Name: Natalya Lavrenchuk (ID: 2141882)
Course: SP25 Machine Learning for Process Engineering
Instructor: Professor Pierantonio Facco
Assignment: Homework #2 – DoE
Date: June 19, 2025

UNIVERSITÀ DEGLI STUDI DI PADOVA

The purpose of this assignment is to apply principles in DOE, linear regression, and process optimization to build an empirical model for a multi-response industrial pigment extraction process. All questions were answered with the use of softwares MATLAB and MiniTab.

<u>Question 1:</u> What type of design was planned and carried out in PigmentOne?

The PigmentOne experiment is a three factor, three level Box-Behnken design with a total of 17 runs. Each factor, temperature (30/40/50 °C), time (20/70/120 min), and mass (0.5/1.0/1.5 g), was coded at levels −1, 0, +1. In this design, the 12 non-center runs lie at the midpoints of the edges of the factor cube, and there are five replicates at the center point (0,0,0) that help to estimate pure error or check for curvature. The absence of any corner runs and the fact that each run has exactly one factor at the intermediate level is characteristic of a Box-Behnken design. Unlike a full 33 factorial design, which would require 27 runs, this study reduces the number of required experiments yet still allows for efficient estimate of the first- and second-order terms.

A

	Factor		
Run	A (Temp.)	B (Time)	C (Mass)
1	-	-	0
2	-	0	+
3	-	+	0
4	+	+	0
5	0	0	0
6	0	0	0
7	0	+	+
8	+	0	+
9	-	0	-
10	0	0	0
11	+	-	0
12	0	-	+
13	+	0	-
14	0	-	-
15	0	0	0
16	0	+	-
17	0	0	0

B

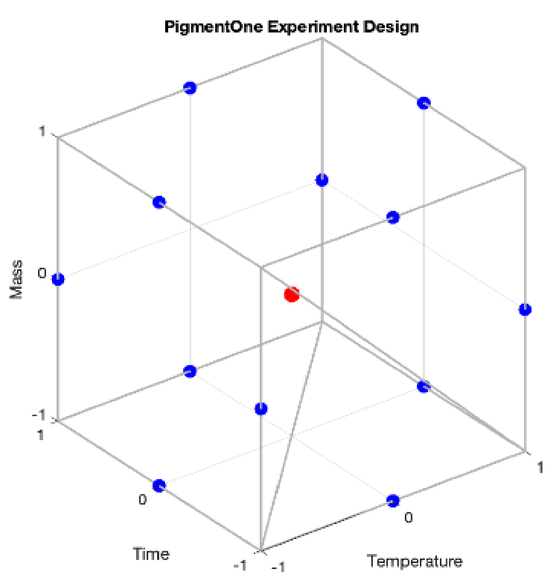


Figure 1. (A) Coded design matrix for three-factor Box-Behnken design (BBD) experiment (temperature, time, and mass) showing all 17 runs. Runs 5, 6, 10, 15, 17 (shaded rows) are the five center-point replicates at (0,0,0). The remaining 12 runs correspond to the midpoint of each face of the $[-1,+1]^3$ cube. (B) 3D view of the coded design space where 12 runs, in blue, lie on the face center position (±1,±1,0), (±1,0, ±1), and (±1, ±1,0), and the red point marks the center point replicates at (0,0,0).

Question 2: **Build the main effect plot and comment on it.**

The main effects plots for the PigmentOne experiment were made in MATLAB and can be visualized in Figure 2A for betacyanin, and Figure 2B for betaxanthin. Similar trends were observed across the three factors on the response of both pigment extraction yields. Clear trends demonstrate that higher recovery of these pigments was achieved with increased time of extraction and increased mass of pear fruit. Generally, these two factors have a positive linear impact on the pigment recoveries. Temperature, on the other hand, had less significant impact to the pigment extraction process, particularly to that of betacyanin, however a peak in recovery is observed at the intermediate temperature of 40 degrees Celsius for betaxanthin. Temperature, unlike the other two variables, seems to have a quadratic impact on the response.

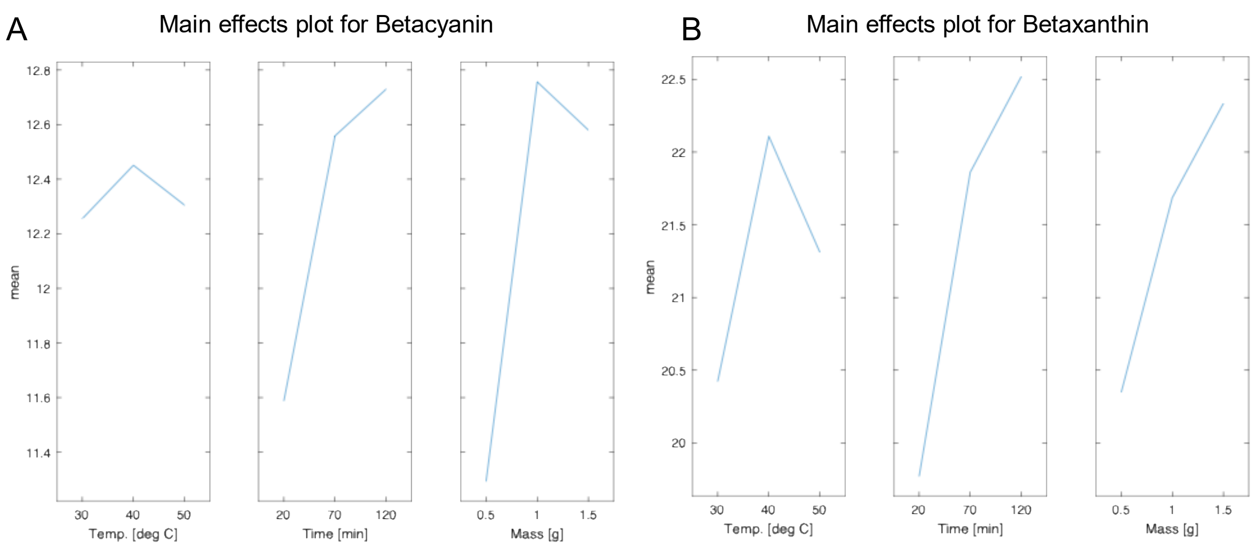


Figure 2. Main effects plots illustrating the impact of the experimental factors (Temperature, Time, and Mass) on the mean recovery of pigments (A) Betacyanin, and (B) Betaxanthin.

Question 3: **Build the interaction plot and comment on it**

The interaction effects between the three factors of the experiment were plotted in Figure 3A for betacyanin and Figure 3B for betaxanthin. Interaction plots with non-parallel lines indicate that the experimental factors interact and together have an impact on the response. For betacyanin, a clear interaction effect is seen in the Time x Mass panel, as the longer the time of extraction, the more pronounced the impact higher mass of pear fruit has on the pigment recovery. On the other hand, Temperature x Mass have nearly parallel lines meaning that changing the mass of the pear fruit has the same effect regardless of the temperature of the extraction process, therefore these factors have minimal interaction effects. Temperature x Time, longer extraction times have a somewhat greater benefit at higher temperatures. For betaxanthin, similar significant interaction effects are observed with Time x Mass, with longer times, mass of pear fruit has a positive impact on betaxanthin recovery. Temp. x Mass plot indicates that there is an interaction. Lastly, Temp. x Time are nearly parallel showing minimal interaction effects.

