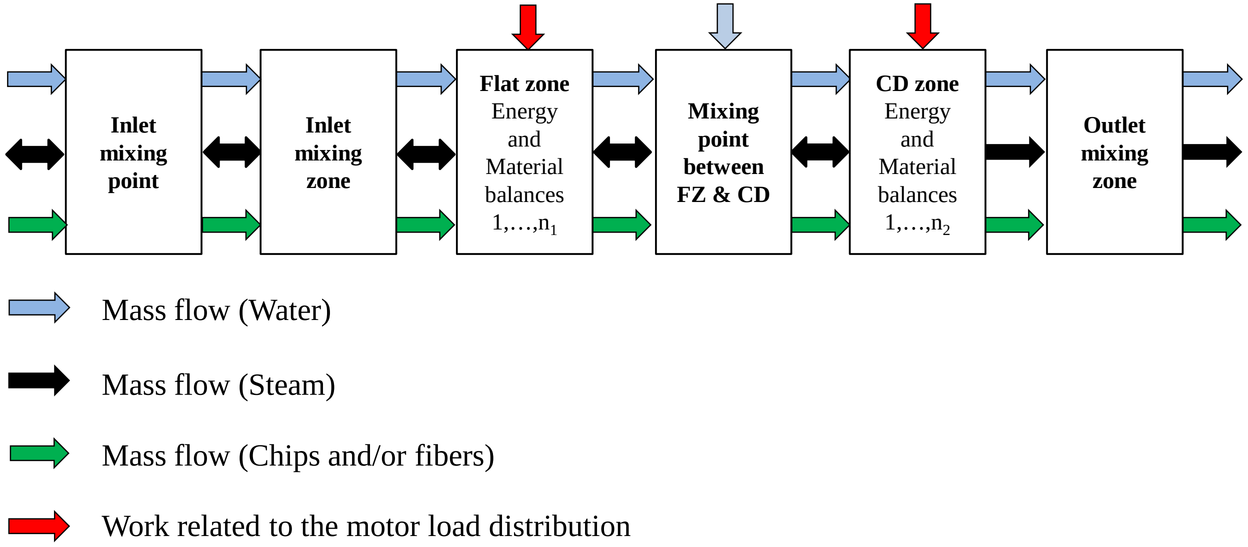
Refining models for optimization and control - How to use such models for pulp and handsheet properties estimation.

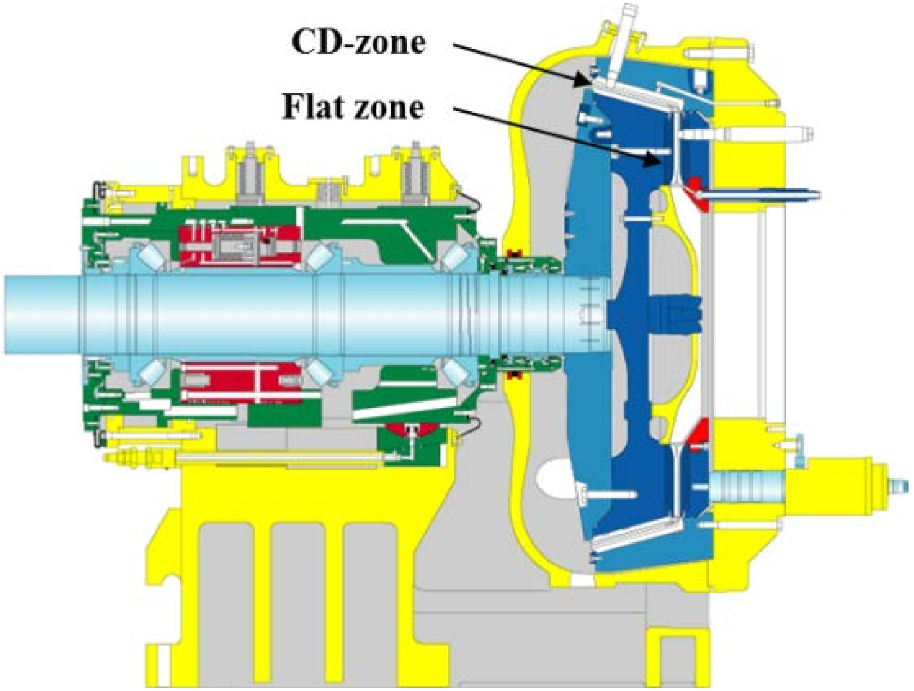
Fredrik Bengtsson*, Anders Karlström*, Jan Hill**, Lars Johansson***,

