A DATA-DRIVEN EXAMINATION OF THE EFFECTS OF PROCESS VARIABLES ON SHEET CALIPER DURING PAPERMAKING

by

Lucas Knill

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of

Master of Science

Boreal Ecosystems and Agricultural Sciences

School of Science and the Environment

Grenfell Campus

Memorial University of Newfoundland

May 2025

Corner Brook, Newfoundland and Labrador

Abstract

Paper thickness, known as caliper, is a critical feature in producing newsprint. The traditional newsprint made at Corner Brook Pulp and Paper Mill (CBPPL) undergoes a calendering ('flattening') process at the end of the line, where the sheet caliper is controlled to meet a specific outcome. A potential market for CBPPL involves paper that does not go through the calendering process. This necessitates a better understanding of the relationship between the paper caliper and other variables present in the manufacturing process so that CBPPL can diversify its production line. This thesis is an observational study examining the relationship between the caliper and other process variables. This is done using baseline statistical techniques, including covariance/correlation analysis, regression, principal component analysis, and a neural network. These techniques show that variables such as species, basis weight, CD tear and KSI have a statistically significant relationship with the caliper. This understanding of variables within the papermaking process is beneficial to CBPPL and the paper-making industry as a whole; further research could be done to explore the effect of tighter control of wood species mix and the interactions of variables within the process.

General Summary

When reading a newspaper, one may notice that the thickness of the paper is consistent among the different pages. This is thanks to a process called calendaring, which occurs at the end of the papermaking process and creates a smooth, consistent paper thickness. Other paper types do not necessarily go through this calendaring process but still require consistent paper thickness and quality. Imagine reading a book where every page has a different thickness. We seek to explore factors in the papermaking process that may influence this thickness. In paper making, this thickness is referred to as caliper. We explore variables in the process that may influence caliper through traditional statistics such as covariance/correlation analysis, regression, and principal component analysis, as well as using a neural network. These techniques show that variables such as species, basis weight, CD tear and KSI have a statistically significant impact on the caliper.

Acknowledgments

I would like to express my sincere gratitude to my family and friends who have been instrumental in my journey and who have consistently provided encouragement in all my endeavours. I extend my appreciation to Dr. Rob Gallant for his outstanding supervision, constructive feedback, and steady support throughout this process. Furthermore, I wish to convey special recognition to Nadia Simmons and the Office of Research and Graduate Studies for their invaluable assistance in facilitating the completion of my master's degree.

Table of Contents

List of Tables

List of Figures

Chapter 1: INTRODUCTION

1.1 Background of Study

1.1.1 Corner Brook Pulp and Paper

Corner Brook Pulp and Paper Limited (CBPPL) is located in Corner Brook, Newfoundland and Labrador. With a population of 19,333 in 2021 (Statistics Canada, 2021), Corner Brook relies heavily on the mill and its residual effect as a driving force of the economy. The mill in Corner Brook has traditionally focused its production efforts on newsprint, but product diversification is highly interesting. Other paper types do not necessarily undergo a calendering process, which helps 'press' the paper into desired caliper ranges (Rättö & Rigdahl, 1998). As such, how the paper caliper varies during manufacturing is of significant interest to CBPPL.

1.1.2 The Papermaking Process

Paper is a spatial network of fibres, fines, and possibly fillers, with tiny voids in between, and many of the properties of the paper are influenced by this structure (Holmstad et al., 2001). Paper production turns fibrous material into pulp, which is then converted to paper (Bajpai, 2018). An 8 ½" x 11" sheet comprises millions of fibers bound by hydrogen bonds (Koivo, 2009). Koivo (2009) discusses the history of papermaking, which was invented around 105 A.D. by Chinese Emperor Tsai Lun, with the first continuous paper-making machine being constructed in France in 1798.

Drost et al. (2003) contend that geographic location and growth conditions are important in determining fibre composition. The age of the trees being harvested is also essential to the process. It has been shown that juvenile wood can have a lower fibre yield than mature wood, by anywhere from 0.4% to 4.4% (Drost et al., 2003).

Along with age, wood species also play a crucial role in the quality of the paper. The

CBPPL woodyard is primarily filled with black spruce (*Picea mariana* (Mill.) B.S.P.) and Balsam Fir (*Abies balsamea* (L.) Mill), the primary feedstocks for papermaking in the province (Norcliffe & Bates,1997). Although recommendations are provided to woodyard workers on the amount of spruce and fir to put into the hopper for pulp at any given time, this is a complex variable to control.

From the woodyard, the wood goes through a debarking and chipping process before becoming pulp. Joutsimo (2004) discusses the next steps where pulping separates cellulose from other materials in the wood, as cellulose is the fibrous material needed to create paper. Several pulping methods exist, including mechanical, chemical, and semi-mechanical methods. Mechanical pulping uses disks and water to grind the wood into pulp while relying on heat and friction (Biermann, 1996). Whereas chemical pulping uses chemicals to break up the wood. Mechanical pulping usually yields more than 90%, whereas chemical pulping converts about half of the dry mass of the wood into fibers (Håkansson, 2014). However, mechanical pulping usually requires much more energy than chemical pulping.

Mechanical pulping leaves the presence of lignin, a dark-colored organic material that binds cellulose fibers; this causes the mechanically pulped paper to be more opaque, weak and easily discoloured (Ververis et al., 2004). Smook (1992) argues that mechanical pulping is more suitable for paper with low brightness but may not be as appropriate for printing, copying, and packaging paper. CBPPL uses a thermomechanical pulp (TMP) where the wood is chipped and squeezed between steel disks in a steam-heated process. The resulting slurry is diluted, and the fibrous mix is pumped to storage tanks before finally being pumped to the paper machine. Mertens et al. (2017) found that TMP is advantageous compared to chemical or mechanical, as it provides higher strength, yields and can have a lower environmental impact.

The paper machine turns this pulp into a sheet of paper, moving the material at speeds of

around 2,000 metres a minute (Pikulik, 2011). The pulp will go through three main sections on the machine: forming, pressing, and drying. After that, the paper may be calendered before being rolled onto a reel. Figure 1 shows the paper machine with different sections of the process outlined. Some steps in the process are not shown, such as the flow of TMP. “A modern paper mill alone has several thousands of basic control loops controlling different flows, temperatures, tank levels, pH of stock, consistencies, etc.” (Koivo, 2009).

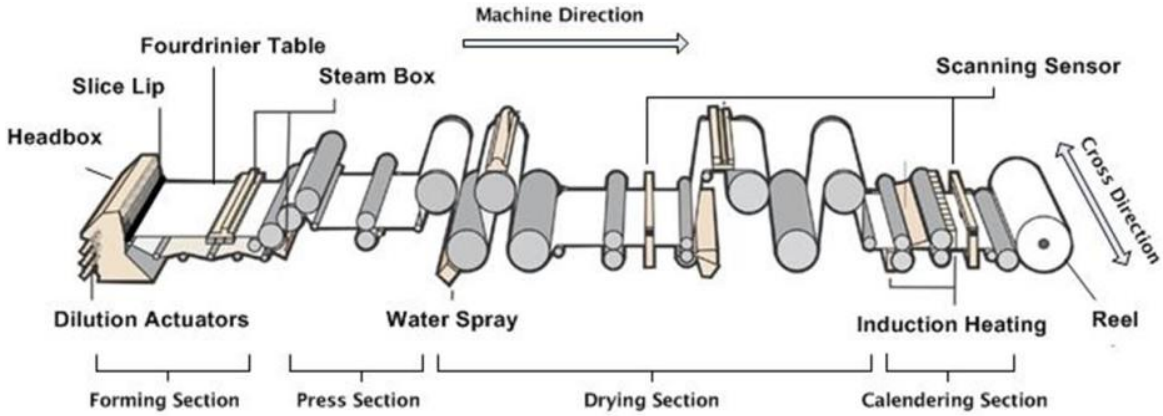


Figure 1 Paper-Making Machine (Source: Chu, Forbes, Backstrom, Gheorghe & Chu, 2011)

In the forming section, the headbox distributes the pulp mixture across the forming fabric/wire (Håkansson, 2014). The headbox needs to produce uniform quality and velocity across the width of the machine to ensure a constant basis weight. Pikulik (2011) outlines the next steps where the mixture will go from 1% solids to 13%-25% by draining water to create a strong fibre sheet. After this, the pressing section brings the product to 38%-50% solids content (Pikulik, 2011). The pressing section is essential for water removal and sheet compaction, considerably increasing the wet sheet's strength from the forming section (Bajapi, 2018). This strength is vital to ensure no sheet breaks, which limits productivity and is costly to the operation. Then comes the drying section, where the final moisture content would be about 2%-10% (Pikulik, 2011). This section removes moisture through evaporation. From here, the newsprint goes through the calendering section before going onto the reels. This process results in a jumbo of paper and sludge or biosolids. Not all types of paper go through the calendering

process, so it is vital to understand the impact of the calendering process on the paper. During the calendering process, the paper goes through a combination of heated rolls before being wound onto a reel. Li (1994) articulates that the calendering process creates a smooth surface for printing, reduces thickness, and corrects irregularities in the paper. There are many types of calendering, such as hard-nip, soft-nip, multi-nip, and long-tip. The final caliper of the paper is controlled by the calendering process, using the temperature of the calendered rolls to manipulate it (Pikulik, 2011).

