DATAWorks

Defense and Aerospace Test and Analysis Workshop

ADSURGO
ascend with analytics

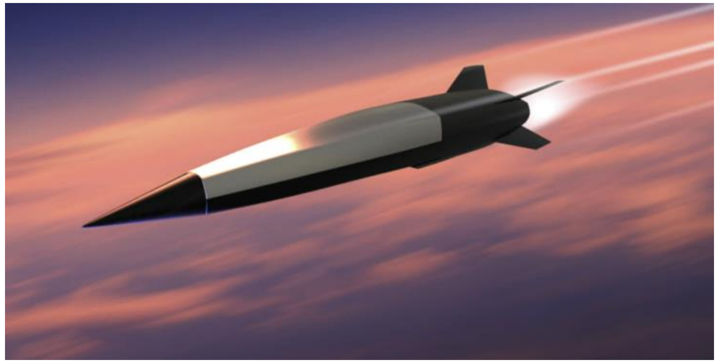
Effective Application of Self-Validated Ensemble Models in Challenging Test Scenarios

Jim Wisnowski
Andrew Karl
Adsurgo

Jim Simpson
JK Analytics

26 April 2023

The Problem



- We are behind in hypersonics while China and Russia aggressively demonstrate fielded capabilities
- High priority for DoD to field more operational systems in near future
- Tech Readiness Levels have to be accelerated to meet IOC goals
- Key challenge for many contenders is the complexities of fuel development to meet range and > Mach 5 requirements
- Fuel performance is a function of many ingredient classes, material choices within classes, process factors, and chemical & physical constraints

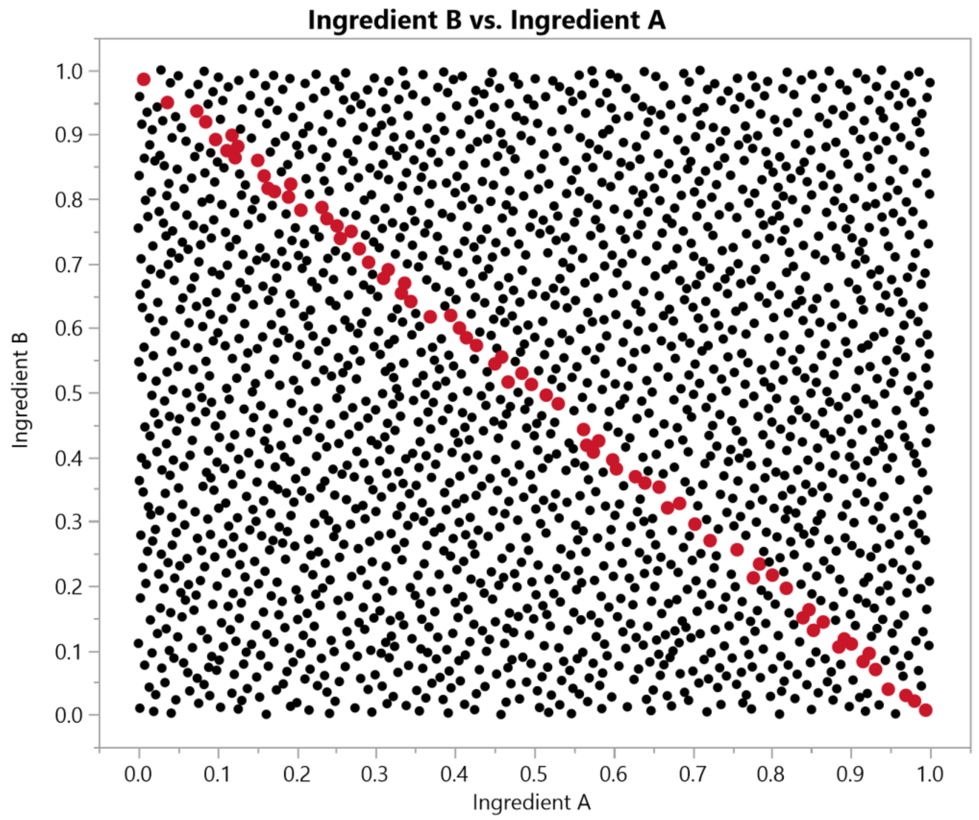
How Did We Get Here?

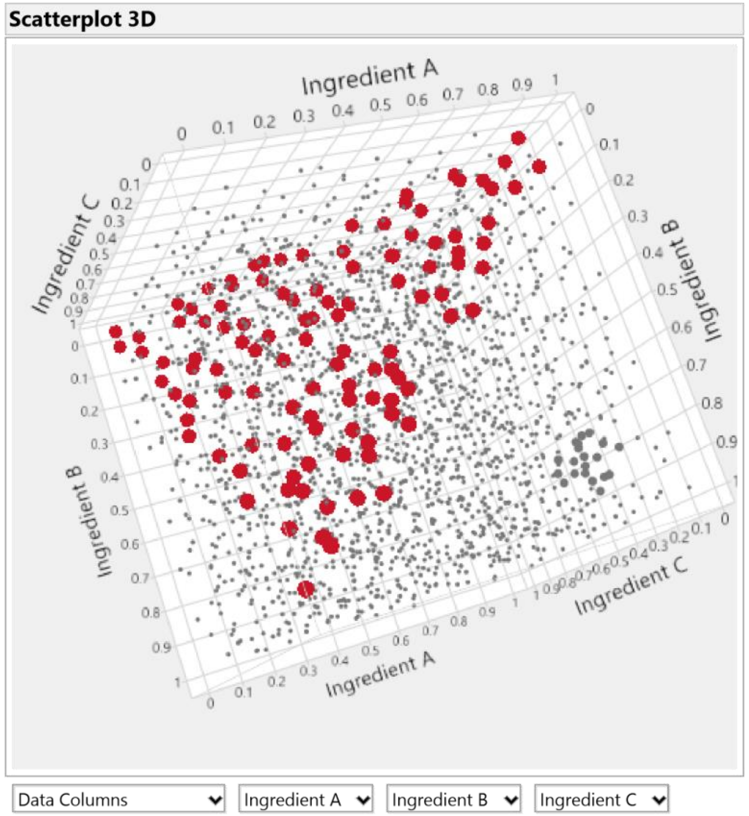
- PM: I just got my Master’s in IE from ASU and wanted Doug Montgomery to consult with us on some propellant formulations. He’s too busy and we tried several other recommendations but no luck. Can you help?
- Jims: Probably. What are you trying to do?
- PM: Did I tell you we wanted Doug Montgomery?
- Jims: Yes you did. So tell us about your test program.
- PM: We need to figure out the right percentage of chemicals to formulate a fuel that is twice as good as anything we’ve ever made.
- Jims: Sounds like a mixture design. Not sure we can get you to 2x, but we can come up with some good design and analysis approaches. How many mixture components do you have?

How Did We Get Here?

- PM: We have 7 main classes (A, B, C…G) and can blend anywhere between 3 and 10 chemical components within each class. But it's not as simple as that because we have stoichiometry constraints within and across classes as well as a lot of other subject matter expertise logic like if you use E2 you can't have more than 20% of A1. See, that's why we need Doug Montgomery.
- Jims: Hmmm. And you said it needs to be twice as good. What are the responses?
- PM: Well we have dozens. Depends on if you want to optimize manufacturability, mechanical performance, tactical performance, range, weight...etc. And not all of them are equally important. Can you help?
- Jims: Certainly. Write this down…doug.montgomery@asu.edu

Mixture Models: Design Implications



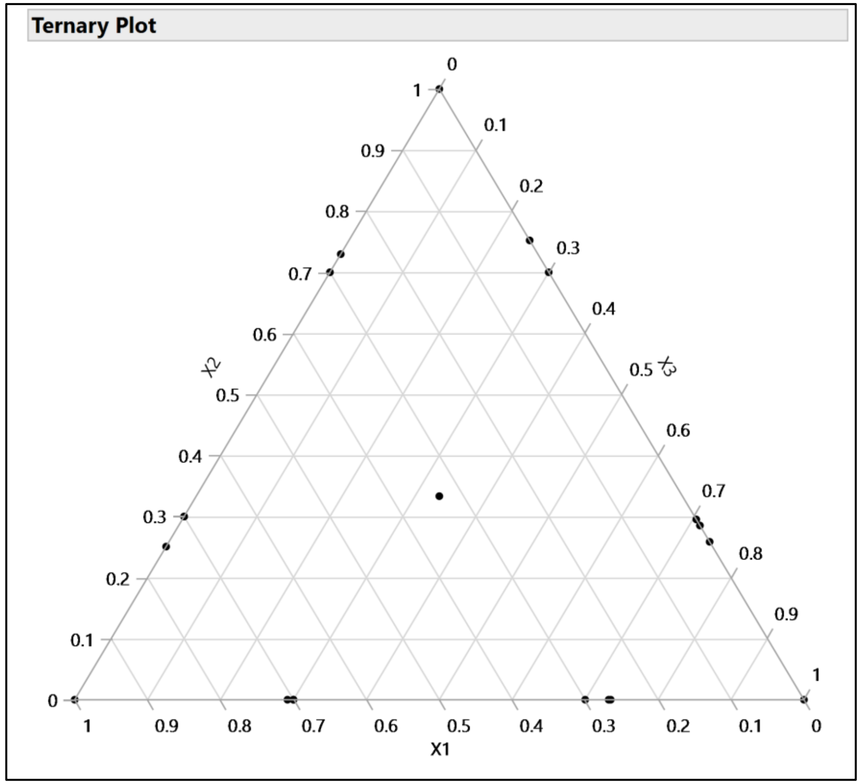


Constraint that *n* mixture factors add up to constant (usually 1) means that the allowable factor space is an *n-1* dimensional subset

