JMP CASE STUDY: IMPURITY ANALYSIS CASE STUDY

Prepared by: CHEW CHIA YIK

017 585 4932 ccyik0205@gmail.com

July 2025

Executive Summary

This case study focuses on determining the key process variable which affect the impurity levels in a polymer production system with JMP Design of Experiment and Prediction Profiler. The goal was to achieve the average impurity levels of 3% while ensuring all outputs remain below the **upper specification limit of 7%**.

The study revealed that the impurity level is significantly affected by **catalyst concentration**, **temperature**, **reaction time**, and specific **reactor-shift combinations**. In particular, **lower catalyst concentration, lower temperature**, and longer reaction times are critical for minimising impurities.

The Prediction Profiler identified the optimal conditions:

- **Lowest temperature**
- **Lowest catalyst concentration**
- **Maximum reaction time**
- **Reactor 1, Shift 2**

Under these conditions, the impurity level is predicted to be 2.6 % which is well below 3%, achieving both objectives of process improvement and quality assurance objectives.

Contents

Introduction

Upholding product purity is crucial for ensuring high yield and consistent quality of polymer production. This case study is adapted from a tutorial provided by JMP. A catalyst is used to enhance the chemical reactions for the formation of a polymer molecule, but it also enhances the side reaction which leads to impurities in the final product. This case study will focus on process optimisation for minimising the impurity level.

The improvement team has been assigned with two tasks:

1. **Reduce the average impurity to 3%**, and
2. **Ensure all products consistently stay below the 7% limit**.

To achieve these goals, statistical modelling and data analysis tools such as **Design of Experiments (DOE)**, **Prediction Profilers**, and **Contour Plots** in the JMP software were utilised to comprehend the process variable interaction and determine the optimal operating conditions. The interactive HTML and CSV dataset used in this case study is available at the appendix section.