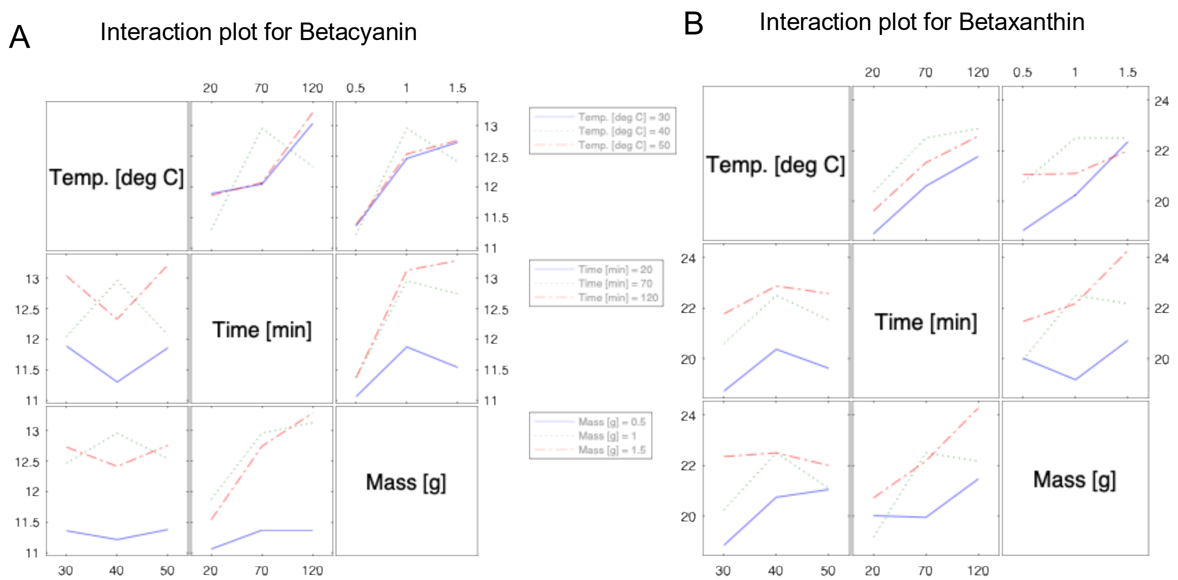


Figure 3. Interaction plots illustrating the impact of the crossed experimental factors Temperature x Time, Temperature x Mass, and Time x Mass on the recovery of pigments (A) Betacyanin, and (B) Betaxanthin.

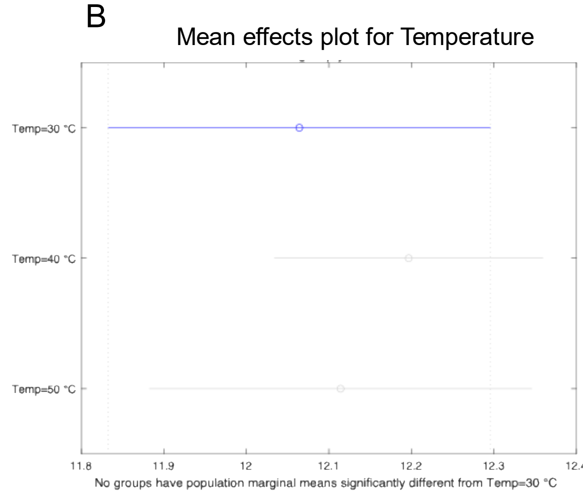
<u>Question 4a:</u> Build a response surface model. estimate the regression model and comment on its structure, considering the parameters uncertainty.

For each pigment, an estimated linear regression model with polynomial of degree two, interaction effects, and main effects were built and n-way ANOVA were performed. For Betacyanin, the results of the ANOVA test, as seen in Figure 4A, indicate that Time ($p < 1E-07$) and Mass ($p < 1E-08$) have highly significant effects on the pigment yield, Temperature proved to be statistically insignificant ($p=0.391$), Time x Mass is the only significant interaction term ($p=3.24E-05$), and all quadratic effects are significant based on a significance level of 5%. Tukey HSD plots (Figure 4B–D) shows that both 70 min and 120 min extractions, as well as both 1 g and 1.5 g masses, yield significantly more pigment than their low-level counterparts (20 min and 0.5g) with no difference between the intermediate and high levels. The ANOVA and multiple-comparison results are consistent with the earlier main-effects and interaction analyses performed.

A

ANOVA for Betacyanin					
	SumSq	DF	MeanSq	F	p-value
Temp	0.005	1	0.005	0.84	0.391
Time	2.61	1	2.61	436.4	1.45e-07
Mass	3.32	1	3.32	554.2	6.34e-08
Temp x Time	0.011	1	0.011	1.843	0.21674
Temp x Mass	2.5e-05	1	2.5e-05	4.18e-03	0.95026
Time x Mass	0.52	1	0.52	86.7	3.42e-05
$Temp^2$	0.049	1	0.049	8.14	0.0246
$Time^2$	0.52	1	0.52	86.2	3.48e-05
$Mass^2$	2.66	1	2.66	444.85	1.36e-07
Error	0.042	7	0.042		

B



C

D

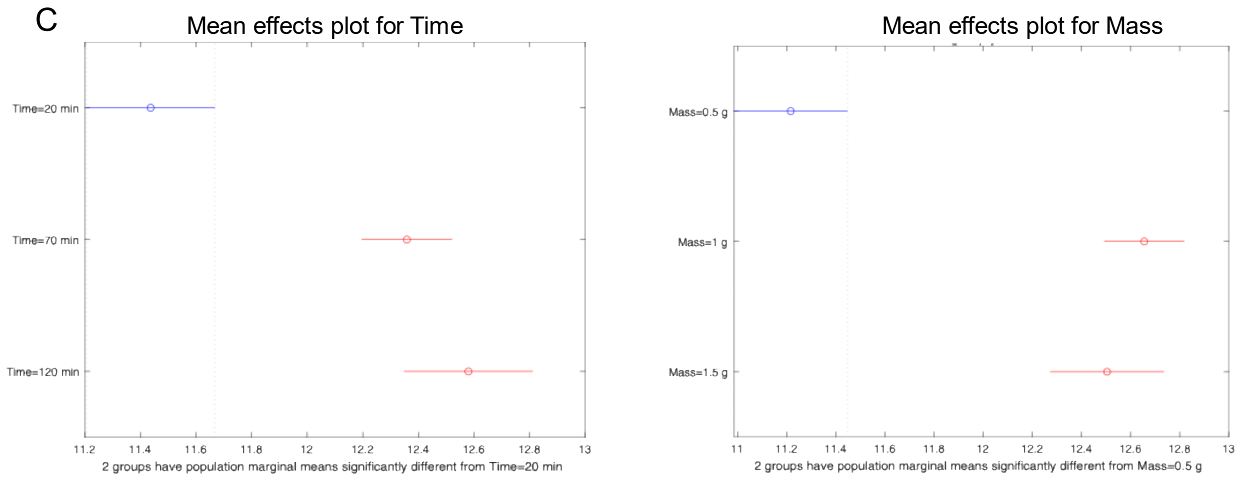


Figure 4. ANOVA and multiple comparison of means plots for betacyanin extraction. (A) ANOVA table showing sum of squares (SumSq), degree of freedom (DF), F-statistics and p-values for each main effect, interaction, and quadratic term in the full response-surface model. (B-D) Tukey's HSD multiple-comparison plots for (B) Temperature, (C) Time, and (D) Mass with 95% confidence interval marginal means.

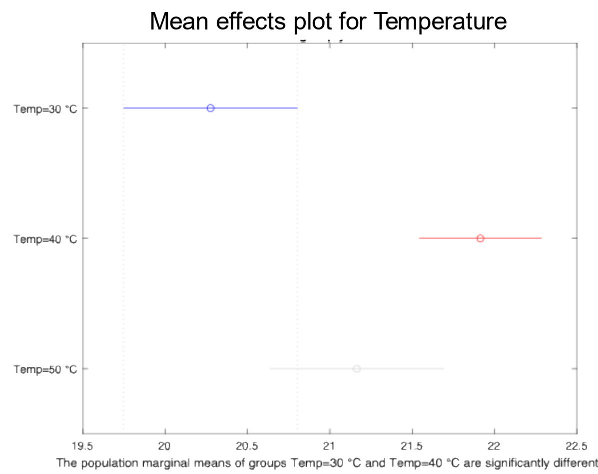
For Betaxanthin, the full second-order model showed strong main-effect significance for Temperature ($p = 2.6E-04$), Time ($p < 1E-07$) and Mass ($p < 1E-06$), as well as significant Temperature x Mass, Time x Mass, and all quadratic terms; only the Temp x Time interaction was non-significant ($p = 0.80$) (Figure 5A). Tukey HSD plots (Figure 5B–D) reveal that both 30 degrees Celsius vs. 40 degrees C and 30 degrees C vs. 50 degrees Celsius differ significantly for betaxanthin yield. Likewise for Time and Mass, similar to that seen in the Betacyanin model, 70 min and 120 min extractions each outperform the low level 20 min, and 1 g and 1.5 g masses each outperform 0.5 g, with no differences between the intermediate and high levels.

