DOE-FIU SCIENCE & TECHNOLOGY WORKFORCE DEVELOPMENT PROGRAM

STUDENT SUMMER INTERNSHIP TECHNICAL REPORT-SUMMER 2009

June 22, 2009 to August 22, 2009

Wiped Film Evaporator Pilot- Scale Experimental Design Data Analysis

Principal Investigators:

Duriem Calderin (DOE Fellow)
Florida International University

Robert A. Wilson, MS, Mentor
Columbia Energy and Environmental Services, Inc.

Florida International University Collaborators:
Leonel Lagos Ph.D., PMP®

Prepared for:

U.S. Department of Energy
Office of Environmental Management

Under Grant No. DE-FG01-05EW07033

ABSTRACT

This report analyzes data collected from a previous pilot-scale experiment of a Wiped Film Evaporator (WFE) used to reduce the amount of liquid from three simulant wastes with compositions representative of the Hanford site tanks. The three simulants were analyzed, and three parameters identified as critical in the process of operation: feed flow rate, temperature at the inlet jacket, and vacuum pressure. The direct influence of these parameters on the specific gravity (response) was analyzed. The data was studied by applying a linear regression model to the observations. The equation of the model was obtained by using the least square algorithm in order to minimize the margin of error of the observed value and the predicted value. The difference in the results obtained in two consecutive sets of data lead us to suspect a delay in the recording of the response, influenced by the test configuration and the associated recording method. A new location for recording the response was proposed. The factors were set to implement a region, or range of variation, of the parameters (vacuum pressure, temperature and feed flow) useful to developing a response surface methodology (RSM). The RSM will be used in future experiments for detecting a region that will converge all responses of the system. The variables or factors were arranged for a 2^3 (three factor at two levels) central composite design (CCD) with six axial points and five replications at the center point.

TABLE OF CONTENTS

ABSTRACT iii
TABLE OF CONTENTS iv
LIST OF FIGURES v
LIST OF TABLES vi
1. BACKGROUND 1
2. INTRODUCTION 2
3. EXECUTIVE SUMMARY 3
4. RESEARCH DESCRIPTIONS 4
5. DATA ANALYSIS 5
5.1 Simulant 1 Data Analysis 5
5.2 Simulant 2 Data Analysis 8
5.3 Simulant 3 Data Analysis 11
5.3 First Set of Data for Simulant 3 11
5.4 Second Set of Data for Simulant 3 15
6. LINEAR REGRESSION MODEL 18
6.1 Regression Model Diagnostic 22
6.2 Test of Regression coefficients 25
6.3 Linear Regression Model & Data Analysis 26
7. CONCLUSIONS 29
8. RECOMMENDATIONS 31
10. REFERENCES 35

LIST OF FIGURES

Figure 1. Schematic representation of the WFE process. 2

Figure 2. Simulant 1 graph of pressure versus time of study. 5

Figure 3 Simulant 1 graph of feed flow versus time. 6

Figure 4. Simulant 1 graph of temperature versus time. 6

Figure 5. Simulant 1 graph of specific gravity versus time. 7

Figure 6. Simulant 2 graph of vacuum pressure versus time. 9

Figure 7. Simulant 2 graph of feed flow versus time 9

Figure 8. Simulant 2 graph of temperature versus time. 10

Figure 9. Simulant 2 graph of specific gravity versus time. 10

Figure 10. Simulant 3 first data set graph of pressure versus time. 12

Figure 11. Simulant 3 first data set graph of feed flow versus time. 12

Figure 12. Simulant 3 first data set graph of temperature versus time. 13

Figure 13. Simulant 3 first data set graph of specific gravity versus time. 13

Figure 14. Simulant 3 second data set: graph of vacuum pressure versus time. 15

Figure 15. Simulant 3 second data set: graph of feed flow versus time. 15

Figure 16. Simulant 3 second data set: graph of temperature versus time. 16

Figure 17. Simulant 3 second data set: graph of specific gravity versus time. 16

Figure 18. Normal probability plot of residual versus fitted value. 23

Figure 19. Residual versus fitted value. 24

Figure 20. Residual versus Temperature. 24

Figure 21. Optimization process using RSM. 31

Figure 22. Central Composite design for a 2^3 model. 32

LIST OF TABLES

Table 1. Summary of Parameters for Simulant 1 8

Table 2. Summary of Data for Simulant 2 11

Table 3. Summary of the First Dataset for Simulant 3 14

Table 4. Summary of the Second Dataset for Simulant 3 17

Table 5. Matrix Form of the Factors 19

Table 6. Summary of ANOVA Calculations 21

Table 7. Summary of Results from the Regression Model 22

Table 8. ANOVA Summary Results - Set 1 27

Table 9. ANOVA Summary Results - Set 2 27

Table 10. Nominal Parameters 29

Table 11. Nominal Parameters 29

Table 12. Set of Parameters Selected for the Model 30

Table 13. Response Surface Methodology 33

1. BACKGROUND

Most of the waste generated at the Hanford site is stored in 177 underground tanks [4]. The tanks were built following two designs: the single shell tanks (SSTs) and the double shell tanks (DSTs). The SSTs are 149 in number [4] and were made of a single layer of stainless steel reinforced with concrete and covered with over 10 ft of soil. The DSTs are 28 in number [4] and have two steel liners separated by a space called the annulus; the annulus provides a margin of safety in case of leakage. With the DSTs, the leak can be detected and the waste removed before it seeps into the underlying soil. As with the SSTs, DSTs are reinforced with concrete and covered with over 10 ft of soil.

The total amount of waste stored at the Hanford's tanks is approximately 56 million gallons [4]. This volume is subject to change due to water evaporation, waste transfer between tanks, waste discharge from laboratories and cleanup production facilities and pipeline flushes. It is necessary to add flush water to prevent line plugging. Consequently, in 1994, 66,000 gallons of water were added to the DSTs by flushed pipelines [4].

The 149 SSTs are aging and currently hold approximately 30 million gallons of mixed waste [5]. Additionally, the Hanford site has 28 newer double-shell tanks (DST) with a total capacity of 31 million gallons [4]. The DSTs currently hold approximately 26 million gallons of waste [5]. Waste stored in SSTs is systematically being retrieved and transferred into the DSTs, but, due to the volume of liquid required to retrieve the SST waste, the available DST space is projected to be exhausted in 2014. The Hanford site currently uses the 242-A Evaporator to reduce liquids in the DSTs and create additional space to support SST waste retrieval. However, because of limitations placed on waste transfers to and from the 242-A Evaporator, the additional DST space created by the 242-A Evaporator campaigns is a fraction of the potential space savings [5]. Columbia Energy has proposed the use of a Wiped Film Evaporator (WFE) to help with the process of reducing the amount of liquid in the waste.

The process of reducing the liquid consists of passing the waste throughout a WFE. The system operates at high vacuum pressure with the goal of reducing the boiling point of the water within the waste. A rotor centrifuges the waste along heated walls in the evaporator. The heat causes the volatile components in the solution to turn into vapor and separate from the heavier, more concentrated components.

2. INTRODUCTION

A schematic view of the WFE process is shown in Figure 1, where the green and red arrows represent the direction of the concentrate and the vapor, respectively. Once the vapor leaves the WFE, it goes to a demister. The demister is a type of filter that retains large drops of water and particulate matter within the vapor stream, allowing a drier and less contaminated vapor stream to pass. The vapor passes through the demister and goes into the condenser, where the majority of water in the vapor stream is condensed. The non-condensable gas continues to the vacuum pump where it is discharged to additional off-gas treatment equipment. The condensate is collected and transported by the condensate pump to temporary storage. The condensate is then sampled and sent to the laboratory for analysis of contaminants.

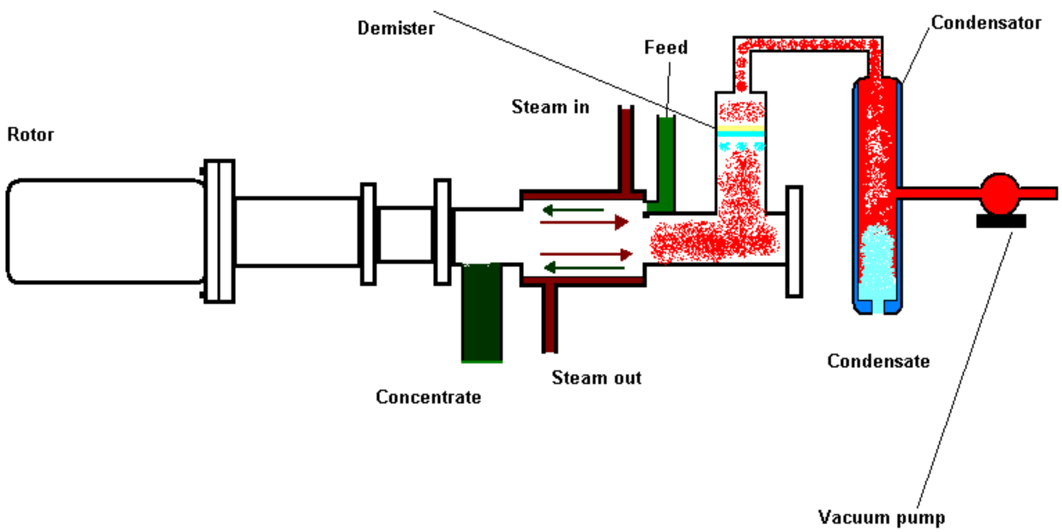

Figure 1. Schematic representation of the WFE process.

The condensate quantity and quality are affected by three major parameters:

1. Feed flow rate
2. Temperature
3. Vacuum pressure

These parameters need to be optimized in order to obtain a high quality and quantity of the condensate and concentrate (measured by the specific gravity).

3. EXECUTIVE SUMMARY

This research has been supported by the DOE-FIU Science and Technology Workforce Development Initiative Program. During the summer of 2009, a Florida International University (FIU) DOE Fellow (Duriem Calderin) spent 10 weeks performing a summer internship at Columbia Energy and Environmental Services Inc, under the supervision and guidance of Robert A.Wilson. This internship was coordinated by Dr. Leonel Lagos. The intern's project was initiated on June 22nd, 2009 and continued through August 22nd 2009, with the objective of research in new waste treatment and disposal technologies applied to reduce the volume of radioactive waste in the Single Shell Tanks (SSTs) and Double Shell Tanks (DST) present at Hanford.

4. RESEARCH DESCRIPTIONS

Three simulants were used to represent the Hanford tank waste. Simulant 1 represents a mixed waste from DST 241-AN-105; Simulant 2 represents a mixed waste from DST 214-AN-107; and Simulant 3 represents a composite of dissolved saltcake waste from S and U tank farm SSTs. Simulant 1, 2 and 3 were prepared in accordance with *Hanford waste Simulants Created to support the Research and Development on the River Protection Projection Waste Treatment Plant* [1]. In the Simulant 1 preparation process, the concentration of cesium and iodine was increased to 160 mg/l and 50 mg/l [2], respectively. The concentration of cesium and iodine in Simulant 2 was increased to 186 mg/l and 50 mg/l, respectively. Finally, in Simulant 3, the concentration of cesium and iodine was increased to 100 mg/l in and 50 mg/l, respectively.

Experiments on a WFE with Simulants 1, 2 and 3 provided data that can be used to set a start point in our test plan and to establish the range of parameters used in the statistical model. The selected factors suspected to affect the flow rates of condensate and concentrate quality are: feed flow rate, vacuum pressure and temperature. The temperature was defined as:

$$\Delta T = \frac{\Delta T_2 - \Delta T_1}{ln(\frac{\Delta T_2}{\Delta T_1})} \quad 4.1$$

Where:

ΔT_2= Temperature of the WFE oil jacket inlet minus temperature of the WFE film

ΔT_1=Temperature WFE oil jacket outlet minus temperature of the feed after leaving the pre-heater.

The data was recorded at intervals of 1 second, and the time of study was between 2 to 3 hours.

The data was analyzed by graphic interpretation and the mean, standard deviation and variability of each factor was obtained in support of the results. The response was characterized by the slope of the trend line of the graph of specific gravity vs. time. The set of parameters (feed flow, temperature and vacuum pressure) were readily identified once the region in which the slope remained highest was obtained.

A linear regression model was applied to the sections of the data in which the model is predictable, and a test of significance was conducted to evaluate the weight of the factors within the model. The least square algorithm was used to find the regression coefficients and to determine their significance within the linear regression model. Finally, the response surface methodology matrix was populated on the region of maximum response (specific gravity) according to the data previously recorded. A central composite design (CCD) was drafted with the results obtained.