*Chalmers University of Technology, Göteborg, Sweden,
**QualTech AB, Tyringe, Sweden
***RISE-PFI, Trondheim, Norway

Pulp refiners

- Focus on a CD refiner
  - Contains two refining zones - flat zone and conical zone.
  - Physical model for two serially linked





Modeling of Tensile index – Issues to consider

- Time consuming procedure to measure
  - Requires manual measurement.
  - Measurements are stochastic, requiring numerous measurements for accurate results.
  - Hard to link to process data.
- Access to tensile measurements is necessary to implement models for automatic control.

Creating a model

- The goal is to predict tensile index from other process measurements.
  - Can be quicker and cheaper than measuring it manually.
- Measurements which can be used to estimate tensile index:

External variables

- Specific energy
- Production rate
- Dilution water flow rate
- Plate gaps

Internal variables

Variables derived from physical models, e.g.

- Consistencies
- Residence times

Deriving a linear model

$$\text{Tensile Index} = \theta_1 x_1 + \theta_2 x_2 \ldots \theta_N x_n$$

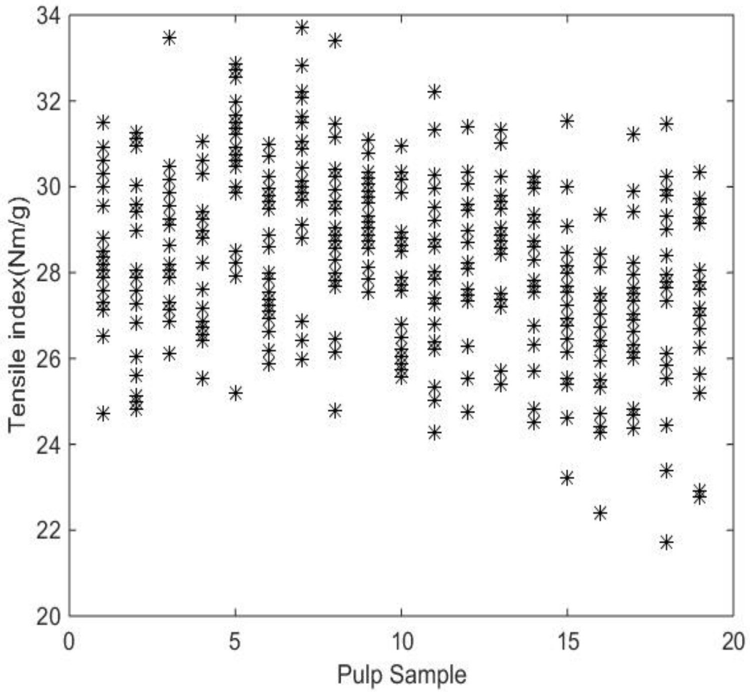
- x_1, x_2,…, x_n are the predictors (available measurements)
- Using a data set, we will train a θ to best describe the relationship between the predictors and the tensile index.

Tensile index measurements

- To create a model we acquired a set of measurements from a CD-refiner.
- 19 Pulp samples were used.
  - Taken during a period of 5 days.
  - For each pulp sample, 20 strips were used for tensile strength measurements.

Tensile Index measurements

- There was considerable variation between tensile index measurements for each sample
  - Outliers were removed;
    - Values more then 3 standard deviations from the mean were disregarded.
  - Thereafter, the average of the remaining measurements was chosen to estimate the tensile index for each sample.



