Optimization of a batch reactor process using statistical analysis

by

Kara DiNatale

B.S., Clarkson University, 2014

A REPORT

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Chemical Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2019

Approved by:

Major Professor
Dr. John Schlup

Abstract

A batch reactor is used to remove an environmental concern (EC) from the wastewater stream of an industrial process at Novelis, Inc.'s, Oswego, NY, facility. Between May 2017 and February 2018 major process changes occurred across the machine centers that produce the wastewater. This led to expected decreases in the concentration of EC in the wastewater, how much wastewater was produced, and several other changes. As a result, the process engineer responsible for the reactor decided that the amount of Chemical A and Chemical B, the reactants, added to each batch would be reduced. Additionally, batches are run less frequently, often with smaller volumes. These changes were not expected to have any effects on reactor efficiency. However, after some time it was noted that the reaction was consistently taking longer to complete than previously and the final concentration of B was often higher than before the changes were made.

With multiple changes having been made around the same time, to both the source of the wastewater and the reactor, it is difficult to understand which changes, or combination of changes, caused this shift. This report details a series of statistical analyses which were used to gain a better understanding of the connections between the changes observed, the resulting shifts, and optimization of the reactor operation. Through the use of factor analysis it was found that the seven potentially relevant input variables could be reduced to three components. Through ANOVA four variables were determined to have significant impacts on both the length of time for the reaction to complete and the final outcome of the reaction. Based on these findings adjustments were made, additional data was collected, and new analyses were run. From this second round of analyses the following change was recommended: the 42 liters of A and 36 liters of B should be used for every cubic meter of wastewater to be treated.

Table of Contents

List of Figures iv
List of Tables v
Acknowledgements vi
Dedication vii
Chapter 1 - Introduction 1
Background 1
Batch Reactor Process and Equipment 4
Equipment 4
Process 5
General Data Collection and Sample Testing 6
Objectives 7
Chapter 2 - STAT 730 Final Project: Examining a Shift in an Industrial Process through Multivariate Statistical Methods 9
Methodology 9
Results and Analysis 11
Conclusion 15
Chapter 3 - Additional Statistical Analysis of RXN 16
Design of Experiment and Methodology 16
Results and Analysis 18
Conclusions 23
Chapter 4 - Effects on B Concentration 24
Design of Experiment and Methodology 24
Results and Analysis 25
Conclusions 27
Chapter 5 - Conclusions and Future Work 28
Summary and Conclusions 28
Recommendations and Future Work 29
Chapter 6 - References 31

List of Figures

Figure 1.1 Simplified diagram of the batch reactor system for the removal of EC 2
Figure 1.2 Individual Values Control Chart for RXN 3
Figure 1.3 Vial of indicator strips for B concentration with gradient scale shown 7
Figure 2.1 Histogram and quartile plot of IN_conc for the whole dataset (left) and the A_vol=34 L/m^3_{ww} population (right). 12
Figure 2.2 Factor pattern plot for factors 1 and 2 of a three-factor analysis..... 14
Figure 3.1 Control chart of IN_conc split into stages based on A_vol (L/m^3_{ww}) 17
Figure 3.2 Histogram and quartile plot of IN_conc (top) and RXN (bottom) for the original dataset (left) and the expanded dataset (right). 19
Figure 3.3 Factor pattern plot for factors 1 and 2 of a three-factor analysis..... 21
Figure 4.1 Plot of B Concentration over time for all collected batches..... 25
Figure 4.2 Plot of B Concentration over time for batches in IN_range d (left) and c (right) 26

List of Tables

Table 1.1 Variables in the dataset, their abbreviations, and their units of measure..... 1
Table 2.1 Mean values of variables used for analysis for the complete dataset and the two subpopulations 13
Table 3.1 Pairings of A and B volumes 16
Table 3.2 Variables used in statistical models and their definitions..... 17
Table 3.3 IN_Range values and their corresponding IN_conc ranges..... 18
Table 3.4 Number of batches and average RXN by IN_Range and A_vol 20
Table 3.5 Number of batches and LSMEANS RXN by IN_Range and A_vol with Standard Error and Confidence Intervals 22

Acknowledgements

I would like to thank my colleagues at Novelis, Inc's Oswego plant for allowing me to use this project for my Master's Report. I would also like to thank the managers I've had since we commissioned this reactor for allowing me to be involved with it even though it is not in the regular scope of my responsibilities. Specifically, I would like to thank Jim M. and Ted M. for allowing me the opportunity to function as the process engineer for this reactor, starting prior to its commissioning in 2016, disregarding the fact that I was a quality engineer at the time.

Dedication

I dedicate this to all the friends I didn’t hang out with, the campsites I didn’t enjoy, and the hockey games I didn’t watch because I was working on schoolwork instead.

Chapter 1 - Introduction

Background

A batch reactor is used to remove an environmental concern (EC) from the wastewater stream of an industrial process at Novelis, Inc.'s, Oswego, NY, facility. A variety of both input and output variables are recorded by the reactor operators and tracked by the process engineer responsible. This list can be found in Table 1.1 below. It should be acknowledged that these have been scrubbed of any potentially proprietary or identifying information. A simplified flow diagram illustrating the basics of the equipment and batch reactor process can be seen in Figure 1.1. A more detailed description of the reactor process and equipment can be found later in this chapter.

Abbreviation	Variable	Abbreviation	Variable
Batch	Batch number	Pre_pH	pH before chemical C
Date	Date batch ran	Final_pH	pH after chemical C
MY	Month/year of Batch	Delta_pH	Difference between Pre_pH and Final_pH
Days	Days since last batch	IN_conc	Concentration of EC in wastewater before treatment (ppm)
Centers	Number of Machine Centers w/process change	PR_conc	Concentration of EC in wastewater after treatment (ppm)
IN_vol	Volume of wastewater in batch (L)	Ve_conc	Verification of concentration of EC in wastewater after treatment (ppm)
A_vol	Volume of chemical A used (L/m^3_{ww})	% R_PR	% reduction in EC based on PR_conc
B_vol	Volume of chemical B used (L/m^3_{ww})	% R_VE	% reduction in EC based on VE_conc
C_vol	Volume of chemical C used (L/m^3_{ww})	RXN	Amount of time it took for the reaction to complete (min)

Table 1.1 Variables in the dataset, their abbreviations, and their units of measure

Between May 2017 and February 2018 major process changes occurred across the three aluminum strip chemical treatment machine centers that produce the wastewater. These changes are known to have caused shifts in the concentration of aqueous EC in the influent (IN_conc), the number of work centers affected (Centers), and the number of days between batches (Days). As a result, changes were also made reducing the volume of chemicals A and B used per volume of influent wastewater (A_vol and B_vol) since their values were stoichiometrically based on the previous maximum expected value of IN_conc.

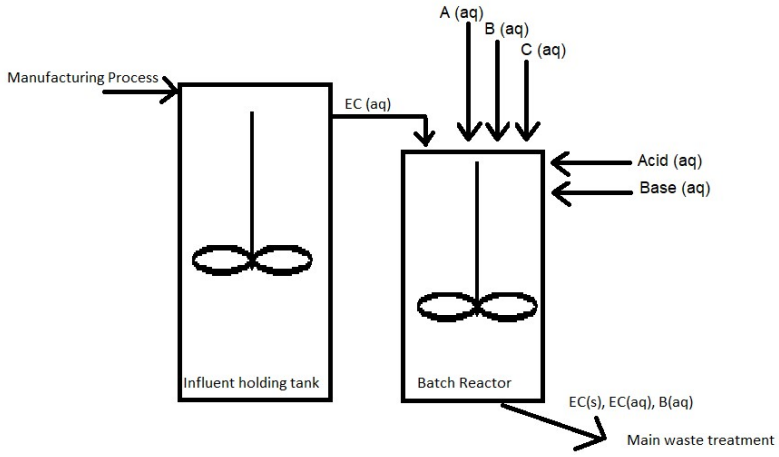

Figure 1.1 Simplified diagram of the batch reactor system for the removal of EC

These adjustments were done in order to reduce material costs and safety concerns related to the batch reactor process. Chemical B, (represented by B_vol) has high risks in regards to both reacting with other chemicals and the physical safety of the operators, so it is desirable to minimize the amount required to be kept onsite. It was not expected that these adjustments would have any impact on the other variables. However, a shift in the time it takes for the reaction to be completed

for each batch (RXN) can be seen in the control chart[1] in Figure 1.2 around batch number 360. Also, chemical B was noted to be present in higher concentrations upon batch completion than previously. With several changes having occurred around that same time frame, it is difficult to determine if any one of them caused the observed shift and higher output concentrations of B.