5. DATA ANALYSIS

5.1 Simulant 1 Data Analysis

Simulant 1 data was studied from 12:50:00 pm to 14:57:00 pm. Graphs of vacuum pressure, feed flow rate, temperature and specific gravity over this period are shown in Figures 2 through 5.

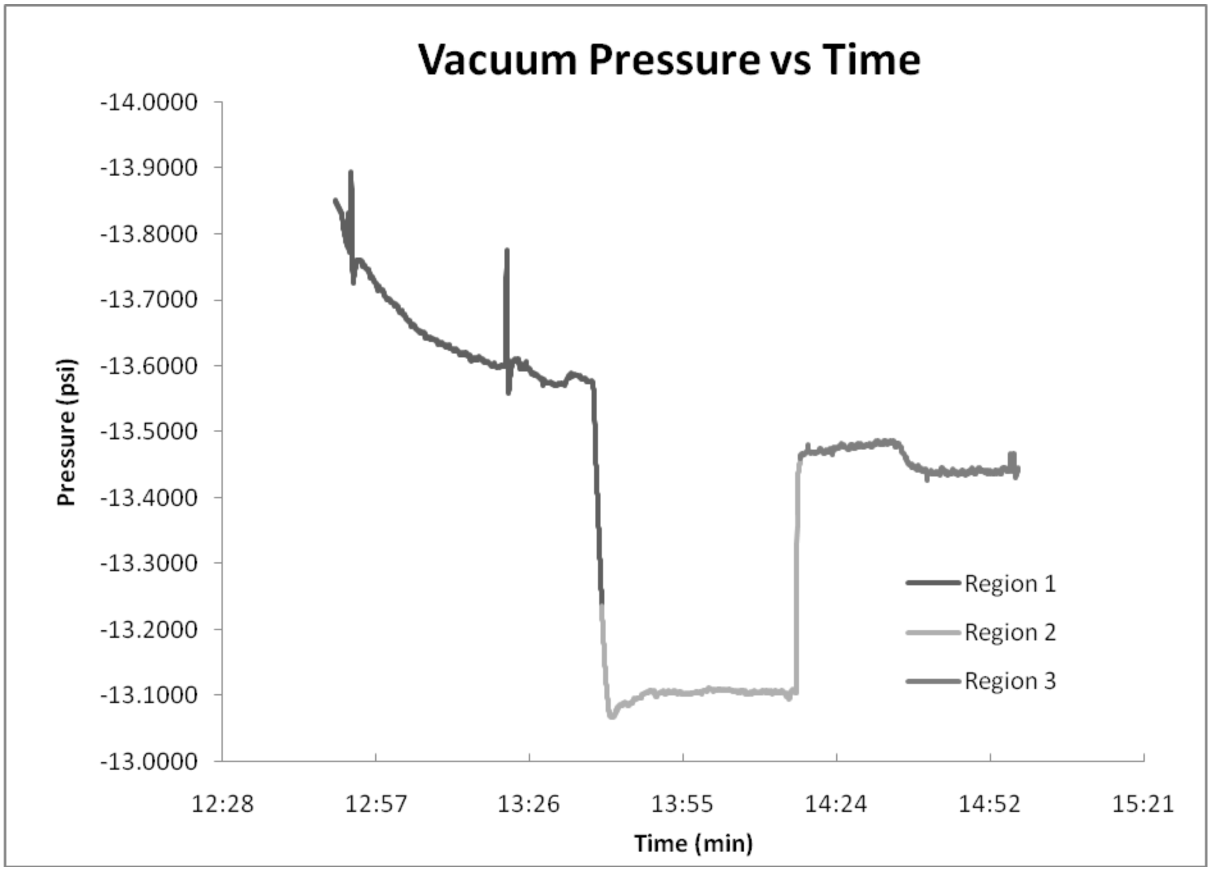

Figure 2. Simulant 1 graph of pressure versus time of study.

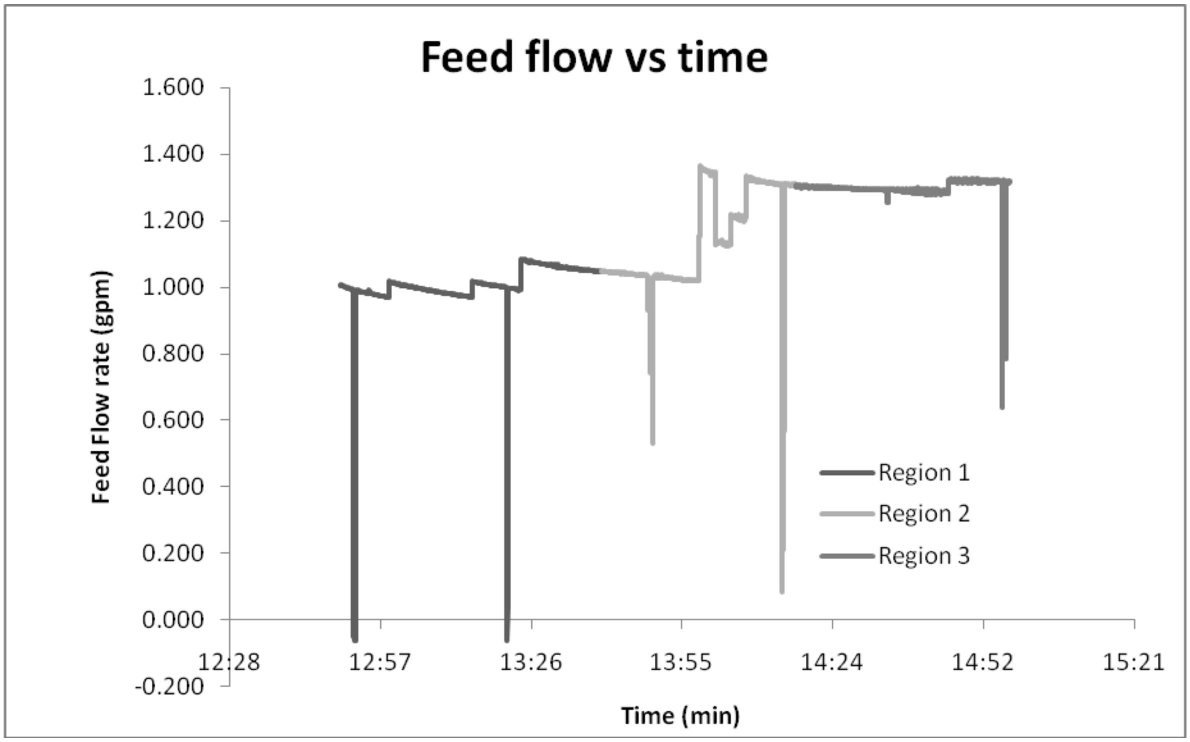

Figure 3 Simulant 1 graph of feed flow versus time.

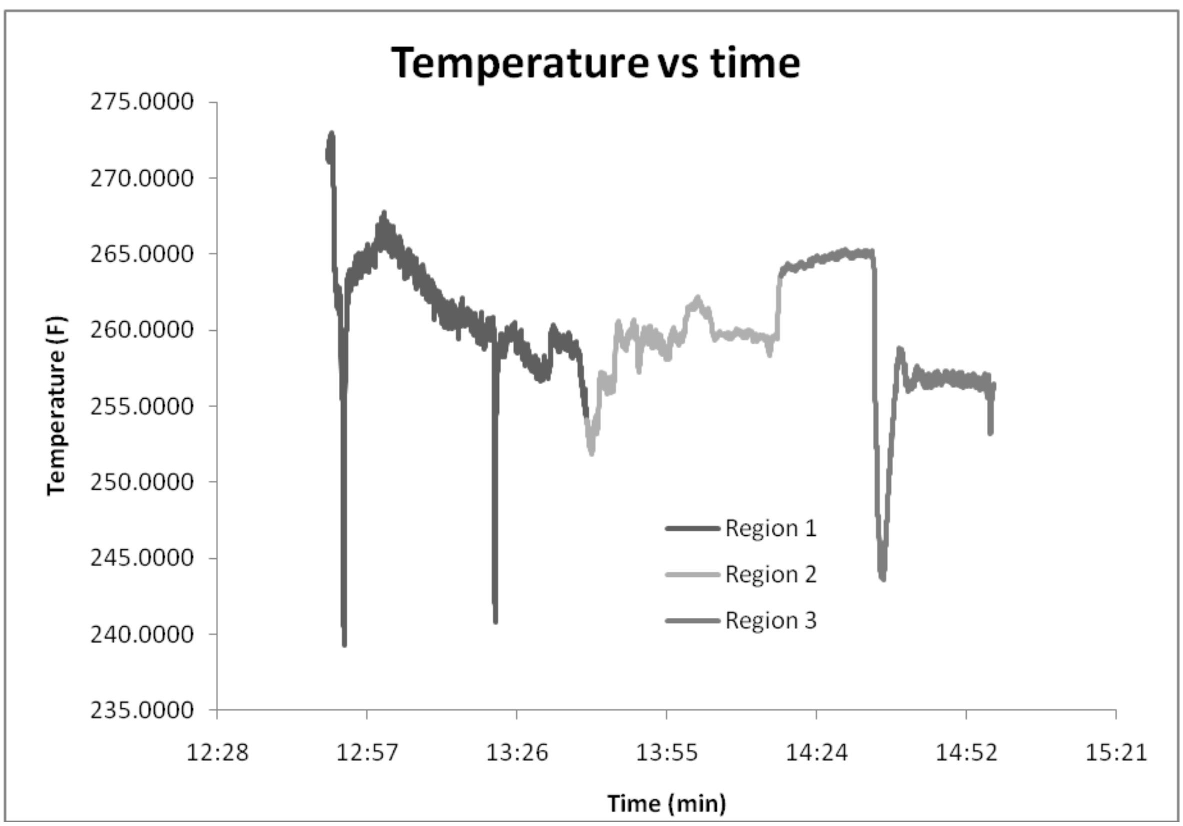

Figure 4. Simulant 1 graph of temperature versus time.

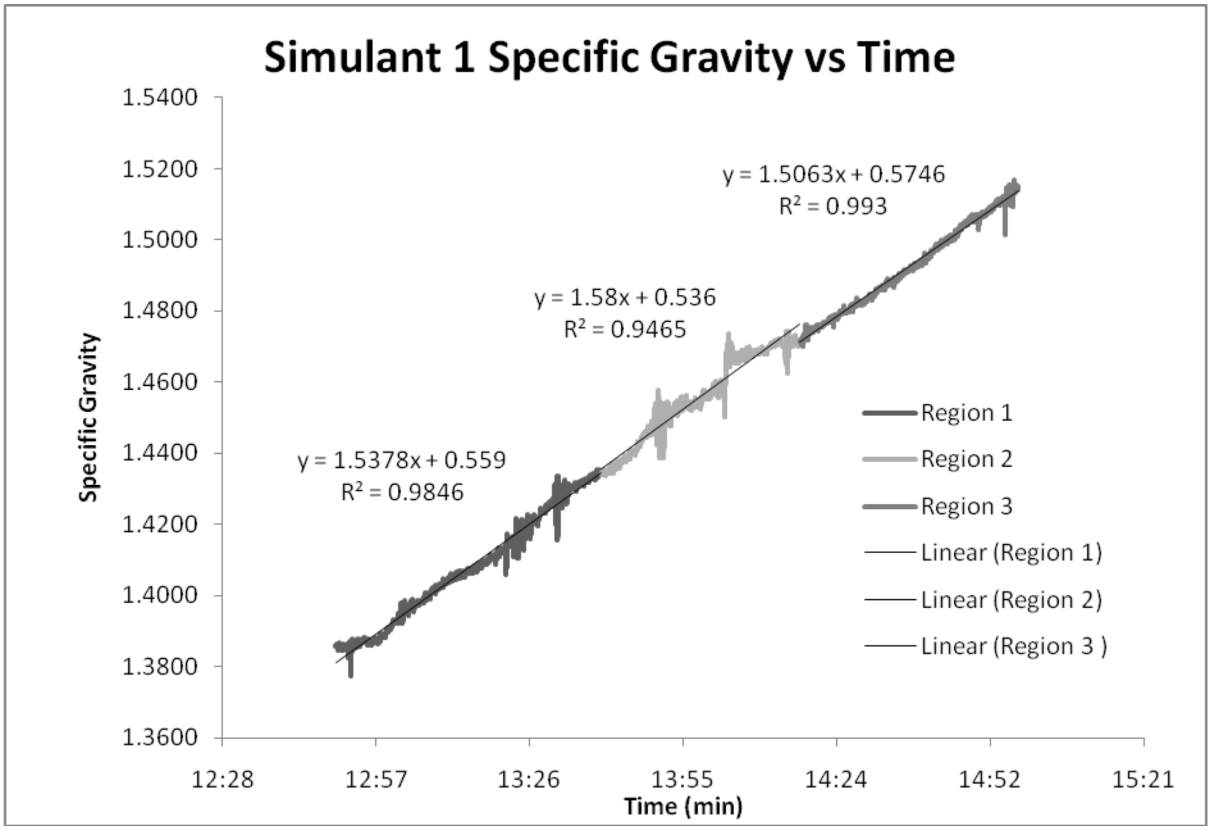

Figure 5. Simulant 1 graph of specific gravity versus time.

