Fouling modelling in a UHT unit based on plant data.

Author: Thai Lam

University supervisor: Andreas Håkansson

Tetra Pak supervisor: Hamid Ghanbari

Examiner: Maria Glantz

Abstract

Heat treatment of dairy products is essential for food safety, one way to perform this is ultra-high temperature (UHT) processing. During the process, macronutrients and minerals form fouling deposits which impairs heat transfer. Eventually the fouling needs to be removed chemically by alkaline and acidic detergents. Fouling buildup and removal are dependent on both internal parameters such as the physical properties of the product and external parameters such as time, flow rate, temperature, and concentration. Modelling fouling development and removal on an industrial scale is valuable from an economical and environmental perspective. By optimizing how long production and cleaning cycles should be the productivity can be increased and the chemical waste as well as the energy consumption can be decreased. Although many models for fouling have been developed, each model is highly dependent on the type of heat exchanger, the product, and the operational parameters. Furthermore, they do not account for upstream and downstream processes which can affect the operator's decision. This thesis takes a data-driven approach to construct fouling development and removal models from 27 months' worth of data that was retrieved from a specific heat exchanger. The raw dataset was divided into production and cleaning in place (CIP) cycles. Each cycle had specific requirements that had to be fulfilled e.g., for production they should be longer than 5 hours. CIP cycles were further divided into caustic and acidic cycles. However, due to deviations from standard operating procedures (SOP) and unreliable conductivity meters, caustic and acidic cycles could not be reliably differentiated, therefore 10 reliable CIP cycles were used instead. Regression models were developed for the production and cleaning cycles, where the fouling during production increased linearly with time while cleaning cycles had a more complicated relationship with fouling.

Popular science summary

Dairy products need to be heat-treated to be safe for consumption, one way to do so is by using ultra high temperature (UHT) treatment, which involves heating products to temperatures of 135°C. This ensures that harmful bacteria are eliminated and allows for storage at room temperature, in contrast to pasteurization. However, at this temperature large deposits of minerals, fats, carbohydrates, and proteins form which is known as fouling. As a fouling layer is created between the hot and cold fluid, the pressure drop increases which leads to a deteriorated performance. To compensate for the increased pressure drop, the flow rate of the heating medium can be increased but after a certain point the fouling layer becomes too thick and chemical cleaning is required. Chemical cleaning is performed using alkaline and acidic solutions to dissolve the proteins and minerals respectively.

In a UHT plant there are many heat exchangers and other units, each with their own sensors. Each sensor takes a measurement every minute which results in a very large dataset. Based on Tetra Pak know how, one specific heat exchanger which had the highest amount of fouling was chosen. The sensors selected was the temperature of the product and heating medium, the flow rates, and the conductivity. In general, higher temperatures lead to more fouling because more reactions can occur between carbohydrates, protein, minerals, and fat, while higher flow rates lead to less fouling due to the forces exerted on the fouling layer by the liquid. Other factors that affect fouling is time and concentration of the detergents. The conductivity was selected to monitor which detergent is used: acid or base.

There are different approaches to modelling fouling: data driven models or physical models. Data-driven models is a pure statistical approach while physical models use mass and energy balances. It is often difficult to find a physical model that is appropriate for industrial applications due to variations in equipment, cleaning solution, product, and operational parameters. A data driven approach has been taken here but this approach also has difficulties such as customers not adhering to the recommended standard operating procedure (SOP) given by Tetra Pak. Another issue is that operator errors might occur. Hence, the data must be cleaned and filtered.

The primary goal of this thesis was to model fouling development during milk production and the secondary goal was to model fouling removal. Fouling development increased linearly over time while fouling removal was close to linear during alkaline treatment and exponential during acidic treatment. Understanding how quickly fouling builds up and how readily it can be removed, can decrease the chemicals used during cleaning and the energy consumption which would improve the sustainability of the process.

Acknowledgements

I would like to thank my supervisors Andreas Håkansson at Lunds tekniska högskola (LTH), and Hamid Ghanbari at Tetra Pak, for their valuable advice and feedback.

Table of Contents

1.Introduction

1.1 An overview of dairy processing

Dairy products may contain pathogenic microorganisms that must be deactivated or destroyed during food processing to ensure food safety. A widely used method is pasteurization which involves heating the product to temperatures between 60 – 100 °C to eliminate harmful bacteria. However, heat resistant bacteria present in raw milk can survive pasteurization and therefore refrigeration is necessary to prevent proliferation. The limiting factors for the shelf life are the growth of microorganisms, enzymatic changes, and flavour degradation. Standard shelf lives are between 8-10 days in a sealed package during refrigeration. Another method to process dairy products is through ultra-high temperature (UHT) processing, in which the product is heated to temperatures above 135 °C for a few seconds, either in indirect mode using heat exchanger or in direct mode using steam. This process ensures that thermoduric bacteria are eliminated and if the product is packaged aseptically, it can be stored at room temperature with shelf lives up to several months. Since refrigeration is not needed, storage and distribution are facilitated, especially in countries where it is difficult to maintain refrigeration. Consequently, the energy consumption is lowered, which is beneficial from both an economic and environmental perspective. Furthermore, a longer shelf life can enable production of larger batches, leading to increased operational efficiency (Bylund, 2015).

During the processing of dairy products, soiling of the equipment occurs which deteriorates the performance, this is also known as fouling. Carbohydrates and proteins may stick to the surface of heat exchangers as a result of high temperatures which lead to cross-linking of proteins and Maillard reactions. Minerals such as calcium phosphate also precipitate at high temperatures. To eliminate the soil, equipment is usually cleaned in multiple steps using alkaline and acidic solutions. The optimization of cleaning cycles is interesting from an economical and environmental perspective, because productivity can be increased, and waste can be decreased (Tetra Pak, 2015).

1.2 Aim

There are two main approaches to making fouling development and fouling removal models: physics-driven models and data-driven models. In this thesis the data driven approach has been taken using 27 months of large-scale production data where 3% and 1.5% fat milk are the products. Given previous studies on fouling by Tetra Pak, trends for fouling development and removal have been observed in controlled conditions, the research question here is: whether this is applicable to industrial data as well? The hypothesis is that if the operators follow standard operating procedure (SOP), the data should look similar.

The aim of this thesis is to develop a predictive model for fouling development in a tubular heat exchanger at UHT conditions, with milk as a product. This will be done in the following steps:

- Providing a literature review of fouling development
- Providing a workflow for cleaning and pre-processing the data
- Understanding the industrial data in comparison to Tetra Pak's SOP
- Develop a predictive model for fouling development.
- Investigate how time, temperature and flow rates affect fouling removal.

The goal for Tetra Pak is to develop models that can be implemented for dynamic simulation and deterministic optimisation for short term scheduling.

1.3 Shell and tube heat exchangers

Shell and tube heat exchangers consist of an outer shell and inner tubes. One fluid flows through the inner tubes while another flows inside the shell and an indirect exchange of heat occurs. In this application, milk flows through the tubes and water flows through the shell. The advantage of shell and tube heat exchangers in comparison to plate heat exchanger is that the channel width is much larger which makes it more resistant to fouling (Fryer et al, 1996). The disadvantages are that it is large and thus the capital expenditure will be high, and that the energy efficiency is only 70% as compared to 95% for plate heat exchangers (Edreis et al, 2020).

1.4 Fouling development in Heat exchangers

During operation of heat exchangers, fouling, which is the deposition of materials onto the surface of equipment, occurs. This leads to decreased heat transfer and increased pressure drop which must be compensated for by either increasing the flow rate or increasing the temperature of the heating liquid. Otherwise, the outgoing temperature will be lower which affects the shelf life and safety of the product. Another alternative to deal with fouling is by cleaning with alkaline and acidic solutions. Fouling resistance is the additional thermal resistance caused by the foulant and it can be determined by measuring the overall heat transfer coefficient after fouling and the overall heat transfer coefficient of a clean heat exchanger (Muller-Steinhagen et al, 2011).

$$R_f = \frac{1}{U_d} - \frac{1}{U}$$

R_f is the fouling resistance [m^2*K / W], U_d [W/ m2* K] is the overall heat transfer coefficient after fouling and U [W/ m2 * K] is the clean overall heat transfer coefficient (Muller-Steinhagen et al, 2011). Fouling development can be monitored using pressure or temperature sensors, as fouling leads to increases in pressure drop and in heat transfer.

There are different types of fouling; one example is crystallization fouling which refers to the deposition of salts such as calcium phosphate. This process has been shown to be linear with an induction period at the start where the fouling resistance is zero (Ritter, 1983), but there have also been studies where fouling resistance decreases gradually and sometimes to an asymptotic state, where the rate of removal is equal to the rate of deposition (Bohnet, 1987). Linear fouling relationships are found in cases where strong deposits are formed, where the removal rate is low, while asymptotic cases are found when weaker deposits are formed. In addition to the deposit characteristics, the type of heat exchanger and operational parameters such as flow rates and temperature, affect how fouling will develop.

1.5 Fouling types and removal

When the fouling level has passed a critical threshold, the production plant must be cleaned. The detergents used for cleaning depends on the type of fouling. Fouling can be classified into two different types: A and B. Type A can be found in temperatures between 75°C – 115 °C and consists of 50-60 % protein, 30-35 % minerals and 4-8 % fat while type B can be found in temperatures above 115 °C. The main type of protein in type A is beta lactoglobulin. Type B fouling is harder and more brittle consisting of 70 – 80 % minerals, 20-25% protein and 4-8% fat (Goode et al, 2016).

There are four different forces that affect fouling removal: mechanical force, thermal force, chemical force, and contact time (Figure 1). For cleaning in place (CIP) the flow must be turbulent and a flow velocity of at least 1.5 m/s must be maintained. The chemicals used are dependent on the type of fouling: alkaline solutions (e.g NaOH 0.5 – 2 wt %) are used to remove proteins and fats while acidic solutions (e.g H_3PO_4 or HNO_3 0.5 – 1.5 wt %) are used to remove minerals (Goode et al, 2016). Another study examining the structural and compositional changes in UHT milk fouling showed that proteins are depolymerized by the caustic solution and when acid is applied, both the minerals and the proteins are washed away. The most important parameters were the temperature during caustic cleaning and the concentration of the cleaning agent (Hagsten, 2016).

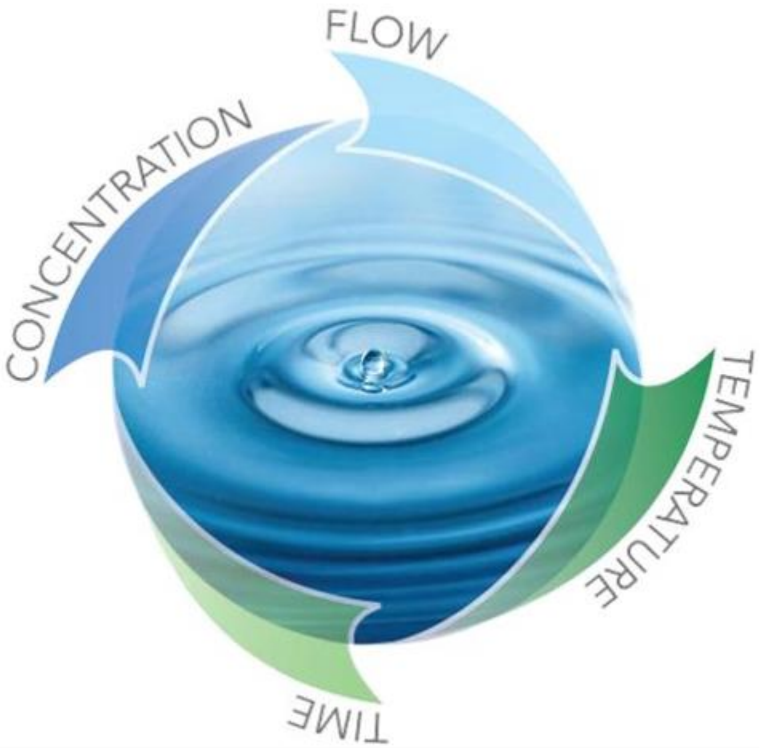


Figure 1. The four factors affecting fouling removal: concentration, flow, temperature, and time. Source: “Production and CIP optimization methodology” by Hamid Ghanbari (2021-12-10).

