Project 3: Multiple Response Optimizations in Beer Brewing

Maj Beau Nunnally, Capt Kevin Farley, 2Lt Stuart Corbett

November 28, 2016

Background

Barley, the most widely adapted cereal grain in the world, is an ancient crop that has been used for thousands of years for feed, food, and the production of beer. The process by which barley is converted into a usable form for brewing is called malting.

After discussion with theoretical subject matter experts, the following dependent variables were determined to be desirable in production of malted barley: Amylase level, measured in U/g is maximized, Superoxide Dismutase, measured in U/g is maximized, and Dymethyl Sulphide, measured in parts per billion (*ppb*) is targeted for 65 *ppb*. The independent variables (with associated ranges) that are used an inputs for this process were determined to be: Steeping Temperature, $14° - 18°$, Steeping Time, $24 - 48$ hours, Germination Temperature $12° - 20°$, Hydrogen Peroxide Concentration in Steeping Water $0 - 0.2$ g/L, Germination Time, $4 - 8$ hours, Withering Temperature, $35° - 50°$, Drying Temperature, $55° - 65°$, Kilning Temperature $70° - 90°$, and IDK Level, $1 - 11$ Tufnels.

The goal of this experiment is to find the ideal operating conditions to maximize Superoxide Dismutose and Amylase level while maintaining a target rate of 65 *ppb* on Dymethyl Sulphide by altering the independent variables listed above.

Methodology

This experiment was divided into stages as listed below.

Stage 1: Variable Screening

In this stage, we conducted a variable screening experiment using the minimum and maximum feasible values on all independent variables. This had an overall purpose of determining the inependent variables with an affect on each of the dependent variables.

This portion of the experiment was conducted with a Definitive Screening Design, using JMPs formulation. This design was chosen because it will contain non-aliased main effects using a small number of design points. There is an underlying correlation structure among the independent variables using a DSD, which should not impact the determination of main effects as described.

In addition to finding main effects of importance, this stage will allow the reduction of the design space based on regions of desirability; if one of the independent variables moves each of the dependent variables in the desirable direction with the same sign change, we can reduce the space associated with that independent variable.

For this stage an $\alpha = 0.20$ was used for sceening main effects.

Stage 2: Model Fitting

In this stage, we conducted an experiment to fit an appropriate model to relate the independent variables resulting from the screening design to the dependent variables of interest. This model was run with a CCD including 4 center points (2 in each experimental block). A CCD was selected because it allows a full capture of the design region. With the reduction in dimensional from Stage 1, a CCD was possible without excessive sampling from the design space.

For this stage, an $\alpha = 0.1$ was used for screening second order effects.

Stage 3: Analysis

No additional experiments were run for this Stage, but, using the model selected from Stage 2, desirability functions were created from the dependent variable. Finish This Section

Experimental Results

This section will discuss the results from the analysis of data for each Stage described in the Methodology Section.

Stage 1: Variable Screening

Beginning by importing the data generated by model input.

We chose a reduced model through stepwise regression against each of the parameters. The ANOVA table for a reduced model for Amylase Level is included below:

Table 1: Anova Table for reduced model for Amylase Level

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Steeping_Time	1	46481	46481	3.747	0.06973
Germination_Temp	1	117151	117151	9.443	0.006893
Kilning_Temp	1	495600	495600	39.95	7.679e-06
Residuals	17	210907	12406	NA	NA

From this, we can see that the independent variables for Steeping Time, Germination Temperature, and Kilning Temperature are significant for this model.

The ANOVA table for Superoxide Dismutase Level is induced below:

Table 2: Anova Table for reduced model for Superoxide Dismutase Level

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Steeping_Time	1	35806	35806	11.15	0.003883
Germination_Temp	1	424695	424695	132.3	1.92e-09
Kilning_Temp	1	493813	493813	153.8	6.066e-10
Residuals	17	54577	3210	NA	NA

Again, we can see that the independent variables for Steeping Time, Germination Temperature, and Kilning Temperature are significant for this model.

Because the optimal value for Dymethyl Sulphide is a target, we fit the data using both the original value and a transformed value based on the distance from the target value (note that this is a minimization, not a maximization).