The data was analyzed in three regions in order to study how the response (specific gravity) varied with the factor (vacuum pressure, feed flow rate and temperature) changes.

Figure 5 shows that the slope of the trend line is 1.5, observing a slight increase when the vacuum pressure is close to -13.100 psi; overall, however, the slope is consistent over the time period without appreciable changes among regions. The peaks observed in Figure 3 were caused by air within the feed line; note that the peaks are almost instantaneous. In Figure 4, the first two peaks appeared in response to the air in the feed line because when the temperature of the heated oil remains constant and the feed flow diminishes, the temperature of the surface will reach a value closer to the heated oil and the temperature, as expressed in equation 4.1, will decrease. The third peak in Figure 4 is due to a manual decrease of the temperature of the heated oil that enters the WFE. When this happens, the temperature of the surface and the oil jacket get closer, and the temperature, as expressed in equation 4.1, decreases.

Table 1 presents a summary of the data collected from Simulant 1.

Table 1. Summary of Parameters for Simulant 1

Simulant 1			
	Region 1		
	Pressure (psi)	Feed Flow (gpm)	ΔT(F)
Mean	-13.63681481	1.00599	261.29966
Standard Deviation	0.08350	0.09979	3.65526
Variability (%)	*0.612315734*	*9.919958115*	*1.398875*
	Region 2		
	Pressure (psi)	Feed Flow (gpm)	ΔT(F)
Mean	-13.10998767	1.140444313	259.16689
Standard Deviation	0.045336086	0.151882399	1.9890142
Variability (%)	*0.34581334*	*13.31782685*	*0.767465*
	Region 3		
	Pressure (psi)	Feed Flow (gpm)	ΔT(F)
Mean	-13.45816502	1.296579609	259.50707
Standard Deviation	0.017956259	0.057349267	5.2224673
Variability (%)	*0.133422789*	*4.423119595*	*2.012457*
	Simulant 1Full		
	Pressure (psi)	Feed Flow (gpm)	ΔT(F)
Mean	-13.42658184	1.137967439	260.10748
Standard Deviation	0.223840937	0.16269385	3.9991827
Variability (%)	*1.667147597*	*14.29688097*	*1.537512*

The factors show a low variability throughout the study of simulant 1. Nevertheless, the feed flow rate has the largest variability, exceeding 10 percent. At this point, we can conclude that the fluctuation of the vacuum pressure, temperature and feed flow were not significant enough to affect the response. This can be shown by inspecting the slope of the trend line in Figure 5, which was approximately the same for each region.

5.2 Simulant 2 Data Analysis

Simulant 2 data was studied from 15:20:00 pm to 17:04:59 pm. The vacuum pressure, feed flow, temperature and specific gravity graph obtained from the data are shown in Figures 6 through 9.

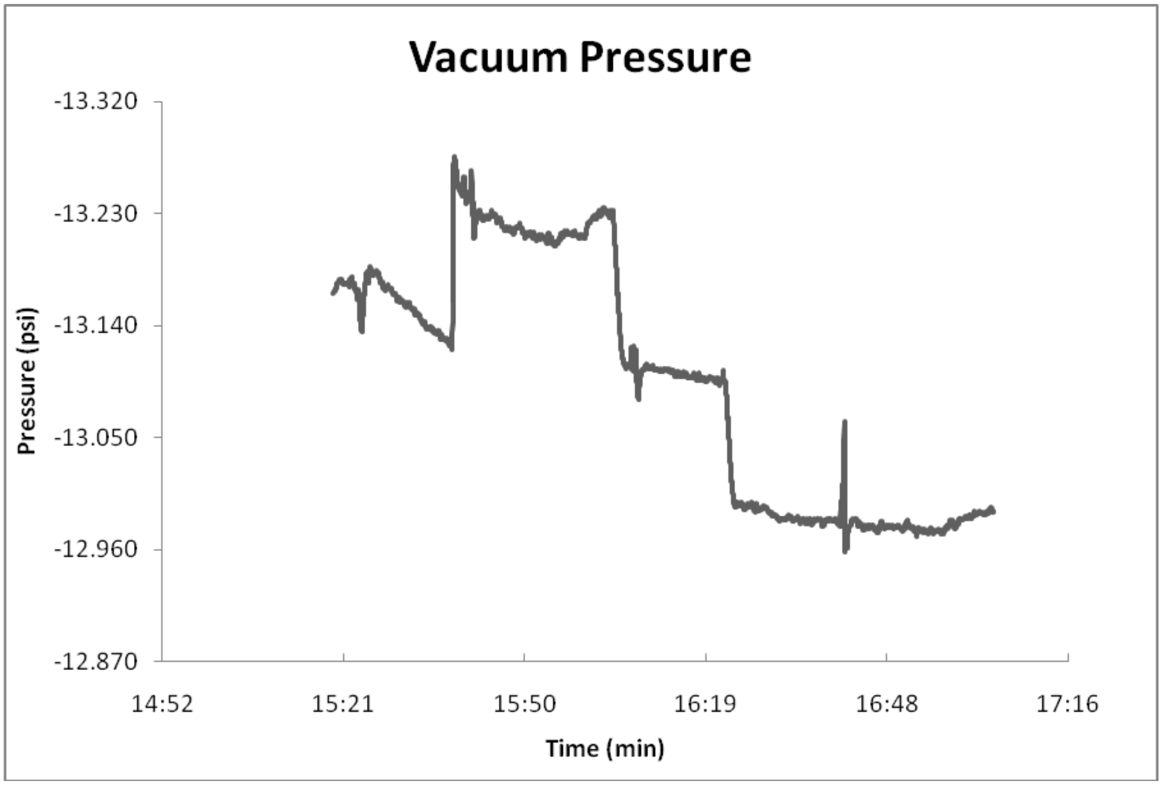

Figure 6. Simulant 2 graph of vacuum pressure versus time.

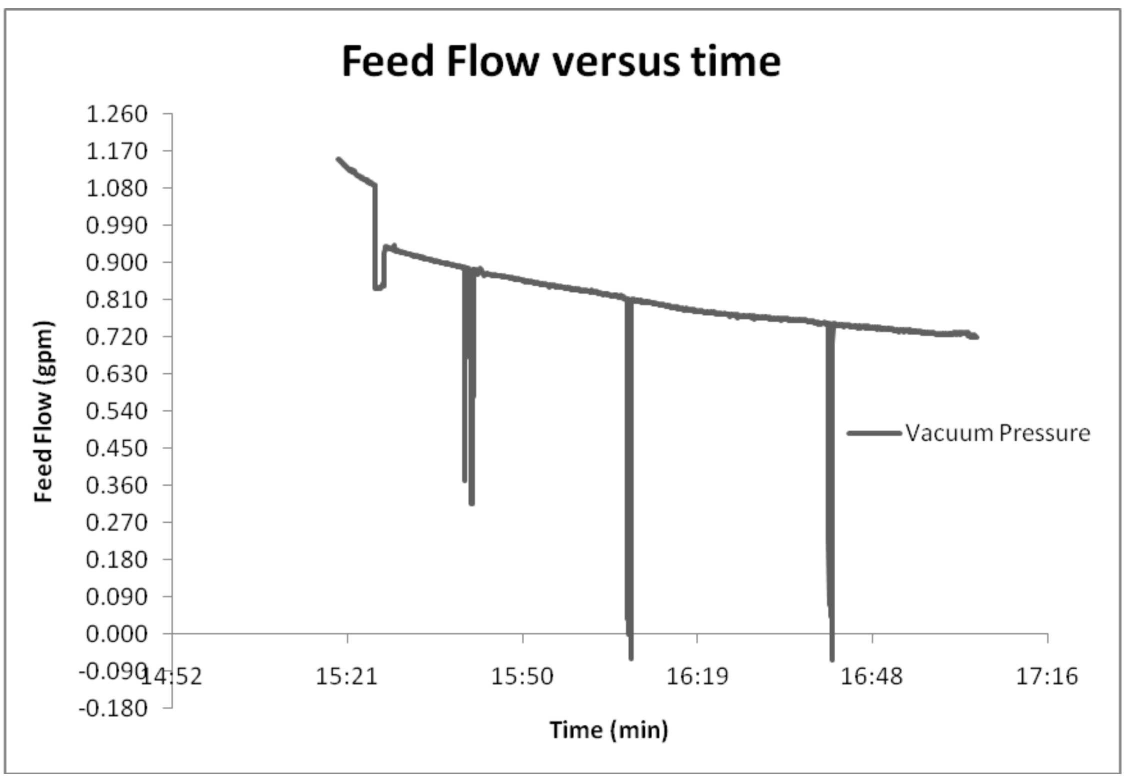

Figure 7. Simulant 2 graph of feed flow versus time

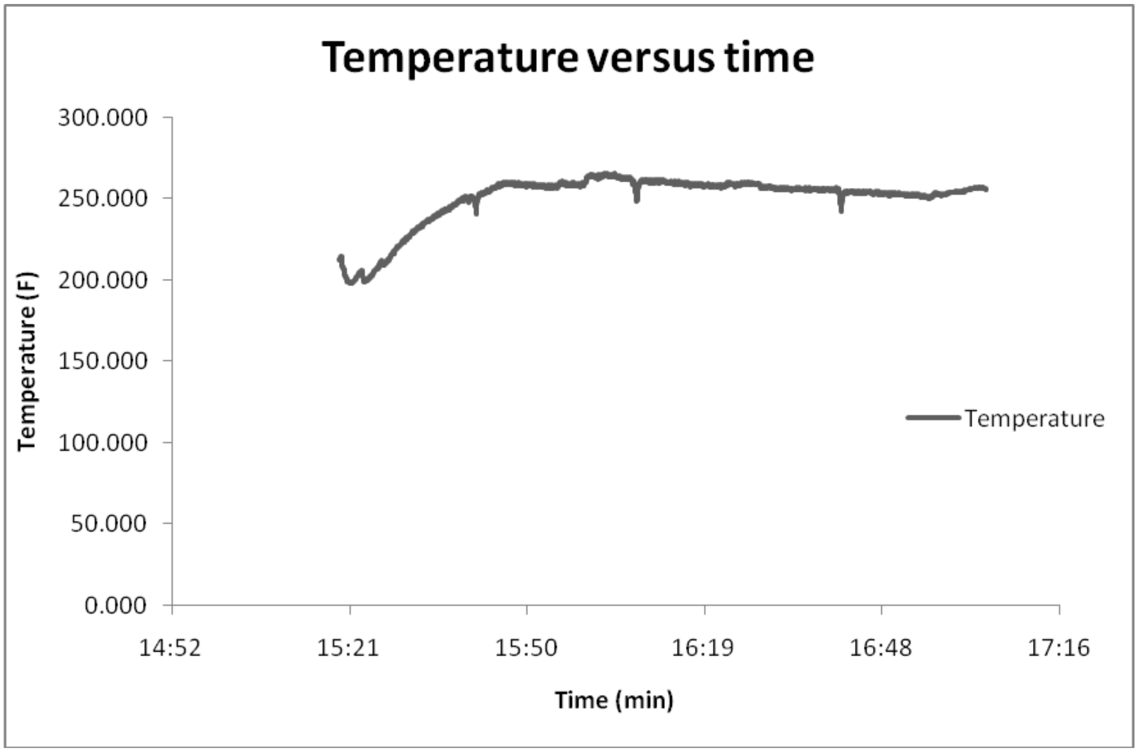

Figure 8. Simulant 2 graph of temperature versus time.