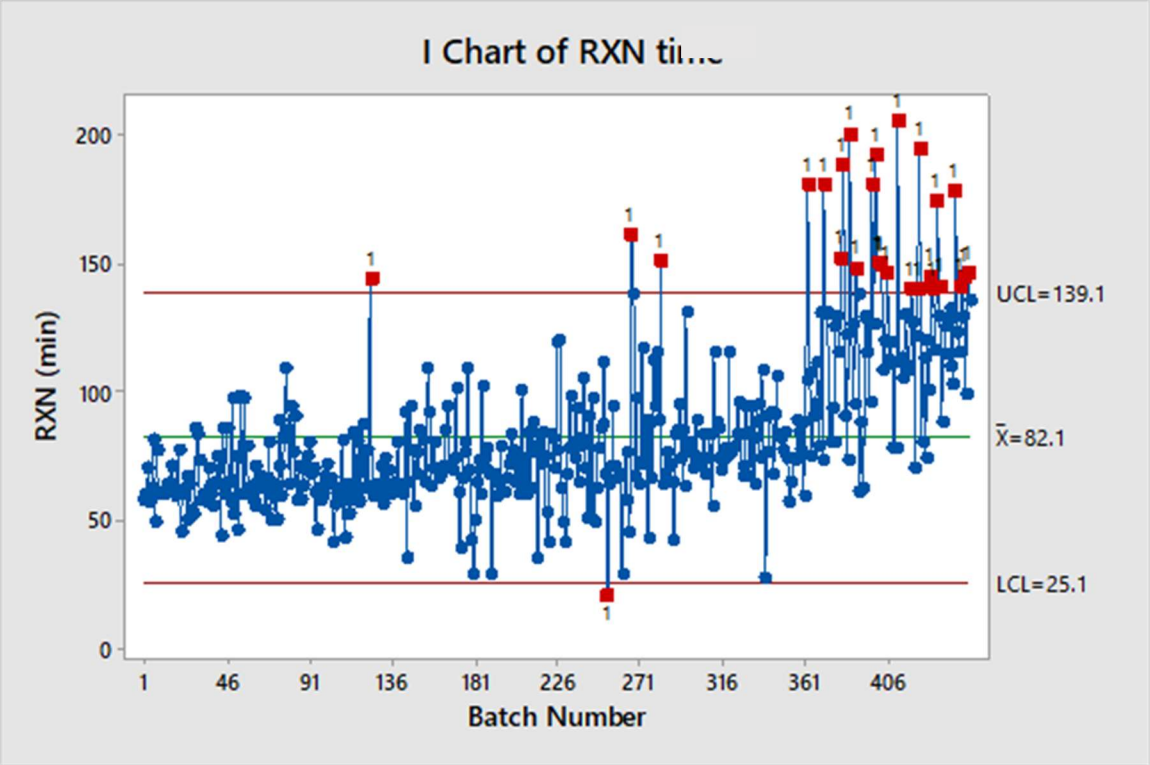

Figure 1.2 Individual Values Control Chart for RXN

[1] Control charts are a tool commonly used in manufacturing and business applications to monitor process stability. In this case it is a time series plot of individual measurements. The middle line, labeled $\bar{x}$, represent the mean value of the measurements. The lines labeled LCL and UCL are the Lower and Upper Control Limits, which are defined as 3 sigma from the mean. Values outside these limits are considered to be "out of control" and marked differently, in this case with red. A process which has a string of out of control values, such as the shift being examined in the project, is said to be out of statistical control. (American Society for Quality, 2005)

It should be noted, however, that the stoichiometric ratios of chemicals A and B to EC is an estimate as the process chemical containing EC is proprietary to Novelis' main competitor and the exact composition is not known. Additionally, the exact details of the chemical reactions in the batch reactor are not well understood by the process engineer as the system was designed and tested by a contracted company specializing in wastewater treatment systems and is partially proprietary to that contracted company.

Batch Reactor Process and Equipment

Equipment

The batch reactor system is comprised of a 50 m^3 influent tank, a 25 m^3 reactor tank, three smaller day tanks each less than 5 m^3, and one tote pumping station. The influent tank, the batch reactor, and the day tank for chemical C have motorized agitators to stir their contents. They also have digital tank level read outs; the other two day tanks have mechanical tank level indicators. Additionally, the reactor has a built in pH probe with a digital read out. Large centrifugal pumps are used to transfer influent from the holding tank to the reactor and to transfer the completed batch from the reactor to the main wastewater treatment process. Smaller pumps are used to transfer each chemical into the reactor. Chemical D is not kept in a day tank and is pumped directly from nearby bulk storage, where the other day tanks are filled from. Chemical B is highly reactive and received in 1 m^3 totes. It is added to the reactor from the tote pumping station using a metered pump. All pipes are plastic. This system is completely manual and all volumes to be pumped are controlled and recorded by the reactor operator.

Process

This batch reactor process treats wastewater from four streams- three manufacturing machine centers and one water purification system. The stream received from each machine center varies in both volume and EC concentration based on what product is being produced due to different products requiring different amounts of the chemical containing EC. The stream from the purification system is the waste from the system's periodic regeneration (cleaning) cycle and has varying concentrations. The machine center streams are acidic while the purification system stream is basic. It should also be noted that there is no standard procedure regarding when the agitator in the holding tank should be running. As a result, it is not always mixing so there may be concentration gradients within the tank.

Influent is pumped from the holding tank to the reactor in volumes ranging from 12-17 m^3 per batch as determined by the operator. The agitator in the reactor is turned on once approximately 5 m^3 have transferred. At that point the operator will also begin pumping Chemical A, which is acidic, into the tank in a set amount based on the previously decided influent volume. The reaction is pH dependent and requires a starting pH of around 3.5. After both the influent and chemical A have been added to the reactor, the pH is read and chemicals D and E, which are an acid and a base, respectively, are added in small volumes as needed. A designated amount of chemical B is added to the reactor once the correct pH is achieved.

As mentioned earlier in this chapter, the exact reactions are proprietary to the equipment supplier and not well known. A best-guess of the series of reactions which occur within each batch is below.

$$A + B \rightarrow AB$$

$$AB + EC_{(aq)} \rightarrow AB + EC_{(s)}$$

As you can see, the byproduct of the reaction between chemicals A and B interacts with EC and transforms the EC from a soluble form to a non-soluble form which precipitates out of solution. The interaction between the byproduct and EC occurs rapidly upon the formation of the byproduct. The consumption of the reactants, observable as a decrease in measured concentration, can be used as an indicator of reaction completion.

Once all reactants are added to the batch reactor, some time is allowed to pass before the operator begins to monitor the progress of the reaction using indicator strips, which read the concentration of chemical B. Once the strips indicate that the concentration of B is less than 50 ppm, a sample is neutralized and filtered at the operator's station and then taken to the onsite chemical lab. At the lab ICP is used to test the concentration of aqueous EC. If this concentration is over 50 ppm, then the operator adds additional A and B to the reactor and repeats the process. If it is under 50 ppm the operator adds chemical C, which is a base, to the reactor until a pH between 7.4 and 8.5 is achieved. At that point the reactor contents are pumped to a tank in the main wastewater treatment process. In the main wastewater system a coagulant is added and the precipitated EC is filtered out.