1. Actual vs. Predicted Plot

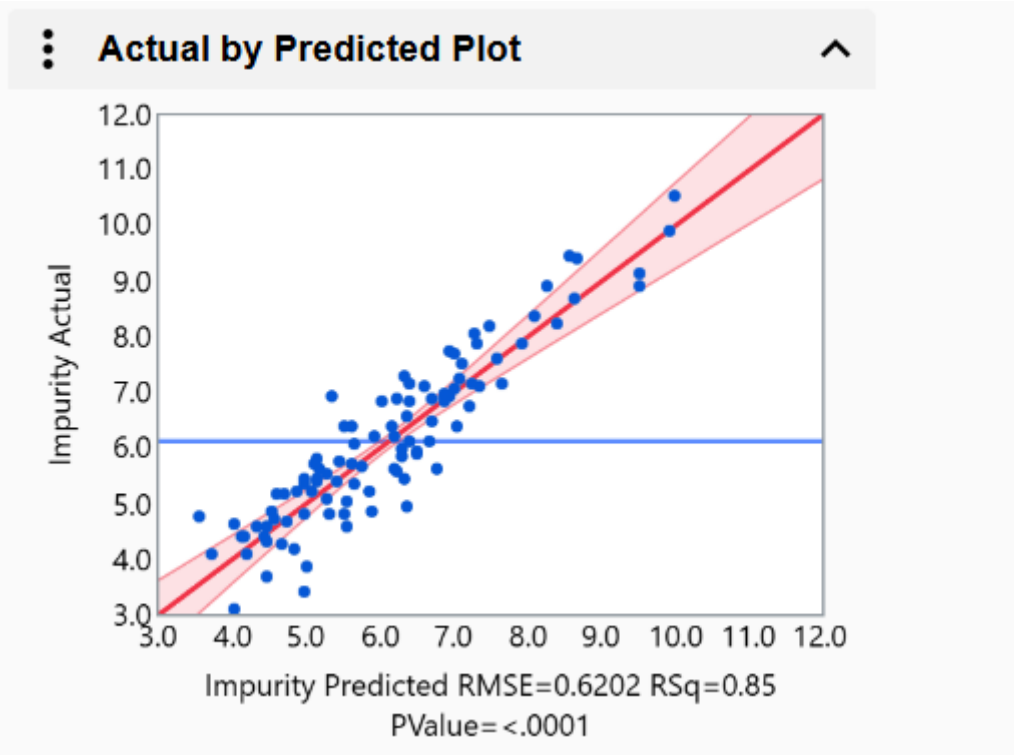

Figure 1: Actual vs Predicted Plot

The Actual vs. Predicted plot serves as a diagnostic tool for the evaluation of performance and accuracy of the developed regression model. The predicted values closely align with the actual experimental data, suggesting a good model fit as shown in Figure 1. The model achieves a **coefficient of determination (R^2) of 0.85**, indicating that **the selected predictors explain 85% of the variability in the response variable**. Additionally, the **Root Mean Square Error (RMSE) of 0.62** signifies a relatively low average deviation between predicted and actual values, reinforcing the model's precision. Moreover, the p-value less than 0.0001 signifies statistical significance, further validating the overall relevance of the model and firmly rejecting the null hypothesis.

2. Effect Summary

Effect Summary

Source	Logworth	PValue
Catalyst Conc	13.727	0.00000
Temp	9.570	0.00000
Catalyst Conc*Reaction Time	2.688	0.00205
Reactor	2.486	0.00327
Temp*Reaction Time	2.325	0.00473
Reactor*Shift	2.185	0.00653
Shift	0.400	0.39830 ^
Reaction Time	0.006	0.98661 ^

Remove Add Edit Exclude Undo FDR

('^' denotes effects with containing effects above them)

Figure 2: Effect Summary of the Model

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Temp	1	1	19.487802	50.6637	<.0001*
Catalyst Conc	1	1	32.173877	83.6445	<.0001*
Reaction Time	1	1	0.000109	0.0003	0.9866
Reactor	2	2	4.698960	6.1081	0.0033*
Shift	1	1	0.277092	0.7204	0.3983
Temp*Reaction Time	1	1	3.229334	8.3955	0.0047*
Catalyst Conc*Reaction Time	1	1	3.881053	10.0898	0.0020*
Reactor*Shift	2	2	4.098363	5.3274	0.0065*

Figure 3: Effect Tests of the Model

Figure 2 provides an overview of the significance of each term and interaction using the LogWorth and associated p-values. The LogWorth represents the negative log10 of the p-value, offering a simple visualisation of each term and sorting them based on their importance in the model.

Figure 3 displayed the **F ratio** and **exact p-values** for each effect in more detail. The **F ratio** reflects the ratio of model variance (signal) divided by the term to the residual variance (noise), indicating how much the term contributes to explaining the response variable. A larger

F ratio indicates a larger signal, coupled with a small p-value (typically < 0.05), suggests a statistically significant effect and not by random fluctuation.

The key findings from Figure 2 and Figure 3 were evaluated as below:

- **Catalyst Concentration** (CatalystConc) and **Temperature (Temp)** are the **most significant main effects**, both with p-values < 0.0001, indicating a **strong influence** on impurity levels.
- Several **interaction effects** are also significant:
    - **Catalyst Concentration * Temperature** (p = 0.002)
    - **Temperature * Reaction Time** (p = 0.004)
    - **Reactor * Shift** (p = 0.0065).

    These interactions suggest that the effect of one variable depends on the level of another, highlighting the importance of considering combined effects in process optimisation.
- The **Reactor** factor (p = 0.003) is quite significant, implying that impurity level is significantly affected among different reactor units.
- In contrast, **Shift** (p = 0.3983) and **Reaction Time** (p = 0.9866) are **not statistically significant**, suggesting they have minimal direct impact on the model. However, they are retained in the model to preserve hierarchical structure. This is due to their presence in statistically significant interaction terms:

    - **Temperature * Reaction Time** (p = 0.004)
    - **Reactor * Shift** (p = 0.0065)

The Principle of Model Hierarchy (Montgomery et al., 2005) implies that :

The main effects or parent effect should be kept in the model if its interaction with other variables is significant, even if the main effect is insignificant by itself.

