



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

---



# **Modelling and Optimization of an Aldolization System**

Master's thesis in Innovative and Sustainable Chemical Engineering

Marcus Åberg



MASTER'S THESIS 2018

# Modelling and Optimization of an Aldolization System

Marcus Åberg



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

Department of Chemistry and Chemical Engineering  
*Division of Chemical Engineering*  
CHALMERS UNIVERSITY OF TECHNOLOGY  
Gothenburg, Sweden 2018

Modelling and Optimization of an Aldolization System  
Marcus Åberg

© Marucs Åberg, 2018.

Supervisor: Robert Larsén, Perstorp Oxo AB

Examiner: Gunnar Eriksson, Chemistry and Chemical Engineering

Master's Thesis 2018  
Department of Chemistry and Chemical Engineering  
Division of Chemical Engineering  
Chalmers University of Technology  
SE-412 96 Gothenburg  
Telephone +46 31 772 1000

Cover: Aerial view of the Site Stenungsund, by Charlotte Wrennfors, Perstorp Oxo AB.

Gothenburg, Sweden 2018

## Abstract

In this thesis the prospect of optimizing and modeling an aldolization system at Perstorp, site Stenungsund, was investigated. This aim was achieved by studying how production of 2-propyl-heptenal, 4-methyl-2-propyl-2-hexenal and 2-isopropyl-2-heptenal were effected by the recirculation ratio, NaOH concentration and temperature into the reactor. The relationship between the factors and the response were explored by utilizing regression and design of experiment tools, to ultimately create a model for how these factors and response interact. This model can then be used to find the optimal working point for the system at full production. This thesis is a first step to finding this model and a continued plan to find this model are found in this report. Some important discoveries during this thesis were that the hypothesized threshold reaction rate relationship between NaOH concentration into the reactor and the products mentioned above did not occur. Also that the possibility of using a heat exchanger to increase the production rate when the feed flow comes from the tank should be investigated further.

Keywords: Aldolization, Design of experiments, Optimization, 2-propyl-heptenal, 4-methyl-2-propyl-2-hexenal, 2-isopropyl-2-heptenal, Perstorp



## **Acknowledgements**

I would like to extend an extra thanks to everyone that supported me during this work both at Perstop and at Chalmers. There is to many to mention you all, just remember that i appreciate your support but I have some persons i would like to thank a little extra. Below follow three names other then my supervise and examiner that made this work to a success.

Stefan Nyh, COWI

Patrik Svensson, Perstorp Oxo AB

Kai Jauhiainen, Perstorp Oxo AB





# Contents

|   |             |
|---|-------------|
| <b>List of Figures</b>  | <b>xi</b>   |
| <b>List of Tables</b>   | <b>xiii</b> |
| <b>1 Introduction</b>   | <b>1</b>    |
| 1.1 Background . . . . .  | 1           |
| 1.2 Aim . . . . .   | 1           |
| 1.3 Limitations . . . . .   | 1           |
| 1.4 Specification of Issue under investigation . . . . .          | 2           |
| <b>2 Theory</b>   | <b>3</b>    |
| 2.1 Plant layout . . . . .  | 3           |
| 2.2 Reactions pathway . . . . .                                   | 4           |
| 2.3 Kinetic model for aldolization . . . . .                      | 6           |
| 2.4 Byproducts . . . . .  | 7           |
| 2.5 Experimental models . . . . .                                 | 7           |
| 2.5.1 Regression analysis . . . . .                               | 9           |
| 2.5.2 Evaluation of a model . . . . .                             | 9           |
| 2.5.3 Design of experiment . . . . .                              | 11          |
| 2.6 Mass and heat balances for the investigated factors . . . . . | 12          |
| <b>3 Methodology</b>  | <b>17</b>   |
| 3.1 Historical data . . . . .                                     | 17          |
| 3.2 Preparations for full scale experimental planning . . . . .   | 17          |
| 3.3 Full scale experimental plan . . . . .                        | 19          |
| <b>4 Results and discussion</b>                                   | <b>21</b>   |
| 4.1 Evaluation of historical data . . . . .                       | 21          |
| 4.2 Evaluation of the full scale experiments . . . . .            | 23          |
| 4.3 Evaluation of continued plan . . . . .                        | 26          |
| <b>5 Conclusions</b>  | <b>28</b>   |
| <b>Bibliography</b>   | <b>29</b>   |
| <b>A Abbreviations list</b>                                       | <b>I</b>    |
| <b>B Plant abbreviations list</b>                                 | <b>II</b>   |
| <b>C Definitions of regression terms</b>                          | <b>III</b>  |
| <b>D Data restrictions</b>  | <b>IV</b>   |
| <b>E Cooling capacity of the heat exchanger at TEST</b>           | <b>V</b>    |
| <b>F Experimental List</b>  | <b>VIII</b> |

|          |                              |           |
|----------|------------------------------|-----------|
| <b>G</b> | <b>Bubble point</b>          | <b>IX</b> |
| <b>H</b> | <b>Historical data model</b> | <b>X</b>  |

## List of Figures

|    |  |    |
|----|--|----|
| 1  | Part of the plant where the focus of this thesis takes place . . . . .   | 3  |
| 2  | Overview of the site . . . . .   | 4  |
| 3  | Matching the simulated data to historical data. Values are scaled by subtracting the lowest value in each axis. . . . .  | 15 |
| 4  | Matching the historical data of PBA concentration against A) NaOH concentration into the reactor, B) temperature, C) recirculation ratio and D) time with calculated ones. Values are either ranged scaled or scaled by subtracting the lowest value in each axis. . . . . | 22 |
| 5  | Scaled mass percent of what enters the reactor for the different experiments. A) Estimated inflow of NaOH in FNAOH, both calculated and measured B) IVAL, C) NVAL and VAL, D) LE, HE and pentanol. Values are scaled by subtracting the lowest value in each axis. . . . . | 24 |
| 6  | Scaled mass percent of what leaves the DEC1, in the different phases, for the different experiments measured both in SP and TEST. A) 2-mbal, B) NVAL, C) PBA, D) HE. Values are scaled by subtracting the lowest value in each axis. . . . .                               | 25 |
| 7  | A) Relationship between the calculated NaOH concentration into the reactor, with NaOH in FRECI being measured in both SP6 and TEST, and A) the different experiments, B) PBA concentration. Values are scaled by subtracting the lowest value in each axis. . . . .        | 26 |
| 8  | Temperature of sample stream at different cooling mass flows . . . . .   | VI |
| 9  | Scaled lowest bubble point for the different component of the plant. Values are scaled by subtracting the lowest value in each axis. . . . .   | IX |
| 10 | Residual between historical data and formula. Values are scaled by subtracting the lowest value i each axis. . . . .   | X  |



## List of Tables

|   |   |      |
|---|---|------|
| 1 | Values of $t_{\alpha/2, n-p}$ and $F_{\alpha, m-p, n-m}$ with $n, m \rightarrow \infty$ . . . . . | 11   |
| 2 | Required number of experiments for different experimental designs . . . . .                       | 12   |
| 3 | Abbreviations in alphabetic order. . . . .  | I    |
| 4 | Flow meters, thermometer and pressure measurement spot used . . . . .                             | II   |
| 5 | Definitions of regression terms . . . . .   | III  |
| 6 | Boundaries for the historical values . . . . .  | IV   |
| 7 | Changing parameters during the experiments . . . . .  | VIII |
| 8 | Values of $\theta_h$ and $t_{obs, h}$ . . . . .   | X    |



# 1 Introduction

Perstorp is an international chemical manufacturing company that have been around for 130 years, with 1500 employees combined in Europe, North America and Asia. The company has an annual turnover of about 11 billion SEK and their special chemicals are used to make products like computers and shoes both safer, stronger and lighter. Perstorp also has a high focus on responsibility, reliability and innovation which influences the corporate culture. Making Perstorp an ideal working environment for chemical engineers all over the world. A list of the abbreviations used in this report can be found in appendix A.

## 1.1 Background

Site Stenungsund, the plant where the practical part of this thesis took place, is located 50 km north of Gothenburg, on the west coast of Sweden. It became a part of Perstorp Group in 2001 but was first commissioned back in 1980. Main products produced here are aldehydes, alcohols, organic acids and renewable biofuel.

At site Stenungsund there is a fully operational aldolization unit, commissioned in early 2015. This unit has not yet been fully optimized which leads to the aim of this thesis. Because this aldolization unit is part of a plant that takes isomers of pentanal or valeraldehyde, VAL, as it is also called and first turn it into isomers of 2-propyl-heptenal, or 2-propyl-3-butyl-akrolein, PBA, as it is also known, see section 2.2 [1]. Down the line PBA is turned into alcohols with 10 carbon atoms where the main product is 2-propyl-heptanol, 2-PH. 2-PH can be used as a plasticizer alcohol, as raw material for acrylates and lubricants in adhesive applications [1, 2]. Also as a raw material for the production of surfactants used in industrial cleaning agents applicable on hard surfaces and detergents [2].

## 1.2 Aim

This master thesis has the goal to optimize the aldolization unit by finding the working point with the highest economical benefits at specific throughput by optimizing selected control variables, then develop a plan for optimizing these factors which results in the creation of a model depicting how PBA production changes with these factors. Finally perform as many planned experiments that time allows and leaving the rest of the experiments as future optimization.

## 1.3 Limitations

Since there is a lot of different variables that could be optimized for this system, a limitation on the control variables to only include NaOH concentration, temperature and recirculation ratio has been selected. As a starting point, existing equipment was considered but a new test site was built to lower the time between measurements. The reactor is assumed to be ideal.

The production unit is also driven with two different inlet ratio between normal-valeraldehyd, NVAL, and iso-valeraldehyd, IVAL [1]. Therefore a decision to optimize the reactor for the normal ratio was made and then suggest a plan optimizing the higher IVAL inlet ratio.

#### **1.4 Specification of Issue under investigation**

The issue with the aldolization reactor is that if the concentration of NaOH is too high, then more product will be wasted in the upcoming separation steps see section 2.1. The NaOH concentration should not be too high to avoid increased production of so called heavy ends, HE, see section 2.4. Too low NaOH concentration lead to insufficient convergence even if the selectivity of the PBA is increased. The relationship between the desired aldolization and the NaOH concentration is hypothesized to be of a threshold nature. Above the threshold the reaction rate is almost constant and below it decreases rapidly. This all lead to the hypothesis that the NaOH concentration should be kept as low as possible but still above the threshold. [3]

Too high temperature is hypothesized to result in more byproduct in the form of HE, while too low temperature results in insufficient convergence [4]. The conclusion that higher reaction temperature leads to an increase in HE production might be incorrect [4]. A conflicting report showed that the HE concentration was lowered by higher temperature so this require some investigation [3], since if the production of HE does not increase with increased temperature, then there is no upper limit to the temperature as long as the reactants remains in liquid form and the NVAL does not form any extra HE. The first report mentioned also states that the speculations of higher end production at higher temperature are just speculations and not verified [4]. This makes temperature a very interesting parameter to optimize since the optimal temperature might be outside the previously tested values. This might mean that the temperature can be raised without any boundaries as long as the pump can keep the stream in liquid phase.

If the recirculation ratio is too high, the reactor capacity will be smaller and if the recirculation ratio is too low, then the convergence might be insufficient and PBA selectivity lowered. This parameter is hypothesized to have the smallest impact on the overall reaction rate of the system assuming the water phase flow rate is kept constant, because it is mostly water that is recirculated which will change the water and organic ratio in the stream. Solubility of the organic compound in water is assumed to have a bigger impact on the overall reaction rate compared to the organic and water ratio, because of the fact that the PBA reaction is assumed to mostly take place in the water phase. Since temperature and NaOH concentration is assumed to have a larger impact on the solubility of the organic compound in the water phase the water to organic ratio are hypothesized to be of less importance. [3]



## 2 Theory

Below follows some extra information about the plant layout in big strokes and the chemistry within, together with the theoretical tools needed to fulfill the aim of this thesis.

### 2.1 Plant layout

A more detailed version of the area of the site where the focus of the thesis took place can be seen in figure 1. Not every heat exchanger, pump or measurement equipment can be seen in the figure, but it is detailed enough for the reader to be able to follow along. The only measurement equipment that can be seen in figure 1 is the SP1 to SP6, which together with the temporary test point, TEST, are the only places where concentration can be measured. A less detailed version of the rest of the site can be seen in figure 2. This is just to show how the plant turn the produced PBA into the final product 2-PH. See appendix B for a short description of all the different points in figure 1 and 2.

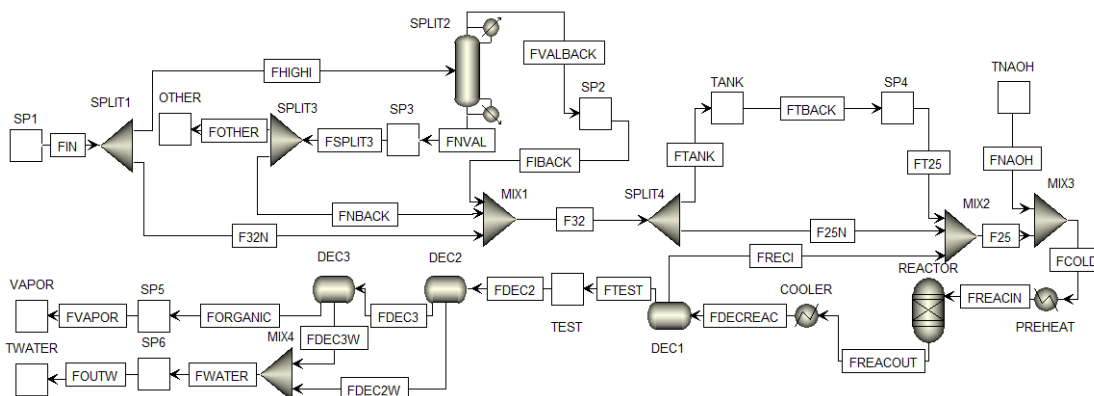


Figure 1: Part of the plant where the focus of this thesis takes place

The feed stream into the plant, FIN, in figure 1 consists of almost pure VAL and a smaller amount of HE [1]. In SPLIT1 the feed is turned into two streams. But in normal production, where the standard amount of NVAL are transported to the PBA reactor, only a small amount of feed are flowing in FHIGHI. The FHIGHI flow that is separated in SPLIT2 are then reunited with the main flow of F32N. In other words the flow in FOTHER is put to zero. This is to ensure the SPLIT2 column does not dry out and also that the concentration in SP1 is enough to measure during normal production. If NVAL is needed in an other part of the site, it is taken out and separated in SPLIT3 and only a smaller amount of NVAL is returned to F32. Leading to a higher IVAL concentration overall. F32 is separated again to either the storage tank, FTANK, or directly to the PBA reactor, F25N. F25N is mixed with flow from the tank FT25 and recirculated flow from the decanter after the reactor, FRECI. NaOH is added from a mixing tank, TNAOH. The flow leaving MIX3, FCOLD, is temperature and pressure regulated before entering the tube reactor, REACTOR.

