On the modeling of tensile index from larger data sets

Paper technology

Anders Karlström*, Lars Johansson and Jan Hill

On the modeling of tensile index from larger data sets

https://doi.org/10.1515/npprj-2018-0019
Received June 14, 2018; accepted November 8, 2018; previously published online June 29, 2019

Abstract: The objective of this study is to analyze and foresee potential outliers in pulp and handsheet properties for larger data sets. The method is divided into two parts comprising a generalized Extreme Studentized Deviate (ESD) procedure for laboratory data followed by an analysis of the findings using a multivariable model based on internal variables (i. e. process variables like consistency and fiber residence time inside the refiner) as predictors. The process data used in this has been obtained from CD-82 refiners and from a laboratory test program perspective, the test series were extensive. In the procedure more than 290 samples were analyzed to get a stable outlier detection. Note, this set was obtained from pulp at one specific operating condition. When comparing such "secured data sets" with process data it is shown that an extended procedure must be performed to get data sets which cover different operating points. Here 100 pulp samples at different process conditions were analyzed. It is shown that only about 60 percent of all tensile index measurements were accepted in the procedure which indicates the need to oversample when performing extensive trials to get reliable pulp and handsheet properties in TMP and CTMP processes.

Keywords: CTMP; energy efficiency; fiber residence time; modeling; pulp and handsheet properties; pulp consistency; temperature profile; tensile index; TMP.

Introduction

Forgacs (1963) reflected on the necessity of linking the variations in the mechanical pulping process variables to the composition of particle shapes and sizes in the pulp. He stated that *"Ideally, the measurements made on the pulp should be such that they can be interpreted in terms of the mechanical pulping operation, and at the same time be used to predict the paper or board making potential of the pulp."* However, the laboratory test procedures have been discussed for decades and due to tedious and complex procedures the analyses of pulp and handsheet properties sometime tend to be based on too few samples. This makes it difficult to verify data sets statistically even though many robust techniques such as modified Z-score, adjusted Boxplot, sample kurtosis and the Shapiro-Wilk W test (Barnett and Lewis 1994) can be natural tools when improving measurement quality. To tackle such a problem, modified detection algorithms based on a generalized Extreme Studentized Deviate (ESD) procedure can be used (Rosner 1983). This method was primarily used for environmental pollution monitoring to avoid the problem of masking (Gilbert 1987).

To maximize insight into a data set and detect discordant outliers and anomalies in laboratory samples it is believed that the ESD approach can be an important add-on to the normal test procedures when analyzing pulp and handsheet properties. To show how to use the methodology, we focus on one property, tensile index, and analyze seven handsheets including three sampling strips each for fourteen different operating points which yields a set of about three hundred samples to analyze.

Even though reliable laboratory data sets are available it is still a challenge, from a process dynamics perspective, to link such data to process information from computers, on-line pulp sampling devices etc.

Traditionally, external variables (such as specific energy, dilution water added to the refiners, disc clearance measurements etc.) have been used for process follow-up and estimation of pulp and handsheet properties (Härkönen et al. 2000, Strand 1996, Sabourin et al. 2001, Härkönen et al. 2003, Strand and Grace 2014, Nelsson 2016). One challenge when using external variables as predictors is that the process non-linearities are not handled in an appropriate way. To cope with that soft sensors, describing physical phenomena in the refining zone, have been developed during the last decade (Karlström and Eriksson 2014a, 2014b, 2014c, 2014d). The soft sensors can be seen

***Corresponding author: Anders Karlström,** Electrical Engineering, Chalmers University of Technology, Göteborg, Sweden, e-mail: anderska@chalmers.se
Lars Johansson, Rise PFI, Trondheim, Norway, e-mail: lars.johansson@rise-pfi.no
Jan Hill, QualTech, Tyringe, Sweden, e-mail: jan.hill@qtab.se

as internal variables (such as fiber residence time, consistency profile, forces on bars, distributed defibration, thermodynamic work etc.) which are difficult to measure directly in the process. Typically, such soft sensors are nonlinear and have become important for advanced process optimization. Specifically, consistency and fiber residence time have been candidates for such activities for some years, as they provide a link to e. g. tensile index, mean fiber length and Somerville shives (Karlström et al. 2015, 2016a, 2016b).

The next challenge is to find soft sensors for other pulp and handsheet properties as well. This has been a key issue for decades but, due to difficulties in assuring the laboratory measurements' relation to process conditions, the efforts continue. Better process models are assumed to be important when handling that problem, although this also causes a data deluge in mill-wide systems. This means that it is essential to handle both laboratory data (obtained from pulp samples provided at non-equidistant sampling intervals) on the same time frame as the process variables from the distributed controllers (which are normally equidistantly oversampled). These challenges are relatively easy to handle using modern technology. Another and perhaps more challenging task is to understand which laboratory data should be combined with the process data and which data we can assume to be outliers in this context.

This contribution focuses on a methodology to find and validate laboratory test results from eighteen pulp and handsheet properties. The main idea in applying the methodology on many properties is to understand the weaknesses and limitations in the measurement procedures and how many samples we need to get "statistically assured" laboratory test results and process data.

Materials and methods

In this paper, two consecutive steps are introduced 1) to handle detect outliers in laboratory samples in order to provide reliable data to 2) to select pulp and handsheet property candidates for process modeling purposes.

Detection of outliers in laboratory samples

To show the outlier detection principals, samples for analyzing tensile index are used in this section. Assume that the dynamic variations in the process, during each pulp sampling, can be considered as small and that the obtained average of each pulp sample (fourteen samples in our case) is representative for the process conditions during each sampling interval. By preparing handsheets from each pulp sample this means that possible outliers are related to the handsheets and not necessarily to the variations in the process. Suppose that seven handsheets are prepared. From each handsheet, three strips are provided for analysis according to Figure 1.

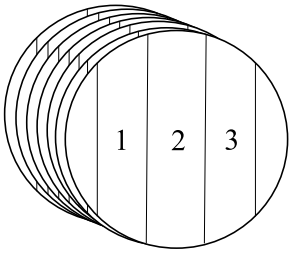

Figure 1: A schematic drawing of three strips obtained from each of the seven handsheets.

This means that mainly three different approaches can be formulated when analysing the tensile index

$$\left.\begin{array}{ll} \text{Case A:} & \tau_{ij} = \theta_{ij}/\bar{\mu} \\ \text{Case B:} & \tau_{ij} = \theta_{ij}/(\frac{1}{l}\sum_{k=1}^{l}\mu_k) \\ \text{Case C:} & \tau_{ij} = \theta_{ij}/\mu_{kj} \end{array}\right\} \quad (1)$$

where θ is the tensile strength while the denominator in Case A is the average basis weight for handsheets, i. e. one measure for the complete batch of handsheets. For Case B the denominator can be seen as the most logical average basis weight to use for each handsheet. Case C covers each tensile index sampled from each handsheet.

Thus, for the samples $j = 1, \ldots, m$ we have to consider $i = 1, \ldots, 3l$ tensile strength (θ) measurements.[1] Introduce $n = 3l$, i. e. in our case $nm = 294$ elements to analyze in the vector. In our example $\boldsymbol{x} = [x_1, \ldots, x_{nm},]$ and when the generalized ESD is applied on the data series the multiple discordant outliers can be recursively identified if the dynamic process variations between the pulp sampling intervals are handled carefully. This statement will be further penetrated in the next section but we need to describe the basic outlier detection procedure first.

Sort the nm observed values in the vector x from the mean $\bar{x}$ in ascending order and calculate statistics for up to $nm - 1$ outliers, i. e. $i = 1 : nm - 1$.

$$R_i = \max_i \left\{ \frac{|x_i - \bar{x}|}{s} \right\} \quad (2)$$

1 For clarifications the example comprises $m = 14$ pulp samples, $l = 7$ handsheets and three strips.

where the denominator s represents the standard deviation. The recursive process is continued until $R_1, R_2, \ldots, R_{nm}$ have been computed.