1.6 Business unit UHT

Ultra-high temperature treatment involves heating a product to 135 – 150 °C for a few seconds to eliminate any microorganisms (e.g *Bacillus sterothermophilus* and *Bacillus sporothermodurans*) (Grijspeerdt et al, 2004). There are two ways of heating: indirect and direct; in direct heating steam is injected directly into the product while in indirect heat is transferred through heat exchangers (Bylund, 2015). Indirect is more energy efficient since it allows reuse of the water/steam while direct heating is less fouling prone (Grijspeerdt et al, 2004). UHT operation can be divided into four main steps: plant pre-sterilization, production, aseptic intermediate cleaning (AIC) and CIP. AICs are used to remove fouling during long productions. It takes 30 minutes and aseptic conditions are maintained. CIP is a full cleaning cycle that takes 70 – 90 minutes and usually takes place after the production step is finished (Tetra Pak, 2015).

The UHT plant used in this study utilizes indirect tubular heat exchangers which is the most common type of heat exchanger in UHT systems. These plants are suitable for products with medium viscosity such as puddings and deserts but also for milk products with longer processing times. A typical plant is shown in Figure 2. In the first step product is transferred from the balance tank (1) that stores the product to the pump (2). Product is pumped through the first heat exchanger where it is preheated regeneratively using heated product that is flowing out (3). The preheated product is transferred to the homogenizer (4) and thereafter heated in the second heat exchanger (5). Following the heat exchanger there is a stabilizing holding tube to stabilize milk proteins and decrease fouling. Another heat exchanger is used for further heating (6) followed by cooling (7). The final product passes through first heat exchanger (3) where it is cooled and passes to the aseptic tank (Bylund, 2015).

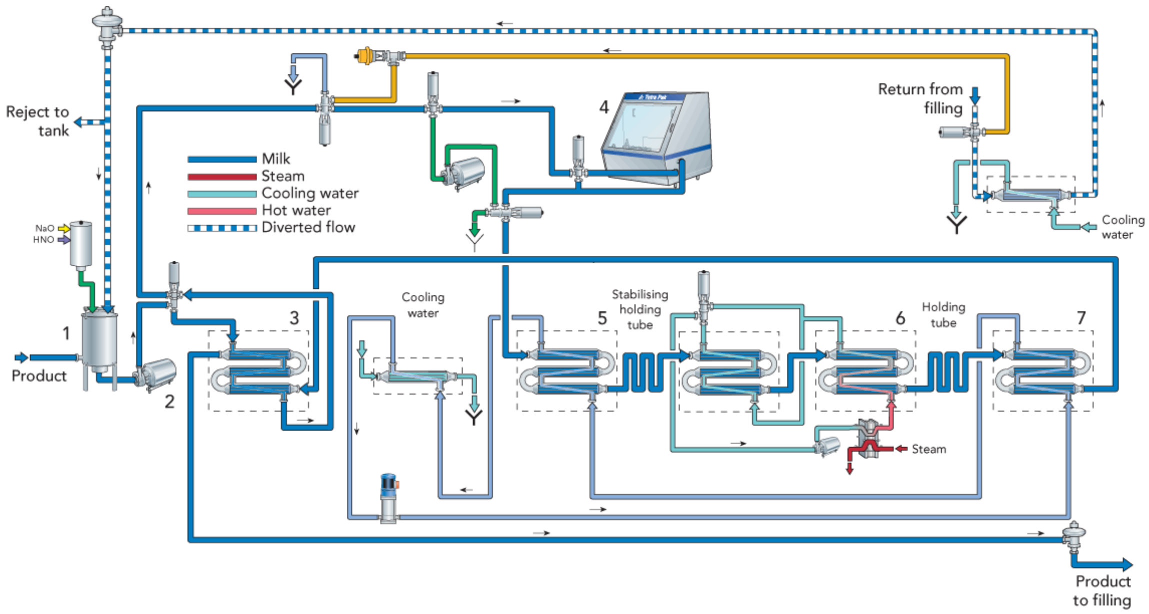


Figure 2. An indirect tubular UHT plant. Source: Dairy processing handbook, Tetra Pak, 2015. 1) Balance tank. 2) Feed pump. 3) Tubular heat exchanger, regenerative and cooler. 4) Non-aseptic homogenizer. 5) Tubular heat exchanger. 6) Tubular heat exchanger final heater. 7) Tubular heat exchanger, cooler.

1.7 Fouling modelling in heat exchangers

Understanding how fouling occurs and modelling it can decrease cleaning cycles and thus increase productivity (Jun et al, 2004). There are many different models for fouling, some being complete 3D models while others are simple correlations. Models that require solving partial differential equations are considered outside the scope of this thesis. While they describe a system completely, a 3D models require large computing power and only short periods can be simulated. Additionally, they are not useful for short term scheduling. From an applications perspective, correlations may be more useful as correlations can capture what happens during longer periods. Some correlations for fouling are 1) The fouling resistance, R_f is linearly correlated with foulant concentration C_b where k is a constant, 2) fouling resistance is proportional to the velocity v to the power of –1.5 with k as a constant, 3) fouling increases with temperature according to Arrhenius equation where K is a pre-exponential factor, E is the activation energy, R is the gas constant and T is the temperature (Muller-Steinhagen et al, 2011).

1. $R_f = k \cdot C_b$
2. $R_f = k \cdot v^{-1.5}$
3. $R_f = Ke^{\frac{-E}{RT}}$

Although fouling has been studied extensively in plate heat exchangers (PHE), there are very few studies on shell and tube heat exchangers (THE) focused on dairy products. One such study shows that there is an initial induction period after which the fouling increases linearly over time until saturation (Paterson et al, 1988). The biot number is a measurement of fouling and is defined as Bi = $h^0 x_d$ / λ, where h^0 is the clean heat transfer coefficient, x_d is the fouling thickness and λ is the deposit thermal conductivity.

The fouling rate, r_i, is described as a function of time and fluid velocity, where u is the fluid velocity, E_a is the activation energy, β is a constant, R is the gas constant and T is the temperature (Paterson et al, 1988).

$$r_i = \frac{\beta e^{-E_a/RT}}{u}$$

Another study uses a statistical model for fouling by using multiple regression (Fryer et al, 1996). In the study PHE and THE are compared with regards to fouling. The results indicated that THE have a longer induction period which is reasonable since PHE have thinner passages which render them more susceptible to fouling. Furthermore, the flow rate had a larger contribution to fouling in tubes than in plate. These models are based on the thickness of the fouling layer, which is not measured in the plant data, therefore they cannot be used.

2. Methodology

2.1 Description of a cycle following standard operating procedure and relative soil level calculation

A full cycle following SOP is preceded by sterilization to ensure product safety. Thereafter a production phase (production cycle) is initiated. This is where the desired dairy product (in this case milk) is heat treated. During the production cycle the pressure drop will increase with time due to increased fouling. When the pressure drop has reached a certain threshold, cleaning must be initiated, which can occur in two ways 1) full cleaning: the system is thoroughly cleaned 2) intermediate cleaning: a brief cleaning is performed and afterwards production starts again. The difference between intermediate cleaning and full cleaning is that in the former, the pressure drop does not decrease back to its original value, which is clean state, but the advantage is that aseptic conditions are maintained. However, since in practice there may be oscillations and slight variations in relative soil level (RSL), it is easier to distinguish between intermediate cleaning and full cleaning based on the StepNO (Table 2). StepNOs are used to identify what is currently happening in the plant for example if production or cleaning is occurring.

It is important to consider that the pressure drop will vary with the product flow rate. To compensate for variations in flow rate, the RSL can be defined as:

$$RSL = \frac{\Delta P}{\Delta P_0} = \frac{\Delta P}{K_{sys} \cdot v^2}$$

Where ΔP [bar] is the measured differential pressure, ΔP_0 [bar] is the differential pressure when the system is clean, K_{sys} [kg/m^3] is a variable that is dependent on the physical properties of the fluid such as temperature, viscosity and density, and v is the velocity [m/s]. K_{sys} was in this case calculated using Quantum tool, a software used for design of heat exchangers, but it can be also calculated from Darcy-Weisbach correlations. Darcy-Weisbach equation can be expressed as:

$$\Delta P_0 = f_d \cdot \frac{\rho}{2} \cdot \frac{v^2}{D_H} \cdot L$$

Where ΔP_0 [bar] is the clean pressure drop, f_d is Darcy friction factor, ρ [kg/m^3] is the density of the medium, v [m/s] is the mean velocity, D_H [m] is the hydraulic diameter and L is the length [m]. Everything on the right-hand side except v^2 lumped together equals K_{sys}. Based on Reynold's number and the pipe's relative roughness ε/D, Darcy friction factor can be estimated, with a Moody diagram for example (Shashi Menon, 2015).

Using K_{sys} values from Quantum tool for different temperatures and velocities, and performing a regression, a correlation for how K_{sys} depends on temperature and velocity can be estimated, which was:

$$K_{sys} = 28.95 - 0.059 \cdot T - 1.15 \cdot v$$

This correlation was used for the calculation of RSL for all datasets. K_{sys} is normally calculated for each product and detergent, usually by stopping production and running the clean system with detergents. The detergents are often formulated detergents which means that the properties are not known, however, since they are diluted, their physical properties will be similar to water. The properties of milk are also similar to water and therefore water was used as reference media for all calculations. The consequence of this is that RSL will not be equal to 1 when the plant is clean, but the cleaning progress can still be monitored by the slope of RSL. In this case the plant was clean after the first cycle of alkaline and acidic treatment, but this is not always the case.

A SOP cleaning cycle is shown in Figure 3, where there is an initial increase in RSL during dosing of alkaline solution due to swelling. Subsequently, the RSL drops as the alkaline solution is circulated since it removes protein fouling. Next the heat exchanger is washed with acidic solution and during the circulation of acid, a large drop in RSL can be observed. After one treatment of alkaline and acidic solution, rinsing is done with water. A RSL of around 1 indicates that the heat exchanger is clean. According to SOP, another round of alkaline and acidic treatment should be performed.

A *fullCIP* cycle is defined as a production cycle followed by a cleaning cycle. A cleaning cycle consists of two rounds of alkaline dosing and circulation and two rounds of acidic dosing and circulation. An intermediate cleaning CIP (*ICCIP)* cycle is similar but contains intermediate cleaning between production cycles.

