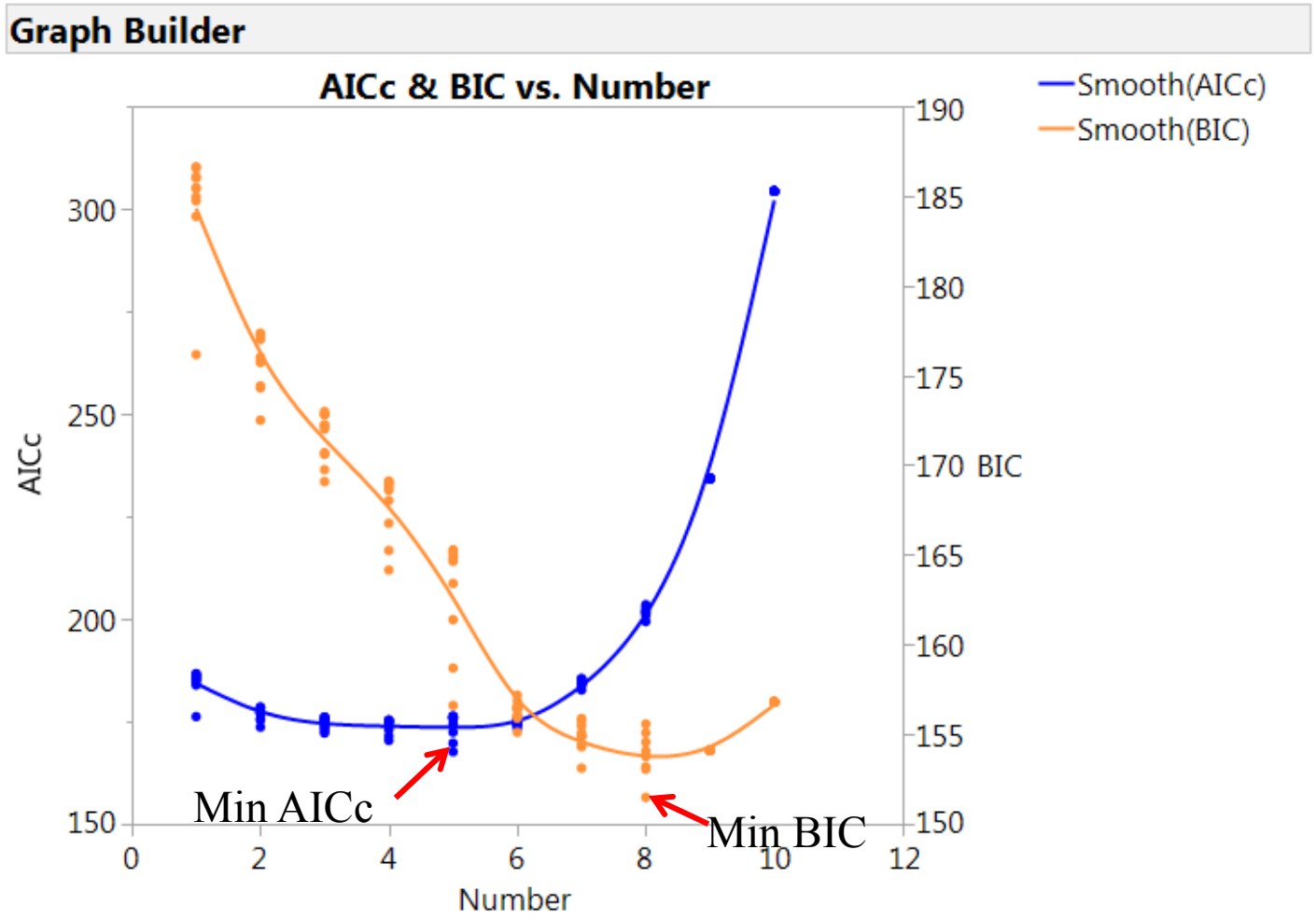
Different choices of the tuning parameters and estimation criterion would no doubt lead to different solutions.

Future work includes studying multiple solutions.

γ=20	DSD	CCD
Main Effects	-33.983	
	-35.9943	
	95.66023	102.8894
	19.64886	
Quadratic Effects	62.44375	53.4563
	42.31875	113.5155
	81.11875	60.82316
	20.61875	43.08352
	177.6188	25.45926
Interaction Effects		
		24.87
		-54.155
		52.37375

The Fermentation Experiments

Below is a plot of AICc and BIC from All Possible Subsets Regression.



The Fermentation Experiments

Below are the best AICc and BIC models. Heredity was not enforced in the All Subsets Regression analysis.

Effect Summary Best AICc Model p = 5

Source	LogWorth	PValue
Feed rate, mL/hr(1.9,3.5)	4.669	0.00002
Feed rate, mL/hr*Feed rate, mL/hr	3.169	0.00068
Induction Temperature C*Feed rate, mL/hr	2.855	0.00140
%DO(20,40)	2.484	0.00328
pH(6.8,7.2)	1.806	0.01563

Effect Summary Best BIC Model p = 8

Source	LogWorth	PValue
Feed rate, mL/hr(1.9,3.5)	4.130	0.00007
Feed rate, mL/hr*Feed rate, mL/hr	2.966	0.00108
Induction Temperature C*Feed rate, mL/hr	2.850	0.00141
%DO(20,40)	1.946	0.01133
pH(6.8,7.2)	1.839	0.01448
Induction OD600(20,40)	1.014	0.09682
%DO*Feed rate, mL/hr	0.963	0.10887
Induction Temperature C(39.5,42.5)	0.686	0.20588 ^

The Fermentation Experiments

The results for the best 8 candidate models using AICc and BIC, the best models are highlighted in the table below.