Compare each R_i with the critical value λ_i for a prespecified significance level α, defined as

$$\lambda_i = \frac{(nm-i)t_{nm-i-1,p}}{\sqrt{(nm-i-1+t^2_{nm-i-1,p})(nm-i+1)}} \quad (3)$$

where $t_{nm-i-1,p}$ is the inverse of Student's t-cumulative distribution function with $nm-i-1$ degrees of freedom and the percentile values of the t-distribution

$$p = 1 - \frac{\alpha}{2(nm-i+1)} \quad (4)$$

where α is the significance level, see further EPA (2006). By this definition, the critical value λ_i represents decision cutpoint to label whether an observation is a potential outlier, see Barnett and Lewis (1994).

The null hypothesis (i. e. no outliers) can be rejected if $|R_i > \lambda_i|$ which results in i extreme values classified as outliers, EPA (2006). This process is continued until $i = nm-1$ and we can conclude that there are a certain number of outliers, or until all the tests have been performed and none were found to be significant. In other words, if none of the tests are significant, i. e. $R_i < \lambda_i$, then there are no outliers in x and the null hypothesis holds. Note, this procedure is not linked to possible outliers in a dynamic perspective, i. e. laboratory samples related to process variations. Therefore, it is necessary to develop it further to be useful in a broader perspective which will be discussed later in this paper.

Selection procedure for reliable pulp and handsheet property candidates for process modeling

As the process data are oversampled (every second) with respect to pulp sampling (equidistantly for 3 minutes), it is wise to maintain the high-frequency information to get a possibility to illustrate the process noise and its impact on the pulp property variations. However, it is unknown whether the measured laboratory samples of the pulp and handsheet properties are reliable in a dynamic perspective as rapid process changes or natural fluctuations are not available. Therefore, it is necessary to strengthen the hypothesis that the dependent variables that are selected can be predicted.

Consider that there are m different test series to study and that each test series comprises i different pulp samples. As the process data are recorded every second during the time interval for each batch of the pulp sample, both vector $\mathbf{x}_{m_i}$ including internal and/or external variables and the sampling rate are known and thereby also the number of samples, N. By estimating the mean values of the internal and external variables for each time interval, a common timeframe for the analysis of the pulp and handsheet property f_{m_i} is obtained.

$$\bar{x}_{m_i} = \frac{1}{N}\sum_{j=1}^{N} x_{jm_i}; \quad f_{m_i} = \bar{f}_{m_i} \quad (5)$$

As seen in Equation 5, the pulp and handsheet property is only defined as an averaged measure during the sampling interval and most likely is non-linearly dependent on the process conditions. This means that it is natural to model the selected pulp and handsheet property as a collection of piece-wise linear functions of the form

$$\hat{f}_m(\bar{\mathbf{x}}_m) = \theta_{m1}\bar{x}_{m1} + \cdots + \theta_{mk}\bar{x}_{mk} + b_m \quad \forall m = 1, 2, \ldots, q \quad (6)$$

where $\{\theta_{m1}, \ldots, \theta_{mk}\}$ represents the parameter vector and k the number of predictors (Lowe and Zohdy (2010)).

The number of linear regions into which the nonlinear function is broken up is represented by q. By using the parameters in Equation 6, the pulp property is assumed to be predictable for sampling rates related to the internal and external variables i. e. $\hat{f}_m(\bar{\mathbf{x}}_m) \rightarrow \hat{f}_m(t)$ and this extension of course requires minimized process fluctuations during the sampling interval.

In the work reported here, the test was performed over a limited period (three days), see Figure 18, and it is assumed that the test series can be modelled by one piecewise linear function, i. e. $q = 1$.

A refinement using the adjusted R^2 will be included to set a penalty for the number of predictors in the model, i. e.

$$\text{adj. } R^2 = 1 - \left(\sum(f-\hat{f})^2 / \sum(f-\bar{f})^2\right)(n-1)/(n-d-1) \quad (7)$$

where $\sum(f-\hat{f})^2$ represents the sum of the squared residuals from the regression[2] and $\sum(f-\bar{f})^2$ the sum of the squared differences from the mean of the dependent variable, while n is the number of observations and d is the degree of the polynomial, see further Draper and Smith (1998).

Compare two test series (1) and (2) of the same dependent variable using the Kolmogorov-Smirnov test (Stephens (1974)) to determine if the test series differ sig-

2 Here the notation refers to $\hat{f} = \hat{f}_1(\bar{\mathbf{x}}_1)$.

nificantly[3] as each element i in the test series is related to the same pulp sample in the blow-line.

Real non-zero values are considered, and the comparison is made to find all pulp or handsheet properties $f_{i(1)}$ and $f_{i(2)}$ that fulfill the relation

$$|f_{i(1)} - f_{i(2)}| < c \quad \text{where } i = \{1, \ldots, \Psi\} \tag{8}$$

Constraint c is related to the accepted variation in the laboratory equipment and the number of samples required for further analysis.

The procedure results in a reduced vector f_j where $j = \{1, \ldots, 2\psi\}$ and ψ is the length of $f_{j(1)}$ and $f_{j(2)}$. All other pulp or handsheet properties are saved separately for later use. Select different combinations of predictors according to Equation 6 and perform a polynomial fit of the models m using the 2ψ accepted samples.

To provide the initial models, a low constraint is introduced on the adj. $R^2 \geq 0.3$. This constraint is normally not acceptable in modeling procedures but in this step we want to find enough data for further analysis.

Note, when selecting the permitted residuals between two sets the elements to be compared can be out of range and still be selected as candidates for analysis if they are both outliers. To reduce that risk, each sample can be tested iteratively by estimating the adj. R^2 for the remaining $2\psi - k$ samples. If the model is improved, the rejected k samples are left for further analysis.

The best model fulfilling the constraint will be used by estimating the dependent variables $\hat{f}_{j_m}$. A vector of the differences is created between measured and estimated variables of all $2\psi - k$ samples. Find the smallest and largest elements multiplied by r, i. e. the coefficient of multiple correlation, and use these scalars as constraints.[4] Estimate the dependent variables $\hat{f}_{l_m}$ based on the data rejected and define the differences $\xi_l = f_l - \hat{f}_{l_m}$. If the elements of ξ_l are within the constraints, i. e.

$$\min(f_j - \hat{f}_{j_m}) \leq \xi_l / r \leq \max(f_j - \hat{f}_{j_m}) \tag{9}$$

and $l = \{1, \ldots, 2(\Psi - \psi) + k\}$, the corresponding measures f_l are accepted for further analysis.

If the assumptions introduced above are incorrect for the given data set, the methods will likely give erroneous results. This means that more laboratory measurements should be performed and analyzed before changing the data in the test series studied.

However, if it is not accepted to re-run the tedious laboratory testing it is still possible to improve the models by ranking the absolute differences between the measured and estimated properties in ascending order. Thereby, new models of the pulp and handsheet properties can be derived and validated. In this step the samples rejected are also tested to find other acceptable measures.[5]

The selection of predictors has been discussed in several articles and the reader is referred to Karlström et al. (2015, 2016a, 2016b) for more details. In this paper, only internal variables like consistency and fiber residence time will be considered as they outperform the external variables as independent variables (predictors) when making polynomial fits of pulp and handsheet properties, see Karlström and Hill (2017a, 2017b, 2017c).

Results and discussion

When analyzing pulp and handsheet properties from TMP and CTMP processes it is well known that the spread in accuracy can deviate considerably. In this example, the TMP process accuracy in pulp and handsheet measurements is slightly better compared with samples obtained from CTMP processes for board making. This is a consequence of different energy inputs used in the processes, i. e. how the fiber development is performed.

It is also interesting to compare e. g. tensile index versus specific energy from CTMP and TMP, as illustrated in Figure 2. In the TMP samples the accuracy in measurements was considerably better compared with Test A and B (CTMP), which varied quite much, which will be discussed in more details below. Note that in both cases, CD-82 refiners were used, see further Karlström et al. (2016a, 2016b) and Karlström and Hill (2017a, 2017b, 2017c).