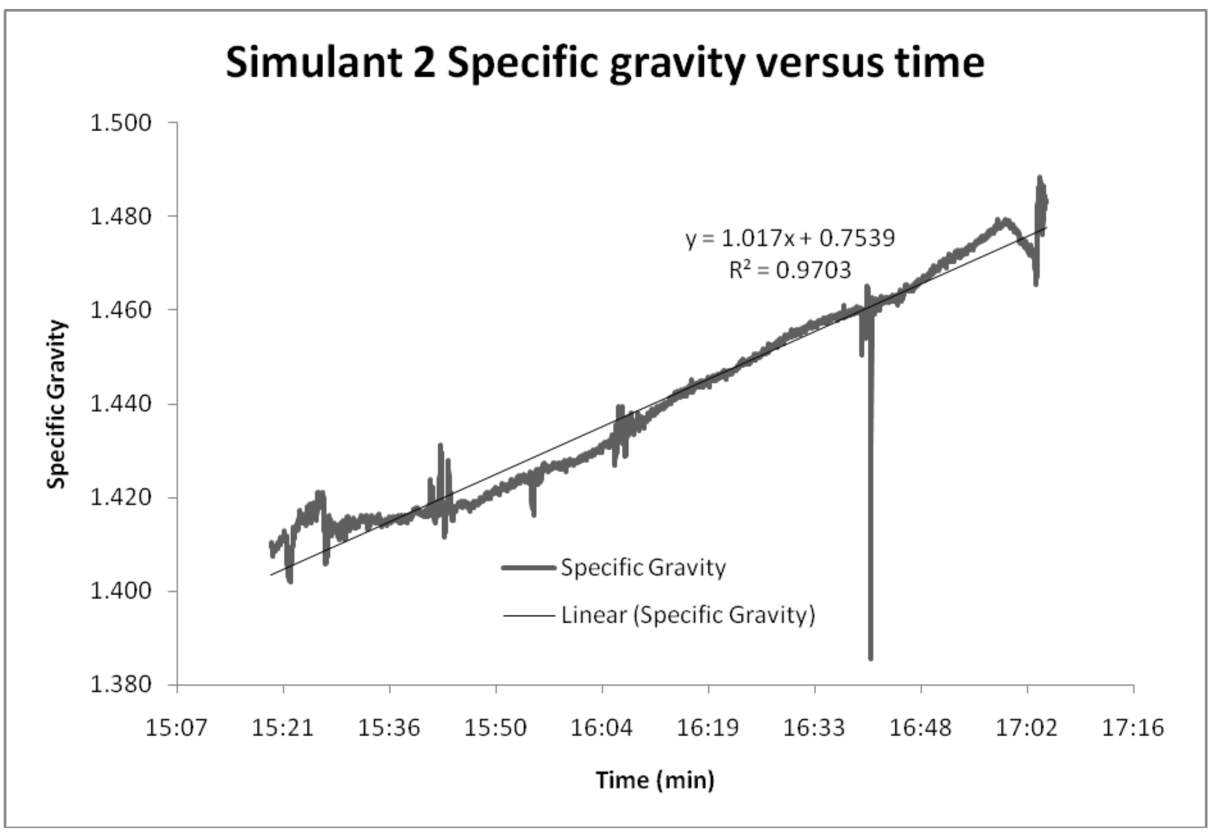

Figure 9. Simulant 2 graph of specific gravity versus time.

The system shows strong peaks in the feed flow rate (Figure 7); therefore, it can be observed that the feed flow has some air in it. Note that every time a peak appears it is reflected on the vacuum pressure (Figure 6), temperature (Figure 8) and specific gravity (Figure 9) graphs.

The vacuum region of operation is between -13.27 and -12.95. At this region, the value of the slope of the trend line of the specific gravity graph (Figure 9), even when positive, is less than the slope of Simulant 1, which means that Simulant 1 will reach the desired specific gravity faster at its mode of operation than Simulant 2. So the factor values for Simulant 1 are still recommended.

The standard deviation, mean and variability of the data are shown in Table 2.

Table 2. Summary of Data for Simulant 2

Simulant 2			
	Pressure (psi)	**Feed Flow (gpm)**	**ΔT(F)**
Mean	-13.09556733	0.81723	250.25161
Standard Deviation	0.09937	0.10991	15.71396
Variability (%)	*0.758843827*	*13.44931073*	*6.279263*

The vacuum pressure was stable throughout the process of study. The variability of the feed flow rate once again exceeded ten percent. These factors have shown the greatest variability among the factors.

5.3 Simulant 3 Data Analysis

Simulant 3 was the largest set of data recorded from 8:20:00 am to 12:30:59 pm at intervals of 1 second. A total of 15,059 discrete data points were taken. Nevertheless, the data was interrupted at 11:26:25 am and started again at 11:28:07 am. The analysis of this Simulant will be divided into two sets of data: the first from 8:20 am to 11:26:25 and the second from 11:28:07 am to 12:30:59 pm.

5.3 First Set of Data for Simulant 3

The first set of data from Simulant 3 is summarized in Figures 10 through 13.

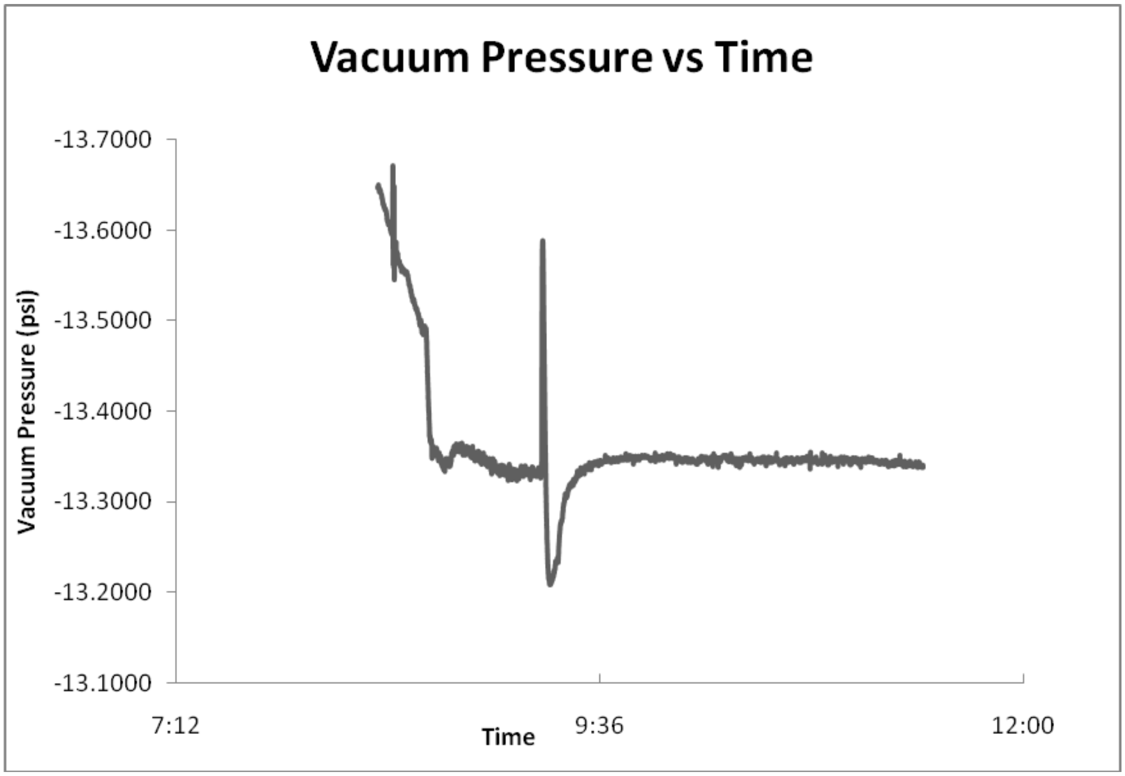

Figure 10. Simulant 3 first data set graph of pressure versus time.

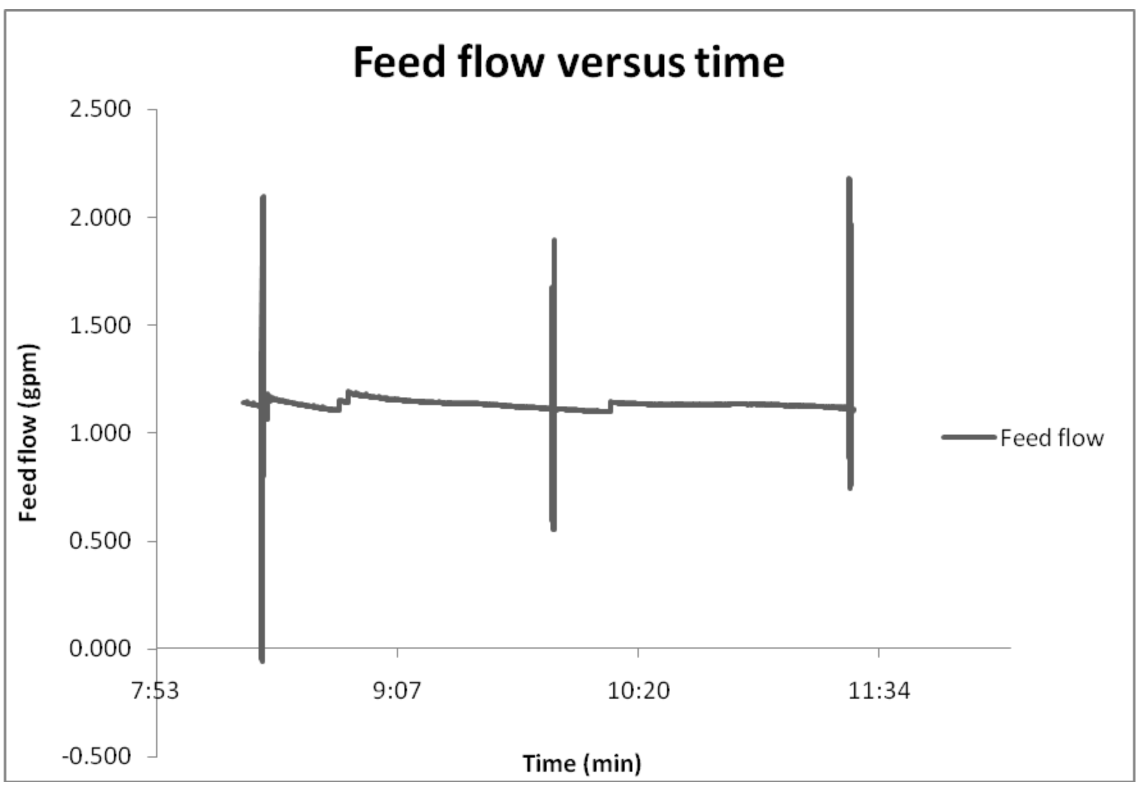

Figure 11. Simulant 3 first data set graph of feed flow versus time.

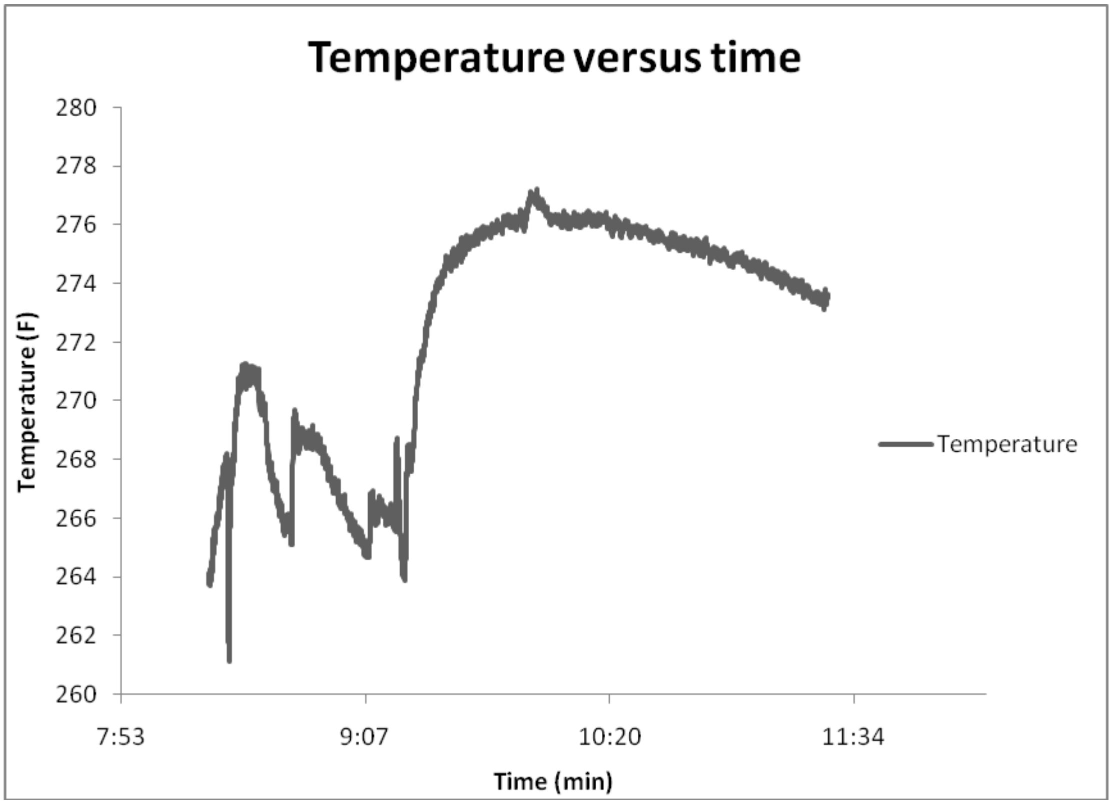

Figure 12. Simulant 3 first data set graph of temperature versus time.