General Data Collection and Sample Testing

Data for each batch is collected through three means. The standard procedure required that dates, volumes, pH's, B concentrations, and start/stop times were recorded on paper by the operators while each batch was being run. These recordings were later entered into a master spreadsheet by a process engineer. Also, per standard procedure, the operators used a port on the reactor to collect

samples from which B and EC concentrations were tested. The concentration of chemical B is tested using indicator strips, shown in Figure 1.3.

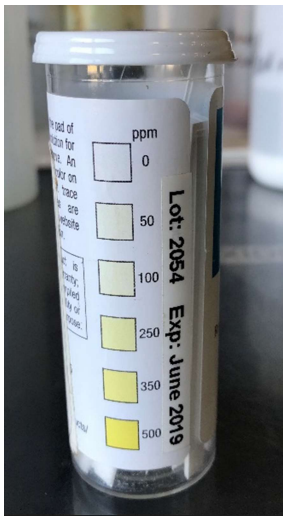

Figure 1.3 Vial of indicator strips for B concentration with gradient scale shown

EC concentration is tested using Inductive Coupled Plasma- Optical Emission Spectroscopy (ICP-OES), the results are recorded electronically by the onsite chemical testing lab, and the process engineer copies this information from the database into the master spreadsheet. The lab uses an Aligent Technologies 5100 series ICP-OES machine. For the purpose of this project, the number of days between batches and the amount of time it took a batch to complete the reaction, as indicated by the complete consumption of chemical B, were calculated by the process engineer based on information recorded by the operators.

Objectives

In the fall of 2018, an initial analysis was conducted regarding the shift in RXN and submitted as the final project for STAT 730: Multivariate Statistical Methods taught by Prof. Perla Reyes. These

results will be discussed in detail in Chapter 2. Based on the findings of that work, a series of additional objectives were developed which will be addressed in Chapter 3 and Chapter 4.

The first additional objective of this report is to continue the work of the statistical analysis regarding the shift in RXN. The second objective is to study how the changes to the process affected the final concentration of chemical B. The third objective is to use the results of objectives 1 and 2 to optimize the batch reactor in regards to both reaction time and final B. This optimization will allow the operators to be better able to plan their other tasks around the steps of this non-automated process, to determine a realistic amount of each chemical that should be stored on site, and to reduce risks resulting from unreacted B to both process efficiency and equipment reliability in the downstream wastewater processes.

Chapter 2 - STAT 730 Final Project: Examining a Shift in an Industrial Process through Multivariate Statistical Methods

In the fall of 2018 an initial analysis was conducted regarding the shift in RXN and submitted as the final project for STAT 730: Multivariate Statistical Methods taught by Prof. Perla Reyes. The objectives of this work were to find what, if any, relationship exists between RXN and the other variables and to determine which changes (or combination of changes) caused the shift observed in RXN.

To achieve these objectives a series of statistical methods were used on a dataset which included values for roughly 450 batches from January 2017 through September 2018 with known cause outliers removed. This dataset includes values for each variable listed in Table 1.1 for each batch. It should be noted that the variables C_vol, pre_pH, final_pH, delta_pH, Ve_conc, and %R_Ve are related to process steps which occur after the reaction has completed and were not included in this analysis. Additionally, %R_PR was also excluded as it is a function of IN_conc and PR_conc. The identifying variables Batch, date, and MY were also excluded from statistical considerations.

Methodology

It should be noted that some portions of the methodology are dependent on the results of prior analyses. Such instances are noted in the descriptions below, multiple methods for the same task may be described. Details regarding which method was chosen will be included in the Results and Analysis section of this chapter. All analysis for this chapter was performed using SAS/STAT™ software version 9.4 in the SAS Studio® web based platform.

Before beginning any multivariate analysis, exploratory statistics were examined. Univariate normality was reviewed through histograms and quartile-quartile (QQ) plots for each variable[2] which were generated via PROC UNIVARIATE. Additionally, the mean value for each variable was calculated using PROC MEANS. If certain forms of non-normality were observed, such as bimodal histograms or skewness, the dataset was split into multiple populations and this exploratory analysis was repeated for each population.

Following this, several multivariate methods for variable reduction were used. The first was Principal Components Analysis (PCA), which creates uncorrelated linear combinations of variables. These combinations are known as the Principal Components[3]. PCA was performed using PROC PRINCOMP on the whole dataset as well as on populations the dataset has during the exploratory analysis. The number of components to be considered in each iteration was determined by examining scree plots and using the rule of thumb that components with eigenvalues over 1 should be kept.

Factor Analysis (FA) was used employing the variables which weighed heavily in the PCA via PROC FACTOR. Similar to PCA, FA finds groupings of variables, referred to as factors, that are related to each other within the group but have limited correlations to variables in the other groups[4]. If the results of the PCA were roughly the same for the whole dataset and any populations, FA was performed for only the complete dataset. Otherwise, like with PCA, it was performed for the

---

[2] For more on univariate normality see (Mendenhall & Sincich, 2012). For more on QQ plots see (Ford, 2015).

[3]For more on Principal Components Analyses see chapter 8 of Applied Multivariate Statistical Analysis (Johnson & Wichern, 2007)

[4] For more on Factor Analysis estimation methods see chapter 9 section 9.3 of Applied Multivariate Statistical Analysis (Johnson & Wichern, 2007)

individual populations as well. If the exploratory statistics indicated that the data was non-normal FA was performed using the Primary Components (PC) method and a varimax rotation; if the data was found to be normal the Maximum Likelihood Estimation (MLE) was used, also a varimax rotation[5] was applied.

The final pair of methods used for objectives one and two were Analysis of Variance and Multivariate Analysis of Variance (ANOVA and MANOVA) which were performed using PROC GLM. For these RXN and PR_conc were treated as the response variables with the variables which weighed heavily in the PCA as the inputs in the model. SSIII and Wilke's Lambda results were examined to determine which variables were significant at an $\alpha=0.05$ level for the individual response variable ANOVAs and the MANOVA respectively.[6]

Results and Analysis

Upon examining the results of the exploratory analysis, it was clear that the complete dataset did not reflect univariate normality. The majority of the variables have discreet datapoints rather than being continuous; in fact, only that data for IN_conc, PR_conc, and RXN could be considered continuous. This was not unexpected given the batch nature of the process and the manual method of operation and data collection. It should be noted that the histograms for all three continuous variables were skewed and the QQ plots do not follow a straight line. The skewness of the data,

[5] For more on Factor Rotations see "Factor Analysis: A Short Introduction, Part 2- Rotations" (Rahn, n.d.) and chapter 9 section 9.4 of Applied Multivariate Statistical Analysis (Johnson & Wichern, 2007).

[6] For more on ANOVA and MANOVA see (Mendenhall & Sincich, 2012) and chapter 6 of Applied Multivariate Statistical Analysis (Johnson & Wichern, 2007).

combined with what was known about the effects of the manufacturing process changes, indicated that the collected data was likely a combination of two populations.

Since the process changes occurred gradually across multiple machine centers, it was decided to use the value of A_vol as the distinguisher to split the data into two populations, as that was not adjusted until after all process changes were completed. The exploratory analysis was rerun using the two populations, A_vol=66 L/m^3_{ww} and A_vol=34 L/m^3_{ww}, and univariate normality was reexamined.

