Prediction and Optimization of Paper Quality

HENRIK EKHOLMER
Prediction and Optimization of Paper Quality

HENRIK EK HOLMER

Master’s Thesis in Optimization and Systems Theory (30 ECTS credits)
Master Programme in Applied and Computational Mathematics (120 credits)
Royal Institute of Technology year 2016
Supervisor at ABB: Carl-Fredrik Lindberg
Supervisor at KTH: Johan Karlsson
Examiner: Johan Karlsson

TRITA-MAT-E 2016:73
ISRN-KTH/MAT/E--16/73--SE

Royal Institute of Technology
SCI School of Engineering Sciences
KTH SCI
SE-100 44 Stockholm, Sweden
URL: www.kth.se/sci
Abstract

A problem in the paper industry is that most paper quality properties can only be measured in lab after a full tambour of paper is produced. The tambour length is normally about 20-40 km and takes about one hour to produce. This may lead to one hour production and several km of paper becomes wasted due to poor paper quality. To reduce this problem, prediction models can be used to estimate the paper quality properties online. By using these models, a control strategy can be developed, which make sure that the paper quality properties are fulfilled. Optimization has here been used to find a control strategy that minimizes the cost to produce paper of with desired paper quality properties.

In the thesis, focus has been to find models for prediction of paper quality properties, which includes synchronizing data in different parts of the paper machine and lab, variable selection and filtering. Focus has also been on minimizing production cost, utilizing the models of paper quality properties. A sensitivity analysis has been done for a number of variables in order to increase the understanding of the optimization.

Prediktering och optimering av papperkvalitén vid papperstillverkning

Sammanfattning


Contents

1 Introduction .................................................. 4
  1.1 The Paper Mill .............................................. 4
    1.1.1 Stock preparation ................................... 4
    1.1.2 Paper machine ......................................... 5
  1.2 Paper qualities ............................................ 6

2 Modeling of paper properties ................................. 9
  2.1 Theory ...................................................... 9
    2.1.1 Prediction .............................................. 9
    2.1.2 Variable selection .................................... 10
    2.1.3 Outlier removal ....................................... 12
    2.1.4 Short fiber signal .................................... 13
    2.1.5 Tambour data .......................................... 14
  2.2 Model of tank dynamics .................................... 15
  2.3 Prediction model .......................................... 17
  2.4 Results .................................................... 20

3 Optimization .................................................. 23
  3.1 Problem formulation ...................................... 23
  3.2 Costs ....................................................... 24
  3.3 Reformulating of the problem ............................. 25
  3.4 Convexity .................................................. 26
  3.5 Results ..................................................... 26
  3.6 Discussion ................................................ 38

4 Conclusions and future work ................................. 40
  4.1 Optimize the change in fiber distribution ............... 40
  4.2 Improve the prediction model ............................. 40
  4.3 Improve the optimization model .......................... 41
1 Introduction

A problem in the paper industry today is that most paper qualities can only be measured in a lab after a full tambour is produced. The tambour length is normally about 20-40 km and takes about one hour to produce. This may lead to one hour production and several km of paper becomes wasted due to poor paper quality. To reduce this problem, prediction models can be used to estimate the paper quality properties on-line. By using these models, a control strategy can be developed, which make sure that the paper quality properties are fulfilled. Optimization has here been used to find a control strategy that minimizes the cost to produce paper of with desired paper quality properties.

A brief introduction on paper making and the paper mill at Gruvön is treated in section 1. The prediction model is presented in section 2 and will work as a foundation to the optimization problem in section 3. Section 4 gives suggestions of future work.

1.1 The Paper Mill

The paper mill at Gruvön has 6 different paper machine that together produces 685 000 ton paper per year. The focus on this thesis is on paper machine 2 which produces kraft paper. This kind of paper has high tensile strength. The paper machine is running around the clock. The paper consists of two layers which are pressed together. The bottom layer has a more smooth surface with good printability while the top layer stands for the strength properties of the paper. All fibers going in to paper machine 2 is bleached, this is done to smoothen the paper and to get a brighter color. [5] [6]

Paper making can be divided into three part, pulp production, stock preparation and the paper machine. In the pulp production the wood chips is processed into pulp. This is done either manically or chemically. The pulp is then refined and mixed together with chemicals in the stock preparation part. The paper machine main purpose is to form and drain the water from the pulp. In this thesis the focus will be on the stock preparation and the paper machine.

1.1.1 Stock preparation

Paper making is a complex process with lots of parameters. The manufacturing of paper can be divided in to two section, the stock preparation and the paper machine. In the stock preparation section the pulp is refined and mixed together with chemicals. The pulp consists of long fiber, short fiber and broke. The long fibers are used to give the paper strength and the short fibers give the paper a smooth surface. Broke is already used paper that did not fulfill the paper quality demands and is recycled to pulp and used again. [4]

All fibers go through refiners. When you refine fibers they get more fuzzy and will easier get stuck to each other, which increase strength but will decrease the stretch of the paper. However, to much refining can destroy the fibers and
they will be of no good use. The fibers in the broke has already been refined once, but you refine them again to refresh them. [7]

After the different pulp flows has gone through the refiners they will get mixed in the mixing tanks and chemicals are added, see figure 2. The paper at this paper machine has two layer, bottom and top layer. The bottom layer is the outside of the package and require a smooth surface. The top layer will give the paper strength. This means that the pulp is divided in to two mixing sections. After the pulp is mixed chemicals are added. [5]

1.1.2 Paper machine

The paper machine can be divided in to four sections, forming, wet press, drying and calendering, as in figure 3. In this paper machine the forming section consists of two head boxes, top and bottom, which are pressed together in the wet press section. In the forming section the pulp is spouted out on a wire with a moisture content of 99.7 %. The wire velocity is the velocity of the wire and also the speed of the production of the paper. The difference in velocity of the jet and the wire will effect the fiber orientation and therefore the strength difference in the machine direction and the cross direction. On the wire there are vacuum suction boxes which suck out water from the wire. The pulp leaving the forming section has decreased its moisture content to 80 %. [1]

In the wet pressing the bottom and top layer are press together and reduces the moisture content to 60 %. More wet pressing should increase the tensile strength. After the wet press the pulp looks more like a paper and is transported in to the drying section. Here the paper is dried by passing several steam heated
cylinders down to a moisture content of 7%. Before the paper is reeled on a large tambour it get pressed by two hot calenders in order to get a smoother surface. [1]

1.2 Paper qualities

The variables mentioned to impact on the paper properties below, is according to theory, and may not always match with the findings we have made. Most of the information is from the visit at Gruvön [5] and Innventia [7], but also from [3] and [4].