Removing these terms would potentially distort the interpretation of interaction effects by violating the model hierarchy. In other words, **these variables influence the response, but only in combination with other factors.**

3. Residual vs. Predicted Plot

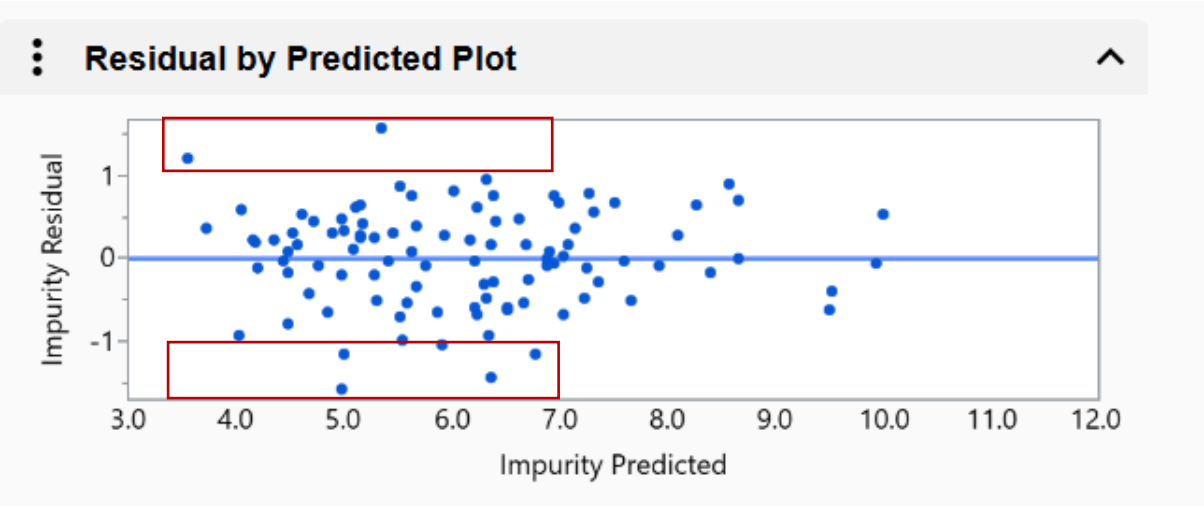

Figure 4: Residual by Predicted Plot (*Red box indicates outliers)

The Residual vs. Predicted plot reveals how residuals (differences between observed and predicted values) are distributed across the prediction range. This method evaluates the appropriateness of the model by observing on potential bias between the model and the actual results.

Out of the 100 data points:

- **2 points have residuals of +1**
- **8 points have residuals of –1**
- The **remaining 90% (90/100) of points** fall within a tight range between –1 and +1

This distribution suggests that:

- The residuals are **fairly symmetrically spread** around zero, with **slightly more negative residuals**, but not enough to indicate strong bias.
- There is **no visible funnelling (heteroscedasticity)** as the residuals do not fan out as predicted values increase or decrease.
- No obvious **curvature patterns** in the residuals, implying that the linear model is fitted,
- The **low count of high-magnitude residuals (±1)** out of 100 total points indicates **no significant outliers** affecting model performance.

Overall, the residual plot supports the assumptions of linear regression and further validates the robustness of the model.

4. Parameter Estimates

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-18.2951	4.995863	-3.66	0.0004*
Temp	0.7916865	0.111226	7.12	<.0001*
Catalyst Conc	1.8498432	0.202263	9.15	<.0001*
Reaction Time	-0.000746	0.044308	-0.02	0.9866
Reactor[1]	-0.289212	0.0934	-3.10	0.0026*
Reactor[2]	-0.038882	0.095072	-0.41	0.6835
Shift[1]	0.0531726	0.062648	0.85	0.3983
(Temp-27.4669)*(Reaction Time-91.0086)	-0.230141	0.079427	-2.90	0.0047*
(Catalyst Conc-1.47897)*(Reaction Time-91.0086)	0.4339683	0.136621	3.18	0.0020*
Reactor[1]*Shift[1]	0.2842803	0.089941	3.16	0.0022*
Reactor[2]*Shift[1]	-0.202025	0.089324	-2.26	0.0262*

Figure 5: Parameter Estimates of the Model (*Orange Coloured font indicates SIGNIFICANT terms)

The **Parameter Estimates** table provides information on how each factor contributes to the response variable, impurity concentration. Both **Temperature (Estimate = 0.79, p < 0.0001)** and **Catalyst Concentration (Estimate = 1.85, p < 0.0001)** show strong, statistically significant positive effects on impurity, indicating that increases in these variables tend to increase impurity levels as well.