Table 3: Anova Table for reduced model for Dymethyl Sulphide

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Steeping_Time	1	980.5	980.5	28.99	4.946e-05
Germination_Temp	1	7613	7613	225.1	3.088e-11
IDK_Level	1	77.34	77.34	2.286	0.1489
Residuals	17	575	33.82	NA	NA

Table 4: Anova Table for reduced model for absolute value of Dymethyl Sulphide

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Steeping_Time	1	446.7	446.7	19.87	0.0003042
Germination_Temp	1	5940	5940	264.2	3.334e-12
Residuals	18	404.7	22.48	NA	NA

The modification of the dependent variable did not have a significant effect on model selection. From this we can see that Steeping Time and Germination Temperature are clearly significant. Because it is a main effect, and has a relatively high p-value, we suspect that there may be a higher order effect involving IDK Level, so the we kept that in for a future design analysis, realizing that the inclusion of this variable could result in a more significant run budget.

Stage 2: Model Fitting

We fit a model using a full CCD on the independent variables of Steeping Time, Germination Temperature, and Kilning Temperature, as discussed in the preceding section and in accordance with the methodological approach presented in the Methodology Section. A full second order model was fit for each of the dependent variables, and the resulting model was reduced through stepwise selection and p-value screening.

Model for Amalyse

The resulting model, after reduction, is given below.

Table 5: Anova Table For Amylase Level

	Estimate	Std. Error	t value	Pr(>|t|)
Steep_Time	37.9	21.1	1.8	0.097
Germ_Temp	-121	21.1	-5.75	9.13e-05
Kiln_Temp	138	21.1	6.55	2.72e-05
I(Germ_Temp^2)	-90	31.6	-2.85	0.0147
Steep_Time:Germ_Temp	45.3	23.5	1.93	0.0781
(Intercept)	975	23.5	41.4	2.54e-14

And the model equation is given by:

$$y_1 = 975.036 + 37.909x_2 + -121.124x_3 + 137.996x_8 + -89.978x_2^2 + 45.341x_3^2$$

Model for Superoxide Dismutase

The resulting model, after reduction, is given below.

Table 6: Anova Table For SOD Level

	Estimate	Std. Error	t value	Pr(>|t|)
Steep_Time	-51.9	17.6	-2.94	0.0134
Germ_Temp	114	17.6	6.47	4.62e-05

	Estimate	Std. Error	t value	Pr(>|t|)
Kiln_Temp	-146	17.6	-8.3	4.59e-06
I(Germ_Temp^2)	50.9	26.5	1.92	0.0807
Steep_Time:Germ_Temp	-60.9	19.7	-3.09	0.0104
Steep_Time:Kiln_Temp	-36.1	19.7	-1.83	0.0945
(Intercept)	1029	19.7	52.1	1.59e-14

And the model equation is given by:

$$y_1 = 1028.691 + -51.893x_2 + 114.178x_3 + -146.493x_8 + 50.915x_3^2 + -60.901x_2x_3 + -36.101x_2x_8$$

Model for Dymethyl Sulphide

The resulting model, after reduction, is given below.

Table 7: Anova Table For DSD Distance from 65

	Estimate	Std. Error	t value	Pr(>|t|)
Steep_Time	-1.02	1.29	-0.798	0.438
Germ_Temp	-8.35	1.29	-6.5	1.41e-05
I(Steep_Time^2)	4.81	1.93	2.5	0.0256
(Intercept)	9.99	1.44	6.95	6.75e-06

And the model equation is given by:

$$y_1 = 9.989 + -1.025x_2 + -8.35x_3 + 4.814x_2^2$$

Model Results

The figures provided below represents a contour plot of each of the response variables.