**Starch** is a chemical that is positive correlated with the strength qualities of the paper, but is negative correlated with the stretch. However, to much starch will create foam which is bad for the paper.

**Chalk** and **Talc** are cheep and are used to fill out the paper in order to get more paper to a relatively lower price. But it is also used to get better print ability. Copy paper consist of relatively much chalk and clay.

**Glue** is added to keep the resulting paper more resistance to water.

**Machine direction (MD) and Cross direction (CD)**
Most of the strength properties are measured in both the machine direction and cross direction of the paper. This is important since a paper is less useful if it is much stronger in one direction than the other. But it is not necessary to have equal strength in both directions, a good example is a grocery bag. It is easier to get more strength in the machine direction than the cross direction. The main parameter effecting the CD/MD distribution is the difference in velocity between the jet and the wire. Another parameter that will effect the CD/MD distribution is the strain of the paper machine, i.e. the difference in velocity between different operating groups. The data for the strain is not included in the data set which this report is based on.

**Tensile strength MD/CD**
The tensile strength is the largest pull force a paper can handle before it breaks, measured in $kN/m$. The tensile strength is positively correlated with refining, wet pressing, improved formation and starch. Formation is the smoothness of the paper, i.e. how the basis weight is distributed over the paper.

**Stretch MD/CD**
The extension of the paper before burst, measured in % compared to the original length. All refining and strain of the paper machine, i.e. difference in velocity between different part of the paper machine, are negative correlated with the stretch. Improved formation leads to increased stretch.

**Breach work MD/CD**
The total amount of work to pull apart a paper, measured in $J/m^2$. The breach work is positive correlated with the tensile strength and stretch, especially the stretch. It is very formation dependent.
Figure 4: Test for tensile strength, stretch and breach work.

**Burst strength**
The burst strength is the largest perpendicular pressure a paper can handle, measured in kPa. It benefits of increased tensile strength and stretch, especially in the MD direction.
2 Modeling of paper properties

In this chapter (2) we will go through how the prediction model is created. The theory for this part is about variable selection and outlier removal. This is presented in section 2.1.

The signal treatment before we can use the signals in the prediction model is presented in section 2.2 and 2.3. Where we in section 2.2 replace and remove bad signals. In section 2.3 we sync the signals in time by considering the tank dynamic.

In section 2.4 we explain the prediction model and show the result in 2.5.

2.1 Theory

We will introduce some new variables which we will use throughout this section. Let \( y \in \mathbb{R}^{n \times 1} \) be the dependent variable, \( X \in \mathbb{R}^{n \times m} \) contains the independent variables and \( \beta \in \mathbb{R}^{m \times 1} \) is the coefficients for the equation

\[
y = X \beta
\]

where \( n \) is the number of samples and \( m \) is the number of variables.

Now we define two notations,

\[
\|y\|_1 = \sum_i |y_i|
\]

and

\[
\|y\|_2 = \sqrt{\sum_i y_i^2}.
\]

2.1.1 Prediction

There are several ways of preforming a regression, in this report the normal least square regression will be used. But the regression will not be done on the whole data set. Instead we will use a modeling set and a validation set. The proportion between the modeling set and the validation set will affect the result. This will reduce the problem of overfitting, i.e. be good at prediction on the test data but bad at the test data. Then \( X_{train} \) is the modeling part of \( X \) and \( X_{test} \) is the validation part of \( X \), and in the same way for \( y_{train} \) and \( y_{test} \).

The regression is made by the minimizing the sum of squares of the training set,

\[
\min_{\beta} \|y_{train} - X_{train} \beta\|_2
\]

where \( \beta \) is a vector with the model coefficients. The performance is then measured on the test set

\[
Performance = \|y_{test} - X_{test} \beta\|_2
\]
where \( \beta \) is the minimizing vector in equation (4).

To validate a set of variables chosen for the prediction, one good method is to look at the sustainability under a variation of test sets for the variables. If a variable performs bad on the test set then it is removed from the model. But if there are variables that does not make a big difference, one must take the physical aspect and the variation into account in the decision.

### 2.1.2 Variable selection

A big problem when you have a large amount of variables to chose from in the prediction model is selecting which variables that are the most important. Below is two methods for solving that problem, The Lasso and The Subset Selection. The best way is of course to be physically correct, but in a complex system it is not always easy to see the physical connection and new unknown variables may be found that improves the model.

#### The Lasso \([2]\)

In a normal regression, the sum of squares of the prediction error is minimized.

\[
\min_{\beta} \| y - X\beta \|_2
\]  

(6)

In the Lasso regression a regularization term, \( \lambda \| \beta \|_1 \), is added,

\[
\min_{\beta} \| y - X\beta \|_2 + \lambda \| \beta \|_1
\]  

(7)

where \( \lambda \) is a user defined parameter. In Matlab a suitable set of different \( \lambda \) are calculated. The idea is to punish the usage of many nonzero coefficients, but it will also punish large coefficients. As the \( \lambda \) parameter is increased more coefficients become zero. The covariates with nonzero coefficients are the most important ones. Note that the Lasso algorithm does not minimize the same criteria as the least square. Hence, other variable combinations than the ones found by Lasso may give a better fit to the data.