In the tube reactor the VAL is turned into PBA where the NaOH acts like a catalyst. The stream leaving the rector, FREACOUT, is used to create steam in a heat exchanger, COOLER. This cooled down flow, FDECREAC, is separated easier than a warmer flow

in decanter, DEC1. Mostly water but also some heavier molecules are separated from FDECREAC stream in DEC1 before it is recirculated, FRECI, into F25 [5]. The main product that leave DEC1 is treated in two water separators, DEC2 and DEC3, where the water flow is treated, TWATER, while the higher PBA concentrated stream, FVAPOR, continues on in the plant. Part of the plant that is described above is where the focus of this thesis takes place. See figure 2 for more details.

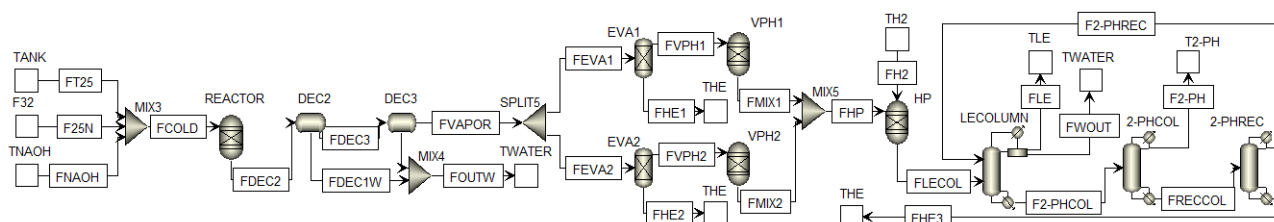
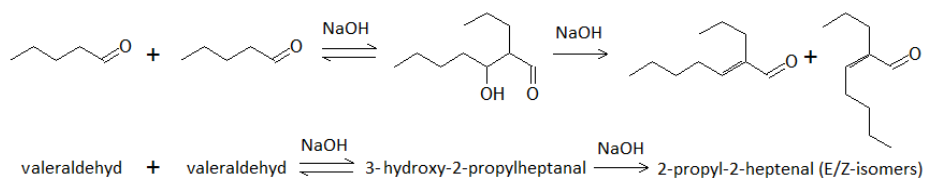


Figure 2: Overview of the site

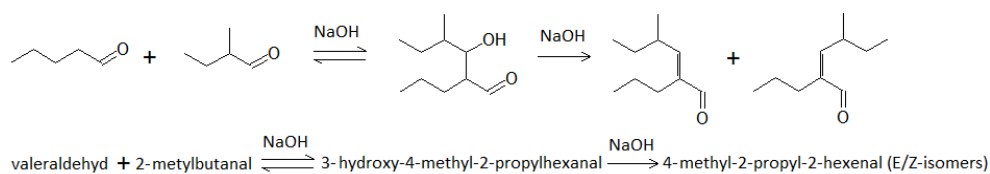
As can be seen in figure 2, FVAPOR is split into two streams, FEVA1 and FEVA2, and processed through two different vaporization units, EVA1 and EVA2, where most of the HE are taken out and shipped to storage, THE. The next step in the production line is two parallel vapour phase hydrogenation units, VPH1 and VPH2, where PBA is turned into 2-PH with the help of a catalyst. The two streams are mixed together, MIX5, and led into a high pressure reactor, HP, where the last of the PBA is turned into 2-PH with the help of hydrogen gas, FH2, and a catalyst. Then the separation process begins where light ends, LE, are distilled out at the top of the first distillation column, LECOLUMN, and sent to a tank, TLE. The heavy distillation stream, F2-PHCOL, consists of mostly 2-PH and HE. This stream goes on to another distillation unit, 2-PHCOL, where most of the product, 2-PH is separated out in the light stream and taken to storage, T2-PH. The heavy stream, FRECCOL, consists not only of HE but some 2-PH too, to keep down the distillation temperature in 2-PHCOL. This is to avoid additional HE production. FRECCOL goes into a smaller distillation column, 2-PHREC, where HE are taken out at the bottom and taken to storage, THE, and mostly 2-PH is recirculated into LECOLUMN.

## 2.2 Reactions pathway

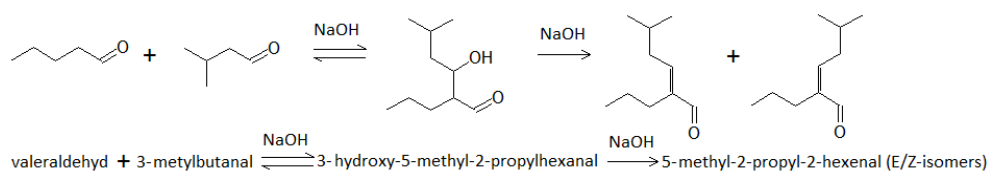
To create an aldol from VAL it first has to become an anion in the presence of a catalyst like NaOH [6, 7]. Depending on which VAL that has become an anion, different aldols are created. If NVAL becomes an anion then three main reactions can occur to create aldols. It all depends on what the anion of NVAL is reacting with. It can react with an NVAL or with an IVAL represented by 3-methylbutanal, 3-mBal, and 2-methylbutanal, 2-mBal, in the presence of the catalyst NaOH [1]. This process is called an aldolization process or an aldol condensation process [6, 7, 8]. The result of these reactions are not too different from the heavier documented butanal aldolization reactions [1]. The NVAL reactions can be seen in reaction 1 to 3 [3, 8, 9, 10, 11].



Reaction 1: Aldolization reaction for NVAL + NVAL

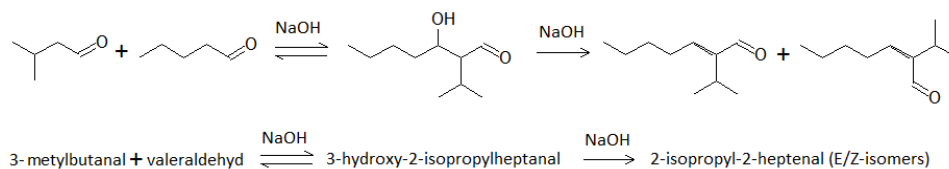


Reaction 2: Aldolization reaction for NVAL + 2-mBal

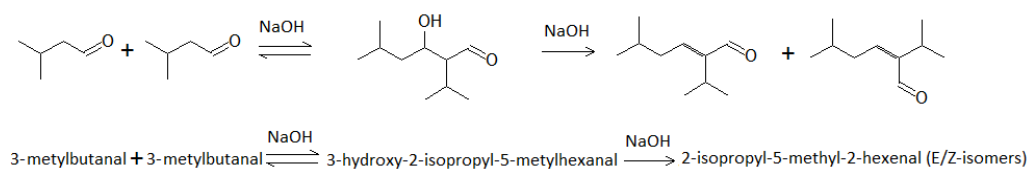


Reaction 3: Aldolization reaction for NVAL + 3-mBal

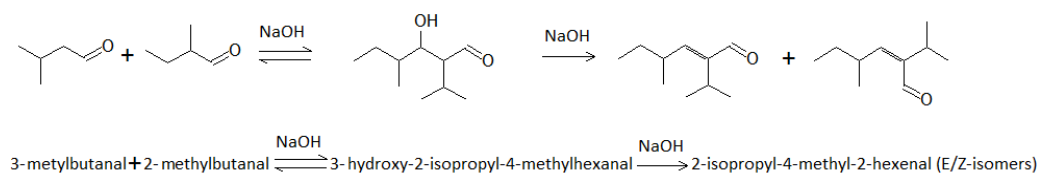
If it is instead 3-mBal that is ionized, then reaction 4 to 6 occur in the presence of NaOH. [1, 3, 6, 8, 10, 11]



Reaction 4: Aldolization reaction for 3-mBal + NVAL



Reaction 5: Aldolization reaction for 3-mBal + 3-mBal



Reaction 6: Aldolization reaction for 3-mBal + 2-mBal

Ionization of 2-mbal can also theoretically lead to the creation of an aldol when it reacts with NVAL, but in the aldol intermediate there is a shortage of  $\alpha$ -hydrogen which results in difficulties creating the conjugated double bond system of an aldol. Without the possibility of creating this bond in the water elimination or aldol condensation reaction further down the product line, the retro-aldolization reaction takes over and nearly no aldols are created with the ionization of 2-mBal. [3, 6, 8]

Only reactions 1, 4 and 2 are classified to create PBA molecules. That is since the concentration of 3-mBal and  $\alpha$ -hydrogen is kept low. Reactions 3, 5 and 6 still exist but in so small amounts that they along with 2-mBal ionization reactions are not classified as PBA reactions. Only reaction 1 eventually lead to the creation of 2-PH. The other two reactions, 2 and 4, end up producing extra alcohol molecules with 10 carbon atoms in the end product.

### 2.3 Kinetic model for aldolization

As described in section 2.2, only three reactions are classified as PBA reactions, depending on what is reacting. To describe the first step, where aldols are created for the three main PBA producing reactions, an expression for the reaction rate can be seen in equation 1, 2 and 3. The reaction rate displayed in equation 1 correspond to reaction 1, equation 2 correspond to reaction 2 and equation 3 correspond to reaction 4.

$$r_{aldol_1} = k_1 \left( c_{NVAL}^2 - \frac{c_{aldol_1}}{K_{aldol_1}} \right) \quad (1)$$

$$r_{aldol_2} = k_2 \left( c_{NVAL} * c_{2-mBal} - \frac{c_{aldol_2}}{K_{aldol_2}} \right) \quad (2)$$

$$r_{aldol_3} = k_3 \left( c_{NVAL} * c_{3-mBal} - \frac{c_{aldol_3}}{K_{aldol_3}} \right) \quad (3)$$

The forward reaction in the equilibrium reactions 1 to 6 can be thought of as termolecular and the backwards reaction then becomes bimolecular, but only if NaOH is taken into account in the reaction rate expression [7, 12, 13]. But since NaOH acts like a catalyst in this reaction, and an in-house report show that the reaction rate is more or less independent of the NaOH concentration in the existing concentration interval, resulting in that the forward reaction have the reaction rate of a bimolecular reaction or a second order reaction rate while the backward reaction have the reaction rate of an unimolecular reaction or a first order reaction rate [12, 13, 14].

After the creation of the aldol, an elimination reaction where the aldol loses a water molecule is needed to create PBA [6]. The reaction rate for this elimination reaction can be seen in equation 4 to 6, where equation 1 correspond to equation 4 and so on, in the presence of sufficient levels of NaOH. This reaction is also assumed to be independent of the NaOH concentration which turns the otherwise bimolecular reaction into a unimolecular reaction [3, 12, 14].

$$r_{PBA1} = k_4 c_{aldol_1} \quad (4)$$

$$r_{PBA2} = k_5 c_{aldol_2} \quad (5)$$

$$r_{PBA3} = k_6 c_{aldol_3} \quad (6)$$

## 2.4 Byproducts

There are three different classifications of byproducts leaving the PBA reactor. The three classifications are HE, LE and byproducts from an earlier stage of the plant and the product of the byproduct reactions in the PBA reactor. LE are VAL that did not react in the PBA reactor and HE are PBA that continue to react into heavier molecules. Increased amount of HE production leads to higher overall production costs and losses in potential revenue. This is because PBA is needed to create HE and because more of PBA is wasted in the vaporization unit, see section 2.1, to ensure that the bottom flow does not clog up.

There are a lot of potential HE that can be formed if NVAL and IVAL are allowed to continue to react in the presence of NaOH. Some are created with continued aldol condensation of PBA which is the case for the reactions described with the reaction rates shown in equation 7 and 8, where equation 7 shows the reaction rate when a PBA continue to react with a VAL to form a heavy end with 15 carbon atoms [7]. Here  $i \in [1, 3]$  symbolizes the number for the aldols created in equation 1 to 3. The reactions are assumed to only go in one direction and be dependent on the water and NaOH concentration since the reaction is treated as termolecular, which is why a lower NaOH concentration could lead to less HE production [3, 12, 13].

$$r_{C15Heavies} = k_7 \frac{c_{NaOH}}{c_{H_2O}} \left( \sum_{i=1}^3 c_{aldol_i} \right) c_{VAL} \quad (7)$$

Equation 8 shows the reaction rate if a PBA react with an other PBA [3, 12, 13].

$$r_{C20Heavies} = k_8 \frac{c_{NaOH}}{c_{H_2O}} \left( \sum_{i=1}^3 c_{aldol_i} \right)^2 \quad (8)$$

## 2.5 Experimental models

Both regression analyze and design of experiment tools was used in this thesis. Below follows some theory concerning the different tools used. When using most experimental models the collected data is split up into objects, X-variables and Y-variables [15, 16]. In table 5 in appendix C follows a short descriptions of some different terms when working with experimental models.