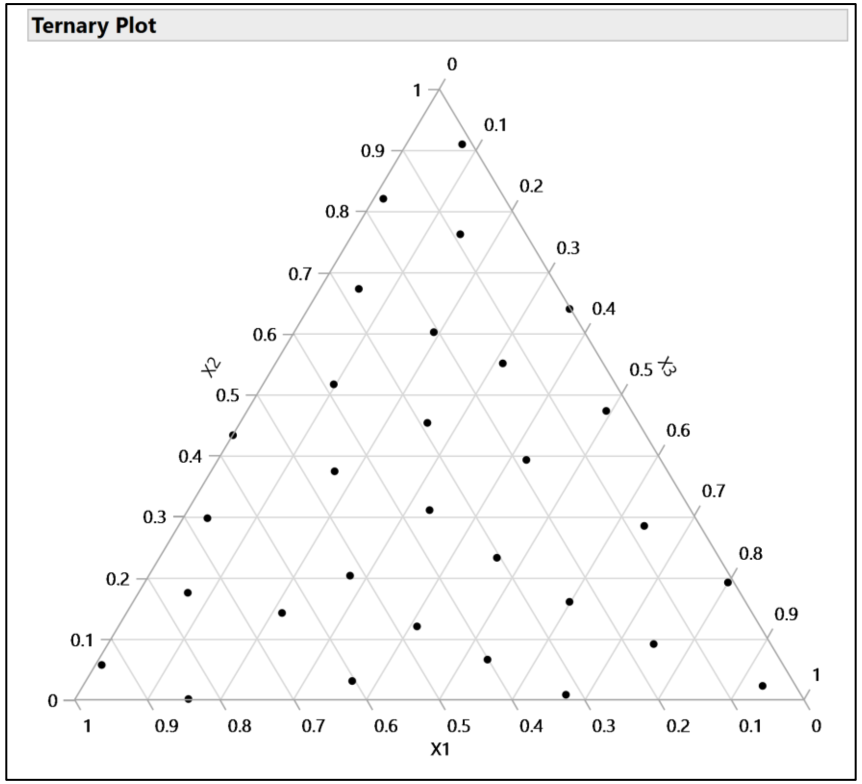
Mixture component dependency = multicollinearity =>increased variance=>Type 2 errors not finding active factors

Mixture Constraint: Design Choices

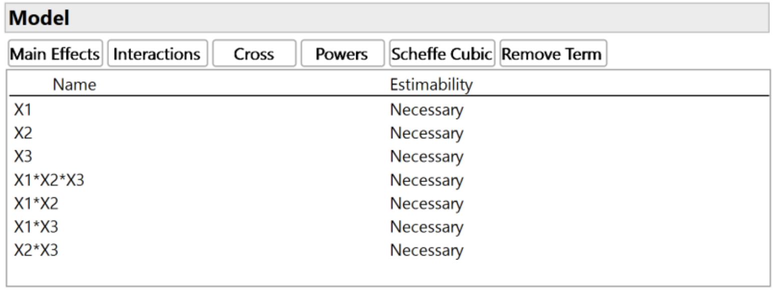
30 runs, D-Optimal for Scheffe Cubic



30 runs, Space Filling



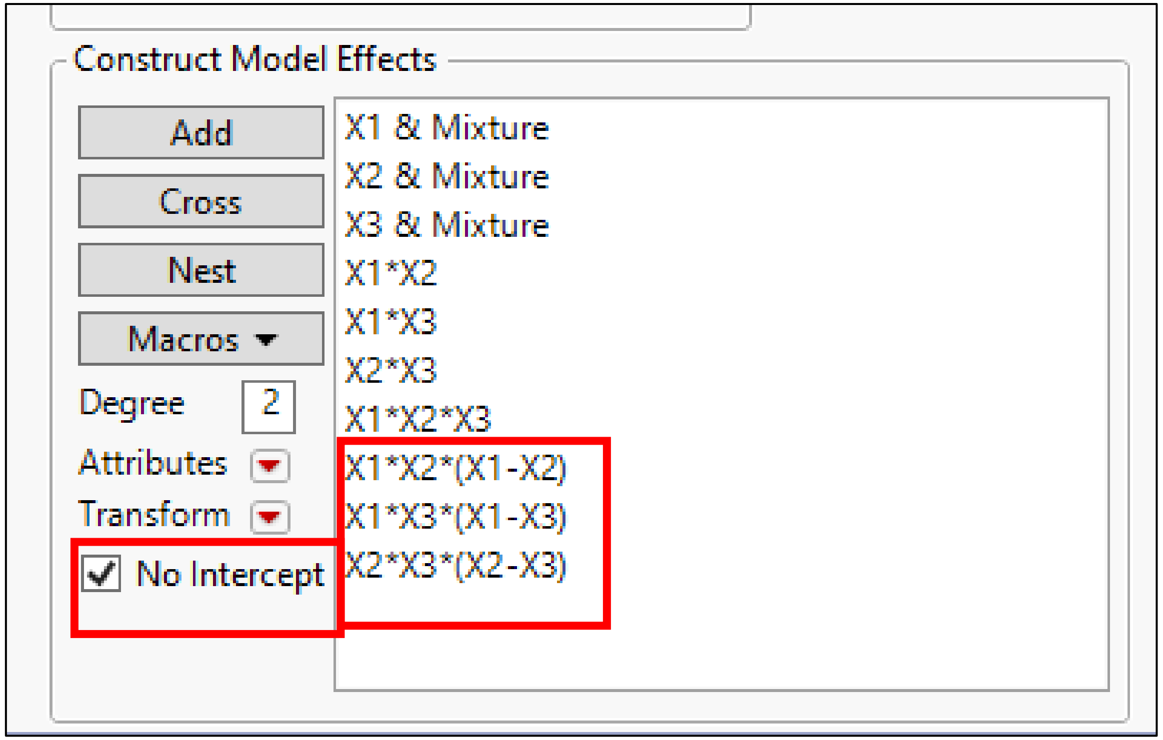
Optimal Mixture Designs



- Optimal designs tend to place most of the runs on the boundaries of the factor space, according to the hypothesized model
- However, in mixture experiments, it is more common than in non-mixture experiments for there to be important high-order terms (third or fourth order). If these are not specified during the design phase, they may be aliased and impossible to detect.
- We also see more instances of experiments where there is a nonlinear drop-off in the response (or outright process failure) near one or more of the boundaries. This could be remedied with tighter factor limits, but this can be expensive.

Mixture Constraint: Analysis Implications

- Linear regression model is used, just like for analysis of a non-mixture design. However, some restrictions are necessary.
- The intercept cannot be fit simultaneously with all of the main mixture effects.
- Scheffé Cubic terms are used instead of quadratic terms



Singularity Details

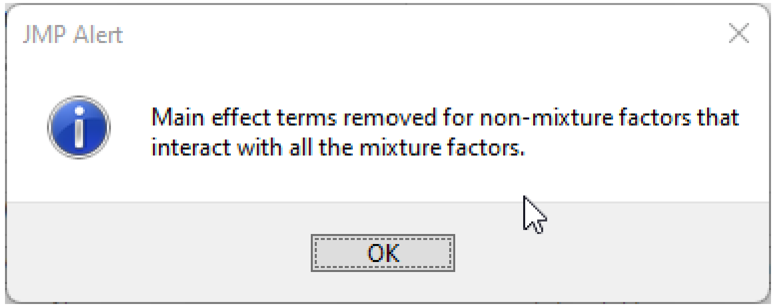
Term	Details
Intercept	=X1(Mixture) + X2(Mixture) + X3(Mixture)

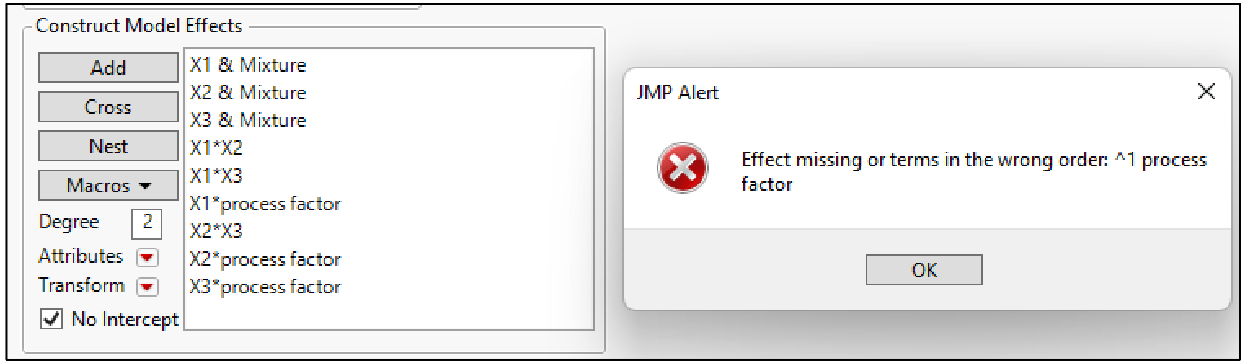
Singularity Details

Term	Details
X1(Mixture)	=X1*X2 + X1*X3 + X1*X1
X2(Mixture)	=X1*X2 + X2*X3 + X2*X2
X3(Mixture)	=X1*X3 + X2*X3 + X3*X3

Mixture Constraint: Analysis Implications

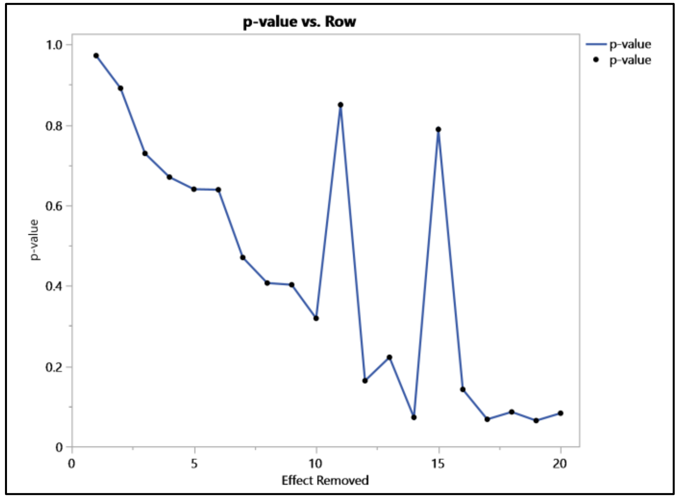
- Non-mixture factors that interact with the mixture factors must have their main effect excluded from the regression, similar to the intercept
- This leads to extra work for base JMP users wanting to use the Stepwise personality of Fit Model
- This is not an *a priori* theoretical constraint from the process, but simply a compromise made so that the regression problem is not ill-conditioned