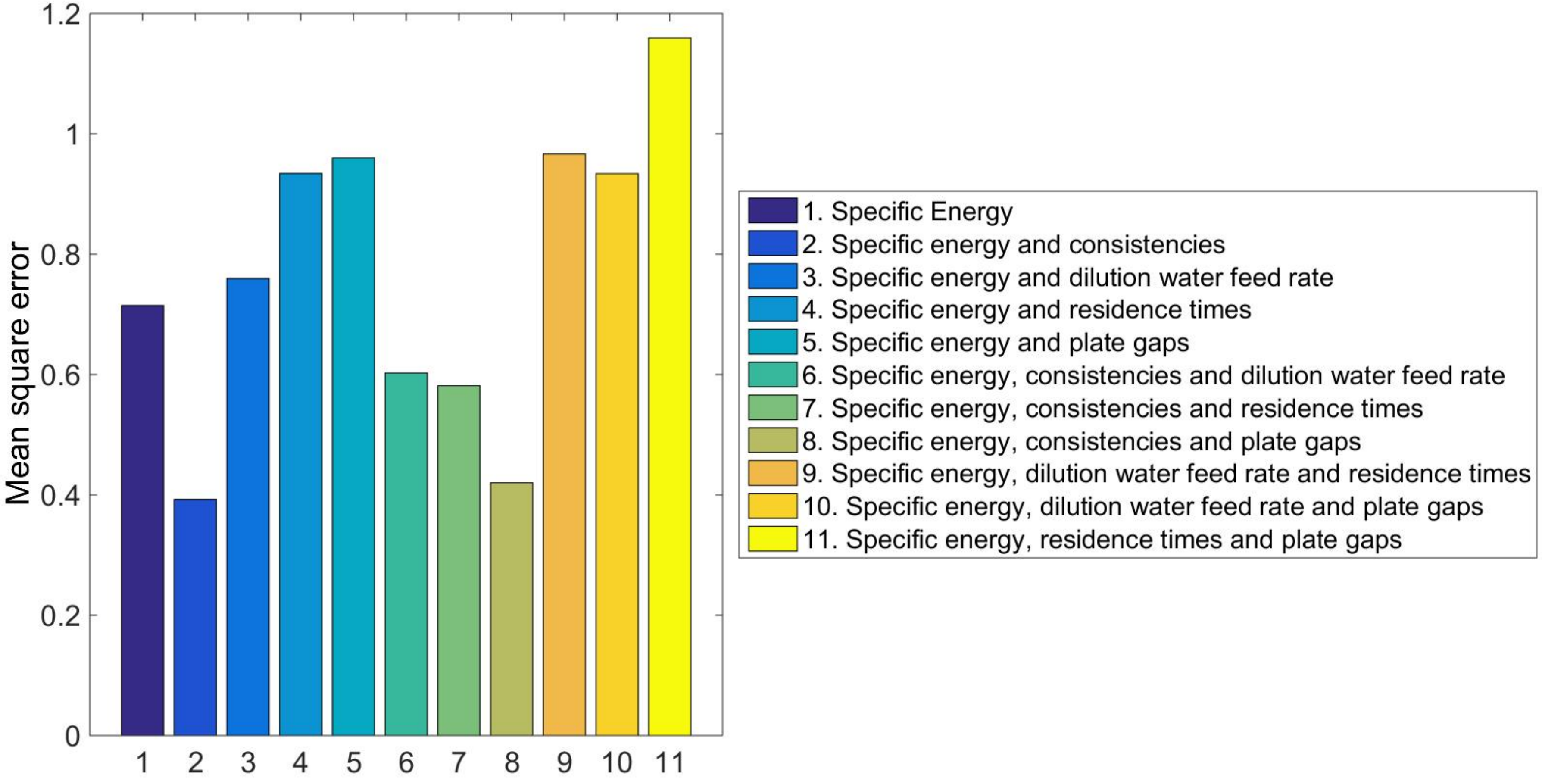
Determining which predictors to use

- The model should preferably use only a few predictors.
  - Using many different predictors will lead to needlessly complex models and overfitting.
    - Especially as the different predictors are not entirely independent from each other.

Determining which predictors to use

- The dataset was randomly split into two parts, a *training set* and a *validation set*.
  - With 12 data points in the *training set* and 7 in the *validation set*.
- Models were trained with different predictors using the training set.
- Each model structures' mean prediction error was calculated on the validation set.
- This procedure was repeated 10 000 times to get an average mean prediction error for each set of predictors.