In addition to the caliper, the strength of the output paper is important to consider. Marklund et al.(1998) summarize two commonly used strength indicators, tensile and tear indexes. Strength parameters can be influenced as early as the wood-cutting phase; fibre dimensions, single-fibre strength, fibre flexibility, and fibre collapsibility are critical factors in the final paper's strength (Marklund et al., 1998). Variation in tree species has also been found to influence strength parameters, with balsam fir having the highest tensile strength, followed by spruce and trailed by jack pine and others (Drost et al., 2003).

1.2 Objectives and Significance of Study

The paper-making process is a riddle of variables that change over time and involve many physical and operational components of the paper machine and process. The Corner Brook Pulp and Paper mill is outfitted with numerous sensors that monitor the papermaking process discussed in the previous section. This research centers on data obtained from CBPPL, capturing paper production sensor data during three periods of interest. We analyze this data using standard statistical methods as well as a neural network-based approach. We interpret the findings with the aim of better understanding how the caliper changes during the paper-making process and what process variables may influence the end caliper of the paper.

Chapter 2: DATA AND METHODS

2.1 Data Collection

Three datasets were provided by CBPPL, each containing process variable data from three periods. The principal data set was taken during a nominal papermaking period and is referred to as the baseline dataset. The two other datasets correspond to periods of operational interest to CBPPL, wherein the species mix of wood used to produce the paper was at two possible extremes – one period where only fir is used and another where only spruce is used. Each dataset consists of 52 sensor measurements, each one minute apart. More information about these 52 variables is outlined in Appendix A. The paper runs were completed on paper machine 2, which is why the variables have the abbreviation PM2; this prefix is sometimes left out when discussing the variables.

The baseline data included measurements for every minute from 8 AM on Feb 1, 2019, to 8 AM on March 12, 2019, thus providing 56099 samples. The second dataset, where only spruce wood was used, included measurements for every minute from 7 PM on March 20, 2019, to 11 PM on March 21, 2019, thus providing 1680 samples. The final dataset, where only fir wood was used, includes measurements for every minute from 3 PM on June 15, 2019, to 3 PM on June 17, 2019, thus providing 1464 samples. These different datasets were chosen to investigate how the individual wood species, compared to the mixture of species used by CBPPL, impact the paper created, with the variation in dataset size due to the trial sizes run by CBPPL.

2.2 Experimental Design and Factor Effects

For the investigation, the machine caliper or actual caliper (PM2 Mx_Caliper) was used as the response (or 'dependent') variable, with the remaining variables as the predictor variables. When a mill operator wants to run a jumbo of a new caliper, the paper machine

operator changes settings, resulting in the paper machine transitioning to a new operating state. This results in changes to the paper machine's various components and sensor readings. The term independent regarding predictor variables is avoided as it may suggest these values are arbitrarily independent, which is certainly not true. For example, Drost et al. (2003) have shown that basis weight is known to be correlated with the caliper for certain paper types.

The variables outlined in Appendix A are labelled as either continuous, discrete or appear to be discrete. The continuous variables can take on an infinite number of values. The discrete variables only take on a finite number of possible values. The variables labelled "Appears to be discrete" take on a value for a period of measurements and then proceed to change to another value, where it stays constant for a certain number of measurements before repeating this process.

2.3 Missing Data and Error Data Points

Some samples within the CBPPL datasets contained missing values, which appeared as "No good Data" within the datasets. We cannot always infer or determine why the data is missing. Such missing data is to be expected when monitoring a complicated process. Sensors can be temporarily knocked offline by transient events or abnormal process states, resulting in missing or invalid data that should be treated as outliers.

How one deals with missing data depends mainly on the study design, analysis goals, and pattern and type of missing data (Soley-Bori, 2013). Missing data can be categorized into three different categories. Sainani (2015) outlines that missing completely at random data has no systemic pattern in how the data is missing. Missing at random refers to data that may be missing within a specific subgroup (Sainani, 2015). Finally, data that is missing not at random is missing due to factors not measured (Sainani, 2015).

One way to deal with the missing data is to remove all observations where data is missing, referred to as complete case analysis (Bennett, 2001). The next group of methods is called single imputation methods, which use an estimate to replace the missing data. Bennett (2001) goes on to discuss the "last value carried forward" method, which replaces the missing data with a value from the last recorded measurement. The "mean Substitution" method replaces the missing value with the mean of all the other measurements of that variable; this can lead to underestimating the variance (Bennett, 2001). "Regression methods" replace the missing data with the outcome of a regression model where the missing data is the outcome variable (Bennett, 2001). Hot-deck imputation replaces the missing value with values with a matching covariate; Cold-deck imputation is similar but depends on knowledge from external information (Bennett, 2001). Sainani (2015) argues that these methods could underestimate the variance because they impute the same value for each missing data point. Rather than imputing one value, multiple imputation methods generate various values for the missing data and combine the results into a single estimate (Sainani, 2015). Statistical algorithms can be used as an alternative to imputation methods. There are many methods for this, including Markov-chain imputation, expectation-maximization approach, and Raw Maximum Likelihood methods, to name a few. There are many ways to deal with missing data points; one solution does not work for all.

Within the mill datasets, the variables representing wet and dry end breaks were removed because these variables were riddled with many "No good data tags" values or were binary variables that jumped between zero and one. To deal with missing or erroneous data points, our datasets were limited to only those jumbos of paper with 30 to 90 observations per jumbo; any fewer or greater number of observations was taken as a sign that the machine was not working correctly, as, on average, there were 60 observations per jumbo. The jumbos with the highest and lowest 5% of the variance were then removed to ensure the data was not influenced by

instances where the machine state was irregular.

2.4 Correlation Analysis

Variance describes the variability in a single variable, while covariance describes how multiple variables vary together (Schober et al. 2018). In contrast, correlation measures how one variable changes in response to another. One measure of the correlation between two variables, X and Y, is the sample correlation coefficient or Pearson correlation coefficient. This coefficient is represented by r_{xy} and is found by dividing the sample covariance by the product of the sample standard deviations; see equation 1 below. In this formula, r_{xy} is the Pearson correlation coefficient between x and y, two variables with n numbers of observations; x_i is the value of x for the ith observation, and y_i is the value of y for the ith observation, $\bar{x}$ is the mean of the x variable and $\bar{y}$ is the mean of the y variable.

$$r_{xy} = \frac{\sum_{n}^{i=1}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{n}^{i=1}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}} \qquad \text{(Eq 1.)}$$

For any two random variables, X and Y, the value of r_{xy} is always between -1 and 1 and can be described using strength and direction (Lee Rodgers & Nicewander, 1988). Strength is defined as strong, moderate or weak, while direction is positive or negative. While the definition of strong, moderate or weak largely depends on the user and field, if r_{xy} is near 1, the value of X and Y will increase or decrease proportionally and strongly. Similarly, when r_{xy} is near -1, the values increase/decrease oppositely but still proportionally. When r_{xy} is near zero, the increases and decreases in X and Y do not match linearly. If variables X and Y are truly independent, then $r_{xy}=0$, though the converse does not hold.

Shiina (2016) explores how correlation has remained relatively unchanged since its inception. Some of the earliest work that contributed to the idea includes *Analyse Mathématique sur les*

Probabilités des Erreurs de Situation d'un Point by Bravais from 1846, in which Bravais was trying to determine errors in the determination of the coordinate points in space, but was not aware of its implications on correlation and is thus not credited with contributing (Piovani, 2008). Also, in the early 1800s, Gauss theorized "the normal surface of n-correlated variates" but did not dive into the concept of correlation (Rodgers & Nicewander, 1988). In 1888, Galton used the term "Co-re-lation" when measuring closeness. Finally, in the mid-1890s, Pearson finalized the concept and developed today's formula (Asuero, Sayago & Gonzalez, 2006).

Correlation shows that one variable is thought to increase or decrease to a certain extent with another variable, but not that the change in one variable is caused by another (Samuel & Okey, 2015). If two random variables are statistically independent, their correlation coefficient is zero. However, the converse is invalid, i.e., if $r = 0$, this does not necessarily imply that x and y are statistically independent. For example, they could be non-linearly related. Gogtay and Thatte (2017) argue that the judgment of cause and effect requires more knowledge of the data set from the investigator as well as further analysis of the data.

Seel (2015) reminds the reader that significance does not mean relevance when interpreting results. There will be a mathematical correlation if two variables increase (or decrease) simultaneously over time. For example, one could compare an increase in traffic lights with an increase in popsicle sales and find they have a strong correlation. This analysis may be true, but it is most likely conceptually irrelevant to data like that. In this example, it would be helpful to examine other variables, such as the increase in population.

2.5 Regression Analysis

Correlation is mainly used to study interdependence, whereas regression is often used to research dependence (Asuero et al., 2006). Poole and O'Farrell (1971) investigate how regression can be used for various analyses, such as descriptive analysis, parameter estimation, predictive capacities

and process control purposes. There are many types of regression, including linear, simple, multiple, ordinal, and many others. The type of regression used depends mainly on the data type being analyzed and the outcome being pursued.

Linear regression models how various predictor variables can estimate a response variable using a linear relationship. When there is one response variable y, and n predictor variables x_1, x_2, …, x_n, the linear regression predictor has the form $y_p = w_1x_1 + w_2x_2 + \ldots + w_nx_n + b$, where the coefficients w_i are chosen to minimize the error between the predicted value of the response variable y_p and the actual value of the response variable y. In the case of a single predictor and single response variable, this prediction can be thought of as the line that best fits through the data. In this work, y_p is an estimate of the caliper, and w_1 to w_n are the weights associated with the process variables. There are various ways to define the error between y and y_p. The most common method is least squares, where the values w_i and b are chosen to minimize the sum of squares of $y-y_p$ over the sample values of the response and predictor variables. The error term $y-y_p$ is sometimes called the 'residual error' or simply a 'residual.' When analyzing a linear regression output, the coefficients w_i can be thought of as a measure of the size of the effect of each predictor variable on the response variable, as well as whether the influence is positive or negative. The value b (the 'intercept') represents what the dependent variable would be if all other influences (independent variables) were eliminated.

