INTERNSHIP AT THE UNIVERSITY OF TENNESSEE; KNOXVILLE

EMPLOYMENT REPORT

Statistical Analysis of the Retention Testing Protocol for Treated Lumber

performed at the course of studies

Forest Products Technology and Wood Construction

Salzburg University of Applied Sciences/Kuchl Campus

submitted from

David Christoph Juriga

Head of degree Course: FH-Prof. DI (FH) Dr. Alexander Petutschnigg

FH-supervisor: FH-Prof. DI (FH) Dr. Alexander Petutschnigg

Business-supervisor: Prof. Timothy M. Young, PhD

Knoxville, October 31, 2016

Acknoledgements

I wish to express my sincere gratitude to Dr. Timothy M. Young for support and encouragement throughout the internship. Furthermore, I'd like to thank him for providing me an opportunity to do my internship, here at the University of Tennessee, Knoxville.

I would also like to thank Dr. Adam Taylor for supporting my work and for his valuable guidance. I also want to thank Anton Astner who not only helped me with statistical questions, but also mentored me at every stage of the work. Special thanks also to the Langdale Forest Products Company of Sweetwater, Tennessee which helped a lot a with tour through the treatment plant and always answered my questions. Funding for this project were provided by the Austrian Marshall plan Foundation and given as a Scholarship.

Finally, I would like to thank my family and friends for their support. I would like to thank my mother and father for their encouraging words, without which this internship could not have been completed.

Kurzfassung

Jedes Unternehmen möchte konkurrenzfähig sein, in diesem Fall sind einige der Holzimprägnierungsfirmen besorgt, dass bei einigen derer Messungen verglichen zu den Messungen von Holzagenturen unterschiedliche Werte gemessen werden, beziehungsweise manche Chargen bei der Messung des Unternehmens über dem im AWPA Standard gegebenen Wert sind und bei der Messung durch die Agentur unter dem Wert. Frühere Arbeiten haben gezeigt, dass Statistical Process Control (SPC) ein geeignetes Werkzeug zum Analysieren der Daten und der Datenqualität ist. In diesem Report wurde zuerst der Prozess der Messung betrachtet. Als nächstes wurden die Datenpaare der Imprägenierungsfirma und der Fremdagentur verglichen und analysiert. Diese wurden in x-individual und in moving Range Diagrammen dargestellt, um die Datenqualität zu visualisieren. Ein großes Problem in diesen Daten Sets ist die große Varianz, verursacht durch die Wiederholung der Messungen, die unter dem gegebenen Wert im AWPA Standard sind. As letzte wichtige Information wurde die Verteilung der Daten angeschaut. Mit diesen Werten war es möglich einige Empfehlungen, wie man den Prozess verbessert und wie man diesen statistisch unter Kontrolle bekommt, zu geben.

Abstract

Every Company wants to be competitive, in this case some of the lumber treating mills are concerned that they measure their treated residential lumber and sometimes when a third-party agency measures they measure values under the given value of the standard. Previous work has demonstrated that statistical process control is a suitable tool to analyze data and look at the data quality. In this report the way of measurement was analyzed. As next the data pairs of the treatment plant and the third-party agency were compared and analyzed. These were also plotted on an x-individual and moving range chart to visualize the data quality. A huge problem in the datasets is the large variance problem, caused by the retesting of the measurements under the given level in the AWPA Standard. As another important information, the distribution of the data was looked at. With these values, it is possible to give some recommendations, how to improve the process and how to get the process under statistical control.

Index

Introduction

All companies should stay competitive to create a good business. The treating plants should optimize their process to gain more knowledge and a better output. Up to now the companies hardly have any statistical control over their process. In this case the treating mills are concerned about the fact that some of their retention measurements pass the given value of the standard, but when a third-party agency measures, the same batch does not pass the given retention value. Previous work in other sectors has shown that a lot of optimizations through statistical process control can be made. All those companies could save in one year a lot of money, just by installing the statistical process control in their companies. They proved that this tool is a suitable tool to visualize and analyze data quality. In this report the process of the pressure treatment was looked at and the way of sampling according to the AWPA Standard. Than different statistical tools where applied to the datasets. At first the difference between the treating mill data and the third-party agency data was looked at with a summary of statistics and a distribution plot. This lead the project to the large variance problem, which is caused by the possibility to retest batches, when the retention value is under the given value of the AWPA Standard. The data was also plotted in an x-individual and moving range chart to visualize the process and see changes far earlier. The data was imagined in Boxplots, which is a tool to compare different processes or retention values. Important for statistical analysis is the distribution of the dataset. This was analyzed with a life cycle analysis to compare different distributions and find out which one is the most accurate one on the dataset. Finally a dataset with multiple measurements of one batch was analyzed. This can show the variance in one batch and the compared variance of one retention level, material or preservative. With the gained information, the goal is to make recommendations how to improve the process and how to get the process statistical under control. New operation targets should be created as well to minimize the values under the given retention level by the Standard.

Research Hypotheses

This research should show the data quality of the lumber treating plants and the third-party agency's which measure the traded lumber. The hypotheses in this study are:

- Because preservative retention in wood varies within and between boards, subsequent sampling of a charge may produce higher or lower retentions than the initial measurement.
- As a result, charges which have been found acceptable at the treating plant are sometimes reported to be inadequately treated during a subsequent inspection.
- This has caused some concern with the wood treating industry, and there is interest in better understanding how much variability can be expected when a charge is measured multiple times.

Objectives

The long-term objectives of this project is the determination of following values based on statistical process control on industrial and third party data from the retention values of pressure treated residential lumber, with the goal of recommendations to minimize the variation, a statistical control of the process and a possible tool to predict the process. Primary goal of this study is to analyze the data quality, to characterize the variability expected in retention values when treated wood charges are measured multiple times and to determine operational target levels given the observed variability.

Report Organization

The literature review in the main part describes traded lumber in the past years. After this part the pressure treatment is presented. The way of measurement and the statistical analysis, involving distribution plots, summaries of statistics, capability analysis, Saphiro W-test, life data analysis, control charts and boxplots, are described in the Materials and Method part of this work. The Results and Discussion Part describes the gained insights from the research. Conclusions are given in the last chapter with a discussion on future direction.

Main Part

Materials and Method

Pressure Treatment

The pressure treatment occurs in six main steps as showed in Figure 1. A, the untreated wood is placed in the cylinder. B, a vacuum is applied to pull air out of the wood. C, the wood is immersed in solution while still under vacuum. D, pressure is applied to force the preservative into the wood. E, the preservative is pumped out and a final vacuum is pulled to remove the excess preservative. F, excess preservative is pumped away and the wood is removed from the cylinder. (Lebow 2010)

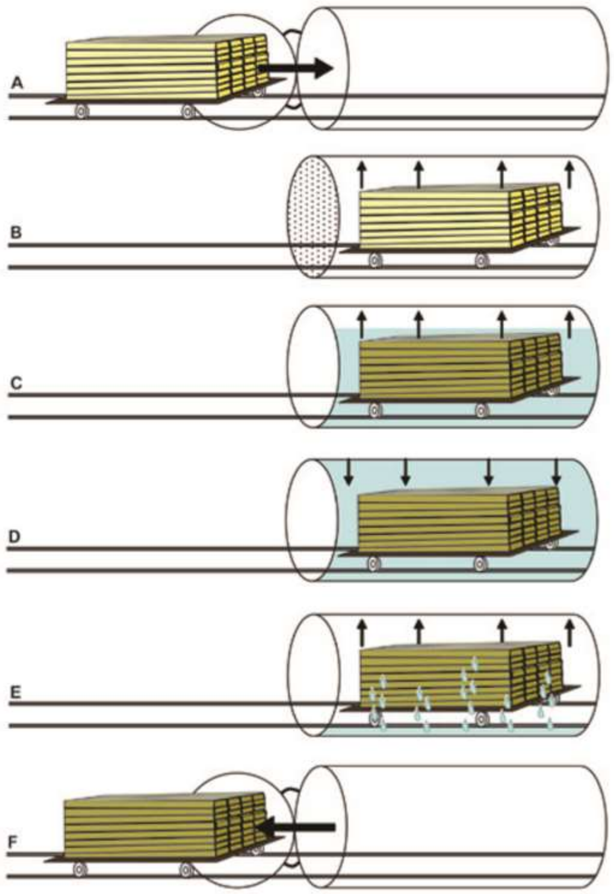


Figure 1: Schema of the pressure treatment process

Sampling of Traded Lumber in the industry according to AWPA Standard

This study investigates the sampling process of wood preserved lumber manufactured at the Langdale Forest Products Company of Sweetwater, Tennessee. In general, two major products are being produced at the site, such as micronized copper azole treated residential lumber (MCA) and copper chrome arsenate treated poles (CCA). The sampling procedure is performed according the AWPA standard M2 20.

In the following, a brief description of the sampling process of Southern Yellow pine is presented. After treatment, the treated lumber was removed from the treatment cylinder (Figure 2 and Figure 3). According to the AWPA standard, boards were randomly chosen and core samples were taken with the minimum length of 2.48 inch (63mm) (AWPA T1 12) (Figure 4 and Figure 5). A borer with an inner diameter of 0.2 inches (0.508 mm) was applied to extract core samples (Figure 7) out from the middle section of the board (AWPA M2-4.2.1). The core samples were placed in a holder seen in Figure 6. After sampling, the test holes of the boards were closed with similar treated wood plugs (AWPA M2- 4.2.3).