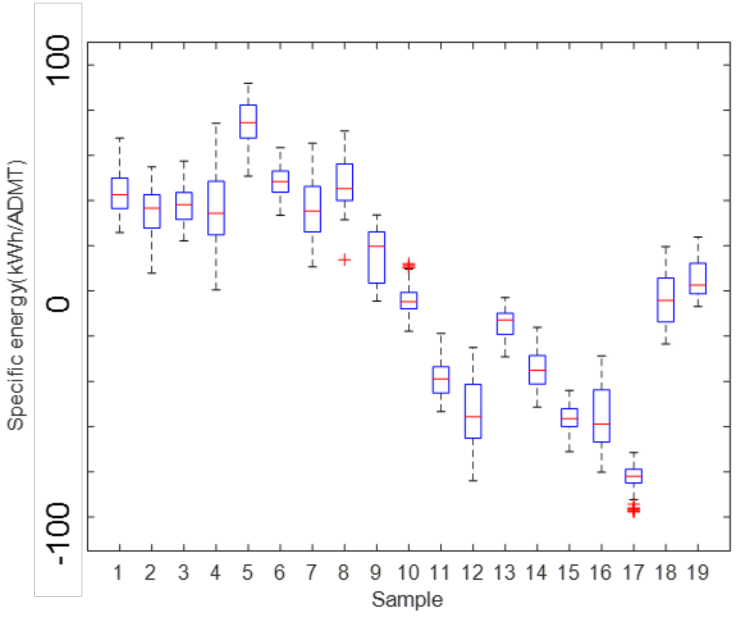
Determining which predictors to use

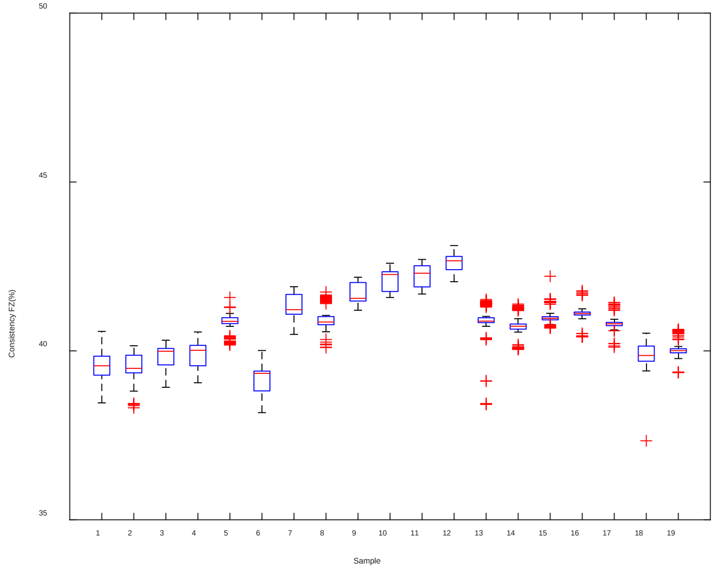


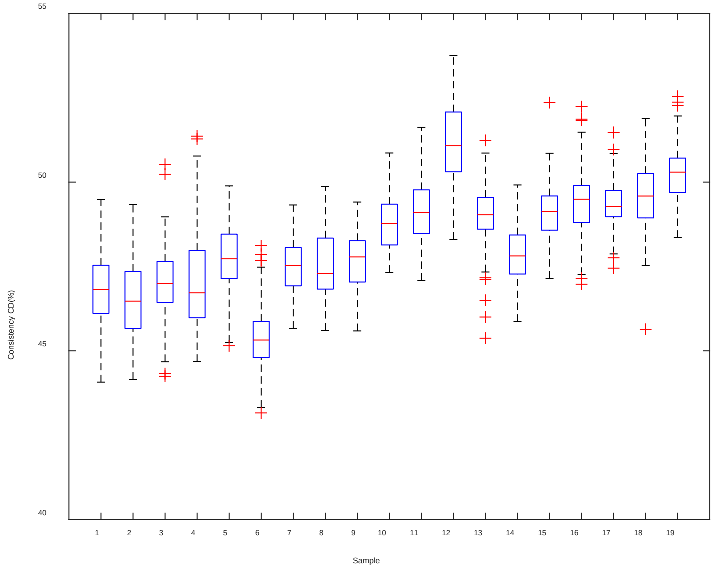
Determining which predictors to use

- Using specific energy and consistencies yields the best model for this data set.
- Further investigations found that using consistency measurements from the conical zone did not improve the model.
  - This is a consequence of the fact that some consistency values are strongly linked to the changes in the flat zone. In future tests this can be handled by designing the tests with specific changes in consistency in the conical zone.
  - This can also be a consequence of that the characteristics of tensile index are strongly related to the defibration in the flat zone (hence, further research needed).