Gauss is generally given credit for providing the foundation for regression, but it is believed that Legendre published first on the method (Kumari & Yadav, 2018). In the early 1800s, Gauss and Legendre were interested in objects orbiting the sun, leading to methods similar to the least squares method. Stanton (2001) writes that the term regression was first introduced after this and came about through Sir Francis Galton's inquiry into pea plant genetics. His version of regression slightly differed from today's concept because he used the median for

measuring central tendency and the semi-interquartile range to estimate variability, as these values were easy to obtain (Barnes, 1998). Even though his estimates were not the most mathematically sound, Galton pursued the idea that differences amongst the variability contributed to the differences in regression slope (Barnes, 1998). Galton was more interested in the deviations around the mean than the mean itself, which leaves room for further advancement. Pearson further developed the regression concept into a more general statistical model similar to the modern formulation of regression, which he published in *Philosophical Transactions of the Royal Society of London* and later proved that the optimum regression slope values were from the product-moment (Stanton, 2001).

Whenever we attempt to predict or approximate a dataset, there is the question of how good the approximation is. How accurately and adequately the prediction model represents the actual data is called the goodness of fit. Common measures of fit are the mean square error (MSE) and the square root of the MSE, called the standard error. The standard error is measured in the units of the response variable and measures the absolute distance data points lie from the regression line. Finally, the so-called coefficient of determination, denoted R^2, represents the percentage of the variance in the dependent variable that the model explains.

A concept similar to model adequacy is model validation, which is used to gain assurance about the reliability and robustness of a model. Montgomery et al. (2012) investigate a common way to perform model validation, which is to split the sample data into training(estimation) and testing(prediction) subsets. The training data is seen by the model and used to determine its parameters (the values w_i and b). The resulting prediction model is then applied to the testing data. Because the testing data has not influenced the model, the model's performance on the testing data is more indicative of its predictive power on unseen data from the population.

Figure 2 shows Anscombe's quartet, four datasets with nearly the same summary

statistics and regression line. In three of the four graphs, the line of best fit provided by linear regression may not be the best suited for modelling the data.

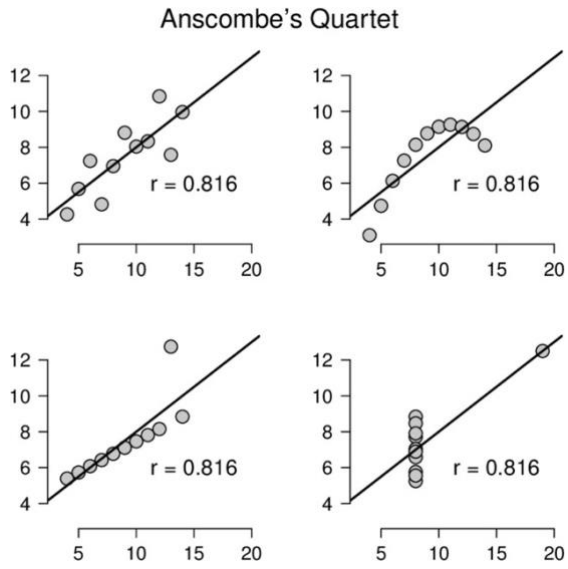


Figure 2 Anscombe's Quartet

One may be using a linear regression simply as a convenient linear predictor. However, further assumptions allow one to infer broader population statistics using the regression. This section will describe the most common assumptions associated with linear regression, the reasons behind them, and the limitations caused by not meeting these assumptions.

One of linear regression analysis's most natural assumptions is that the variables are (roughly) linearly related. Some simple ways to check for linearity in the data are through a scatterplot, examining the residuals or examining the least square error for high accuracy. When reviewing the residual plots, studying the predicted values vs. standardized residuals, and observed values against each predictor is necessary to ensure the data is linear (Osborne & Waters, 2002). This assumption is inherent when examining the equation that linear regression is trying to fit the data; it is the equation of a line. If this assumption is not met, the model may be a poor fit and not capture authentic relationships or provide misleading coefficient estimates.

Another underlying assumption of linear regression is that the outcome variable is continuous. If the outcome variable is categorical/discrete, this becomes a classification problem, and one would be better off using logistic regression, a binary classification method (Tripepi et al., 2011). Casson and Farmer (2014) show that a linear regression model with a single categorical outcome variable that can take 1 of 2 values is equivalent to a simple t-test. Similarly, if the categorical predictor variable has more than two values, the linear regression model is comparable to an ANOVA.

Another common assumption is that the data is drawn from a normal distribution. Osborne and Waters (2002) discuss how this can be checked by a normality test or a histogram of the data that would look like a bell-shaped curve for normal data. This assumption is mainly used for making statistical inferences about the regression model, for example, confidence intervals or showing that the estimator from the regression is the maximum likelihood estimator. Not meeting the normality assumption may not be a problem due to the central limit theorem, which states that large sample sizes will be approximately normal. Still, smaller datasets may be able to be transformed.

Ideally, predictor variables in a linear regression should be linearly independent. This is called a multicollinearity assumption and is essential because correlated (non-independent) predictor variables will make it hard to determine the specific contributions of each variable (Tripepi et al., 2011). This assumption can be checked with a correlation matrix of the variables or via tolerance tests; the most common measure of tolerance is simply $1\text{-}R^2$, where R^2 is the coefficient of determination obtained from the regression model. Tabachnick and Fidell (2001) argue that a higher value for $1\text{-}R^2$ is desirable, with a minimum being .10. If one's data shows multicollinearity, the highly correlated variables could be removed or combined into one variable.

Another assumption is that the error terms, sometimes referred to as residuals, are normally distributed. Checking the normality of the error term can be done through a histogram of the residuals, a Q-Q plot or other formal statistical tests. This assumption allows one to make inferences from the model or estimate the probability that a given error will exceed some threshold, such as a confidence interval. A violation of this assumption means a skewed error distribution that can disproportionately influence parameter estimates. (Osborne & Waters, 2002).

Another property the error terms may satisfy is that they are uncorrelated with a mean of zero. This can be checked by examining the mean of the residuals, examining residual plots, or using tests such as the Durbin-Watson test. An average error value of zero shows the error is unbiased; if the mean is not zero, the model would constantly be under/overpredicting the values. If this assumption is unmet, one can use tools such as Newey-West estimators, also known as Heteroskedasticity and autocorrelation consistent (HAC) standard errors (Montgomery et al., 2021).

One final assumption involving the error term is homoscedasticity, which refers to equal variance of the error terms. The error term represents the distance between each data point and the fitted regression equation; since each point is supposed to contribute to the regression equation equally, this variance should also be equal, thus making the predictive capability the same for each data point (Casson & Farmer, 2014). The easiest way to check this is by plotting the standardized residuals vs. the predicted values. Poole and O'Farrell (1971) write that this assumption can also be checked through tests such as Hartley's Fmax test or Bartlett's test; it should be noted that these tests are susceptible to non-normality in the data. Not meeting this assumption can increase Type 1 errors (Osborne & Waters, 2002).

2.6 Principal Component Analysis

Principal component analysis (PCA) is a linear dimensionality reduction method that can be applied to a dataset with many variables to aid in visualizing complex, high-dimensional data (Mishra et al., 2017). Many terms are used more or less interchangeably for PCA, such as factor analysis, eigenvector analysis, or latent vector analysis (Jolliffe, 2002). "Principal component analysis seeks to reduce the dimensionality of the data along what are called principal components; these components are ordered by the amount of variation from the original data set they retain" (Ringnér, 2008). PCA projects high-dimensional data onto a smaller dimensional subspace – the principal components are an orthogonal basis for this subspace (Mishra et al., 2017). Ringnér (2008) investigates different applications of PCA, such as classification, compressing images and facial recognition.

To understand the basic idea of PCA, suppose there is a dataset of n samples, each consisting of p values. Perhaps these values represent the readings of p different sensors at a given moment. When the relative magnitudes of the variables are different, it is common to standardize the data values by centering and scaling the data. Replacing the matrix product with the correlation matrix of the variables has the same effect as standardizing the data first.

Therefore, it is important to center the data so that the mean of each of the p-values is zero. The data can then be represented using an $n \times p$ matrix X, where the n rows of X can be thought of as the n samples and the columns as the values for each sensor. The dataset can then be linearly reduced by choosing any unit basis of R^k. Then a matrix, W, of form $p \times k$ can be created where the columns are the basis vectors. The $n \times k$ matrix XW can then be thought of as linearly projecting each point in the data X to a k-dimensional space. There are many ways to choose the basis vectors W. PCA uses a basis W that is orthogonal and has the property that the variances of the projections onto the basis

vectors W are maximized. For example, if w is the first basis vector (and so the first column of W), then the matrix product $(Xw)^t$ (Xw) represents the variance of the transformed data when projected onto the first basis vector (because the transformed data is still centered). To maximize this variance, the unit vector w that maximizes the value $(Xw)^t$ (Xw) is desired; the vector solving this optimization problem is the first basis vector in W. The next basis vector of W is chosen to maximize a similar matrix product, but it must now be orthogonal to w. Continuing in this manner, one can iteratively find each k basis vector in W. These *k* vectors are the 'principal components' of the transformed space.