A

ANOVA for Betaxanthin

	SumSq	DF	MeanSq	F	p-value
Temp	1.58	1	1.58	45.8	2.6e-04
Time	15.1	1	15.1	440	1.4e-07
Mass	7.90	1	7.90	230	1.3e-06
Temp x Time	0.0025	1	0.0025	0.072	0.80
Temp x Mass	1.63	1	1.63	47.3	2.4e-04
Time x Mass	1.10	1	1.10	32.1	7.6e-04
$Temp^2$	6.00	1	6.00	175	3.3e-06
$Time^2$	1.70	1	1.70	49.6	2.0e-04
$Mass^2$	0.25	1	0.25	7.28	0.031
Error	0.24	7	0.034		

B



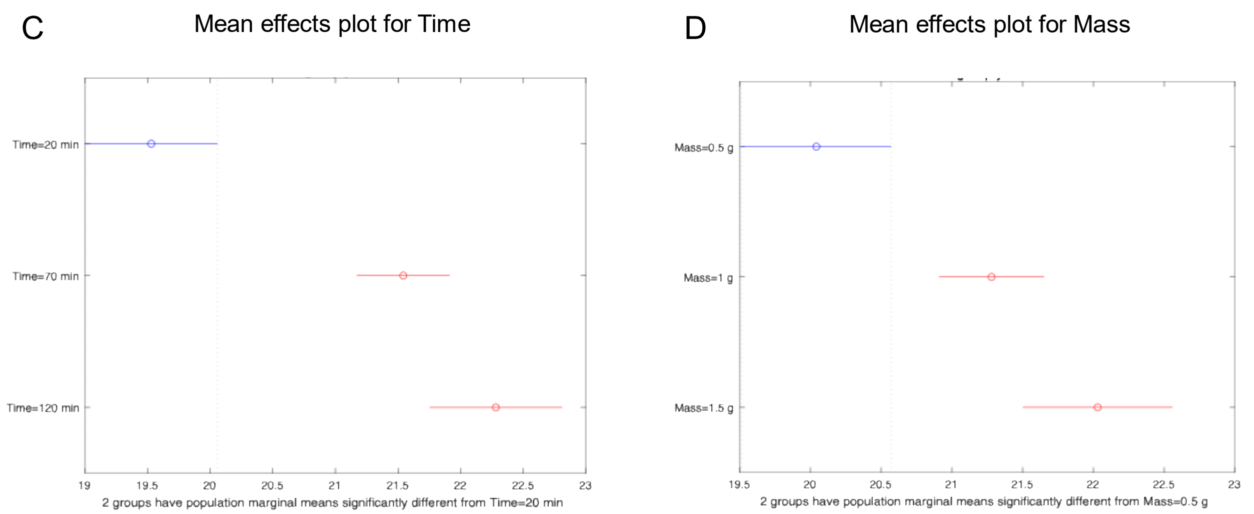


Figure 5. ANOVA and multiple comparison of means plots for betaxanthin extraction. (A) ANOVA table showing sum of squares (SumSq), degree of freedom (DF), F-statistics and p-values for each main effect, interaction, and quadratic term in the full response-surface model. (B-D) Tukey's HSD multiple-comparison plots for (B) Temperature, (C) Time, and (D) Mass with 95% confidence interval marginal means.

For both pigments, the estimated coefficients (Figure 6) show that Time and Mass has the largest positive impact on the response, confirming that longer extraction times and higher fruit mass strongly increase pigment yield. Temperature's linear effect is negligible for Betacyanin since its

95% confidence interval crosses the zero line, but significant for Betaxanthin. Among interactions, only the Time x Mass term is reliably positive in both models, while the Temperature x Time term is essentially zero. Lastly, all three negative quadratic coefficients indicate that each factor exhibits diminishing benefits, meaning the improvement in yield is boosted from low to intermediate levels but not as much improvement in recovery is seen between the intermediate and high levels. The red bars indicating the 95% confidence interval for each regression coefficient indicates the certainty the model has on the effects impact to the response. A narrow red bar, like in the case of Mass and Time indicate higher accuracy in the effect estimators whereas a wide bar indicates higher uncertainty in the strength of those effects.

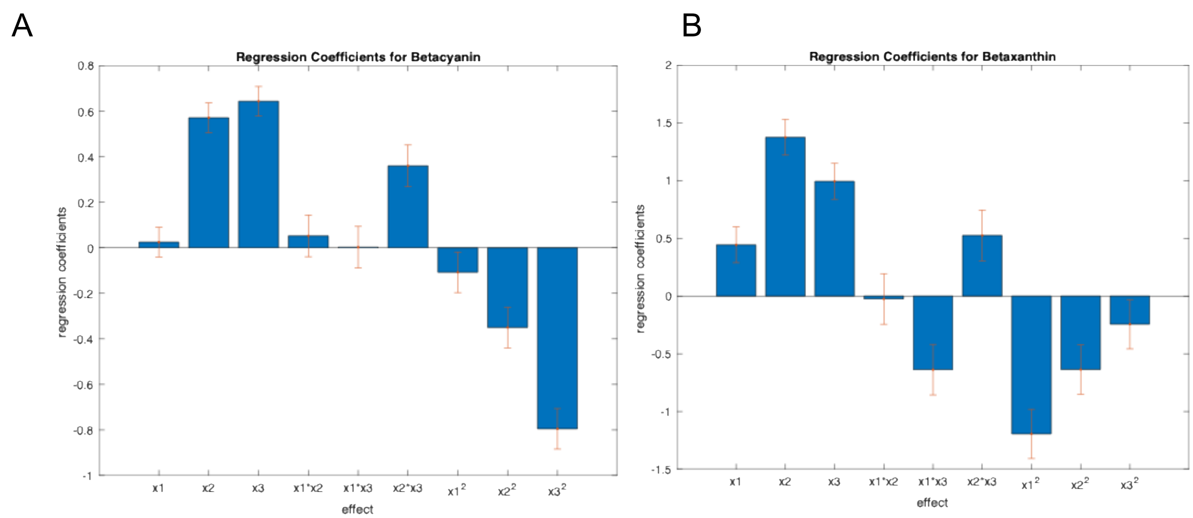


Figure 6. Estimated regression coefficients for the full response-surface models: (A) betacyanin and (B) betaxanthin. Each bar corresponds the magnitude of the model term-main effects(x_1, x_2, x_3), two-way interactions (x_1x_2, x_2,x_3) and quadratic terms (x_1^2, x_2^2, x_3^2) with the red error bar indicating the ±95% confidence interval.

<u>Question 4b:</u> discuss the outcome (in terms of regression parameters) considering what were the results of both the main effect plot and the interaction plot.

The estimated regression coefficients for both pigment models match what is predicted by the main effects and interaction effect plots discussed in Question 2 and 3, respectively. The regression coefficients that have the largest magnitude and impact to the model, particularly Time and Mass, also exhibited high slopes in the main effect plots, demonstrating the importance of the factors to pigment recovery. On the other hand, factors that have minimal coefficients, particularly those whose 95% confidence intervals cross zero, like Temperature in the Betacyanin model, have no significant impact to the model, which is reflected by a nearly flat line across all levels in the main effects plot. In terms of interaction effects, similar trends could be seen. Those with large regression coefficients, like Time x Mass in both models, are reflected by a pronounced fan-out of the Time x Mass lines in the interaction plots, confirming that longer extraction times enhance the benefits seen with greater masses of pear fruit. In contrast, the nearly parallel lines in the Temp × Time interaction plots, seen in both models, and Temp × Mass panels, in the Betacyanin model, mirror the near-zero β_{12} and β_{13} estimated coefficients.