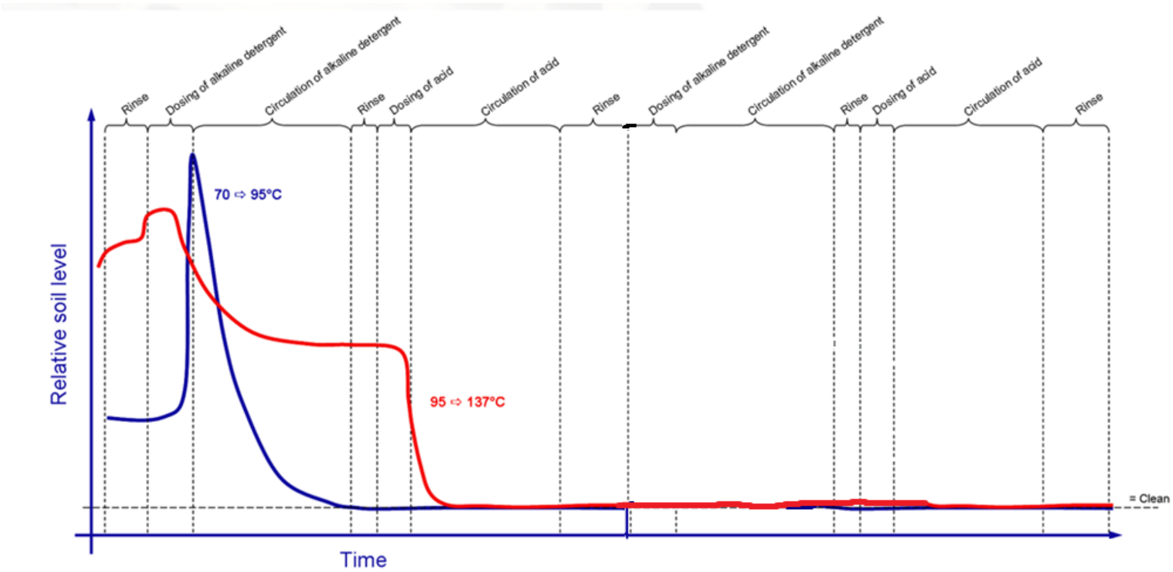


Figure 3. An SOP cleaning cycle. Relative soil level (RSL) as a function of time. Blue low temperature. Red high temperature. Source: “Production and CIP optimization methodology” by Hamid Ghanbari (2021-12-10). SOP: standard operating procedure.

2.2 Data processing

There were two datasets acquired from a dairy plant was used in this study: one consisting of 3 months milk production data which takes measurements every minute and another with 27 months milk production data taking measurements at the same interval. The 3 months milk production data were conducted under supervision and therefore the data adheres to SOP while this is not guaranteed in the 27 months dataset which consists of the 3 months dataset and another 24 months data. The dataset is for one heat exchanger in a complete production line, and it was chosen based on Tetra

Pak's knowledge that if this section is clean then the whole loop is clean. This assumption reduces the dataset immensely. The relevant sensors are product flow rate, heating medium flow rate, product temperature, heating medium temperature, conductivity and pressure drop (Table 1). The pressure drop is calculated by subtracting the ingoing pressure from the outgoing pressure (Figure 4).

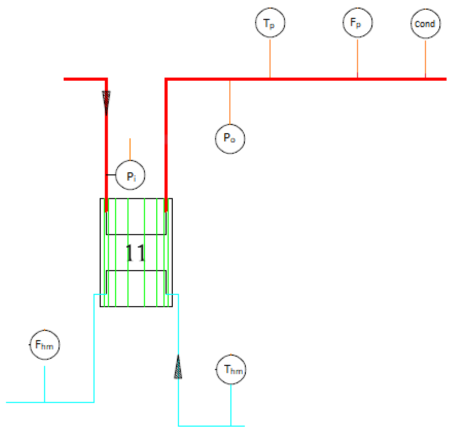


Figure 4. A simplified process flow chart and where the sensors are placed in relation to the heat exchanger. P_i pressure in, P_o pressure out, T_P product temperature, T_{hm} heating medium temperature, F_P product flow rate, F_{hm} heating medium flow rate. Red is the product and blue is the heating medium.

Table 1. All sensors that are important for studying the fouling development and removal.

Sensors	Name
F_p [m^3/h]	Flow rate product
F_{hm} [m^3/h]	Flow rate heating medium
T_p[°C]	Product temperature
T_{hm} [°C]	Temperature heating medium
Cond [mS/cm]	Conductivity
ΔP (bar)	Pressure drop
t (min)	Time

The Pandas library in Python was used to generate data subsets (Figure 5). The raw data was in the form of a matrix where each column corresponds to a sensor variable and each row corresponds to a measurement (a measurement occurs every minute). In total there are about 1 300 000 measurements for each sensor.

The dataset is complex consisting of more than 70 possible steps. Each step is used as a tag to identify what is currently happening in the plant. A certain combination of step/steps comprises a cycle. The most important steps are summarized in Table 2. One important factor to consider is that the only way to separate caustic dosing from acidic dosing and caustic circulation from acidic circulation is based on the conductivity. Another consideration is that there is no way to distinguish between the first and the second dosing/circulation, hence it must be assumed that they always occur in order. The following assumptions were made when creating the datasets:

- Each cycle starts with production and ends with final rinse.

- Every CIP cycle must follow the pattern: caustic dosing --> caustic circulation --> acidic dosing --> acidic circulation --> caustic dosing 2 --> caustic circulation 2 --> acidic dosing 2 --> acidic circulation 2.
- Conductivities of 40 – 70 mS/cm indicates acid while conductivities of 90 - 130 mS/cm indicates caustic.

However, since the UHT process is dependent on upstream and downstream processes, operators can sometimes make unusual changes in steps which can have unpredictable effects on the dataset construction. Another consideration is deviations from SOP either due to the sensors manufacturer's preferences or due to human errors.

The dataset was constructed in the following way: first NaN values were removed by filtering away all rows that contain at least one NaN value. Approximately 4.5% of all values were removed in this filtering step. Next the data was inputted into *separateFullCIPfromIC* which generates cycles with intermediate cleaning (IC) and cycles with *fullCIP*. The separation is based on whether there is an intermediate cleaning step between production and final rinse, if there is the cycle is assigned to the IC cycles. The *fullCIP* cycles are then separated into production cycles and CIP through *ExtractProdtimes* which was done based on the production step. The CIP cycles are used as input to *findCaustic* which is then divided into caustic dosing 1 & 2 and caustic circulation 1 & 2. The same process is done with *findAcidic,* the difference here is in the conductivity. For acids the conductivity is between 40 – 70 mS/cm while bases have conductivities of 90-130 mS/cm. The separation into first and second dosing/circulation assumes that it alternates between 1 and 2.

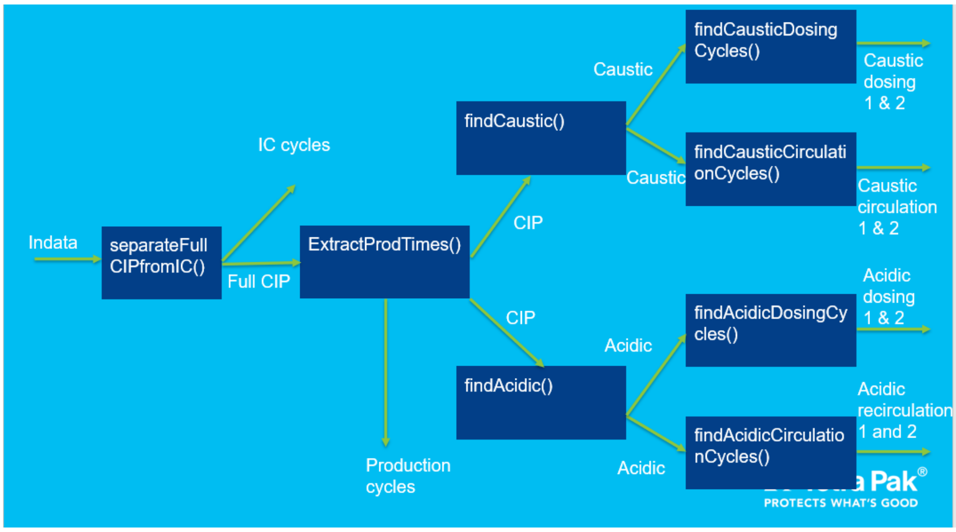


Figure 5. Flowsheet for generation of subsets. First fullCIP is separated from IC cycles, then fullCIP is divided into production and CIP cycles. CIP cycles are then separated into caustic and acidic cycles. These are in turn separated into dosing and circulation. CIP: Cleaning in place. IC: intermediate cleaning.

Table 2. The most important steps during production and cleaning. StepNOs are used as identifiers for what is currently happening in the plant.

Name	Explanation	StepNO
Production	This is where a cycle starts.	135
Intermediate cleaning	This step is used to differentiate between *ICCIP* and *fullCIP*.	225
Water circulation	Sometimes water circulation occurs in the middle of production, thus making it difficult to determine if it should be counted as one or two production cycles. Example: production --> water circulation --> production.	95
Dosing	Where either caustic or acidic solution is added gradually. The only way to distinguish between them is through the conductivity.	435
circulation A + circulation B	Every dosing step is followed by at least one circulation either A or B and sometimes A and B.	445
Intermediate rinse	Occurs after circulation.	465
Hibernation	Sometimes the process needs to go into hibernation due to downstream issues or other reasons	600
Final rinse	This is where a cycle ends.	480
Sterilization	A step used to sterilize the equipment.	45
Idle	Used when the operators are on a break.	0

2.3 Statistical models

Using the production dataset and cleaning dataset, models were developed for predicting fouling formation and removal. Three different statistical models were tested to see which one gives the most accurate results: multiple adaptive regression splines (MARS), multiple linear regression (MLR), and partial least squares (PLS). Each model and their advantages and disadvantages are briefly described in this section.

2.3.1 Multiple adaptive regression splines

MARS is a non-linear regression method that uses piecewise functions called hinge functions to approximate the dependent variables based on the independent variables. A hinge function can be described as:

$$h(x-t) = \begin{cases} x-t, & x > t \\ 0, & x \leq t \end{cases}$$

i.e a function that assumes the value zero for all values smaller than equal to t and x-t for all values where x is larger than t. Different degrees can be used, when a first-degree spline is used, piecewise linear functions will be used to estimate the dependent variable. Higher degrees usually lead to a higher R^2 in the model, however, there is a risk of overfitting which means that the R^2 in validations will be considerably worse. The advantage of MARS in comparison to MLR and PLS, is that it can capture non-linear relationships that are more complex, which is the case for CIP (Friedman, 1991).

2.3.2 Multiple linear regression

MLR is a statistical method used to estimate a dependent variable based on the assumption that it has a linear relationship with the independent variables. Other assumptions include: no multicollinearity between independent variables, the observations are independent, and normally distributed residuals (Allison, 1999). MLR is useful for capturing simple linear relationships which is expected in the production model based on previous studies (Ritter, 1983).

2.3.3 Partial least squares

PLS is an extension of multiple regression but instead of maximizing the overlap between dependent and independent variables one at a time for each independent variable, it considers all independent variables at once. The advantage of PLS is that it performs better than MLR if there is collinearity between the independent variables (Geladi et al, 1986). PLS involves projection into another space, thus the independent variables will become components instead. Using all components will give the same results as multiple regression, however, it is not guaranteed that more components equal a better model, rather that should be decided based on the validations.

2.4 Scaling of the production cycles

Scaling variables is especially important if the independent variables are of different orders of magnitude. The independent variables are product flow rate F_p, heating medium flow rate F_{hm}, change in temperature ΔT ($T_{hm} - T_p$) and time t since the start. Typical values for flow rates are 30 m^3/h, ΔT = 4°C and time increases from 1 to 1600 min, which might indicate a need for scaling.

The RSL was scaled based on the initial value of each cycle to establish the same starting point for all cycles. When scaling of the independent variables was performed, a standard scaler using the equation $z = (x - u)/s$, was used, where z is the scaled number, x is the sample value, u is the mean of all samples and s is the standard deviation.

2.5 Filtering of production cycles

The following filters were applied to production cycles:

- $RSL_{initial} < 1.6$. All production cycles that had an RSL of 1.6 or higher were removed since it was decided that values exceeding this is indicative of an unclean system.
- Production cycles with independent variables that did not change over time were removed, constant variables indicate that sensors malfunctioned.
- $t_{end} > 300$ min. Production cycles that did not last longer than 5 hours are most likely errors during the operation of the plant.