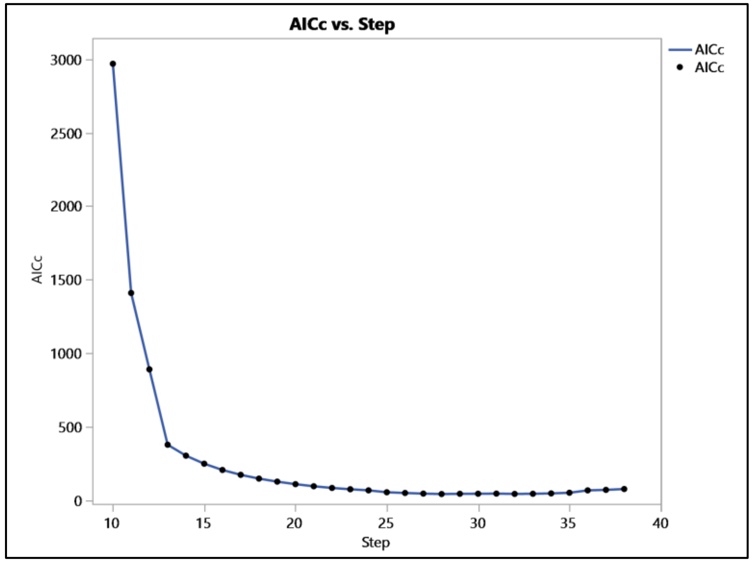


Mixture Constraint: Analysis Implications

- P-values for all terms are "unstable" when reducing the model, due to multicollinearity
- In base JMP, we prefer using min AICc as opposed to using the p-values for backward selection
- In JMP Pro, Self-Validating Ensemble Models (SVEM) provide a solution that is both easier and more accurate (both with respect to finding the optimal *X* and predicting its *Y*)



Backwards selection from full model using p-value threshold of 0.05. Graph shows p-value of removed effect.



Backwards selection from full model using p-value min AICc. Graph shows AICc with each removed effect.

Practical Questions for Mixture Designs

- Should I use a space filling or an optimal design?
- If optimal, should I use D-, I-, or A-Optimal?
  - Should I expend any runs on replicates or center points?
  - What is the marginal benefit (to the goal of optimization) of adding one or more points, if power is not a useful metric?
- How should I analyze the results?
  - Do any of the design choices listed above affect how I should analyze the results? Or is there a near-uniformly best modeling approach?
  - “No Intercept” is the default. Should I consider including it?
  - Main effects are forced by default. Should they be if an intercept is allowed?

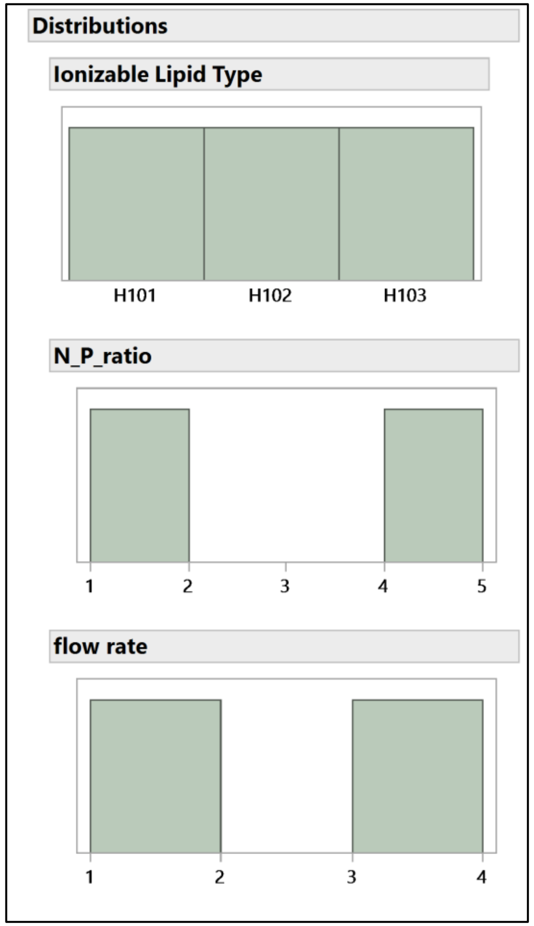
Simulation Models for Mixture Designs

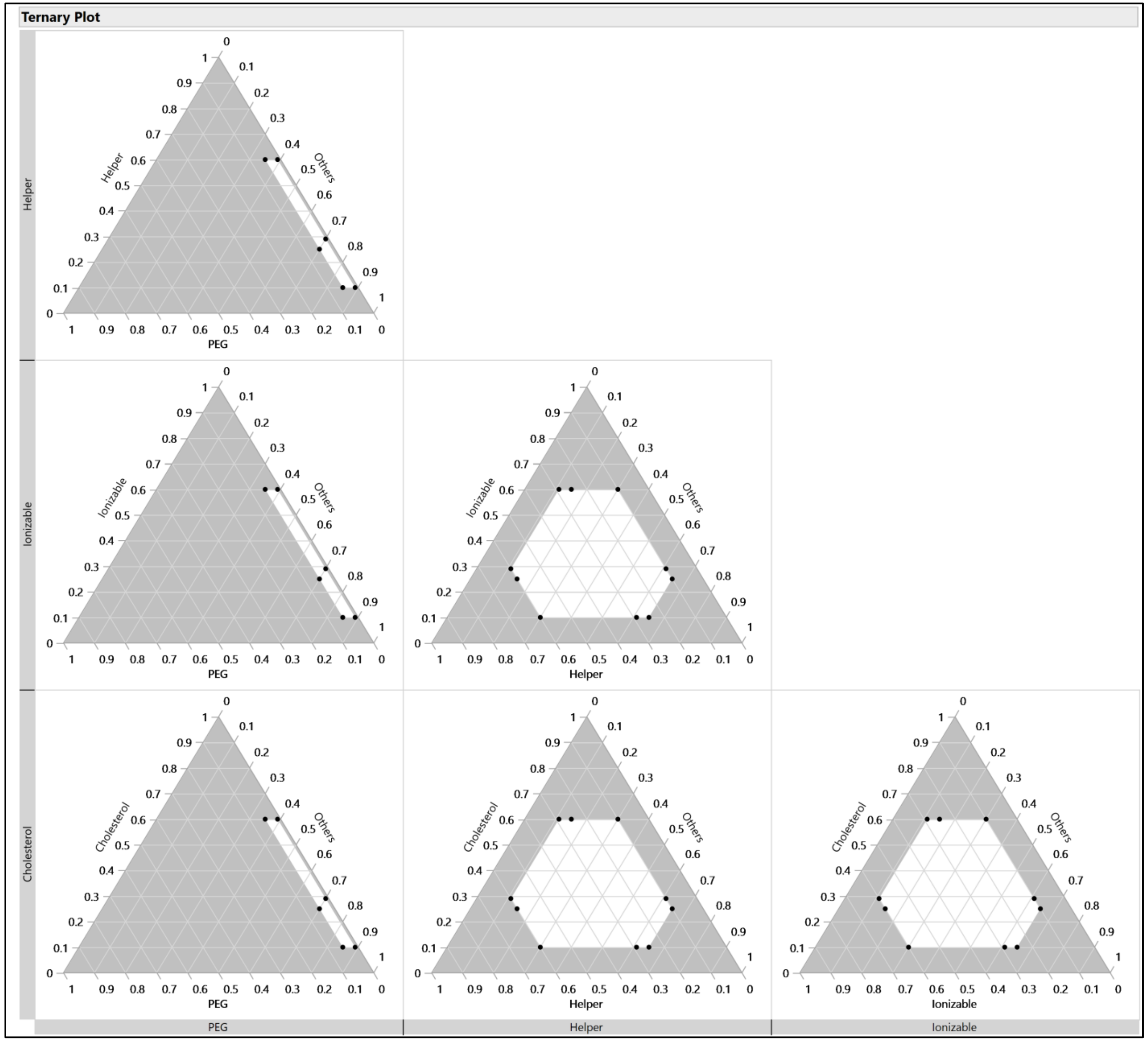
- Simulation used to answer the practical questions posed.
- Use a fairly complex test scenario with 4 mixture factors, 2 continuous process factors, and 1 three-level categorical factor.
- We consider two different situations
  - 12 run design (optimal designs are built around main effects)
  - 40 run design (optimal designs are built around second order model plus third order in mixture effects).
- We also briefly consider the difference in performance between a 12 and a 17 run space filling design.
- Our goal is to compare the impact of different design and analysis options in these scenarios to provide practical guidance.

Simulation Model Metric

- We typically measure the performance of predictive models by considering the difference between the *predicted response* and the *actual response* (e.g. RMSE, RMPSE).
- By contrast, for formulation optimization, we are more concerned with the difference between the *predicted optimal location* and the *actual optimal location*.
  - Since there may be a ridge in the response surface, the important difference here is not the distance between these points, but the difference in actual responses at these points
- In the simulations, we record the “percent of maximum” of each candidate optimum, as well as the RMSE of the prediction at that candidate point

12 Run D-optimal (Default Design)





"Default" Settings, model against main effects

- All linear methods restricted to looking at main effects.
- D-optimal design around main effects. Min runs would have been 10, this used 12.
- Non-SVEM model selection makes things worse because most of the effects are active. Interesting that SVEM brings these methods at least back up to the performance of the full model.