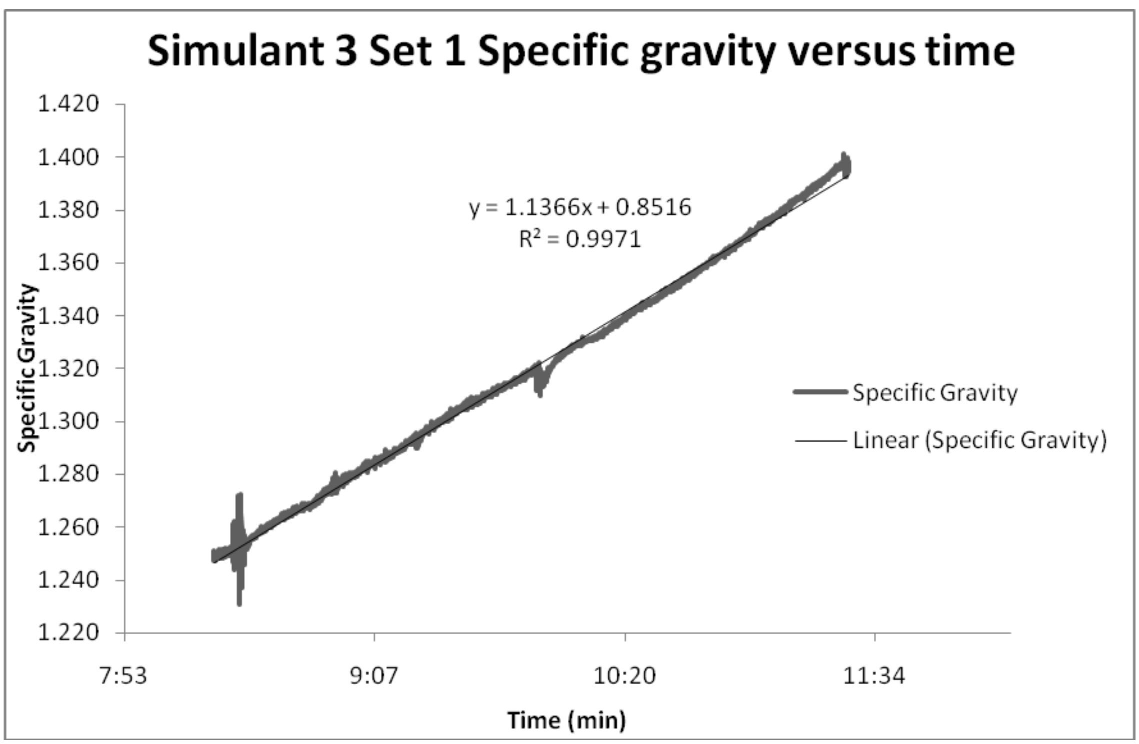

Figure 13. Simulant 3 first data set graph of specific gravity versus time.

Table 3. Summary of the First Dataset for Simulant 3

Simulant 3 First Data Set			
	Pressure (psi)	**Feed Flow (gpm)**	**ΔT(F)**
Mean	-13.36274352	1.13014	272.5818
Standard Deviation	0.07066	0.05740	3.93327
Variability (%)	*0.528754413*	*5.07857065*	*1.44297*

The standard deviation and mean of vacuum pressure, feed flow rate and temperature are shown in Table 3 and summarized below:

- Pressure ($P_v = -13.36 \pm 0.07\ psi$): This means that the vacuum pressure remained constant. The fluctuations observed in the graph were due to the scale of plot used (the variability was very low at 0.5%). Then, we can conclude that the pressure remained stable during the study time. However, special attention should be paid when the feed flow falls abruptly because it causes deep peaks in both of the other factors and in the specific gravity.
- Feed Flow ($F = 1.13 \pm 0.06\ gpm$): The feed flow shows a very low variability. However, the instantaneous peaks have been shown to affect every factor and the response; this behavior could be attributed to some air in the feed flow.
- Temperature ($\Delta T = 272.58 \pm 3.9$): The temperature has a variability of 1.4%, which means it is quite stable.
- The specific gravity under this condition can achieve a maximum or desirable value, but it will take longer to reach this point than the Simulant 1 set of values.

5.4 Second Set of Data for Simulant 3

Figures 14 through 17 present the variations of vacuum pressure, feed flow and temperature over time for the second set of data for Simulant 3.

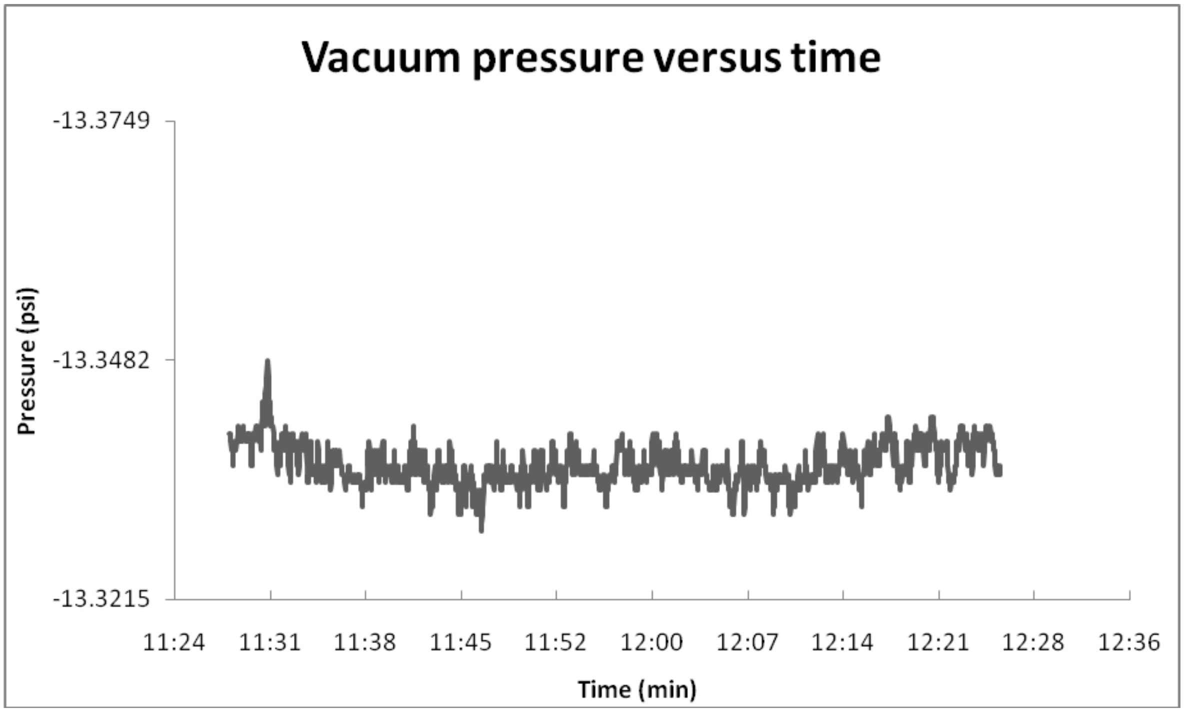

Figure 14. Simulant 3 second data set: graph of vacuum pressure versus time.

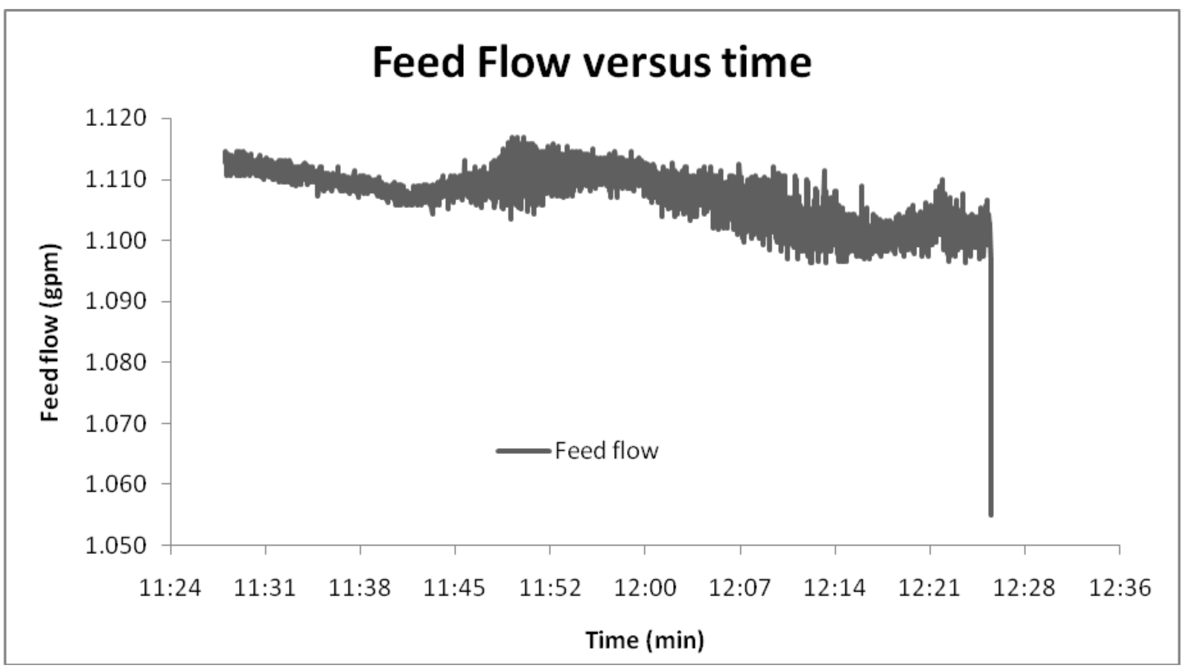

Figure 15. Simulant 3 second data set: graph of feed flow versus time.

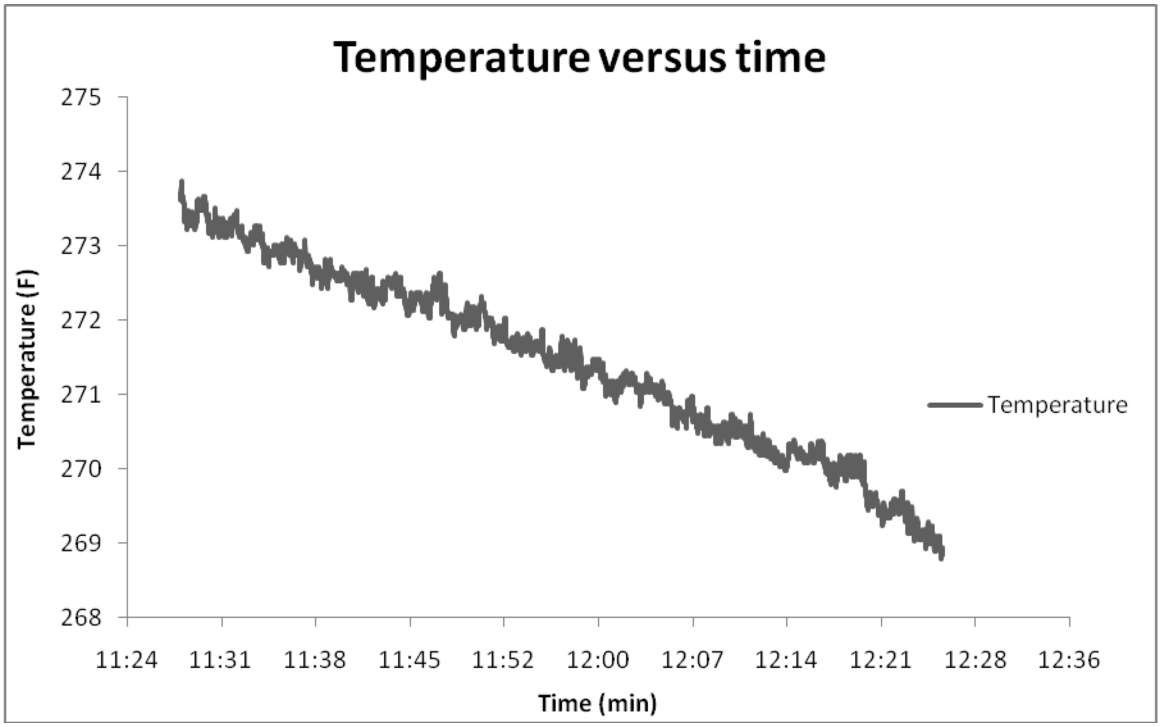

Figure 16. Simulant 3 second data set: graph of temperature versus time.