While **Reaction Time** is not significant solely as a main effect (p = 0.9866), it participates in two statistically significant interactions — with **Temperature (p = 0.0047)** and **Catalyst Concentration (p = 0.0020)**. The same goes for the **Shift** factor (p = 0.3983),its interaction with reactor type is significant for both Reactor 1 (p = 0.0022) and Reactor 2 (p = 0.0262). Additionally, **Reactor 1** shows a significantly lower impurity output with estimate = -0.29, as a negative sign indicates lower compared to the baseline (p = 0.0026), while **Reactor 2** does not differ significantly. Table 1 below summarises the interpretation for each term and interaction involved in the model.

Table 1: Interpretation of Each Term of Parameter Estimates

Term	Estimate	t Ratio	p-Value	Interpretation
Intercept	–18.30	–3.66	0.0004	The baseline level where all predictors are at reference values.
Temp	+0.79	7.12	<.0001	Highly significant; for each 1°C increase, impurity increases by 0.79 units.
Catalyst Conc	+1.85	9.15	<.0001	Strongest effect: each unit increase in catalyst concentration raises impurity by 1.85 units.
Reaction Time	~0	–0.02	0.9866	Not significant alone — negligible direct effect.
Reactor[1]	–0.29	–3.10	0.0026	Significantly lower impurity in Reactor 1 vs baseline.
Reactor[2]	–0.04	–0.41	0.6835	Not significant.
Shift[1]	+0.05	0.85	0.3983	Not significant on its own.
Temp * Reaction Time	–0.23	–2.90	0.0047	Significant interaction — effect of Temp depends on Reaction Time.
Catalyst * Reaction Time	+0.43	3.18	0.0020	Also significant interaction.
Reactor[1] * Shift[1]	+0.28	3.16	0.0022	Significant interaction.
Reactor[2] * Shift[1]	–0.20	–2.26	0.0262	Significant interaction.

(***Remarks: In the model, we included both Reactor and Shift as categorical variables. Because Reactor has three levels and Shift has two, JMP automatically selects one level from each as a reference point (baseline) — in this case, Reactor 3 and Shift 2. The effects of other levels are then compared against these baselines. That's why you won't see Reactor 3 or Shift 2 directly listed in the parameter estimates — they're the points of comparison.)