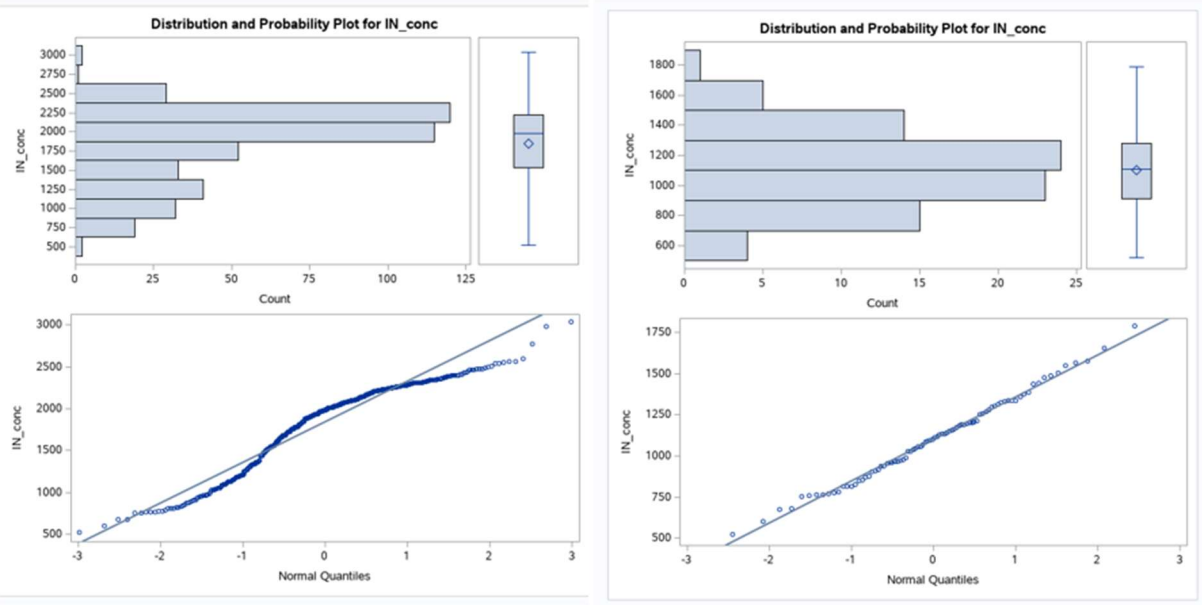

Figure 2.1 Histogram and quartile plot of IN_conc for the whole dataset (left) and the A_vol=34 L/m^3_{ww} population (right).

While there was no change to the normality of the discrete variables or PR_conc, both IN_conc and RXN are notably more normally distributed when the data is split. Figure 2.1 compares IN_conc for the whole dataset against that of the A_vol=34 L/m^3_{ww} population. The effects of

splitting the data can also be seen by the differences in the mean value of each variable displayed in Table 2.1

Variable	**Complete**	**A_vol=66** L/m^3_{ww}	**A_vol=34** L/m^3_{ww}
Days	1.42	1.11	2.73
Centers	1.27	0.86	3
IN_vol (L)	15931	16210	14756
A_vol (L/m^3_{ww}*)*	59.86	66	34
B_vol (L/m^3_{ww}*)*	51.63	57	29
IN_conc (ppm)	1842.6	2019.3	1103.1
PR_conc (ppm)	12.85	11.41	18.92
RXN (min)	81.64	72.23	121.26

Table 2.1 Mean values of variables used for analysis for the complete dataset and the two subpopulations

Following this, two different variable reduction methods were used. Through principal components analysis (PCA) it was determined that the variables could be reduced from the seven listed in Table 2.1 (excluding the response variable RXN) to 2 or 3 variables (principal components). The results of the PCA also indicated that the complete dataset, not the two populations, should be used for any further analysis.

Next, factor analysis (FA) was conducted for two and three-factor systems based on the findings of the PCA. It was determined that the three-factor system made the most sense as the factors could easily be named based on the variables they contain: “Inputs”, “Frequency”, and “Batch Volume”. As you can see in the factor pattern plot in Figure 2.2, which is a plot of the weighted values of each variable in relation to Factors 1 and 2. Inputs consist of A_vol/B_vol and IN_conc, Frequency contains Days and Centers, and Batch Volume is synonymous with IN_vol. The same groupings were observed in the plots of Factors 1 and 3 and Factors 2 and 3. A factor analysis keeping only

two factors was also examined but its results are less distinct than the three factors version and could be summarized as “Batch Volume” and “Everything Else”.

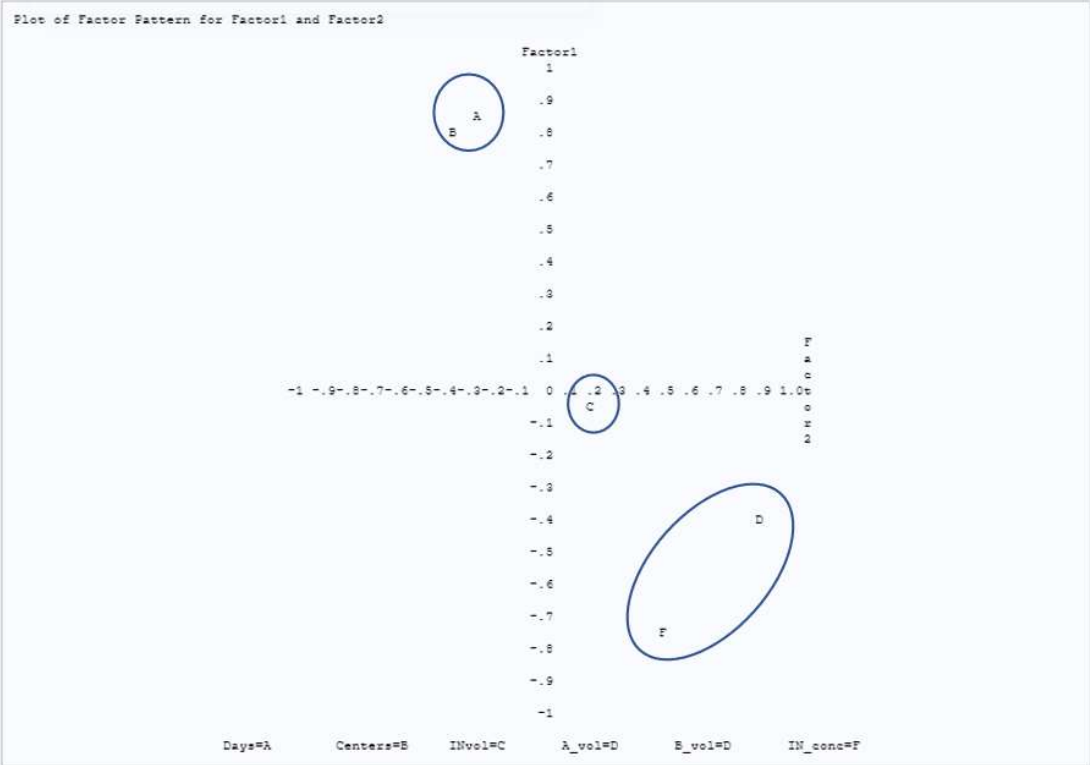

Figure 2.2 Factor pattern plot for factors 1 and 2 of a three-factor analysis

Following this an analysis of variance (ANOVA) was performed for the response variable RXN as well as for PR_conc, which also had experienced a shift in average as can be noted in Table 2.1. The models being examined were:

PR_conc = Days + IN_vol + IN_conc + A_vol + Centers,

RXN = Days + IN_vol + IN_conc + A_vol + Centers, and

PR_conc*RXN = Days + IN_vol + IN_conc + A_vol + Centers.