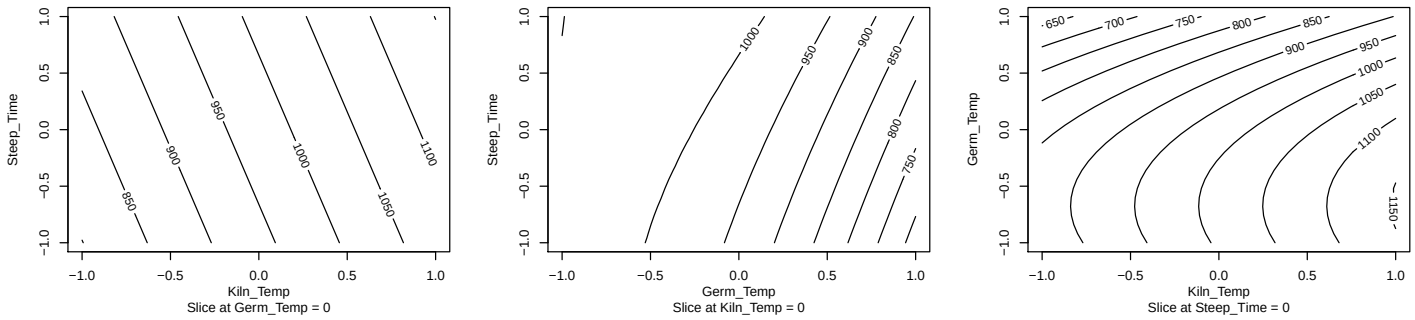

Figure 1: Contour Plots for Amylase

As we can see by these plots, there is a complex surface represented by the contours. Amylase level is likely maximized within the surface at a large Kilning Temperature, large Steeping Time, and a small Germination Temperature. SOD is likely maximized at low Kilning Temperature, Steeping Time, and Germination Temperature. The projected distance for DSD from 65 is maximized at a high value of Germination Temperature, and a low value of Steeping Time.

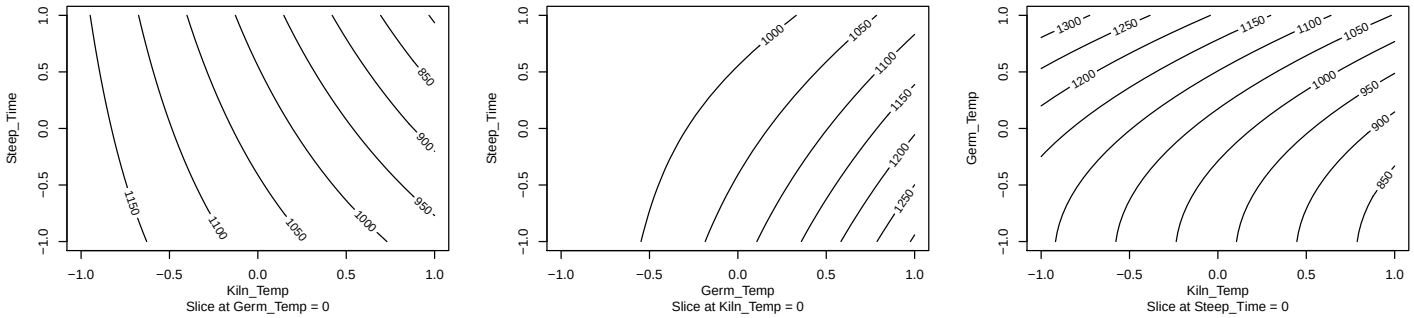

Figure 2: Contour Plots for SOD

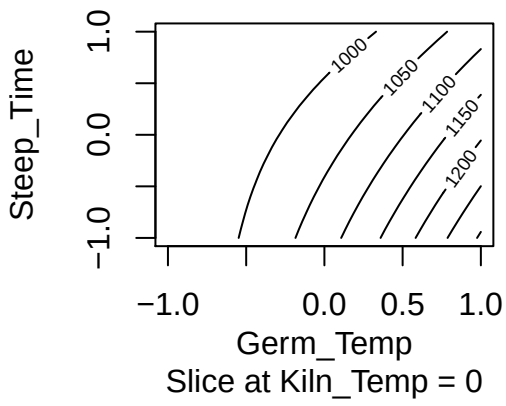

Figure 3: Contour Plots for DSD

Stage 3: Analysis

This section will provide the necessary steps to develop optimal operating conditions for the brewing process.

Pareto Front

To begin, we accept the equations from the previous section as truth and create a grid of points on which to evaluate those expressions. For this grid, we chose to evaluate each term from -1 to 1 using a seperator of .1. This created 21 points over each independent variable, creating a grid of 9261 points. Evaluating this dataset in each of the functions gives the objective values and using outer products (we can actually get the pareto rank, which provides more information, but we digress), we were able to generate the pareto front, containing 1621 points. The plots below show the 3D scatterplot and each 2D projection of the scatterplots with points on the Pareto Front in blue.