5. Interaction Profiles

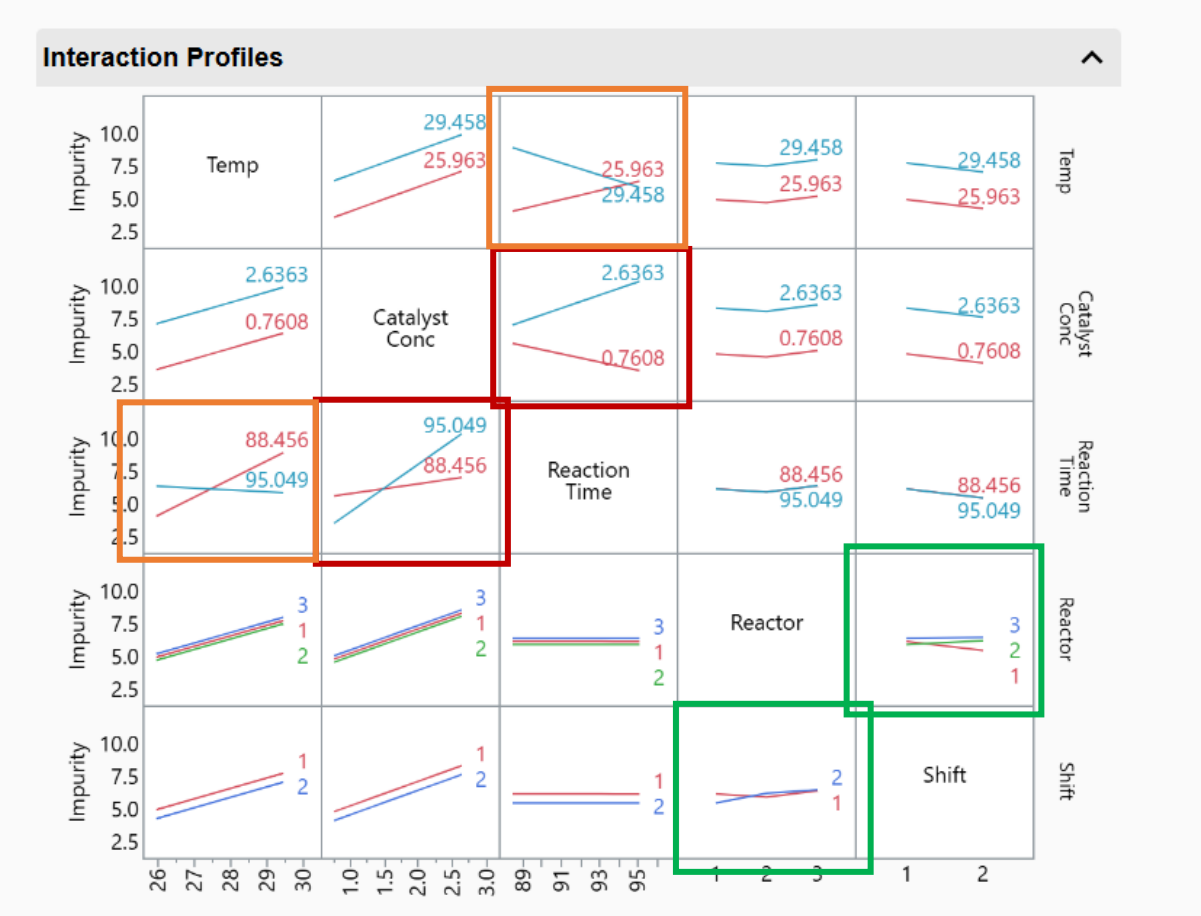

Figure 6: Interaction Profile

An **Interaction Plot** helps us understand how two factors influence a response **together**, rather than in isolation. To explain this simply, imagine you're cooking. You're adjusting both **temperature** and **cooking time**. An interaction means:

- The **best cooking time depends on the temperature** used.
- Or: Changing the **temperature has different effects depending on the cooking time**.

In the interaction plot, **non-parallel lines** suggest that the effects of one factor depend on the level of another. This means their **combined effect matters** and cannot be optimised independently.

In **Figure 6**:

- The **red box** highlights the interaction between **Catalyst Concentration and Reaction Time**.
- The **orange box** shows the interaction between **Temperature and Reaction Time**.
- The **green box** indicates interaction between **Reactor and Shift**.

These lines are **not parallel**, and in some cases, they **cross,** which indicates a **statistically significant interaction**. In contrast, the rest of the plots with **parallel lines** suggest little to no interaction between those factor pairs.

6. 3D Contour Plot

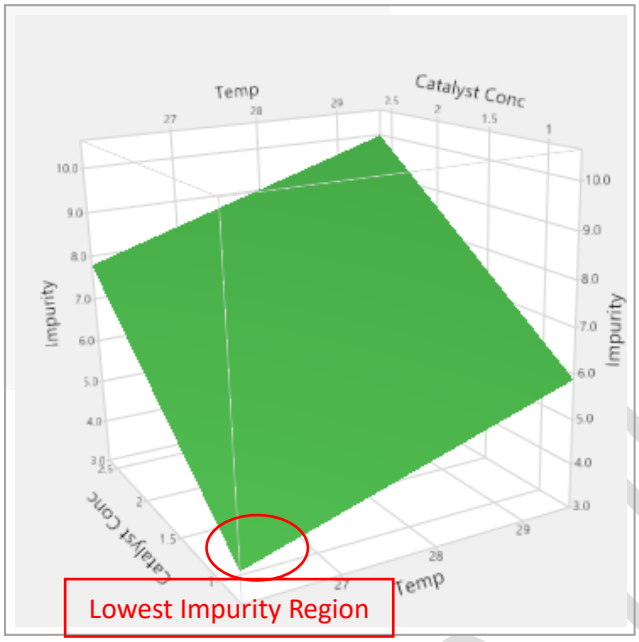

Figure 7: 3D Contour Plot of Impurity to Catalyst Concentration and Temperature

A **3D Contour Plot** visually illustrates how two independent variables simultaneously affect a response variable. In this case, the **x-axis and y-axis represent the two selected factors** (Catalyst Concentration, Temperature), while the **contour levels indicate the predicted values of the response** (Impurity level).