Process variations in predictors







- Specific energy interval : +/-100 kWh/T
- Larger variations in the conical zone consistency compared with the consistency in the flat zone. Note! In this case only the outlet consistencies from each zone are considered.

Outlier detection

- With only 19 data points the impact of each data point on the model is considerable.
- Measurements are unreliable and not all information is included in the model.
  - High probability of outliers.

Outlier detection

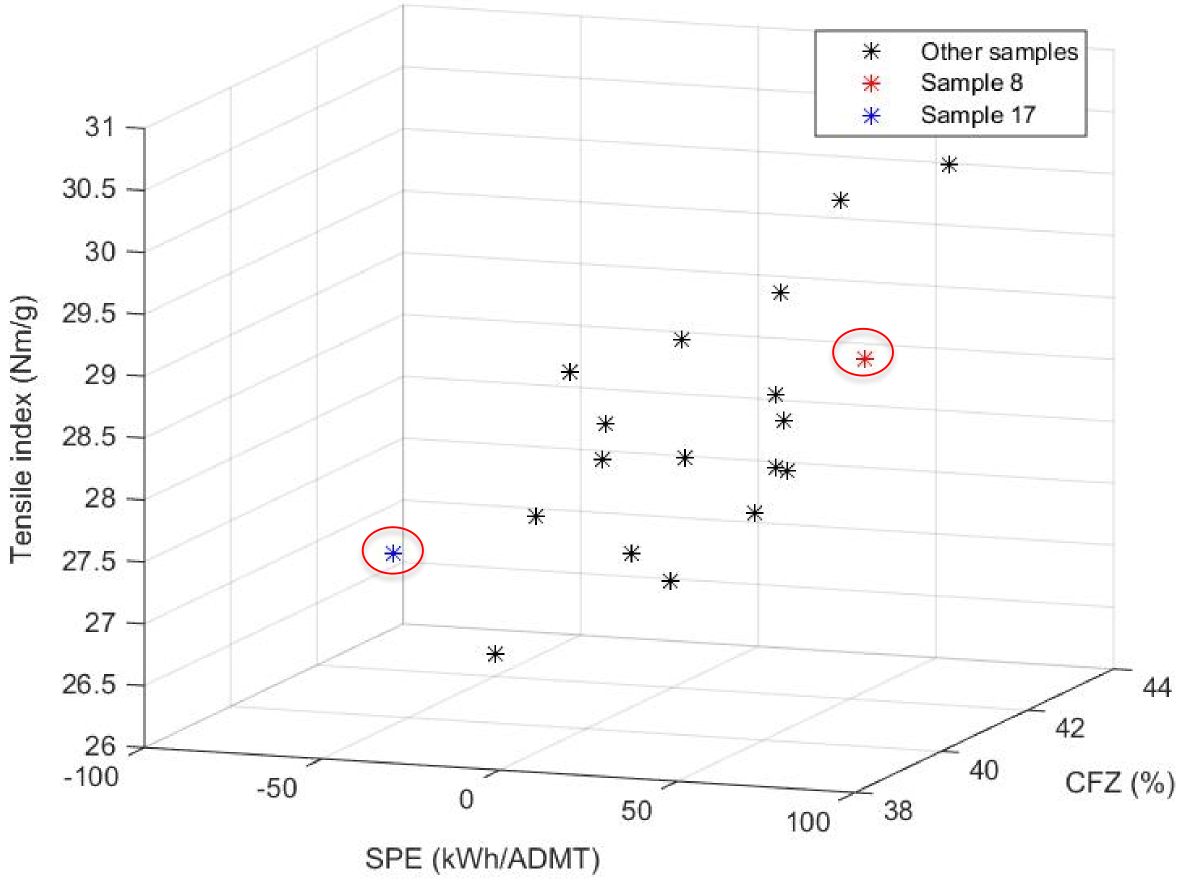
- To detect outliers we designed models using only a subset of the data.
  - All possible combinations of ten or more samples tested.
  - In total, **262144 combinations** were tested.
- The adjusted R^2 was calculated for each model.
  - We noted how many times each sample occurred in models with an adjusted R^2 greater than 0.9.