To detect outliers in laboratory samples we start with data for tensile index measurements, obtained from a TMP process. As indicated above discordant outliers can be detected by two consecutive sequences where the first iteration is a rudimentary check by visualizing the measured pulp properties from each handsheet. Measures far from the mean value are automatically rejected, see Figure 3.

3 The Kolmogorov-Smirnov test has the advantage of making no assumption about the distribution of data and this can be seen as a first check of the measurement accuracy. These test series comprise all measurements obtained from Test A and Test B.

4 The introduction of r can be seen as an extra penalty between zero and one. A higher value close to one indicates a better predictability. A poor predictability tends to introduce a large interval for the elements of ξ_l in Equation 9 if r is not introduced. The coefficient of multiple correlation (Draper and Smith (1998)) is the square root of the coefficient of determination R^2 in the linear regression model described by Equation 6 which includes an intercept.

5 This cross-check can be skipped if the lower and upper constraints are defined symmetrically.

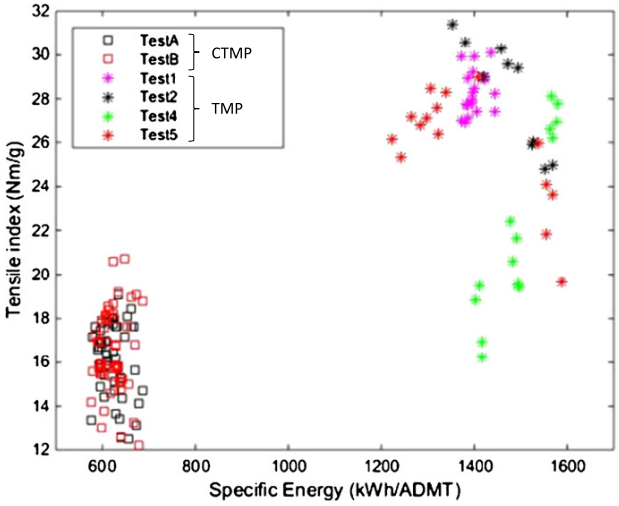

Figure 2: Tensile index versus specific energy for typical CTMP and TMP operations.

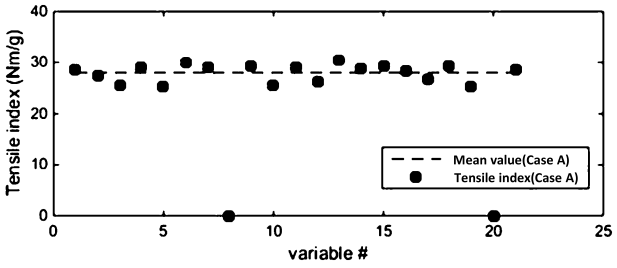

Figure 3: Tensile index estimated for Case A. Outliers detected by visual inspection.

Detection of outliers in laboratory samples

In our example $3nm = 291$ elements in the sample vector are acceptable for further analysis.[6]

To find out if the selected variables are acceptable for further analysis a second iteration based on the modified generalized ESD procedure, Equation 1–Equation 4 can be applied.

The assumption that the pulp property measurements, excluding the suspected outliers, are approximately normal distributed is appropriate when the vector size is greater than or equal to 25 Rosner (1983). In the pulp and paper industry, this required vector size can be a problem due to tedious laboratory tests.[7] Therefore, other complementary methods where the sampling vectors are extended must be added to get a reliable set of data. In our example the condition is thereby fulfilled if we can handle the process dynamics when extending the vector size by adding new measurements from other sampling intervals. However, most often such procedures result in situations where the mean values obtained from each sampling interval can differ considerably see Figure 4. If these deviations are caused by uncertain test procedures it is important to handle the data set with care.

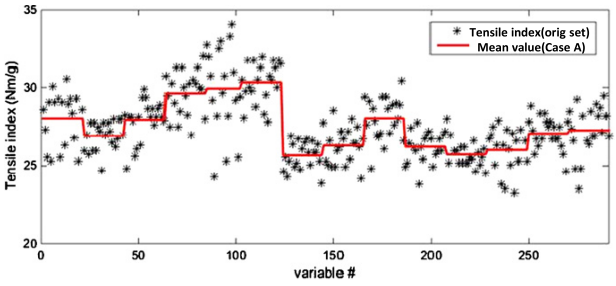

Figure 4: Tensile index for Case A for the entire data series when all outliers caused by measurement errors have been rejected (1st iteration).

Obviously as seen in Figure 4, the standard deviation differs quite much for each sample as well as for Case A–Case C in Table 1. Which case to choose as a standard when analyzing tensile index is hard to pre-specify as the means for all samples are almost equal.

We can conclude that the number of samples in each sample is ≤21 which means that the generalized ESD most likely does not approach a normal distribution. Moreover, as seen in Equation 2 and Equation 3 the outlier criterion $R_j \geq \lambda_i$ differs from the discretized $R_j \geq \lambda_j$. To overcome such problems we introduce a procedure where the mean of each sample is extracted from each measurement. This is shown in Figure 5 for Case A and Case C. The discrepancies between the two cases are small and we can expect that Case B, which is in between the two cases in Table 1, has similar characteristics.

In Figure 5, it is also shown that several potential outliers can exist depending on the chosen significance level α.

Following the procedure outlined in Appendix A, outliers can be detected in samples of pulp and handsheet properties. When the deviations in the mean values are caused by variable process conditions it is even more relevant to introduce the procedure above. A good checkpoint is to see if the distributions tend to be skew. When the skewness is far from zero this indicates that some of the tensile index measurements are in the lower region of

6 In this example the vector size of n is reduced to n'.

7 Moreover, the normal procedure is that only one average value of each sample is provided when comparing laboratory data with process data. This means that important information about pulp property variations can be lost.

acceptable measures. Nevertheless, the key questions are whether or not it is acceptable with a deviation in tensile index of ±3 Nm/g and if it is acceptable to use measurements with such spread for modeling purposes, i. e. link the results to different types of process dynamic evaluations.

Table 1: Standard deviation in tensile index (Nm/g) for Case A, Case B and Case C for all samples studied.[8]

Sample	Case A	Case B	Case C
1	2.90	2.88	2.46
2	0.82	0.81	0.69
3	1.59	1.57	1.44
4	2.00	2.00	2.31
5	8.25	8.14	7.39
6	1.67	1.65	1.58
7	0.76	0.76	0.77
8	1.37	1.36	1.39
9	1.83	1.81	2.07
10	0.96	0.95	0.87
11	0.68	0.68	0.68
12	1.91	1.89	2.71
13	1.03	1.03	1.29
14	2.51	2.47	2.52
Mean	2.02	2.00	2.01

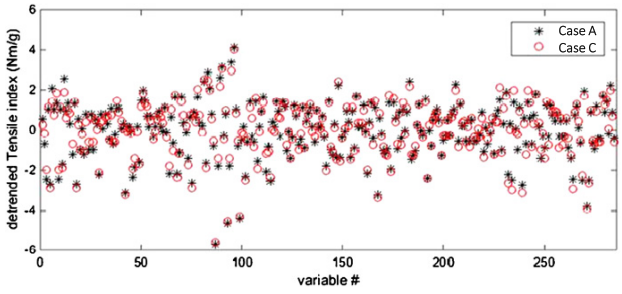

Figure 5: Detrended tensile index for Case A and Case C after the 1st iteration. Each sample is detrended individually.

Selection procedure for reliable pulp and handsheet property candidates for process modeling

Even though, an appropriate generalized ESD-procedure is used to detect outliers it is not guaranteed that the laboratory data will be useful in a dynamic perspective. This is best illustrated by studying the time plot for specific energy where the time for pulp samples are included. As seen in Figure 6 the specific energy and most likely also the pulp and handsheet properties can vary considerably during the sampling. This is of course a challenge when it comes to validation of pulp and handsheet properties in a dynamic perspective.