It was found that, using a 95% confidence interval, Days and IN_vol were statistically significant[7] to PR_conc. They were also statistically significant in regards to RXN along with A_vol and IN_conc. A multivariate analysis of variance (MANOVA) for RXN and PR_conc yielded similar results. These results are not what was expected, particularly Centers' lack of significance. It could be speculated, from a statistics standpoint, that this is the result of the data's lack of normality and failure to meet the ANOVA and MANOVA assumptions. However, it could also be speculated, from a chemical engineering standpoint, that IN_vol, which is somewhat related to Days, which is related to Centers, may have an impact on mixing efficiency within the batch reactor, which in turn could have an impact on reaction rate.

Conclusion

In conclusion, the data can be split into two populations but it is more reasonable to analyze it as one for the purpose of this project. Through both PCA and FA, the seven potentially relevant input variables can be reduced to 3 components. From FA these components can be interpreted as "Inputs", "Frequency", and "Batch Volume". In regards to the first and second objectives, using both ANOVA and MANOVA, it is clear that some of the input variables have a significant impact on RXN. Specifically, Days, IN_vol, A_vol, and IN_conc are significant which also reflects the gist of the factors from the FA.

[7] "Statistically significant" denotes that there is a high probability that a variable has an impact on the response variable.

Chapter 3 - Additional Statistical Analysis of RXN

Design of Experiment and Methodology

In order to meet the objectives detailed in Chapter 1, several steps were employed. The first was to tackle objective one by running a series of batches with varying values for A_vol and B_vol, shown in Table 3.1 below. The values were determined by following a linear trend between the amounts that were used before the changes to the machine centers (Pair 1) and the original post-changes amounts (Pair 5) as these should be in a stoichiometric ratio with the expected IN_conc.

Pair #	**A_vol (**L/m^3_{ww})	**B_vol (**L/m^3_{ww})
1	66	57
2	58	50
3	50	43
4	42	36
5	34	29

Table 3.1 Pairings of A and B volumes

It was planned that each pairing would be used for a minimum of 1 month and that they would be implemented sequentially. However, further changes to the machine centers resulted in the expected value of IN_conc shifting several times. After each shift, the stoichiometrically appropriate pairing from Table 3.1 was implemented. This shown in the control chart in Figure 3.1, which has been broken into stages. The start of each shift is marked by a series of red points indicating that IN_conc is out of statistical control. The use of each pairing is marked by vertical dashed lines and is labeled with the A_vol value for that pair. It should be noted that no batches were run using Pair 2 and that all batches with Pair 1 were run prior to any adjustments being made to the reactor process.

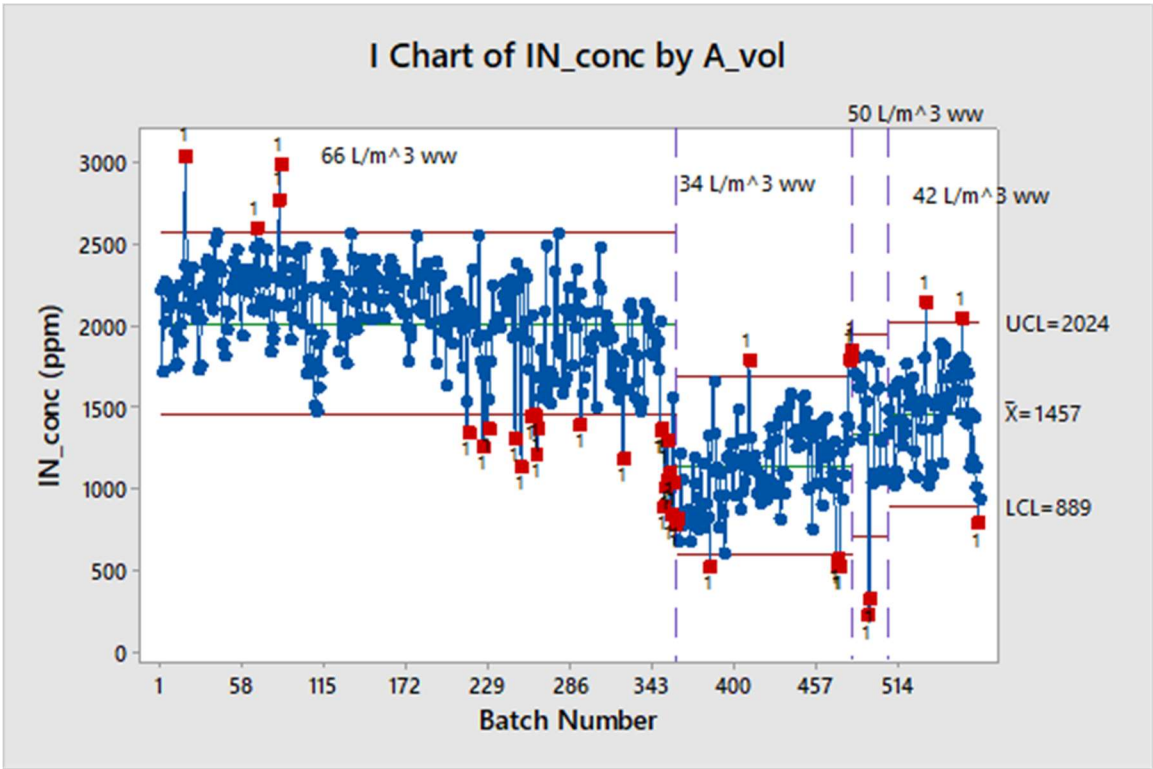

Figure 3.1 Control chart of IN_conc split into stages based on A_vol (L/m^3_{ww})

In addition to the use of specified pairings, the variables to be included in the statistical models were adjusted based on the analyses from Chapter 2 and a new classification variable, IN_Range, was added. The ranges for each value of IN_Range can be found in Table 3.3 below. The new list of variables, their units, and definitions can be seen in Table 3.2 below.

Abbreviation	Variable
Batch	Batch number
Days	Days since last batch
IN_vol	Volume of wastewater in batch (L)
A_vol	Volume of chemical A used (L/m^3_{ww})
IN_conc	Concentration of EC in wastewater before treatment (ppm)
PR_conc	Concentration of EC in wastewater after treatment (ppm)
RXN	Amount of time it took for the reaction to complete (min)
IN_Range	Groupings of batches based on their IN_conc values

Table 3.2 Variables used in statistical models and their definitions

IN_Range	IN_conc (ppm)
a	> 1900
b	1600<[EC]<1900
c	1300<[EC]< 1900
d	<1300

Table 3.3 IN_Range values and their corresponding IN_conc ranges

These variables were analyzed through exploratory statistics, variable reduction methods, and ANOVA in a manner similar to that described in the methodology section of Chapter 2. The dataset used for this analysis includes 570 batches from January 2017 through September 2019. The analysis was performed using SAS/STAT™ software version 9.4 in the SAS Studio® web based platform and Microsoft® Excel® 2016, version 1909 for Windows©.

Results and Analysis

As in Chapter 2 upon examining the results of the exploratory analysis, which were generated using PROC UNIVARIATE, it was clear that the expanded dataset did not reflect univariate normality. Since the only new variable, IN_Range, is a classification variable[8], again only the data for IN_conc, PR_conc, and RXN could be considered continuous for the same reasons provided previously. However, compared to the original dataset in Chapter 2, the expanded dataset approaches a normal distribution for both IN_conc and RXN, as can be seen in Figure 3.2.

Additionally, the data was broken down by IN_Range and the number of batches for each range and each value of A_vol were counted using the pivot table feature in Excel®. The average values of RXN for each combination of IN_Range and A_vol were also calculated in the pivot table. This

[8] Classification variables are variables that have set "classes" or levels, rather than continuously collected datapoints. These values can be set numerical intervals (5ft, 10ft, 15ft… etc) or categorical values (a, b, c…etc).

is displayed in Table 3.4. There are three main takeaways from this information. First, the average reaction time was notably lower for all ranges when A_vol=66 L/m^3_{ww}, which was expected. Second, there does not immediately appear to be a significant difference in reaction times between A_vol=34 L/m^3_{ww} and A_vol=42 L/m^3_{ww} for ranges b and d, but A_vol=42 L/m^3_{ww} had slightly lower reaction times for range c. Finally, the dataset was not collected as a balanced design, which was not unexpected due to its nature and origin. Given these, and the closer resemblance to normality, it was determined that it was not necessary to split the data for analysis as separate populations.