Although both had different motivations, Mishra et al. (2017) credit Pearson and Hotelling with inventing PCA. Pearson was interested in fitting lines and planes to points in P-dimensions. In contrast, Hotelling was interested in finding a smaller set of independent variables that determine the values of the original variables (Bro & Smilde, 2014).
Others credited with the advancement of PCA include Girshick and his alternative derivations of principal components, as well as Rao, who provided other interpretations and extensions of PCA (Jolliffe, 2002). It was only with the advent of computers that PCA became widely used due to the high computational needs of the technique (Jolliffe, 2002).

Shlens (2003) examines how principal components (PCs) are closely related to matrix eigenvectors and eigenvalues and can be computed through many algorithms, including factor analysis, singular value decomposition (SVD), and eigenvector analysis. Principal components are arranged in order of highest eigenvalue to lowest, which gives the order of principal components with the most variation to least. Once the principal components have been computed, one can decide how many principal components to keep for further analysis. This choice reflects a trade-off between reducing the data dimensionality and retaining as much data variation as possible. Jolliffe (2002) discusses how one can use a scree plot, which shows the

variation along each principal component, to subjectively decide where the plotted points go from 'steep' to 'not steep' and retain the components before the change. An alternative method to decide if the number of PCs included provides an accurate representation is to use cross-validation of the data by splitting the data into testing and training sets and then predicting the test set (Bro & Smilde, 2014).

Rodionova et al. (2021) state that the most common outputs from PCA are scores, loadings, and the associated plots. The scores represent where the original observations are projected onto each component. Thus, they can be thought of as the coordinates on the plane created by the new principal components. Because the matrix in PCA is orthonormal, the relationships from the untransformed data are held. Thus, a score plot allows one to examine clusters, outliers and relationships between individual measurements (Bro & Smilde, 2014). The next significant output is loadings. Bro and Smilde (2014) define loadings as the weight of each variable on the respective principal component. Therefore, each principal component has a loading vector that shows how a coordinate on that principal component axis combines the values from each of the original variable axes. Thus, loading plots show the relationship between each principal component and the original variables.

2.7 Neural Networks

The advancement of technology has made techniques like those described earlier easier and has also created new data analysis fields such as machine learning. As a field within artificial intelligence, machine learning has been evolving since as early as 1949 when Donald Hebb published "The Organization of Behaviour" about neurons and their interactions (Tarca et al., 2007). Fradkov (2020) reports that the term machine learning was coined by Arthur Samuel, who programmed a computer to play chess and record all previous positions. Frank Rosenblatt

used similar ideas to create the perceptron, which recognized the letters of the alphabet and was the predecessor of neural networks. Since then, and especially in the last decade, new techniques and more robust computing power have led to a surge in machine learning applications.

Artificial neural networks are a specific type of machine learning tool. Warner and Misra (1996) propose that these networks are based on the model of a biological neuron created by Warren McCulloch and Walter Pitts in 1943. The purpose of neural networks is to exploit unknown structures in a dataset by making the higher-level representation of the data more abstract while making the data features more invariant to variations in training distribution and preserving as much information about the input as possible (Bengio, 2012). Neural networks can be used for many data exploration techniques, such as dimensionality reduction, regularization, clustering, anomaly detection, and recommender systems (Cunningham & Ghahramani, 2015).

In general, neural networks can be thought of as a specific type of function from R^n to R^k that is constructed using multiple layers. The input is a vector x_0 in R^n and can be thought of as 'layer 0' of the network. The neural network transforms x_0 into a vector x_1, where $x_1 = f(W_1x + b_1)$, W_1 is an h x n matrix of 'weights,' and b_1 in R^h is called a 'bias' vector. The function f is called the 'activation' function and is typically a simple nonlinear function such as ReLU(x): = x if x>0 and 0 otherwise. This layer is meant to model the action of h neurons receiving an input of n numbers, each of the h neurons weighting the importance of the n inputs, and then each neuron 'summing' their total (weighted) inputs and finally 'firing' (creating an output) if the input signal is strong enough. Subsequent layers behave the same way, transforming an input vector into an output vector via a matrix multiplication and activation function (Voulodimos, et al., 2018). The collection of weight matrices and bias vectors for the entire network is called the 'learnable parameters' of the model. The example transforms the input layer x_0 in R^n to a vector x_1 in R^h. The output sizes of each layer can vary, though it is common to have many layers with

the same size of input and output vector. Figure 3 is a representation of a neural network with an input layer of length 4, which is transformed to a vector of length 3 in layer 1, which is transformed to a vector of length 4 in layer 2, which is finally transformed to a vector of length 2, which is the final output of the neural network.

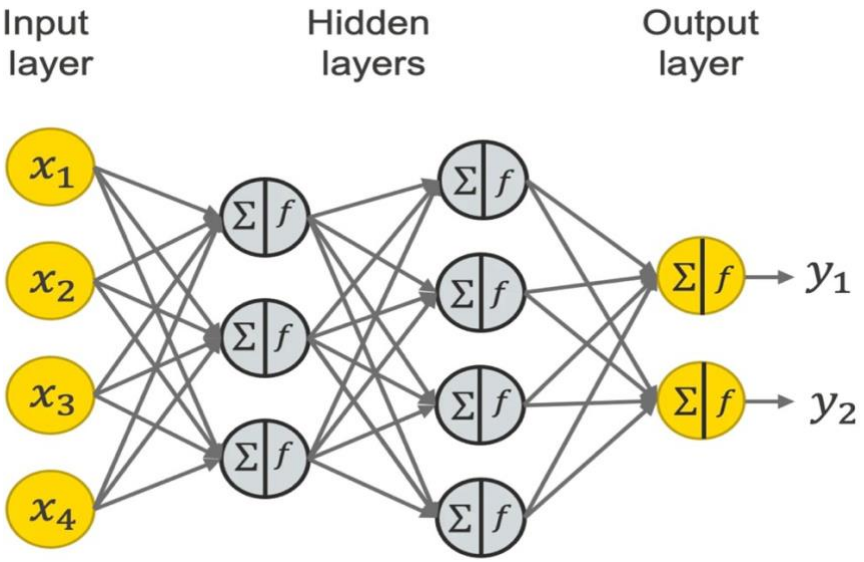


Figure 3 Feedforward neural network (Source: https://www.knime.com/blog/a-friendly-introduction-to-deep-neural- networks)

A neural network N is 'trained' to 'learn' a function during the training process. In this case, the data is a list of input and output pairs (x_i, y_i). The network is given an input x_i, and the output $N(x_i)$ is computed and compared to the 'desired' output y_i. The difference in the desired output y_i and the actual output $N(x_i)$ can be thought of as an error signal that can be used to modify the weight matrices W and bias vectors b for each layer in such a way that the function N with revised weights has $N(x_i)$ closer to y_i (Bebis & Georgiopoulos, 1994). This is done repeatedly for each point in the dataset, and over time, this learning process results in a trained network N where the differences between $N(x_i)$ and y_i are minimized. Mathematically, this happens by iteratively applying a gradient descent algorithm to find parameters that minimize a loss function, such as the mean square error MSE between the desired output y and the network output N(x).

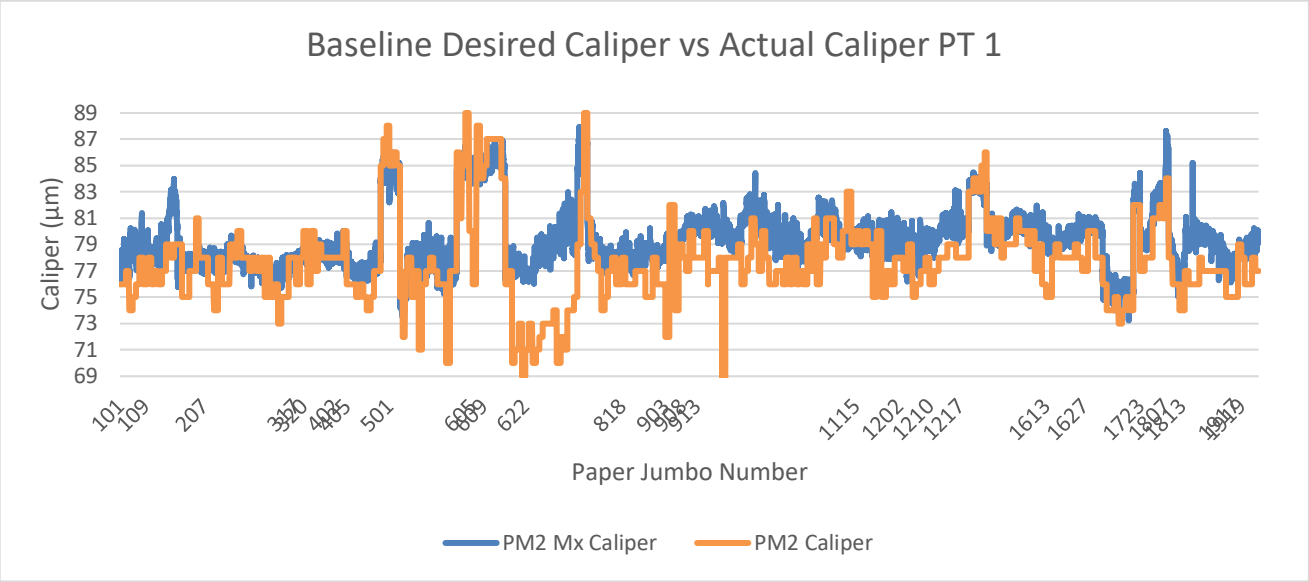
What makes neural networks interesting is that they have shown, as a class of functions, to have a good ability to approximate many complicated functions encountered in practice. It is not well understood whether there is something special about neural networks that affords them this approximative ability. Indeed, part of their success now, compared to earlier decades, is simply due to the availability in the last few decades of massive data sets and vast computational power now available to train such networks and the increased flexibility in approximation possible for large networks with many (billions, in some cases) weights that can be tweaked during the learning process. Ironically, sometimes a network can learn the training data 'too well,' which means the network will continue to tweak weights to decrease the error between $N(x_i)$ and y_i for the test data but at the expense of making the prediction error worse on unseen data from the same population as the original data. This phenomenon is called overfitting and can be a problem when training neural networks with data sets that are too small compared to the number of learnable weight parameters of the model. Monitoring the neural network's performance on a validation set during training on a test set (just as in model validation, discussed earlier) allows one to detect overfitting and halt training.