Figure 2: Treated lumber coming out of the pressure cylinder

Figure 3: Cylinder for the pressure treatment

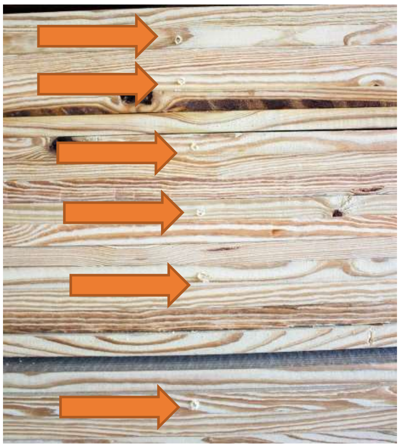


Figure 4: Randomly taken core borings from the treated lumber

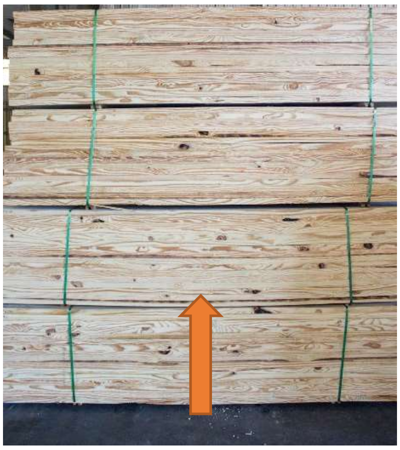


Figure 5: Sample borings locations

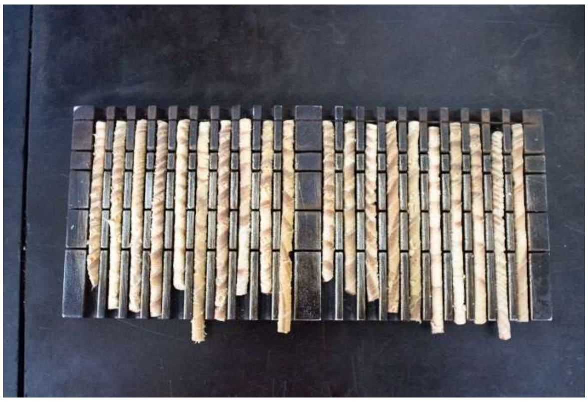


Figure 6: 20 Core samples taken from treated lumber

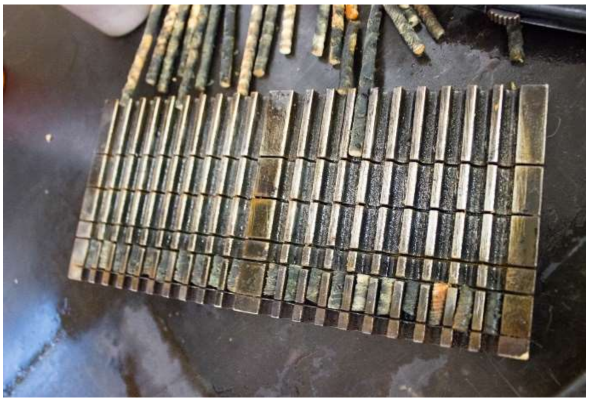
Figure 7: Drill to make core sample borings

To identify the amount of heartwood, an indicator was sprayed on the samples, either a O-anisidine hydrochloride or a 10 percent sodium nitride solution (AWPA M2 - 4.3.1.1). The heartwood turns red when applying this indicator. The red colored parts get cut of the core samples (Figure 8). As a next step, an indicator for copper was sprayed on the core samples. As an indicator a mixture of *chrome azurol S* and *sodium acetate* was used (Figure 9)(AWPA A3- 2.).

Figure 8: Removing the red parts indicating the heartwood on the core samples

Figure 9: Spraying the indicator for copper on the core samples

The indicator turns the treated parts of the core samples into a dark color. According to the color, this test shows cores, which were not penetrated sufficiently by the preservative. The cores with insufficient indication are considered as deficient products. According to the standard commodity specification (AWPA T1 -section), at least 85% of the entire samples are required to pass the test. The cores were cut to a length of 0.6 inches (15.24mm) (Figure 10) (AWPA T1 11). The samples were transferred into a microwave oven and were dried down to a moisture content of 0 % (AWPA A9, 5.1.5 / 8.1.1.1).



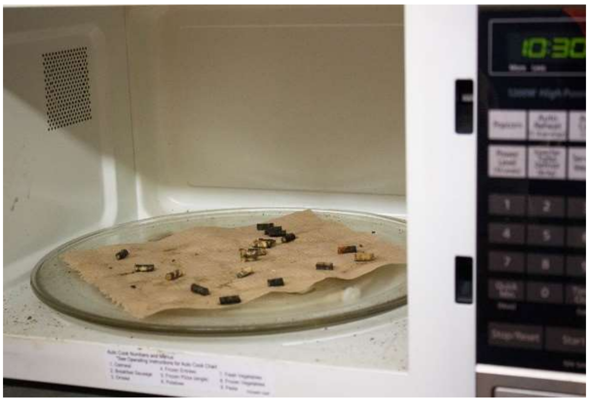


Figure 10: Cutting the samples on a size of 0.6 inches

Figure 11: Drying the core samples in a microwave

When the samples have been dried (Figure 12) the core samples were ground in a grinder to a size of 20 mesh (0.0331 inch /0.841 mm) (A9, 8.1.1.2) (Figure 13).

Figure 12: Dried core samples

Figure 13: Grinding the core samples for the X-ray measurement

Afterwards, ground core material was placed in a sample cup (Figure 14) and was measured in an X-ray machine (Figure 15). It was required to label each sample with appropriate identification. The concentration (%) was evaluated in lb/ft^3 (pounds per cubic foot, pcf) according to the following equations: Equation 1, Equation 2, Equation 3, and Equation 4. Those equations are based on standard AWPA A9, 12.2.

Equation 1: Compressed wood method (AWPA A9 12.2.2)

$$(\%\ of\ preservative\ in\ Sample)x\ \frac{(A12\ Density\ of\ wood)}{100} = Total\ pcf$$

Equation 2: Compressed wood method (AWPA A9 12.2.3)

$$(\%\ of\ preservative\ in\ Sample)x\ \frac{Sample\ Density}{100} = Total\ pcf$$

Equation 3: Sample density (AWPA A9 12.2.3.1)

$$Sample\ Density = W\ x\ F$$

W Sample weight

F Conversion factor from F6, (Table 3)

Equation 4: Sample density (AWPA A9 12.2.3.2)

$$Sample\ Density = \frac{W\ (1.212)}{(D/2)2\ xL\ x\ N}$$

W Sample Weight

D Bit Diameter

L Core Length

N Number of Cores

Figure 14: Compressing the ground core samples into the measurement cup for X-ray measurement

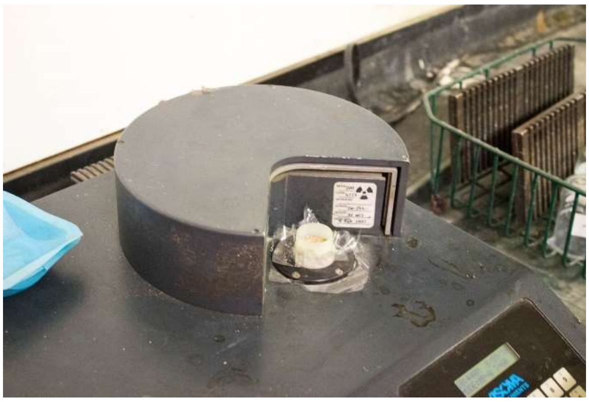
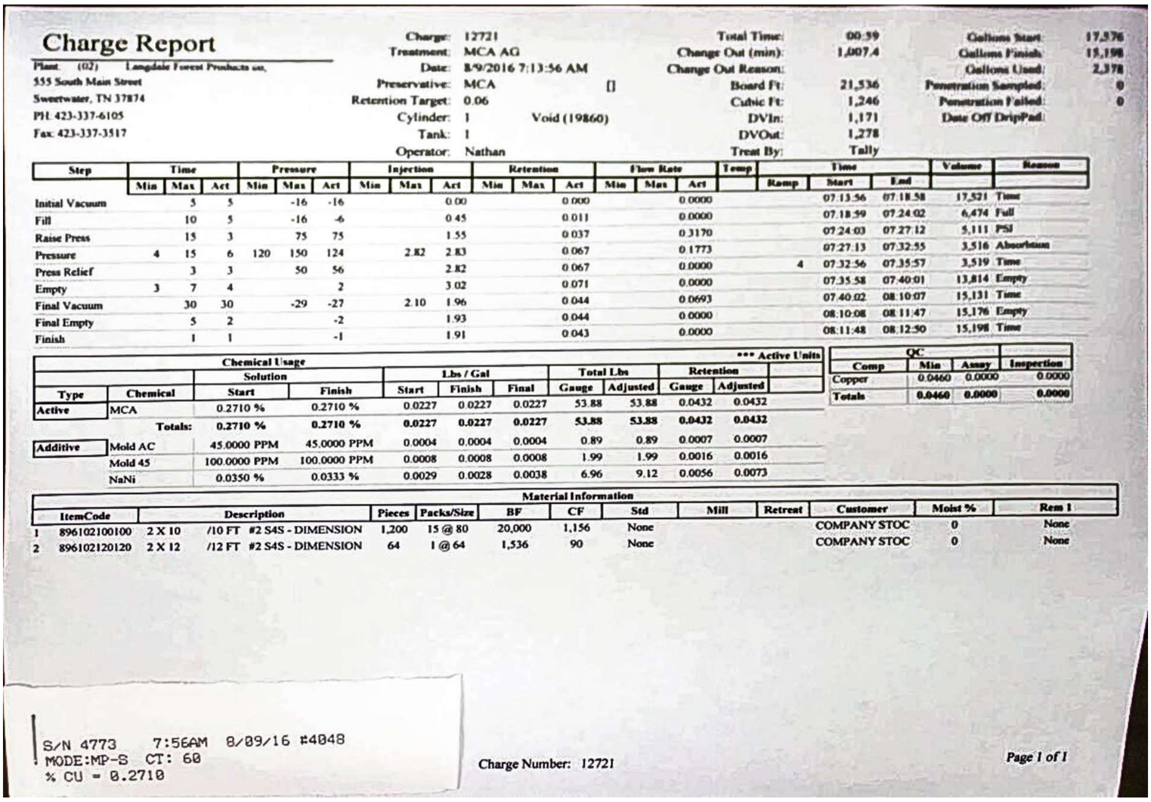


Figure 15: Measuring the core sample in the X-ray device

Levels of the preservative were compared with the retention values from the AWPA standards U1, section 6, commodity specification A and for MCA in the ESR-1721 (Table 1). For every charge the retention and penetration values must be reported. The Figure 16 is an example of a charge report of the Langdale Forest Products Company of Sweetwater, Tennessee. From every charge the retention and penetration values have to be reported. When the sample doesn't pass the retention test it's possible to retest the batch 3 times, when it still fails the batch has to be retreated. This can occur up to 3 times, according to the treating mill of Langdale.