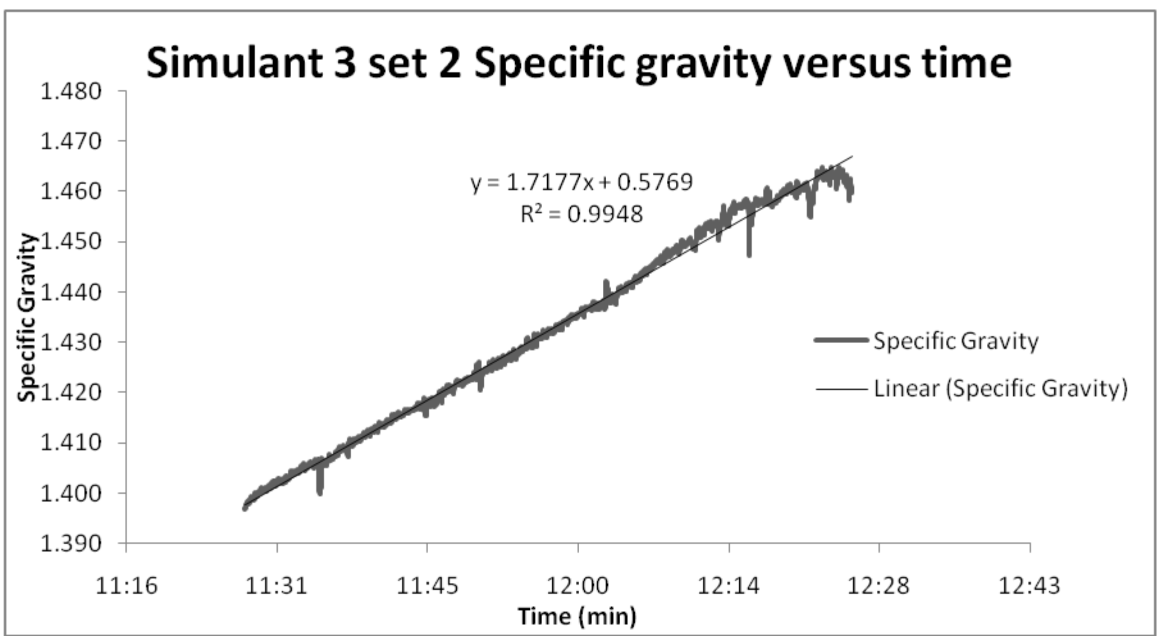

Figure 17. Simulant 3 second data set: graph of specific gravity versus time.

Table 4 is a summary of the mean, standard deviation and variability of the factors:

Table 4. Summary of the Second Dataset for Simulant 3

Simulant 3 Second Data Set			
	Pressure (psi)	**Feed Flow (gpm)**	**ΔT(F)**
Mean	-13.33650037	1.10725	271.43909
Standard Deviation	0.00224	0.00457	1.24059
Variability (%)	*0.016811983*	*0.412906078*	*0.457043*

- Pressure ($P_v = -13.336 \pm 0.002\ psi$): This means that the vacuum pressure remained constant. The variations of the graph observed are due to the scale used to plot the results.
- Feed Flow ($F = 1.1072 \pm 0.005\ gpm$): The feed flow rate remains constant.
- Temperature ($\Delta T = 271.4307 \pm 1.24$): The temperature remains constant.

At this point, a good indicator appears. The slope of the fitted line of the specific gravity has the greatest value of all the specific gravity regressions, meaning that these parameters will lead to high specific gravity values in a shorter period of time.

Based on the results of this second data set for Simulant 3:

1. The set of values (vacuum pressure, feed flow and temperature) gave the greatest slope of the trend line from the specific gravity curve, which means that at these process conditions, higher values of the response can be achieved in less time.
2. The system shows great reproducibility over time as evidence by the low variability of the vacuum pressure (0.017%), feed flow (0.4%) and temperature (0.46%).
3. The values of the factors analyzed show an overall combination of stability.

6. LINEAR REGRESSION MODEL

A linear regression equation can be written as:

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k \qquad 6.1$$

As the coefficients are found empirically, there is an error associated to the solution method, then 6.1 is expressed as:

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k + \epsilon \qquad 6.1a$$

This conclusion was reached by assuming only one expected response (*y*) from the model, but generally the model has several responses in which it is important to find the set of regression coefficients that minimize the error of every response. In this event, 6.1a can be expressed as:

$$y_i = \beta_o + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + \epsilon_i \qquad 6.1b$$

After grouping terms, the error is expressed as:

$$\epsilon_i = y_i - \beta_o - \sum_{k=1}^{n} \beta_k x_{ki} \qquad 6.2$$

The least square (LS) method is meant to minimize the sum of the square of the error, then:

$$L = \sum_{i=1}^{n} \epsilon_i^2 = \epsilon\epsilon' \qquad 6.3$$

Where ϵ is a column vector that represents the error, and it can be shown that the sum square of the error is $\epsilon\epsilon'$ (ϵ' is the transpose of the column vector).

If we express 6.1b in matrix form:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_i \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{1k} \\ \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_o \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_i \end{bmatrix} \qquad 6.4$$

Equation 6.3 expressed in matrix notation is:

$$L = (y - X\beta)(y - X\beta)' \qquad 6.5$$

To find the minimum error, we need to minimize 6.5. First, it is necessary to separate all the terms from 6.5 then differentiate with respect to the regression coefficients.

$$L = yy' - 2(X\beta)'y + (X\beta)'X\beta \qquad 6.6$$

Differentiating 6.6 with respect to βs:

$$\frac{\partial L}{\partial \beta} = -2X'y + 2X'X\beta = 0 \xrightarrow{yields} \beta = (X'X)^{-1}X'y \qquad 6.7$$

As an example, using the mean of 40 consecutive observations 16 times from Simulant 3, the factors would be as shown in Table 5:

Table 5. Matrix Form of the Factors

	X matrix			
Run sequence	***Temperature (°F)***	***Feed flow (gpm)***	***Vacuum Pressure (psi)***	***SpG. (y)***
1	386.6144	1.1390	-13.6509	1.2487
2	387.6385	1.1376	-13.6460	1.2492
3	388.7335	1.1348	-13.6395	1.2495
4	389.8511	1.1324	-13.6294	1.2496
5	390.8651	1.1306	-13.6232	1.2506
6	391.7727	1.1279	-13.6123	1.2504
7	392.7210	1.1258	-13.6075	1.2506
8	393.6780	1.1235	-13.5981	1.2512
9	394.4691	1.0300	-13.5948	1.2528
10	395.3251	1.0611	-13.5756	1.2538
11	395.9204	1.0963	-13.5726	1.2526
12	396.6431	1.1599	-13.5641	1.2526
13	397.2652	1.1589	-13.5583	1.2511
14	397.9279	1.1569	-13.5563	1.2521
15	398.4754	1.1538	-13.5542	1.2533
16	399.0242	1.1514	-13.5503	1.2547

Following equation 6.7:

$X'X =$

16	6296.925	18.0199	-217.533
6296.925	2478441	7091.996	-85609.8
18.0199	7091.996	20.3142	-244.994
-217.533	-85609.8	-244.994	2957.559

$X'y =$

20.0228
7880.223
22.5503
-272.226

Finally:

$$\beta = \begin{bmatrix} 1.1455 \\ 0.0004 \\ -0.0145 \\ 0.0021 \end{bmatrix}$$

Then the linear regression equation takes the form

$$y = 1.1455 + 0.0004x_1 - 0.0145x_2 + 0.0021x_3 \qquad 6.8$$

$$y = 1.1455 + 0.0004Temperature - 0.0145Feed\ Flow + 0.0021Vacuum\ Pressure$$

A test for the significance of regression was conducted to determine if a linear relationship between the regression variables and the response exists. After this test, we set our hypothesis as:

$H_o: \beta_1 = \beta_2 = \cdots = \beta_k = 0$

$H_1: \beta_k \neq 0$

The null hypothesis H_o states that none of the regression variables $x_1, x_2 \dots \dots x_k$ contributes significantly to the model. The alternative hypothesis, H_1, states otherwise.

The analysis of variance (ANOVA) test, which was used to check the validity of H_o, was implemented by:

$$SS_T = SS_R + SS_E \qquad 6.9$$

Where: SS_T is the total sum of squares, SS_R is the regression sum of squares and SS_E is the error sum of squares.

The terms mentioned above were calculated by:

$$SS_E = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}\epsilon_i^2 = \epsilon\epsilon' = y'y - \beta'X'y \qquad 6.10$$

Equation 6.10 was reached following equation (6.6) and simplifying by knowing that $X\beta = y$:

$$SS_T = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} = y'y - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} \qquad 6.11$$

Calculating equation 6.11: $SS_T = 0.000074315$

Finally, combining equations 6.10 and 6.11 into 6.9:

$$SS_R = \beta' X' y - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} \quad 6.12$$

Thus, $SS_R = 0.000041005$ and the sum of squares of the error following equation 6.10 is $SS_E = 0.000006309$.

The mean square of the regression equation and the error was calculated by:

$MS_R = \frac{SS_R}{d} = 0.000013688$ and $MS_E = \frac{SS_E}{de} = 0.000000526$ and $F = \frac{MS_R}{MS_E} \approx 26$

d,d_e: are the degree of freedoms of the regression model and the error, respectively (*d=number of factors= 3* and d_e *= total number of observations- number of factors – 1= 16-3-1=12*).

Table 6 presents a summary of the calculations made above.

Table 6. Summary of ANOVA Calculations

ANOVA table					
Source	Degree of Freedoms	Sum of Squares	Mean Square	F	P
Regression	3	0.000041005	0.000013668	26.000	0
Residual Error	12	0.000006309	0.000000526		
Total	15	0.000047315			

The P-value is extremely low so we can reject H_o and conclude that at least one of the regression variables is significant within the model. Another approach is to look for $F_{\alpha,k,n-k-1} = F_{0.05,3,12} = 3.49$ and then compare $F > F_{0.05,3,12}$, rejecting H_o as well.

6.1 Regression Model Diagnostic

Table 7. Summary of Results from the Regression Model

Run number	Y	y predicted	Residual	St residual	h	D
1	1.248686	1.2485	0.000186	0.4238	0.4208	0.0326
2	1.249172	1.2489	0.000272	0.4857	0.2291	0.0175
3	1.249528	1.2494	0.000128	0.2107	0.1818	0.0025
4	1.24957	1.2498	-0.00023	-0.3547	0.1326	0.0048
5	1.250605	1.2503	0.000305	0.4888	0.1731	0.0125
6	1.250376	1.2507	-0.00032	-0.3982	0.0804	0.0035
7	1.250568	1.2511	-0.00053	-0.7186	0.1603	0.0246
8	1.251194	1.2515	-0.00031	-0.4411	0.116	0.0064
9	1.252799	1.2532	-0.0004	-0.8346	0.6275	0.2933
10	1.253827	1.2531	0.000727	1.4062	0.5353	0.5694
11	1.252583	1.2528	-0.00022	-0.325	0.1975	0.0065
12	1.252639	1.2522	0.000439	0.6375	0.2225	0.0291
13	1.251057	1.2525	-0.00144	-2.1132	0.2509	0.374
14	1.252063	1.2527	-0.00064	-0.9567	0.1851	0.052
15	1.253279	1.253	0.000279	0.4658	0.2235	0.0156
16	1.254694	1.2533	0.001394	2.288	0.2636	0.4684

Table 7 shows a summary of the results obtained from the model where:

y: is the measured response.

y predicted: is the estimated response.

Residual: is the difference between the observed minus the predicted value

($\epsilon_i = y_i - \hat{y}_i$).