Chapter 3: RESULTS AND DISCUSSION

3.1 Caliper Variance

Each roll of paper (referred to as a jumbo) has a desired caliper set by the mill to meet a standard specification, represented by the variable PM2_Caliper. This variable is manipulated by the person running the paper machine and may change by jumbo depending on the paper's specification. Since the machine does not always meet the exact value of the caliper, we also have the variable PM2 Mx_Caliper, which measures the actual caliper of paper on the machine at each observation.

The graphs below explore the difference between the jumbos' desired and actual calipers for the three datasets. The horizontal axis outlines where each jumbo of paper changes; thus, jumbos with fewer observations are closer together, and jumbos with more observations are spread further apart. The vertical axis lists caliper measurements. Due to the large amount of data, the baseline data was split into two graphs for easier viewing. The second part of the baseline data has the suffix MAR after the jumbo number to indicate those jumbos were from March and to not mix them up with jumbos of the same number from the previous month.



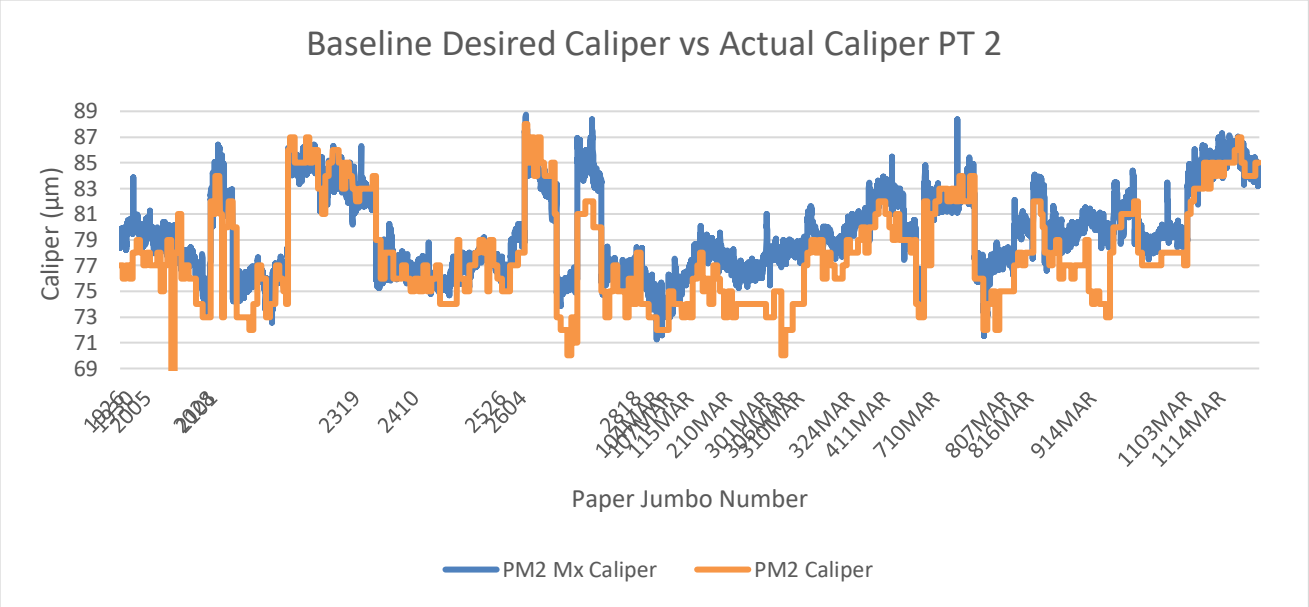


Figure 4 The desired caliper (PM2 Caliper) and the actual caliper (PM2 Mx Caliper) for the jumbos of paper in the baseline dataset, split into two figures due to the size of the dataset

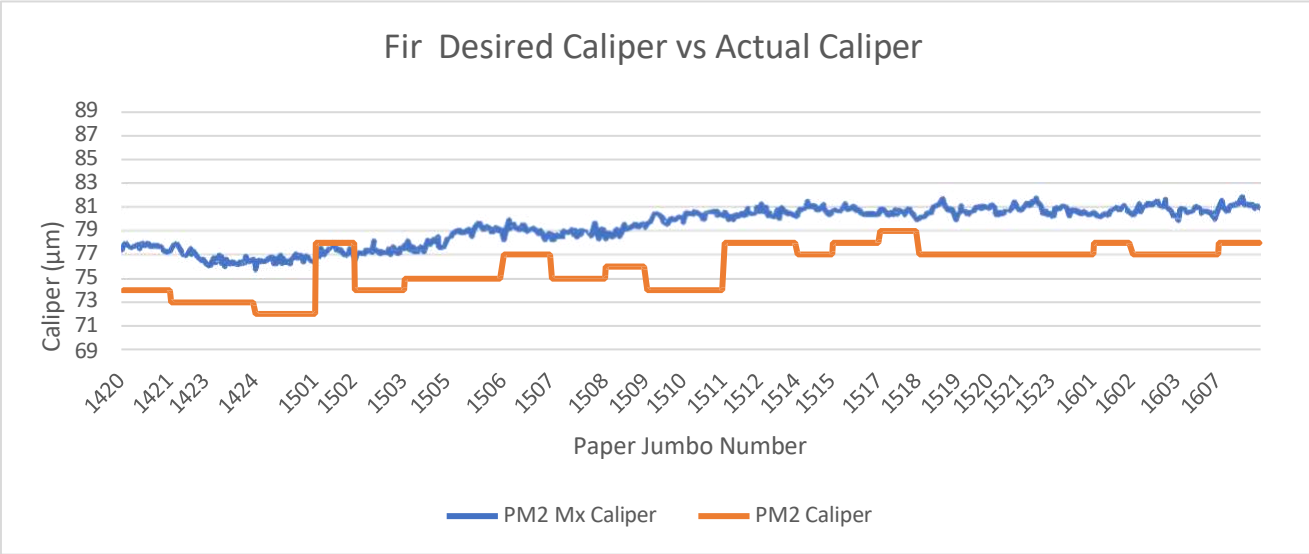


Figure 5 The desired caliper (PM2 Caliper) and the actual caliper (PM2 Mx Caliper) for the jumbos of paper in the fir dataset

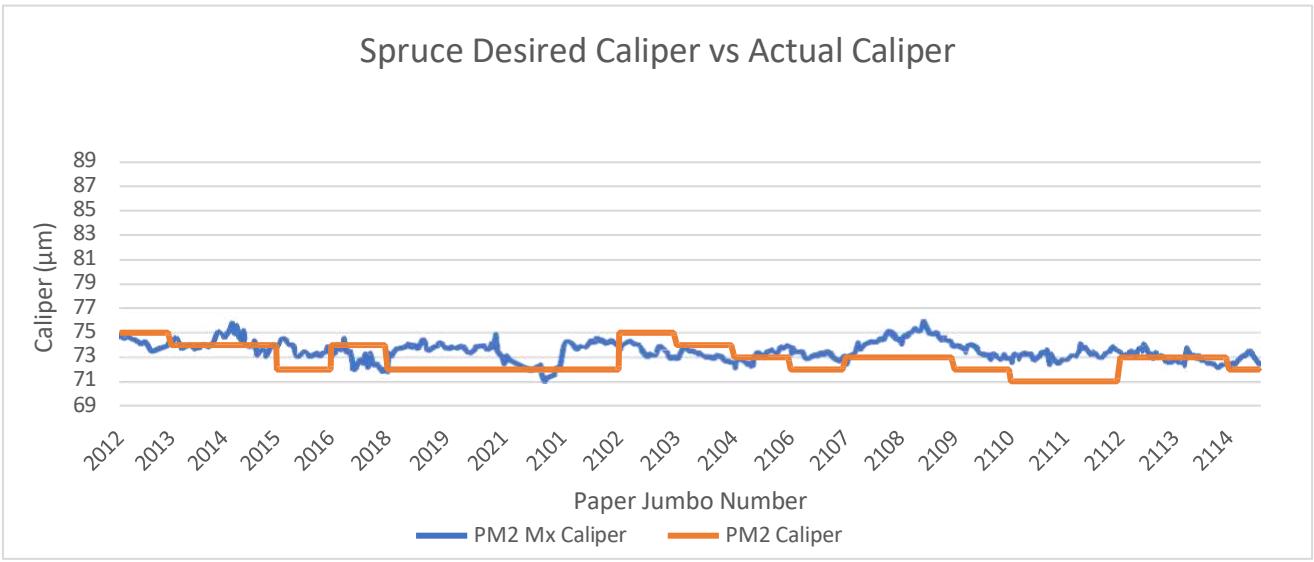


Figure 6 The desired caliper (PM2 Caliper) and the actual caliper (PM2 Mx Caliper) for the jumbos of paper in the spruce dataset

The following conclusions can be drawn by examining the above graphs. Figure 5 shows that the actual measured caliper (PM2 Mx_Caliper) was consistently higher than the desired caliper (PM2_Caliper) during the fir-only trial. Of the three datasets, the machine caliper and desired caliper appear closest to each other in the spruce dataset.