Student's t All Pairwise Comparisons

All Pairwise Differences Connecting Letters

Setting					Least Squares Mean
SVEM-Neural	A				0.92211640
SVEM-LASSO_w_int		B			0.91010924
Full Model		B			0.90881787
SVEM-LASSO_no_int		B			0.90843834
SVEM-FS_w_int		B			0.90768462
SVEM-FS_no_int		B			0.90663050
Lasso_AICc_no_int			C		0.88670002
Backward AICc_no_int			C		0.88355285
Forward_AICc_no int			C		0.88175927
Lasso_AICc_w_int				D	0.82697684

Levels not connected by same letter are significantly different.

Build design around main effects, allow model to consider third order effects

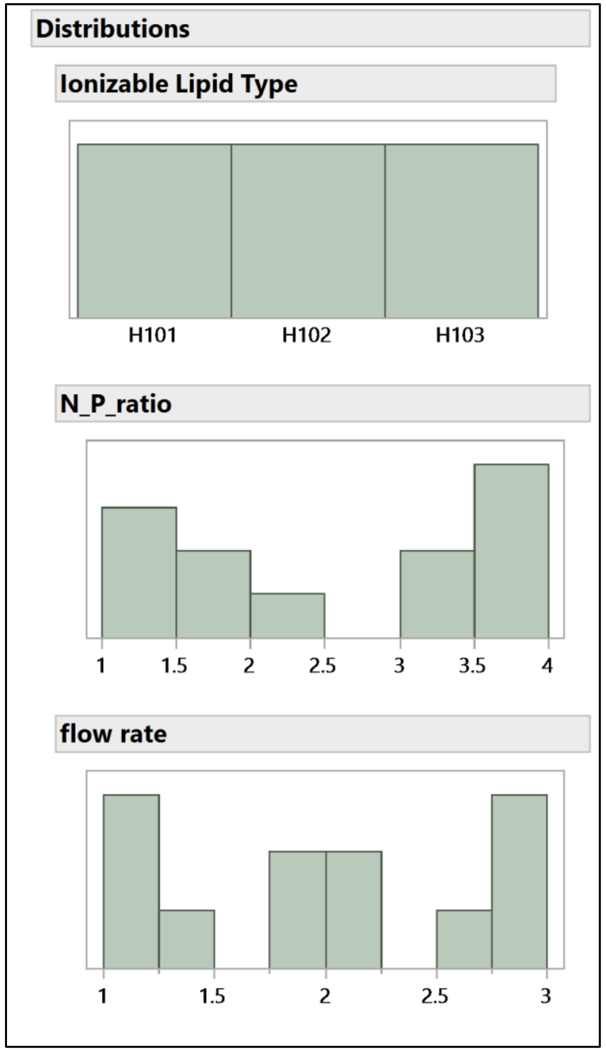
- D-optimal design around main effects. Min runs would have been 10, this used 12.
- SVEM-FS and SVEM lasso now do better than SVEM-Neural.

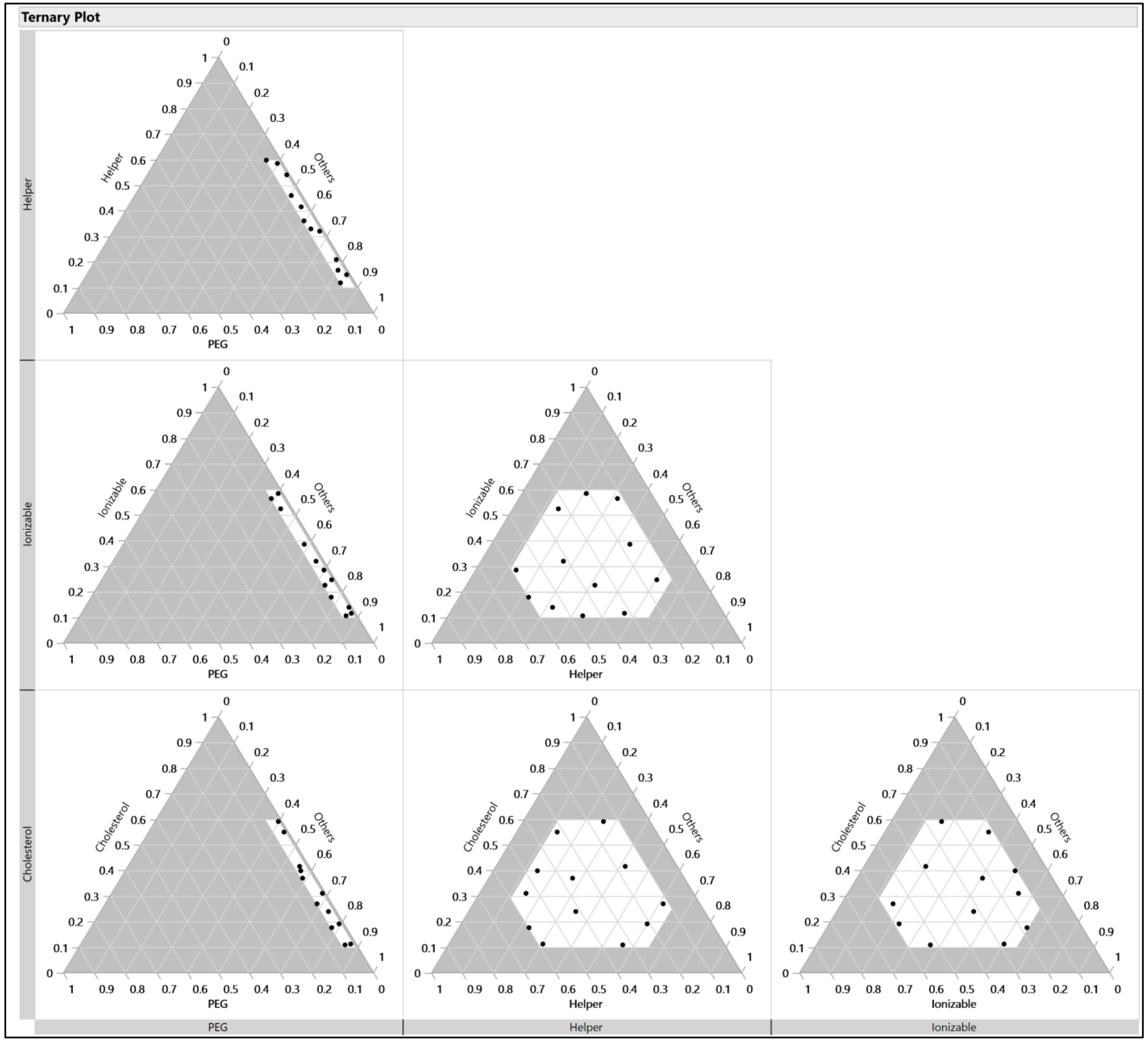
All Pairwise Differences Connecting Letters

Setting								Least Squares Mean
SVEM-FS_w_int	A							0.97297833
SVEM-LASSO_w_int		B						0.96206123
SVEM-FS_no_int			C					0.94001927
SVEM-Neural				D				0.91654380
Lasso_AICc_w_int					E			0.89043952
Forward_AICc_no int						F		0.87960375
Lasso_AICc_no_int						F		0.87797807
SVEM-LASSO_no_int							G	0.86205995

Levels not connected by same letter are significantly different.

12 Run SFD


Distributions
Ionizable Lipid Type
H101
H102
H103
N_P_ratio
1
1.5
2
2.5
3
3.5
4
flow rate
1
1.5
2
2.5
3


Ternary Plot
Helper
Ionizable
Cholesterol
Others
PEG
Helper
Ionizable

Comparing design types for 12 run design when allowing third order effects

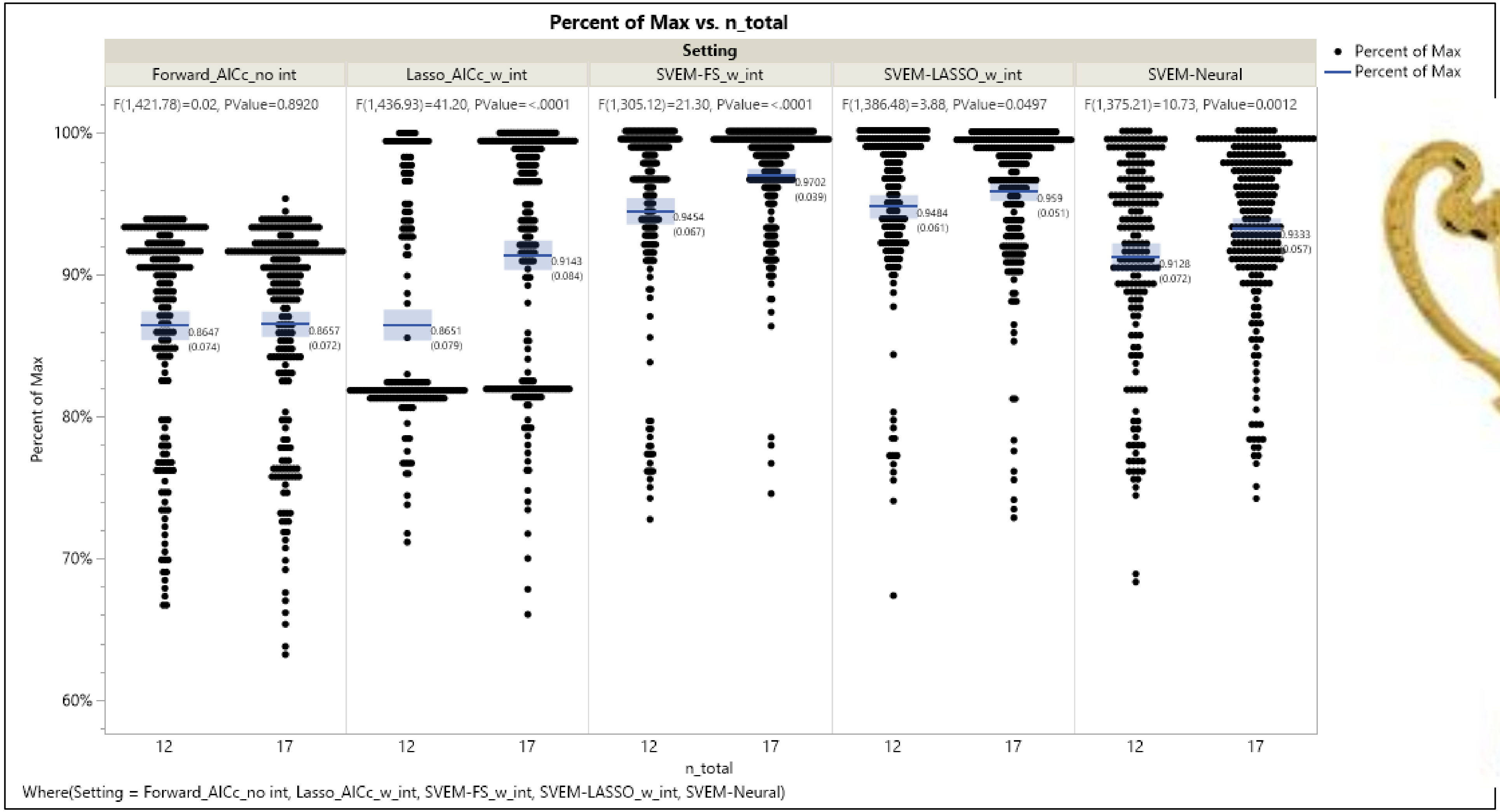
- SVEM methods significantly better than classical
- Beware SVEM Lasso without intercept
- Not a significant difference between the Space Filling and D-optimal