#### The Subset selection \([2]\)

This is a greedy algorithm, which does the best for the current position. It works in two ways, by adding a variable as in method 1 or remove a variable as in method 2.
1. Let $\Omega = \{1, 2, \ldots, m\}$ and $S = \emptyset$.
2. For each element $i \in \Omega \setminus S$ (for each element, $i$, in $\Omega$ not included in $S$)
   - Let $S_i = S \cup \{i\}$.
   - The performance error, $r_i = \min_{\beta_j} \|y - \sum_{j \in S_i} \beta_j x_{i,j}\|_2$.
3. Let $S$ be the set $S_i$ with the lowest performance error, $r_i$.
4. Repeat 2 and 3 until the number of variables in $S$ and/or the performance error of $S$ is satisfied.

Method 1: Subset algorithm, adding.

1. Let $S = \{1, 2, \ldots, m\}$.
2. For each element $i \in S$
   - Let $S_i = S \setminus \{i\}$.
   - The performance error, $r_i = \min_{\beta_j} \|y - \sum_{j \in S_i} \beta_j x_{i,j}\|_2$.
3. Let $S$ be the set $S_i$ with the lowest performance error, $r_i$.
4. Repeat 2 and 3 until the number of variables in $S$ and/or the performance error of $S$ is satisfied.

Method 2: Subset algorithm, removing.

The problem with this algorithm is when many variables perform good together but not alone. For example, let $\{x_1, x_2, x_3, x_4, x_5\}$ be a set of variables where

$$\|y - (\gamma_0 + \gamma_1 x_1)\|_2 < \|y - (\theta_0 + \theta_1 x_i)\|_2 \quad \forall i$$

and

$$\|y - (\alpha_0 + \alpha_2 x_2 + \alpha_3 x_3)\|_2 < \|y - (\beta_0 + \beta_1 x_1 + \beta_i x_i)\|_2 \quad \forall i$$

where $\alpha, \beta, \gamma$ and $\theta$ is the minimizing coefficients for each regression. This means that method 1 will chose the set $\{x_1, x_i\}$ and not the best set, $\{x_2, x_3\}$.

A solution to this problem is to introduce a parameter $k$, which stands for the number of variables for each set in method 3 below.

1. Let $\Omega = \{1, 2, \ldots, m\}$ and $S = \emptyset$.
2. Then let $P$ be the set of all combinations of $k$ elements in the set $\Omega \setminus S$.
3. For each combination $P_i \in P$
   - Let $S_i = S \cup P_i$.
   - The performance error, $r_i = \min_{\beta_j} \|y - \sum_{j \in S_i} \beta_j x_{i,j}\|_2$.
4. Let $S$ be the set $S_i$ with the lowest performance error, $r_i$.
5. Repeat 2 - 4 until the number of variables in $S$ and/or the performance error of $S$ is satisfied.

Method 3: Subset algorithm with parameter $k$, adding.

This can be very time consuming for large $k$ and large set of variables in $X$. The same change can be done on the removing version, method 2.
2.1.3 Outlier removal

The least square regression is constructed to perform best with normal distributed measurement errors. But that is not always the case. Sometimes the data has outliers which deviate strongly from the correct values and this will have a large effect on the estimated coefficients, since the square of the error is minimized. The removing of outliers without losing excitation is a very important part of the modeling.

The Median Filter [8]

The median filter is a good tool for removing of outliers of the kind of data in this report. It has one parameter, \( r \), which controls the range of samples which the median is calculated. The median filter will not change a value that is constant over \( r - 1 \) samples. Note that the parameter \( r \) is only defined for odd numbers.

For example, if \( r = 3 \) then every sample, \( s_t \), is the median of the samples \( s_{t-1}, s_t, s_{t+1} \). This means that it will not change a value that is constant over two samples when \( r = 3 \).

![Figure 5: Median filter, \( r = 3 \). Will remove outliers that are 1 sample wide.](image)

This is very effective for the signals in this data set since the paper machine produces a tambour which is intended to have the same quality all the way through the whole tambour. Which means that most of the signals should be relatively constant through a whole tambour. This means that we do not have to be as careful with the median filter on this data set, as for a data set that do not have this characteristics.

LTS - Least Trimmed Squares [1]

The Median filter works on data that have many samples during a tambour. But there is only one lab sample per tambour, which means that there is not any other samples to compare with in order to detect outliers. However, lab measurements tends to be trustworthy. But if the paper sample that the lab measurement is made on is a bad represent of the whole tambour, e.g. have
been damage, the sample should not be included in the regression. This is not very common. A more common situation where the above filters will not work is when there is a measurement error in the on-line data over too many samples. If the error is too big it will affect the estimated coefficient. These two problems where a tambour needs to be removed from the regression is done by the LTS method.

In a normal regression, the sum of squares of the prediction error is minimized,

$$\min_{\beta} \| y - X\beta \|_2 = \min_{\beta} \left( \sum_{i=1}^{n} (y_i - x_{i,:}\beta)^2 \right)$$

(10)

where $x_{i,:}$ is a row vector containing the variables’ value for sample $i$. In the LTS regression a parameter $k < n$ is added and the criteria is changed to,

$$\min_{\beta} \sqrt{\sum_{i \in I} (y_i - x_{i,:}\beta)^2}$$

(11)

where $I = \{ i : (y_i - x_{i,:}\beta)^2 < K \}$, where $K$ is selected so that $I$ contains $k$ elements. Which means that the regression is made without the $n - k$ samples with the largest prediction error, i.e. those with potential outliers.

2.1.4 Short fiber signal

In figure 6, magnitude of the original signal $f_{118}$ (blue) is much higher than the calculated signal (red). The calculated signal is based on the model in the same figure, the sum of the flows from the refiners should be equal to the flow into the mixing tanks

$$f_{118} = f_{162} + f_{163} - f_{154}$$

(12)

And since the signals from the calculated signal is more trustworthy we replace the original signal with the calculated.
2.1.5 Tambour data

The basis weight is highly correlated with the tensile strength, which makes it easy to see if the on-line data and the lab data is in sync. It seems like there is a delay in the lab data compare to the on-line data. If we shift the on-line data one tambour later in time, i.e. the current tambour’s lab data is matched with the on-line data from the previous tambour. As we can see in figure 7 we will get a much better correlation between the tensile strength and the basis weight. This shift is done on all the on-line data in order to get it synced with the lab data.