Below in equation 9 to 13 follows expressions on how to calculate some of the terms in table 5. In these equations  $i = 1, 2, \dots, n$ . [15, 16]

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (9)$$

If the variance seen in equation 10 is set to zero that means that all measurements are identical. A large variance indicate a large spread in the measurements. [15, 16]

$$Var(x) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (10)$$

Since the  $S_x$  is just the square root of the variance which can be seen in equation 11, a low value of  $S_x$  indicate a low spread of the measurements. The reason why  $S_x$  is used is so the spread of the measurement can be expressed using the same units as the measurements. [15, 16]

$$S_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (11)$$

A large absolute value of the covariance indicate a strong linear dependence between two variables, while a value close to zero indicate the opposite. If the regular covariance seen in equation 12 is positive that means that values of both variables occur together, the variables mostly display similar behavior, while a negative covariance indicate a display of large values of one variable correspond to small values of the other variable. [15, 16]

$$cov(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1} \quad (12)$$

To easier assess if the covaraince or the linear dependence is large between two different variables the expression for the correlation is used instead, seen in equation 13. That way a unit-less and scaled term can be used to evaluate the linear dependence between two variables. The scale for  $r \in [-1, +1]$ , where a value of zero gives no correlation, -1 exactly gives a negative linear correlation and +1 gives a positive linear correlation between the variables. [15, 16]

$$r = \frac{cov(x, y)}{S_x S_y} \quad (13)$$

An other important variable to keep in mind when performing MVDA and other DOE analysis is the so called  $Q^2$ -value. This value represent how well a component can predict the total variance of Y or X, estimated using cross-validation. It is recommended to have this value as high as possible or at least higher then 0.5. Even if it fully confirms that the model has a good predicted effect, it can still be used as a good indicator. [17, 18]

For both of these experimental analysis tools, three x-variables and one y-variable had to be specified to form the X-variable and Y-variable, see section 2.5.1 for more details. The factors are quantitative NaOH concentration in FCOLD,  $x_1$ , temperature in FREACIN,  $x_2$ , and the mass flow ratio between fresh flow into the reactor loop,  $F_{25N} + F_{T25} + F_{NAOH}$  and the recirculation flow,  $F_{RECI}$  also called recirculation ratio,  $x_3$ . The response is the PBA concentration measured either in TEST or SP5. See section 2.1 for more information about the plant layout and section 2.6 for more information about the definition of the calculation of the factors.

### 2.5.1 Regression analysis

When fitting a model to a big amount of data there exist different types of methods. One of the more common one is multivariate data analysis, MVDA and an other is a regression approach [15]. Regression approach is used to relate two sets of variables with each other [15]. This is done by selecting the relevant X-variables and see how they correspond to the chosen Y-variables [15]. There may be different approaches to this such as multiple linear regression, MLR, but as the name suggest this approach is for linear models with independent variables [16]. In some cases an linear model is not applicable and an non-linear approach like nonlinear least square, NLS, have to be used [19].

As for the MLR approach  $y = (y_1, \dots, y_n)'$  stands for a vector with n number of dependent responses,  $x_i = (x_{i1}, \dots, x_{ik})'$ , referees to a vector of k independent variables that correspond to the  $i$ th observation, the  $X = (x_1, \dots, x_n)'$  matrix of independent variable values and the error vector,  $\epsilon = (\epsilon_1, \dots, \epsilon_n)'$  [16]. The only thing that differentiate the nonlinear and linear is the parameter vector  $\beta$ , which is replaced by  $\theta^* = (\theta_1^*, \dots, \theta_q^*)$ .  $\theta^*$  refer to the exact parameter vector value, while the model's estimated value is  $\theta$ . This means the model can be constructed as in equation 14 . [19]

$$y_i = f(x_i, \theta^*) + \epsilon_i \quad (14)$$

Where the function,  $f(x_i, \theta^*)$ , need to be specified. In the model  $\epsilon_i$  is assumed to be uncorrelated, homoscedastic and have a mean zero. The estimation of  $\theta^*$  in the NLS approach is done by trying to minimize the difference between  $\theta$  and  $\theta^*$ . This is done by minimizing the error sum of square,  $SS_E$  seen in equation 15. [19, 16]

$$S^2(\theta) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - f(x_i, \theta))^2 = SS_E \quad (15)$$

If the partial derivative of equation 15 with regards to  $\theta_h$  is put to zero and adequate differentiability of  $f$  is assumed then we get the expression in equation 16. [19]

$$\sum_{i=1}^n (y_i - f(x_i, \theta)) * \frac{\delta f(x_i, \theta)}{\delta \theta_h} = \sum_{i=1}^n (y_i - f(x_i, \theta)) * v_{ih} = 0 \quad (16)$$

For the values of  $j = 1, \dots, k$ , since if  $j = 0$  then  $v_{i0}$  is the derivative of a constant. This can sometime be solved analytically but most of the time an iterative algorithmic solution is required. [19]

An other method to try and minimize the difference between  $\theta$  and  $\theta^*$  is by manually trying to fit the model to experimental data using a combination of the residual,  $e_i$ , and testing the significance of regression coefficients,  $\theta_h$ . The residual is calculated with equation 17. For more information about about the residual and how to evaluate models see section 2.5.2 below.[16, 20]

### 2.5.2 Evaluation of a model

Below follows some simple tools to help evaluate a potential model. This chapter will include a short description of how to use the residual to find potential conceptions between

factors, the t-test and also the lack of fit test.

When the residual is calculated as in equation 17 and plotted against factors, it is important that no clear relationship can be seen between the residual and the factors, because if there is then that indicate that the proposed model is incomplete and missing some factors. [16, 20]

$$e_i = y_i - \hat{y}_i = \sqrt{SS_E} \quad (17)$$

By using the expression for  $SS_E$  seen in equation 15, the variance can be calculated using equation 18. [16, 20]

$$\hat{\sigma}^2 = \frac{SS_E}{n - p} \quad (18)$$

The correlation matrix,  $C_{gj}$ , for a non-linear function can be seen in equation 19. The non linear  $C_{gj}$  uses Jacobian matrix instead of the X matrix for the linear case. [20]

$$C_{gj} = \frac{\{(J'J)^{-1}\}_{gj}}{\sqrt{\{(J'J)^{-1}\}_{gg} * \{(J'J)^{-1}\}_{jj}}} \quad (19)$$

If  $C_{gj}$  in equation 19 is close to zero then there are no correlation between the parameters but if it is instead close to 1 or  $-1$  then there is a strong correlation between the parameters and the experimental design need to be examined.  $\hat{\sigma}$  together with  $\{(J'J)^{-1}\}_{jj}$  may be used to create the standard error,  $se(\theta_h)$ , seen in equation 20. [16, 20]

$$se(\theta_h) = \hat{\sigma} * \sqrt{\{(J'J)^{-1}\}_{jj}} \quad (20)$$

By dividing  $\theta_h$  with  $se(\theta_h)$  a formula for estimating the validity of the null hypothesis is created. In equation 21 the test statistic for the null hypothesis for parameter  $h$ ,  $t_{obs,h}$ , can be seen. The null hypothesis is rejected if the expression in 21 is true, see table 1 for values of  $t_{\alpha/2, n-p}$ . [16, 20]

$$|t_{obs,h}| = \left| \frac{\theta_h}{se(\theta_h)} \right| > t_{\alpha/2, n-p} \quad (21)$$

If the null hypothesis is true on the other hand then  $\theta_h$  can be put to zero since does not have any significant impact on the model. [16, 20]

If repeated experiments are conducted with the same values on the factors then  $SS_E$  calculated with equation 15 can also be split up into two different components, the sum of square due to lack of fit,  $SS_{LoF}$ , and the sum of square due to pure error,  $SS_{PE}$ . If  $i = 1, \dots, m$  represent all the unique measurement points then  $l = 1, \dots, n_i$  represent the number of repeated experiments for the same values of the  $i$ th level. That means  $n = \sum_{i=1}^m n_i$ , which results in the expression seen equation 22. [16, 20]

$$SS_E = SS_{PE} + SS_{LoF} = \sum_{i=1}^m \sum_{l=1}^{n_i} (y_{il} - \hat{y}_i)^2 = \sum_{i=1}^m \sum_{l=1}^{n_i} (y_{il} - \bar{y}_i)^2 + \sum_{i=1}^m n_i * (\bar{y}_i - \hat{y}_i)^2 \quad (22)$$



By dividing the sum of squares in equation 22 with the degree of freedom an expression for the mean sum of squares are created. The pure error mean sum of square,  $MS_{PE}$  can be seen in equation 23. [16, 20]

$$MS_{PE} = \frac{SS_{PE}}{n - m} \quad (23)$$

Equation 24 shows the expression for the lack of fit mean sum of square,  $MS_{LoF}$ . [16, 20]

$$MS_{LoF} = \frac{SS_{LoF}}{m - p} \quad (24)$$

When both  $MS_{LoF}$  and  $MS_{PE}$  are obtained the lack of fit test statistics,  $F_{obs,LoF}$ , can be obtained by just dividing them as can be seen in equation 25 [20, 16]

$$F_{obs,LoF} = \frac{MS_{LoF}}{MS_{PE}} \quad (25)$$

If  $F_{obs,LoF} < F_{\alpha,m-p,n-m}$  then the lack of fit is not significance and the model can be evaluated further, see table 1 for values of  $F_{\alpha,m-p,n-m}$ . If this is not the case then further evaluation of the model and the residual plots are required. [20, 16]

Table 1: Values of  $t_{\alpha/2,n-p}$  and  $F_{\alpha,m-p,n-m}$  with  $n, m \rightarrow \infty$

| Variable                 | value |
|--------------------------|-------|
| $t_{0.025,\infty}$       | 1.645 |
| $t_{0.005,\infty}$       | 2.326 |
| $F_{0.05,\infty,\infty}$ | 1     |
| $F_{0.01,\infty,\infty}$ | 1     |

### 2.5.3 Design of experiment

When designing experiments to try and create a model for a system where at least the second order derivative are fully taken into account, there are many possibilities. Possibilities like a  $3^k$ -factorial design where k is equal to the number of factors investigated and the number 3 since there will be 3 points investigated on each factorial, -1, 0 and +1 [16, 21]. The  $3^k$ -factorial design gives quite accurate representation of the system but require a large amount of experiments especially if more then 2 factors are investigated. Instead a more conservative approach can be used in the form of centered composition design, CCD or Box-Behnken design [22, 16, 21].

For the CCD this turns the  $3^k$ -factorial design into a variant of  $2^k$  with  $2k$  extra points for each factor investigated. These  $2k$  points are located at an axial distance,  $\alpha_r$ , from origo in the factorial plain. CCD and  $3^k$ -factorial design also have a number of repeats of the experiment point in origo,  $n_0$ . Table 2 illustrates the difference in required amount of experiments for the different experimental designs. With  $n_0 = 3$  but since  $3^k$ -factorial design already have a point in origo it uses  $3^k + (n_0 - 1)$  number of experiments. [16, 21, 23]

Table 2: Required number of experiments for different experimental designs

| factors | $3^k$ -factorial | CCD | Box-Behnken design |
|---------|------------------|-----|--------------------|
| 1       | 5                | 7   | -                  |
| 2       | 11               | 11  | -                  |
| 3       | 29               | 17  | 17                 |
| 4       | 83               | 27  | 27                 |

As can be seen in table 2. If the number of factors extends 3 then CCD and the Box-Behnken design is clearly a more conservative models. The problem with the Box-Behnken design is its missing corners which leads to lost information while CCD can be designed to be both rotatable, depending on the value of  $n_0$  and  $\alpha_r$ , and orthogonal, depending of the value of  $\alpha_r$  [16, 22, 21, 23]. There are different types of CCD models which all depends on the value  $\alpha_r$ .

For example if  $\alpha_r$  is put to 1 then the model is called centered composition face-centered design, CCF. CCF is not rotatable but is instead useful when the region of interest appears more cubical than spherical, for example if the maximum is known to be close to the center of the cube, since the variance increase the further away from the center you get. [16, 24]

An other example of an CCD is if  $\alpha_r = n_F^{1/4}$ , where  $n_F$  is the amount of points in the factorial part of the design. This design called a central composition circumscribed design, CCC, and for 3 factors,  $n_F = 8$  which means  $\alpha_r = 8^{1/4}$ . CCC are rotatable which means that the variance is evenly spread over the region of interest. [16, 24]

A third example is one that is somewhere in between a CCF and a CCC. It is called a central composition orthogonal design, CCO, and has an  $\alpha_r = 1.35313$  for 3 factors. This design is not rotatable which mean that the variance wont be evenly spread over the region of interest. If the maximum can be assumed to not exist in the corners of factorial part of the CCO, then the lack of rotatability might be a fair trade off to have the region of interest closer to immeasurable parts of the factors. [16, 25]

## 2.6 Mass and heat balances for the investigated factors

There already exist a thermostat at FREACIN, so no heat or mass balance needed for the factor  $x_2$ . The fact that flowmeters or the ability to measure concentration do not exist in every flow, resulted in that mass balances had to be utilized to comprehend the flows in certain places necessary to create the sought after factors,  $x_1$  and  $x_3$ , see section 2.5. For a better understanding of the different terms in the mass balances see section 2.1 and appendix B. The only assumptions is that the total mass is preserved and that the concentration in FRECI is the same as in the water phase in FTEST and by extension in FWATER.