<u>Question 4c:</u> Comment on model fitting and adequacy

The full model fit has very high accuracy, with R^2 values of 0.9958 for the Betacyanin model and 0.9933 for the Betaxanthin model. This indicates that the models do an extremely good job at explaining most of the variation and predicting the response of pigment extraction recoveries from the factors in the set ranges being analyzed. To further check model adequacy, investigation into the residuals for both models is performed, particularly looking into the residual normality, homoscedasticity, independence, and linearity. Checking these properties in the residuals is crucial because the assumptions underlie the validity of hypothesis tests, confidence interval and prediction intervals in linear regression and ANOVA. Plots investigating these residual properties for Betacyanin are shown in Figure 7. The Q-Q plot, Figure 7A, shows that generally points fall on the 45-degree line, supporting the residual normality assumption. Figure 7B shows the residuals scattered randomly around zero with no obvious trend against the fitted values, indicating constant variance in residuals, or homoscedasticity, and confirms that the relationship is adequately captured by a linear model. Independence of residuals is demonstrated by the randomness in the scatter of the points when plotting across sequential experiments (Figure 7C). Lastly, when analyzing the residuals across every level of each factor, Figure 7D-F, we see random scatter around zero with similar vertical spread and no obvious patterns, confirming that the residual variance is constant across factor settings and the model captured the main relationships without leaving out any obvious patterns in the data. Investigation into the residuals of the Betaxanthin model, Figure 8, yields similar results, concluding that the residuals for both models fit the assumptions of normality, independence, and homoscedasticity. The high R^2 values, together with the residual analysis, confirm that the models fit the data well and meets all the usual assumptions.

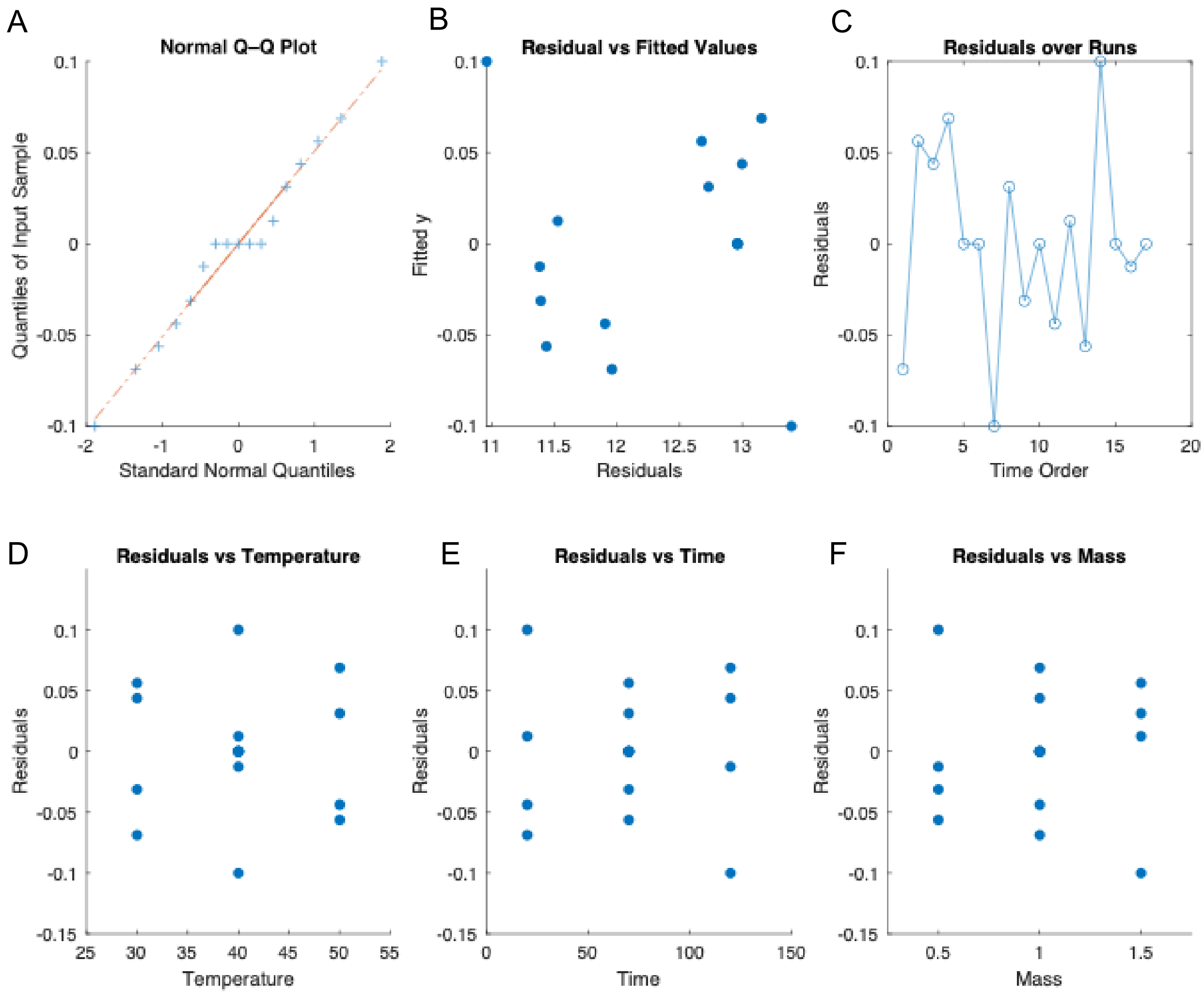


Figure 7. Diagnostic plots for Betacyanin full response-surface model, showing (A) Normal Q-Q Plot of residuals, (B) residuals vs. fitted values, (C) residuals by run order, and (D-F) residuals at each level of Temperature, Time, and Mass, showing no major departures from normality, homoscedasticity, independence, or linearity.