2.6 Data analysis of the production cycles

The production cycles generated from *ExtractProdTimes* using the 27 months data was used for creating a production model using MARS, Multiple regression and PLS. For MARS the python packages sklearn-contrib-py-earth (0.1.0) along with pandas (0.25.3), scikit-learn (0.24.2) and numpy (1.19.5) were used in Python 3.6. The multiple regression and PLS were performed using the sci-kit learn package (1.2.1) in conjunction with Python (3.9.13).

MARS with degree 1 was performed with different variations of scaling: only RSL initial value scaling, scaling independent variables and initial value scaling of RSL, and scaling independent variables and scaling RSL twice (initial value and standard scaler). For the best scaling method, MARS with degree 2 and 3 were also performed to see if there is any improvement in the model and the predictions. The same scaling process was repeated for multiple regression using the same combinations as for MARS. PLS was performed using only initial value scaling and up to four components. For validation, 2 production cycles from the 3 months dataset were used. The training dataset is all production cycles that fulfilled the filtering criteria from the 27 month dataset and the testing dataset are 2 production cycles from the 3 month dataset. The selection of testing dataset was done based on two criteria: long production time and linear increase in RSL.

2.7 Data analysis of CIP

The same Python packages that were used for production were used for analysis of CIP cycles. Since it was not possible to distinguish between CIP cycles due to deviations from SOP and possible errors with the conductivity meter, 10 CIP cycles from the pre-study where the caustic and acidic cycles can be distinguished with certainty were chosen for regression analysis. Different combinations of standard scaling were performed: all unscaled, unscaled independent variables and RSL scaled based on initial value and all independent variables scaled except time and RSL scaled twice. The best performing scaling which is the independent variables scaled except time and RSL scaled twice, was used for both caustic and acidic cycles. MARS with degree 1,2 and 3 was performed for caustic circulation and the acidic cycles. For caustic circulation multiple regression was performed to compare with the third-degree spline. For validation, 4 cycles were used, 2 from the dataset that was used to construct the model and another 2 from outside the dataset. These cycles were chosen based on their similarity to a SOP cycle (Figure 3).

3. Results and discussion

3.1 Overview of process statistics

Figure 6 provides an overview of 820 days of production data. An average of 1 million liters of milk was produced a day and the total energy consumption (including idle time, water circulation etc) was 3.7 GWh (due to the complexity of the P&ID the energy consumption is for two heat exchangers instead of one), which is equivalent to 0.016 kWh/L milk consumption. Approximately 47% of the total time was spent on production, 35% on idle, 7% on water circulation and 5% on CIP. Productivity could be increased immensely if the idle time was decreased, however, it is not known why the operators set the plant to idle. CIP only constituted 5% of the total time, meaning that improvements in CIP cycles, such as decreasing the number of caustic and acidic cycles from 2 to 1, will only have marginal effects on the total productivity.

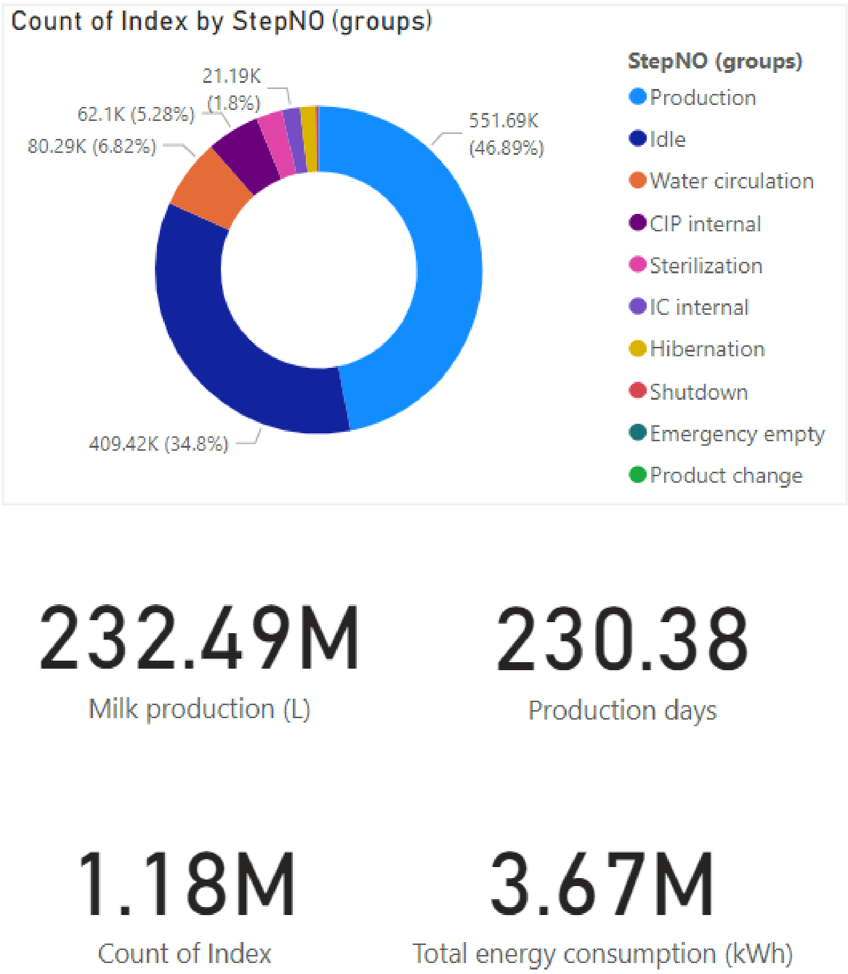
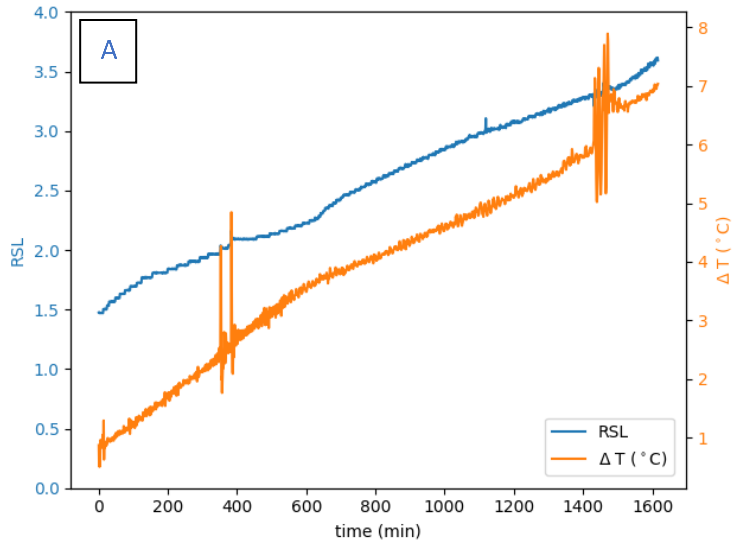


Figure 6. The pie chart shows the distribution of different StepNOs i.e how much time was spent on production, CIP etc. Count of index indicates the total number of measurement points.

3.2 Analysis of Fouling development

3.2.1 An overview of a typical production cycle

A typical production cycle is shown in Figure 7A, where the RSL (left vertical axis) and ΔT (right vertical axis) are shown as a function of time (horizontal axis). As more fouling builds up in the heat exchanger, the RSL linearly increases with time. More fouling deteriorates the heat transfer, and thus the difference between the heating medium temperature and the product temperature must increase to ensure that the product temperature exceeds 135°C. In Figure 7B, the product flow rate and heating medium flow rate are shown as a function of time. The flow rates remain constant for a long period but are sometimes increased stepwise. According to SOP the operator can choose to increase flow rate stepwise, linearly, or exponentially.



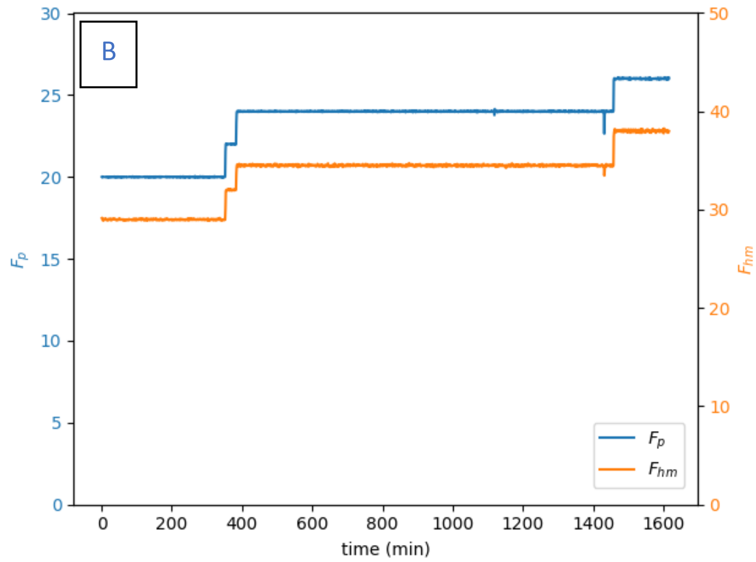


Figure 7. A typical production cycle. The top graph shows the RSL and ΔT as a function of time (A). The bottom graph shows the F_p and F_{hm} as a function of time (B). RSL: Relative soil level. F_p: flow rate product. F_{hm}: flow rate heating medium. ΔT: temperature heating medium – temperature product.

Next, three different statistical models: MLR, MARS and PLS will be used to predict fouling development based on the following operating parameters: F_p, F_{hm}, ΔT and t.

3.2.2 MARS

The results from MARS using degree two is shown by Figure 8A, where the model is compared to the data that was used for training, in terms of RSL as a function of time. Figure 8B-D show a comparison between the model and the validation dataset. The best combination of scaling variables was scaling RSL based on their initial value of each cycle and keeping the independent variables unscaled. Scaling the independent variables resulted in large oscillations in one of the validation data, therefore, it was determined that unscaled independent variables give more reliable results for production data (Figure 8D). In the model, the RSL increases linearly with time which is consistent with the results conducted by Tetra Pak in a pre-study and the results of Ritter et al (1983). However, the slope differs between each cycle which can be attributed to the different flow rates and temperatures of the product and heating medium fluid. The R^2 for different degrees of MARS are shown in Table 3. Since it was known the fouling increased linearly with time from the pre-study, higher degrees of spline with respect to time is not necessary, however, higher degrees of the other independent variables could lead to improvements, as shown in Table 3. After degree 2, the model did not improve further. The oscillations in the prediction in validation1 are due to oscillations in temperature (Figure 8).

Table 3. The RSL during production was modelled as a function of time, temperature and flow rate using MARS. The R^2 values for the model, validation 1 and validation 2 for different degrees of MARS are shown in the table. MARS: multiple adaptive regression splines. RSL: relative soil level.