$x_3$  was regulated by choking the FCOLD with the same combined fresh inflow from F25N and FT25. The definition for  $x_3$  can be found in equation 27 with  $o$  referring to how much

the valve is opened with regards to a chosen maximum FCOLD flow, see section 3.3 for more information about the how the factors were planned to be changed during the experiments.

$$x_3 = \frac{\dot{m}_{FRECI}}{\dot{m}_{F25N} + \dot{m}_{FT25} + \dot{m}_{FNAOH}} = \frac{\dot{m}_{FCOLD} * o - (\dot{m}_{F25N} + \dot{m}_{FT25} + \dot{m}_{FNAOH})}{\dot{m}_{F25N} + \dot{m}_{FT25} + \dot{m}_{FNAOH}} \quad (26)$$

Equation 26 can then be used to craft the final expression for  $x_3$  which can be seen in equation 27.

$$x_3 = \frac{\dot{m}_{FCOLD} * o}{\dot{m}_{F25N} + \dot{m}_{FT25} + \dot{m}_{FNAOH}} - 1 \quad (27)$$

By rearranging equation 27 an expression for the inlet flow that depend on  $x_3$  can be seen in equation 28

$$\dot{m}_{F25N} + \dot{m}_{FT25} = \frac{\dot{m}_{FCOLD} * o}{x_3 + 1} - \dot{m}_{FNAOH} \quad (28)$$

Equation 28 can then be used to construct an expression for  $x_1$  that require knowledge about NaOH concentration in the water phase in the flow  $\dot{m}_{FRECI}$ . This expression can be seen in equation 29, where W denotes only the water phase of the respective flows.

$$x_1 = \frac{\dot{m}_{FCOLD,NaOH}}{\dot{m}_{FCOLD} * o} = \frac{\dot{m}_{FNAOH} * X_{FNAOH,NaOH} + \dot{m}_{FRECI} * X_{FRECI,NaOH}}{\dot{m}_{FCOLD} * o} \quad (29)$$

Since  $\dot{m}_{FRECI} = \dot{m}_{COLD} * o - (\dot{m}_{FNAOH} + \dot{m}_{F25N} + \dot{m}_{FT25})$  equation 29 becomes

$$\begin{aligned} x_1 &= \frac{\dot{m}_{FNAOH} * X_{FNAOH,NaOH} + \dot{m}_{FCOLD} * o * X_{FRECI,NaOH} - (\dot{m}_{FNAOH} + \dot{m}_{F25N} + \dot{m}_{FT25}) * X_{FRECI,NaOH}}{\dot{m}_{FCOLD} * o} = \\ &= \frac{\dot{m}_{FNAOH} * (X_{FNAOH,NaOH} - X_{FRECI,NaOH}) - (\dot{m}_{F25N} + \dot{m}_{FT25}) * X_{FRECI,NaOH}}{\dot{m}_{FCOLD} * o} + X_{FRECI,NaOH} \end{aligned} \quad (30)$$

Using the assumption that  $X_{FRECI,NaOH} = X_{TEST,W,NaOH}$  for equation 30 makes it necessary to guess a value for the concentration of NaOH in the water phase at the test point,  $X_{TEST,W,NaOH}$ . This value is verified when the sample is measured. The guess is calculated with a combination of old simulated values using HYSYS from when the site was built, denoted with H in the name. These values were then matched with historical values from the plant, resulting in an expression for  $X_{TEST,W,NaOH}$  seen in equation 31 with not all the NaOH leaving the system in  $\dot{m}_{FWATER}$  and at steady state.

$$X_{TEST,W,NaOH} = \frac{\dot{m}_{FTEST} * X_{FTEST,NaOH}}{\dot{m}_{FWATER}} = \frac{\dot{m}_{FNAOH} * X_{FNAOH,NaOH}}{\dot{m}_{FTEST,W} * (X_{FTEST,W,H_2O} + X_{FTEST,W,NaOH} + X_{FTEST,W,VAL})} \quad (31)$$

Equation 32, 33 and 34 is used to calculate values for the different terms in the denominator of equation 31.

$$\frac{\dot{m}_{FTEST,W} * X_{FTEST,W,H_2O}}{\dot{m}_{F25N} + \dot{m}_{FT25}} = \frac{\dot{m}_{FWATER,H} * X_{FWATER,H,H_2O} - \dot{m}_{FNAOH,H} * X_{FNAOH,H,H_2O} + \dot{m}_{FORGANIC,H} * X_{FORGANIC,H,H_2O}}{\dot{m}_{F25N,H} + \dot{m}_{FT25}} \quad (32)$$

$$\frac{\dot{m}_{FTEST,W} * X_{FTEST,W,NaOH}}{\dot{m}_{F25N} + \dot{m}_{FT25}} = \frac{\dot{m}_{FNAOH,H} * X_{FNAOH,H,NaOH}}{\dot{m}_{F25N,H} + \dot{m}_{FT25,H}} \quad (33)$$

$$\frac{\dot{m}_{FTEST,W} * X_{FTEST,W,VAL}}{\dot{m}_{F25N} + \dot{m}_{FT25}} = \frac{\dot{m}_{FTEST,W} \sum_{i=1}^n X_{FTEST,W,i}}{\dot{m}_{F25N} + \dot{m}_{FT25}} = \frac{\dot{m}_{FWATER,H} * (X_{FWATER,H,2-mbal} + X_{FWATER,H,NVAL})}{\dot{m}_{F25N,H} + \dot{m}_{FT25,H}} \quad (34)$$

In equation 34  $i$  refer to all the individual organic compounds in the water phase and  $n$  refer to the total number of different organic compound in the water phase. In the simulated values only 2-mbal and NVAL is assumed to exist in  $\dot{m}_{FWATER,H}$ . The older simulated variables, the ones including a H in the name, of the right side of equation 32, 33 and 34 are summarized as a constant,  $k_1$ , in equation 35.

$$k_1 = \frac{\dot{m}_{FWATER,H} * X_{FWATER,H,H_2O} - \dot{m}_{FNAOH,H} * X_{FNAOH,H,H_2O} + \dot{m}_{FORGANIC,H} * X_{FORGANIC,H,H_2O}}{\dot{m}_{F25N,H} + \dot{m}_{FT25,H}} + \frac{\dot{m}_{FNAOH,H} * X_{FNAOH,H,NaOH} + \dot{m}_{FWATER,H} * (X_{FWATER,H,2-mbal} + X_{FWATER,H,NVAL})}{\dot{m}_{F25N,H} + \dot{m}_{FT25,H}} \quad (35)$$

The expression in equation 35 is then used to construct an expression for  $X_{TEST,W,NaOH}$ , seen in equation 36. Here  $k_2$  is just a value to better adapt the approximation to historical values, see figure 3 for the difference.

$$X_{TEST,W,NaOH} = \frac{\dot{m}_{FNAOH} * X_{FNAOH,NaOH}}{(\dot{m}_{F25N} + \dot{m}_{FT25}) * k_1} - k_2 \quad (36)$$

$$x_1 = \frac{\dot{m}_{FNAOH} * (X_{FNAOH,NaOH} - X_{TEST,W,NaOH}) - (\dot{m}_{F25N} + \dot{m}_{FT25}) * X_{TEST,W,NaOH}}{\dot{m}_{FCOLD} * o} + X_{TEST,W,NaOH} \quad (37)$$

Using the relationship for calculating  $\dot{m}_{F25N} + \dot{m}_{FT25}$  found in equation 28 may be implemented into equation 37.

$$x_1 = \frac{\dot{m}_{FNAOH} * (X_{FNAOH,NaOH} - X_{TEST,W,NaOH}) - \left(\frac{\dot{m}_{FCOLD*o}}{x_3+1} - \dot{m}_{FNAOH}\right) * X_{TEST,W,NaOH}}{\dot{m}_{FCOLD} * o} + X_{TEST,W,NaOH} \quad (38)$$

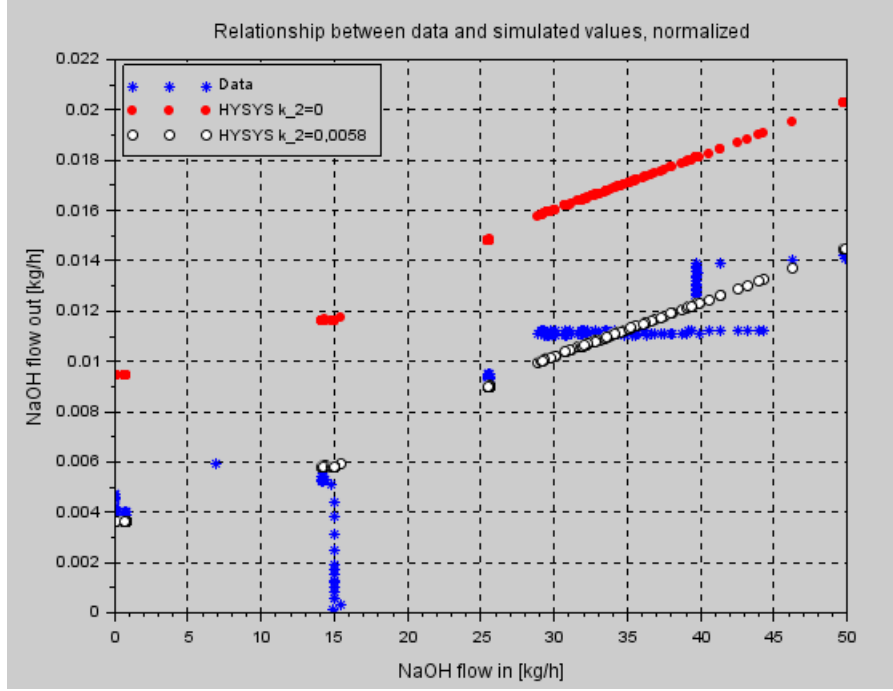


Figure 3: Matching the simulated data to historical data. Values are scaled by subtracting the lowest value in each axis.

After the use of algebra on equation 38 a final expression for calculating  $x_1$  is constructed and can be found in equation 39

$$x_1 = \frac{\dot{m}_{FNAOH} * X_{FNAOH,NaOH}}{\dot{m}_{FCOLD} * o} - \frac{X_{TEST,W,NaOH}}{x_3 + 1} + X_{TEST,W,NaOH} \quad (39)$$

Rearranging equation 39 to get an expression for  $\dot{m}_{FCOLD} * o$

$$\dot{m}_{FCOLD} * o = \frac{\dot{m}_{FNAOH} * X_{FNAOH,NaOH}}{x_1 + \frac{X_{TEST,W,NaOH}}{x_3+1} - X_{TEST,W,NaOH}} \quad (40)$$

With an expression for  $\dot{m}_{FCOLD} * o$  in equation 40 and a relationship containing  $\dot{m}_{FNAOH}$  in equation 28, results in the following expression for  $\dot{m}_{FNAOH}$

$$\dot{m}_{FNAOH} = \frac{\dot{m}_{FCOLD} * o}{x_3 + 1} - (\dot{m}_{F25N} + \dot{m}_{FT25}) = \frac{\frac{\dot{m}_{FNAOH} * X_{FNAOH,NaOH}}{x_1 + \frac{X_{TEST,W,NaOH}}{x_3+1} - X_{TEST,W,NaOH}}}{x_3 + 1} - (\dot{m}_{F25N} + \dot{m}_{FT25}) \quad (41)$$

Since both sides contain  $\dot{m}_{FNAOH}$ , a final expression for  $\dot{m}_{FNAOH}$  can be derived and seen in equation 42

$$\dot{m}_{FNAOH} = \frac{\dot{m}_{F25N} + \dot{m}_{FT25}}{\frac{X_{FNAOH,NaOH}}{X_{TEST,W,NaOH} + (x_3+1)(x_1 - X_{TEST,W,NaOH})} - 1} \quad (42)$$

By guessing a value for  $X_{TEST,NaOH}$  and putting  $\dot{m}_{F25N} + \dot{m}_{FT25}$  at a constant value, the value of both  $\dot{m}_{FNAOH}$  and  $\dot{m}_{FCOLD} * o$  can be known before the experiments are conducted

making it easier to monitor. The only problem is that the  $X_{TEST,NaOH}$  guess has to be verified and does not take temperature into account.

When it comes to calculate the in concentration of VAL from the combination of FT25 and F25N, a simple mass balance is utilized. The formula for this can be found in equation 43.

$$X_{F25N+FT25,specie} = \frac{\dot{m}_{F25N+FT25,specie}}{\dot{m}_{F25N} + \dot{m}_{FT25}} = \frac{X_{SP1,specie} * \dot{m}_{F25N} + X_{SP4,specie} * \dot{m}_{FT25}}{\dot{m}_{F25N} + \dot{m}_{FT25}} \quad (43)$$

### 3 Methodology

A combination of literature review, regression tools, experimental design tools and project planning was used to achieve the aim of this thesis. After the literature review, summarized in section 2 of the thesis, an attempt to create a model for the system was made by working with historical data. Then the project moved on to full scale experimental planning and executing these plans, see section 2.6 for information about the factors.

#### 3.1 Historical data

Relevant historical data was gathered once every hour for the past two years, see appendix B for what points was gathered. Since there was no concentration data for these points a linear correlation between the different concentration measurement points were made to get more data points. This gave about 25000 data points. The data was processed to get rid of stray points or instances where the plant was not running at full capacity to help get more efficient data. The data was then processed in the computer software Simca to try and performed a multivariate data analysis, MVDA [15, 26]. However the data proved to have low linear dependence between the different factors and the response which was combated with tighter restrictions on the collected data. At the end only 301 data points remained, see appendix D for restrictions, and still there were no positive Q2 values from any of the linear models in Simca.

Since Simca is quite restrictive in what terms it allows into its model and the system appeared to be nonlinear a self-made data regression tool was constructed using the computer software Scilab [27]. In this software the factors of the 301 data points were ranged scaled with equation 44.

$$xn_i = \frac{x_i - \bar{x}_i}{\max(x_i) - \min(x_i)} \quad (44)$$

Where  $i = 1, 2, 3$  represent the 3 factors investigated in this thesis, see section 2.6. With a combination of regression tools found in section 2.5.1 and the model evaluation tools found in section 2.5.2, an attempt to fit the data to a nonlinear model was conducted, but was not completely successfully.

#### 3.2 Preparations for full scale experimental planning

The fact that the regression model failed to fully represent the data and represented data also existed outside the region of interest, especially for  $x_1$  where a lower inlet concentration was desired, showed the need for a more controlled set of experiments to create a model for this system. The lack of data in the region of interest together with insufficient information about the reaction coefficients for the reactions rates in equation 1 to 8 lead to that simulations of the system in Aspen Plus would be of lesser value [28]. The hypothesized threshold nature for lower  $x_1$  and the fact that data was lacking for the region of interest for  $x_2$  and  $x_3$  at this low  $x_1$  values, a one-factor-at-a-time method was proposed for  $x_1$  and  $x_2$  [16]. The reason for not using one-factor-at-a-time method for  $x_3$  is to limit the number of experiment, because of the hypothesized smaller impact of this factor which lead to that a large interval could be utilized as long as it contains the optimal working point.

But if time permits use the one-factor-at-a-time method here too.

The hypothesized threshold nature of  $x_1$  also brings other problems when conducting experiments on a full scale plant. If the hypothesized threshold was found and surpassed then the potential losses in product in the form of increased LE production leads to the need to mitigate potential loss in revenue. All systems have a residence time,  $\tau$ , which is calculated with equation 45 .

$$\tau = \frac{\dot{m}}{\rho * V} \quad (45)$$

For the recirculating part of the plant, MIX2 to DEC1 in figure 1, the flow was calculated to flow through 9 to 10 times before steady state could be assumed. The residence time where therefor put to about 45 to 50 min, while the decanters DEC2 and DEC3 had a calculated residence time of about one and a half hour each. The combined residence time then becomes just under 4 hours. The risks of having that high of a residence time is that a lot of potential LE created because of low  $x_1$  values and it also limits the capacity of the number of experiments executed in a day, so the possibility to install an alternative test site was investigated.

Examples of different points that was explored was a valve on the heat exchanger, COOLER. It existed right after the reactor which had advantages but unfortunately the valve was not located on the ground level which made it inconvenient to flush and extract measurements there. So a valve on a heat exchanger after the circulation was chosen since it was on the ground level. By moving the test site to before DEC2 and DEC3  $\tau$  in between the different experiments are only put to 45 to 50 min.