The average, standard deviation and variance were calculated further to explore the difference between PM2_Caliper and PM2 Mx_Caliper. The following formulas were used for

the i^{th} observations within the j^{th} jumbo, with n_j number of observations within each jumbo and N number of jumbos within each dataset. Let X_{ij}=$PM2\ Mx\_Caliper_{(ij)} - PM2\_Caliper_{(ij)}$

Table 1 *Mathematical equations for traditional exploratory statistics*

Variable	Mathematical Equations
Average caliper difference for each data set	$\sum_{j=1}^{N} \frac{\left(\frac{\sum_{i=1}^{n}(X_{ij})}{n}\right)}{N}$
Average standard Deviation in the caliper difference for each dataset	$\sum_{j=1}^{N} \frac{\sqrt{\frac{\sum_{i=1}^{n}(X_{ij}-\overline{X}_{ij})^2}{n}}}{N}$
Average variance in the caliper difference for each dataset	$\sum_{j=1}^{N} \frac{\frac{\sum_{i=1}^{n}(X_{ij}-\overline{X}_{ij})^2}{n}}{N}$

Table 2 *Traditional exploratory statistic results for the three datasets*

Species	Average	Standard Deviation	Variance
Baseline	1.64	0.48	0.28
Fir	3.39	0.29	0.09
Spruce	0.61	0.43	0.20

The results in Table 2 show that the dataset with only fir provided a smaller variance in the difference X between the measured and desired calipers than the other datasets. The largest variance of X occurred in the baseline dataset when both species were mixed, although the variance was not much more than for the spruce dataset. The lower variance for the fir dataset is interesting as it means the difference between the desired and actual caliper varied less throughout the dataset, suggesting a more consistent effect on the caliper with fir. The larger average value for X for the fir confirms what was seen in the graphs, which shows that the difference between the measured and desired caliper was consistently larger for the fir sample. Similarly, the spruce has the smallest average for X, confirming what was suggested by Figure 6.

To test if the average difference in the desired and actual caliper may statistically be

attributed to the species mix, an Analysis of Variance (ANOVA) was run on the data. Leven's test was used to analyze whether the group samples were drawn from populations with the same variance to test the homogeneity of variances. The results in Table 3 below show p=0.005, so we reject the null hypothesis for alpha=0.05 and conclude that the assumption of homogeneity of variances was violated.

***Table** 3 Test of homogeneity of variances of caliper difference amongst the three datasets*

	Levene Statistic	df1	df2	Sig.
Based on Mean	5.24	2	825	.005
Based on Median	5.33	2	825	.005
Based on Median and with adjusted df	5.33	2	788.59	.005
Based on trimmed mean	5.23	2	825	.005

The normality tests in Table 4 show that X is not necessarily drawn from a normal distribution. Due to the Central Limit Theorem, the ANOVA can still be completed because of the large sample sizes.

***Table 4** Normality tests for the caliper difference*

	Kolmogorov-Smirnov			Shapiro-Wilk		
Species	Statistic	df	Sig.	Statistic	df	Sig.
Baseline	.097	771	<.001	.878	771	<.001
Fir	.172	33	.014	.887	33	.003
Spruce	.124	24	.200	.947	24	.230

Since the assumption of homogeneity of variances was not met, a Welch ANOVA was used. The main difference is that Welch's ANOVA modifies the F-statistic and degrees of freedom to account for different variances and sample sizes. The ANOVA's null hypothesis (H_0) is: There is no statistically significant difference in mean caliper difference (X) between species. The alternative hypothesis (Ha) for the ANOVA is that the mean caliper difference (X) of at least one species differs significantly from the overall mean.

Table 5 *ANOVA results for caliper difference amongst datasets*

	Sum of Squares	df	Mean Square	F	Sig.
Between Species	123.51	2	61.75	11.08	<.001
Within Species	4597.17	825	5.57		
Total	4720.67	827			

Table 6 *Welch's robust test of equality of means for caliper difference*

	Statistic	df1	df2	Sig.
Welch	42.26	2	45.92	<.001

Since the P value in Welch's test is less than 0.05, we reject the null hypothesis and accept the alternative hypothesis. Therefore, there is a statistically significant difference in the value of X between species. To investigate this difference further, a Games-Howell post hoc test was completed.

Table 7 *Games-Howell multiple comparison test results for species effect on caliper difference*

		Mean Difference	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Baseline	Fir	-1.75	.23	<.001	-2.30	-1.20
	Spruce	.99	.24	<.001	.39	1.59
Fir	Baseline	1.75	.23	<.001	1.20	2.30
	Spruce	2.74	.31	<.001	2.00	3.49
Spruce	Baseline	-.99	.24	<.001	-1.59	-.39
	Fir	-2.74	.31	<.001	-3.49	-1.99

As seen in Table 7, the p-value is less than 0.05 for the mean difference between all three datasets. Therefore, the difference between the desired and measured caliper was statistically significant between the datasets. Thus, the species may impact how the desired and actual calipers differ.

3.2 Correlation Analysis of Variables

Correlation analysis was completed on all three datasets to understand what variables may have a relationship with the measured caliper. The correlation r_{xy} between variables x and y is calculated using the observations x_i and y_i from the samples in the respective dataset and the

formula outlined in Chapter 2. These correlations were then plotted on a heatmap in Figures 7,8, and 9. Each square represents the correlation between the variables in the corresponding row and column, with the strength and direction of the relationship indicated by the colour intensity and hue; the lighter the colour, the stronger the positive correlation, and the darker the colour, the stronger the negative correlation.

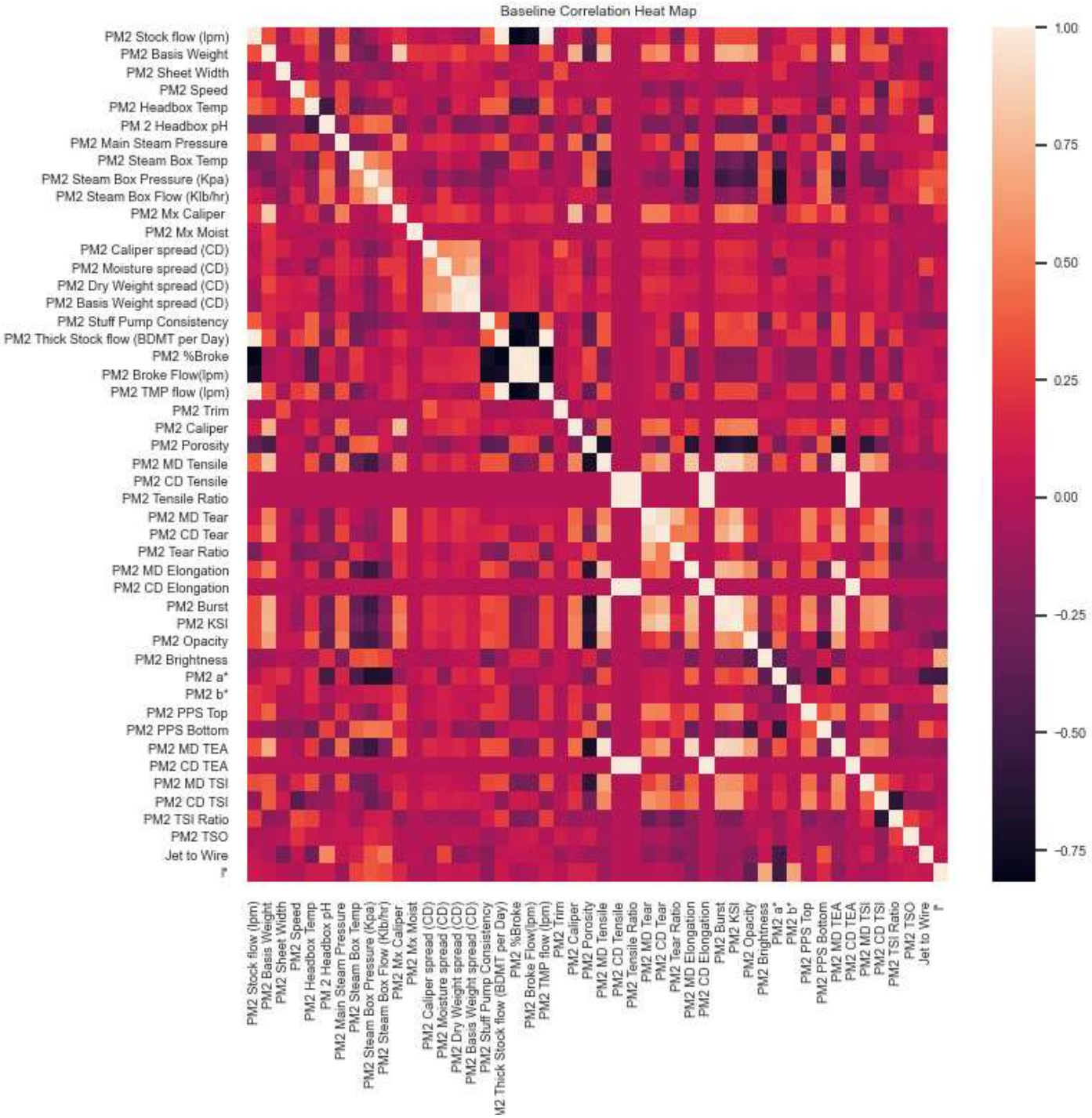


Figure 7 Correlation heatmap for the baseline dataset

Figure 7 shows the caliper heatmap for the baseline dataset. When examining the row for PM2 Mx_Caliper, there appear to be many colours around the middle of the colour spectrum, suggesting the variables have a low correlation with the caliper. Compare this to the row for PM2 Mx_Caliper in the fir heatmap in Figure 8, and one will see more darker and lighter colours, suggesting stronger correlations amongst variables with PM2 Mx_Caliper in the fir dataset.