Outlier detection - Results

Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Number of times it occurs in good models	24059	25783	23078	25126	32084	25424	27989	**6189**	27054	25408	26050	22790	25425	26371	27601	30213	**0**	25380	25843

- Two samples seem to be likely outliers
    - Sample 8 and 17.

Visual inspection



Outliers

- Other methods were used to confirm the outliers:
  - Using validation and training set in the same way as when determining predictors:
    - Removing sample 8 and 17 yielded the largest reduction in validation cost when the impact of removing pairs of samples was examined.
  - Cook’s distance
    - A measure of how much removing samples changes the model.
    - The pair consisting of sample 8 and 17 had the largest Cook’s distance.

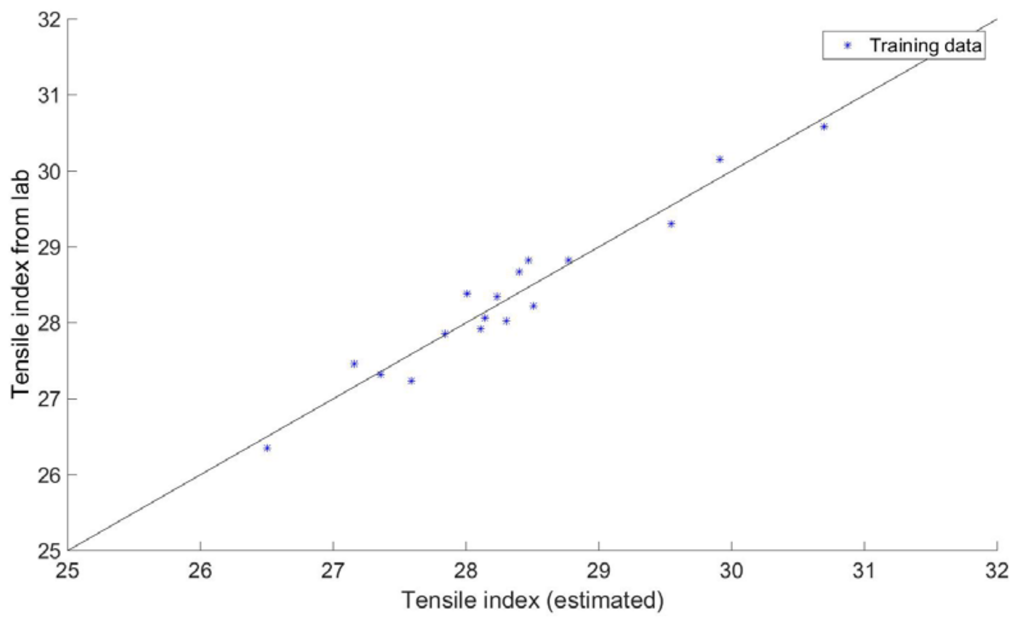
Resulting model

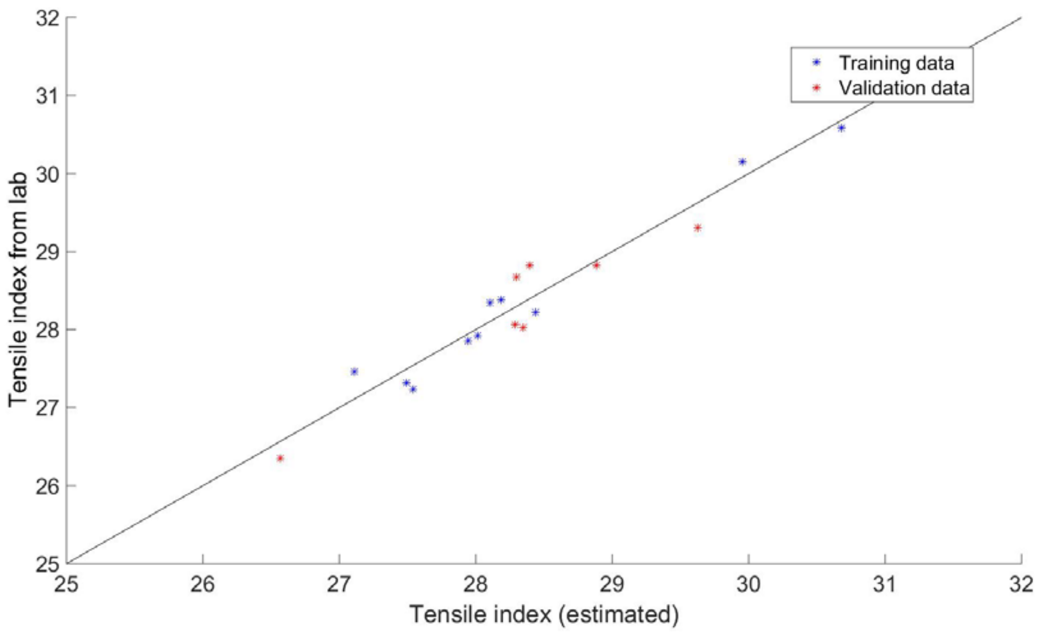
$$\text{Tensile Index} = -56.6304 + 0.034x_1 + 0.9595x_2$$