Charge Report

Plant: (02) Langdale Forest Products co,
555 South Main Street
Sweetwater, TN 37874
PH: 423-337-6105
Fax: 423-337-3517

Charge: 12721
Treatment: MCA AG
Date: 8/9/2016 7:13:56 AM
Preservative: MCA []
Retention Target: 0.06
Cylinder: 1 Void (19860)
Tank: 1
Operator: Nathan

Total Time: 00:59
Change Out (min): 1,007.4
Change Out Reason:
Board Ft: 21,536
Cubic Ft: 1,246
DVIn: 1,171
DVOut: 1,278
Treat By: Tally

Gallons Start: 17,576
Gallons Finish: 15,198
Gallons Used: 2,378
Penetration Sampled: 0
Penetration Failed: 0
Date Off DripPad:

Step	Time			Pressure			Injection			Retention			Flow Rate			Temp	Time			Volume	Reason
	Min	Max	Act	Min	Max	Act	Min	Max	Act	Min	Max	Act	Min	Max	Act		Ramp	Start	End		
Initial Vacuum		5	5		-16	-16			0.00			0.000			0.0000			07:13:56	07:18:58	17,521	Time
Fill		10	5		-16	-6			0.45			0.011			0.0000			07:18:59	07:24:02	6,474	Full
Raise Press		15	3		75	75			1.55			0.037			0.3170			07:24:03	07:27:12	5,111	PSI
Pressure	4	15	6	120	150	124		2.82	2.83			0.067			0.1773			07:27:13	07:32:55	3,516	Absorbtion
Press Relief		3	3		50	56			2.82			0.067			0.0000		4	07:32:56	07:35:57	3,519	Time
Empty	3	7	4			2			3.02			0.071			0.0000			07:35:58	07:40:01	13,814	Empty
Final Vacuum		30	30		-29	-27		2.10	1.96			0.044			0.0693			07:40:02	08:10:07	15,131	Time
Final Empty		5	2			-2			1.93			0.044			0.0000			08:10:08	08:11:47	15,176	Empty
Finish		1	1			-1			1.91			0.043			0.0000			08:11:48	08:12:50	15,198	Time

Chemical Usage *** Active Units

Type	Chemical	Solution Start	Solution Finish	Lbs / Gal Start	Lbs / Gal Finish	Lbs / Gal Final	Total Lbs Gauge	Total Lbs Adjusted	Retention Gauge	Retention Adjusted
Active	MCA	0.2710 %	0.2710 %	0.0227	0.0227	0.0227	53.88	53.88	0.0432	0.0432
	Totals:	0.2710 %	0.2710 %	0.0227	0.0227	0.0227	53.88	53.88	0.0432	0.0432
Additive	Mold AC	45.0000 PPM	45.0000 PPM	0.0004	0.0004	0.0004	0.89	0.89	0.0007	0.0007
	Mold 45	100.0000 PPM	100.0000 PPM	0.0008	0.0008	0.0008	1.99	1.99	0.0016	0.0016
	NaNi	0.0350 %	0.0333 %	0.0029	0.0028	0.0038	6.96	9.12	0.0056	0.0073

QC

Comp	Min	Assay	Inspection
Copper	0.0460	0.0000	0.0000
Totals	0.0460	0.0000	0.0000

Material Information

	ItemCode	Description	Pieces	Packs/Size	BF	CF	Std	Mill	Retreat	Customer	Moist %	Rem 1
1	896102100100	2 X 10 /10 FT #2 S4S - DIMENSION	1,200	15 @ 80	20,000	1,156	None			COMPANY STOC	0	None
2	896102120120	2 X 12 /12 FT #2 S4S - DIMENSION	64	1 @ 64	1,536	90	None			COMPANY STOC	0	None

S/N 4773 7:56AM 8/09/16 #4048
MODE:MP-S CT: 60
% CU = 0.2710

Charge Number: 12721

Page 1 of 1

Figure 16: Charge Report of MCA traded lumber of Langdale forest products

Table 1:Minimum preservative retention requirements (ICC_ES Report – ESR 1721)

End Use	Minimum Actives Retention lb/ft^3 (kg/m^3)			
	CA-B	CA-C	µCA-B (MCA-B)	µCA-C (MCA-C)
Above ground – general use	0.10 (1.7)	0.06 (1.0)	0.06 (1.0)	0.05 (0.8)
Above ground – decking & specialties use				
- Species listed in Section 3.3 (primarily sapwood)	0.08 (1.4)	0.06 (1.0)	0.06 (1.0)	0.05 (0.8)
- Species listed in Section 3.3 (primarily heartwood)	0.21 (3.3)	0.15 (2.4)	0.15 (2.4)	0.14 (2.2)
Ground contact – general use	0.21 (3.3)	0.15 (2.4)	0.15 (2.4)	0.14 (2.2)
Ground contact – heavy duty	0.31 (5.0)	0.25 (4.0)	0.23 (3.7)	0.23 (3.7)
Ground contact – wood fountain systems	0.31 (5.0)	0.25 (4.0)	0.23 (3.7)	0.23 (3.7)
Ground contact – extreme duty	0.41 (6.6)	0.35 (5.7)	0.33 (5.3)	0.33 (5.3)

(ICC Evaluation Service 2016)

Statistical Analysis

For the analysis of the data in this project the computer program JMP Pro 11. was used. The first approach of analysis of each dataset was a histogram, which has the possible values on one axis and frequencies for those values on the other axis (Figure 17). As a next step the Normal Quantile Plot also termed GG-Plot was created. As seen in Figure 18 all data points are plotted on the X-axis with the cumulated probability and on the Y-axis with their value. The red line shows the cumulated normal distribution as a line.

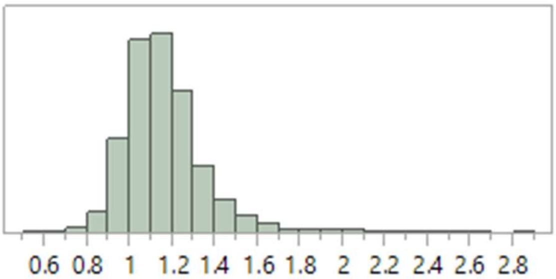


Figure 17: Distribution Plot

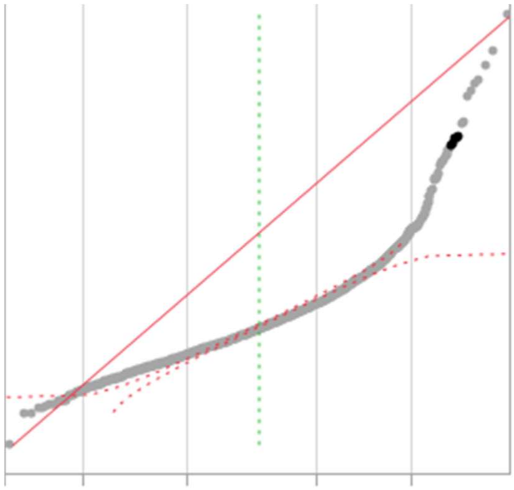


Figure 18: Normal Quantile Plot

As a following step, the summary of statistic was added including the mean, standard deviation, n, variance, CV, min, max, median, mode and range. Those values are explained in Table 2.

Table 2: Summary of Statistic

Value	Explanation	Formula
Mean / Average	The average identifies the center of the mass for the values in the dataset	$\bar{X} = \frac{\sum_{i-1}^{n} x_i}{n}$
Standard deviation	a quantity calculated to indicate the extent of deviation for a group as a whole.	$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$
Variance	Is the squared deviation of a random variable from its mean.	$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$
N	Number of Samples	

CV / Coefficient of Variation	Is a standardized measure of dispersion of a probability distribution or frequency distribution.	$CV[\%] = \frac{s}{\bar{x}} x100 = \frac{\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}}{\frac{\sum_{i-1}^{n} x_i}{n}} x100$
Min	Smallest value of the dataset	
Max	Biggest value of the dataset	
Median	The sample median of a “set of n measurements” is the middle value (“50th percentile”) when the measurements are arranged from smallest to largest.	
Mode	The mode of the Dataset is the Value that occurs most often.	
Range	The range is defined as the Maximum minus the Minimum	$R = \min - max$

(Wheeler, Chambers 2010)

Furthermore, a capability analysis was applied to the dataset. For analysis, a “Lower Specification Limit” (LSL) a Target and a “Upper Specification Level” (USL) must be defined. In this case the LSL is the minimum retention value, which is given in the standard. The target is 1.5 times the LSL and the USL is 2 times the LSL. This data shows the amount which is in and out specification and the amount of data which is above and below the target. The capability analysis gives us the actual data of this dataset and the probability according to a distribution function.