When installing this temporary test site some steps needed to be taken. First of all when proposing a project like this a formal request for the money was needed. In this request a calculation of potential savings for the company was created and a risk assessment was conducted to spot potential danger with this project. When constructing the test site it was important to calculate capacity of the cooling water to lower the temperature of the sample leaving the test site. This was done with an expression according to appendix E. Then the project proposal was discussed with the investors before it was given a green light. Once the project was approved the more practical part of building TEST started. With the consulting of technical designers, instrumentation DCS engineers the setup starting at the valve connected to an flange, connected to a tube pipe which moves on to a heat exchanger leading to a needle valve, was agreed upon.

Installing a new test site, TEST, before the decanters came at a price. The sample included both NaOH and organic compounds, which presented a problem with handling of the waste. There existed no prior way of collecting this type of waste at the site so a cipax container was installed so the waste could be collected and after all the experiments where finished transported to the treatment plant [29]. The sample extracted from TEST contain both a water and an organic phase, presented a problem since both phases needed to be analyzed. Since highly polar substances like NaOH can destroy the column used when performing gas chromatography, GC, the two phases had to be separated. If



it separated for too long a period of time then the test could have been performed at SP5 and SP6 instead so a balance needed to be met. A new set of handling the analysis of the samples from TEST was developed with the collaboration of the laboratory workers. By comparing different separation times, flash volumes and also how to extract the necessary parts of the samples a standard for handling samples was created.

By flushing about 3 dm<sup>3</sup> before extracting two samples of 250 ml each a better representation of the concentration of the flow was gained. These two samples were allowed to separate for 10 to 15 min before a pipet was used to extract a small amount of top layer to be analyzed in the GC. When that were done the sample was put in a separation colon for 20 min, before 10 ml of the bottom part of the water phase was extracted to be analyzed with the help of titration, where the NaOH and organic concentration was measured.

Before tests could be performed on the plant, a risk assessments needed to be conducted to catch the potential risks when the values of the factors are changed. It is mostly about how to monitor for different changes in the plant and also how the plant handles added HE and LE production. Increased HE production is handled in the evaporators, EVA1 and EVA2, while increased LE production is handled in the LE column, LECOLUMN, from figure 2.

### 3.3 Full scale experimental plan

Once all the preparation was conducted an experimental plan had to be formulated. Since data was missing for lower values of  $x_1$  especially closer to the hypothesized threshold, and other parameters had larger fluctuation, especially  $x_2$ , see section 4.1 for more information. This combined with the hypothesis that  $x_3$  might have a lower effect on the results together with a larger region of interest made  $x_1$  a logical first factor to evaluate.

The inflow was put at a value that represented full production, but after experiment 8 the inflow was increased by 0.3ton/h after wishes from Perstorp. This do not seem to have any larger impact on the results obtained in this thesis.  $x_2$  was put to constant values decided by maximum values found when doing the regression analyzed explained in section 3.1.  $x_3$  remained fairly constant but changed a little during the experiments since it was quite hard to get a constant value on  $\dot{m}_{FCOLD} * o$  during normal production and also because it only changed a little for each change in  $x_1$ . Then  $x_1$  was put on a value included in normal operating conditions. The feed concentration was measured at SP1 and SP4 once a day and then kept under mechanical observation during the experimental run to ensure as constant operating condition during the day as possible, see section 4.2 for more information.  $x_1$  that was calculated according to the equation 39 was lowered by first 0.1% then 0.05% until the hypothesized threshold nearly was reached, then a 0.02% decrease was used to avoid to much LE production. But when the concentration where the hypothesized threshold supposed to exist was passed no threshold had been found, see section 4.2 for more information.

Since time was of the essence and the desire to find the most optimal  $x_1$  value still existed, a new plan was used, lowering the NaOH feed, FNAOH from figure 1, by 5kg/h until a

threshold was reached. But as can be seen in section 4.2 there existed no threshold and the calculated  $\tau$  from section 3.2 turned out to be too low and instead an increased LE production started during the night. This unfortunately led to the plant going off spec until morning instead of increasing the  $x_1$  factor. Some of the lower  $x_1$  values had to be retried, but this time they were tried after 4 and 12 hours in both SP5, SP6 and TEST. Purity test were extracted from F2-PH in between these different tests, combined with more extensive instruction, both written and verbally, about how to avoid this from happen again. These instructions emphasized the importance of a more rigorous oversight of the LECOLUMN during the time I was unavailable as well as increasing  $x_1$  when problems arose.

For a more complete summary of the experiments conducted see appendix F. This appendix include the continued optimization of the plant including the evaluation of factor  $x_2$  and the CCO or CCC investigation, see section 2.5.3 for more info about CCO and CCC. These eA xperiments could not be performed due to time constrains but a plan for how to conduct them have been formulated, see section 4.3 for more detail.

## 4 Results and discussion

Below follows a presentation of the results obtained in this thesis. Because of secrecy reasons all the values have been ranged scaled or scaled by subtracting the lowest value in each figure and axis. See appendix A for a abbreviations list and see figure 1 and 2 for the plant layout.

### 4.1 Evaluation of historical data

The attempt to fit the historical data to a nonlinear model in equation 46 can be seen in figure 10. The values of  $\theta_h$ , the residuals, as well as the values of the  $t_{obs,h}$ , from equation 21, can be found in appendix H. As can be seen by comparing table 8 and 1, not all  $t_{obs,h}$  follows  $t_{obs,h} > t_{\alpha/2,np}$ , but without these factors the model would look even worse and the null hypothesis described in section 2.5.2 could therefore not be omitted. This makes the calculation of  $F_{obs,LoF}$  invalid, since even if  $F_{obs,LoF} > F_{\alpha,mp,nm}$ , the model would still be incorrect. When looking at figure 4 it is clear that the lack of fit is significant and the model needs to be changed to accommodate this.

$$\begin{aligned} \hat{y} = & \theta_1 - \theta_2 * xn_2 - \theta_3 * (xn_1 + 0.2)^2 - \theta_4 * xn_1 - \theta_5 * xn_1 * xn_3 + \theta_6 * (0.3)^{xn_2} + \theta_7 * \log_{10}(xn_2 + 1.3) + \\ & \theta_8 * xn_3 + \theta_9 * xn_2 * xn_3 * xn_1 - \theta_{10} * xn_1^2 * xn_3^2 - \theta_{11} * \exp(-4 * xn_1 + 1) * \sin(8 * xn_1 + 3) - \\ & \theta_{12} * \cos\left(\frac{\pi * xn_1}{4}\right) + \theta_{13} * xn_1^2 * xn_3^2 + \theta_{14} * (xn_3 + 0.2)^2 - \theta_{15} * \exp(2 * xn_1) * \sin(10 * xn_1) + \\ & \theta_{16} * (xn_1 + 0.35) * (xn_1 - 0.35) + \theta_{17} * \sin(8.975 * xn_1) - \theta_{18} * \sin(15 * xn_1) \end{aligned} \quad (46)$$

The so called bad time parts comes from the bad fit between the model and the data in figure 4D) around a scaled time value of 200. This is to see where this part of the model exists when plotted against the factors. By doing this some of the largest outlines can be attributed to these data points. As can be seen in the model displayed in equation 46, time differences are not taken into account which means something else must be responsible for these outlines. The fact that these points are gathered together like this shows that these outlines have a common error factor. When evaluating these data more closely the only apparent reason is a smaller amount of 3-mbal inflow at these points. This could explain the results since 3-mbal reacts faster than both 2-mbal and NVAL, so if it is decreased then the total PBA concentration might become lower too. This might account for a large part of the significant lack of fit discussed above.

When evaluating figure 4.A) Y seems to be unaffected by  $xn_1$  for the investigated region. The problem is that the region of interest exist just before the hypothesized threshold and this value could not be seen here. Excluding the bad time part from the model results in a quite reasonable fit for the model. Something could have been done to fix the higher values of  $x_1$ , but as can be seen in equation 46 the expression started to become really long and for every thing added another error seemed to take its place. The model displayed in equation 46

Figure 4B) has an passable resemblance between model and the data, especially if the bad time parts are omitted. This is unexpected since the data had such an irregular shape, especially in the  $xn_2 = -0.2$  which is where the maximum value seems to be located. Figure 4C) has a similar shape as 4A) with a maximum value around  $xn_3 = -0.3$ . This figure also has more compact clusters of data which might explain the wired shape of the model at  $xn_3 = 0.2$  as the lower values are a result of the models handling of values around  $xn_1 = 0.4$  and not an effect off the treatment of  $xn_3$ .

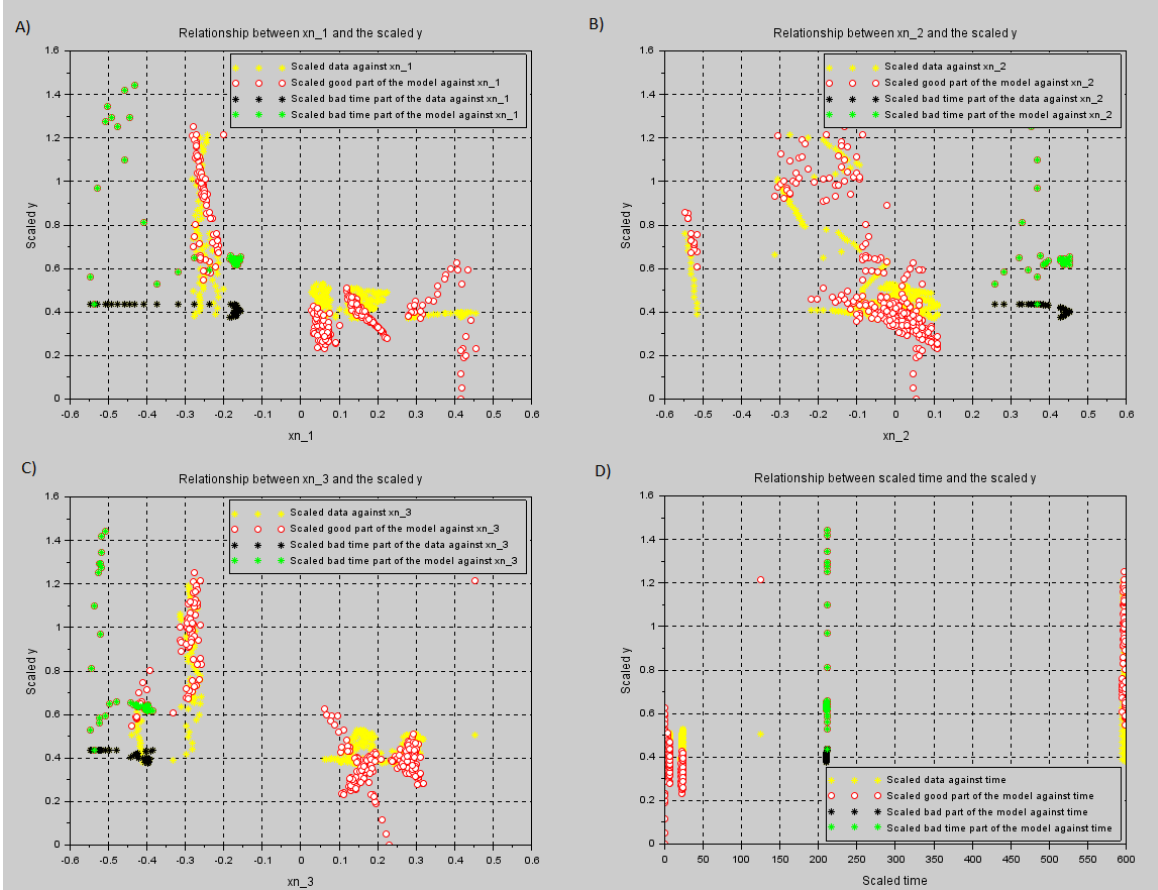


Figure 4: Matching the historical data of PBA concentration against A) NaOH concentration into the reactor, B) temperature, C) recirculation ratio and D) time with calculated ones. Values are either ranged scaled or scaled by subtracting the lowest value in each axis.

Another problem with the model of the historical data is that the correlation matrix,  $C_{gj}$  found in equation 19, are also close to either 1 or  $-1$  which can be seen in equation 47 below. This is a further indication that the model is wrong and it need more work before it can be used.

$$C_{gj} = \begin{bmatrix} 1 & -0.9999 & -1 \\ -0.9999 & 1 & 1 \\ -1 & 1 & 1 \end{bmatrix} \quad (47)$$

Overall the results from this historical data analysis can be summed up by that the model found in equation 46 seems to be a bad fit for the data. The model especially have a hard

time calculating the values of the scaled time of around 200, but even with these points omitted it would still not result in a perfect model. But it gives a good indication of where to put the factors  $x_2$  and  $x_3$  as  $x_1$  is lowered, as well as strengthen the hypothesis that the relationship between  $x_1$  and  $y$  is of a threshold nature and that the reaction rate seen in equation 1 to 6 are correct.

## 4.2 Evaluation of the full scale experiments

When evaluating the full scale experiments it is appropriate to evaluate what goes into the system first. It is important to remember that the NaOH concentration is not lowered by the same amount between each experiments. The exact amount of NaOH for each experiment can be found in appendix F. In figure 5A) the NaOH concentration in FRECI is depicted. As the experiments continues both the measured and calculated concentration of NaOH goes down because of the decrease of FNAOH, but when the inflow data for experiment 18 was used during the night the calculated residence time was deduced to have been too low, resulting in the area going into off spec during the night as mentioned in section 3.3. After this incident some of the lower inflows of NaOH had to be retried with a longer residence time. This seemed to lead to a closer fit between the data of SP6, TEST and the calculated results with  $k_2 = 0.0058$ . After getting back the results from experiment 19, where the PBA concentration was on an all time low, see figure 6C), the inflow of NaOH was increased to the final value and tried a total of 3 times before it was concluded to be the optimal operating concentration.