Figure 6: Original and calculated short fiber flow to the mixing tank, top layer.
2.2 Model of tank dynamics

All the on-line data are sampled at the same time but at different parts of the paper machine. This means that the data are not synced in the aspect of a point on the resulting paper. In order to get that synced, a tank model is used to calculate new signals, based on the model in figure 8. Since the content of the tank are assumed to be mixed completely we can not treat the signal as a plug flow and only shift the signal in time.

Figure 7: Result of with and without the tambour shift.
Figure 8: Simplified model of the refine and the mixing section of figure 2 where $P^{(i)}$ is the refiner for fiber $i$.

The equations for the tank model are based on the assumption that the volume is constant in all tanks, which leads to that the flow into the tank is equal to the flow out from the tank. This is a good assumption since the tank level is roughly constant. There are also measurement errors in the data which are easier reduced with this assumption.

We introduce new variables in table 4 below. They have the following sub-/superscript declaration, $C_{l,k}^{(i)}$ where the $i \in I$ stands for type of fiber, $l \in L$ stands for the level and $k \in K$ stands for type of layer. For example, $C_{2,1}^{(LF)}$ stands for concentration of long fiber in the second level tank for the top layer.
Sets
\[ I = \{ \text{Long fiber, Short fiber, Broke} \} \]
\[ L = \{ 0, 1, 2 \} \]
\[ K = \{ 1, 2 \} \]

Constants
\[ \Delta T \] Sample time, 1 minute. [h]
\[ V_{l,k} \] The volume of tank \( T_{l,k} \). [m³]

Data
\[ F_k \] The total flow going through the tanks for layer \( k \). [m³/h]
\[ P^{(i)} \] The power of the refiner for fiber \( i \). [%]

States
\[ C^{(i)}_{l,k}(t) \] The concentration of fiber \( i \) in tank \( T_{l,k} \). [%]
\[ E^{(i)}_{l,k}(t) \] The refining energy spent fiber \( i \) in tank \( T_{l,k} \). [%]
\[ R^{(i)}_{l,k}(t) \] The refining energy spent on fiber \( i \) going into tank \( T_{l,k} \). [%]

Table 4: Variable declaration for the tank model equations

To follow the concentration through the mixing section the following equation is used,
\[
C^{(i)}_{l,k}(t + 1) = C^{(i)}_{l,k}(t) + \Delta T \cdot \frac{F_k(t)}{V_{l,k}} \left( C^{(i)}_{l-1,k}(t) - C^{(i)}_{l,k}(t) \right)
\] (13)
where \( C^{(i)}_{0,k} \) is the concentration of fiber \( i \) in the refining section going into layer \( k \).

The refining energy is calculated by the following,
\[
E^{(i)}_{l,k}(t + 1) = E^{(i)}_{l,k}(t) + \left( R^{(i)}_{l,k}(t) - \frac{\Delta T \cdot F_k(t) \cdot C^{(i)}_{l,k}(t)}{V_{l,k}} E^{(i)}_{l,k}(t) \right)
\] (14)
where
\[
R^{(i)}_{l,k}(t) = \begin{cases} \sum_{k \in K} F_k(t) \Delta T \cdot P^{(i)}(t) & \text{for } l = 1 \\ \frac{\Delta T \cdot F_k(t) \cdot C^{(i)}_{l,k}(t)}{V_{l,k}} E_{1,k}(t) & \text{for } l = 2 \end{cases}
\] (15)

2.3 Prediction model

In this section the prediction model will be presented. First we will define the variables necessary for constructing the covariates for the model. Then the covariates will be explained and the result will be presented in section 2.4. The model is based on the theory in section 2.1.1 and is calculated by
\[
\min_{\beta} \| y - \sum_{m \in M} x_m \beta_m \|_2
\] (16)
where \( y \in \mathbb{R}^{n \times 1} \) is the paper quality, \( X \in \mathbb{R}^{n \times |M|} \) is the covariates in table 6 and \( \beta \in \mathbb{R}^{|M| \times 1} \) is the model coefficients for the prediction model. \( n \) is the number of samples and \( M \) is the set of the covariates in table 6.

This model enables us to predict the paper qualities based on the control signals from the stock preparation and the paper machine. The prediction model will later be used in the optimization problem in section 3 to lower the production cost. For the optimization problem we also need to predict the steam consumption.

A linear least square model is used. The covariates are selected by using knowledge of the paper qualities from section 1.2 and the subset selection and the Lasso from section 2.1.2.

All the paper qualities are highly dependent on how much materials there are in the paper, in other words the weight of the paper. The measurement for the weight is called basis weight and measures the weight per area. We can see the basis weight as a control signal even though it is controlled by the flow and wire velocity.

### Sets

- \( I \) = \{Long fiber, Short fiber, Broke\}
- \( J \) = \{Chalk, Talc, Starch, Color, Glue\}

### Variables

- \( B \) Basis weight
- \( H^{(j)} \) Flow of chemical \( j \) going into both layers.
- \( F^{(i)} \) Flow of fiber \( i \) going into both layers.
- \( \xi^{(i)} \) Refining energy spent on fiber \( i \) going into both layers.

Table 5: Variable declaration for the constructing of the covariates.