To find out how the data is distributed a Saphiro Wilk-Test was applied. This test proves if the dataset is normal distributed or not. A value under 0.5 is probably not normal distributed a value over 0.5 is probably normal distributed. The Saphiro Wilk-Test is calculated according to Equation 5.

Equation 5: Saphiro W-Test

$$W = \frac{(\sum_{t=2}^{n} a_t y_t)^2}{\sum_{t=1}^{n}(x_t - \bar{y})^2}$$

n Number of Values

y_t Values from parent sample

a_t Tabulated coefficients

H_0 Sample comes from a normal distribution

(Lohninger 2012)

When the dataset was not normal distributed, a lifecycle data analysis was applied. This test compares different distributions (like Lognormal, Weibull, Loglogistic, Frechet, Normal, Logistic and Exponential distribution) and fits them to the dataset. The distribution with the smallest AIC (Akaike's Information Criterion) and BIC (Bayesian Information Criterion) has the highest probability to describe the data distribution. The AIC and BIC are calculated according to Equation 6, Equation 7 and Equation 8.

$$AIC = -2LogLikelihood + 2k + 2k\frac{k+1}{n-k-1}$$

Equation 6: AIC

$$BIC = -2LogLikelihood + k\ln(n)$$

Equation 7: BIC

$$\mathrm{L}(\theta; \mathrm{x}) = \prod_{i=1}^{n} f\ (\mathrm{X}_i; \theta)$$

Equation 8: Likelihood

L	likelihood
k	number of estimated parameters in the model
n	number of observations in the dataset

(JMP 2016) (JMP 2016)

Control Chart (X-Individual, Moving Range)

To prove the data quality, a X-individual and a moving range chart is used. The x-individual chart shows the time ordered data on the x-axis and the value on the y- axis. The LCL was calculated according to Equation 9 and the UCL to Equation 10. They are plus and minus three standard deviations (3σ) away from the dataset average and contain 99.7% of the data. Those are calculated with an unbiased estimator of σ, R/d^2.

$$UCL = \bar{X} + 2.66\ x\ (\overline{mR})$$

Equation 9: Upper Control Limit

$$LCL = \bar{X} - 2.66\ x\ (\overline{mR})$$

Equation 10: Lower Control Limit

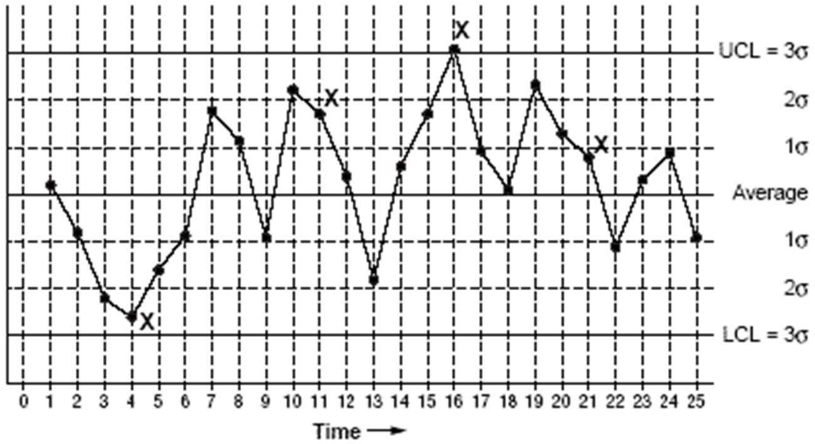


Figure 19: Example for a X-individual chart

(Tague 2005)

A second useful tool for analysis of data is termed moving range chart. This chart shows the moving range between one value and the previous value of time ordered data. In general, for this chart an UCL_{mR} and an average moving range are being calculated. These values contain 99.7% of the data.

$$\overline{mR} = \text{"}average\ moving\ range\text{"}$$

Equation 11: Average moving range

$$UCL_{mR} = 3.268\ x\ (\overline{mR})$$

Equation 12: Upper control limit of moving range

To compare sets of data pairs the data can be displayed in a boxplot. The boxplot shows the median, the 25th and 75th percentile and the whiskers are 1.5 times the interquartile range to show potential outliers.

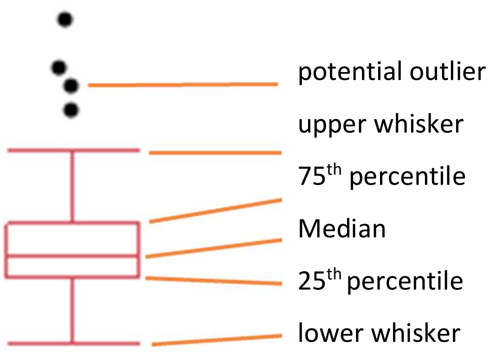


Figure 20: Example for a boxplot

(Young 2016)

Results and Discussion

The results of the analysis, which include distribution, summary of statistics and comparisons between datasets are presented in this chapter. In the first section of the chapter the data quality is evaluated. In the second part the large variance problem is described and applied to this data. The next part describes the statistical difference between the different use classes. The further part looks at the distribution of the retention values. Last the opportunity of multiple measurements of one batch was analyzed.

Data Pairs of the treatment plant and the third-party agency

First the dataset of 4066 data points of the treatment plant was analyzed. All the Data points a multiplied by a factor to set the target to one. For example, the minimum target of micronized copper azole (MCA) is 0.06 lb/ft^3 (pounds per cubic foot) by the use class 2 (UC2 for no water contact). This value is multiplied by 16.66 to gain the value one. Through this multiplication every use class and preservative can be compared easily. In this dataset, the average has the value 1.18, the median at 1.16 and the mode by 1.15. All three values are close to each other, but a slight skewness to the right can be seen. The standard deviation has a value of 0.20 and the variance is 0.04. The coefficient of variation is 17.15%. The minimum value of the treatment plant is 0.11 and the maximum retention measured is 4.54. The range of the dataset is 4.54. Those values show a high spread of the data. A value of 2 means that the double amount of the preservative is in the lumber. In this case a Range of 4.54 means that in this dataset values from 0.11, far under the given level, and 4.65 far over the given level. In both cases this is a bad value. A value under 1 is a mistreated charge and should be retreated. A value fare over 1 means that too much preservative was pumped into the wood. In Figure 21 the distribution is plotted. On the y-axis the retention values are given and on the x-axis the frequency is shown. It is easy to see the spreading of the data. Also seen is the skewness to the right and the fact that hardly any data points are under one. Calculated are 5.6% under the given level in the standard. With a confidence interval of 95% this value should be 2.5% to be statistical under control. This data is very close to the value of 2.5%. The boxplot in Figure 21 shows a lot of potential outliers in this dataset. These can be special events, cause by a defect in the machine or process, or the process and the equipment were correct and the material was hard to treat. In all cases there are too many points out of control. Statistically, there should be in this dataset with a confidence interval of 95% 101 points over and under the boxplot to be under control and predictable. In this case this is not given. The third graph in Figure 21 is the normal quantile plot. The red line shows the line where the data should be to have a normal distribution. Again, there can be seen the skewness to the right, by the data points under the red line. Another strange area in this set are the points under one, there is a hole of the data and the values go faster to the value of zero. This can be an effect of the retesting when the data is under the level of one. This is called the large variance problem and is described in the next chapter.

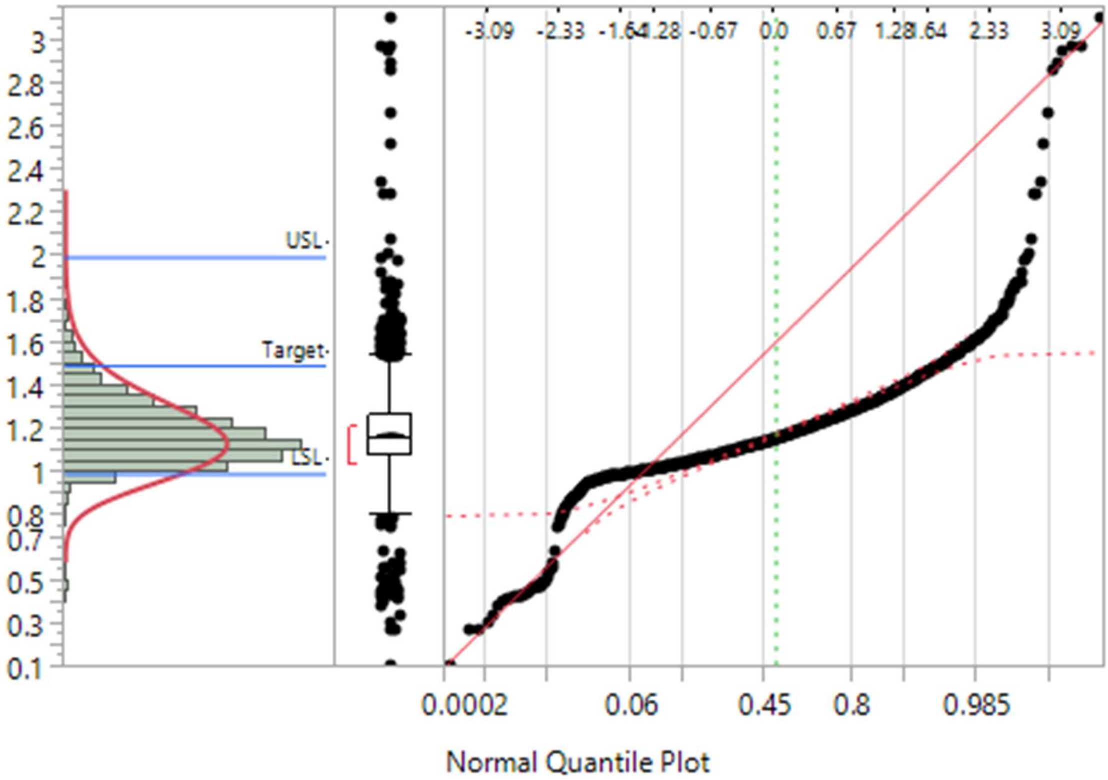


Figure 21: Distribution, Box Plot and Normal Quantile Plot of the Data of the treatment plant