This plot enables easy identification of:

- **Optimal regions** (Zone with lowest impurity).
- **Interactions between variables**, where the shape of the contours can suggest whether factors interact.

In this case, the 3D contour plot exhibits **diagonal and evenly spaced contour lines**, indicating a **linear and additive relationship** between Temperature and Catalyst Concentration. The absence of curvature or distortion in the contour lines suggests that there is **no significant interaction** between these two variables. This implies that the effects of Temperature and Catalyst Concentration on Impurity are independent of each other, and their contributions to the response are simply additive non-interactive. The **optimal condition** corresponding to the **lowest impurity level** is located at the **lowest values of both Temperature and Catalyst**

Concentration as indicated by the red circle. Thus, impurity can be effectively minimised by jointly reducing both parameters.

7. Prediction Profiler

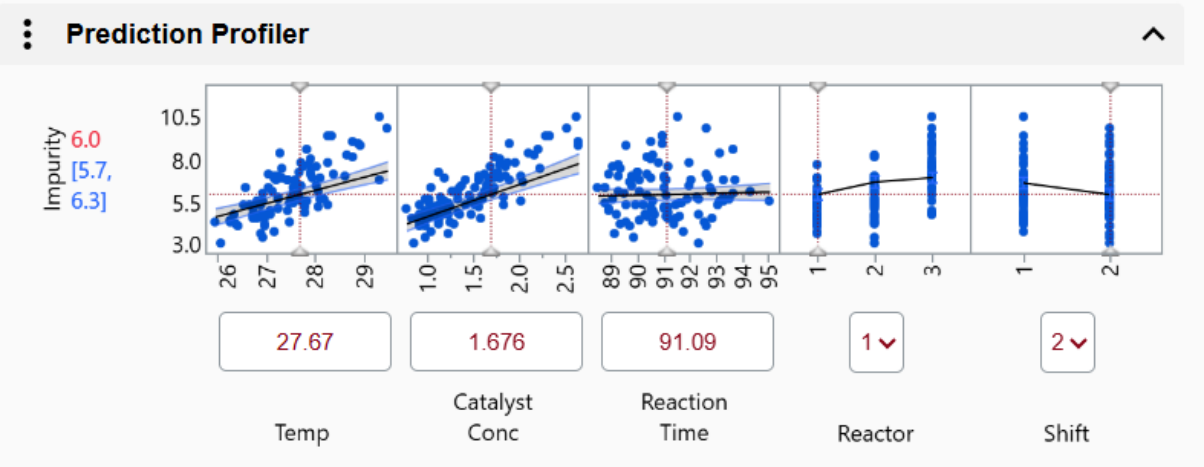

Figure 8: Prediction Profiler

The **Prediction Profiler** visualise how each input variable affects the output variable (Impurity) while maintaining other variables at a constant level. The line in each profile is the trend of the response variable as a function of each input.

In this case, the profiler reveals that:

- **Temperature** and **Catalyst Concentration** both exhibit **positive slopes**, indicating that increasing either variable tends to **increase the impurity level**.
- **Reaction Time** also shows a horizontal trend, suggesting a **non-significant effect on the impurity.**
- **Reactor** and **Shift** are categorical factors. The profiler shows distinct step changes between different levels, suggesting that impurity levels significantly differ among different reactors and shifts.

The slopes and step sizes in each profiler enable fast identification of how each variable influences the response variable, where the steeper slope symbolises a greater effect. Different combinations of factor levels with their effect on the response impurity can be observed by adjusting the profiler sliders.