In table 5 we introduce new variables in order to simplify the description of the covariates in the prediction model. The signal for the basis weight, \( B \), and the chemical flow, \( H^{(j)} \), are taken directly from measurement. But the fiber flow, \( F^{(i)} \), and the refining energy, \( \xi^{(i)} \), are calculated from the results in section 2.3 by the following formulas,

\[
F^{(i)} = \sum_{k \in K} F_k \frac{C^{(i)}_{2,k}}{\sum_{i \in I} C^{(i)}_{2,k}} \tag{17}
\]

\[
\xi^{(i)} = \sum_{k \in K} \frac{\Delta T \cdot F_k \cdot C^{(i)}_{2,k}}{V_{2,k}} E^{(i)}_{2,k} \tag{18}
\]
Covariates

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LF · BW</td>
<td>The weight of long fiber per area of the paper.</td>
</tr>
<tr>
<td>SF · BW</td>
<td>The weight of short fiber per area of the paper.</td>
</tr>
<tr>
<td>Broke · BW</td>
<td>The weight of broke per area of the paper.</td>
</tr>
<tr>
<td>Chalk · BW</td>
<td>The weight of chalk per area of the paper.</td>
</tr>
<tr>
<td>Talc · BW</td>
<td>The weight of talc per area of the paper.</td>
</tr>
<tr>
<td>Starch · BW</td>
<td>The weight of starch per area of the paper.</td>
</tr>
<tr>
<td>Color · BW</td>
<td>The weight of color per area of the paper.</td>
</tr>
<tr>
<td>Ref LF</td>
<td>Refining energy spent on long fiber.</td>
</tr>
<tr>
<td>Ref SF</td>
<td>Refining energy spent on short fiber.</td>
</tr>
<tr>
<td>Ref Broke</td>
<td>Refining energy spent on broke.</td>
</tr>
<tr>
<td>Diff Jet - wire BL</td>
<td>Difference in velocity between the jet and the wire for the bottom layer.</td>
</tr>
<tr>
<td>Diff Jet - wire TL</td>
<td>Difference in velocity between the jet and the wire for the top layer.</td>
</tr>
<tr>
<td>Sum vacuum TL</td>
<td>The sum of the vacuum suction boxes for the top layer.</td>
</tr>
<tr>
<td>Sum vacuum BL</td>
<td>The sum of the vacuum suction boxes for the bottom layer.</td>
</tr>
<tr>
<td>Rotor position 25</td>
<td>The rotor position for the refining of long fiber.</td>
</tr>
<tr>
<td>Rotor position 26</td>
<td>The rotor position for the refining of long fiber.</td>
</tr>
<tr>
<td>Sum wet press</td>
<td>The sum of the wet pressing.</td>
</tr>
</tbody>
</table>

Table 6: The covariates for the prediction model.

The first eight covariates are depending on the basis weight. Where the fraction of the total flow for the fiber or chemical is multiplied by the basis weight. One problem with this model is that the weight of a fiber/chemical on the paper is based on the part of the flow and not part of the weight. This is a problem if the flows have different densities. In this model we assume that all densities are the same.

The first three covariates are calculated on the following formula,

\[(i) \cdot BW = B \frac{F^{(i)}}{\sum_{i \in I} F^{(i)} + \sum_{j \in J} H^{(j)}} \quad (19)\]

and the remaining five covariates by,

\[(j) \cdot BW = B \frac{H^{(j)}}{\sum_{i \in I} F^{(i)} + \sum_{j \in J} H^{(j)}} \quad (20)\]

where \(i \in I\) and \(j \in J\). For example, the first covariate in \(X\) is

\[x_{i,1} = LF \cdot BW = B \frac{F^{(\text{Long fiber})}}{\sum_{i \in I} F^{(i)} + \sum_{j \in J} H^{(j)}}. \quad (21)\]
The refining energy is included in the prediction model as refining energy per fiber and is calculated by,

\[ \text{Ref} (i) = \frac{\xi^{(i)}}{F^{(i)}}. \]  

(22)

The reason for this is to normalize the refining based on the amount of fiber. This is important since the paper machine produces paper with different basis weight and therefore different amount of fibers.

The difference in velocity between the jet and the wire can be implemented in two ways, as the difference in speed or as the ratio of the speed. The difference between the two ways is whether to take the total speed in to account or not. By using the ratio, the total speed will be included. In this report we will use the difference in speed because the force that acts on the fiber should be independent of the total speed. There are one variable for the top layer and one for the bottom layer. The top layer should be the most interesting since there is more focus on the strength there than in the bottom layer. In the top layer the speed of the jet is higher than the speed of the wire and in the bottom layer it is the opposite. This should not make a difference and therefore both signal is the absolute value of the difference.

On the wire there are vacuum suction boxes which is the first stage of removing water from the wire. This is implemented as the sum of the pressure for all vacuum boxes on the wire. There is one variable for the top layer and one for the bottom layer. The problem here is the dependence on how the pressure is distributed, but this is neglected here.

The wet press should be positively correlated with the tensile strength and it is divided in three parts, first, second and third press. The dependence of the velocity should not be as big as for the suction boxes. This is neglected in this report.

The rotor position for refiner 25 and 26, which are used in the refining for long fiber. The rotor position is a measurement on how tight the fibers are refined.

2.4 Results
Table 7: Model coefficients, where BW stands for basis weight.

We start by looking at the fibers and the steam. One can see how short fiber requires more steam to dry the paper than long fiber. This is expected since short fiber is smaller and will therefore be more compact which makes it harder to dry with steam. However, broke is by far the most easiest to dry. All the fibers are also positively correlated with all the paper qualities, except for long fiber and stretch CD.

The difference in speed between the jet and the wire does not affect the CD/MD ratio as expected. Lets look at the difference in the top layer for tensile strength. For both the CD and the MD direction the coefficient is negative, one would expect it to be positive for the CD direction and negative for the MD direction. There is also a difference in sign between the top and bottom layer, which is strange. One would expect it to have a small difference in contribution to the paper quality but not change in sign.

In figure 9 we can also see that the diff jet - wire is not correlated with the CD / MD ratio for tensile strength, the same thing also applies for stretch and breach. This could explain the unexpected coefficients. One could ask if it is necessary to change the diff jet - wire since it does not seems to have any contribution to the paper qualities.
Figure 9: Diff jet - wire affect on the CD/MD ratio.

If we look at the first wet press we can see that the coefficient is positive for the MD and negative for the CD for the tensile strength and for the breach. The reason for this could be that when you press the paper between two large rollers the water is pushed in the MD. This could affect the fiber orientation to get in the machine direction.
3 Optimization

In this section we will present the optimization problem. In section 3.1 the problem is formulated in an optimization problem and in table 9 the constants and variables are explained. The costs in the objective function is formulated in section 3.2. And in section 3.3 we will reformulate the problem into a more smooth problem to get better local optimums. The results in section 3.5 is presented with a sensitivity analysis on the costs in the production process.