The second dataset is the pair to the measurements of the treating mill. The third-party agency has the possibility to measure the quality of the lumber when they come to the plant. Also this dataset of 4066 data points was analyzed. In this dataset the values are again multiplicated with a factor to have one as minimum retention value. The mean is at 1.16, the median at 1.14 and the mode at 1.03. These are close to each other and close to the minimum given in the AWPA Standard. The mode in this case is lower than the mean and the median, but still close enough to the other values. The standard deviation is 0.18 with a variance of 0.03. The coefficient of variation is 15.77%. The minimum value of the third-party agency is 0.53 and the maximum is 2.81. The range of this dataset is 2.28. This is a good spread of the data and means that some values are under the value of 1 and some far over. But the maximum of overtreatment in this case is 2.81 times as high as written is the Standard. The minimum is 0.53 times the value and the mean with 1.16 is slightly over the value of the standard. In Figure 22 the distribution of the dataset is plotted. The retention values are on the y-axis and the frequency on the x-axis. Here is seen a slight skewness to the right. The retention values follow the red curve, which shows the log normal distribution (explained in the chapter "Distribution"). The values under one are to less treated. In this case, there are

15.1% of the charges too less treated according to the Standard. Compared with a confidence interval of 95% 101 points can be under 1, in this case there are 614 measurements to low. This means the target of the treatment plant is set too low and the understanding of variation should be forced. In the middle of Figure 22 the boxplot of the third-party agency data is plotted. This shows a lot of outlays to the top and a few events at the lower end. They could be outliers or event, but in this case, they look like a part of the log normal distribution. In the right part of Figure 22 the normal quantile plot is seen. Also this dataset is skewed to the right, but fits better on the red line, which describes the log normal distribution.

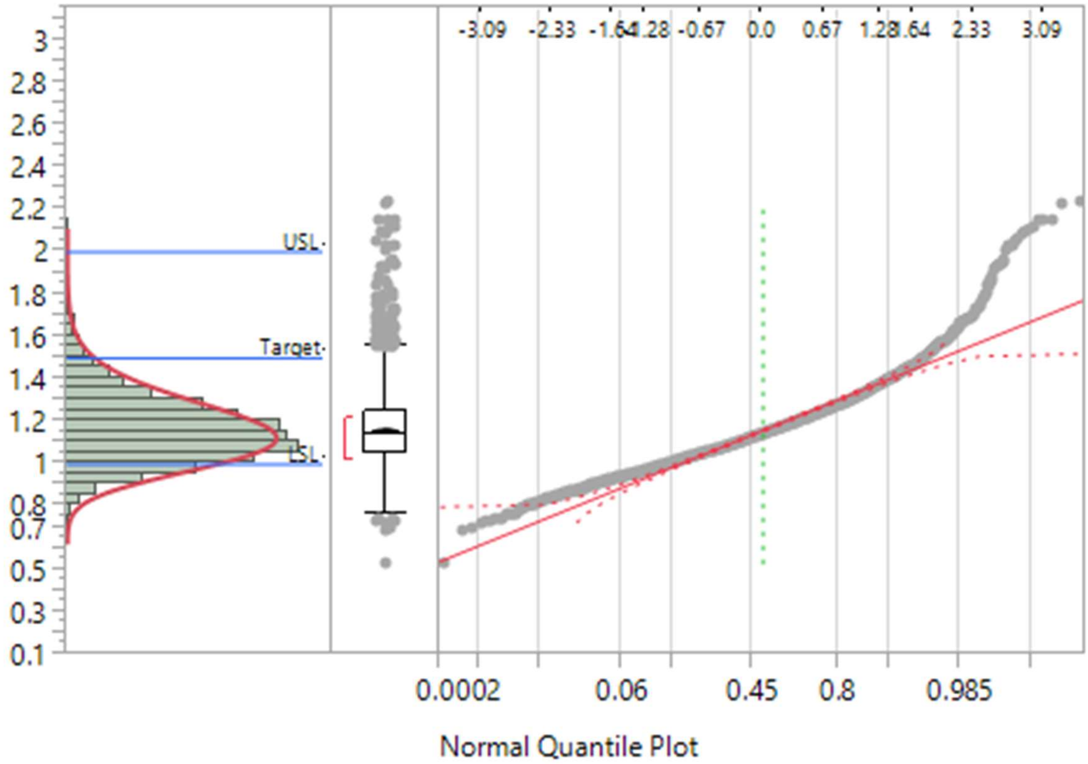


Figure 22: Distribution, Boxplot and Normal Quantile Plot of the third-party agency

When the retention values of the treating mill and the retention values of the third-party agency get compared (Table 3) the mean and the median are 0.02 lower from the third-party agency than of the treating plant. The standard deviation is 0.02 lower at the third-party agency, as well the CV is 1.38% lower. Significant is the difference in the range. The agency has a 2.26 lower range than the treating plant. A reason for this phenomenon can be the different handling of the core samples. The agency measures with the more accurate method than the treating mill, but both are allowed in the Standard. Another value with a huge difference is the percentage of data points that are under the minimum level in the standard. The treating mill has 5.6% and the agency measures 15.1% out of specification. The reason for the huge difference is the fact that the treating mill can measure a batch again, when the value is under the given value in the Standard. The opposite does the agency. They are not allowed to retest a batch.

Table 3: Compared Statistical summary of the retention values of the treating plant and the third-party agency

	Treating plant	Third-party agency
N	4066	4066
Mean	1.18	1.16
Median	1.16	1.14
Mode	1.15	1.03
Min	0.11	0.53
Max	4.65	2.81
Range	4.54	2.28
Standard deviation	0.20	0.18
Variance	0.04	0.03
CV	17.15%	15.77%

Large Variance Problem

In the chapter before, the large variance problem was mentioned. First it is important to understand that every system has a natural variation. That means one value that occurs most frequent and some values higher and lower that occur less often the farer they are away from the middle value. This is in general shown as a Gaussian curve, seen in Figure 23 as the orange line. Every process has a target value they want to reach and a lower and / or higher specification level, seen in Figure 23 as the green vertical lines. When the process has a high variation the orange curve is stretched to the left and right, when the variation is low, the curve is compressed and rises in the middle. The large variance problem occurs when the curve is flat and in this case the lower end of the curve is under the lower specification limit. When the first measurement fails, the treating plant remeasures the batch and easily can get over the lower speciation limit. The reason for that is, that the probability under the orange curve sums up to one. In a Gaussian curve the highest point is the gravity point of the curve and 50% of the values will be below this value und 50% will be above this value. The LSL (lower specification level) has already a lower value than the center point of the curve and the probability to get two measurements lower than the LSL is lower, than getting the first value under and the second value over the LSL. One method to find reasons for variation is called root cause analysis, seen left in Figure 23. Variation has several reasons. The important reasons are machines, material, methods, measurements and the operating people. The root cause analysis looks at one of those points and askes why there is variation and repeats that for 5 times. This process helps to find hidden variation which is mostly hard to find.

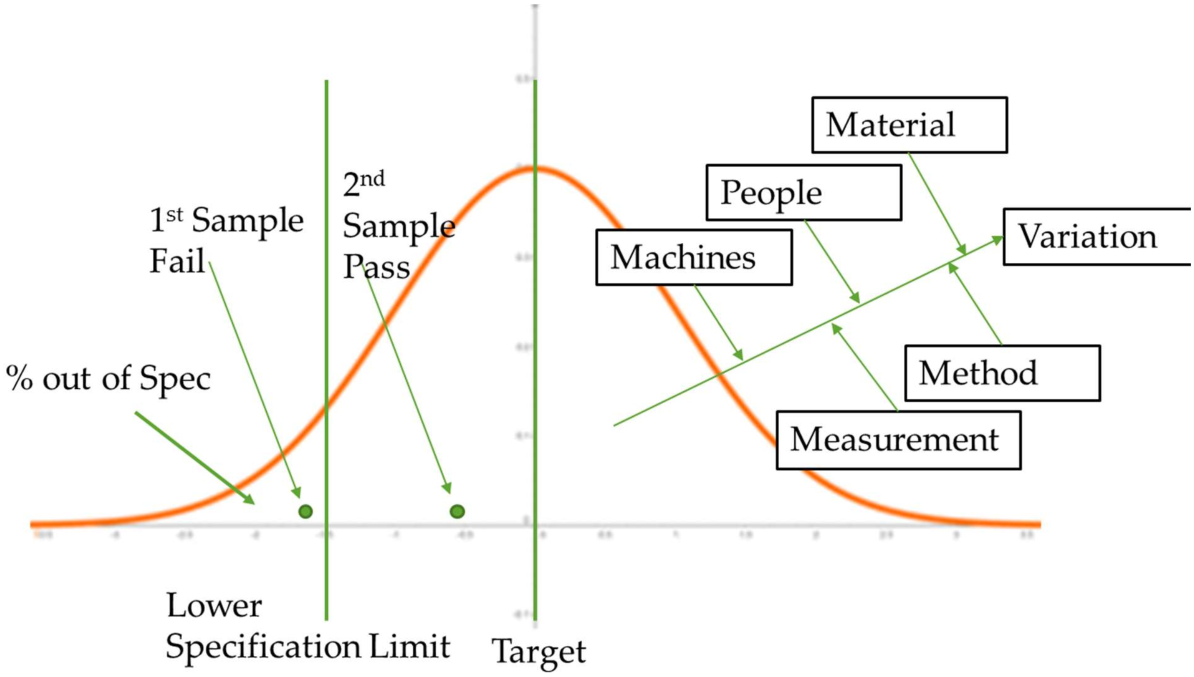


Figure 23: Large variance problem and root cause analysis

Shift target

One opportunity to solve the large variance problem is to shift the target. The Gaussian curve gets shifted to that level that the area under the LCL contains less than 5%. This corresponds to a confidence interval of 95% and means that 5% of all measurements will be under the LCL when measured infinite times. The problem with this method is that the process doesn't change and the variability stays the same. Mostly this is the more expansive method. In this case, more preservative must be pumped into the wood and this causes more costs. This is shown in Figure 24 from the green to the red curve.