Harrell et al. (1985) and Freedman and Pee (1989), who presented a general guideline for the minimum number of events per variable (EPV) in multivariate analysis, demonstrated that overfitting was inflated when the ratio of the number of variables to the number of observations was greater than 1/4, which corresponds to an EPV ≥ 4. Peduzzi et al. (1996) suggested increasing that number to at least ten events per variable analyzed to maintain the validity of the final model. This analysis was based on data from a cardiac trial with good quality data, and, in our situation, this recommendation would result in at least 50 samples if five predictors are used.

Draper and Smith (1998) suggested the use of an EPV of 10 as a good choice but in industrial (especially in pulp and paper industry) applications, this is usually not possible due to tedious laboratory analysis and uncertainties in the measurements which limits the number of reliable samples.

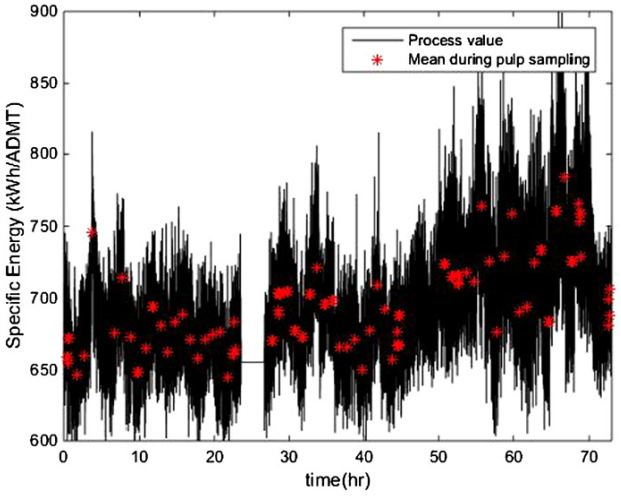

Figure 6: Specific energy and pulp sampling occasions versus time.

Vittinghoff and McCulloch (2007) conducted a large simulation study of other influences on confidence interval coverage relative bias and other model performance measures and found a range of circumstances in which coverage and bias were within acceptable levels despite an EPV less than 10. In short, they concluded that the "one in

8 For details about the cases studied, see the introduction to the section "Materials and methods".

ten" rule can be relaxed and, in this paper, it is assumed that reliable models can be derived using an EPV ≈ 4 if it is possible to confirm that some of the measurements obtained during the major step changes in production, plate gap and dilution water feed rate in Figure 18 (Appendix B) are covered.

In summary, the methodology for large data sets is based on three different steps using double tests of each sample and it is always appropriate to be critical of the results if too few laboratory samples are available relative to the number of predictors analyzed in the model. This is central in this section and the idea is to extract the laboratory measurements which are possible to link to process data. Thereafter the data selected is ranked before training and verifying the models according to the procedure outlined by Karlström and Hill (2017a, 2017b, 2017c).

The idea so far, has been to extract the laboratory measurements which are possible to link to process data. Originally, 160 pulp samples[9] were included in the proposed test series, although only about 100 tensile index measurements were analyzed.

In this paper, the measurements will be ranked by using the absolute difference between the measured property and the estimated property. This is illustrated in Table 2, which gives the accepted measurements of tensile index. As seen in Table 2, about 60 % of the pulp samples are accepted if selecting a constraint of ±2 Nm/g according to EPV = 4. The original data selection is shown in the right column, i. e. the modeling and hold-out sets. The original data selection is shown in the right column, i. e. the modeling and hold-out sets together with some of the rejected samples obtained as a result of the asymmetry in the upper and lower constraints in Equation 9. If such rearranged measurements are used, it is possible to analyze a new set of "accepted" data at different adj. R^2 according to the procedure outlined above, see Table 3.

As stated by Karlström and Hill (2017a, 2017b, 2017c), only internal variables like consistency and fiber residence time are necessary to consider as independent variables when making polynomial fits of pulp and handsheet properties. Thereby, Equation 6 can be expressed as

$$\tau = 40.3 - 2.192C_{FZ} + 1.025C_{CD} - 16.51\eta_{FZ} + 200.98\eta_{CD} \quad (10)$$

where τ corresponds to the tensile index estimation in a CD-82 refiner. The independent variables C and η represents the consistencies and fiber residence times in the flat zone and conical zones $\{FZ, CD\}$ respectively. For details, see Karlström and Hill (2017a, 2017b, 2017c).

9 Eighty pulp samples in Test A and Test B, respectively.

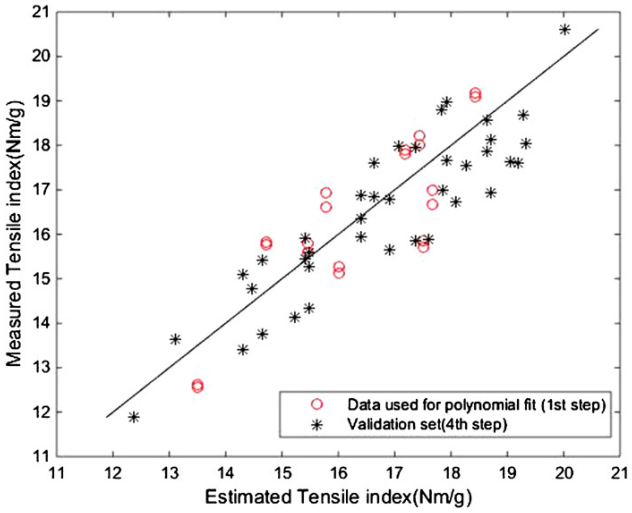

Figure 7: Measured tensile index versus estimates tensile index.

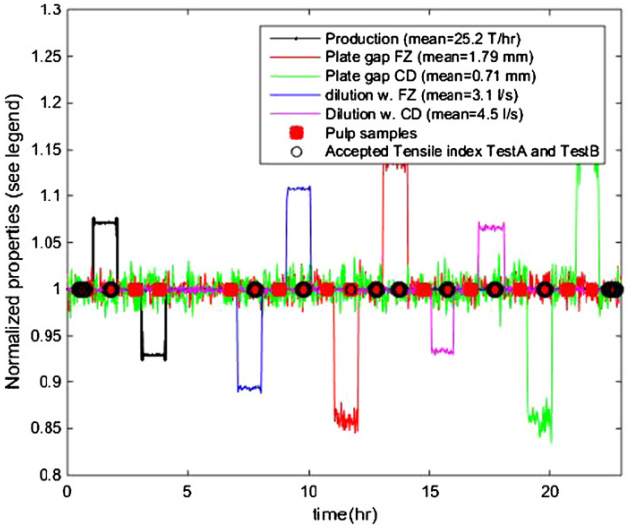

Figure 8: Normalized properties to show where samples for tensile index are taken.

In this paper, we choose to focus on three figures summarizing findings for estimated and measured values for tensile indices.

In Figure 7, it is obvious that the model can estimate the tensile index within ±2 Nm/g. It is also important to confirm that the major dynamics in the predictors are covered by the step changes, see Figure 8. Finally, to get an understanding of the process fluctuations in the estimation of the tensile index it is wise to also include a time plot, see Figure 9.

Thus, to use the tensile index model in on-line applications it is important to verify it over long periods and different process conditions. The reason is of course that

Table 2: Tensile index ranking based on the absolute difference between measured and estimated values.