The optimization problem is based on the prediction model in section 2. The main idea of the optimization is to find ways of running the paper machine that will lower the production cost and fulfill requirements on paper quality properties. It is also of interest to find new ways that will improve the excitation of the data.

3.1 Problem formulation

This is done by optimizing the control variables, in table 6, in the prediction model based on their cost and the cost for the violation of the paper quality boundaries. We will also use the prediction model for the steam consumption to add additional indirect costs for variables that effect the ability to dry the paper. The boundaries for the variables are set to their respectively minimum and maximum in the existing historical data set.

\[
\begin{align*}
\min_{x,r,s,y} \quad & \sum_{i \in I,K,L} c_i^{(z)} x_i + c^{(r)} \sum_{i \in J} x_i r_i + c^{(s)} s + \sum_{i \in J} c_i^{(y)} (\hat{y}_i - y_i) + \\
\text{Subject to} \quad & y = a^{(y)} + B^{(s)} x + B^{(r)} r \\
& s = a^{(s)} + b^{(s)} x + b^{(r)} r \\
& x_{\min} \leq x \leq x_{\max} \\
& r_{\min} \leq r \leq r_{\max} \\
& s_{\min} \leq s \leq s_{\max} \\
& b_{\min} \leq \sum_{i \in I,K} x_i \leq b_{\max}
\end{align*}
\]

Table 8: Optimization problem with variable declaration in table 9.

The first two constraints in the optimization problem in table 8 is the result of the prediction model in section 2. Where \(a^{(y)}, B^{(s)}, B^{(r)}, a^{(s)}, b^{(s)} \text{ and } b^{(r)}\) are from the result in table 7. This is the foundation of the optimization problem.
Sets
\( I \) \{ LF, SF, Broke \}.
\( K \) \{ Chalk, Tale, Starch, Color, Glue \}.
\( L \) \{ Diff jet - wire BL, Diff jet - wire TL, Sum vacuum TV, Sum vacuum BV, sum wet press, Rotor position 25, Rotor position 26 \}.
\( J \) \{ Tensile str CD, Tensile str MD, Burst, Stretch CD, Stretch MD, Breach CD, Breach MD \}.

Variables
\( x \in \mathbb{R}^{15 \times 1} \) Independent control variables in the sets \( I, K, L \).
\( r \in \mathbb{R}^{3 \times 1} \) Independent control variables for the refining energy on the fibers in set \( I \).
\( s \in \mathbb{R}^{1} \) Depending variable for steam flow divided by the wire velocity.
\( y \in \mathbb{R}^{7 \times 1} \) Depending variables for the paper qualities in set \( J \).

Constants
\( c(x) \in \mathbb{R}^{15 \times 1} \) Costs for the variables in the sets \( I, K, L \).
\( c(r) \in \mathbb{R}^{3} \) Cost for refining energy.
\( c(s) \in \mathbb{R}^{1} \) Cost for steam flow divided by the wire velocity.
\( d(y) \in \mathbb{R}^{7 \times 1} \) Cost for violating the paper quality demand.
\( a(y) \in \mathbb{R}^{7 \times 1} \) Intercept for the paper qualities in set \( J \) for the prediction model, as in table 7.
\( a(s) \in \mathbb{R}^{1} \) Intercept for the steam flow, \( s \), as in table 7.
\( B(x) \in \mathbb{R}^{7 \times 15} \) Coefficients for \( x \) for the paper qualities \( y \) in the prediction model.
\( B(r) \in \mathbb{R}^{7 \times 3} \) Coefficients for \( r \) for the paper qualities \( y \) in the prediction model.
\( b(x) \in \mathbb{R}^{1 \times 15} \) Coefficients for \( x \) for the steam flow, \( s \), in the prediction model.
\( b(r) \in \mathbb{R}^{1 \times 3} \) Coefficients for \( r \) for the steam flow, \( s \), in the prediction model.
\( x_{\text{min}} \in \mathbb{R}^{15 \times 1} \) Lower bound for the variables in \( x \).
\( x_{\text{max}} \in \mathbb{R}^{15 \times 1} \) Upper bound for the variables in \( x \).
\( s_{\text{min}}, s_{\text{max}} \in \mathbb{R}^{1} \) Lower and upper bound for the steam flow.
\( b_{\text{min}}, b_{\text{max}} \in \mathbb{R}^{1} \) Lower and upper bound for the basis weight.
\( \hat{y} \in \mathbb{R}^{7 \times 1} \) Target values for the paper qualities.
\( \tilde{y} \in \mathbb{R}^{7 \times 1} \) Take action values for the paper qualities.

Table 9: Variable declaration for the variables and constants for the optimization problem.

3.2 Costs

The price for long fiber is slightly more than for the short fiber. The price for broke is a bit tricky since it does not have any direct costs. But we have a limited amount of broke, which means that the cost should depend on how much broke you have access to. However, since it contains both long fiber and short fiber the price is set to the mean of the long fiber and short fiber price. The basis weight is restricted by an upper and lower boundary since the cost is already implied in the cost for the fibers and chemicals.
The refining energy variable is energy per amount of fiber, which means that we have to multiply with the amount of fiber in order to get how much energy we spent on the paper. Therefore the cost for refining energy calculated by

\[ c^{(r)} \sum_{i \in I} x_i r_i \]  

(23)

where \( c^{(r)} \) is a constant based on the price for electricity.

The boundaries for the paper qualities are divided into three steps, target value, take action value and reject value. The cost for violating a paper quality will be normalized based on the two first boundaries, target value, \( \hat{y} \), and take action value, \( \check{y} \). The cost for the paper quality \( y \) will be

\[ c_j(y) = C_y \frac{\hat{y}_j - y_j}{\check{y}_j - y_j} \]  

(24)

where \( C_y \in \mathbb{R} \) is a constant and \( j \in J \). This is done because the gap between the target and the take action value differs between different paper qualities. The constant \( C_y \) can be changed depending how important it is to reach the paper quality demands.