	Degree 1	Degree 2	Degree 3
Model	0.95	0.97	0.97
Validation 1	0.94	0.96	0.97
Validation 2	0.93	0.93	0.93

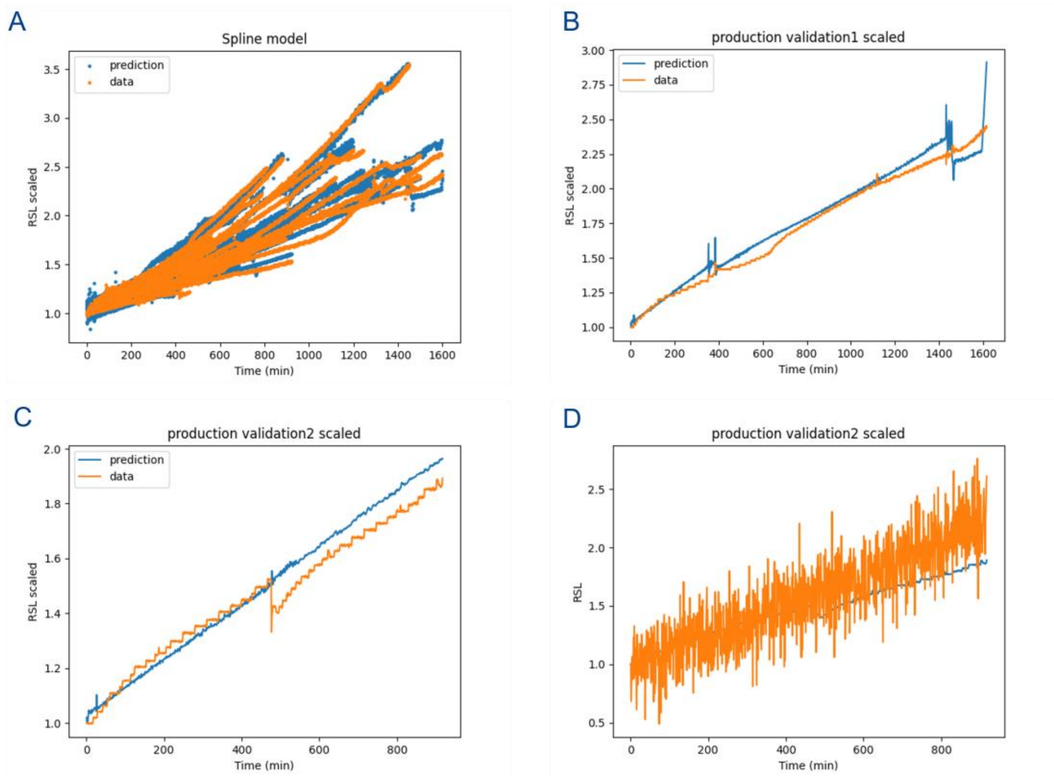


Figure 8. The results from a two-degree spline, data from a filtered subset of all the production cycles in the 27-month data was used to make the model (A), and validation was performed using two cycles from the 3 months data (B, C). When the independent variables were scaled oscillations were present in the second validation dataset (D). RSL: relative soil level.

3.2.3 Multiple regression

Multiple regression with time (min), product flow rate (m^3/h), heating medium flow rate (m^3/h) and change in temperature between product and heating medium ΔT(°C) gave similar results with R^2 values of 0.95, 0.95 and 0.93 for the model, validation1 and validation 2, respectively (Table 4, Figure 9) . Figure 9A shows the model compared to the training data, Figure 9B-C shows the predicted values in comparison to two different validation datasets. Once again, scaling only RSL gave the best results, as the other combinations resulted in oscillations in validation2. As indicated in Table 5, all coefficients were statistically significant, and the equation was:

$$\text{RSL scaled} = 1.39 + 0.0007*t + 0.0695*\Delta T - 0.0410*F_p + 0.0168*F_{hm}$$

Physically it is reasonable to include the time, difference in temperature and flow rates but since the condition number was larger than 30, there is evidence of strong multicollinearity. However, this assumes that all variables have already been scaled, which was not feasible due to oscillations in validation 2. Nevertheless, PLS which explicitly considers multicollinearity, was also performed to compare the results with multiple regression.

Table 4. RSL during production was modelled as a function of time, temperature and flow rates using MLR. The table shows statistics showing the R^2, adjusted R^2 and the F-statistic for the model, and R^2 for the validations. RSL: relative soil level. MLR: Multiple linear regression.

Statistic for the model / validation	Value
R^2	0.949
Adj R^2	0.949
Prob (F-statistic)	0.00
Condition number	5340
R^2 validation 1	0.95
R^2 validation 2	0.93

Table 5. All variables used in the MLR model for production, and their lower and upper confidence intervals. MLR: Multiple linear regression.

Variable	Lower confidence interval (α = 0.95)	Upper confidence interval (α = 0.95)
Constant	1.381	1.404
t	0.001	0.001
F_p	-0.041	-0.041
F_{hm}	0.017	0.017
ΔT	0.068	0.071

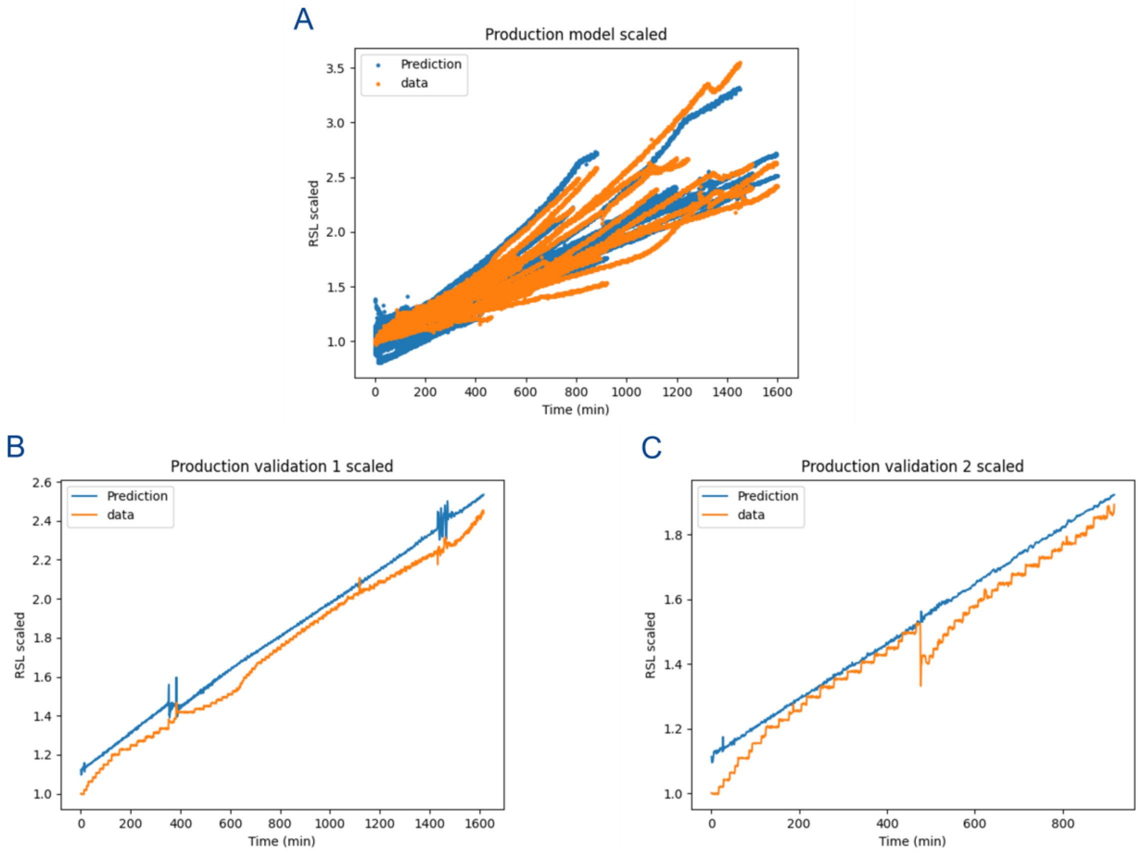


Figure 9 . RSL during production as a function of time. The results from multiple regression: model (A), validation1 (B) and validation 2 (C). RSL: relative soil level.

3.2.4 PLS

The same scaling method was used for PLS as MLR and MARS. The number of components (n = 2) was chosen based on the best R^2 values in the validation which were 0.97 and 0.95, respectively (Table 6). An R^2 of 0.945 was achieved with 2 principal components which only increased to 0.949 using 4 components (Table 6). PLS does not assume that there is no collinearity in contrast to multiple regression, which may explain why it performed slightly better. Figure 10A shows the PLS model in comparison to the data points and Figure 10B-C shows the predicted values using the model compared to two different validation cycles.

Table 6. A PLS model for how RSL changes with time, temperature, and flow rates during production. The table shows the R^2 value for the model, validation1 and validation2 based on the number of components used in PLS. n: number of components. PLS: partial least squares. RSL: relative soil level.

	n= 1	n=2	n=3	n=4
Model	0.928	0.945	0.947	0.949
Validation 1	0.943	0.966	0.964	0.951
Validation 2	0.961	0.952	0.948	0.927

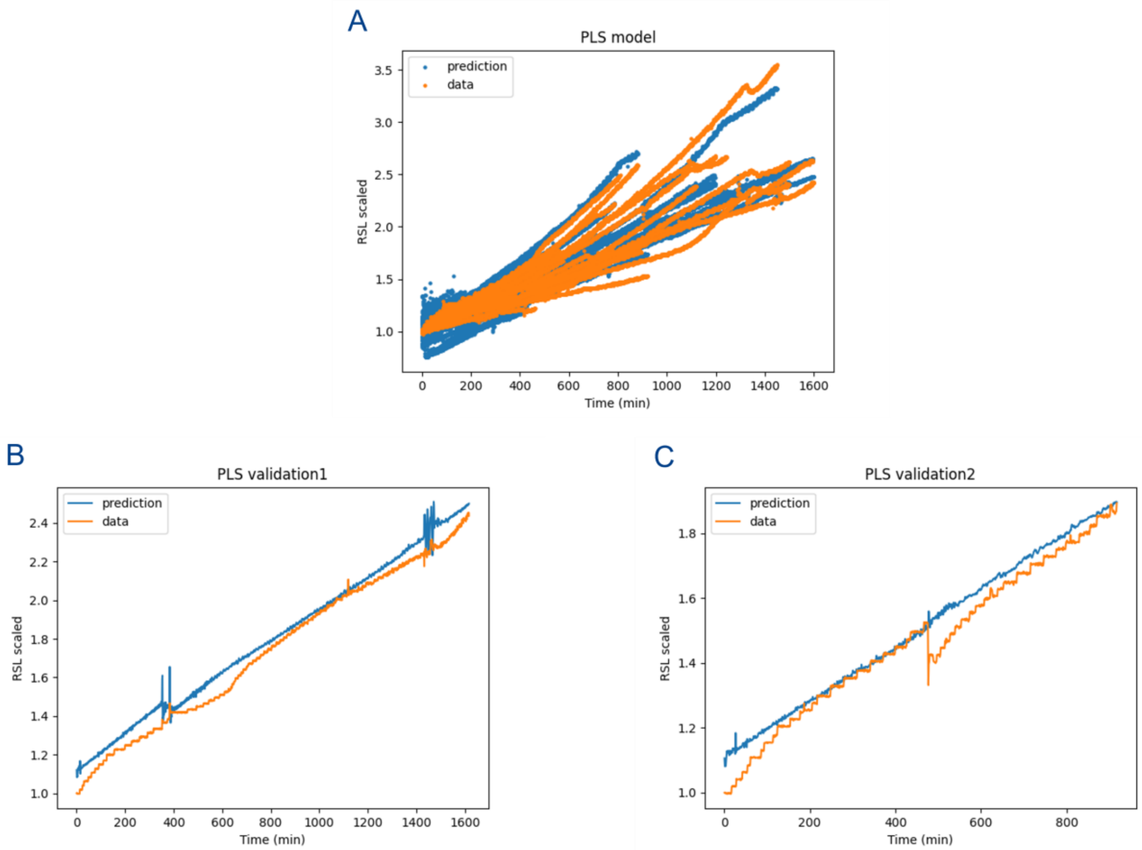


Figure 10. RSL as a function of time. PLS for the model (A) and two validations (B-C) using n = 2. The number of components to use was based on the R^2 value of the validations. The same scaling was used for PLS as for MLR and MARS. RSL: relative soil level. PLS: partial least squares. MLR: multiple linear regression. MARS: multiple adaptive regression splines.