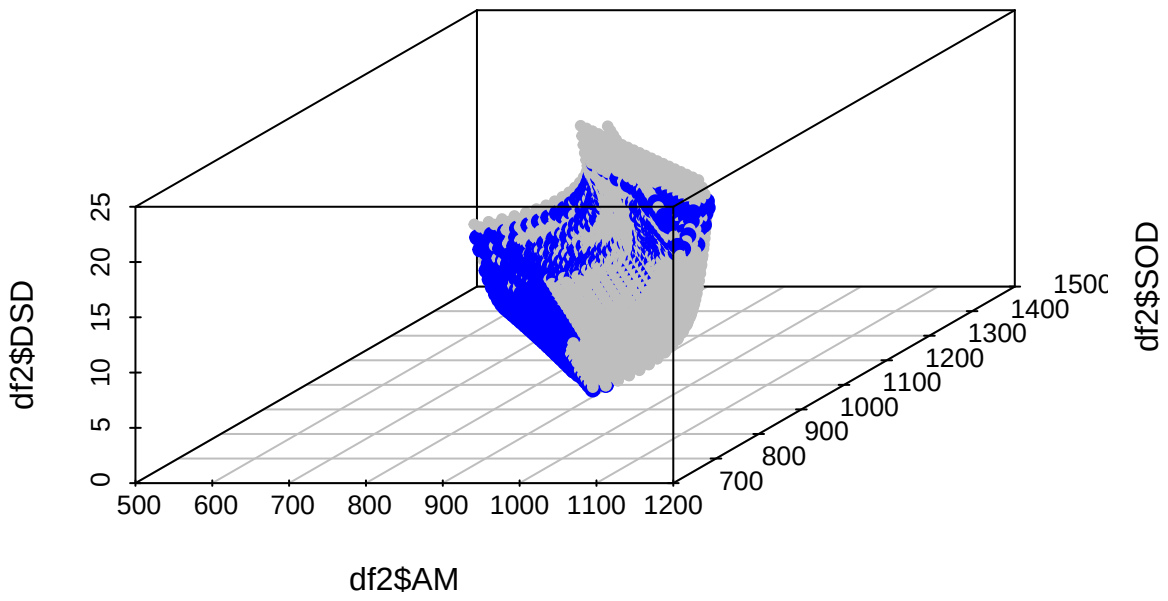

Figure 4: 3D Scatter Plot of Pareto Front, blue points are on the Pareto Front

Mixture Plots

Mixture plots for additive and multiplicative functions are provided below, with the axes represnting the weights on each of the response variables (scaled according to the Pareto front).

However, this doesn't give as much information as one would like. A slightly better way to narrow down points for consideration is to examine each location by the percentage of area covered in the mixture plot.

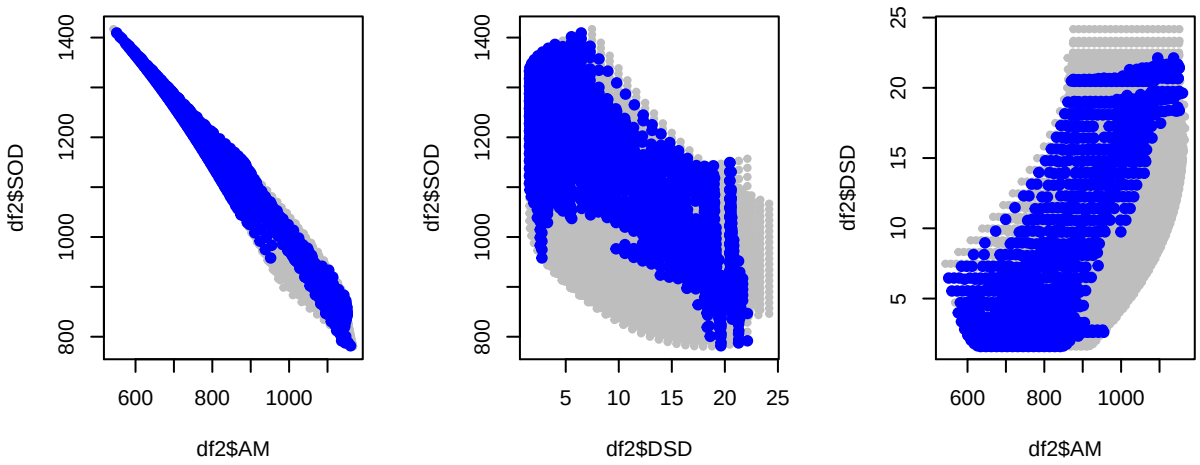

Figure 5: Projections of pareto front into 2D, blue points are on the pareto front