All Pairwise Differences Connecting Letters

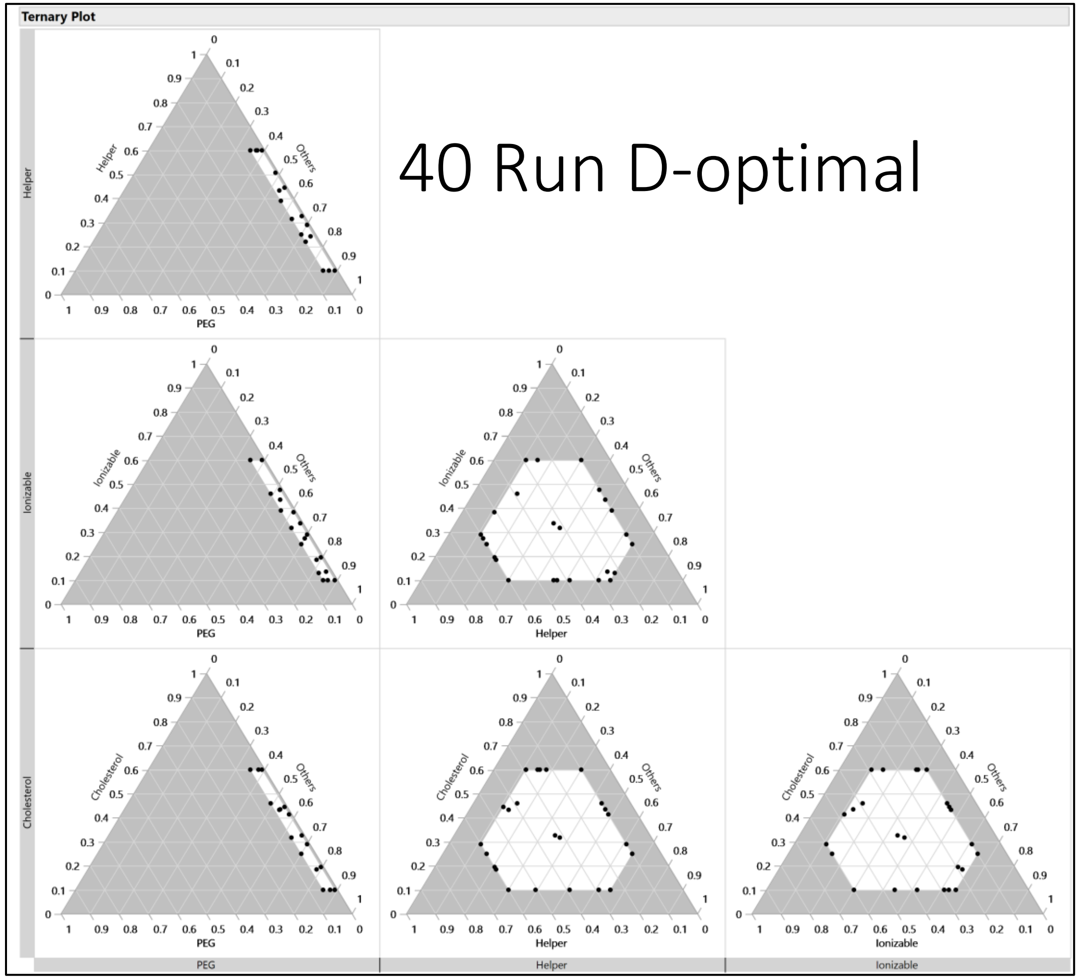
Concatenate[Setting,design_type 2]								Least Squares Mean
SVEM-FS_w_int_D-Optimal	A							0.97297833
SVEM-LASSO_w_int_D-Optimal		B						0.96206123
SVEM-LASSO_w_int_SFD			C					0.94843459
SVEM-FS_w_int_SFD			C					0.94544932
SVEM-FS_no_int_D-Optimal			C					0.94001927
SVEM-Neural_D-Optimal				D				0.91654380
SVEM-Neural_SFD				D				0.91278197
SVEM-FS_no_int_SFD					E			0.89836361
Lasso_AICc_w_int_D-Optimal					E			0.89043952
Forward_AICc_no int_D-Optimal						F		0.87960375
Lasso_AICc_no_int_D-Optimal						F		0.87797807
SVEM-LASSO_no_int_SFD						F		0.87757522
Lasso_AICc_w_int_SFD							G	0.86507188
Forward_AICc_no int_SFD							G	0.86472871
Lasso_AICc_no_int_SFD							G	0.86442280
SVEM-LASSO_no_int_D-Optimal							G	0.86205995

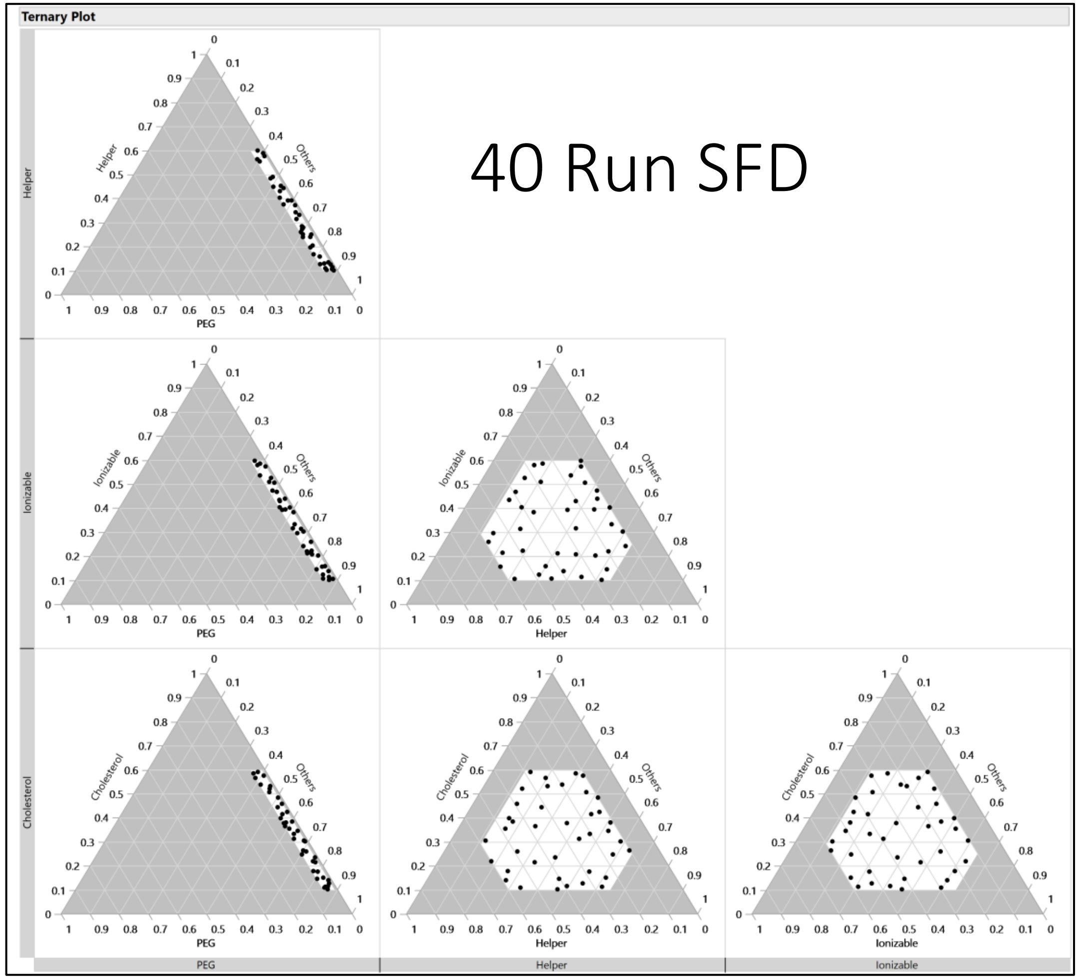
Levels not connected by same letter are significantly different.

Marginal benefit of increasing from 12 to 17 runs in SFD? Actually, not much.