### 3.3 Reformulating of the problem

The problem in table 8 is a nonlinear optimization problem. However, the function \( (\hat{y}_j - y_j)^+ = \max(\hat{y}_j - y_j, 0) \) makes the problem less smooth for most solving methods. Which leads to we will get stuck in local optimum more frequently than with a smoother problem. But as we will see this can be reformulated as a more smooth problem. To solve this we introduce two new set of variables, \( y^+, y^- \geq 0 \) which are equal to

\[ y^+ - y^- = \hat{y} - y. \]  

(25)

This leads to that \( y^+ \) is the violation of the target value for the paper qualities and \( y^- \) is the marginal to the violation. If we substitute \( y^+, y^- \) into the function \( (\hat{y}_j - y_j)^+ \) we get

\[ (y^+_j - y^-_j)^+ = \max(y^+_j - y^-_j, 0) = y^+_j \]  

(26)

since we penalize \( y^+_j \) the optimal solution when \( y^+_j > 0 \) will be to have \( y^-_j = 0 \) and vice versa.

If we substitute \( y^+, y^- \) into the optimization problem in table 8 we get
\[ \min_{x,r,s,y^-} \sum_{i \in I} c_i^{(x)} x_i + c^{(r)} \sum_{i \in I} x_i r_i + c^{(s)} s + \sum_{i \in J} c_i^{(y)} y_i^+ \]

Subject to
\[ y^- - y^+ = a(y) - \hat{y} + B^{(x)} x + B^{(r)} r \]
\[ s = a^{(s)} + \hat{b}(x) x + \hat{b}(r) r \]
\[ x_{\text{min}} \leq x \leq x_{\text{max}} \]
\[ r_{\text{min}} \leq r \leq r_{\text{max}} \]
\[ s_{\text{min}} \leq s \leq s_{\text{max}} \]
\[ y_{\text{min}} \leq \sum_{i \in I} x_i \leq y_{\text{max}} \]

Table 10: Reformulated optimization problem.

### 3.4 Convexity

This is not a convex problem because the objective function contains products in the second term. To see this we create a new vector, \( z \), that only contains the variables in the second term, i.e.

\[ z = [x_{\text{LF}}, x_{\text{SF}}, x_{\text{Broke}}, r_{\text{LF}}, r_{\text{SF}}, r_{\text{Broke}}]. \]  

This gives us the sub objective function,

\[ z \cdot H \cdot z^T = c^{(r)} \sum_{i \in I} x_i r_i \]

where the Hessian, \( H \), is

\[
\begin{bmatrix}
0 & 0 & 0 & c^{(r)} & 0 & 0 \\
0 & 0 & 0 & 0 & c^{(r)} & 0 \\
c^{(r)} & 0 & 0 & 0 & 0 & c^{(r)} \\
0 & c^{(r)} & 0 & 0 & 0 & 0 \\
0 & 0 & c^{(r)} & 0 & 0 & 0 \\
0 & 0 & 0 & c^{(r)} & 0 & 0 \\
\end{bmatrix}
\]

which has negative eigenvalues. This results in a non convex optimization problem and we will only find a local optimum. To get a better local optimum the GlobalSearch function in Matlab is used.

### 3.5 Results

The normalization of the variables in the plots below, figure 10 - 18, means that they are normalized with their min and max values,

\[ \tilde{x}_i = \frac{x_i - x_{i_{\text{min}}}}{x_{i_{\text{max}}} - x_{i_{\text{min}}}} \quad \text{and} \quad \tilde{r}_i = \frac{r_i - r_{i_{\text{min}}}}{r_{i_{\text{max}}} - r_{i_{\text{min}}}} \]

where \( \tilde{x} \) and \( \tilde{r} \) are the normalized variables. The normalization is done in order to get a good overview of the variables. For the paper qualities the normalization
is done based on the target value and the take action value as,

\[ \hat{y}_j = \frac{y_j}{y_j - \hat{y}_j}, \]  

(31)

where \( y = y^+ - y^- \) and \( j \in J \). For the paper quality violation, there is only a punishment for the \( y^+ \) variable. Which means that values below zero will not be punish or gained as mentioned earlier.
Figure 10: Sensitivity analysis for the cost for long fiber.
Figure 11: Sensitivity analysis for the cost for short fiber.
Figure 12: Sensitivity analysis for the cost for broke.
Figure 13: Sensitivity analysis for the cost for starch.
Figure 14: Sensitivity analysis for the cost for refining.
Figure 15: Sensitivity analysis for the cost for steam.
Figure 16: Sensitivity analysis for the cost for color.
Figure 17: Sensitivity analysis for change in target value for the basis weight.
Figure 18: Sensitivity analysis for the cost for paper violation.
Figure 19: The total cost for the sensitivity analysis in figure 10 - 18 above.
3.6 Discussion

The top left and middle left plots in figure 10 - 18 shows the fiber distribution and the refining energy. We can see that they are negatively correlated with each other. The reason for this is that the cost for refining energy is depending on the amount of fibers, i.e. larger part of long fiber will increase the cost for refining long fiber. The reason could also be that with increased refining less fiber is needed to obtain the same quality.

One problem with choosing the refining variable as refining energy per fiber flow is that the effect of the refining depends on how much fiber is used. This becomes a problem in the optimization. For example, lets keep a constant refining energy per fiber for long fiber and increase the amount of long fiber. This will increase the cost for the refining of long fiber, which the optimization will take in to account. But the affect on the paper qualities will only depend on the change in the part of long fiber with an average refining. This is a problem since long fiber with high refining will have different properties than long fiber with low refining.

But if we do not divide the refining energy by the fiber flow the effect of refining per fiber is not taken in to account. For instance, lets keep a constant refining energy for long fiber and increase the amount of long fiber. This will change the refining energy per fiber, but this is not taken in to account. But the cost for refining is taken in to account by the optimization.

As we can see with both this ways of implementing the refining energy has a problem with the prediction part, but is able to handle to cost. However, if the refining does not change the properties of the fiber considerably the current model in this report will be a fine approximation.