St Residual: is the studentized residual, which was calculated by:

$$r_i = \frac{\epsilon_i}{\sqrt{\sigma^2(1-h_{ii})}} \text{ [3]}$$

This parameter was calculated because detects any point with large residual and large h_{ii} value. Those points sometimes are not clearly visible by inspecting the residuals (see the red highlighted values in Table 7).

h: is the diagonal of the hat matrix, which was obtained by $H = X(X'X)^{-1}X'$. The hat matrix relates the fitted values to the observed values [9]. The diagonal elements of the hat matrix are the leverages that describe the influence each observed value has on the fitted value for that same observation [9].

D: is the leverage from each observation obtained from $D_i = \frac{r_i^2}{p} \frac{h_{ii}}{(1-h_{ii})}$ [3]. p is the rank of H and can be easily calculated by summing all the elements of the principal diagonal expressed as $\sum_{i=1}^{n} h_{ii} = rank(H) = p = 4$.

The observations (red highlighted rows in Table 7) show that those points have the largest influence among the other points, but in practice, this is not significant to the experiment. The order of the error is around 10^{-2}, and for a scale industrial process this is quite affordable.

Figures 18 through 20 show the normal probability plot of the residual, residual versus fitted value and residual versus temperature.

The normality plots graph of the residual (Figure 18) shows no departure from the trend, so we can conclude that the residual is normally distributed and there is no correlation between the error and the model. The graph of residual versus fitted value shows no pattern and no correlation with the response (Figure 19). Finally, the residual versus temperature (Figure 20) shows that with the increase in temperature there is an increase in the variability of the response (specific gravity).

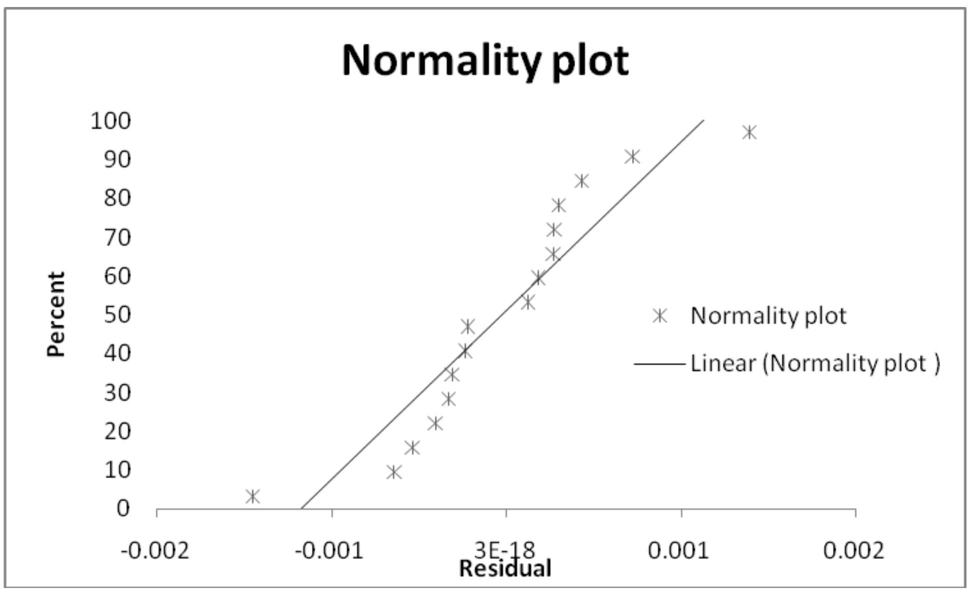

Figure 18. Normal probability plot of residual versus fitted value.

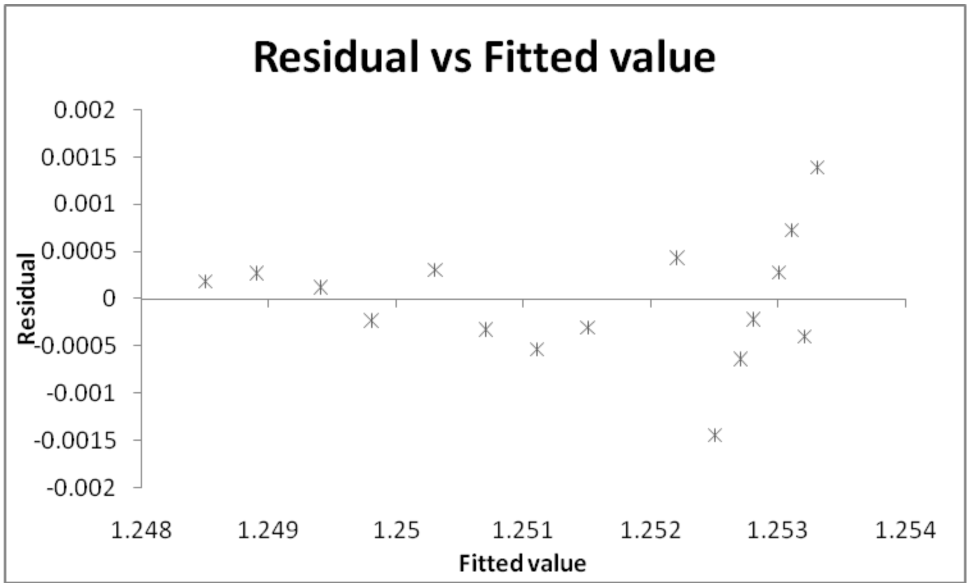

Figure 19. Residual versus fitted value.

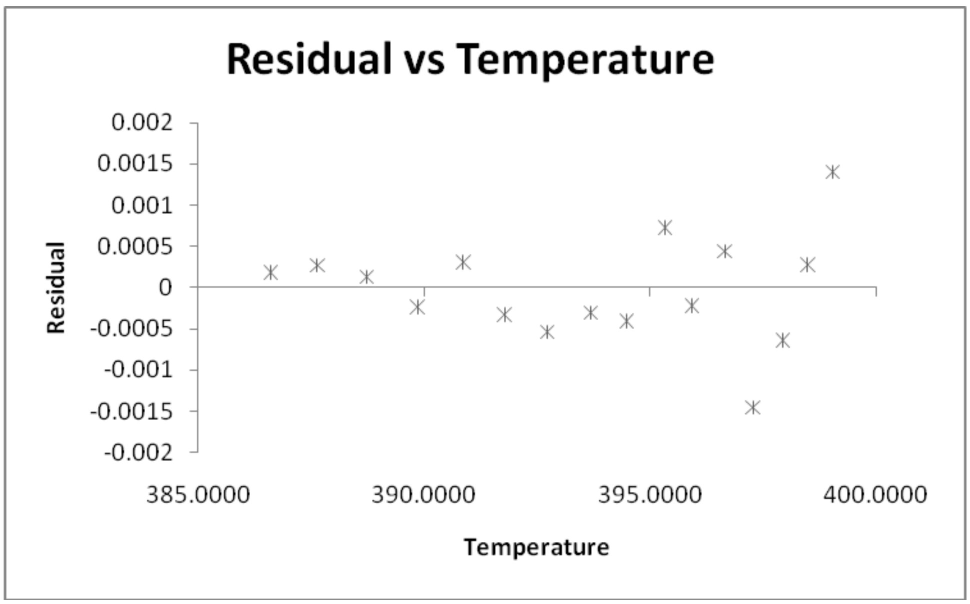

Figure 20. Residual versus Temperature.

Also, the coefficient of multiple determinations was calculated. This coefficient is an indicator of how much variability is explained in the model by fitting the data with the least square algorithm [3]. It can be calculated as:

$$R^2 = 1 - \frac{SS_E}{SS_T} = 0.8724 \tag{6.1.1}$$

It can be shown that this coefficient is biased toward increasing when it increases the number of variables. Then it was also calculated the adjusted coefficient by [3]:

$$R^2_{adj} = 1 - \left(\frac{n-1}{n-p}\right)(1 - R^2) = 0.8404 \quad 6.1.2$$

The decrease of the value of R adjusted is a consequence of terms in the model that do not play a significant role. The values of R^2 and R^2_{adj} are not the optimal expected values because only 84 % of the variability of the data is explained by the regression equation.

A last test was conducted to check the prediction error sum of squares (PRESS). This term is calculated to know how well the model will explain the variability of new observations [3].

$$PRESS = \sum_{i=1}^{n}\left(\frac{e_i}{1-h_{ii}}\right)^2 \text{ and } R^2_{pre} = 1 - \frac{PRESS}{SS_T} = 0.7289$$

This means that the model explains about 72% of the variability in predicting new observations.
By visual inspection, it can be seen that the regression coefficients of the factors have a low significance in the model (value less than 10^{-2} in magnitude), while the feed flow has the greatest significance of the three factors (higher regression coefficient).

6.2 Test of Regression coefficients

We can confirm the statement made above by finding the effect of each coefficient to the model. This can be done by assuming that all others coefficients are already in the model and by setting the hypothesis based on only the evaluated coefficient. Then:

$$H_o: \beta_1 = 0 \text{ and } H_1: \beta_1 \neq 0.$$

H_0 means that the temperature has no effect on the model. Then we can conduct a partial F test to measure the contributions of xj as if they were the last variable added to the model [3]. Then our regression model takes the form:

$$y = X_{23}\beta_{23} + \epsilon \quad 6.2.1$$

Where X_{23} represents the columns of X that are the vacuum pressure and feed flow, respectively. Then the coefficient is calculated by:

$$\beta_{23} = \left(X'_{23}X_{23}\right)^{-1}X'_{23}y \quad 6.2.3$$

The partial F test will be tested by:

$$F = \frac{SS_R(\beta_1|\beta_2\beta_3\beta_o)/r}{MSE} \quad 6.2.4$$

In equation 6.2.4, SS_R signifies the sum of square of β_1 when β_2 and β_3 are already in the model, r is the degree of freedoms and MSE is the mean square of the error of the full system.

Note that $SS_R(\beta_1|\beta_2\beta_3\beta_o) = SS_R - SS_R(\beta_2\beta_3|\beta_0)$ and the second term is the sum of squares calculated by the model expressed in 6.14.

Then the rejection criterion is $F > F_{\alpha,r,n-p} \xrightarrow{yields} 0.87 < F_{0.05,2,16-4} = 3.89$ failing to reject the null hypothesis.

Following the same algorithm for vacuum pressure and feed flow, $F = 0.06$ and $= 6.1247$, respectively, concluding the feed flow was the only significant factor to the model.

Having concluded the significance of the factors within the model, we can rewrite 6.8 as:

$$y = 1.1455 - 0.0145x_2 \qquad 6.2.5$$

The results were still not convincing because the vacuum pressure and temperature variability 0.25 and 1%, respectively, were much lower than the feed flow variability (15%). This led to the hypothesis that the response was sensitive to the change in feed flow at that variability level, but not to the two other factors. A valid statement at this point is that: variability less than 1% in temperature and 0.25% in pressure are not significant to the response when the feed flow varies by 15%.

A second conclusion is that the low values of R^2_{pre}, R^2 and R^2_{adj} means that the model is not completely trustworthy. Usually, a good model can predict at least 90% of the variability in predicting a new observation (R^2_{pre}) and more than 90% by fitting the least square algorithm (R^2_{adj}) to the data.

6.3 Linear Regression Model & Data Analysis

An additional two sets of consecutive data with similar variability were analyzed with a total of 969 points each.

For the first and second data sets the regression equations were:

$$y_1 = 2.58 - 0.0003x_{11} - 0.0001x_{21} - 1.03x_{31} \qquad 6.3.1$$

$$y_2 = -1.3836 + 0.0007x_{12} - 0.2138x_{22} - 0.39x_{32} \qquad 6.3.2$$

x_{11}, x_{21}, x_{31} , y_1 ,x_{12}, x_{22}, x_{32} and y_2, represent the temperature, vacuum pressure, feed flow and fitted value of the first and second data, respectively.

For the first set, $F_1 = 16.98$, $F_2 = 0.1421$, and $F_3 = 9896$, leading to the conclusion that the feed flow and temperature are the most significant factors within the model; however, for the second set of data, $F_{12} = 32.26$, and $F_{22} = 1926.8$ and $F_{32} = 1432.2,$, leading to the conclusion that each of the three factors were significant to the model.

Tables 8 and 9 summarize the results of both regression models.