3.2.5 Fouling development summary

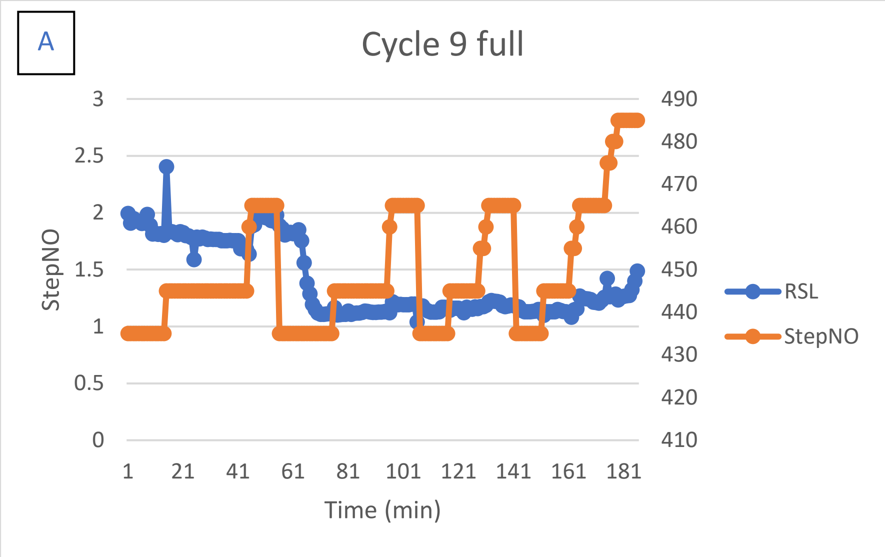
The differences between models both in terms of model fitting and predictions are small. Considering the small differences and prior knowledge that fouling develops linearly over time, MLR is the most suitable model. All independent variables were statistically significant, and they could explain about 94% of the variance in RSL. RSL is slightly overestimated but still close to the real values. An advantage of MLR in comparison to PLS is that it is possible to see the contribution of each physical parameter, which enables prediction of how RSL will change if one of the parameters are changed. The problem with MARS is that there is risk for overfitting, although this was not the case in the two validation sets, using more validation data might show this issue.

3.3 Analysis of fouling removal

This section will first present a typical CIP cycle and then explain deviations from SOP. Subsequently, fouling removal will be modelled and predicted primarily using MARS since CIP is known to be a non-linear process. MLR will be used for parts of CIP that resemble linearity for comparison with MARS.

3.3.1 An overview of a typical CIP cycle

The cycle below indicates an SOP cleaning cycle (Figure 11). It consists of caustic dosing, caustic circulation, intermediate rinse followed by acidic dosing, acidic circulation, intermediate rinse, all these steps are repeated once (see Table 2for the StepNOs associated with each step). In this plot, it can be observed that the largest decrease in fouling, which is indicated by the RSL, occurs during the first acidic dosing [1 – 14 min], and that after the first acidic dosing RSL only changes slightly [56-76 min] (Figure 9A). This was a recurring trend; thus, it was determined that the first caustic cycle and the first acidic cycle should be used in the present study for constructing a fouling removal model. It can also be observed that there are four peaks in conductivity, two of which have conductivity values between 90-130 mS/cm and the others have conductivities of 40 – 70 mS/cm (Figure 11B). The former corresponds to alkaline treatment and the latter to acidic treatment. This is the desired cleaning procedure because there is no ambiguity.



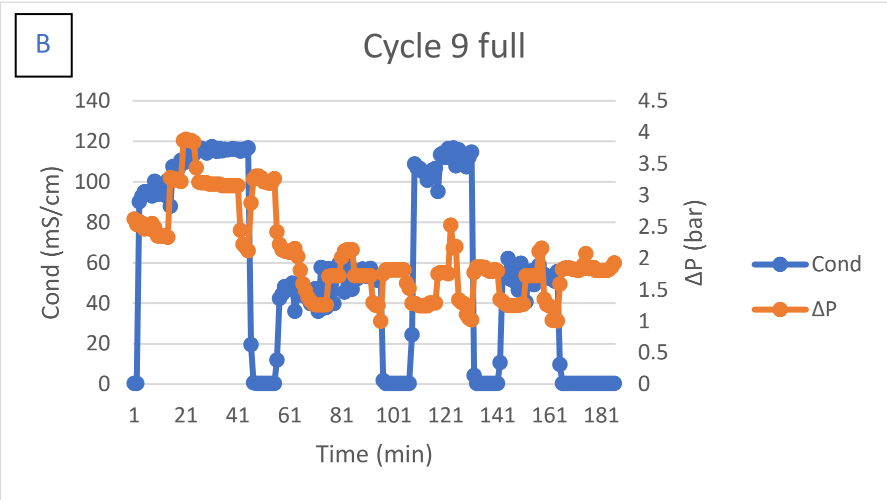


Figure 11. A typical CIP cycle. The cycle alternates between caustic and acidic treatment twice. RSL and stepNO as a function of time (A). Cond and pressure drop as a function of time (B). CIP: cleaning in place. RSL: relative soil level

3.3.2 Investigating cycles that differ from standard operating procedures

When examining the cycles, many cycles deviated from the SOP cycle of caustic, acidic, caustic, acidic treatment. Figure 12 is an example of a cycle that deviates from SOP where the top figure shows the StepNO and RSL as a function of time, while the bottom figure shows the conductivity and pressure drop as a function of time. Figure 12A shows that either caustic or acidic dosing, and circulation occurred, when according to SOP both should be performed. The issue is that there should be a caustic dosing, but the conductivity meter is indicative of an acid (40-70 mS/cm). Note that the conductivity values reported have been compensated for temperature dependency. There are several possible explanations for this: 1) the conductivity data accuracy is not good enough which means that it is not possible to distinguish between acidic and caustic, 2) the caustic detergent concentration changed which affected the conductivity or 3) The operator made a mistake and applied acidic dosing instead of caustic. The deviation from the recommended operation along with the operators stopping at different points during CIP makes distinguishing dosing 1 and dosing 2 as well as circulation 1 and 2 unreliable. To resolve this problem separate tags would have to be introduced for acidic and caustic, and for the number i.e., whether it is acidic dosing 1 or acidic dosing 2. Another alternative is to improve the instrumentation with for example a better conductivity meter or a turbidity meter. Based on these observations, 10 CIP cycles adhering to SOP were chosen for analysis using MARS.

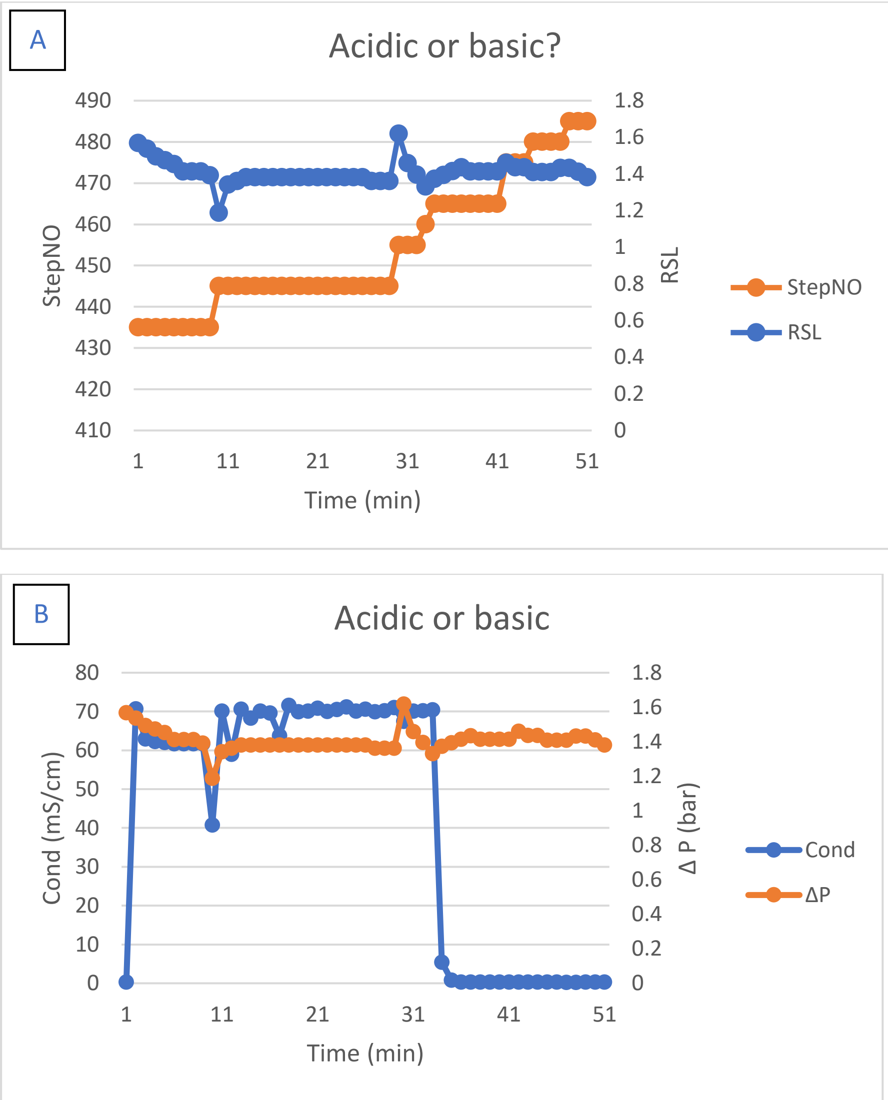


Figure 12. An example where it is impossible to distinguish between acidic and caustic dosing. StepNO 435 is the caustic or acidic dosing. A) shows the StepNO and RSL as a function of time, while B) shows the conductivity and pressure drop as a function of time. SOP suggests caustic, conductivity indicates acidic. RSL: relative soil level. SOP: standard operating procedure.

3.3.3 Analysis of caustic cycles - MARS

Under controlled conditions, a slight increase in RSL is expected due to swelling, however, this could not be observed in the 10 CIP cycles and the data were scattered. Therefore, caustic circulation which is a longer process of around 30 min was studied instead. In caustic circulation, the RSL almost decreases linearly with time except for the initial start where there is a small increase in RSL (Figure 13). Since previous Tetra Pak studies showed that fouling removal is more complex than a linear process, MARS was used but MLR will be used as a comparison. The best results were achieved by using MARS, scaling all independent variables except time and scaling RSL based on initial value followed by standard scaling which resulted in an R^2 of 0.86 (Table 7)(Figure 13). It is possible that this increase is the one that was expected during caustic dosing, but it was delayed due to fouling amounts, flow rates etc. The small change in RSL is consistent with the results of Hagsten (2016), where it was concluded that proteins depolymerize during caustic treatment, but they do not dissociate from the fouling network.

Table 7. A model for RSL as a function of time, temperature and flow rate during caustic circulation was developed using MARS. The R^2 values of the model based on different types of scaling. The best scaling method was used for the remainder of analysis on caustic cycles. RSL: relative soil level. MARS: multiple adaptive regression splines.

	No scaling	Initial value scaling on RSL	All independent variables scaled except time, RSL scaled twice
Model	0.69	0.85	0.86

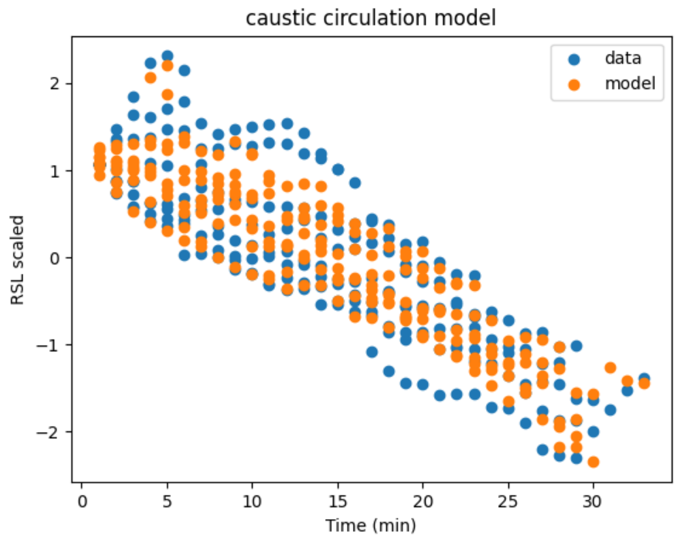


Figure 13. The caustic circulation model where RSL scaled is shown as a function of time. RSL: relative soil level.