Tensile index (ranking) abs (measured-estimated)	TI (Nm/g)	Sample	# at LAB	Corresponding test series
0.02	15.45	27	27	Validation set Test B
0.07	16.34	51	54	Validation set Test A
0.07	18.57	79	84	Validation set Test B
0.09	15.58	5	5	Validation set Test B
0.12	16.79	75	80	Validation set Test B
0.13	15.60	24	24	Set for polynomial fit Test B
0.21	16.84	45	48	Validation set Test B
0.21	15.26	72	77	Validation set Test B
0.28	17.65	68	71	Validation set Test A
0.29	14.76	54	601	Validation set Test B
0.31	15.78	24	24	Set for polynomial fit Test A
045	15.94	25	25	Validation set Test B
048	16.87	25	25	Validation set Test A
049	11.88	66	69	Validation set Test A
0.51	15.91	26	26	Validation set Test B
0.54	13.64	65	68	Validation set Test A
0.57	18.13	14	14	Validation set Test B
0.57	17.95	21	21	Validation set Test B
0.58	18.03	39	40	Set for polynomial fit Test A
0.60	20.62	13	13	Validation set Test B
0.62	18.67	78	83	Validation set Test B
0.62	17.81	15	15	Set for polynomial fit Test A
0.64	19.09	9	9	Set for polynomial fit Test A
0.68	16.99	1	1	Set for polynomial fit Test B
071	17.89	15	15	Set for polynomial fit Test B
0.71	17.55	3	3	Validation set Test B
075	19.19	9	9	Set for polynomial fit Test B
075	15.27	58	61	Set for polynomial fit Test A
076	18.20	39	40	Set for polynomial fit Test B
076	15.41	62	65	Validation set Test A
077	17.86	79	84	Validation set Test A
078	15.09	55	602	Validation set Test B
0.82	16.60	19	19	Set for polynomial fit Test A
0.87	16.98	42	44	Validation set Test A
0.89	15.13	58	61	Set for polynomial fit Test B
0.89	12.61	53	59	Set for polynomial fit Test B
0.90	13.75	62	65	Validation set Test B
0.90	13.41	55	602	Validation set Test A
0.91	17.99	32	334	Validation set Test B
0.95	18.79	71	75	Validation set Test B
0.95	12.55	53	59	Set for polynomial fit Test A
0.99	16.68	1	1	Set for polynomial fit Test A
0.99	17.62	45	48	Validation set Test A
1.04	15.77	56	603	Set for polynomial fit Test A
1.05	18.98	68	71	Validation set Test B
1.10	14.13	60	63	Validation set Test A
1.11	15.84	56	603	Set for polynomial fit Test B
1.14	14.34	72	77	Validation set Test A
1.15	16.92	19	19	Set for polynomial fit Test B
1.27	15.64	75	80	Validation set Test A
1.28	18.05	80	85	Validation set Test A
1.38	16.72	4	4	Validation set Test A
143	17.63	33	34	Validation set Test A
1.51	15.86	21	21	Validation set Test A
1.58	17.61	11	11	Validation set Test A
1.65	15.86	2	2	Set for polynomial fit Test A
173	15.87	35	36	Validation set Test B
176	16.94	14	14	Validation set Test A
1.82	15.70	2	2	Set for polynomial fit Test B

Table 3: Final tensile index ranking based on the absolute difference between measured and estimated values.

Tensile index (Rejected in the 1st ranking)				
abs (measured-estimated)	TI (Nm/g)	Sample	# at LAB	Corresponding test series
1.25	7.65	74	79	Rejected samples from Test A
1.47	545	28	32	Rejected samples from Test B
1.51	6.92	26	26	Rejected samples from Test A
1.69	7.17	5	5	Rejected samples from Test A
1.70	8.10	51	54	Rejected samples from Test B
1.78	6.17	57	604	Rejected samples from Test A
1.94	9.11	73	78	Rejected samples from Test B
1.96	5.13	32	334	Rejected samples from Test A
2.00	7.19	11	11	Rejected samples from Test B
2.03	7.45	27	27	Rejected samples from Test A
2.104	5.81	42	44	Rejected samples from Test B
2.19	4.57	66	69	Rejected samples from Test B
2.24	7.77	13	13	Rejected samples from Test A
2.26	7.19	52	55	Rejected samples from Test A
2.39	6.78	57	604	Rejected samples from Test B
2.57	8.38	17	17	Rejected samples from Test B
2.70	4.90	35	36	Rejected samples from Test A
2.71	5.00	64	67	Rejected samples from Test B
2.77	5.87	65	68	Rejected samples from Test B
2.82	1.66	54	601	Rejected samples from Test A
2.84	6.44	78	83	Rejected samples from Test A
2.90	6.88	28	32	Rejected samples from Test A
3.00	7.95	17	17	Rejected samples from Test A
3.00	2.23	60	63	Rejected samples from Test B
3.10	1.97	76	81	Rejected samples from Test B
3.12	4.72	71	75	Rejected samples from Test A
3.15	3.25	74	79	Rejected samples from Test B
3.36	5.70	33	34	Rejected samples from Test B
3.39	8.46	76	81	Rejected samples from Test A
3.58	4.16	23	23	Rejected samples from Test B
3.84	4.42	3	3	Rejected samples from Test A
3.98	5.35	80	85	Rejected samples from Test B
4.01	3.16	73	78	Rejected samples from Test A
4.23	4.70	69	73	Rejected samples from Test A
4.39	3.35	23	23	Rejected samples from Test A
5.06	8.10	7	7	Rejected samples from Test A
5.10	2.99	4	4	Rejected samples from Test B
5.19	2.52	64	67	Rejected samples from Test A
5.31	5.78	69	73	Rejected samples from Test B
5.55	7.61	7	7	Rejected samples from Test B
5.78	0.71	52	55	Rejected samples from Test B
11.98	6.95	77	82	Rejected samples from Test B
12.88	6.04	77	82	Rejected samples from Test A

the model parameters derived can change when using e. g. other refining segments, production levels etc.

It is finally, interesting to note that the models can be used in other processes based on CD-82 refiners as well. This statement was confirmed by implementing the model (with another intercept of course) in a TMP process, see Karlström et al. (2018).

Concluding remarks

The main purpose of this study is to investigate outliers in laboratory data. It is shown that a generalized ESD procedure can be used. It is also seen that the significance levels do not affect the number of outliers. However, it is questionable if traditional laboratory measurement pro-

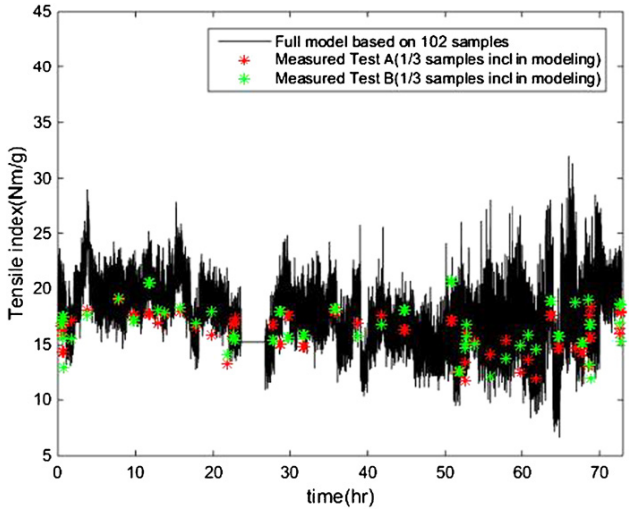

Figure 9: Estimated and measured (TestA and TestB) tensile index.

cedures provide insight enough regarding the accuracy in each measure in the data series. In this paper it is stressed that enough samples must be collected and analyzed to get an acceptable significant level in each measure.

Using temperature profile measurements, it is possible to derive hidden physical phenomena that are impossible to measure inside the refining zones. Such measures are typical internal variables and, in this study, we use the consistency from each refining zone and the fiber residence time in each refining zone.

A procedure, including data selection and rearrangement of data before modeling and validation, is introduced in this paper to cope with larger data sets.

The internal variables perform a stable fit and reproduce the properties studied.

It is an understatement that non-linearities exist in the refining process. It is shown that models using internal variables as predictors can improve the model accuracy considerably. This makes it more interesting to further study the internal variables. It is interesting to see that ranked accepted measurements obtained from the methodology outlined in this paper give a possibility to improve the analysis of the data if enough data from different process operating points are considered.

Finally, it is indicated that tensile index can be optimized by changing the consistency and fiber residence time. The collinearities in the predictors however requires on-line implementation of the extended entropy model derived by Karlström and Eriksson (2014a, 2014b, 2014c, 2014d) to get reliable estimates of the consistency and fiber residence time when changing dilution water flow rates and plate gaps.