The smallest Press occurs for the best BIC model, which is smaller than even the other two 8 effect models.

Number	MS(Press)	AICc	BIC	Δ_i	w_i	B_i	BIC wts
4	61.43	170.42	156.61	2.76	0.2514	12.67	0.0018
4	62.71	171.51	165.26	3.85	0.1457	13.76	0.0010
5	45.74	167.66	156.61	0.00	1.0000	5.12	0.0775
5	48.35	169.75	158.70	2.09	0.3516	7.21	0.0273
7	41.79	182.75	153.13	15.10	0.0005	1.63	0.4428
8	18.21	199.42	151.50	31.76	0.0000	0.00	1.0000
8	33.43	200.98	153.06	33.33	0.0000	1.57	0.4570
8	31.93	201.12	153.20	33.47	0.0000	1.70	0.4265

The Fermentation Experiments

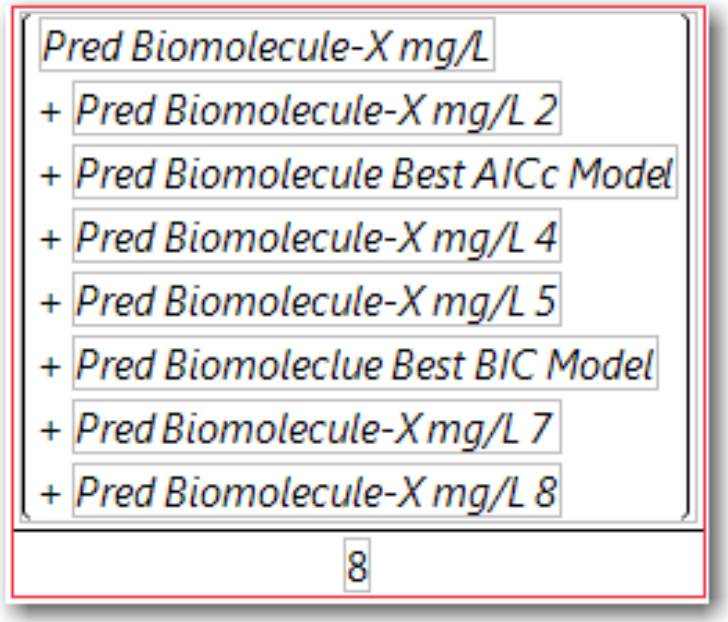
In the present case, we will use the arithmetic average of the predictions from the 8 models selected as best, so models with 4 through 8 effects are used to generate the average.

A simple arithmetic averaging is being used here, but many other weighting schemes can be used.

The averaging has the effect of mitigating under and over fitting.

Little research has been done on prediction averaging for the analysis of designed experiments.

Prediction averaging is now often used in predictive analytics.



The Fermentation Experiments

Finally 8 models fit to the DSD data were compared in terms of how well they predicted on the CCD data.

The models are rated by the root average square error (RASE).

The top three models were comparable with the best BIC model being slightly better.

Model Comparison

Measures of Fit for Biomolecule-X mg/L

Predictor	Creator	RASE	AAE	Freq
Pred Biomoleclue Best BIC Model	Fit Least Squares	85.676	66.797	31
Pred Biomolecule Best AICc Model	Fit Least Squares	88.332	70.754	31
Prediction Average	Model Comparison Model Averaged	88.577	65.173	31
Pred Biomolecule-X mg/L 8	Fit Least Squares	97.963	74.530	31
Pred Biomolecule-X mg/L 2	Fit Least Squares	111.94	86.575	31
Pred Biomolecule-X mg/L 4	Fit Least Squares	115.51	83.734	31
Pred Biomolecule-X mg/L 7	Fit Least Squares	120.96	83.728	31

Conclusions

At this point it is difficult to draw firm conclusions from the simulation study.

In general, the Dantzig Selector using BIC as the estimation criterion seems to perform better than using the AICc criterion.

Unfortunately, on the DSD simulation results there is no clear cut winner overall in terms of model selection between Dantzig Selector, Forward Selection, and All Subsets Regression.

The results are preliminary and need much more analysis.

Generally, the DSD and BDD designs had comparable results in terms of analysis by the Dantzig Selector,

However Forward Selection seems to perform poorly overall on the BDDs.

Conclusions

For the Dantzig Selector more work needs to be done on how to appropriately choose γ based on the problem in study.

The Dantzig Selector performed as well as and in some cases better than All Subsets Regression.

The Dantzig Selector is by far the more computationally efficient of the two methods and therefore seems a better overall choice for analysis given the results (All Subsets is infeasible for large p).

Often multiple solutions are considered in the analysis of experimental results, but almost simulation studies rely on automatic selection of a best solution.

The power of both All Subsets Regression and the Dantzig Selector would no doubt improve if multiple solutions were considered

Implementing multiple solutions in a large simulation is difficult.