40 Run Designs





40 Run D-Optimal Design

Student's t All Pairwise Comparisons

All Pairwise Differences Connecting Letters

Setting							Least Squares Mean
SVEM-FS_w_int	A						0.99142698
SVEM-LASSO_w_int	A						0.99125355
SVEM-FS_no_int	A	B					0.99049913
Forward_AICc_no int	A	B	C				0.98746014
Lasso_AICc_w_int		B	C				0.98678132
Full Model			C				0.98580034
Backward AICc_no_int			C				0.98526983
SVEM-Neural				D			0.97063341
SVEM-LASSO_no_int					E		0.95928419
Lasso_AICc_no_int						F	0.90074564

Levels not connected by same letter are significantly different.

40 Run Space Filling

Student's t All Pairwise Comparisons

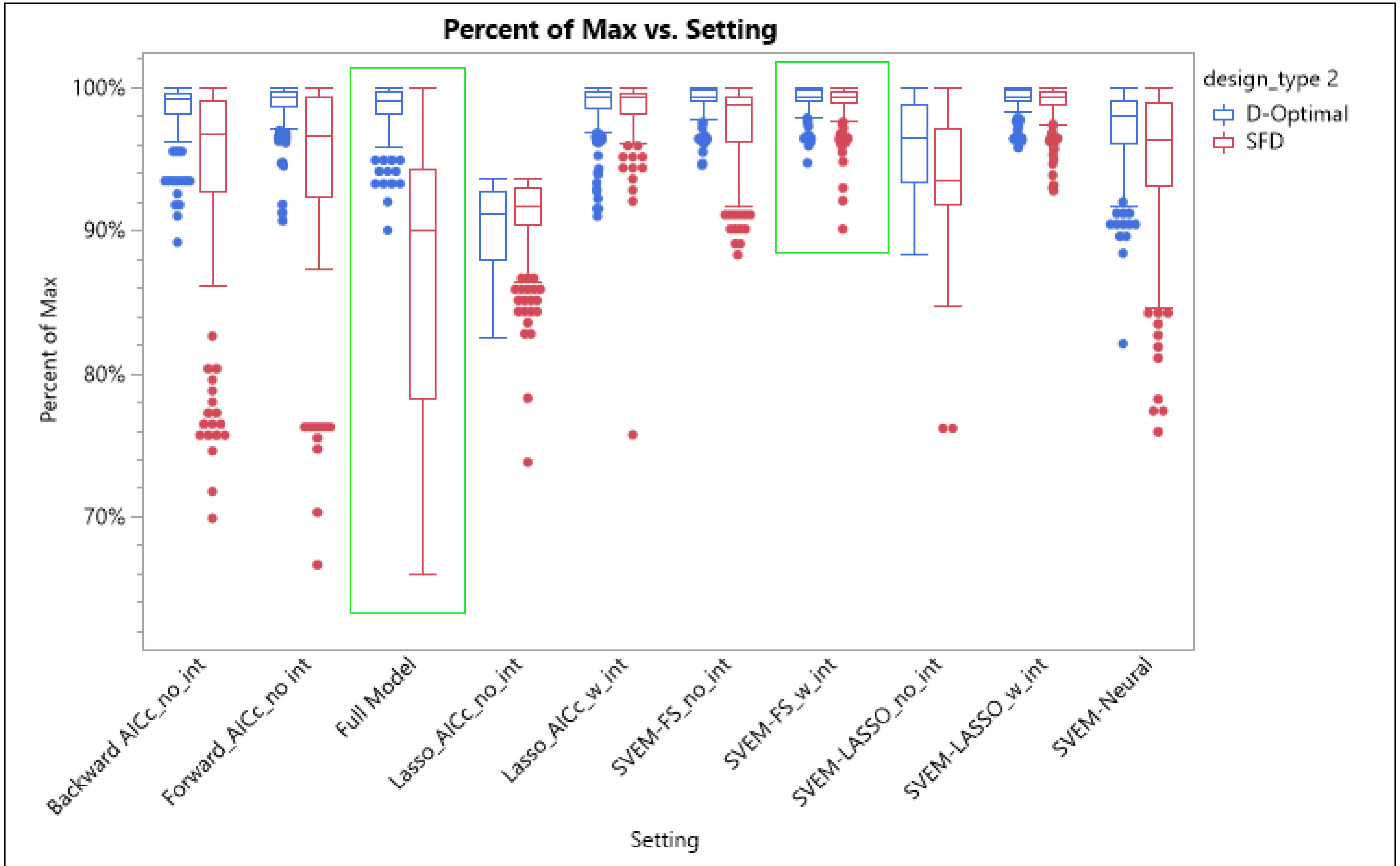
All Pairwise Differences Connecting Letters

Setting							Least Squares Mean
SVEM-FS_w_int	A						0.98893800
SVEM-LASSO_w_int	A						0.98874953
Lasso_AICc_w_int	A						0.98581645
SVEM-FS_no_int		B					0.97360330
Forward_AICc_no int			C				0.95152822
SVEM-Neural			C				0.95109017
Backward AICc_no_int			C	D			0.94565714
SVEM-LASSO_no_int				D			0.93980247
Lasso_AICc_no_int					E		0.91040226
Full Model						F	0.87185363

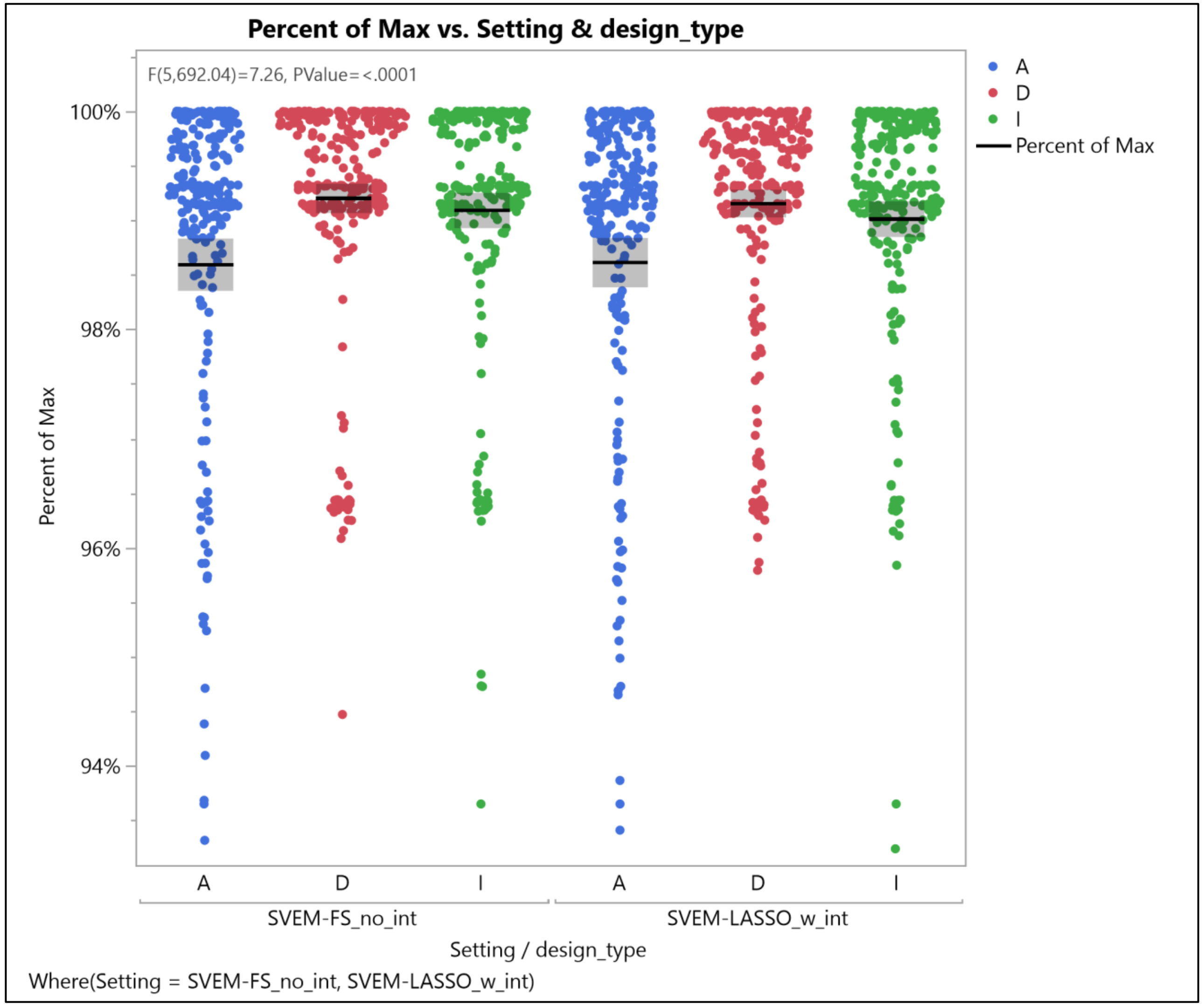
Levels not connected by same letter are significantly different.