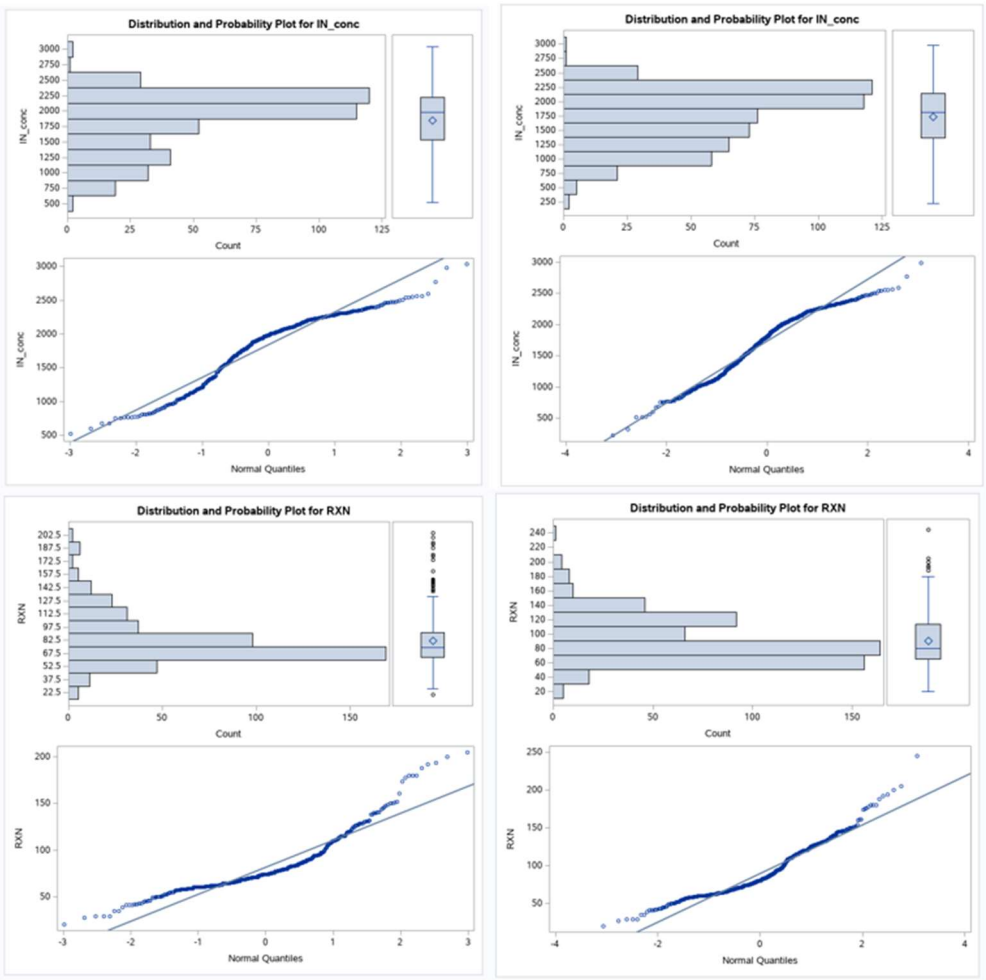

Figure 3.2 Histogram and quartile plot of IN_conc (top) and RXN (bottom) for the original dataset (left) and the expanded dataset (right).

IN_Range	A_vol (L/m^3_{ww})	Number of Batches	Average RXN (min)
>1900 ppm		***259***	***70.9***
	42	2	124
	66	257	70.49
1900>[EC] >1600ppm		***97***	***94.02***
	34	5	128.8
	42	18	126.72
	50	10	127.1
	66	64	76.94
1600>[EC]>1300 ppm		***88***	***107.51***
	34	32	127.19
	42	25	115.4
	50	5	107
	66	26	75.81
<1300 ppm		***126***	***113.06***
	34	85	118.11
	42	18	118.61
	50	10	108.7
	66	13	75.77

Table 3.4 Number of batches and average RXN by IN_Range and A_vol

Next, principal components analysis (PCA) was subsequently run via PROC PRINCOMP and, as in Chapter 2, the results indicated that three components were appropriate. Based on this, factor analysis (FA) was run for 3 factors using PROC FACTOR. It was found that these factors made sense as "Volume" "Stoichiometry", and "Frequency". No rotation was necessary.

However, PCA and FA can only be run on non-classification input variables, and the variable list had already been reduced based on the results of the work in Chapter 2. Thus, this analysis only reduced the variables from 4 to 3. As you can see in Figure 3.3, only "Stoichiometry" (circled in red) contains more than one variable, IN_conc and A_vol.

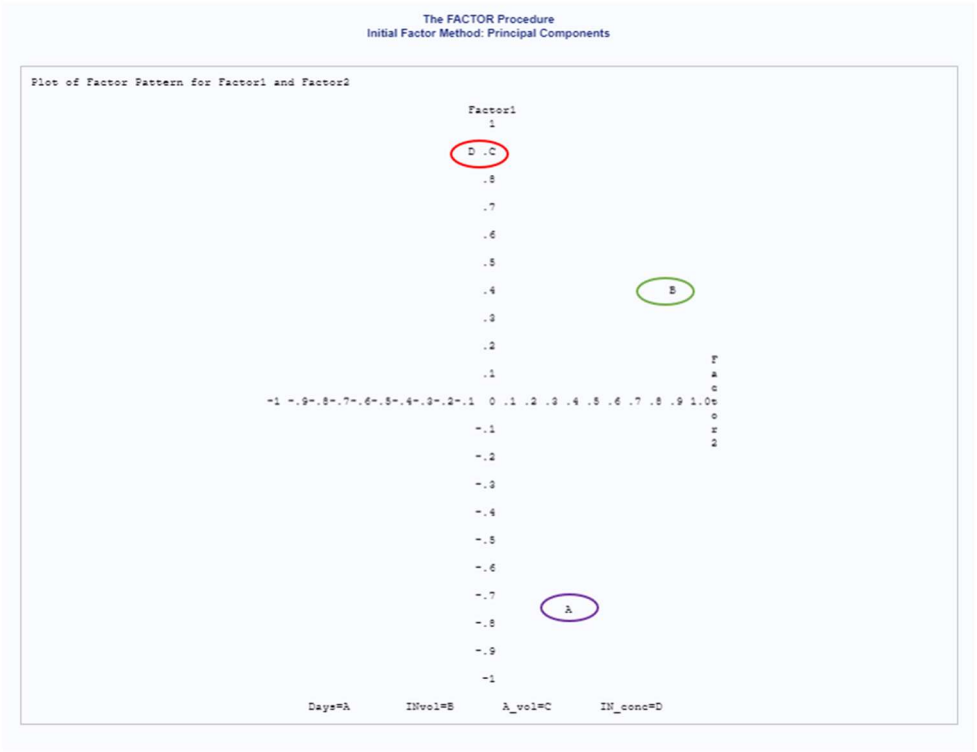

Figure 3.3 Factor pattern plot for factors 1 and 2 of a three-factor analysis

Using PROC GLM, an ANOVA was run with RXN as the response variable and IN_conc, A_vol, IN_vol, and Days as the input variables using the general linear model. The model being fitted was:

RXN = A_vol + IN_conc + IN_vol + Days + A_vol*IN_conc + A_vol*IN_vol + A_vol*Days + IN_conc*IN_vol + IN_con*Days + IN_vol*Days.