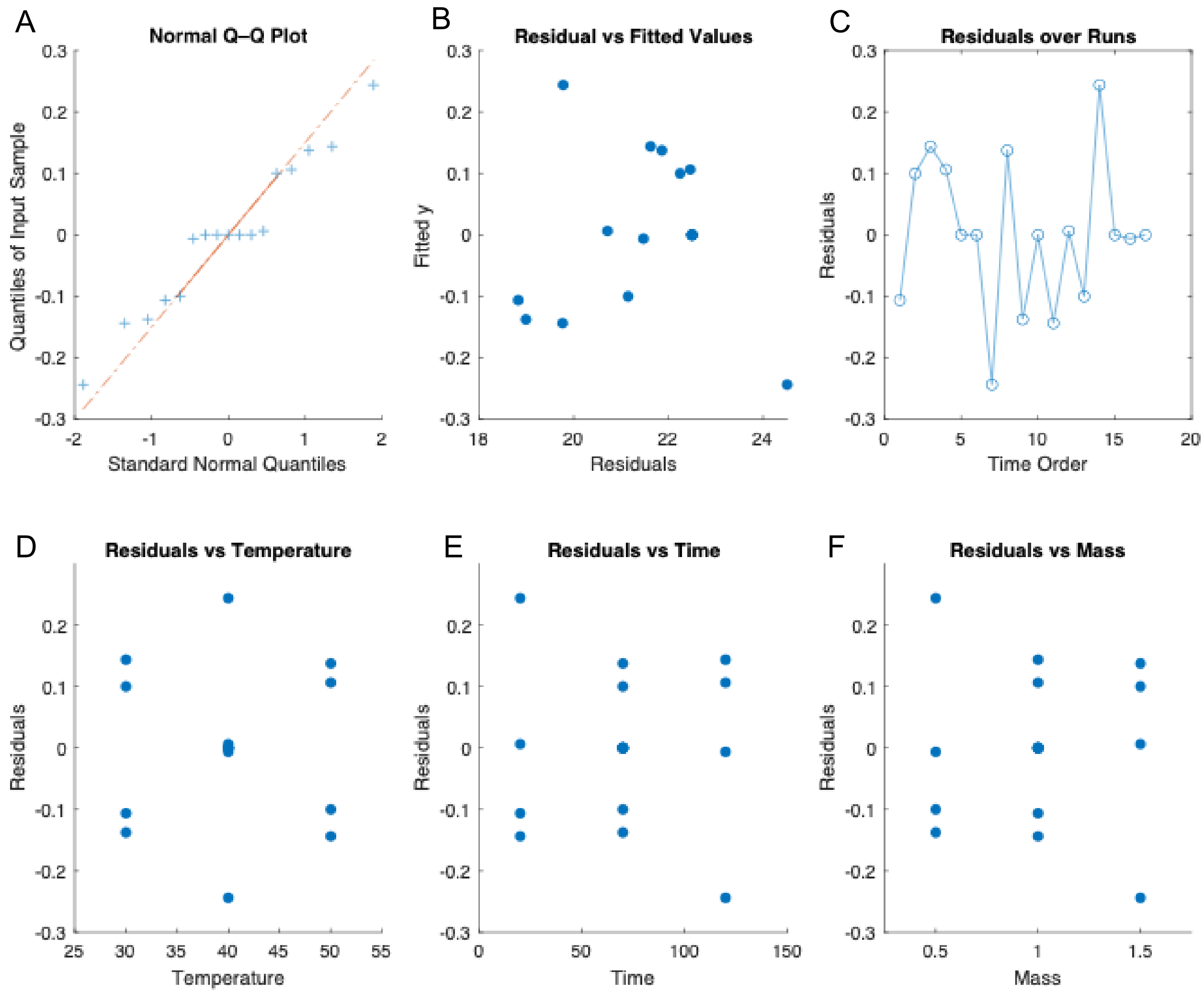


Figure 8. Diagnostic plots for Betaxanthin full response-surface model, showing (A) Normal Q-Q Plot of residuals, (B) residuals vs. fitted values, (C) residuals by run order, and (D-F) residuals at each level of Temperature, Time, and Mass, showing no major departures from normality, homoscedasticity, independence, or linearity.

<u>Question 4d:</u> Refine and update the model structure, if needed, and determine the updated values of regression coefficients, uncertainty of the coefficients, R2, and residuals

All statistically insignificant terms, whose p-value is above the 5% statistical confidence level, were dropped from the models. This included three terms, Temperature, Temperature x Time, and Temperature x Mass, for the Betacyanin model, and just one term, Temperature x Time, in the Betaxanthin model. The models were then re-run without these terms resulting in the final reduced response surface with all statistically significant coefficients displayed in Figure 9.

Written in equation form, the regression model for Betacyanin is shown in Equation (1) and Betaxanthin is shown in Equation (2), where x_1 = Temperature, x_2 = Time, and x_3 = Mass.

$$y_{Betacyanin} = 12.96 + 0.57x_2 + 0.64x_3 + 0.36x_2x_3 - 0.11{x_1}^2 - 0.35{x_2}^2 - 0.80{x_3}^2 \quad \text{......................(1)}$$

$$y_{Betaxanthin} = 22.5 + 0.44x_1 + 1.38x_2 + 0.99x_3 - 0.64x_1x_3 + 0.53x_2x_3 - 1.19{x_1}^2 - 0.64{x_2}^2 - 0.24{x_3}^2 \quad \text{...(2)}$$

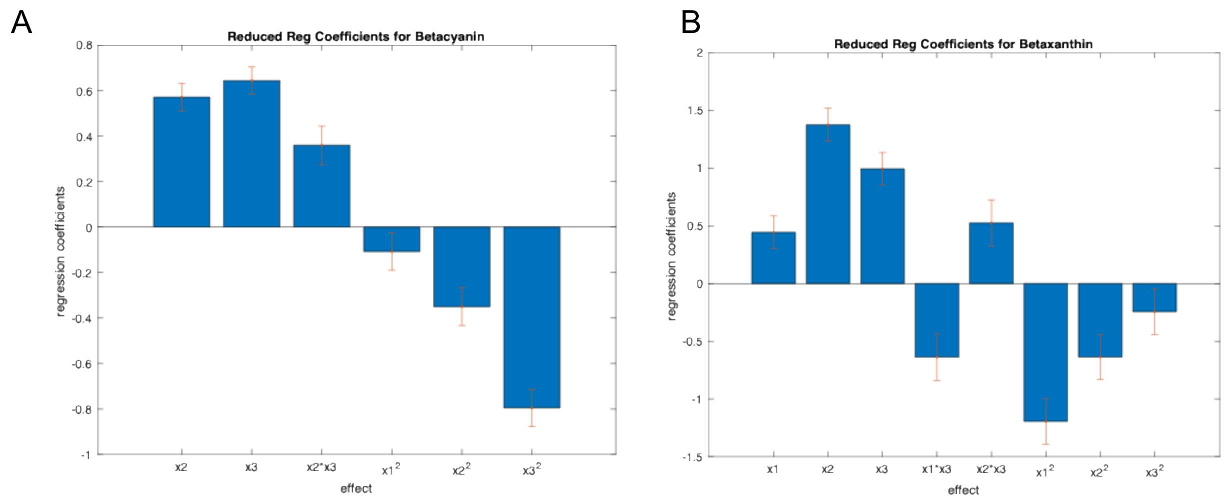


Figure 9. Estimated regression coefficients for the reduced response-surface models: (A) betacyanin and (B) betaxanthin. Each bar corresponds the magnitude of the model term-main effects(x_1, x_2, x_3), two-way interactions (x_1x_2, x_2,x_3) and quadratic terms (x_1^2, x_2^2, x_3^2) with the red error bar indicating the ±95% confidence interval.

The new models have R^2 values of 0.9942 for Betacyanin and 0.9933 for Betaxanthin, indicating no loss to model accuracy in the new reduced versions. Residuals for these models were reanalyzed in a similar way to that discussed in Question 4c and indicate that both models maintain residuals with normality, independence, and homoscedasticity (not shown here, see source code for plots). In combination, we can conclude that the new reduced model fits the data well and is used further in the assignment.

<u>Question 4e:</u> build the response surface plots adding the experimental points

The response surface plots overlayed with true experimental data were built with the reduced Betacyanin and Betaxanthin models (Equation 1 and 2, respectively). Contour plots for Betacyanin, Figure 10, show that increasing extraction time and mass of pear fruit consistently increases the yield of pigment. Temperature as a factor, as seen before, does not play a major role in this pigment recovery since no change in yield is observed across the three temperature levels. Contour plot of Time x Mass, Figure 10C, shows the synergetic interaction between the two factors, where both variables work together to boost the yield of pigment.