Acknowledgments: Special thanks to Rita and Olof Ferritsius (MidSweden University) for the support and encouragement.

Funding: The authors gratefully acknowledge the funding of the Swedish Energy Agency, StoraEnso and Holmen Paper. Special thanks go to the StoraEnso Skoghall mill for running trials and providing the excellent laboratory and process data used in this study.

Conflict of interest: The authors declare no conflicts of interest.

Appendix A. Detection of outliers

The outliers detected by the generalized ESD procedure are given in Table A.1 and it is seen that the significance levels do not affect the number of outliers detected.

However, do we actually have insight enough regarding the accuracy in each measure in the data series or do we set the significant levels based on traditions?

Is the spread in tensile index, as given in Figure 5, maybe too large?

Consider an example where $\alpha = 0.05$ and 0.1, i. e. by tradition we are assumed to be 95 % and 90 % confident that we have no outliers. This means that we on beforehand assume that we can be wrong with a probability 0.05 and 0.1.

When using the generalized ESD procedure however, the percentile values of the t-distribution is not only dependent on the selected significance level but also on the degrees of freedom $(nm - 1)$ we have in the data set.[10] For example, if $\alpha = 0.05$, the lower and upper percentile in

Table A.1: Indices for the outliers detected in the second iteration using the generalized ESD procedure.

$\alpha = 0.05$		$\alpha = 0.1$	
Indices Outliers detected		Indices Outliers detected	
Case A	Case C	Case A	Case C
87	87	87	87
93	93	93	93
96	96	96	96
99	99	99	99
271	181	271	181
	271		271

10 It is also notable that Equation 4 describes a two-sided outlier distribution. To describe one-sided outlier problems we substitute $\alpha/2$ by α in the value of p.

Equation 4 will be {99.17, 99.99}. This is indeed a conservative setting and the use of the traditional concept with a pre-specified significance level can to some extent become misleading in the analysis. This is best illustrated by setting the significance level (or whatever we call it in this specific case[11]) to 0.9 and 0.99 which yields a completely different picture as seen in Table A.2 and we can conclude that this results in a less conservative limit which stretch out the definition of outlier detection procedures. This is also seen in Figure 10 where the histograms for R, $\lambda(\alpha = 0.05)$ and $\lambda(\alpha = 0.99)$ are given. Another way to illustrate this is to plot the sorted data for R, $\lambda(\alpha = 0.05)$ and $\lambda(\alpha = 0.99)$ see Figure 11.

Note, $\lambda(p = 0.998)$ has been included in Figure 11 as well and can be seen as an alternative outlier setting at a constant percentile value in the t-distribution. In summary we can see that both Figure 10 and Figure 11 visualize the intercepts for $R_j \geq \lambda_i$ clearly together with the potential outliers.

It is also interesting to see that no difference was detected between Case A and Case B in Table A.2, while the number of detected outliers is doubled for Case C. This is most likely a consequence of how the use of average basis weight in Case A and Case B smooth out the variations while in Case C the basis weight for each handsheet is used.

The use of the measure defined in Case C causes a risk for a larger spread in the data set. However, which case to use in the analysis is not obvious as the tensile strength and basis weight are expected to be inherently correlated at the same time as we are looking for possible outliers necessary to analyze further. Nevertheless, to illustrate the methodology we will use Case C as a reference below.

The outlier detection procedure can be illustrated in a number of different ways and in Figure 12, Case C (2nd iteration) is compared for two levels $\alpha = 0.05$ and 0.99. The upper and lower limits are changed marginally when the confidence reduces which is also seen in Figure 13.

From a laboratory perspective, the lower limits are of certain interest and therefore it is tempting to analyse the distribution to see how many outliers we get in this region. In Figure 14, the number of expected outliers are given versus the significance level for Case C.

The two cases in Figure 12 approach a normal distribution. This statement is strengthened by the normal probability plot and the histogram for Case C ($\alpha = 0.05$) in Figure 15 and Figure 16, respectively.

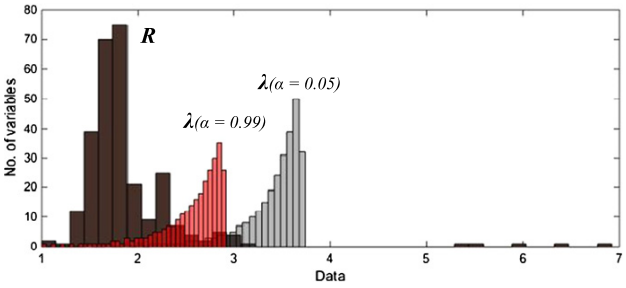

Figure 10: Histogram for R, $\lambda(\alpha = 0.05)$ and $\lambda(\alpha = 0.99)$ in Case C for the detrended tensile index (data).

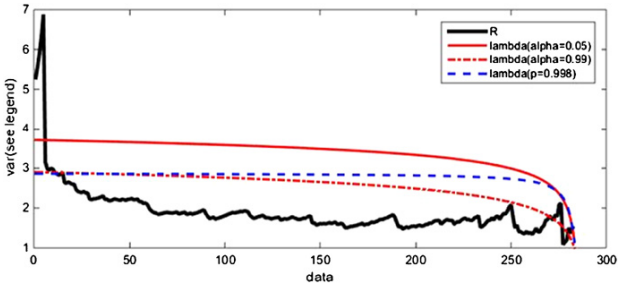

Figure 11: R, $\lambda(\alpha = 0.05)$ and $\lambda(\alpha = 0.99)$ in Case C for the detrended tensile index (data).

Table A.2: Indices for the outliers for all three cases detected in the second iteration using the generalized ESD procedure.

$\alpha = 0.9$			$\alpha = 0.99$		
Indices Outliers detected			**Indices Outliers detected**		
Case A	**Case B**	**Case C**	**Case A**	**Case B**	**Case C**
21	21	5	21	21	5
42	42	18	42	42	18
63	63	42	63	63	42
87	87	75	87	87	45
90	90	80	90	90	75
93	93	87	93	93	80
95	95	93	95	95	87
96	96	95	96	96	90
99	99	96	99	99	93
167	167	99	167	167	95
181	181	129	181	181	96
271	271	161	271	271	99
		167			129
		233			161
		234			167
		239			233
		249			234
		264			239
		271			249
					264
					271

11 Note, to reach the limit of p, i. e. $[0, 100]$, $\alpha \approx 6$. These consequences relate to extreme situations which most often are ignored in the research literature as knowledge about the data set often is assumed to be secured.

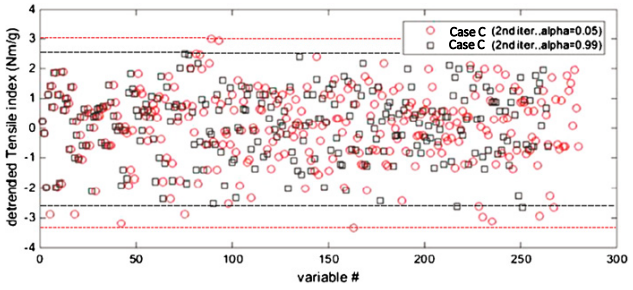

Figure 12: Detrended tensile index for Case C after the 2nd iteration for different significance levels.

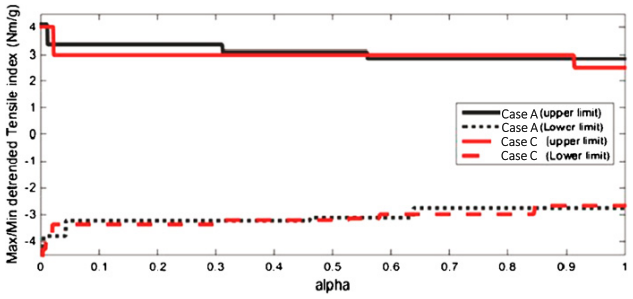

Figure 13: Upper and lower limits for the detrended tensile index versus α.