Reduce variation

The second opportunity to solve the large variance problem is to reduce variation. This can happen through a root cause analysis. Better equipment, better instructed operators, a more accurate measurement method, or better graded material are some of the solutions to reduce the variation. Mostly this method costs more at the beginning, but it will save money in long term than shifting the target. In this case the target hasn't to be shifted, just through the reduction of the variation the Gaussian curve will have less percentage under the LSL. This can be seen in Figure 24 from the green to the orange line.

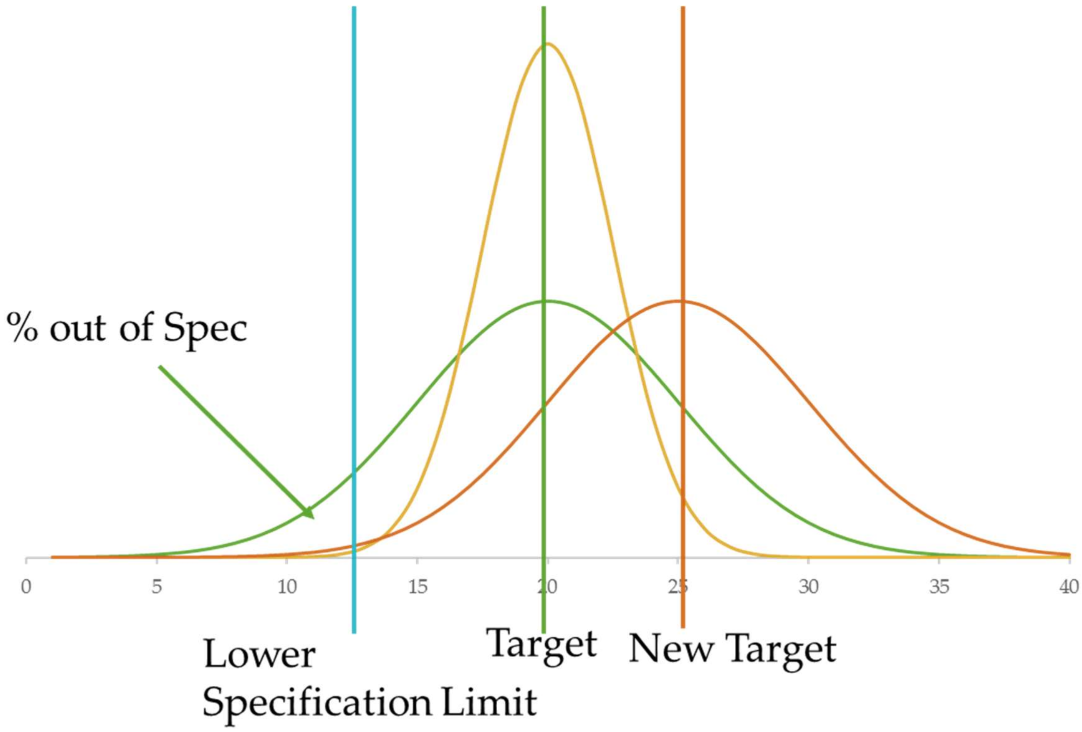


Figure 24: Solution fort the large variance problem, shifting target or reducing variation

Control Chart (x-individual, moving range)

A tool to visualize the data quality the the x-individual and moving range chart. The x-individual is the upper graph on Figure 25. The dataset includes 341 data points, with a mean, median and mode of 0.07. The standard deviation is low with 0.01 but it is a high CV with 14.11%. The range is 0.06, this can be seen from the highest to the lowest data point. The green line illustrates the mean. The upper red line explains the UCL (upper control limit) calculated in Equation 9 and the lower red line shows the LCL (lower control limit) and is calculated in Equation 10. The probability that data points are outside of the red line is 0.3%, in this case 1.02 points. On this chart there are 5 points out of control, but in this case it is not important because all data points are outside the upper control limit. In this case the orange line at 0.06 lb/ft^3 is the lower specification limit, which shows the minimum retention value given by the AWPA Standard. The dataset has 62 points (17.88%) under the lower specification limit. The chart shows immediately the data quality and trends in the process. This process data is under statistical control and predictable between the LCL and the UCL, but the process is not in the given specification limit. The lower chart in Figure 25 is the moving range chart. The upper red line is the UCL and calculated in Equation 12 and is under control, because there are hardly any outliers and it has a continuous look.

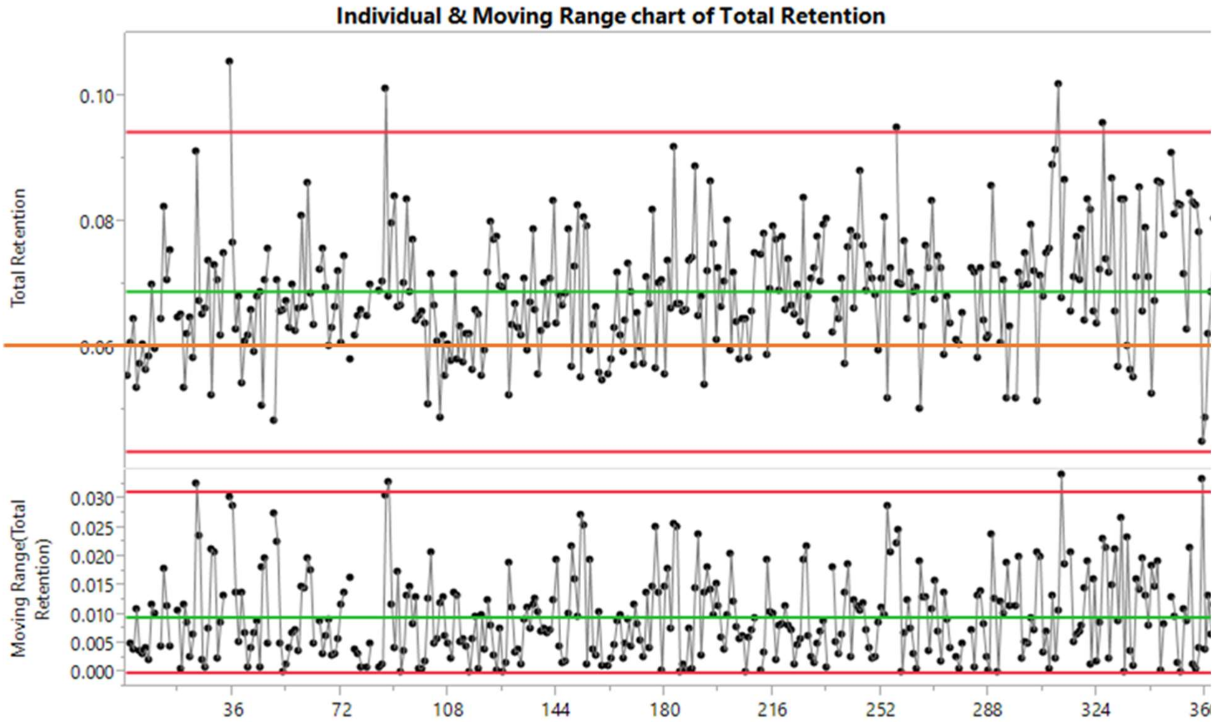


Figure 25: X-individual and moving range chart

The second x-individual and moving range charts show the difference between the retention values of the different use classes. In the case of MCA (micronized copper azole) the minimum retention values are for use class 2 0.06 lb/ft^3, use class 3B 0.06 lb/ft^3, use class 4A 0.15 lb/ft^3 and use class 4B 0,23 lb/ft^3, those are written in Table 1. This dataset has 416 data points. The mean is 1.25, the median is 1.20 and the mode is 1.14. The standard deviation is 0.21 with a variance of 0.04 and a coefficient of variance is 17.09%. The range is 1.58. Same as in the previous dataset the red lines are the LCL and UCL and the green line represents the mean. The blue line at 1.00 is the LSL. This dataset comes from a treating plant. This can be seen at the fact that hardly any data points are lower than the LSL, caused by the retesting of the data points under the LSL. This dataset shows also that different materials or preservative levels have a different natural variation, seen in Figure 26. The moving range has some data point which are out of control, this can be caused by a shift change, another operator or a special event. For sure this is a point where a root cause analysis can help to decline the variation.