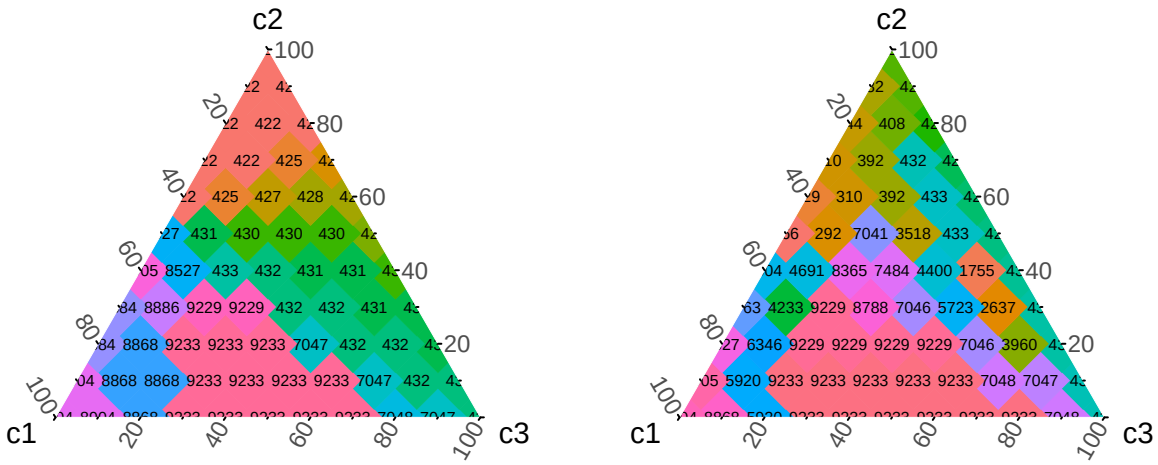

Figure 6: Additive and Multiplicative Mixture Plots (respectively)

Table 8: 5 best addative scores

9233	422	432	431	430
0.182	0.136	0.121	0.0909	0.0606

Table 9: 5 best multiplicative scores

9233	9229	432	310	392
0.167	0.0758	0.0455	0.0303	0.0303

After calculating these points, we narrowed down the selection of points to those that have a relatively large share (above 10%) of the value on addative mixtures. Plots of additive and multiplicative contours are shown below for each of those points.

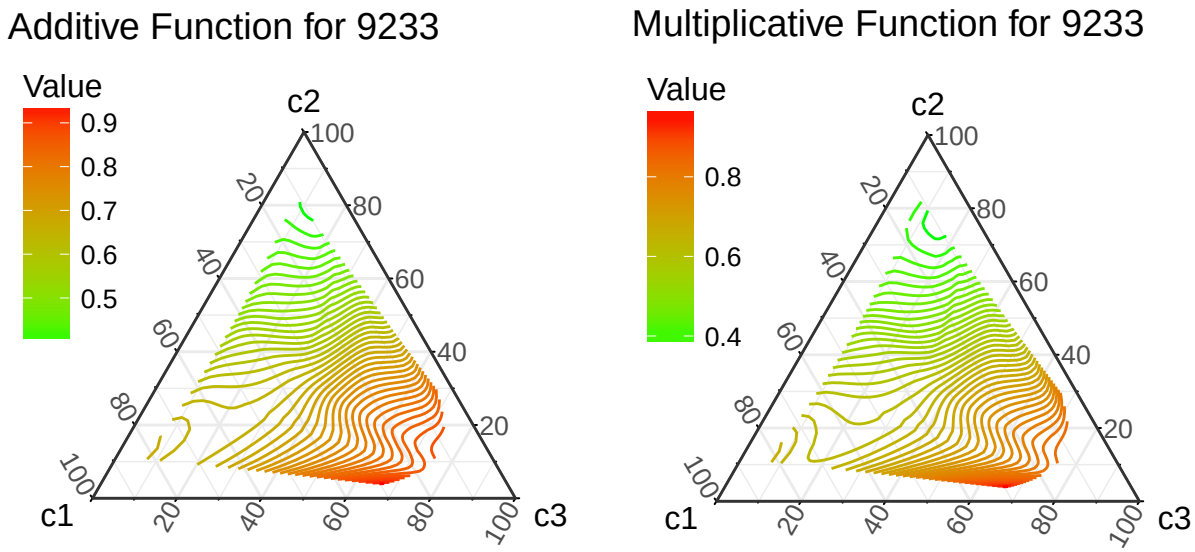

Figure 7: Point Desireablity

Of these points, point 9233 and 432 have the best consistency in value for both additive and multiplicative functions. Their values are given in tables below (in coded units for all variables):