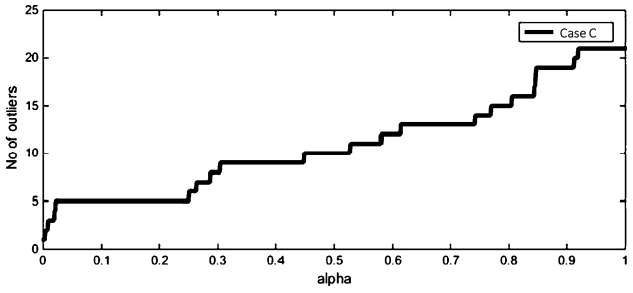

Figure 14: Number of detected outliers versus α.

As seen in Figure 15, the interval between the 25th and 75th percentiles indicates a tensile index distribution of about +/−0.8. Moreover, in Figure 16 the distribution is somewhat skew to the left (skewness is about −0.42) while the kurtosis is about 2.9. In other words, the kurtosis indicates that we approach a normal distribution as the peakedness of the distribution approach 3 (which is the value for a strictly normal distribution) at the same time as the number of tensile index values in the lower region of acceptable measures is higher than expected.

For symmetric distributions the skewness is zero but in our example, the skewness is far from zero when α = 0.05 which indicates that some of the tensile index measurements are in the lower region of acceptable measures.

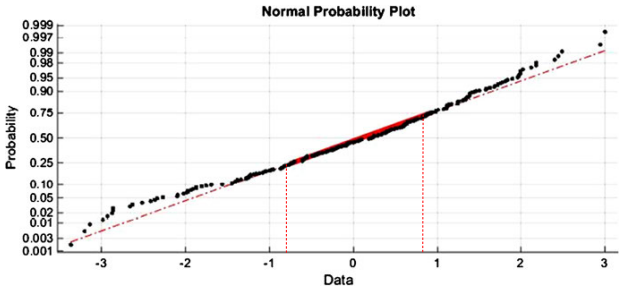

Figure 15: Probability versus the detrended tensile index (data) for α = 0.05.

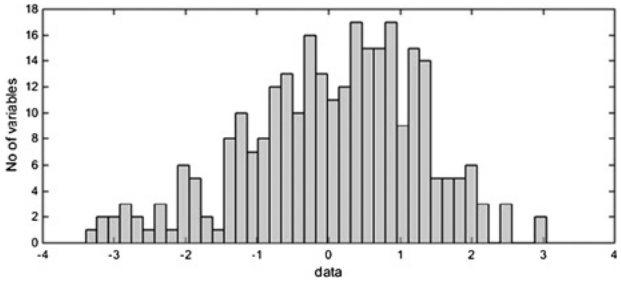

Figure 16: Histogram for Case C for the detrended tensile index (data) when α = 0.05.

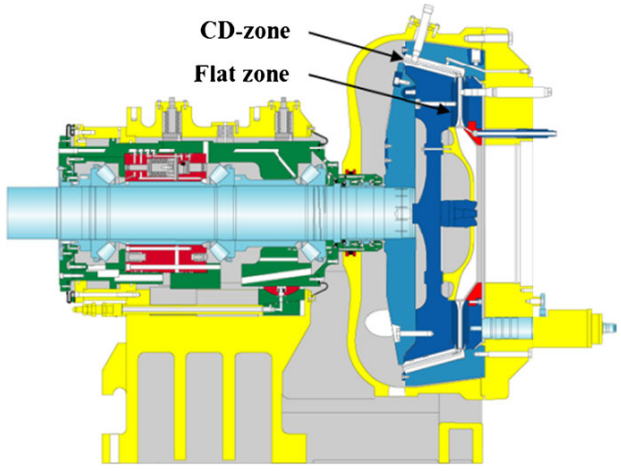

Figure 17: A schematic drawing of a CD refiner. The vertical flat zone (FZ) is directly linked to the conical zone (CD) via an expanding point.

If α = 0.99 the skewness is reduced significantly to −0.29 as expected but this also means that the number of outliers to be analyzed increase furthermore as indicated in Figure 10 and Figure 14.

It is obvious that the lower tails in Figure 15 and Figure 16 are interesting to study further as they relates to the tensile strength in the paper which is considered to be as high as possible.

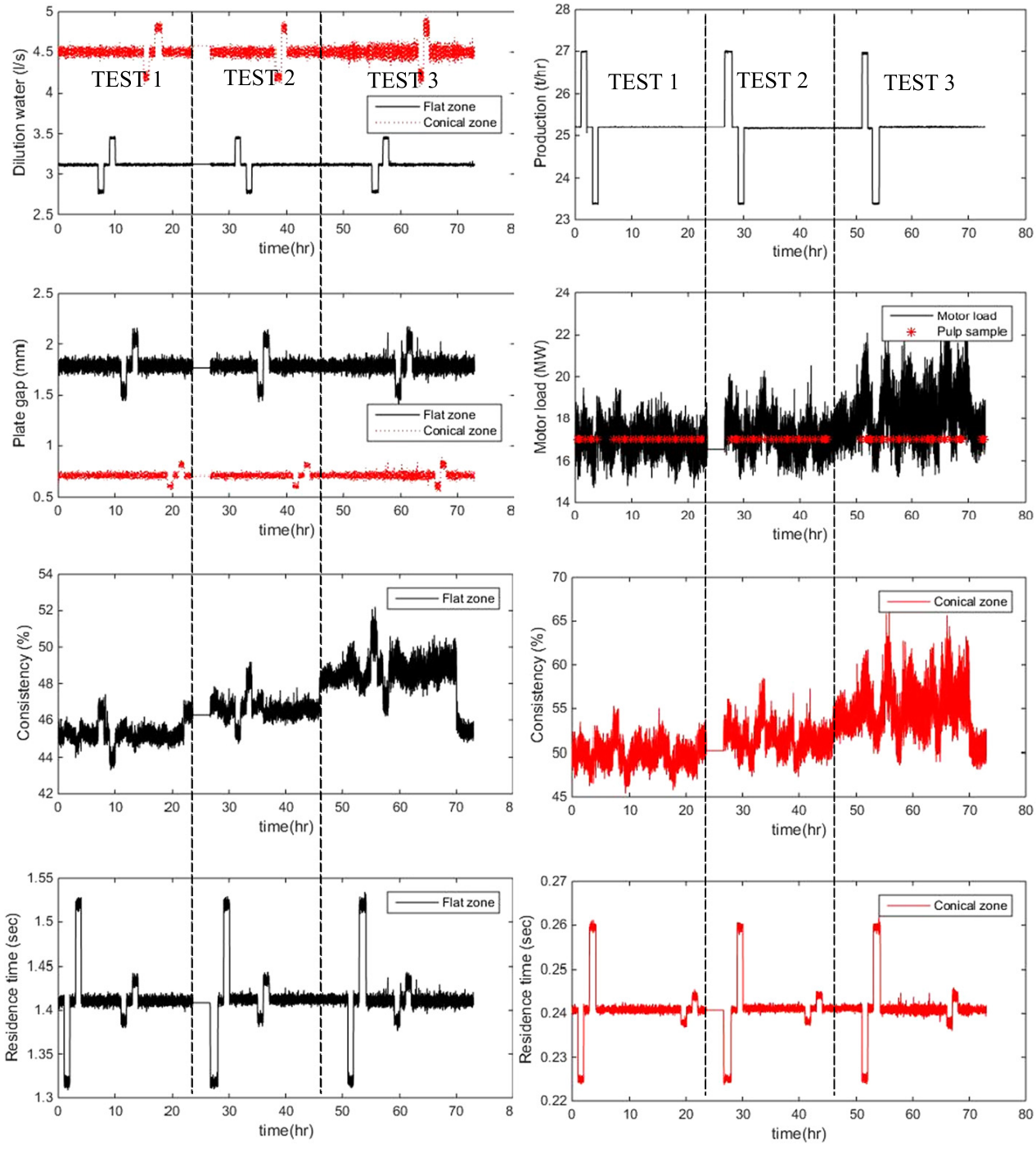

Figure 18: Step changes performed in the external variables dilution water (upper left), production (upper right) and plate gap (middle left); response in motor load including time for each test point (middle right). Responses in the internal variables consistency and residence time (lower figures).