- x_1 is the specific energy (kWh/T)
  - Modelled on data in range of -100 to +100 kWh/T; (Mills specific info)
- x_2 is the consistency of the flat zone (%)
  - Modelled on data in range of 39.2-42.6 %

Evaluating the model structure

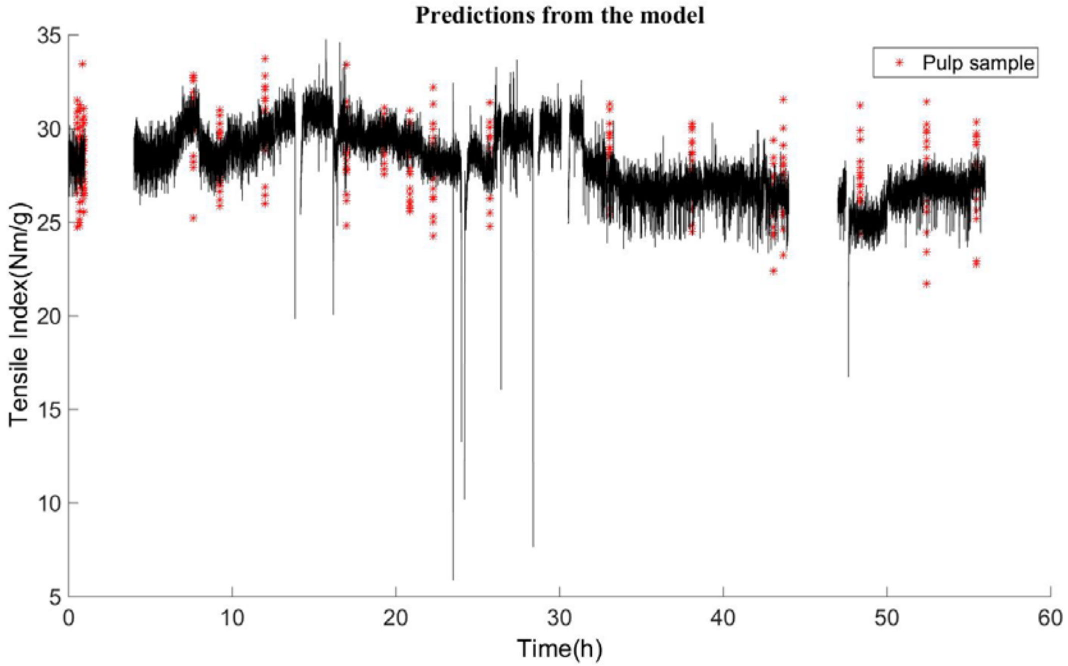
- The model had an adjusted R^2 of 0.939





Utilizing the model

- Predicting the tensile index
- Controlling the tensile index
  - Consistencies can be controlled independently in each zone
  - Model should be expanded to include impact of consistency in the conical zone.



Conclusions

- The best model of the tensile index was found using the specific energy and the consistency of the flat zone.
  - So, a combination of internal and external variables yielded the best result.
- After outliers were removed the model showed good results.
  - However evaluating the model on a new set of data may give new insights.

Further Work

- Use more data to further evaluate and develop the model
  - Evaluate how well the model explains this new dataset.
  - With additional data a more complex model may show to be more appropriate
- With more accurate measurements for the consistencies in the conical zone, the model can be expanded to include the impact of these.
- Examine other handsheet properties.