Equation 33 fits the SP data points depicted in figure 5A) quite well when  $k_2 = 0.0058$ , but the constant might have been a little large since the SP data points mostly exist above the calculated values. This combined with the fact that the difference in SP and TEST measurements are not constant might depend on two reasons. One is the fact that the residence time of the system seemed to be larger then anticipated but that would not fully explain why the difference between the calculated values and the measured TEST values changes as they do. The second reason why the difference is not constant might be because the lower total NaOH concentration in the system might have affected chemical reaction 1 to 6 so that the water elimination step occurs faster, which frees a lot of NaOH that previously was bound in aldols and did not separate in the decanters. If NaOH is bound for less time in each reaction then each NaOH molecule can be used to produce more PBA. It can also be separated easier in the decanters and will not follow along bound to organic compound in FORGANIC. Instead it can be measured in the water phase either in SP6 or TEST. The decreased amount of NaOH in FORGANIC leads to less NaOH in the evaporators where it clogs the HE separation streams.

Figure 5B) shows how the VAL and NVAL concentration changed during the experiments. While the overall VAL concentration is mostly constant the NVAL is not. Especially at experiment 14 and 15 where the majority the inflow was taken from the TANK. This resulted in 0.005 mass% lower NVAL. In figure 5C) this decrease in NVAL concentration can be seen being supplemented by a proportional increase in 2-mbal and a small increase in 3-mbal. Figure 5D) on the other hand that shows some of the by-products going into the reactor remains quite constant. The only small difference is the Pentanol concentration

that changes a little. This change do not seem to be significant enough since the over all VAL concentration is still the same. That means that this Pentanol difference occur while some other by-product that is not measured with the GC is decreased.

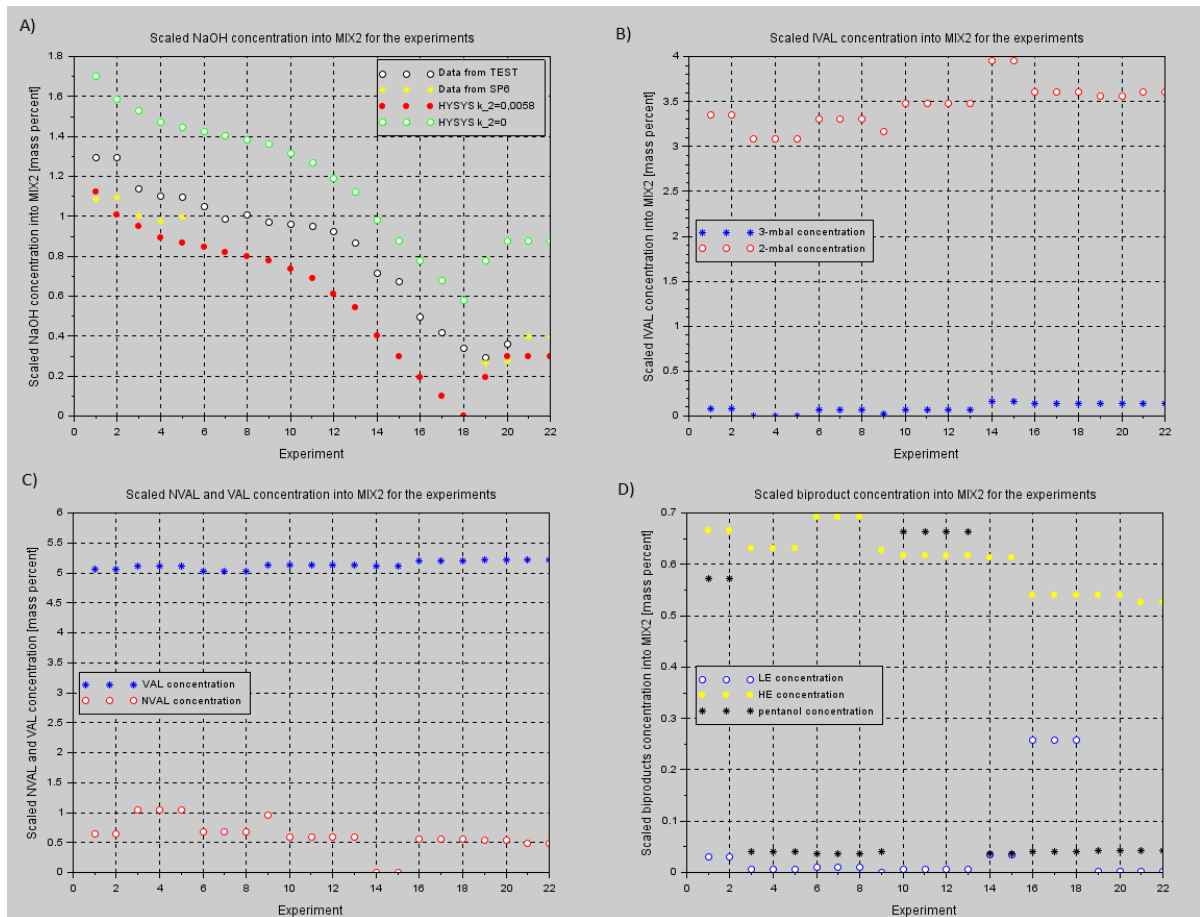


Figure 5: Scaled mass percent of what enters the reactor for the different experiments. A) Estimated inflow of NaOH in FNAOH, both calculated and measured B) IVAL, C) NVAL and VAL, D) LE, HE and pentanol. Values are scaled by subtracting the lowest value in each axis.

In figure 6 the outflow concentration measured at both TEST and SP6 for the experiments is displayed. The first thing to note when examining figure 6B) is that the increase in NVAL does not start in just one point, but is something that happens gradually as the NaOH concentration is decreased. The highest concentration of NVAL can be seen in experiment 19 and not 18 which is the lowest inflow of FNAOH. That is another reason to believe the assigned residence time was insufficient, because when experiment 19 was performed, a longer residence time was used which resulted in a higher NVAL concentration. The change in inflow relationship in experiment 14 and 15 where more inflow was taken from the tank can not be fully seen even if experiments 15 could have been a little higher to follow the overall trend.

In figure 6A) the effect of the increased tank inflow can be seen as the 2-mbal concentration is high for experiment 14 and 15, but the decrease of NaOH concentration do not seem

to have effected the 2-mbal concentration nearly as much as can be seen in experiment 16 to 22. This indicates that something else is effecting the reaction 2 and 6. The same relationship can be seen in figure 6C) where the PBA concentration goes down by over 2 mass percent, which is 4 times the loss that the decreased NVAL concentration, seen in figure 5B), should result in if everything else was equal. The decrease in PBA concentration at experiment 19 is probably as discussed earlier because of the increased residence time for these measurements, but the result does not feature the hypothesized threshold but instead displayed a more gradual decrease. This is probably because the FRECI is much larger then F25N+FT25 and the fact that the reaction rate seems to increase when the NaOH concentration is lower. When it comes to figure 6D) the HE concentration decreases linear as expected when looking at the rate expression for HE seen in equation 7 and 8, but there do not seem to be any difference regarding where the inflow comes from.

The large decrease in PBA concentration at experiment 14 and 15 could be explained by the fact that the temperature in FT25 compared to F25N is lower. This can lead to a less efficient reaction in the pump in FCOLD, where the increased mixing might increase the reaction rate. But more about that in section 4.3.

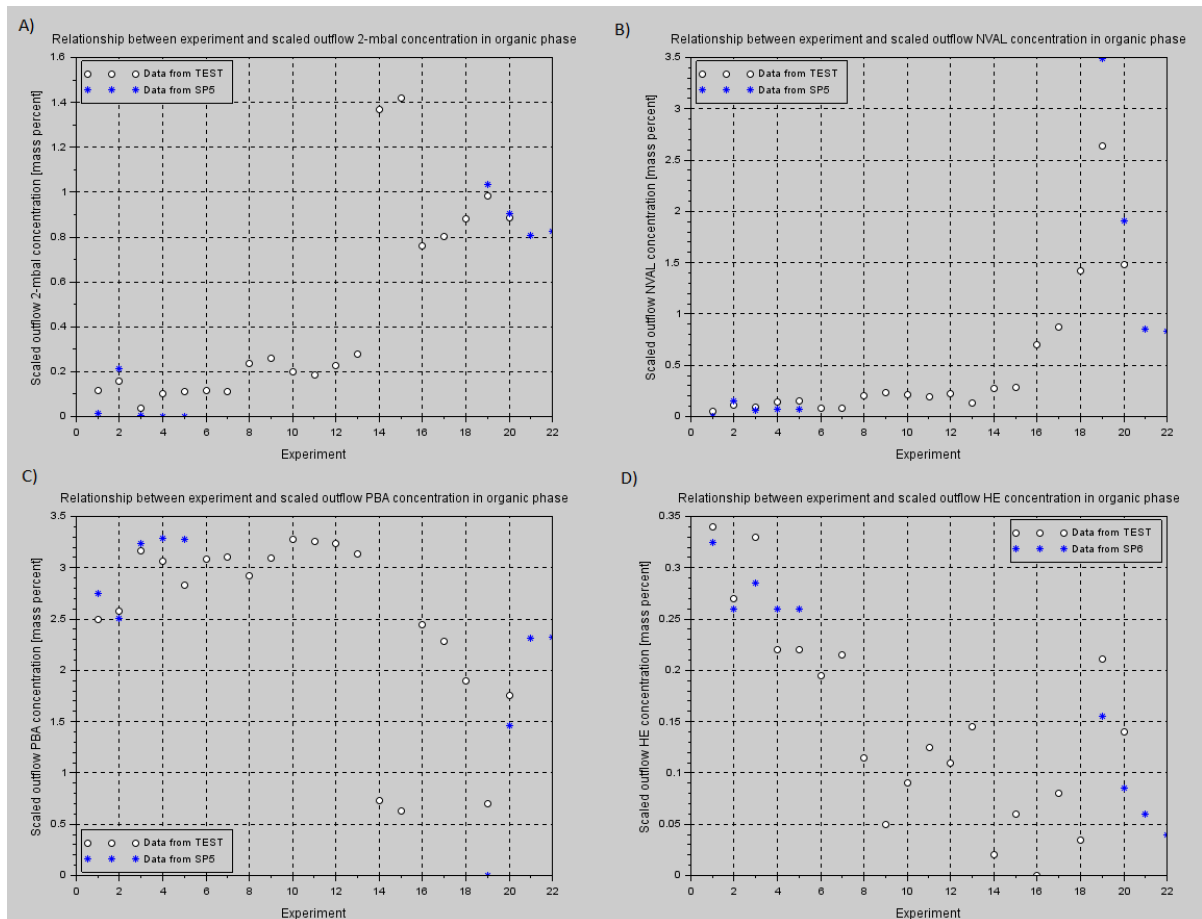


Figure 6: Scaled mass percent of what leaves the DEC1, in the different phases, for the different experiments measured both in SP and TEST. A) 2-mbal, B) NVAL, C) PBA, D) HE. Values are scaled by subtracting the lowest value in each axis.

To get a better understanding of how much  $x_1$  was lowered during the experiments, figure 7A) shows the  $x_1$  with the NaOH concentration in FRECI being equal to the measured NaOH concentration in TEST and SP6 and then calculated with equation 39.  $x_1$  was lowered by almost 0.7 mass% which is nearly double of the expected decrease. This resulted in the need for an increased amount of experiments to determine this factors optimal working region, but the relationship between experiments and  $x_1$  seems to be mostly linear when plotted against the experiments but since the experiments lowered the  $x_1$  with a different amounts, as can be seen in figure 7B) which reveals a more second order relationship between  $x_1$  and  $y$  and not the hypothesized threshold nature. This means that the reaction rate is closer to the way described in the literature and not the RD report hypothesized. This result also means that large amount of NaOH was wasted each year which resulted in problems down the product line especially in the evaporators, EVA1 and EVA2, where a larger proportion of the product had to be used to keep the bottom of the evaporators from clogging. The two outlines in figure 7 are the once where most of the VAL is taken from FT25 which have been discussed above.

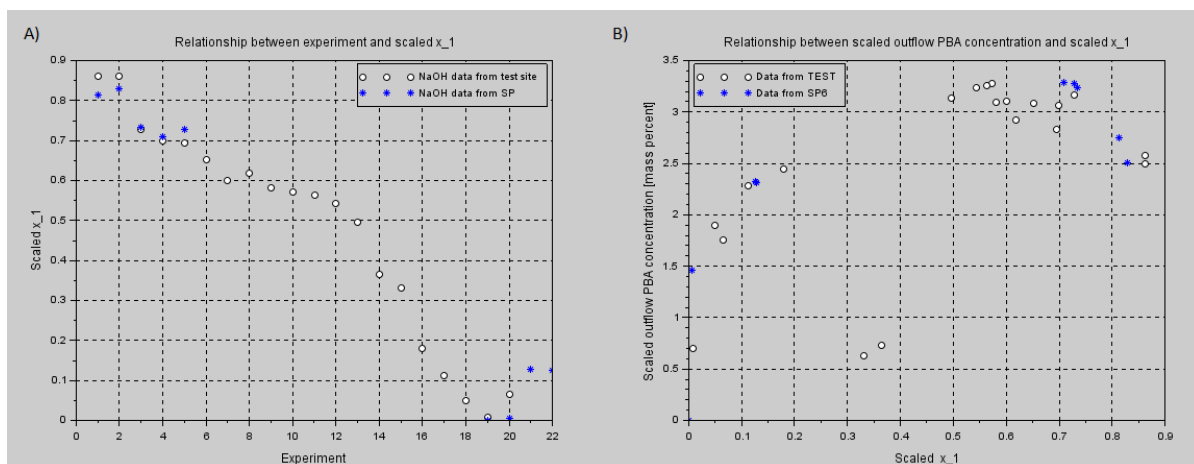


Figure 7: A) Relationship between the calculated NaOH concentration into the reactor, with NaOH in FRECI being measured in both SP6 and TEST, and A) the different experiments, B) PBA concentration. Values are scaled by subtracting the lowest value in each axis.

### 4.3 Evaluation of continued plan

The increased amount of experiments needed to determine the optimal working interval for  $x_1$  and instability at the site, left, because of time restrictions, other factors than  $x_1$  unexplored. To continue on with the optimization of this system temperature,  $x_2$ , is the next factor that should be tried. It should not require to many experiments to find an optimal working interval, since  $x_2$  could be increased by 4 °C to see if there is any difference, then if Y is increased continuing by increasing by 4 °C more until a maximum can be reached. If an optimal working interval for  $x_2$  is not found by increasing with 4 °C from the original value with the same amount as it was increased should be investigated. It is important to check the temperatures of the pumps during the time  $x_2$  is increased. If it becomes necessary to increase the temperature more the pressure could always be increased, also keeping in mind that an increased residence time of at least 2 hours for TEST and 5 hours



for SP6 should be used.