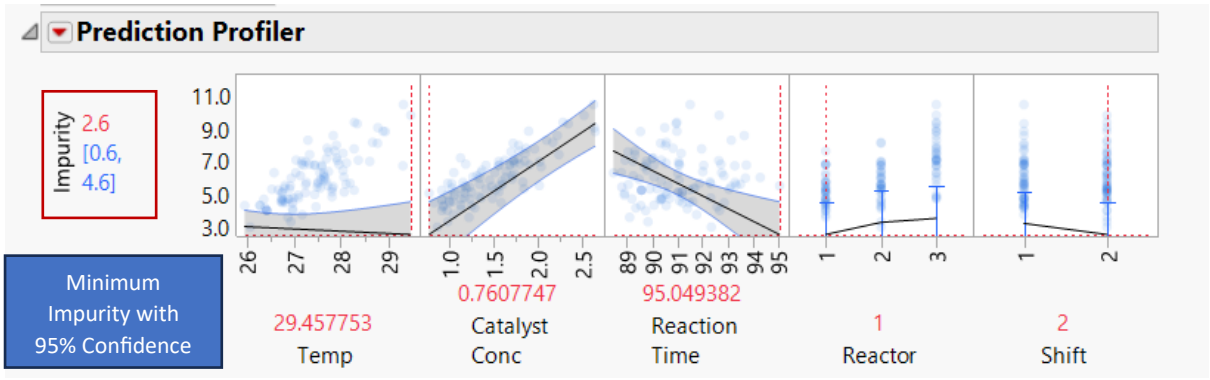

Figure 9: Optimised Prediction Profiler

Figure 9 shows the optimal setting of the Prediction Profiler to minimise impurity level to value of 2.6:

- **Lowest temperature, 29.46**
- **Lowest catalyst concentration, 0.76**
- **Maximum reaction time, 95.05**
- **Reactor 1**
- **Shift 2**

The following behaviours were observed in the optimised profiler:

- **In Figure 8, Temperature** initially shows a **positive relationship** with impurity, but the slope flattens as **reaction time increases**. This suggests an interaction effect—longer reaction time reduces the sensitivity of impurity to temperature changes.
- **Catalyst concentration** maintains a **consistently positive slope**, indicating that higher concentrations of catalyst contribute to higher impurity levels, regardless of other settings.
- **Reaction time** changes from a flat slope to a **negative relationship**, where longer times reduce impurity level.

These observation highlights the implementation of DOE and Prediction Profiler in optimisation while visualising the interactive effect between the process variables. (In this case, longer reaction time leads to insensitivity of impurity to temperature). To conclude the findings of this optimisation study, impurity can be effectively reduced while staying within specification limits by maximising reaction time and minimising both temperature and catalyst

concentration. The minimised impurity level is around 2.6 with CI of [0.6,4.6], which successfully achieved the desirable target value and within specs.

Conclusion

The analysis demonstrated that the impurity levels in the polymer production can be minimised by tuning two main variables, which are the catalyst concentration and temperature. Interaction effects, such as temperature with reaction time suggest that the process behaviour is not purely additive, highlighting the significance of interaction in process modelling and optimisation.

We observe that increasing both temperature and catalyst concentration tends to increase the impurity level regardless of the reactor. This is in contrast with general chemical engineering principles, where higher temperatures and catalyst loading are typically employed to boost reaction rates and yield. However, the relationship appears linear, meaning impurity consistently rises with both parameters in this case. Suggesting that while higher temperature and catalyst doses may enhance production capacity, they likely promote side reactions or thermal degradation, leading to a rise in impurities as well.

Reducing temperature and catalyst concentration would lower impurities but also compromise production throughput. Unfortunately, data on yield, throughput or production constraints are unavailable, making it impossible to determine an optimal balance between product capacity and quality. This contradiction emphasises a common industrial dilemma: maximising purity often comes at the cost of production rate and vice versa. Future studies should incorporate process throughput or yield metrics to enable multi-objective optimisation strategy.

References

Montgomery, D. C., Myers, R. H., Carter, W. H., & Vining, G. G. (2005). The hierarchy principle in designed industrial experiments. *Quality and Reliability Engineering International*, *21*(2), 197–201. https://doi.org/10.1002/qre.615

Appendix

The interactive HTML of the JMP can be viewed here: https://cchiayik.github.io/Impurity-Analysis-Case-Study/Impurity%20JMP%20-%20Fit%20Least%20Squares.htm

The dataset .csv file available here: https://github.com/cchiayik/Impurity-Analysis-Case-Study/blob/impurity-analysis-case-study/data/Impurity.csv