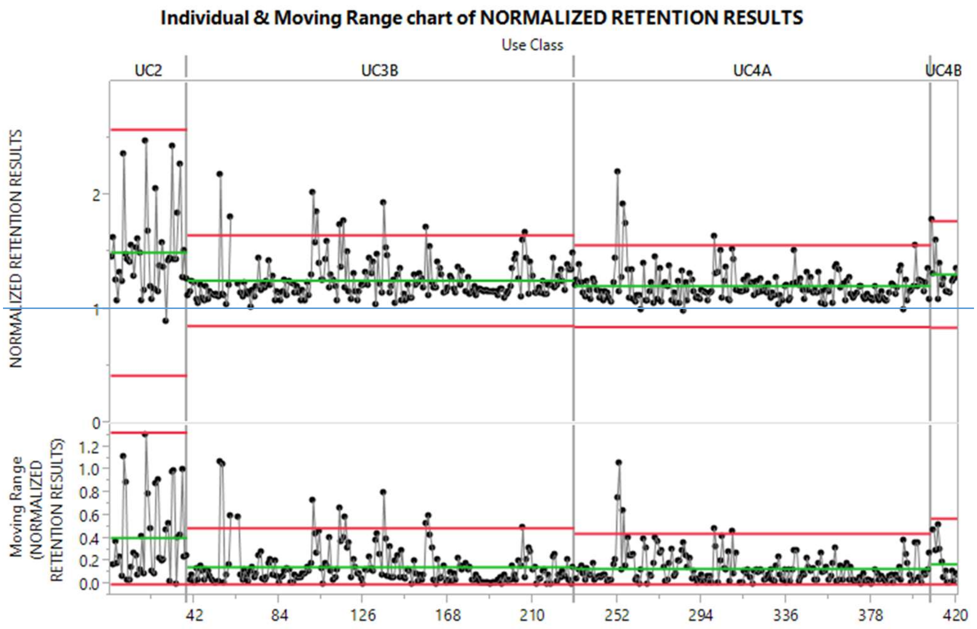


Figure 26: Moving range - and x-individual chart

Boxplot of Data Pairs

Another great tool to visualize the data are boxplots, those are described in Figure 20. In this case the dataset is with the data pairs of the retention values from the treating mill and the third-party agency. Again, the dataset with 4066 data points is used, seen in Figure 27. The y-axis shows the retention values. The data is split into the treatment and third-party agency. The treating mill data is in blue and the third-party agency is visualized in red. Also the retention values are split into the target retentions of 0.06 lb/ft^3, 0.15 lb/ft^3 and 0.23 lb/ft^3. The values were multiplied 16.66, 6.66 and 4.35 to gain the level of 1.00, through this multiplication the values get comparable. The boxplots show that the 3 different retentions have similar values of mean and range. What can be seen is, that the potentional outliers of the datasets are mostly above the upper whisker and hardly below the lower whisker. Also can be realized that the range of the treating mill data is in all three retention values higher than those of the third-party agency. Significant is that the standard deviation is always lower from the third-party agency. Of the retention value from 0.06 lb/ft^3 the difference is 0.03, at the value of 0.15 lb/ft^3 the difference is 0.01 and from the retention value of 0.23 lb/ft^3 the difference is 0.04. This is probably caused by the different handling of the core samples, which is a more accurate handling at the third-party agency.

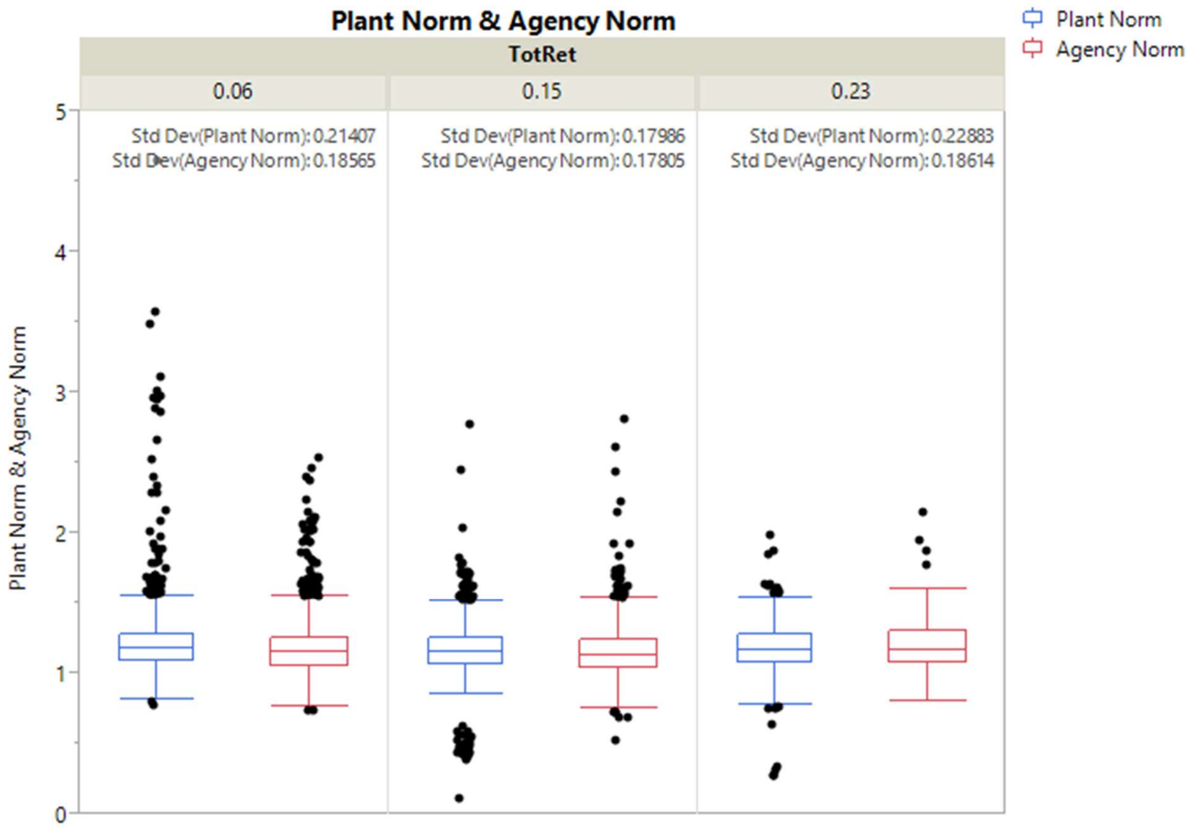


Figure 27: Boxplot of treating plant and third party agency retention values

Distribution of the Retention Values

Another information is the distribution of a dataset. For this analysis, the data pair set of retention values from the treating mill and the third-party agency are used. With the lifecycle analysis calculated in Equation 6, Equation 7 and Equation 8 the best fitting distribution was figured out. Important for this comparison is that the third-party agency data is more accurate than the treating mill data. The reason for that is the retesting of some data points in the treating mill, this will distort the distribution. The analyzed distributions where Log-logistic distribution, LEV distribution (Lévy distribution), Log-normal distribution, Logistic distribution, Fréchet distribution, Normal distribution, Weibull distribution, SEV distribution (smallest extreme value distribution) and the Exponential distribution. For the data of the third-party agency the distribution, according to Table 4, is a log-logistic distribution followed by the LEV distribution. Those 2 distributions have the lowest AICc and the lowest BIC values, which are indicators for the goodness of fit. In Table 5 the distribution of the treating mill data is shown. In this case the lowest AICc and BIC values are at the log-logistic distribution, followed by the logistic distribution. The most likely distribution and probably the distribution of the dataset is the log-logistic distribution.

Table 4: Distribution of the retention values from the third-party agency

Distribution	AICc	-2Loglokelihood	BIC
Log-logistic	-3142.85	-3146.86	-3130.23
LEV	-3000.06	-3004.07	-2987.45
Lognormal	-2913.58	-2917.59	-2900.97
Logistic	-2855.25	-2859.25	-2842.63
Fréchet	-2492.78	-2496.78	-2480.16
Normal	-2257.79	-2261.79	-2245.17
Weibull	-902.65	-906.64	-890.03
SEV	1271.93	1267.93	1284.54
Exponential	9355.13	9353.13	9361.14

Table 5: distribution of the retention values from the treating mill

Distribution	AICc	-2Loglokelihood	BIC
Log-logistic	-3203.84	-3207.84	-3191.22
Logistic	-3030.40	-3034.40	-3017.78
Lognormal	-1781.08	-1785.10	-1768.48
Normal	-1394.83	-1398.83	-1382.21
LEV	-983.05	-987.05	-970.43
Weibull	646.64	642.65	659.27
Fréchet	3383.83	3379.82	3396.45
SEV	5178.49	5174.49	5191.11
Exponential	9531.16	9529.16	9537.47

In Figure 28 an example for the Fréchet - (blue line), normal – (orange line), log normal – (grey line) and log-logistic distribution (yellow line) is given. On the y-axis is the frequency and on the x-axis is the value given. When looked at these distributions with some certainty can be said, that the analyzed dataset must be skewed to the right, which already could be seen in the comparison of the treatment mill data and the third-party agency retention values.