The  $x_3$  value might also need to be tested to investigate effect of the  $o$  parameter of equation 27 on  $y$ , especially when  $o$  is set to 0.9 while  $x_1$  is put to its lowest value, since a lower value of  $x_3$  also gives a lower value of  $x_1$  according to equation 39 which could force the plant to go off spec again. If this is the case increase the original point, 0, for  $x_1$  so that the plant at least do not have to go off spec. It might also be interesting to increase  $x_3$  and see if that help to reach the higher historical values achieved when analyzing historical data in section 4.1. It might be that the effect seen in figure 4 is not just because the concentration of NVAL and 3-mbal in the tank are higher then during the experiments.

Lastly perform experiments according to the CCO if one-step-at-a-time is not used on  $x_3$  and use CCC if it is preformed. CCC and CCO is explained in section 2.5.3 and evaluate the result using Modde GO [25]. The reason for using CCC is because there where no hypothesized threshold nature between  $x_1$  and  $y$ , and if a if the region of interest can be specified more accurate then the rotatability of a CCC model is preferable. In Modde GO a somewhat nonlinear model can be created to fit the results from the experiments. This model will then have a maximum value which corresponds to the optimal operating conditions for the three factors. When doing this it is important to wait for lab results since equation 36 turned out to be somewhat unreliable when approximating  $X_{TEST,W,NaOH}$ . So to make sure of this a measured value of NaOH should be used instead of calculating  $x_1$ .

## 5 Conclusions

The most important results discovered when doing this work is that even if the reactions to create PBA seems to be independent of NaOH at higher concentrations, the threshold relationship does not occur when NaOH concentration is lowered. So instead of the hypothesized threshold relationship between  $x_1$  and  $y$  the yield instead begins to gradually become lower after a certain concentration of NaOH. It is impotent to remember that  $x_1$  is dependent on  $x_3$ , and  $x_3$  need to be further investigated.

The temperature of the inflow, F25N and FT25, seems to have a larger effect than the NaOH concentration as long as there are enough for the reaction to occur. Which leads to the conclusion that heat could be taken from a heat exchanger located at FDEC2, where the product flow is cooled by utilities. Some of this wasted heat could be used to preheat the inflow. Since alot of the reactions seem to occur in the pump located at FCOLD.

Lastly trying to follow the plan laid out in section 4.3 to achieve a more complete optimization of the site. Might even want to consider using a one-factor-at-a-time approach when finding the optimal value of  $x_3$ . Just remember to recalculate  $x_1$  with the new value. Unfortunately as can be seen when trying to create a model with historical data this might be hard even with a CCD approach.

## Bibliography

- [1] Knovel [Internet]. *Knovel Critical Tables*, 2nd edition, 2003, <https://app.knovel.com/> (cited: 2018-01-29).
- [2] Nilsson, A. Sjölund, E. *2-Propylheptanol*. <https://www.perstorp.com/products/2-propylheptanol> (cited: 2018-01-24).
- [3] Ouni, T. Nikkilä, K. Lahtinen, L. Eriksson, J. *Production of 2-propylheptanol from valeraldehyde*. Confidential in house report written in 2007, about results obtained from experiments conducted on a pilot plant of the production plant of interest for this report.
- [4] Andersson A. *Anläggningsbeskrivning för area 25 (2PH)*. Confidential in house report written 2017 about general information about the site of interest.
- [5] Seader, J D. Henley, E J. Roper, D K. *Separation process principles: chemical and biochemical operations*. 3rd. John Wiley Sons, Inc., 2011.
- [6] McMurry, J. *Fundamentals of Organic Chemistry*. 4th. Brooks/Cole Publishing Company, 1984.
- [7] Clayden, J. *Organic chemistry*. 1st. Oxford Univ. Press, 2001.
- [8] Allinger, N L. *Topics in Stereochemistry*. Vol. 13. John Wiley Sons, inc, 1982.
- [9] Kaizik, A. Fridag, D. Lueken, H. Bueschken, W. "Process for preparing alpha, beta-unsaturated C10-aldehydes". US8581008 B2 (US). 2012.
- [10] Madan, R L. *Chemistry for degree students*. 1st. S. Chand company PVT. LTD., 2011.
- [11] Madan, R L. *S. Chand's success guides (Questions and answers) Organic chemistry for B.Sc. 1, 2 and 3 year*. 1st. S. Chand company PVT. LTD., 2001.
- [12] Kerr, J A. Drew, R M. *Handbook of Bimolecular and Termolecular gas reactions*. Vol. 3.A. CRC Press, Inc, 1987.
- [13] Olbergts, J. "Termolecular reaction of nitrogen monoxide and oxygen: A still unsolved problem". In: *International journal of chemical kinetics* 17 (8) (Aug. 1985), pp. 835–848.
- [14] Baer, T. Hase, W L. *Unimolecular Reaction Dynamics: Theory and Experiments*. New York: Oxford University Press, 1996.
- [15] Esbensen, K. Guyot, D. Westad, F. Houmøller, L P. *Multivariate data analysis : in practice : an introduction to multivariate data analysis and experimental design*. 5ed. CAMO ASA, 2001.
- [16] Montgomery, D C. *Design and analysis of experiments*. 8th. John Wiley Sons, Inc, 2013.
- [17] Golbraikh, A. Tropsha, A. "Beware of  $q^2$ !" In: *Journal of Molecular Graphics and Modelling* 20 (4) (Jan. 2002), pp. 269–276.
- [18] Hawkins, D M. Basak, A C. Mills, D. "Assessing Model Fit by Cross-Validation". In: *Journal of chemical information and computer sciences* 43 (2) (Mar. 2003), pp. 579–586.
- [19] Sen, A. Srivastava, M. *Regression Analysis Theory, Methods, and Applications*. 1st. Springer-Verlag New York Inc., 1990.
- [20] Rasmuson, A. Andersson, B. Olsson, L. Andersson, R. *Mathematical Modeling in Chemical Engineering*. 1st. Cambridge University Press., 2014.
- [21] Khuri, A I. Mukhopadhyay, S. "Response surface methodology". In: *Wiley interdisciplinary reviews. Computational statistics* 2 (2) (Mar. 2010), pp. 128–149.

- [22] Box, G E P. Behnken, D W. “Some New Three Level Designs for the Study of Quantitative Variables”. In: *Technometrics* 2 (4) (Nov. 1960), pp. 455–475.
- [23] Box, G E P. Wilson, K B. “On the Experimental Attainment of Optimum Conditions”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 13 (1) (1951), pp. 1–45.
- [24] Verseput, R. [Internet]. *Digging Into DOE*, <https://www.qualitydigest.com/june01/html/doe.html> (cited: 2018-03-28).
- [25] Sartorius Stedim Biotech. *Modde GO*. <https://umetrics.com/product/modde-go> (cited: 2018-01-29).
- [26] Sartorius Stedim Biotech. *Simca*. <https://umetrics.com/products/simca> (cited: 2018-05-29).
- [27] Scilab Enterprises. *Scilab*. <https://www.scilab.org/> (cited: 2018-05-29).
- [28] Aspen tech. *Aspen Plus*. <http://home.aspentech.com/products/engineering/aspen-plus/>(cited: 2018-01-29).
- [29] CIPAX. *cipax*. <http://www.cipax.se/industri> (cited: 2018-05-29).
- [30] Mörtstedt, S. Hellsten, G. *Data och Diagram, Energi- och kemitekniska tabeller*. 7st. Liber AB., 2012.
- [31] Welty, J R. Wicks, C E. Wilson, R E. Rorrer, G L. *Fundamentals of Momentum, Heat, and Mass transfer*. 5st. John Wiley Sons, Inc., 2008.
- [32] Engineering department Crane Co. “Flow and fluids through valves, fittings and pipes”. In: (2009). Technical paper No. 410.



## A Abbreviations list

In tabel 3 follows a list of all the abbreviations used in this report.

Table 3: Abbreviations in alphabetic order.

| Abbreviation | Meaning                           |
|--------------|-----------------------------------|
| 2-mBal       | 2-Methylbutanal                   |
| 2-PH         | 2-Propyl-Heptanol                 |
| 3-mBal       | 3-Methylbutanal                   |
| HE           | Heavy Ends                        |
| HP           | High Pressure reactor             |
| IVAL         | Iso-Valeraldehyde                 |
| LE           | Light Ends                        |
| NVAL         | Normal-Valeraldehyde              |
| PBA          | isomers of 2-propyl-heptenal      |
| VAL          | Valeraldehyde                     |
| VPH          | Vapour Phase Hydrogenation units  |
| $x_1$        | Factor one, NaOH concentration    |
| $x_2$        | Factor two, Temperature           |
| $x_3$        | Factor three, recirculation ratio |
| $y$          | Response, PBA concentration       |

## B Plant abbreviations list

Below follows a list of the flows in figure 1 that had flow meters, thermometer or pressure measurement spot used in this work.

Table 4: Flow meters, thermometer and pressure measurement spot used

| Flow      | Measurement equipment      |
|-----------|----------------------------|
| FREACIN   | Thermometer                |
| FDECREAC  | Thermometer                |
| FRECI     | Thermometer                |
| FDEC2     | Thermometer                |
| FVAPOR    | Pressure measurement spot  |
| F25N+FT25 | Thermometer and flow meter |
| FT25      | Thermometer and Flow meter |
| F25N      | Thermometer and Flow meter |
| FTANK     | Flow meter                 |
| FNAOH     | Flow meter                 |
| FCOLD     | Flow meter                 |
| FWATER    | Flow meter                 |

## C Definitions of regression terms

Table 5 shows definitions on some of the regression terms used in section 2.5 [15, 16].

Table 5: Definitions of regression terms

| Term                   | Definition  |
|------------------------|---|
| Object                 | All observations from one sample                                  |
| X-variables            | The different factors off the objects                             |
| Y-variables            | The different responses off the objects                           |
| $n$                    | Total amount of objects   |
| $p$                    | Total amount of parameters  |
| $k$                    | Total amount of independent x-variables                           |
| $q$                    | Total amount of parameters  |
| $i$                    | Unique sample observation   |
| $j$ or $g$             | $j$ th or $g$ th independent x-variable                           |
| $l$                    | Repeated sample observation                                       |
| $m$                    | Total amount of unique measurement points                         |
| $n_i$                  | Total amount of repeated objects in one measurement point $i$     |
| $h$                    | $h$ th parameter number   |
| $x$                    | independent variable  |
| $y$                    | response variable   |
| $\hat{y}_i$            | Fitted model  |
| $\bar{x}$              | Mean value of a variable  |
| $\text{Var}(x)$        | Variance of a variable or measurements of variable values spread. |
| $S_x$                  | Standard deviation of a variable                                  |
| $\text{cov}(x,y)$      | covariance or measure of variables linear association             |
| $r$                    | correlation or scaled, unit-less covariance measure               |
| Quantitative factor    | A factor describing measurements based on quantities              |
| Qualitative factor     | A factor describing measurements based on qualities               |
| $\epsilon$             | Model error   |
| $f$                    | Model function without error                                      |
| $SS_E$                 | Residual or error Sum of square                                   |
| $e$                    | Residual  |
| $\hat{\sigma}^2 = S^2$ | Variance of a Model and experimental responses                    |
| $\alpha_r$             | Axial distance  |
| $\epsilon$             | Error vector of between model and data                            |
| $C_{g,j}$              | Correlation matrix  |
| $se(\theta_h)$         | standard error  |
| $t_{obs,h}$            | Test statistic for the null hypothesis for parameter $h$          |
| $\theta_h$             | Parameter $h$   |
| $Q^2$                  | Prediction of Y or X variance                                     |
| $SS_{PE}$              | Sum of square pure error  |
| $SS_{LoF}$             | Sum of square lack of fit   |
| $MS_{PE}$              | Mean sum of square pure error                                     |
| $MS_{LoF}$             | Mean sum of square lack of fit                                    |
| $F_{obs,LoF}$          | Observable F-distribution   |



## D Data restrictions

As can be seen in table 6 where the final boundaries for the historical values are displayed. The boundary for  $\dot{m}_{FNAOH}$  might have been a little high and some of the temperatures might have been a little high too but this was the final data where 25000 points were lowered to 301 to try to make the data more linear. The y boundary might have been a little high too and are higher than what the experiments got. A reason why the value got up this high was not found but as stated in section 4.3 it might have been because of higher concentration into the reactor but since no data of the tank was available this could not be checked. No other obvious reason for it could be found. Because of secret reasons the numbers in table 6 will be omitted.

Table 6: Boundaries for the historical values

| measurement           | boundary | value   |
|-----------------------|----------|---------|
| $T_{FDEC2}$           | >        | – °C    |
| $T_{FDECREAC}$        | >        | – °C    |
| $T_{FRECI}$           | >        | – °C    |
| $T_{F25N+FT25}$       | >        | – °C    |
| $\dot{m}_{F25N+FT25}$ | >        | – ton/h |
| $\dot{m}_{FNAOH}$     | >        | – kg/h  |
| $\dot{m}_{FCOLD}$     | >        | – ton/h |
| $\dot{m}_{FWATER}$    | >        | –       |
| $P_{FVAPOR}$          | €        | –       |
| $X_{SP6-SP1,HE}$      | >        | –       |
| $x_1$                 | <        | –       |
| $x_2$                 | >        | – °C    |
| $x_3$                 | >        | –       |
| y                     | >        | –       |

## E Cooling capacity of the heat exchanger at TEST

Below follows the calculations used to ensure that the heat exchanger at TEST had enough cooling capacity to lower the samples temperature to around  $T_{H,out} = 45$  °C.  $T_{H,out} = 45$  °C are assumed to be safe working conditions for the operating technicians. The pressure was assumed to be lost in the valves and flange. Because of secretes reasons some of the numbers will be omitted.