In figure 10 we can see that the breakpoint when it is better to use a bigger part long fiber than refining them more is when the price for long fiber is below 115. The substitute for long fiber is short fiber. When the short fiber has reach its maximum the amount of broke is increased the basis weight is also increased in order to obtain a good paper quality.

The price for short fiber is changed in figure 11. The substitute for short fiber is broke and increased basis weight. One would expect it to be long fiber since short fiber is the substitute for long fiber in figure 10. There are also a bunch of other changes in variables as the cost changes. The marginal for the paper qualities to the target value will also decrease as the short fiber cost increases.

Broke is a tricky fiber since the contain will vary depending on the paper that is recycled. The refining per fiber flow for broke varies and the broke is refined to refresh and separate the fibers. Which means that the refining of broke could be a measurement on how good the broke is. The price is kind of abstract since it does not have any direct cost. But it can not be free since the access is limited. The prices will also vary over time depending on the access.

In figure 13 we can see that the price for starch must multiply several times...
in order to decrease the amount of starch. The reason for this could be that the original price is too low. The reason could also be that starch is positively correlated with the tensile strength in CD, which is the hardest to fulfill. There is also no substitute for starch, instead the tensile strength in CD is violated when the basis weight is maxed out.

As the cost for the steam is increased, in figure 15, the amount of short fiber is substituted with broke. As mentioned earlier broke is more easy to dry than short fiber. To compensate for the substitution the basis weight is increased together with the amount of broke. This also leads to that the marginal for the paper qualities decrease. It is also interesting that the steam flow decreases as the basis weight increases. The reason for this is that the short fiber is replaced by long fiber and broke, which are easier to dry.

Color is a surprisingly good chemical as we can see in figure 16. In order to decrease the amount one must increase the price to unreasonable large values. However, that the color will have these properties is hard to believe. The reason that it has this good properties could be that the kind of paper that has a lot of colors also has higher demand on specific paper qualities. These can be fixed with more excitation in the data for the color when other variables are fixed.
4 Conclusions and future work

We have synchronized the signals in time with a model of the tank dynamic. Then a prediction model was calculated based on signals from a variable selection. The prediction model was used as a foundation to the optimization problem.

The result of the optimization problem is highly dependent on how well the prediction model performs. Most time has therefore been spent on the prediction model. There is also a lot of freedom in constructing the prediction model which will affect the structure of the optimization problem. For example, the steam consumption is predicted as the steam flow per wire velocity, which makes it easier to compare with the basis weight.

4.1 Optimize the change in fiber distribution

Another way to minimize the production cost is in the change in the distribution of fibers. Today when a change of the fiber distribution there are no actions to optimize the change. Bigger changes takes about 15-20 minutes from the change in the refining section has gone through to the mixing section and reach the paper machine. This is known by looking at the change in the steam demand, when it has stabilized the change is done.

This could be optimized by lowering the tank levels when a change is made, which will speed up the change as we can see in the tank model equation for the amount of fiber $i$ in tank $T_{l,k},$

$$C_{l,k}^{(i)}(t + 1) = C_{l,k}^{(i)}(t) + \Delta T \cdot \frac{F_{k}(t)}{V_{l,k}} \left( C_{l-1,k}^{(i)}(t) - C_{l,k}^{(i)}(t) \right)$$

But by lowering the tank level the probability that something can go wrong will increase. There is also a trade off in how high flow rate that is optimal for the change. A higher flow rate will speed up the time for a change but it will be harder to dry.

4.2 Improve the prediction model

The prediction model can be improved further by making experiments in the paper manufacturing process and by adding more signals to the data set.

The paper qualities is now predicted by the basis weight as a foundation. The basis weight can be substituted with the flow divided by the wire velocity. This will however give a worse prediction. But for the optimization problem it would be more interesting to include the wire velocity in this way if the prediction can be improved, by adding more variables.

As mention earlier the affect of the refining energy is hard to model. There is also a problem that the refining does not seems to have the good properties
that the it should in theory. This could depend on that the excitation in the data is not so good due to large margins on the paper qualities properties.

A on-line measurement of the pulp quality can be installed in order to get more information of the fibers going into the paper machine. Test can be made by only change the refining and the fiber amount respectively in order to see the contribution to the paper qualities.

With this new data a new covariate can be added to the prediction model that will replace the fiber amount and the refining by combining the two in a formula depending on the refining energy, amount of fiber and the paper quality. For example, tensile strength should be improved by more refining and stretch should be worsen by more refining. This means by increase the refining the amount of fiber will have a positive boost by more refining for tensile strength and a negative boost for the stretch. But more amount of fiber should increase both tensile strength and stretch.

The felt on the wire is changed from time to time. This is done since the felt ability to drain water gets worse over time. This will affect the drying process in the paper machine and therefore also the prediction model. Today there is no signal for the felt change, but it will be good to include such a signal.

In the data set the signal for formation is out of sync. To get this signal on-line will be good for the prediction since the formation play a big part of for the paper qualities. It will also be interesting to predict the risk for bad formation. For example, long fiber is harder to form than short fiber. This means that with a bigger part of long fiber the prediction will vary more that it would if we use more short fiber. This also applies for the quality through out the tambour.

4.3 Improve the optimization model

There is a lot of ways of improving the optimization problem. The wire velocity can be added in order to take the production and the profit in to account. The wire velocity depends on the paper machines ability to dry the paper. Increased wire velocity will also increase the steam flow. And since short fiber is harder to dry than long fiber the production will also be slower.

Broke is almost free, the direct cost is only cost for washing and bleaching. But since we only have a limited amount of it, we have to take that in to account. This can be done by optimize the production for several tambours or relate the price to how much we have access to. One advantage of looking at the production over several tambours is that if the tambours have different paper quality demands. Since if one type of paper quality demand is easier to achieve with more broke and another one is not.

We can also introduce an uncertainty on the paper qualities. This uncertainty can depend on how good the prediction model performs or/and by also predicting the risk for bad formation. The uncertainty of the prediction model will
punish us to have a small margin to the paper quality demand. The risk for bad formation will favor us to chose variables that have a lower risk for bad formation.
References


[8] MATLAB - medfilt1