- SVEM Forward Selection and SVEM Lasso with intercept best
- Lasso no intercept notably worse (SVEM or not)
- Full model performs poorly for Space Filling

40 Run D-optimal vs SFD

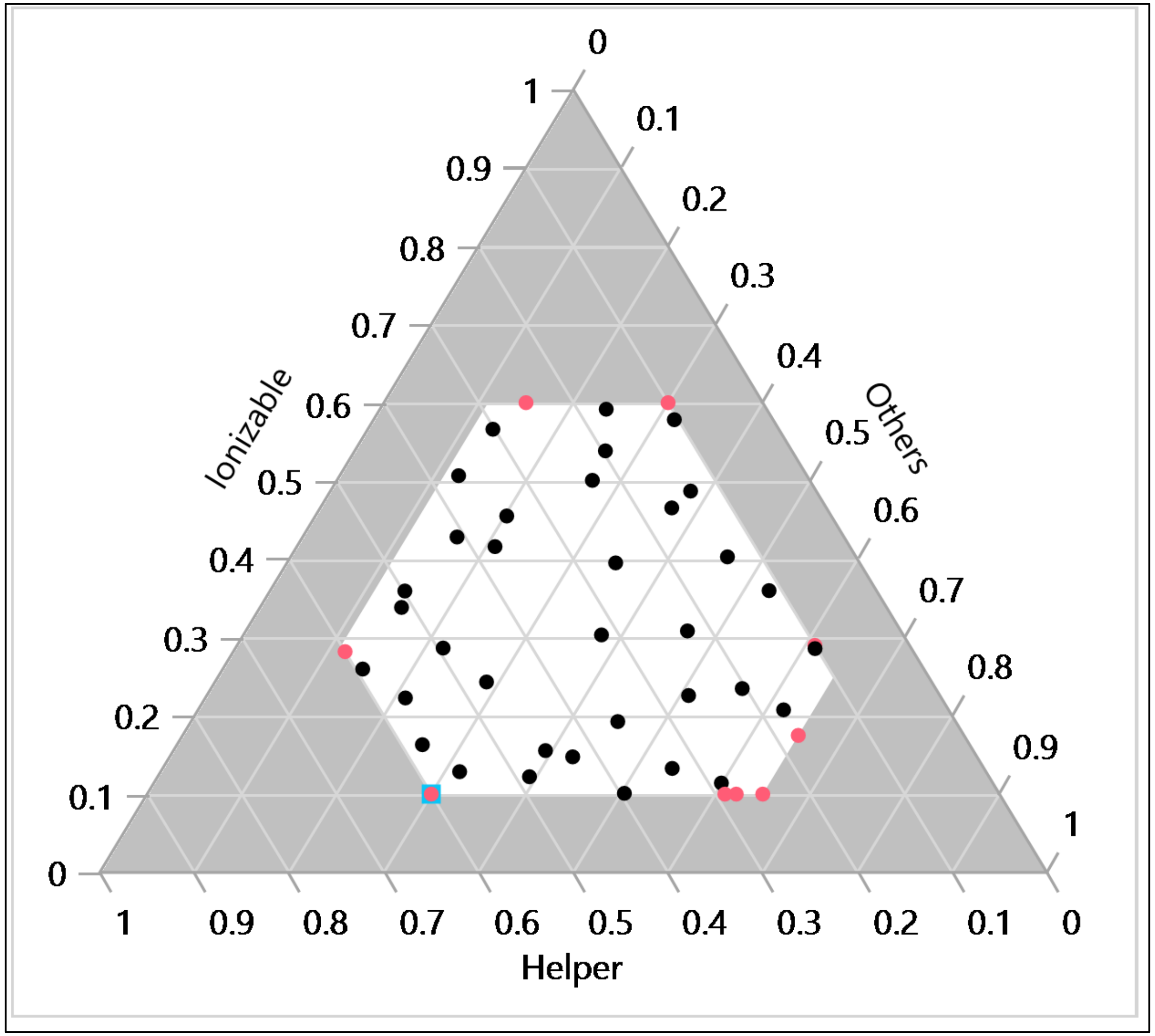


40-run Changing Optimality Criterion

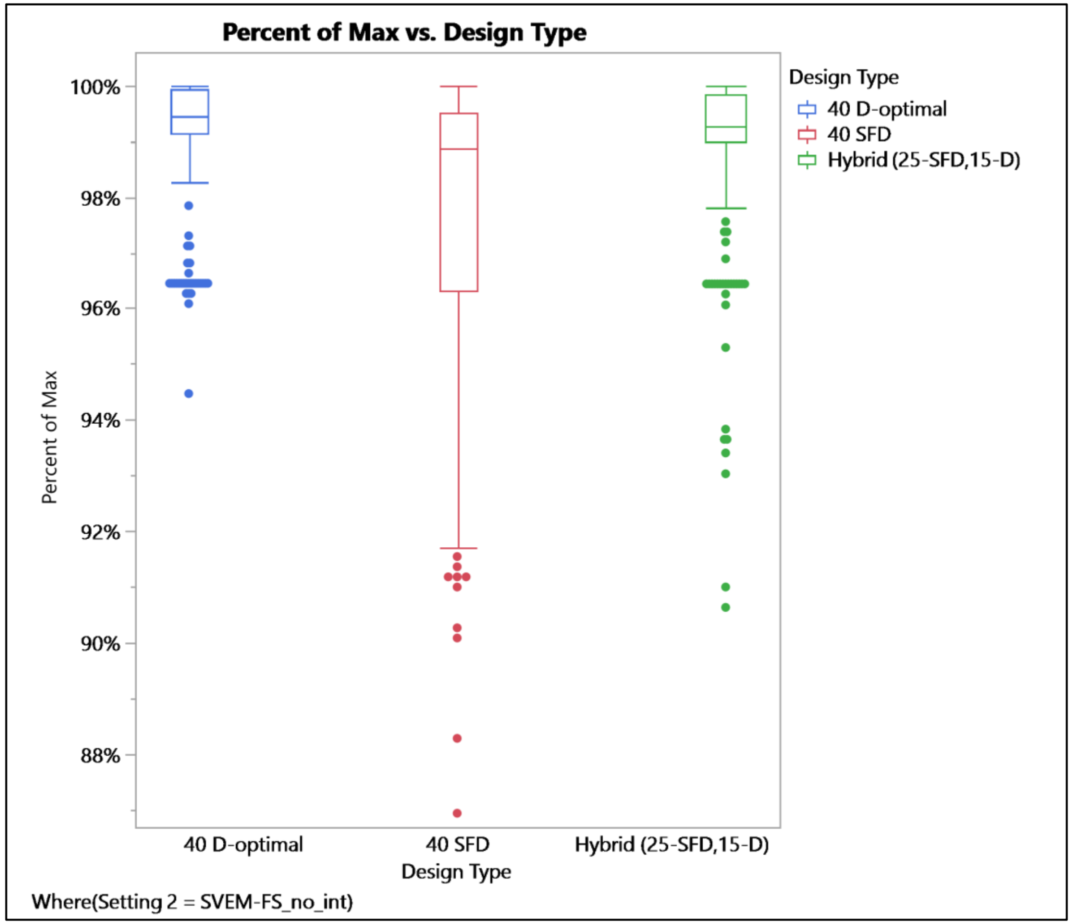


- D vs I are similar
- We really like A-optimal designs for non-mixture studies, but they do tend to give worse results for mixture studies.
- Consistent with what the diagnostics show (e.g. FDS and prediction variance) and visual comparison of ternary plots after years of applying in practice.

Hybrid Design (Space filling with D-augment)



Hybrid designs can match D-optimal results while still offering some SFD advantages



Student's t All Pairwise Comparisons

Quantile = 1.96314, DF = 747.0

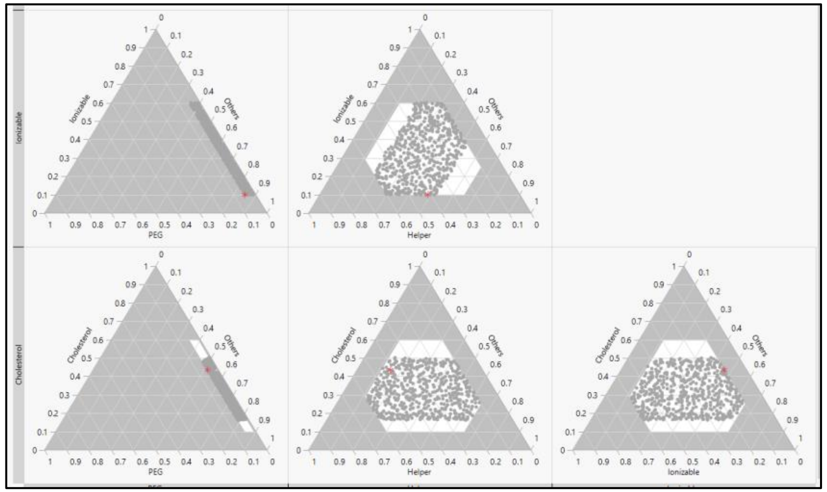
All Pairwise Differences Connecting Letters

Design Type			Least Squares Mean
40 D-optimal	A		0.99207744
Hybrid (25-SFD,15-D)	A		0.98937619
40 SFD		B	0.97581075

Levels not connected by same letter are significantly different.

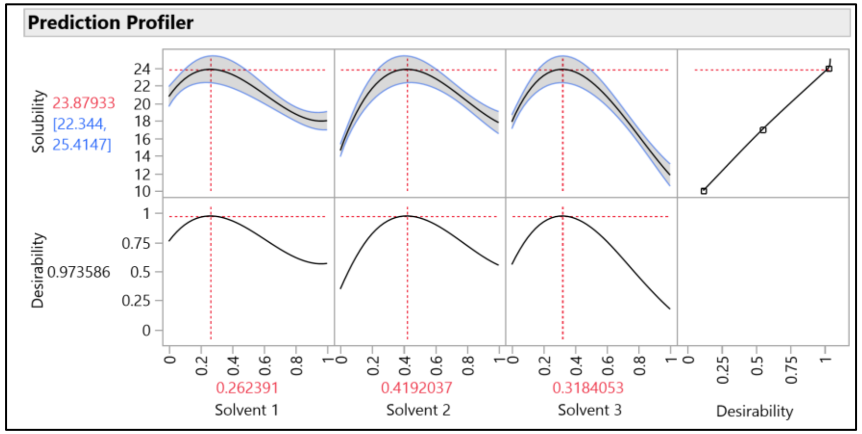
Other Excursions

- If we augment 40 run design (D-optimal), no difference in rankings and percent of max values if add 5 center points, 5 replicated points, or 5 random points. There are other benefits that should be weighed against confirmation runs.
- When add constraint (complex) to D-optimal or SFD, this does not impact the results of percent of max rankings among the techniques