Appendix B. Step changes in internal variables and responses in internal variables

Data from a full-scale CTMP production line (CD82-refiner) have been used, see Figure 17. In both the flat zone (FZ) and the conical zone (CD), sensor arrays with eight sensors have been mounted to measure the entire temperature profiles. The temperature measurements can be seen as internal variables that are measured together with traditional process variables, such as production rate, dilution water flows, plate gaps and motor load (external variables), and vary considerably when changes are made in process conditions and the refining segment pattern.

Both internal and external variables are used in the extended entropy model (Karlström and Eriksson (2014a, 2014b, 2014c, 2014d)), which can be used for estimation of e. g. the consistency profile and the fiber residence time in the FZ and CD zones (Karlström and Hill (2017a, 2017b, 2017c).

The test was performed according to Figure 18, and the time for each test point was well-documented. From a laboratory test program perspective, the test program was extensive and covered 80 test points where pulp samples were taken from the blow-line valve over a period of 3 minutes each. As seen in Figure 18, the test was performed using three distinct sets of pulp samples with different chip mixtures (TEST1; 100 % saw mill, TEST2; 65 % saw mill and 35 % roundwood and TEST3; 100 % roundwood) following the same step changes in the manipulated variables: dilution water feed rate, production and plate gap. Besides recording the time[12] for pulp sampling, ten grab samples were taken to get a reliable and synchronized mean value of each pulp sample. The pulp samples were then homogenized carefully and double tested. In total, 2×51 samples were analyzed.

The process sampling rate was 1 second, which resulted in a sampling matrix for this test of the size 300×260000.

The original idea of using internal variables instead of external variables to find proper piece-wise linear models was to cope with non-linearities in the process, see Karlström et al. (2015, 2016a, 2016b). By using information from the estimated consistency profile and the fiber residence time it was shown that the internal variables outperformed the external variables as independent variables (predictors) when making polynomial fits of pulp and handsheet properties.

Fully understanding the relationships in refining zone conditions, when the external variables are changed, is a challenge. For instance, an increased production rate will result in a reduced residence time while an increased dilution water feed rate has a limited effect on the residence time. On the other hand, when the plate gap is increased, the residence time will be reduced. Moreover, an increased dilution water feed rate in the FZ will reduce the consistency in the FZ and CD while an increased plate gap in FZ has a minor impact on the consistency, see Figure 18. Hence, when the plate gap in the CD zone is changed, the consistency is not affected linearly. This is most likely a consequence of the non-negligible changes in the fiber pad.

In Figure 18, three test series are available for analysis, where each test series comprises double tested pulp samples which means that the procedure outlined above for small data sets is not suitable.

12 When matching laboratory variables to a set of process variables, it is essential to record the time for pulp sampling and the sampling interval.

References

Barnett, V., Lewis, T. Outliers in Statistical Data. Wiley, Chichester, 1994.

Draper, N.R., Smith, H. Applied Regression Analysis. 3rd ed. Wiley, New York, 1998.

EPA. (2006) Data Quality Assessment: Statistical Methods for Practitioners EPA QA/G-9S, EPA/240/B-06/003, U. S. Environmental Protection Agency, Office of Environmental Information, Washington DC.

Forgacs, O.L. (1963) The characterization of mechanical pulps. Pulp Pap. Mag. Can. 89–118.

Freedman, L.S., Pee, D. (1989) Return to a note on screening regression equations. Am. Stat. 43:279–282.

Gilbert, R.O. Statistical Methods for Environmental Pollution Monitoring. Wiley & Sons, Inc., New York, NY, 1987.

Harrell, F., Lee, K.l., Matchar, D.B., Reicert, T.A. (1985) Regression models for prognostic prediction: Advantages, problems and suggested solutions. Cancer Treat. Rep. 68:1071–1077.

Härkönen, E., Huusari, E., Ravila, P. (2000) Residence time of fibre in a single disc refiner. Pulp Pap. Can. T:330–335.

Härkönen, E., Kortelainen, J., Virtanen, J., Vuorio, P. (2003) Fiber development in TMP main line. In: International Mechanical Pulping conference, Quebec, Que, Canada, 2–5 June 2003. pp. 171–178.

Karlström, A., Eriksson, K. (2014a) Fiber energy efficiency Part I: Extended entropy model. Nord. Pulp Pap. Res. J. 29(2).

Karlström, A., Eriksson, K. (2014b) Fiber energy efficiency Part II: Forces acting on the refiner bars. Nord. Pulp Pap. Res. J. 29(2).

Karlström, A., Eriksson, K. (2014c) Refining energy efficiency Part III: Modeling of fiber-to-bar interaction. Nord. Pulp Pap. Res. J. 29(3).

Karlström, A., Eriksson, K. (2014d) Refining energy efficiency Part IV: Multi-scale modeling of refining processes. Nord. Pulp Pap. Res. J. 29(3).

Karlström, A., Hill, J. (2017a) CTMP process optimization Part I: Internal and external variables impact on refiner conditions. Nord. Pulp Pap. Res. J. 32(1).

Karlström, A., Hill, J. (2017b) CTMP process optimization Part II: Reliability in pulp and handsheet measurements. Nord. Pulp Pap. Res. J. 32(2).

Karlström, A., Hill, J. (2017c) CTMP process optimization Part III: On the prediction of Scott-Bond, Z-strength and tensile index. Nord. Pulp Pap. Res. J. 32(2).

Karlström, A., Hill, J., Ferritsius, R., Ferritsius, O. (2015) Pulp property development Part I: Interlacing undersampled pulp properties and TMP process data using piece-wise linear functions. Nord. Pulp Pap. Res. J. 30(4).

Karlström, A., Hill, J., Ferritsius, R., Ferritsius, O. (2016a) Pulp property development Part II: Process nonlinearities and its influence on pulp property development. Nord. Pulp Pap. Res. J. 31(2).

Karlström, A., Hill, J., Ferritsius, R., Ferritsius, O. (2016b) Pulp property development Part III: Fiber residence time and consistency profile impact on specific energy and pulp properties. Nord. Pulp Pap. Res. J. 31(2).

Karlström, A., Hill, J., Johansson, L. (2018) An overview of some efforts to understand CD-refiners. In: International Mechanical Pulping Conference, Trondheim.

Lowe, G.K., Zohdy, M.A. (2010) Modeling nonlinear systems using multiple piecewise linear equations. Nonlinear Anal.: Model. Control 15(4):451–458.

Nelsson, E. (2016) Improved energy efficiency in mill scale production of mechanical pulp by increasing wood softening and refining intensity, ISSN: 1652-893X, ISBN: 978-91-88025-59-3, PhD thesis, Mid Sweden University.

Peduzzi, P., Cocato, J., Kemper, E., Holford, T.R., Feinstein, A.R. (1996) A simulation study of the number of events per variable in logistic regression analysis. J. Clin. Epidemiol. 49(12):1373–1379.

Rosner, B. (1983) Percentage points for a generalized ESD many-outlier procedure. Technometrics 25:165–172.

Sabourin, M., Wiseman, N., Vaughn, J. (2001) Refining theory considerations for assessing pulp properties in the commercial manufacture of TMP. In: 55th Appita Annual Conference. pp. 195–204.

Stephens, M.A. (1974) EDF statistics for goodness of fit and some comparisons. J. Am. Stat. Assoc. 69(347):730–737.

Strand, B.C. (1996) Model based control of high consistency refining. Tappi J. 79(10):140–146.

Strand, B.C., Grace, B. (2014) Implementation of advanced supervisory control within a TMP refiner quality control system. In: International Mechanical Pulping Conference, Helsinki, Finland.

Vittinghoff, E., McCulloch, C.E. (2007) Relaxing the rule of ten events per variable in logistic and Cox regression. Am. J. Epidemiol. 165:(6).