The pipes are assumed to be of a rough nature of the length,  $L = 12$  m and have a diameter,  $d = 0.005$  m. The density of the sample,  $\rho_H = \_$  kg/m<sup>3</sup>, is an old value calculated using HYSYS when the plant was created and is assumed to be independent of pressure and temperature, since it is a liquid. The temperature inside the system is put to  $T_{H,in} = \_$  °C which is where the pressure needs to be increased, as can be seen in figure 9 in appendix G. Assuming the inflow temperatures of the cooling water  $T_{C,in} = 23$  °C. The viscosity of the stream,  $\mu_H = \_$  Pa s, and is assumed to be constant during the pipe, This value is taken from simulated values from HYSYS when the plant was built. The mean temperature of the cooling flow,  $\bar{T}_C = 45$  °C, and the pressure outside is set to 1 bar [30]. Density of the cooling water,  $\rho_C(\bar{T}_C) = 990.15$  kg/m<sup>3</sup>, heat capacity for the cooling water,  $cp_C(\bar{T}_C) = 4.176$  kJ/kg [30]. While the heat capacity for the warm stream,  $cp_H = \_$  kJ/kg, since it is assumed to be the same as for the stream is inside the plant, this value was calculated using HYSYS when the plant was created.

$$v_H = \frac{1}{\rho_H} \quad (48)$$

By assuming complete turbulent rough pips the relative roughness,  $e/D = 0.05$ , an iterative method where equation 49 to 52 was used to create a solution. [31, 32]

$$K = \frac{f_v * L}{D} \quad (49)$$

$$\dot{m}_H = \frac{\sqrt{\frac{\Delta P_H * (D * 1000)^4}{0.6253 * K * v_H}}}{3600} \quad (50)$$

$$Re = \frac{\dot{m}_H * 4}{\mu_H * D * \pi} \quad (51)$$

$$f_v = \frac{-1}{3.6 * \log_{10} \left( \frac{6.9}{Re} + \left( \frac{e}{3.7 * D} \right)^{10/9} \right)} \quad (52)$$

When a solution was reached the calculated Reynolds number,  $Re = 29620.13$ , with equation 51, which is a little lower then the range of equation 52, since the lower limit of the equation is  $Re = 4 * 10^4$  [31]. But since only an approximate value for the out temperature of cooling stream,  $T_{H,out}$ , was sought the calculated Re was assumed to be close enough to the allowed range. With mass flow of the heated stream,  $m_H = 0.0352$  kg/s, the total heat released by the system,  $Q_H$ , could be calculated with equation 53.

$$Q_H = \frac{cp_H * \dot{m}_H}{T_{H,in} - T_{H,out}} \quad (53)$$

Since  $Q_H$  have to be equal to the heat taken up by the cooling stream,  $Q_C$ , equation 53 can be used to calculate the mass flow of the cooling stream,  $\dot{m}_C$ .

$$\dot{m}_C = \frac{Q_C}{cp_C * (T_{C,out} - T_{C,in})} \quad (54)$$

$T_{H,out}$  are then evaluated at an interval between 30 °C and 60 °C and the result of how  $\dot{m}_C$  are changed by doing this can be found in figure 8.

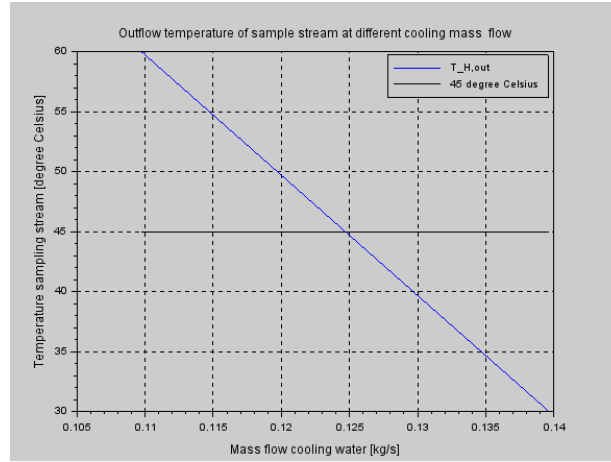


Figure 8: Temperature of sample stream at different cooling mass flows

To keep  $T_H = 45$  °C then  $\dot{m}_C$  have to be about 0.125 kg/s which since the diameter of the cooling pipe,  $D_C = 0.012$  m, making the inner cross section area,  $A_C = 7.85 * 10^{-5}$  m<sup>2</sup> calculated with equation 55

$$A_C = \frac{D_C^2 * \pi}{4} \quad (55)$$

This indicates that the heat exchanger at TEST should have more then enough cooling water to cool the sample stream since the cooling water are regulated to deliver colling water to a heat exchanger further down the pipe. The total cooling flow,  $\dot{m}_{C,tot}$ , was therefor calculated with a heat balance over that heat exchanger at FDEC2 but at lower than full production, with  $\dot{m}_{F25N+FT25} = \_$  ton/h, since some thermostats are only local historical values could not be used, see equation 56 for the calculation. The heated product stream temperature into the heat exchanger,  $T_{H,tot,in} = \_$  °C, heated temperature out of the heat exchanger,  $T_{H,tot,out} = \_$  °C, heat capacity of the heated stream,  $cp_{H,tot} = \_$  kJ/(kg \* K), the cool stream outflow temperature,  $T_{C,tot,out} = 40$  °C.

$$\dot{m}_{C,tot} = \frac{\dot{m}_{F25N+FT25} * cp_{H,tot} * (T_{H,tot,in} - T_{H,tot,out})}{cp_C * (T_{C,tot,out} - T_{C,in})} \quad (56)$$

The diameter of the total cooling pipe over the heat exchanger,  $D_{C,tot} = 0.048$  m, making the inner cross section area,  $A_{C,tot} = 0.00181$  m<sup>2</sup> calculated with equation 57.

$$A_{C,tot} = \frac{D_{C,tot}^2 * \pi}{4} \quad (57)$$

The amount of cooling water needed only reflects about 1.46 % of the total mass flow through the connected cooling flow pipe, while the difference in cross section area between the two pipes are 4.34 %, calculated with equation 58 and 59. This means that there should be more than enough cooling water for TEST to be operating at safe conditions.

$$\dot{m}_{C,quota} = \frac{\dot{m}_C}{\dot{m}_{C,tot}} \quad (58)$$

$$\dot{A}_{C,quota} = \frac{\dot{A}_C}{\dot{A}_{C,tot}} \quad (59)$$

## F Experimental List

There are some parameters that was kept constant during the experiments. These where pressure measured at FVAPOR,  $P_{FVAPOR} = \_ \text{ bar}$ , see appendix G for more information on why,  $\dot{m}_{FCOLD} = \_ \text{ ton/h}$  as well as  $X_{FNAOH} = \_$  which is standard. But there was also some things that changed during the experiments, and these thing can be found in table 7 which of secrecy reasons was omitted from this report.

Table 7: Changing parameters during the experiments

## G Bubble point

Below follows the approximate bubble point calculations for the production stream, where the bubble points of most of the different substances are calculated according to formulas, equation 60 to 63, provided by Perstorps RD department. The picture on the bottom shows how the lowest bubble point changes with temperature and pressure. The maximum pressure is put to \_ bar which gives a bubble point of about \_ °C which is lower then the bubble point for water at this pressure [30].

$$P_{NVAL}^* = \frac{\exp\left(149.58 - \frac{8890}{T} - 20.697 * \log_{10}(T) + 0.022101 * T\right)}{1000} \quad (60)$$

$$P_{2-mbal}^* = \exp\left(9.2656 - \frac{4161.8}{T - 4.2046}\right) * 1000 \quad (61)$$

$$P_{PBA}^* = \frac{\exp\left(85.726 - \frac{9944.52}{T} - 21.366 * \log_{10}(T) + 4.71 * 10^{-18} * T^6\right)}{1000} \quad (62)$$

$$P_{pentanol}^* = \frac{\exp\left(168.96 - \frac{12659}{T} - 8.6977 * \log_{10}(T) + 1.16 * 10^{-5} * T^2\right)}{1000} \quad (63)$$

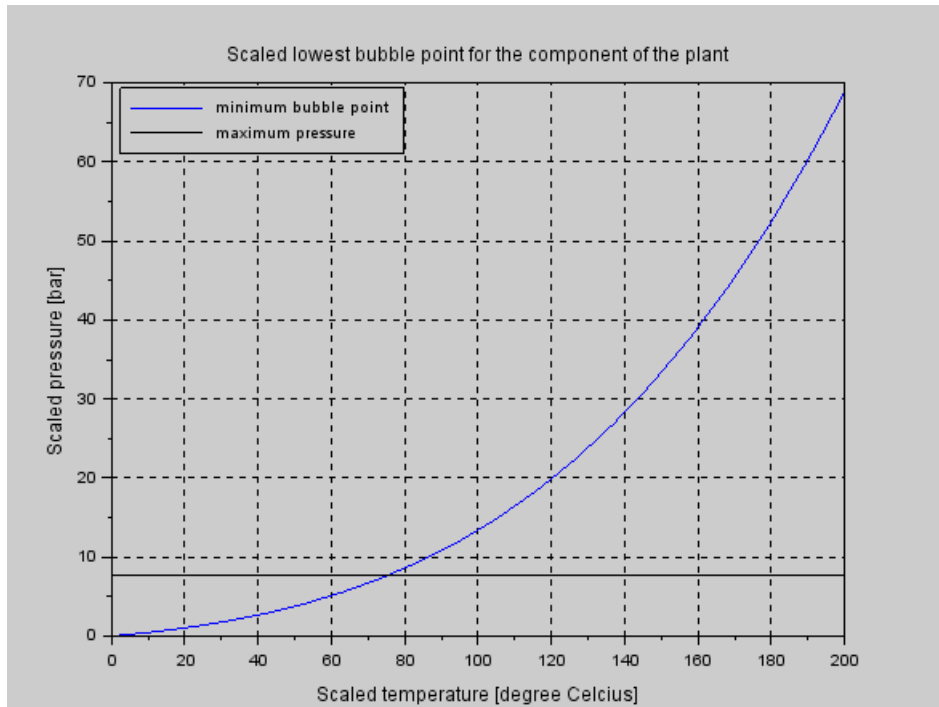


Figure 9: Scaled lowest bubble point for the different component of the plant. Values are scaled by subtracting the lowest value in each axis.

## H Historical data model

Below follows a more detailed summary of the historical data analysis. In figure 10 the residual, calculated with equation 45, are plotted against the factors and time, while in table 8 the final results of the parameters of equation 46 can be found as well as there  $t_{obs,h}$  value which is below  $t_{\alpha/2,n-p}$  from table 1 which indicates that the null hypothesis can not be neglected but if these parameters are taken away then the model has a worse fit to the data. For secrecy reasons 8 is omitted from this report.

As can be seen in figure 10A), NaOH concentration, and 10C), recirculation ratio, the residual still seem to have a second order relationship between the factors and the PBA concentration,  $y$ . Figure 10B) and 10D) display a more random spread of the residual, indicating a less additional dependency on these factors.

Table 8: Values of  $\theta_h$  and  $t_{obs,h}$

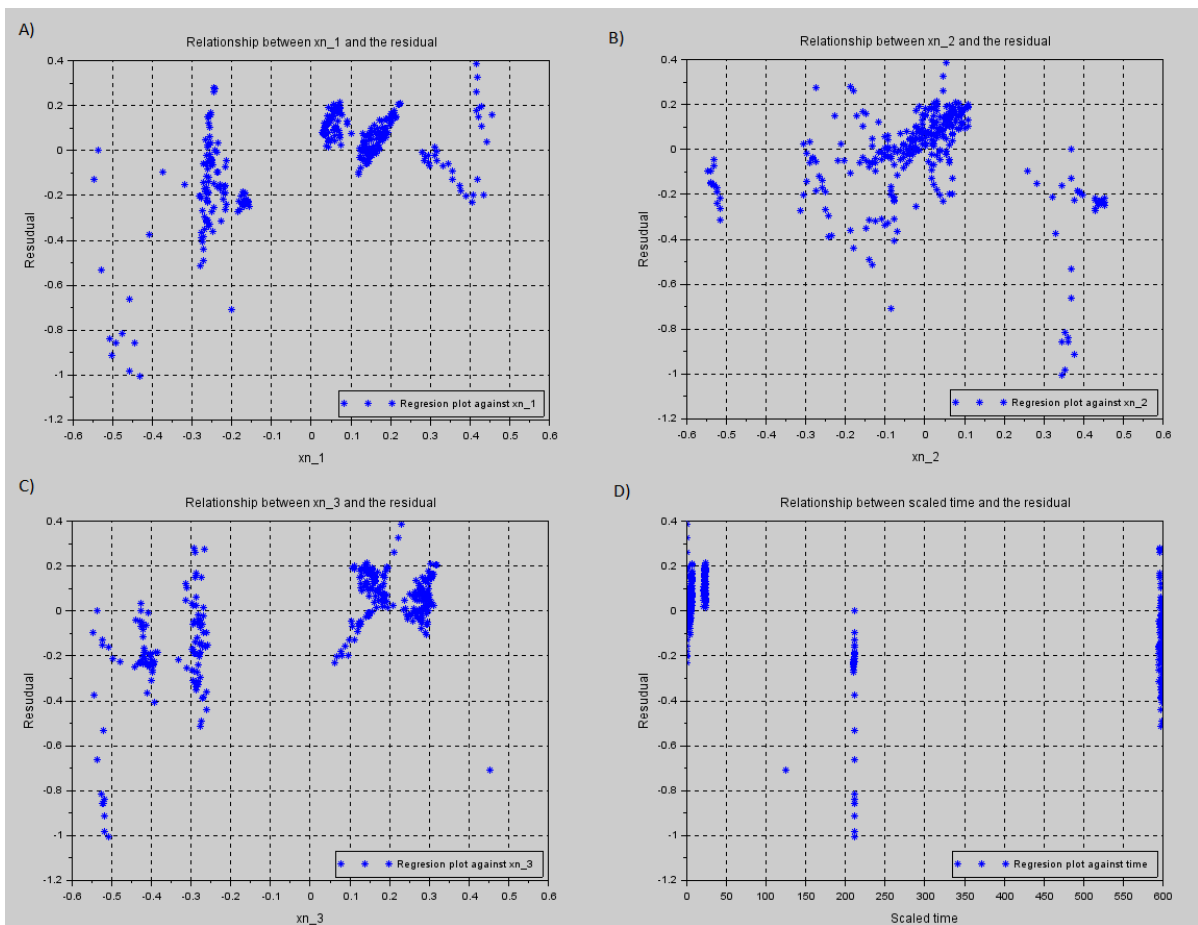


Figure 10: Residual between historical data and formula. Values are scaled by subtracting the lowest value i each axis.