The results indicated that at a 95% confidence level IN_conc and its interaction with A_vol had significant impacts on reaction time. However, neither IN_vol, Days, nor A_vol had significant impacts on their own, unlike in Chapter 2's analysis.

A second ANOVA was run, this time using the mixed model (PROC MIXED), with IN_range and IN_conc denoted as random variables. This ANOVA examined the interaction between A_vol and IN_range, as well as the interaction between A_vol and IN_conc, and found that there was no significant impact on reaction time as a result of either interaction. The model being fitted was:

RXN = A_vol*IN_conc + A_vol*IN_Range.

IN_Range	A_vol	RXN LSMEAN	Standard Error	Confidence Interval
>1900 ppm				
	42	123.47	19.47	(85.22, 161.72)
	66	75.27	3.67	(68.0, 82.49)
1900>[EC] >1600ppm				
	34	127.42	9.95	(107.87, 146.96)
	42	126.73	5.24	(116.44, 137.03)
	50	126.42	7.09	(112.49, 140.34)
	66	77.19	2.78	(71.73, 82.66)
1600>[EC] >1300 ppm				
	34	137.17	5.55	(126.27, 148.06)
	42	115.78	9.42	(97.29, 134.28)
	50	102.21	11.92	(78.79, 125.62)
	66	73.03	4.78	(63.63, 82.42)
<1300 ppm				
	34	140.42	9.08	(122.58, 158.26)
	42	119.59	21.84	(76.68, 162.49)
	50	96.98	17.56	(62.48, 131.48)
	66	68.35	8.11	(52.42, 84.27)

Table 3.5 Number of batches and LSMEANS RXN by IN_Range and A_vol with Standard Error and Confidence Intervals

The least squared means (LSMEANS) estimates of RXN were also generated for each combination of A_vol and IN_range. These values, shown in Table 3.5, along with the associated standard errors and confidence intervals, are slightly different than the averages in Table 3.4. Based on the

LSMEANS, there appears to be a notable difference between A_vol=34 L/m^3_{ww} and A_vol=42 L/m^3_{ww} in both IN_ranges c and d. When looking at the pairwise comparisons of the LSMEANS for these combinations, there is not a statistically significant difference in either range at the 95% confidence level. However, in range c, where the p-value was 0.0509 it can be argued that there is a *practical* difference between A_vol=34 L/m^3_{ww} and A_vol=42 L/m^3_{ww}.

Overall, when looking at the LSMEANS values, it can be noted that for IN_ranges b, c, and d, the value for A_vol=42 L/m^3_{ww} is lower than that for A_vol=34 L/m^3_{ww}. The values for A_vol=50 L/m^3_{ww} and A_vol=66 L/m^3_{ww} are also lower, but the higher volumes of A (and subsequently B) are undesirable.

Conclusions

In conclusion, the additional analysis of RXN using an expanded dataset indicate that neither the frequency nor the size of batches have a significant impact on how long the reaction will take. In agreement with the earlier analysis, the incoming concentration of EC and the volume of reactants used do have an impact. Further examination of these factors, using the LSMEANS, show that the higher the A_vol value, the shorter the reaction time. However, the operators, engineers, and others at the plant would prefer to use less of the reactants, so A_vol=50 L/m^3_{ww} and A_vol=66 L/m^3_{ww} are undesirable despite their shorter times. Given that, A_vol=42 L/m^3_{ww} has the most favorable values and should be used.

Chapter 4 - Effects on B Concentration

Design of Experiment and Methodology

In order to meet the second objective, batches were sampled throughout the reaction every 15 minutes and tested for concentration of chemical B using the indicator strips shown in Figure 1.3. These concentrations were recorded on paper at the time of testing and later transcribed into a spreadsheet by the process engineer, along with the corresponding IN_conc, IN_range, and IN_vol for each batch. Due to the additional work this task requires of the operators, it was not practical to do this for every batch; so this was planned to be done for two batches per week for six weeks. However, as the result of a combination of maintenance activities and process development trials utilizing the machine centers but not the chemical containing EC, insufficient wastewater was generated during the first two intended weeks of data collection. The collection period could only be extended one additional week due to other time constraints. As a result, B concentration data was collected for ten batches between September 16, 2019 and October 14, 2019.

It should be noted that, in some instances, data were recorded every 30 minutes rather than every 15 minutes. In these cases, the missing concentration was interpolated using the two adjacent points. It should also be noted that for one batch, 91919, the operator did not wait until the B concentration was below 50 ppm, as per procedure, and took the final sample to the lab after 150 minutes when the concentration was still 350 ppm. They did not collect any concentration data after that point. Similarly, batch 101019 took over 1000 minutes to reach but the data collection sheet only had spaces for up to 225 minutes. Therefore, B concentrations were not recorded between 225 minutes and when 50 ppm was reached for that batch. Additionally, it should be acknowledged that all batches were run with A_vol=42 L/m^3_{ww}.

The analysis was performed using SAS/STAT™ software version 9.4 in the SAS Studio® web based platform and Microsoft® Excel® 2016, version 1909 for Windows©. The collected B concentration values were plotted in Excel® over time for each batch to see if any rate curve stood out. Additionally, ANOVA using a general linear model (PROC GLM) was conducted with IN_conc, IN_range, and IN_vol as the input variables and the B concentration after 60 minutes as the response variable. ANOVA using a general linear model was also run with the same input variables and the number of minutes it took each batch to reach 100 ppm as the response variable.

Results and Analysis

Upon reviewing the plots of the B concentrations for each batch, displayed in Figure 4.1, no common slope is immediately obvious. When separating the plots by IN_range, shown in Figure 4.2, a slight pattern is visible in range d but not in c. Overall, even when examining based on IN_range, no consistent rate curve can be isolated though most batches reach 50 ppm B within 135 minutes.