The R^2 of the validation cycles were 0.67, 0.72, 0.07 and 0.53 respectively (Table 8). Lower degrees of MARS resulted in a worse fit for the model and slightly worse for the validations. The validations that were a part of the model performed better than the other ones as expected (Figure 14A-B and C-D). The predictions are worse than the predictions in the fouling development which is expected since fouling removal is a more complicated process, where in addition to the process conditions, the composition and amount of detergent play an important role. Nevertheless, the predictions are within the same order of magnitude which suggests that there is no overfitting. Using more cycles to construct the model may improve the predictions as ten cycles may not capture the variation in the flow rates and ΔT.

Table 8. A model for RSL as a function of time, temperature and flow rate during caustic circulation was developed using MARS. The R^2 values of the model and four different validations based on MARS degree 1, 2 and 3 are shown. RSL: relative soil level. MARS: multiple adaptive regression splines.

	Degree = 1	Degree = 2	Degree = 3
Model	0.711	0.785	0.861
Validation 1	0.63	0.63	0.67
Validation 2	0.67	0.86	0.72
Validation 3	0.55	-0.1	0.07
Validation 4	-0.3	0.48	0.53

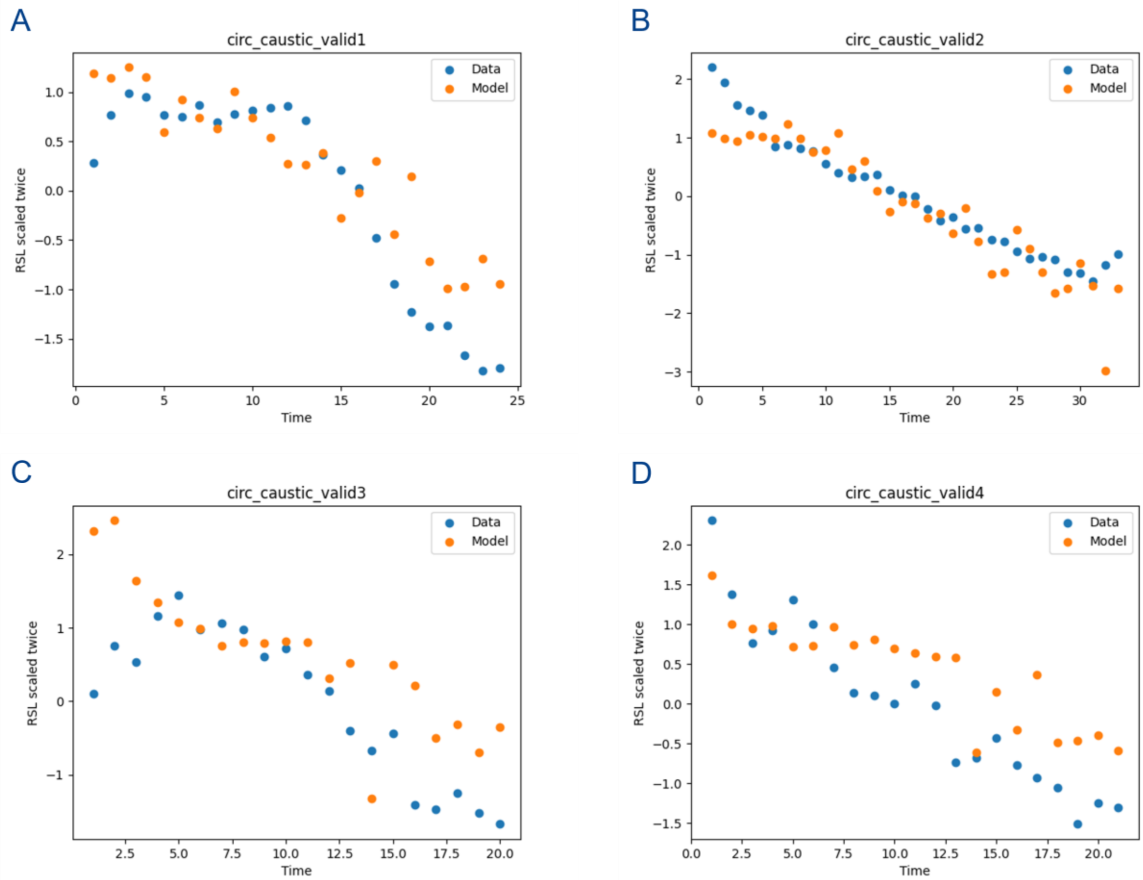


Figure 14. RSL as a function of time. The different validations used for caustic circulation. Validation 1 and validation 2 are from the same dataset as the model (A-B), while the other two are from a different dataset (C-D). RSL: relative soil level.

3.3.4 Analysis of caustic cycles - MLR

Multiple regression was also performed since caustic circulation looked close to linear, using the same independent variables and scaling as MARS, to investigate if it resulted in better predictions. The model performed notably worse (R^2 = 0.77) which is expected since there were increases in RSL at the start of caustic circulation which caused deviations from linearity (Figure 15). The R^2 value of the validations were: 0.69, 0.80, -119 and 0.32, respectively, indicating that the spline model was better

(Figure 16). Once again, the validation cycles that were a part of the dataset used for the model (Figure 16A-B) performed better than the ones from a different dataset (Figure 16C-D)

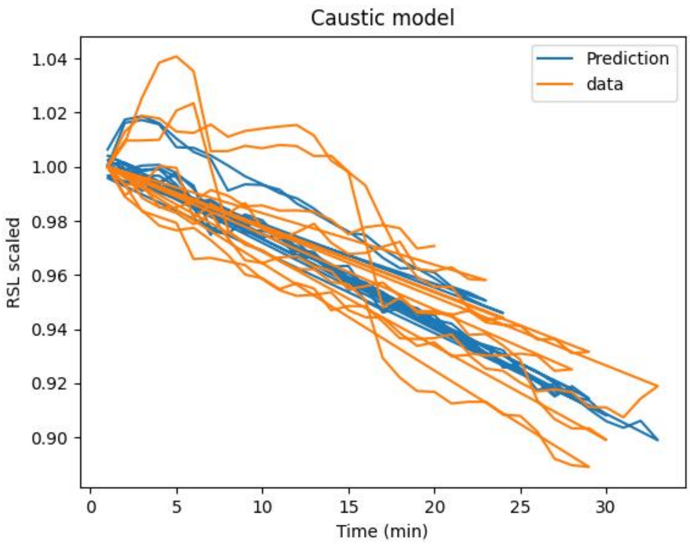


Figure 15. RSL as a function of time. The caustic model generated from multiple regression. RSL: relative soil level.

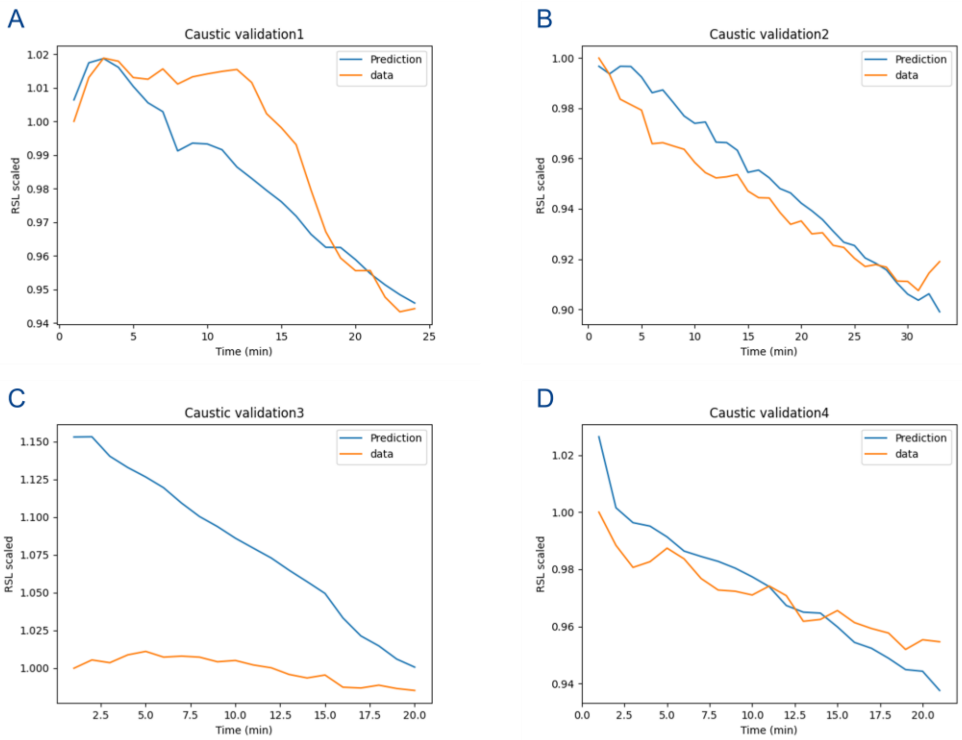


Figure 16. RSL as a function of time. The caustic validation cycles using multiple regression. Validations from the same dataset as the model (A-B), validations from a different dataset (C-D). RSL: relative soil level.

3.3.5 Analysis of acidic cycles

Based on the SOP cycle and previous studies, an exponential decrease in RSL is expected during the acidic circulation (Hagsten, 2016). However, when examining the 10 cycles closely, the decrease in RSL appeared to be connected to how much time had passed since acidic treatment started rather than the cycle step, since cycles with longer acidic dosing saw an exponential decrease in RSL during the dosing phase. With this hypothesis, it was reasonable to analyze dosing and circulation together, to capture the exponential decrease. Since both MLR and PLS are linear models, they would not capture the exponential decrease in RSL, therefore only MARS was used. First, different types of scaling were performed to investigate which one performed the best (Table 9). Subsequently, different degrees of MARS were used, with the third-degree MARS model capturing the acidic dosing and circulation the best (R^2 =0.91) (Figure 17).

Table 9. A model for RSL as a function of time, temperature and flow rate was developed using MARS for the acidic cycles. The R^2 values for the model for different types of scaling, the best method was chosen. RSL: relative soil level. MARS: multiple adaptive regression splines.

	No scaling	Initial value scaling on RSL	All independent variables scaled except time, RSL scaled twice
Model	0.86	0.91	0.91

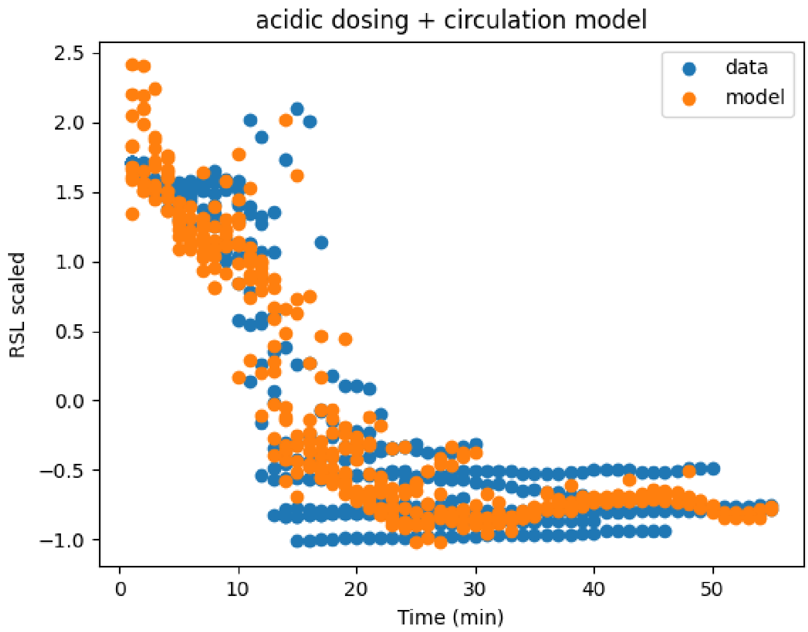


Figure 17. RSL as a function of time. Acidic dosing and circulation model that was obtained based on 10 CIP cycles. RSL: relative soil level.