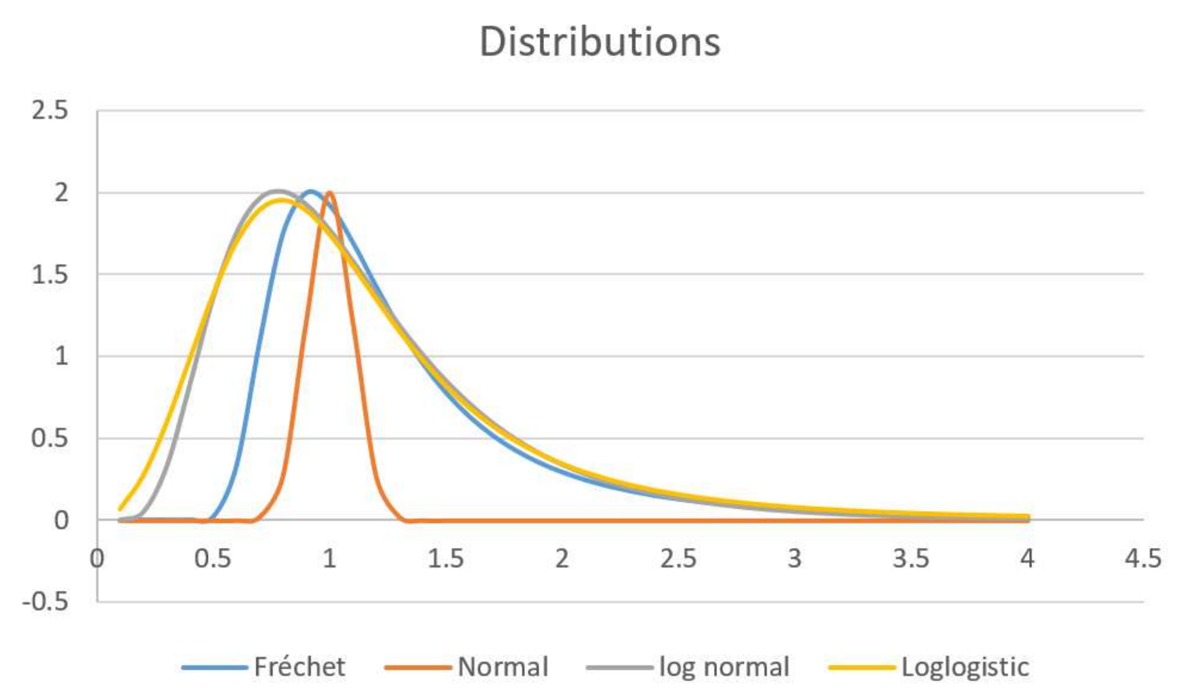


Figure 28:Example for different distributions

In Figure 29 and Figure 30 the cumulated probability on the y-axis and the retention values on the x-axis can be seen. The grey dots are the data points, the red line is the log-normal distribution, the blue on the log-logistic, the orange one the Fréchet, the turquoise one is the LEV distribution and the green one is the normal distribution. When those two Figures are compared it can be seen that the data of the agency fits better on the log-normal distribution than the dataset of the treating mill. This can be caused by the different handling or the retesting of the values underneath the LSL.

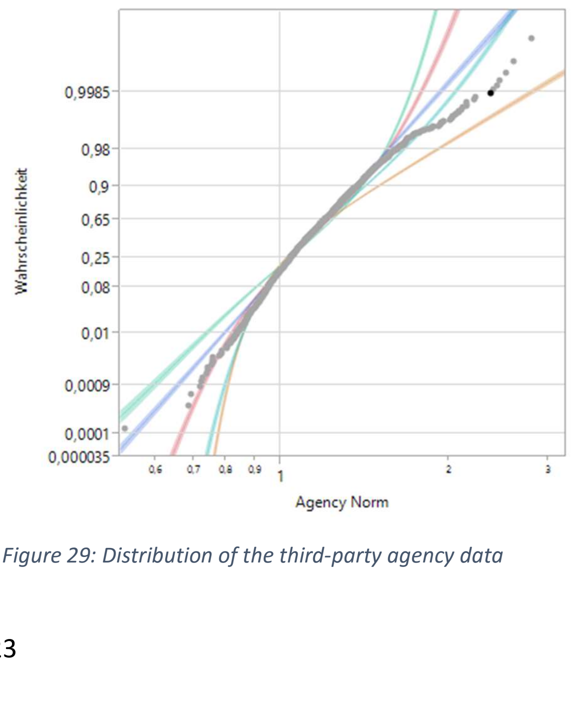


Figure 29: Distribution of the third-party agency data

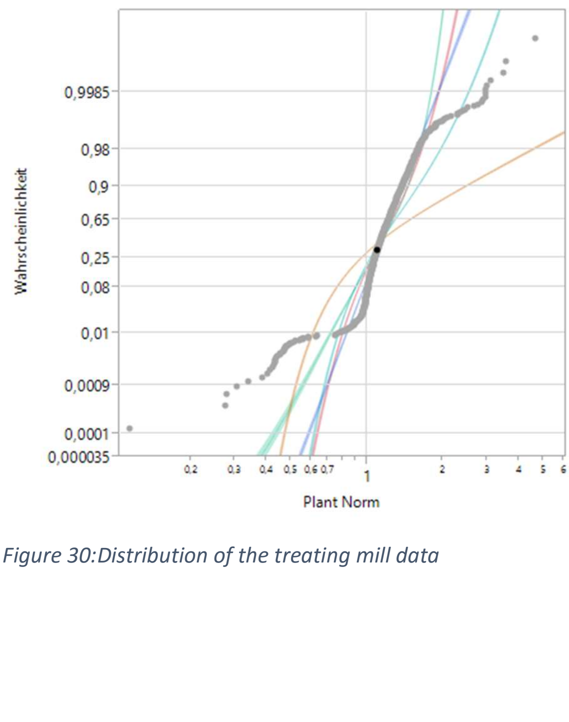


Figure 30:Distribution of the treating mill data

Variance in- and between Batches and multiple Measurements

The fact that in one batch of treatment are over 1000 boards and 20 core borings are made to one retention value of the whole batch is statistical very imprecise. In the one measurement, just one value can be seen and according to variation, probability and distribution nothing can be said from the one value. In this case this dataset measured ten batches each five times to gain more information of variation. The Samples from 01 to 05 are treated with EL-2 (eco live 2 or DCOI-Imidicloprid- Stabilizer) and the samples from 06 to 10 are treated with MCA C (micronized copper azole). In Figure 31 the retention values of the batches are visualized. On the x-axis the sample number is given and on the y-axis the retention values are given. What can be seen is that one data point alone can't say anything about the variation in the batch, here is visualized the range and the assumed spread of this batch. When on point more is measured, it can be assumed that this data point will be in this range. Through the multiple measurement the treatment gain new operating targets which are based on the variation data of multiple measurements. It also helps the company to stay with a 95% (two standard deviations plus and minus the mean) chance over the LSL. The short-term data will show the variation in one batch. The long-term data will show the variation of the preservative, material and treating method over longer time and helps to create the operation targets. The long-term and short-term data can be seen in the Figure 32 and Figure 31.

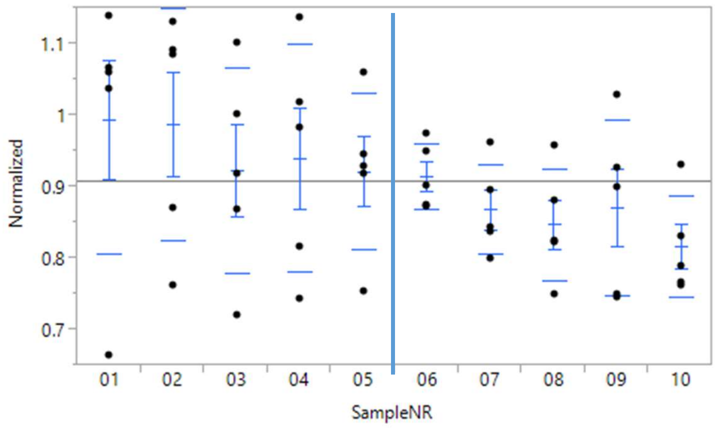


Figure 31: multiple measurements of retention data of a batch

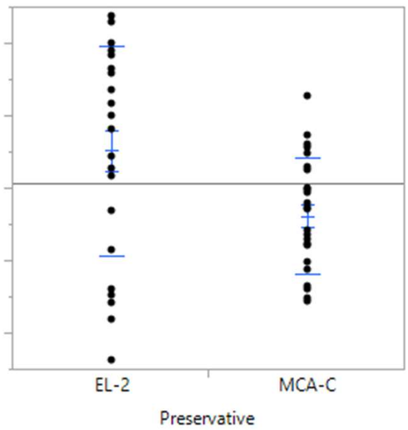


Figure 32: multiple measurements of the retention of a preservative

Conclusion

The study performed, applied statistical process control on industrial and third party data from the retention values of pressure treated residential lumber, with the goal of recommendations to minimize the variation, a statistical control of the process and a possible tool to predict the process. Primary goal of this study was to analyze the data quality.

The important factors for this analysis were the minimum, maximum, range, standard deviation and the coefficient of variation. The time ordered data was plotted in a moving range - and x-individual chart, those could point out the statistical difference of the data under the lower specification level. The agency had 15.1% and the treating plant 5.6% out of specification. This phenomenon can occur through the retesting of data points which are out of specification and is called large variance problem. Through the comparison of Data Pairs the variation of the third-party agency always was lower than the variation in the data of the treating mill. This can be a result of the attention paid on the measurement and handling of the core samples. When the sample is treated every time the same the values will have a smaller variance. A further important result was the distribution, where the log normal distribution fitted the best on this data.

This Results show us a huge potential of optimizations on the side of the treating mill. Through the appliance of statistical process control the mill can predict the process and minimize the values out of specification.

Some further recommendations are to create operational target levels given by current short term and long term variability relative to the lower specification limits. Recommendations can be made on the data quality. Especially the handling of the core samples and the handling of the remeasured data. Recommendations on root cause analysis for reducing variability based on several mill tours. At last recommendations on a new measurement system based on more core samples to evaluate the variance in one batch.

Bibliography

2016 AWPA Book of Standard, American wood Protection Association

ICC Evaluation Service (2016): *Wolmanized® outdoor® preservative-treated wood*, Atlanta: ICC Evaluation
Service

JMP (2016): Overview of the Life Distribution Platform
http://www.jmp.com/support/help/Overview_of_the_Life_Distribution_Platform.shtml#143089 [03.10.2016]

JMP (2016): Likelihood, AICc, and BIC
http://www.jmp.com/support/help/Likelihood_AICc_and_BIC.shtml [04.10.2016]

Lebow, Stan T. (2010): *Research forest products technologist, Wood Handbook, Wood as an Engineering Material*, Madison: USDA united states Department of Agriculture

Lohninger, Hans (2012): Shapiro-Wilk-Test
http://www.statistics4u.com/fundstat_germ/ee_shapiro_wilk_test.html [08/18/2016]

Targue, Nancy R (2005): The Quality Toolbox, Second Edition, ASQ Quality Press

Wheeler, Donald J; Chambers David S (2010): Understanding Statistical process control, third edition, Knoxville: SPC Press

Young, Timothy M (2016): *Statistical process control*, Skript

List of figures

List of tables

List of formulas