Table 8. ANOVA Summary Results - Set 1

Analysis of Variance Set 1					
Source	**DF**	**SS**	**MS**	**F**	**P**
Regression	3	0.0147769	0.0049256	4363.42	0
Residual Error	965	0.0010893	0.0000011		
Total	968	0.0158662			
	R=0.94	***R-adj=0.90***	***R-Press=0.99***	**Significance**	
Regression Coefficients	**Value**	**Partial F test**		**Yes**	**No**
Temperature	0.0003	16.98		**X**	
Pressure Vacuum	-0.0001	0.1421			**X**
Feed Flow	-1.03	9896		**X**	

Table 9. ANOVA Summary Results - Set 2

Analysis of Variance Set 2					
Source	**DF**	**SS**	**MS**	**F**	**P**
Regression	3	0.0105324	0.0035108	6325.74	0
Residual Error	965	0.0005356	0.0000006		
Total	968	0.011068			
	R=0.94	***R-adj=0.92***	***R-Press=0.99***	**Significance**	
Regression Coefficients	**Value**	**Partial F test**		**Yes**	**No**
Temperature	0.0007	32.26		**X**	
Pressure Vacuum	-0.2138	1926.8		**X**	
Feed Flow	-0.3876	1432.2		**X**	

Both regression models concluded that at least one of the factors within each model is significant. The issue is that the vacuum pressure is not significant and is then confirmed as significant. At first, this seems contradictory, but to explain this, it is necessary to go back to the technical design of the process. The specific gravity (response) is measured with the same instrument as the feed flow, and a change in the factors will not be detected in the response according to the moment the change occurs. The process is a closed cycle in which the concentrate goes back to the feed tank and then again into the WFE. Consequently, changes in specific gravity take time to detect, especially if the feed tank is full at the beginning of the experiment. As a result, we proposed to add a second coriolis-meter between the Wiped Film Evaporator and the feed-tank with the objective of detecting the variations of the factors at the moment it occurs.

The other concern is that the specific gravity will increment with time even when the parameters remain constant, making our response conditional to the time of study. We could solve this issue by taking, as a response, the slope of the trend line of the graph of specific gravity versus time; the higher the slope means less time arriving to our goals of desired specific gravity, and it also means optimization of the process. The same amount of time to study each response needs to be set. If in all Simulants, the study time for each Simulant is approximately 5 hours, then 0.25 hours (5 hours divided by the numbers of responses 20) per response can be accepted as a good range for the study time.

7. CONCLUSIONS

After analyzing the specific gravity from all the Simulants of the pilot test, the one with the best results (expressed in terms of high specific gravity in less time) is the dataset for Simulant 3 (second dataset), which shows the highest increment of this response (Sp.G) over time. Based on analyzed values, we propose choosing this set of parameters as a nominal standard value to create our low and high level for the statistical model.

Table 10 summarizes the expected nominal parameters of the statistical model:

Table 10. Nominal Parameters

Nominal parameters		
Pressure (psi)	**Feed Flow (gpm)**	**ΔT (F)**
-13.34	1.10	271.43

However, temperature is a factor we would like to reduce because it will mean a high heat transfer between the surface and hot oil. Therefore, we can change the temperature value to the one obtained with Simulant 2, ΔT= 250.25 ºF. The final nominal values are provided in Table 11.

Table 11. Nominal Parameters

Nominal parameters		
P(psi)	**FF(gpm)**	**ΔT(F)**
-13.34	1.10	250.25

There are some concerns about the results:

1. The small range of variation of the factors is an issue because the measurement and reading equipment should be precise enough to measure the difference in vacuum pressure (0.1 psi), feed flow (0.1 gpm) and temperature ($1 \leq \Delta T \leq 5$) in order to collect the data.
2. The specific gravity is a response that varies over time; therefore, it is necessary to find another response to record for our model. The slopes of the graph of specific gravity versus time provide a hint of a usable trending parameter. If the slope is positive and its value is high, it means that, with the passage of time, the specific gravity increases faster than a positive slope with a lesser value.
3. ΔT expressed as 4.1 cannot be used as a factor because it varies over time. Even when the temperature of the oil that enters into the WFE remains constant, ΔT decreases over time. Then the factor that we could control is the temperature of the oil that enters into the WFE.

Further research into the measurements leads to the following recommendations for the equipment:

1. Additional proportional control valve needs to be added to the model of the WFE, to be controlled by the PLC.
2. The control valves will be set to control temperature and vacuum pressure digitally.

The final sets of recommended parameters are shown in Table 12.

Table 12. Set of Parameters Selected for the Model

	Low level value	Center value	High level value
Feed Flow (gpm)	1.0	1.10	1.20
Temperature (F)	385	395	405
Vacuum Pressure (psi)	-13.58	-13.34	-13.10

The ranges of variations of the factors were $\pm 0.1\ gpm$ for the feed flow, $\pm 10\ °\mathrm{F}$ for the temperature and $\pm 0.24\ psi$ for the vacuum pressure.

8. RECOMMENDATIONS

Base on the set of parameters obtained in section 7 (table 12) it is proposed to use a Response Surface Methodology (RSM). The response surface methodology (RSM) is useful for problems in which a response (or several responses) is influenced by certain variables, and the objective is to optimize the process. This process is represented in Figure 21.

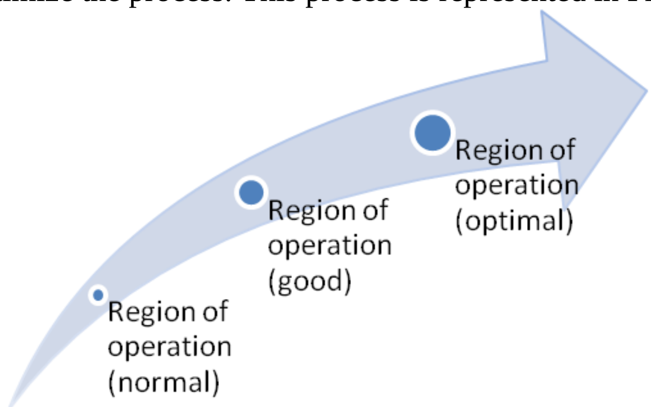

Figure 21. Optimization process using RSM.

The idea of the process is to find a linear regression equation that can represent the response of the system. A second-order model is proposed as an initial estimation and can be expressed as:

$$y = \beta_o + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum\sum_{i<j} \beta_{ij} x_i x_j + \epsilon \quad (8.1)$$

Where

y: response(s)

x: represents the coded variables (number of factors)

β: represents the regression coefficients

ε: is the error or noise associated to the expected response y and the independent factors (x_n).

Note that no matter the order, the equation will always be a linear regression equation given respect to the regression coefficients.

The model is a 2^3 factorial design in which the three factors (feed flow rate, temperature and vacuum pressure) are at two levels.

The range of factors was found based on the analysis of previous experiments. The two levels in which the factors will work are acceptable, and there are not any constraints in the design. A central composite design (CCD) can be used. The center, low level and high level points of the model are the values selected in section 7 (see Table 12). The axial points of the model will be following $= (2^3)^{1/4} = 1.68179$.

A schematic design of the CCD is shown in Figure 22.

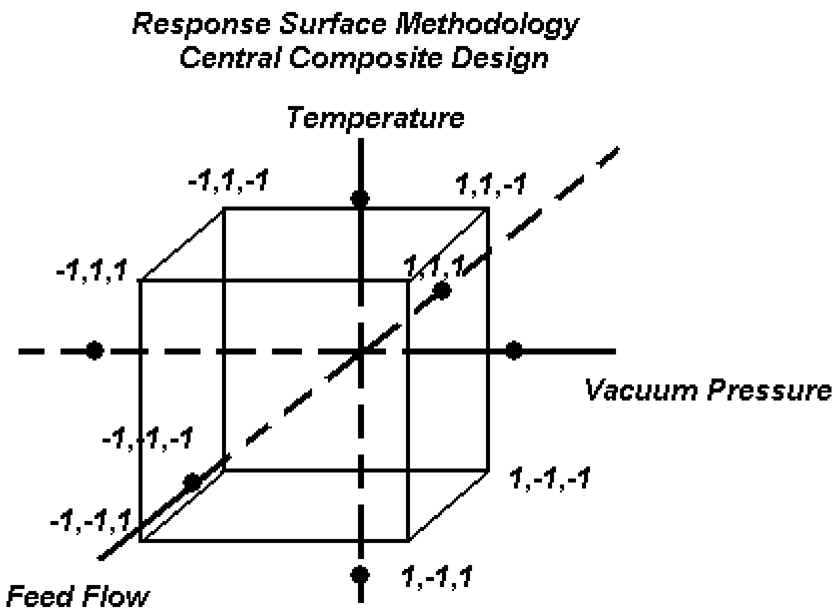

Figure 22. Central Composite design for a 2^3 model.

The following table shows the method planned to be used in future experiments following the values from Table 12.

Table 13. Response Surface Methodology

Natural Variables				***Coded Variables***			***Response(s)***
No.	*Feed Flow*	*Temperature*	*Pressure*	x_1	x_2	x_3	y_i
1	1.0	385	-13.58	-1	-1	-1	
2	1.20	385	-13.58	1	-1	-1	
3	1.0	405	-13.58	-1	1	-1	
4	1.20	405	-13.58	1	1	-1	
5	1.0	385	-13.10	-1	-1	1	
6	1.20	385	-13.10	1	-1	1	
7	1.0	405	-13.10	-1	1	1	
8	1.20	405	-13.10	1	1	1	
9	0.93	395	-13.34	-1.68179	0	0	
10	1.26	395	-13.34	1.68179	0	0	
11	1.10	378	-13.34	0	-1.68179	0	
12	1.10	412	-13.34	0	1.68179	0	
13	1.10	395	-13.74	0	0	-1.68179	
14	1.10	395	-12.93	0	0	1.68179	
15	1.10	395	-13.34	0	0	0	
16	1.10	395	-13.34	0	0	0	
17	1.10	395	-13.34	0	0	0	
18	1.10	395	-13.34	0	0	0	
19	1.10	395	-13.34	0	0	0	
20	1.10	395	-13.34	0	0	0	

The coded variables were assuming 1 or -1 for simplification purposes, but they can be calculated by:

$$x_{imin} = \frac{L-N}{\Delta} \qquad \text{and} \qquad x_{imax} = \frac{H-N}{\Delta}$$

Where:

X_{imin}: is the coded variable that represents the minimum level.
X_{imax}: is the coded variable that represents the maximum level.
N: represent the nominal value (used as the center point).
L: is the low value of the factor x_i.
H: is the high value of the factor x_i.
Δ: is the increment of each factor respectively.

The light and dark grey cells represented in Table 13 are the axial and center points of the cube respectively (Figure 22).

To go further in the optimization process, following the response surface methodology, there are the following recommendations:

1. Follow the matrix specifications for the factors levels summarized in Table 13.
2. If, in the test configuration, the concentrate goes back to the feed tank, then the responses (specific gravity and condensate flow) should be recorded as the slope of the graph of these responses versus time.
3. If, in the test configuration, the concentrate does not go back to the feed tank, then the responses mentioned above should be recorded as the mean value of the response during the observation time.
4. The geometry of the system should remain constant for every measure of the response. In other words, keep the same mode of operation of the WFE during the recording time and just vary the identifying factors.
5. Once the data is collected, process the results by using statistical software such as MINITAB, SPSS, SAS, etc.

10. REFERENCES

1. CEES-0373, Rev. 0. "*Sampling and Shipping Procedure for Wiped Film Evaporator Pilot-Scale Testing"*, Columbia Energy & Environmental Services, Inc., Richland, Washington 2007.
2. WSRC-TR-2000-00338, 2001, *Hanford Waste Simulants Created to Support the Research and Development on the River Protection Project – Waste Treatment Plant*, Rev. 0,Westinghouse Savannah River Company, Aiken, South Carolina.
3. Montgomery C. Douglas. "*Design and analysis of Experiments"* 6th edition. John Wiley & Sons Inc, 2005.
4. Gephart R.E., Lundgren R.E. "*Hanford Tank Cleanup: A Guide to Understanding the Technical Issues.*" Pacific Northwest Laboratory, Richland, WA 1992.
5. CEES-0374 Rev.0. "*Pilot Scale Thin Film Evaporator Test Procedure*." June 13, 2007.
6. Department of Energy & the Office of Environmental Restoration and Waste Management. "*Legend and Legacy: Fifty years of Defense Production at Hanford Site."* (1992) .
7. Zhao Jianzhi. Jin Baosheng, Zhon Zhaoping. "*Study of the separation efficiency of a demister vane with response surface methodology*." Elsevier (2007). Journal of Hazardous Materials 147 (2007) page. 363-369.
8. Cvengros Jan, Pollak Stefan, Micov Miroslav. "*Film wiping in the molecular evaporator*." Elsevier (2000) Chemical Engineering Journal 81 (2001) pages 9-14.
9. Hoaglin, David C.; Welsch, Roy E. (February 1978), "The Hat Matrix in Regression and ANOVA", The American Statistician 32 (1): 17–22, doi:10.2307/2683469, JSTOR: 2683469.