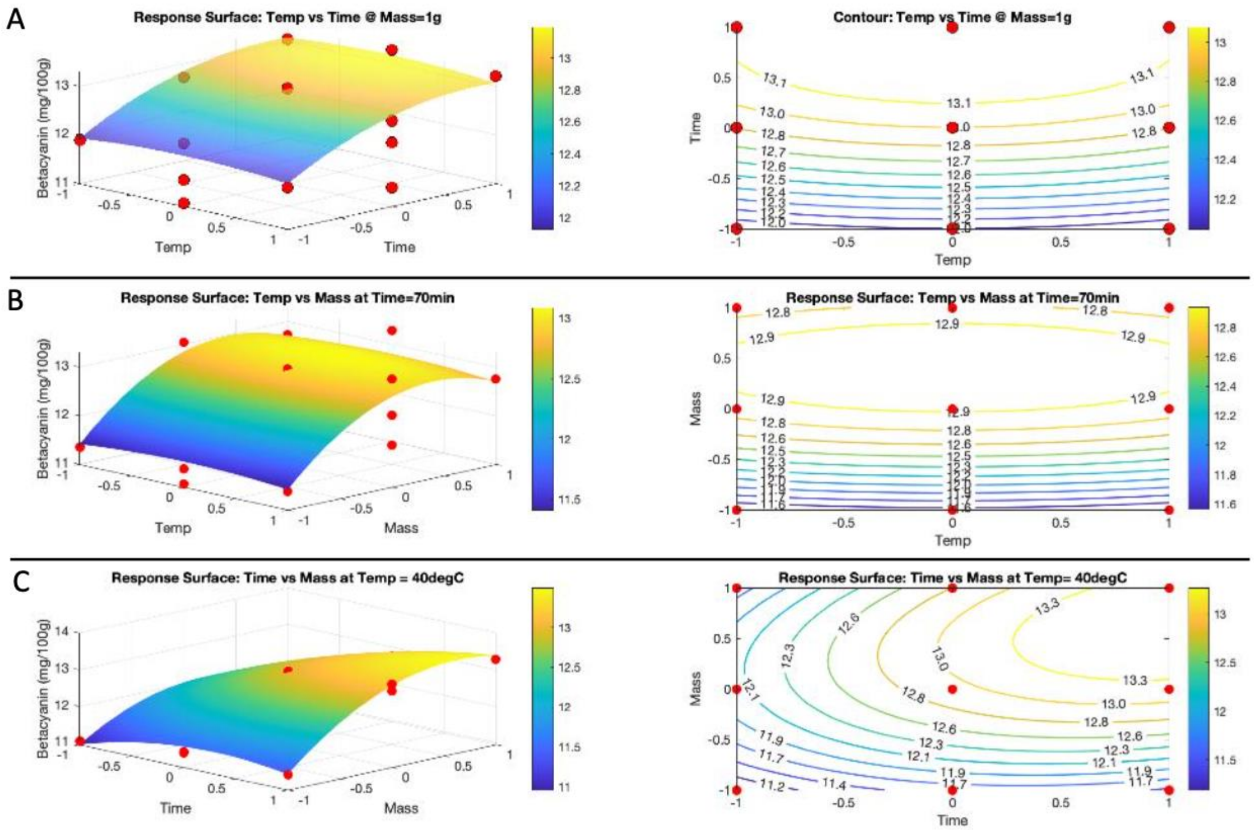


Figure 10. Contour maps of the fitted Betacyanin response surface, showing predicted yield (mg/100 g) as a function of: (A) extraction time vs. temperature at 1 g mass, (B) temperature vs. mass at 70 min, and (C) time vs. mass at 40 °C. Experimental data points are overlaid in red; the color bar on the right indicates the predicted yield.

The contour plots for Betaxanthin, Figure 11, show a slight difference to those of Betaxyanin, especially the impact temperature has on the pigment recovery. In Figure 11A, we see that Temperature and Time work together, to an extent, to increase recovery with intermediate to high temperatures boosting yield at longer extraction times. Similarly, we see intermediate temperature and high mass also cooperate to increase yield, as seen in Figure 11B. Also, at high temperatures, the mass of fruit does not seem to have an impact on the response. Lastly, similar to that of Betacyanin, Time and Mass have a strong interaction, with higher times and masses significantly boosting yield. We also see that time impacts the yield relatively linearly while mass exhibits a slight curvature, indicating mass of fruit slows down its benefits when going from low to high. Overall, the response-surface and contour analyses consistently highlight that longer extraction times and greater sample masses maximize both pigment yields while temperature plays a secondary but still impactful role.

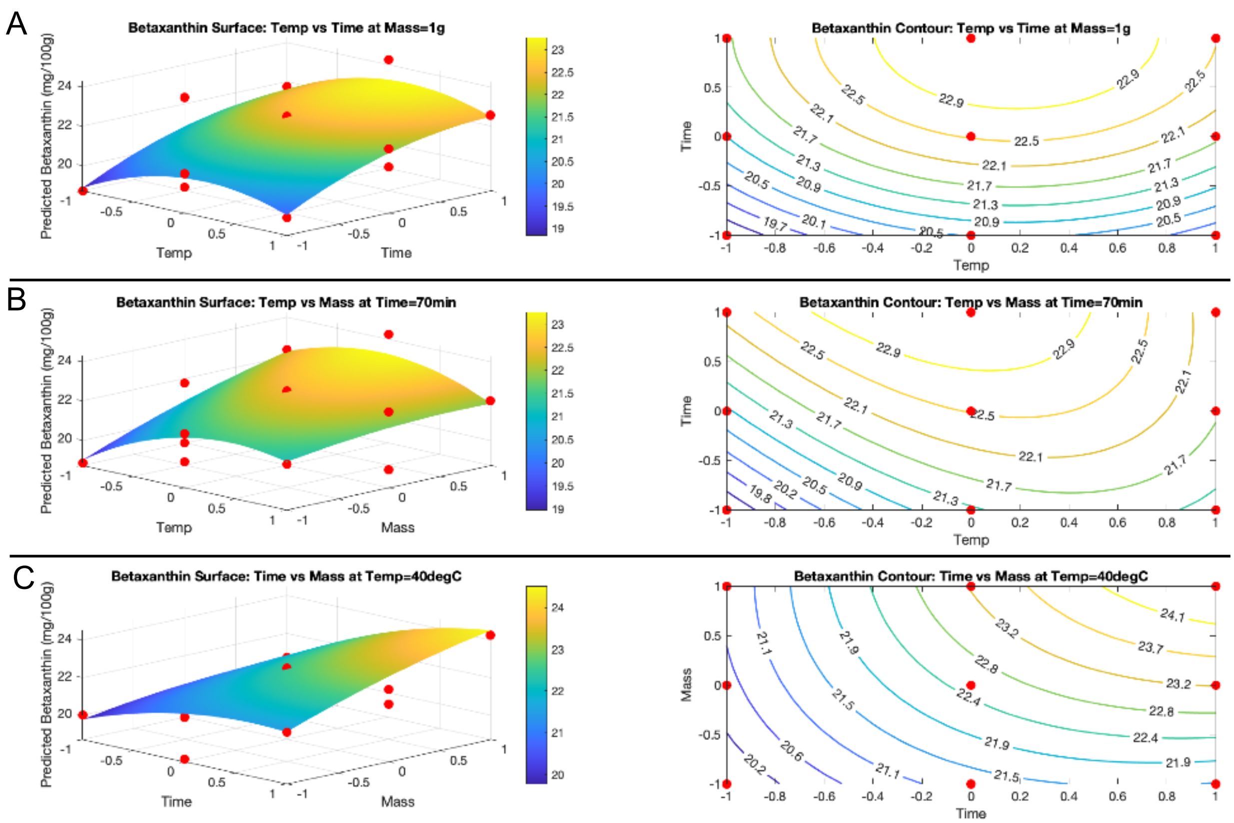


Figure 11. Contour maps of the fitted Betaxanthin response surface, showing predicted yield (mg/100 g) as a function of: (A) extraction time vs. temperature at 1 g mass, (B) temperature vs. mass at 70 min, and (C) time vs. mass at 40 °C. Experimental data points are overlaid in red; the color bar on the right indicates the predicted yield.

<u>Question 5:</u> Identify the optimal formulation which guarantees that the maximum extraction of betacyanin and betaxanthin

Optimization was performed on each pigment individually in MATLAB to find the factor levels that correspond the highest individual level of pigment yield. For Betacyanin, this corresponded to a coded factor level of 0 for Temperature, 1.0 for Time, and 0.613 for Mass, corresponding to the real values of 40 degrees Celsius, 120 min extraction time, and 1.32 g mass of fruit. This combination of factors predicts to obtain the maximum recovery in yield of 13.5 mg/100g of Betacyanin. Plots depicting the optimal conditions are shown in Figure 12A-C. For Betaxanthin, the factors predicted to maximize pigment yield are at coded factor level of -0.0812 for Temperature, 1.0 for Time, and 1.0 for Mass, corresponding to real values of 39.2 degrees Celsius, 120 min extraction time, and 1.5 g mass of fruit. This combination of factors predicts to obtain the maximum recovery in yield of 24.5 mg/100g of Betaxanthin. Plots depicting the optimal conditions are shown in Figure 12D-F. A single, formal multi-response optimum was not determined, but both surface models suggest that operating at approximately 40 °C, 120 min, and 1.3–1.5 g of fruit simultaneously maximizes extraction of both pigments.