Table 10: Table values for Point 9233

	Steep_Time	Germ_Temp	Kiln_Temp	AM	SOD	DSD
9233	0.3	0.9	1	0.663	0.321	0.951

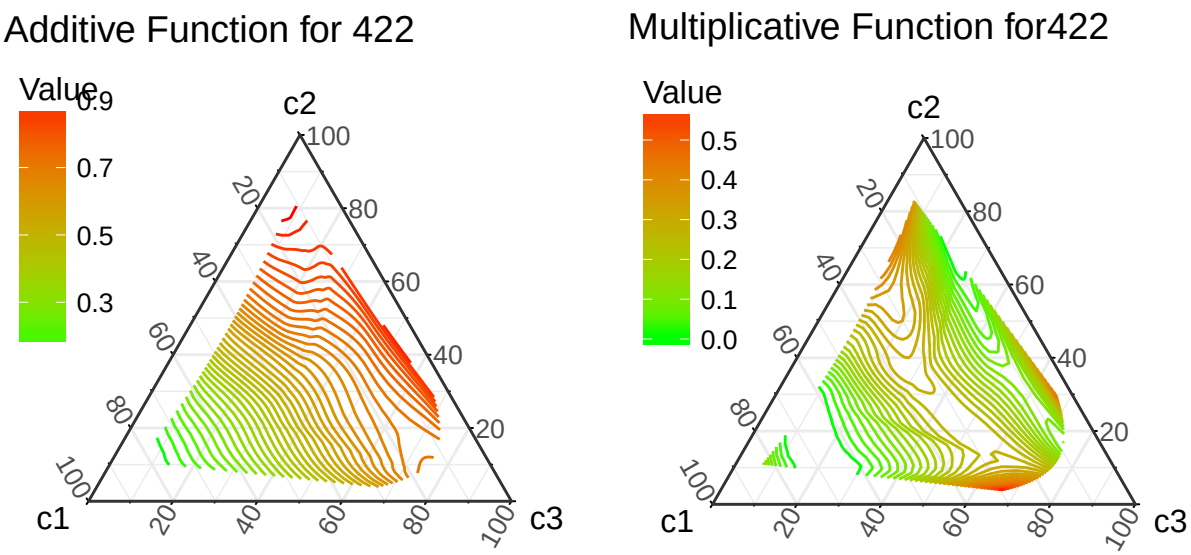

Figure 8: Point Desireablity

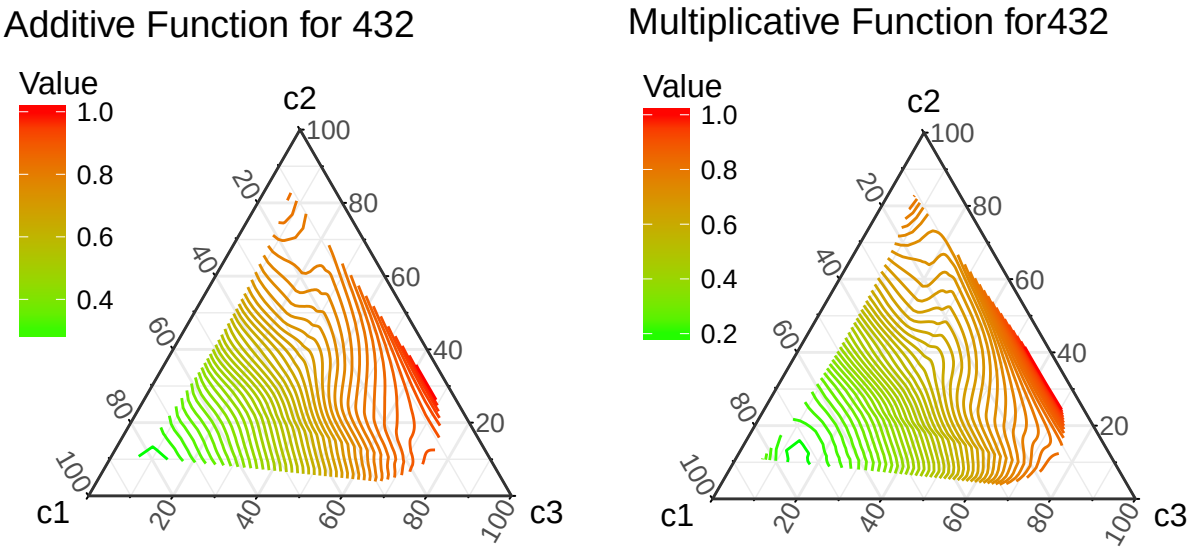

Figure 9: Point Desireablity

Table 11: Table values for Point 432

	Steep_Time	Germ_Temp	Kiln_Temp	AM	SOD	DSD
432	0.1	1	-1	0.137	0.878	1

Conclusions

Recommendation

Considering that the points perform equally well, the reccomendation is to operate at the values associated with point 9233. This point achieves similar results with less sacrifices in each of the Response Variables. This point corresponds to the values of Steeping Time 39.6 hours, Germination Temperature of 23.2°C, and a Kilning Temperature of 90°C.

Lessons Learned

Appendix: Model Data

For the purposes of presentation in this appendix, names of variables have been reduced to numeric operators based on order presented in the project instructions and an X indicating an independent variable or Y indicating a dependent variable.

Stage 1

X1	X2	X3	X4	X5	X6	X7	X8	X9	Y1	Y2	Y3
0	1	1	1	1	1	1	1	1	967.19	940.05	68.45
0	-1	-1	-1	-1	-1	-1	-1	-1	373.23	859.45	5.68
1	0	-1	-1	-1	-1	1	1	1	795.14	690.85	22.09
-1	0	1	1	1	1	-1	-1	-1	668.81	1294.93	56.01
1	-1	0	-1	1	1	-1	-1	1	742.36	1037.56	25.56
-1	1	0	1	-1	-1	1	1	-1	1182.40	799.90	39.61