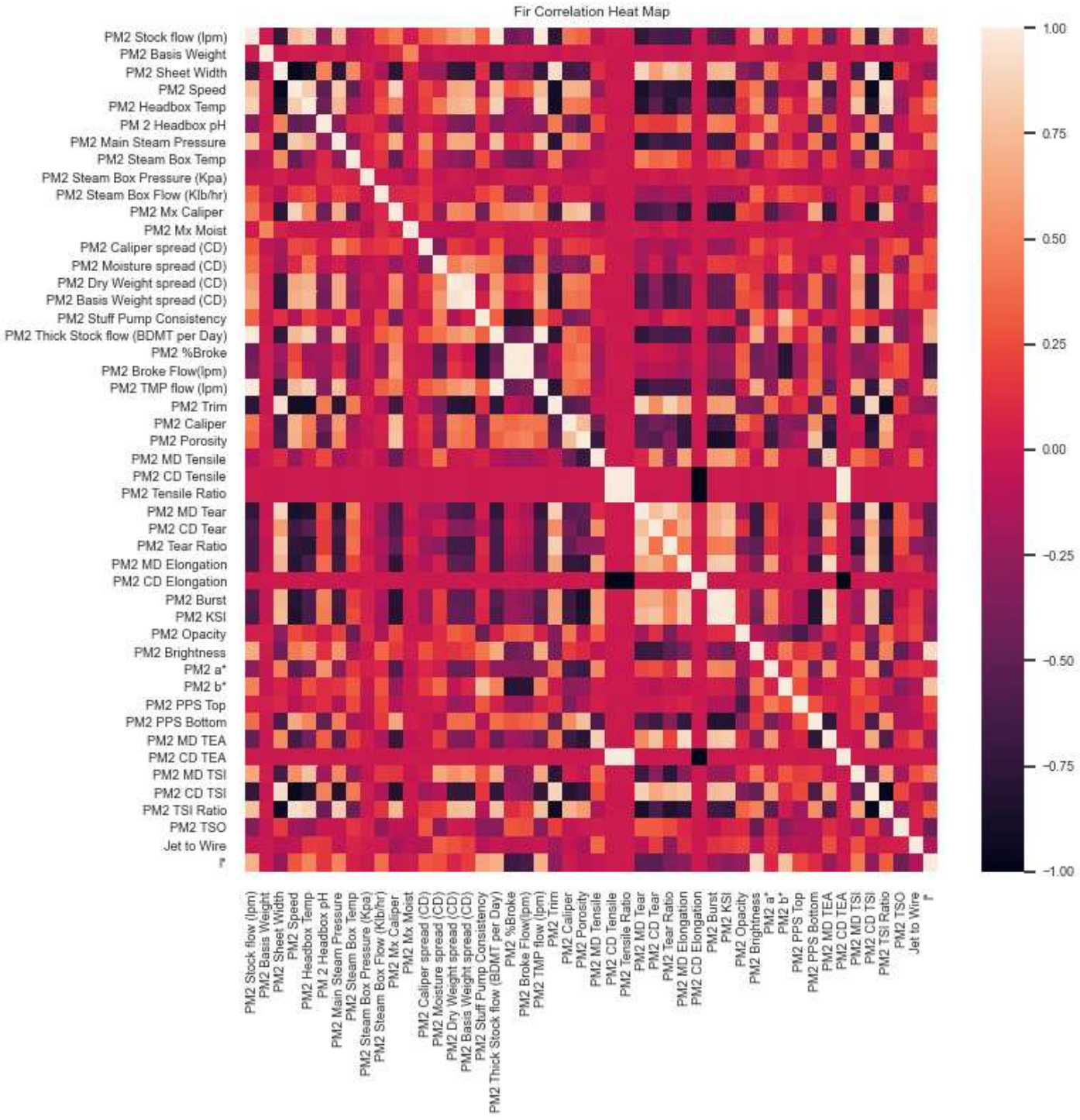


Figure 8 Correlation heatmap for the fir dataset

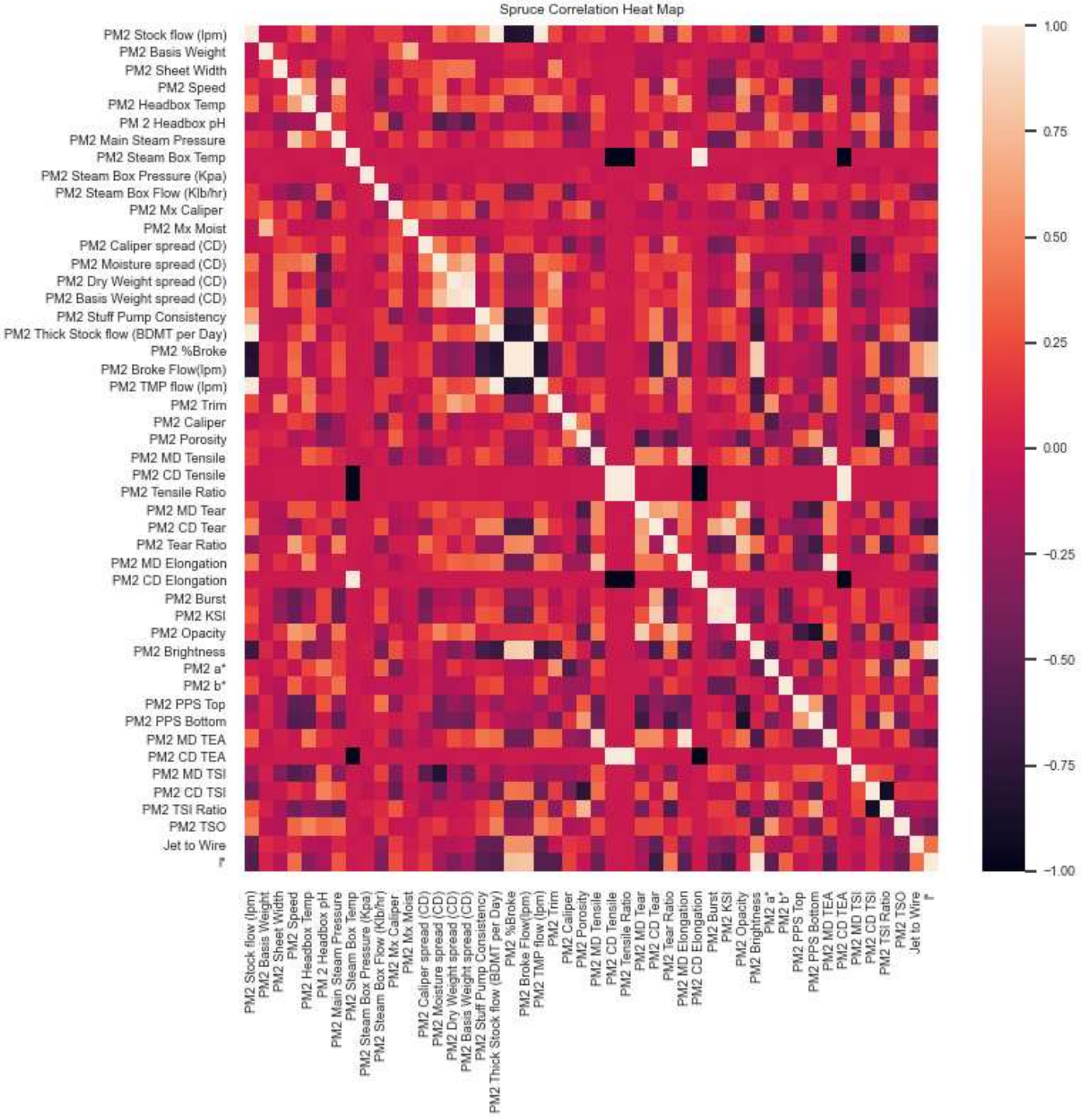


Figure 9 Correlation heatmap for the spruce dataset

Comparing Figures 7,8, and 9, there appear to be more variables that were positive, albeit not strongly correlated to the machine caliper for the baseline dataset, than the other two datasets. It also appears that variables in the baseline and fir datasets have stronger correlations to the machine caliper than the spruce dataset. The table below shows the correlation value between the machine caliper and each other variable for the three datasets to understand better the variables correlating with the caliper.