When validating the model with four different cycles, each cycle had a few datapoints which were two orders of magnitude off from the expected values, indicating that there is an overfitting (Figure 18). Interestingly, the cycles that were from the same dataset as the model also had large deviating values (Figure 18A-B). The overfitting could be caused by certain hinge functions in the spline model or by a combined effect of the independent variables. All validation data had independent variables that were within the range of the model's values. Decreasing the degree to one gave similar results (

Table 10). Since the trend was not linear, multiple regression and PLS would not give accurate results for the acidic cycles. Another explanation could be that there is a large variation between cycles and more cycles are needed to capture this property.

Table 10. MARS was used to construct a model for RSL over time, temperature, and flow rate for acidic cycles. The R^2 of the model and four different validations based on MARS degree 1 – 3 are shown. RSL: relative soil level. MARS: multiple adaptive regression splines.

	Degree = 1	Degree = 2	Degree = 3
Model	0.765	0.896	0.913
Validation 1	-506	-1017	-8858
Validation 2	-71.1	-540	-397
Validation 3	-1707	-111931	-4827
Validation 4	-1540	-125000	-2915

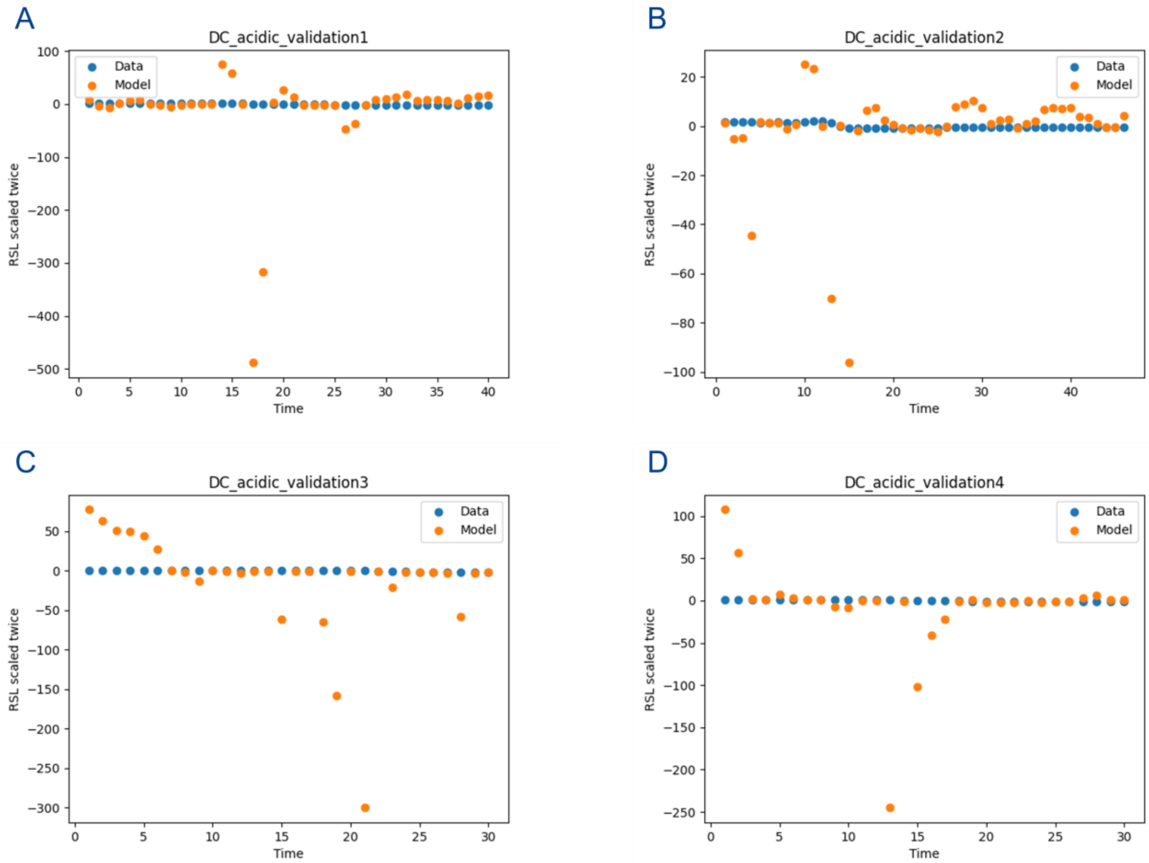


Figure 18. RSL as a function of time. Acidic validation data. Validation 1 and 2 are from the same model (A-B) while validation 3 and 4 are from an independent dataset (C-D). RSL is first scaled by initial value and then by standard scaling. RSL: relative soil level.

3.3.6 Fouling removal summary

For caustic circulation, MARS performed better than MLR, however, the predictions were not very accurate which may be due to the low number of cycles used for training or overfitting. Despite this, the general trend seems to be captured by the model. For the acidic cycles only, MARS was used since the data suggested that the process is exponential. The R^2 for the model is good (0.913) but the validations contained several predicted values that deviated with a factor of 100 from the expected value, suggesting that there is an overfitting in certain hinge function in the MARS model or that the combination of independent variables in the validation data generates incorrect values by the model. A larger sample of CIP cycles may amend these issues.

4.Technical and environmental relevance

A fouling development model can help with short-term scheduling, by using the RSL as an indicator of when cleaning is necessary, less cleaning cycles will be required which reduces energy consumption and increases productivity. A fouling removal model facilitates understanding of how long each caustic and acidic cycle should be to ensure that the equipment is clean. By optimizing cleaning cycles, the consumption of chemicals and energy can be decreased.

The models derived in this study are data-driven, in contrast to the physical models found in literature. The main difficulty encountered with this data-driven approach was that the industry data deviated from Tetra Pak's SOP. Customers tend to develop their own procedures as they figure out what works and what does not, in certain cases shortcuts are taken and in other mistakes are made. Consequently, identifying SOP CIP cycles reliably is troublesome, however, if SOPs were followed and better tags introduced, the data-driven approach is feasible. Human errors will always remain a factor which is why certain filters will have to be applied, given a sufficiently large dataset with the right filters, reliable models for fouling development and removal could be constructed.

5.Conclusion

This study concluded that fouling increases linearly with time during production which is consistent with previous studies. A workflow for cleaning and pre-processing the data was established, however due to deviations from SOP, a smaller subset of CIP cycles had to be used. Fouling decreases almost linearly during caustic treatment and exponentially during acidic treatment. The predictions were fairly accurate for the production, however, there were large errors in the caustic and acidic validation. This could be caused by overfitting and using more cleaning cycles may ameliorate regression results. Alternatively, other non-linear regression models can be used to capture the changes in RSL during cleaning. However, first the tags differentiating caustic and acidic cycles would have to be introduced.

6.Future work

The primary step in a future study would be to develop fouling removal models using a larger dataset to see if the models improve. Furthermore, it would be interesting to see what effect water recirculation has (which is done after each production cycle). Another topic to consider is the two different types of cleaning cycles: *fullCIP* and *ICCIP*, which of these modes of cleaning is the most efficient in terms of productivity, energy consumption and chemical use.

7.References:

Allison, P.D. (1998). *Multiple regression: a primer*. Thousand Oaks, Ca: Pine Forge Press.

Bohnet, M. (1987). *Fouling of Heat Transfer Surfaces*. Chemical Engineering Technology, 10: 113125. DOI :10.1002/CEAT.270100115

Bylund, G. (2015). *Dairy processing handbook*. 3rd edition. Tetra Pak Processing systems AB SE 221 86 Lund, Sweden.

Edreis, E, Petrov A (2020). *Types of heat exchangers in industry, their advantages and disadvantages, and the study of their parameters*. IOP Conf. Ser.: Mater. Sci. Eng. 963, 12-27.
DOI: https://doi.org/10.1088/1757-899X/963/1/012027

Friedman, J. (1991). *Multivariate Adaptive Regression Splines*. Annals of Statistics.19(1), 1-67. DOI: https://doi.org/10.1214/aos/1176347963

Fryer, P. J., P. T. Robins, C. Green, P. J. R. Schreier, A. M. Pritchard, A. P. M. Hasting, D. G. Royston, and J. F. Richardson (1996). *A Statistical Model for Fouling of a Plate Heat Exchanger by Whey Protein Solution at UHT Conditions*. Trans. I. Chem. E. Purt C. 74(4), 189-199. DOI: https://doi.org/10.1205/096030896531181

Geladi P, Kowalski Bruce R. (1986). *Partial least-squares regression: a tutorial*. Analytica Chimica Acta. 18, 1-17. DOI: https://doi.org/10.1016/0003-2670(86)80028-9.

Goode K.R, Christian G.K.,. Fryer P.J. (2016). *Chapter 32 - Improving the Cleaning of Heat Exchangers*, Editor(s): Huub Lelieveld, John Holah, Domagoj Gabrić, In Woodhead Publishing Series in Food Science, Technology and Nutrition. Handbook of Hygiene Control in the Food Industry (Second Edition). Woodhead Publishing, Pages 465-489, ISBN 9780081001554. DOI: https://doi.org/10.1016/B978-0-08-100155-4.00032-7.

Grijspeerdt K, Mortier L, De Block J, Van Renterghem R (2004). *Applications of modelling to optimise ultra high temperature milk heat exchangers with respect to fouling*. Food Control. 15(2), 117-13. DOI: https://doi.org/10.1016/S0956-7135(03)00023-9.

Hagsten, C. (2016). *Cleaning of ultra-high temperature milk fouling -Structural and compositionl changes*. Doctoral Thesis (compilation), Physical Chemistry. Department of Chemistry, Lund University.

Jun, S. and Puri, V.M. (2005), *Fouling models for heat exchangers in dairy processing: a review.* Journal of Food Process Engineering. 28(1), 1-34. DOI: https://doi.org/10.1111/j.1745-4530.2005.00473.x

Kim J. H. (2019). *Multicollinearity and misleading statistical results*. Korean journal of anesthesiology, 72(6), 558–569. DOI: https://doi.org/10.4097/kja.19087

Muller-Steinhagen H, Zettler HU. (2011). *Heat exchanger fouling. Mitigation and cleaning technologies.* 2nd edition. PP publico publications, Israel.

Ritter, R.B. (1983). *Crystalline Fouling Studies*. ASME journal of heat and mass transfer, 105(2), 374–378. DOI: https://doi.org/10.1115/1.3245588.

Shashi Menon E.(2015). *Chapter Five - Fluid Flow in Pipes*. Editor(s): E. Shashi Menon, Transmission Pipeline Calculations and Simulations Manual. Gulf Professional Publishing. Pages 149-234. ISBN 9781856178303. DOI: https://doi.org/10.1016/B978-1-85617-830-3.00005-5.

Tetra Pak. (2015). *Cleaning in place handbook*. Tetra Pak processing systems AB. Lund, Sweden.

W.R. Paterson, P.J. Fryer. (1988). *A reaction engineering approach to the analysis of fouling.* Chemical Engineering Science. 43(7),1714-1717. https://doi.org/10.1016/0009-2509(88)85166-2.