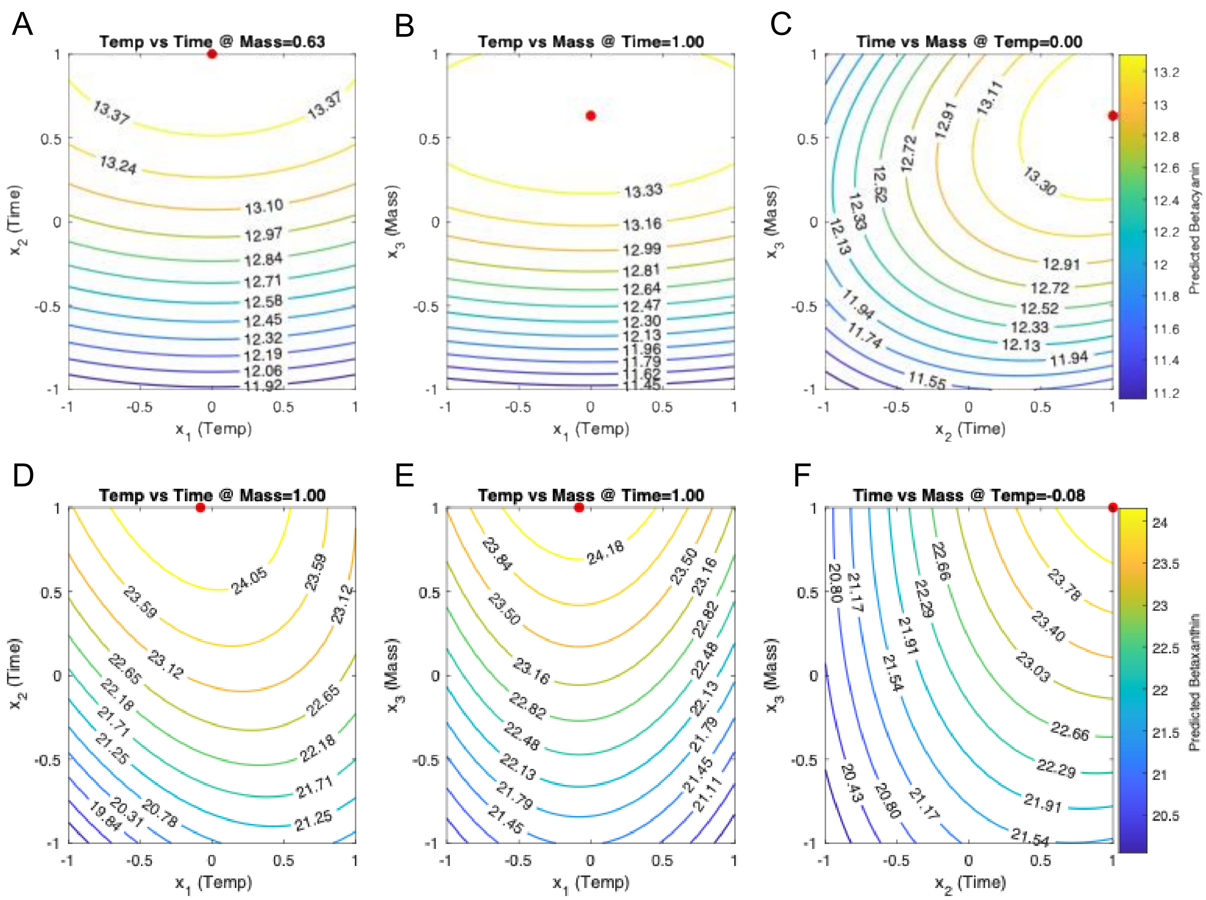


Figure 12. Contour plots of the optimized response surfaces for betacyanin (A–C) and betaxanthin (D–F), with the red marker highlighting each process's predicted optimum operating condition.

<u>Question 6:</u> Comment on what are the differences you find in the main effect plots, in the interaction plots, and in the response surface models built with Matlab® and Minitab®.

The same analysis of the PigmentOne extraction experiment was performed in Minitab to compare the overall results to those generated in MATLAB. Plots for main effects, interaction effects, and response surface models can be found in Figure 13 for Betacyanin and Figure 14 for Betaxanthin. Both in Minitab and MATLAB figures, the betacyanin and betaxanthin surface-and-contour plots are essentially indistinguishable. Longer extraction times and larger sample masses consistently boost pigment yields, while temperature has a milder, slightly curved effect peaking around 40 °C. The 3D roofs and 2D contour hills align perfectly between software, highest pigment levels sit in the upper-right corner of each contour plot (high time, high mass) with only a gentle arch in the temperature direction.

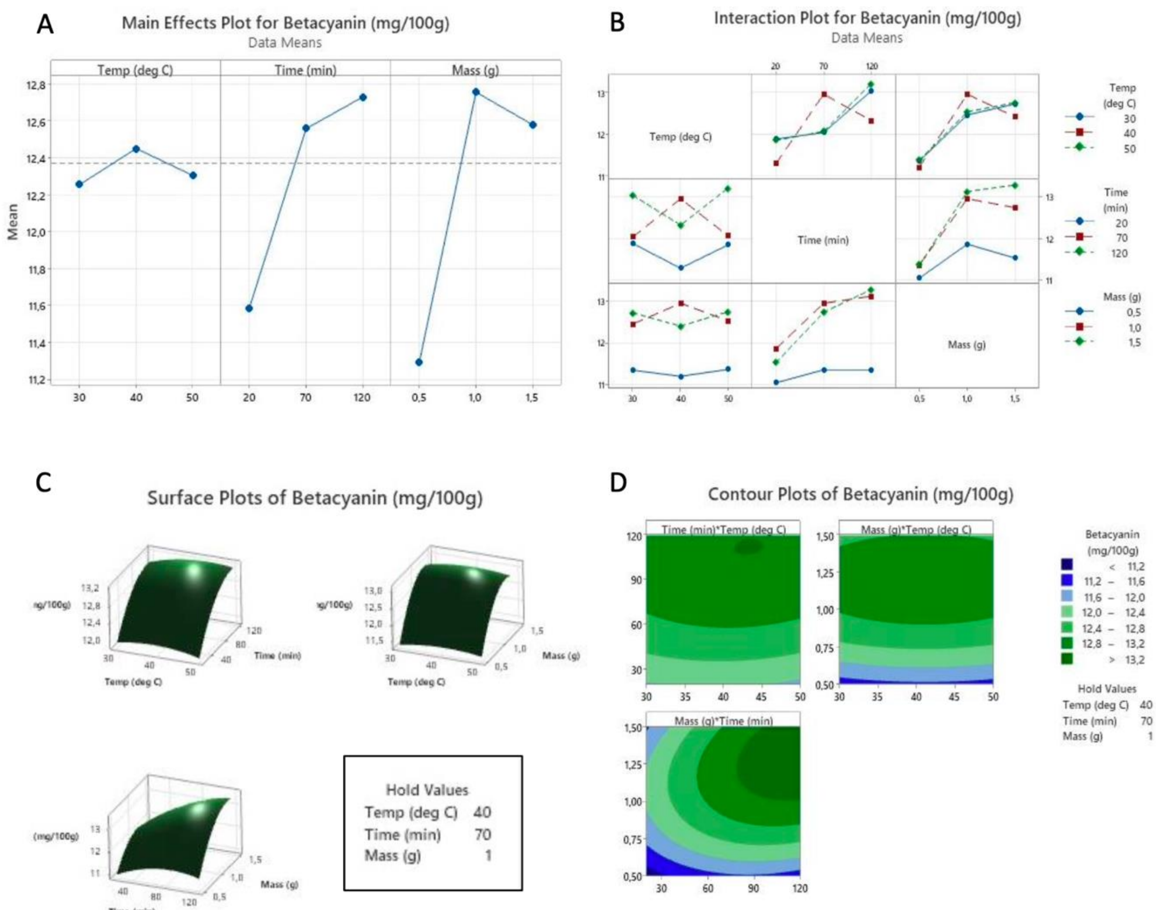


Figure 13. Minitab visualization of the Betacyanin extraction response-surface model with (A) Main-effects plot showing the individual impact of temperature, time, and mass; (B) Interaction plot highlighting significant factor interplay; (C) Three-dimensional response surfaces at the center point (T = 40 °C, Time = 70 min, Mass = 1 g); (D) Contour plots of the same slices with color- coded yield regions.