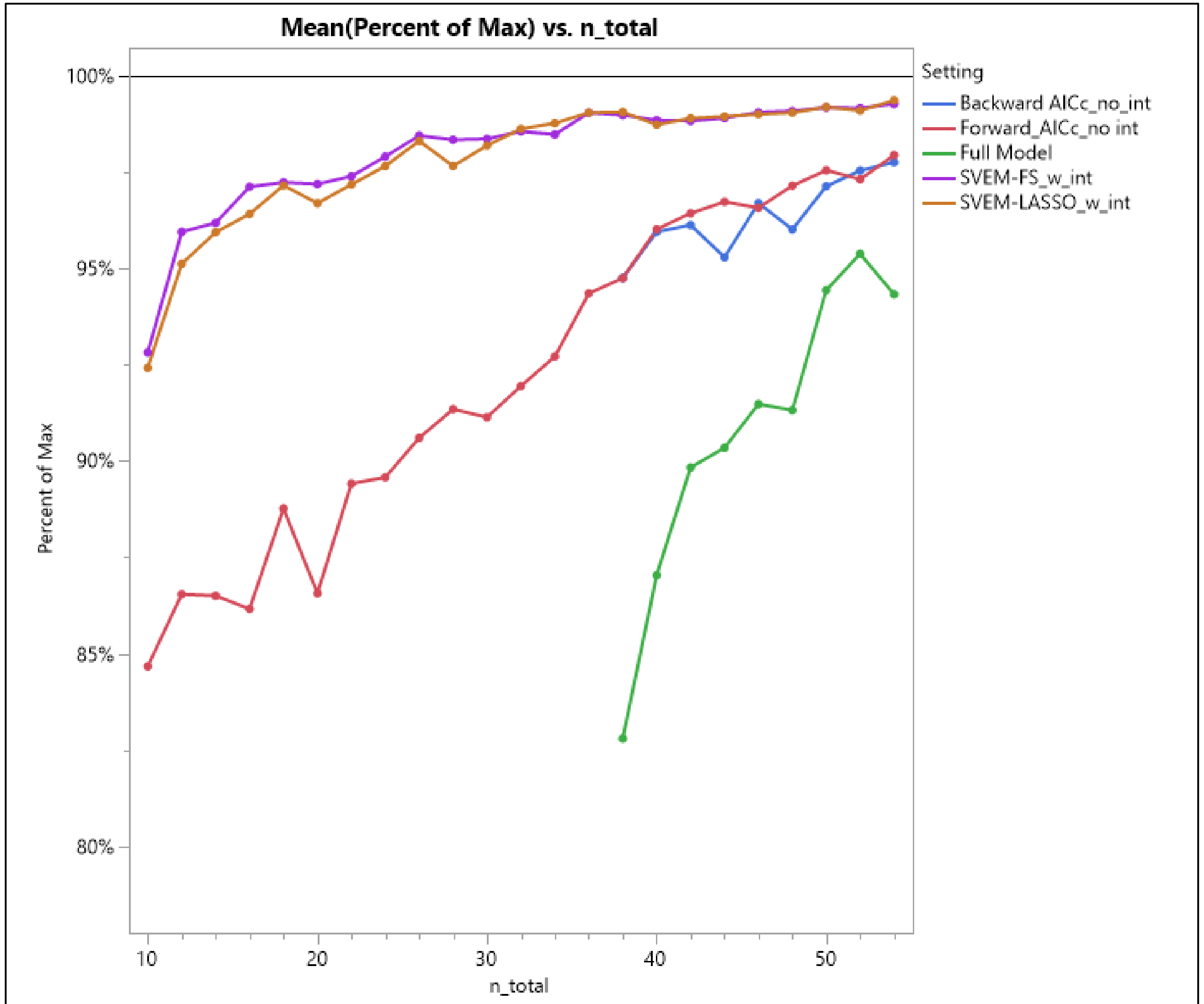


Prediction Easier than Factor Interpretation

- "In the final analysis we may have to learn to live with collinearity in our fitted model, and as a result, we must refrain from trying to interpret the unstable coefficient estimates. Instead, we must rely on the final model form as a tool for plotting the response trace along the component axis...This may be the best that we can do." (John Cornell, p.489, **Experiments with Mixtures**, 2002).



SVEM vs Single Shot model reduction



Highlights

- Space Filling Designs are most promising with smaller designs and/or smaller process variation. For large designs (e.g. capable of supporting a third order model) with high process variance, the optimal (or hybrid) designs outperform as long as there is not concern about failure along boundaries.
- Avoid A-optimal designs with mixture effects present. I-optimal designs can take longer to generate and are prone to convergence failures when a mixture factor has a narrow range and Scheffe cubic terms are included.
- SVEM-FS (with non-default settings for mixture analysis : uncheck “No Intercept” and do not force main effects) provides universally good performance across a wide variety of design sizes and options, and is extremely easy to implement in JMP 17 Pro.
- SVEM-Lasso provides an inferior fit to SVEM-FS unless the No Intercept option is disabled, in which case SVEM-Lasso is competitive with SVEM-FS. A similar contrast is seen in the single-shot Lasso AICc models with and without an intercept included
- SVEM-Neural shows some promise in the case of nearly saturated main effect screening designs, but has a high model variability that can be hedged by also considering the SVEM-FS candidate.
- SVEM-FS produces confidence intervals with a closer-to-nominal coverage rate than single shot FS or backward selection.
- With native SVEM options, JMP Pro provides the opportunity to obtain significantly better optimal formulation candidates than base JMP.

Modeling Options in Base JMP

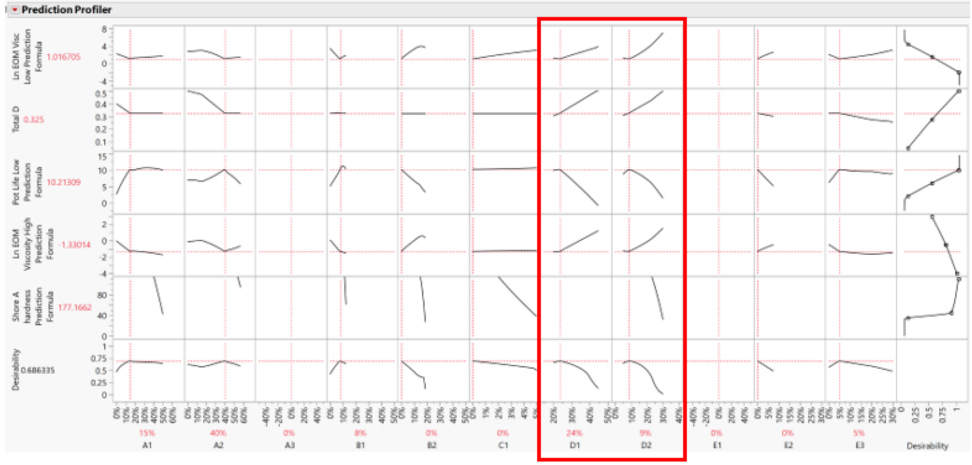
- Full model, backward p-value selection (Effect Summary), AICc or BIC (forward or backward).
- Not as clear of a pattern of a “winner” as was observed with SVEM-FS or SVEM-Lasso (with intercept).
- Seems to be a stronger dependence on the number of active/inactive effects in the candidate model.
- In a large design with third-order candidate effects that includes all of the true effects, AICc (forward) tends to do best.
- In a smaller design where most of the second order candidate effects (plus main effects) are active in the process (and there are also unincluded, active third order effects), then simply sticking with the Full Model “wins”.
- The difficulty is that we do not know how many effects are active before analyzing the experimental results!

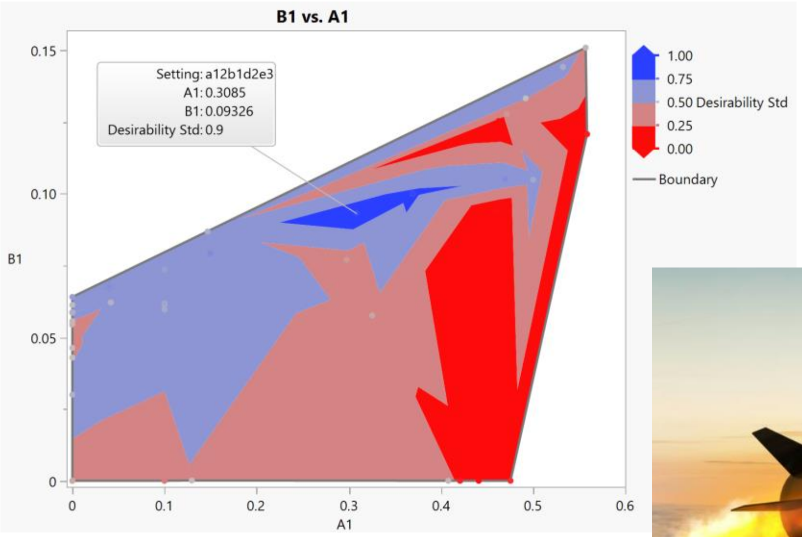
Some rules of thumb for mixture designs

- Preference for space filling designs if there is concern that process may fail along boundaries and there is a desire to map this failure boundary.
- Preference for space filling designs if goal is optimization and number of runs is close to the number of main effect d.f.
- Preference for (D-) optimal designs if the process is understood well enough to be confident that space spanned by specified model effects provides reasonable approximation of response surface.
- Preference for optimal designs if there is a lot of process variability and there are enough runs to support a richer (perhaps second order?) model.
- Also consider hybrid designs when the run-budget is sufficient to consider third order mixture terms.
- Use Profiler > Ouput Random Table with Graph Builder, Ternary Plot, etc, to analyze results

Epilog…how did we do on the hypersonic fuel?

- SVEM was critical to understanding drivers and interactions—dynamic profilers indispensable
- Predictor screening with bootstrapping components insightful
- Old friends of decision trees and neural networks at a minimum complemented understanding
- Most important was working with the SME to interpret weak signals and interactions





SVEM gave a hypersonic boost to test program!