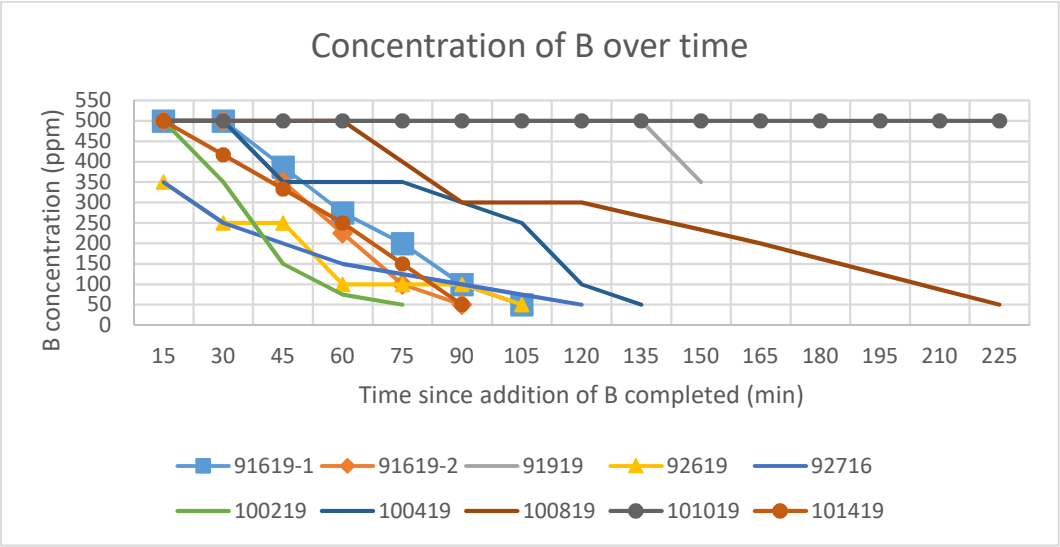

Figure 4.1 Plot of B Concentration over time for all collected batches

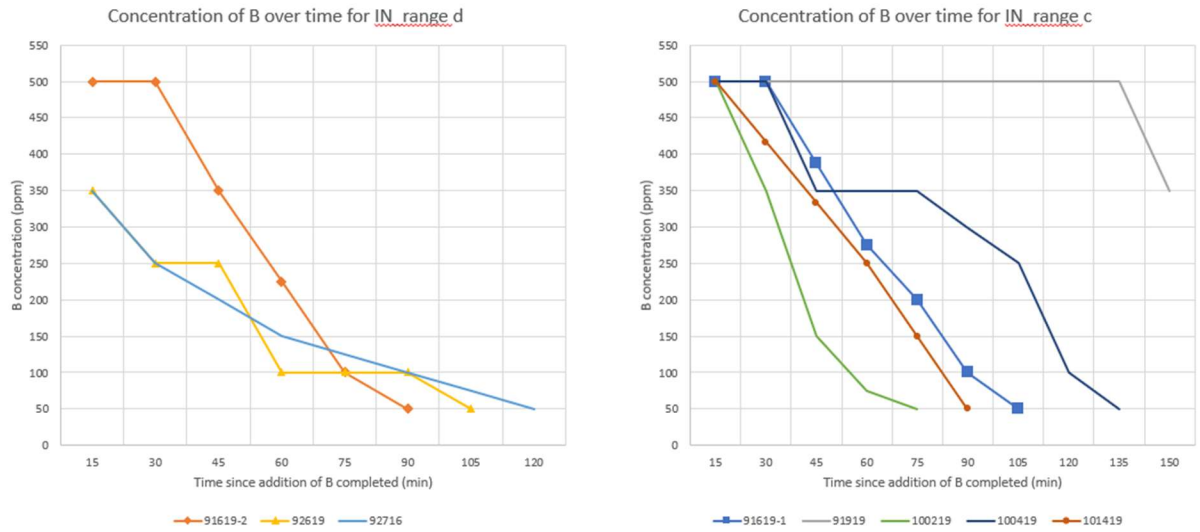

Figure 4.2 Plot of B Concentration over time for batches in IN_range d (left) and c (right)

Due to there only being ten datapoints for each variable, univariate normality could not be determined. ANOVA results were examined based on the standard 95% confidence interval. Again, the dataset was not collected in a balanced design, so the Type I (sequentially calculated) results were used.

The first ANOVA indicated that none of IN_conc, IN_range, or IN_vol had an impact on the B concentration after 60 minutes nor did the interaction between IN_conc and IN_vol or IN_range and IN_vol. On the other hand, the second ANOVA, which used the number of minutes it took for the B concentration to reach 100 ppm as the response variable, also indicated that none of the input variables had a significant impact at the 95% confidence level though IN_range was significant at an 85% confidence level with Pr>F value of 0.1157 and an F value of 4.82. The models being fitted, in order, were:

[B] at 60 min = IN_range + IN_conc + IN_vol and

Min to 100 ppm = IN_Range + IN_conc + IN_vol.

Looking at the LSMEANS and pairwise differentials for IN_range with the time to 100 ppm as the response, it is clear that no two IN_ranges were significantly different from each other. The LSMEAN values for IN_ranges a, c, and d were 233.01, 100.31, and 69.91 minutes, respectively. Recalling Table 3.3, IN_range = a for the highest IN_conc values, IN_range = d for the lowest IN_conc values, and IN_range = c for the second lowest range of IN_conc values. Thus, these results indicate that when A_vol = 42 L/m^3_{ww}, then the lower the IN_conc value the more quickly B will be consumed.

Conclusions

In conclusion, based on the data collected, no distinct rate curve could be determined though most batches reach 50ppm B within 135 minutes. ANOVA results indicated that IN_Range may have an impact on the rate of B consumption. When looking specifically at how many minutes it took for the B concentration to be reduced to 100 ppm, it was found that, when A_vol = 42 L/m^3_{ww}, batches with lower IN_conc consumed B more quickly than batches with higher IN_conc values. However, these conclusions are based on a limited dataset and that more data should be collected and the analysis rerun before making any recommendations in regards to reactor operation.

Chapter 5 - Conclusions and Future Work

Summary and Conclusions

Three aluminum strip processing machine centers at Novelis, Inc's, Oswego, NY, facility produce wastewater containing the water-soluble form of environmental concern (EC). A batch reactor processes the wastewater stream to remove the EC. After changes were made to the machine centers that reduced the concentration of EC in the wastewater, adjustments were made to the amount of both reactants A and B used in the reactor. This was done for financial and safety reasons.

However, after the adjustments were made an unexpected shift in reactor performance was observed, specifically that batches were taking notably longer to consume the reactants. The number of changes, including the reduction in the volume of reactants used, which occurred around the time of the shift, made it difficult to pinpoint the root cause of the shift. A range of statistical methods was used to investigate the issue.

The first round of analysis was performed as the final project for STAT 730: Multivariate Statistical Methods in the Fall 2018 using data for batches from January 2016 through September 2018. This analysis determined, through both PCA and FA, that seven potentially relevant input variables can be reduced to 3 components. From FA these components can be interpreted as "Inputs", "Frequency", and "Batch Volume". Using both ANOVA and MANOVA it is clear that some of the input variables have a significant impact on RXN. Specifically, Days, IN_vol, A_vol, and IN_conc are significant, which also reflects the factors from the FA.

Based on these results, additional data were collected for batches using different amounts of each reactant over the course of several months. The changes in reactant volumes were implemented in response to shifts in reaction time. Using this expanded dataset, which included information for batches from January 2016 through September 2019, a similar analysis to the one performed in the fall of 2018 was performed. It concluded that neither the frequency nor the size of batches has a significant impact on how long the reaction will take. In agreement with the earlier analysis, the incoming concentration of EC and the volume of reactants used do have an impact. Further examination of these factors, using the LSMEANS, show that the higher the A_vol value (representing the amount of reactant used), the shorter the reaction time.

The rate of consumption of reactant B was examined for ten batches run in the fall of 2019. Though no consistent reaction rate curve could be determined, it was noted that seven out of ten batches reached the desired concentration of B, 50 ppm, within 135 minutes. ANOVA was used to examine the relationship between batch volume, the incoming concentration of EC, and the amount of time each batch took to reach 100ppm B. Batches with lower starting EC concentrations were found to consume B more quickly.

Recommendations and Future Work

Based on these conclusions, it is recommended that the volumes of reactants represented by A_vol=42 L/m^3_{ww} should be used for all batches going forward. A_vol=50 L/m^3_{ww} and A_vol=66 L/m^3_{ww} took less time for the reaction to complete but are undesirable due to the high amounts of reactants used.

While some conclusions could be drawn from the B consumption data, the dataset is small. It is likely that these conclusions would change if more batches were included. Due to this, collection

of additional data is recommended, at least another 10 batches; then the analysis should be repeated, with the addition of a response surface model, before making any recommendations in regards to reactor operation.

Chapter 6 - References

American Society for Quality. (2005). *Control Chart- Statistical Process Control Charts | ASQ*. Retrieved from ASQ.org: https://asq.org/quality-resources/control-chart

Ford, C. (2015, August 26). *Understanding Q-Q Plots*. Retrieved from University of Virginia Library Research Data Services + Sciences: https://data.library.virginia.edu/understanding-q-q-plots/

Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis.* Pearson Prentice Hall.

Mendenhall, W., & Sincich, T. (2012). *A Second Course in Statistics: Regression Analysis, Seventh Edition.* Boston: Prentice Hall.

Rahn, M. (n.d.). *Factor Analysis: A Short Introduction, Part 2–Rotations*. Retrieved from The Analysis Factor: https://www.theanalysisfactor.com/rotations-factor-analysis/