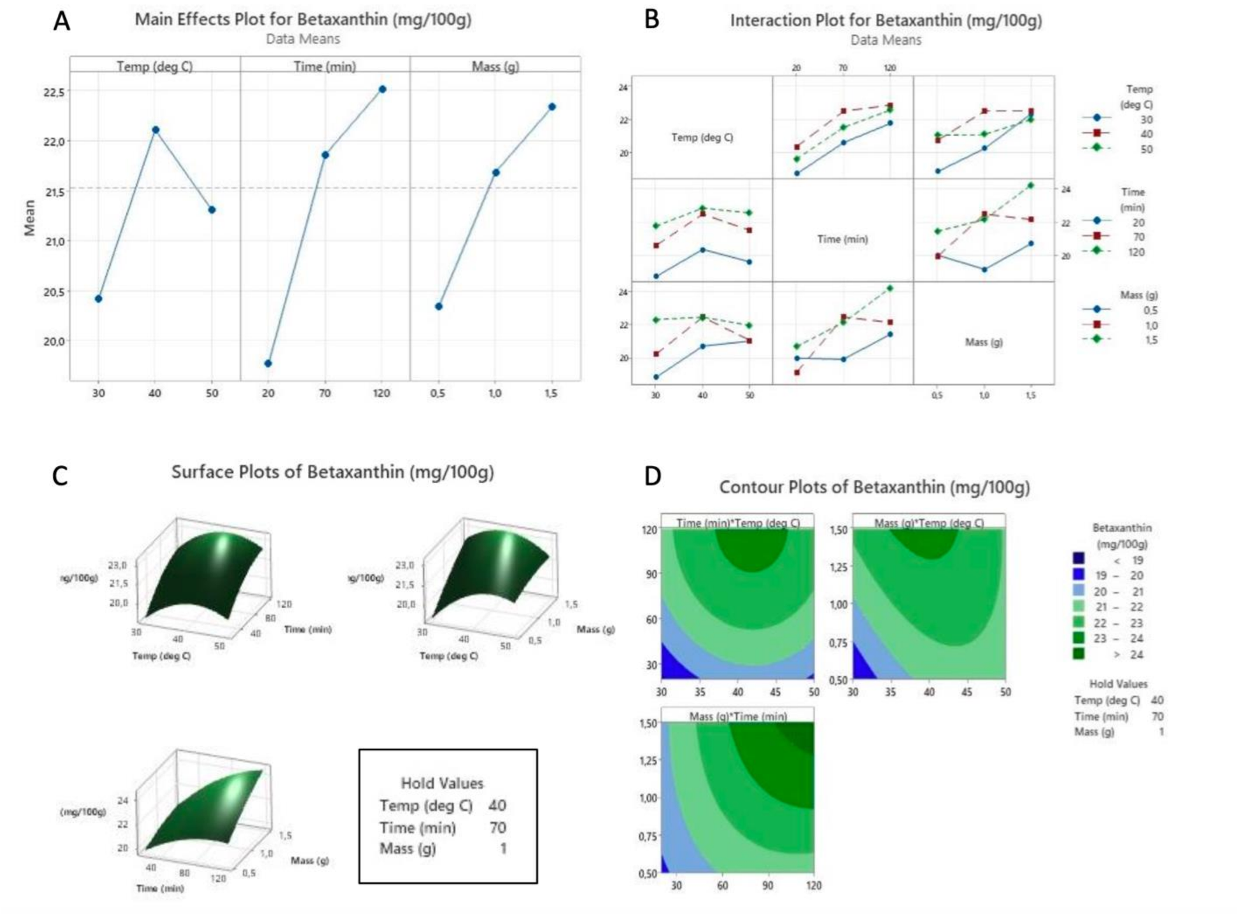


Figure 14. Minitab visualization of the Betaxanthin extraction response-surface model with (A) Main-effects plot showing the individual impact of temperature, time, and mass; (B) Interaction plot highlighting significant factor interplay; (C) Three-dimensional response surfaces at the center point (T = 40 °C, Time = 70 min, Mass = 1 g); (D) Contour plots of the same slices with color-coded yield regions.

In short, aside from minor stylistic differences Minitab's default plotting and the figures generated in MATLAB convey the same trend and regions for maximizing the pigment yield.

<u>Question 7:</u> Comment on what are the differences you find in the optimal points found with Matlab® and Minitab®

A multi response prediction model was made in Minitab to find the optimal conditions of the pigment process for the extraction of both pigments, Betacyanin and Betaxanthin, simultaneously. The results from the optimization can be found in Figure 15. The Minitab multi-response optimization process consolidates the individual pigment optima into a single compromise solution by maximizing a composite desirability function. While the MATLAB analyses independently found the best betacyanin yield at (39.2 °C, 120 min, 1.50 g), the Minitab approach selects (≈39.1 °C, 120 min, 1.50 g) to jointly optimize both. This slight temperature adjustment (from 40 to ~39 °C) and the adoption of the higher mass level reflect the balancing of competing objectives: push mass high enough to favor betaxanthin without substantially sacrificing betacyanin, and tune temperature to a compromise midpoint. The resulting predictions (betaxanthin ≈24.5 mg/100 g, betacyanin ≈13.4 mg/100 g) demonstrate that a multi-response desirability framework can effectively lead to a practical and implementable set of conditions that optimize pigments simultaneously instead of just individually.

Parameters

Response	Goal	Lower	Target	Upper	Weight	Importance
Betaxanthin (mg/100g)	Maximum	18,72	24,27		1	1
Betacyanin (mg/100g)	Maximum	11,06	13,29		1	1

Solution

Solution	Temp (deg C)	Time (min)	Mass (g)	Betaxanthin (mg/100g) Fit	Betacyanin (mg/100g) Fit	Composite Desirability
1	39,0909	120	1,5	24,5238	13,3818	1

Multiple Response Prediction

Variable	Setting
Temp (deg C)	39,0909
Time (min)	120
Mass (g)	1,5

Response	Fit	SE Fit	95% CI	95% PI
Betaxanthin (mg/100g)	24,524	0,161	(24,143; 24,904)	(23,943; 25,104)
Betacyanin (mg/100g)	13,3818	0,0671	(13,2231; 13,5406)	(13,1397; 13,6240)

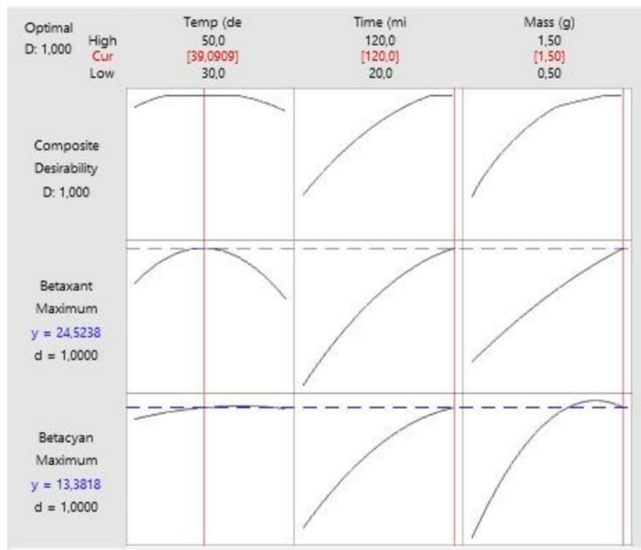


Figure 15. Optimized settings from the Minitab desirability analysis showing target and weight functions for each factor, composite-desirability solution at T≈39 °C, Time=120 min, Mass=1.5 g, and predicted betaxanthin and betacyanin yields with 95 % confidence intervals.

Conclusion

This project applied DOE and response-surface methodology to model and optimize the extraction of two fruit pigments, betacyanin and betaxanthin, as functions of temperature, time, and fruit mass. Main-effects and interaction plots revealed that time and mass strongly increase both yields, with modest quadratic curvature indicating diminishing benefits at extremes, while temperature played a smaller role. Separate single-response optimization identified best-case settings for each pigment, and a multi-response desirability approach then reconciled those into a unified operating point (≈40 °C, 120 min, 1.5 g) that delivers high yields of both pigments simultaneously. Overall, the workflow demonstrated how empirical modeling, ANOVA, and desirability functions can guide efficient process tuning when multiple quality criteria must be balanced.