1	-1	-1	0	1	1	1	1	-1	763.89	518.29	12.99
-1	1	1	0	-1	-1	-1	-1	1	650.79	1231.00	73.70
1	-1	1	1	0	-1	-1	1	-1	888.23	926.97	43.23
-1	1	-1	-1	0	1	1	-1	1	485.07	1018.77	19.63
1	-1	1	1	-1	0	1	-1	1	568.87	1286.22	42.79
-1	1	-1	-1	1	0	-1	1	-1	861.48	728.33	18.81
1	1	-1	1	-1	1	0	-1	-1	487.69	1072.48	19.09
-1	-1	1	-1	1	-1	0	1	1	891.50	912.64	57.89
1	1	-1	1	1	-1	-1	0	1	656.62	823.36	23.50
-1	-1	1	-1	-1	1	1	0	-1	694.17	1077.27	45.85
1	1	1	-1	-1	1	-1	1	0	964.81	899.47	69.85
-1	-1	-1	1	1	-1	1	-1	0	385.08	844.81	20.83
1	1	1	-1	1	-1	1	-1	-1	677.20	1258.24	69.59
-1	-1	-1	1	-1	1	-1	1	1	711.23	505.58	14.56
0	0	0	0	0	0	0	0	0	993.97	958.31	40.16

Stage 2

X2	X3	X8	Y1	Y2	Y3	Target.65
-1	-1	-1	839.33	1129.40	38.36	26.64
-1	-1	1	1157.43	787.34	41.13	23.87
-1	1	-1	613.86	1342.93	61.57	3.43
-1	1	1	668.81	1294.93	56.01	8.99
1	-1	-1	828.76	1123.48	45.93	19.07
1	-1	1	1182.40	799.90	39.61	25.39
1	1	-1	669.01	1256.29	72.84	7.84
1	1	1	990.79	901.00	73.70	8.70
-1	0	0	974.11	1042.65	48.79	16.21
1	0	0	961.67	997.65	57.11	7.89
0	-1	0	1022.99	987.02	43.58	21.42
0	1	0	877.20	1173.77	61.07	3.93
0	0	-1	821.29	1253.68	50.16	14.84
0	0	1	1152.78	857.68	69.86	4.86

X2	X3	X8	Y1	Y2	Y3	Target.65
0	0	0	990.22	1004.56	54.58	10.42
0	0	0	955.45	1022.31	57.82	7.18
0	0	0	969.69	1013.83	50.95	14.05
0	0	0	975.08	1037.17	61.79	3.21