Table 8 *Correlations between PM2 Mx_Caliper and other variables*

Variable	Baseline	Fir	Spruce
PM2 Stock flow (lpm)	0.21	0.45	0.17
PM2 Basis Weight	0.84	0.12	0.34
PM2 Sheet Width	-0.01	-0.78	0.05
PM2 Speed	0.05	0.86	0.23
PM2 Headbox Temp	0.19	0.49	-0.03
PM 2 Headbox pH	-0.18	-0.44	-0.42
PM2 Main Steam Pressure	0.51	0.69	0.18
PM2 Steam Box Pressure (Kpa)	-0.05	0.00	0.02
PM2 Steam Box Flow (Klb/hr)	0.17	0.08	-0.16
PM2 Mx Caliper	1	1	1
PM2 Mx Moist	0.09	0.04	0.22
PM2 Caliper spread (CD)	0.18	0.23	0.26
PM2 Moisture spread (CD)	0.24	-0.24	0.38
PM2 Dry Weight spread (CD)	0.27	0.52	0.19
PM2 Basis Weight spread (CD)	0.18	0.48	0.33
PM2 Stuff Pump Consistency	-0.04	-0.41	-0.33
PM2 Thick Stock flow (BDMT per Day)	0.21	0.45	0.17
PM2 %Broke	0.12	0.52	0.09
PM2 Broke Flow(lpm)	0.19	0.59	0.13
PM2 TMP flow (lpm)	0.21	0.45	0.17
PM2 Trim	-0.02	-0.66	-0.15
PM2 Caliper	0.78	0.74	0.21
PM2 Porosity	-0.10	0.80	0.33
PM2 MD Tensile	0.52	-0.47	-0.32
PM2 MD Tear	0.51	-0.63	-0.09
PM2 CD Tear	0.51	-0.58	-0.14
PM2 Tear Ratio	0.19	-0.48	0.03

PM2 MD Elongation	0.23	-0.75	-0.13
PM2 Burst	0.46	-0.79	-0.11
PM2 KSI	0.53	-0.81	-0.14
PM2 Opacity	0.45	0.24	0.06
PM2 Brightness	-0.01	0.18	0.27
PM2 a*	-0.14	-0.58	-0.36
PM2 b*	0.25	-0.25	-0.03
PM2 PPS Top	0.38	-0.11	-0.27
PM2 PPS Bottom	0.16	0.66	-0.01
PM2 MD TEA	0.41	-0.72	-0.29
PM2 MD TSI	0.12	0.16	-0.34
PM2 CD TSI	0.22	-0.81	-0.46
PM2 TSI Ratio	-0.14	0.74	0.30
PM2 TSO	0.01	-0.05	-0.03
Jet to Wire	-0.02	-0.20	0.18

Of interest in Table 8 above are the variables with a correlation coefficient above or below .5 and -.5. In the baseline dataset, these variables are basis weight, main steam pressure, PM2_Caliper, MD tensile, MD tear, CD tear, and KSI. The fir dataset shows sheet width, speed, main steam pressure, dry weight spread (CD), % broke, broke flow, trim, caliper, porosity, MD tear, CD tear, MD elongation, burst, KSI, a*, PPS Bottom, MD Tea, CD TSI, and TSI ratio as strongly correlated. No variables were shown to be highly correlated to the machine caliper in the spruce dataset according to the outlined specifications.

3.3 Regression Analysis of Variables

To perform linear regression on the mill datasets, PM2 Mx_Caliper is treated as the response (dependent) variable, with the rest as the predictor (independent) variables. PM2 CD Tensile, PM2 Tensile Ratio, PM2 CD Elongation, and PM2 CD TEA were removed from the analysis because they are constants; thus, there is no variation for the regression to explain. TMP Flow was also excluded from the analysis as it shows multicollinearity with other variables in all

three datasets, meaning its contributions to the model cannot be distinguished from those of different variables. Within the fir dataset, CD TSI was removed for multicollinearity. Within the spruce dataset, Stock flow, basis weight spread, CD Tear, KSI, Opacity, brightness and TSI Ratio were all excluded from the analysis due to multicollinearity. The variables were standardized to ensure they evenly contributed, and variable units did not impact the model.

Baseline Dataset

The first regression model was run on the baseline dataset. Figure 10 below plots the residuals vs. predicted values, where the residuals are the vertical distance from the observed data to the regression line. The residuals are located randomly around the zero line, suggesting a linear relationship. Also, the points are generally in a horizontal band, constantly spread across the fitted values, suggesting homoscedasticity of the residuals.

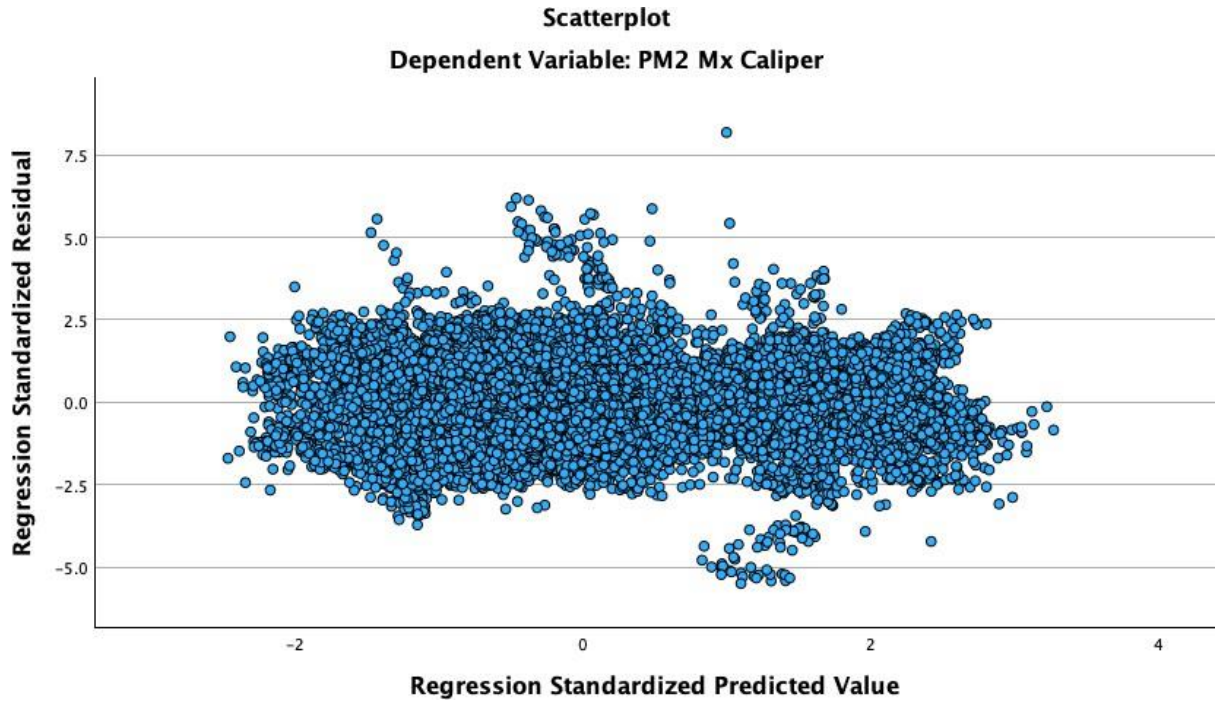


Figure 10 Scatterplot of the residual vs. predicted values for the baseline dataset

Figure 11 shows a histogram of standardized residuals with a normal curve imposed on it. Figure 12 is a normal probability plot of the residuals. Both figures show that the residuals are normally distributed with a mean close to 0 and a standard deviation close to 1.

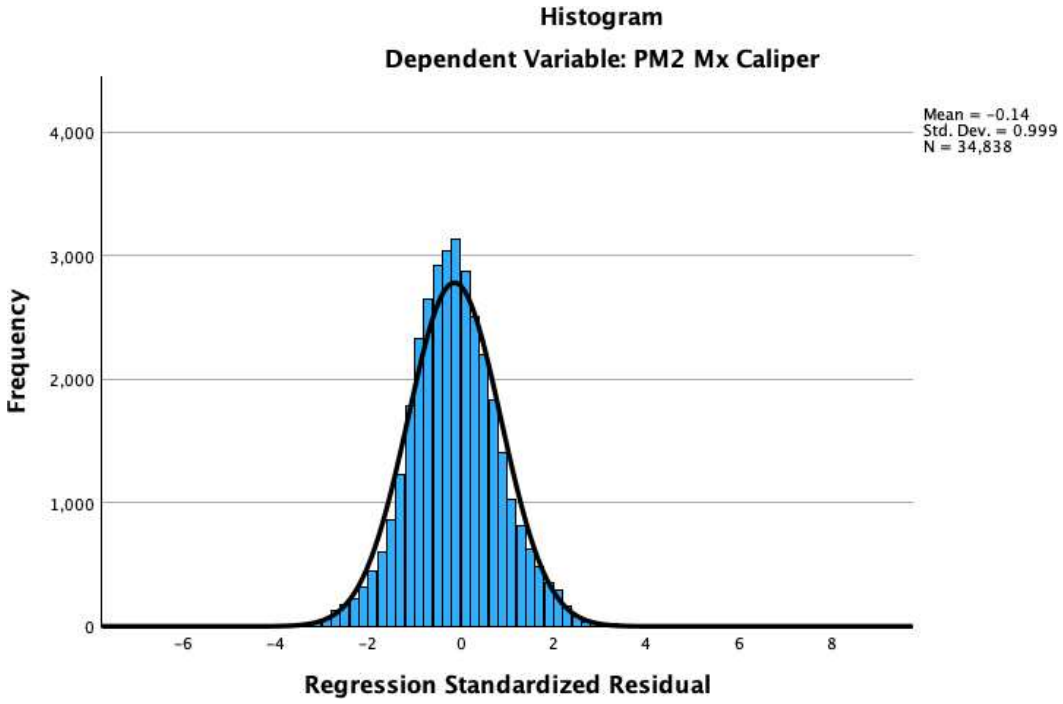


Figure 11 Histogram of the standardized residuals imposed with a normal curve for the baseline dataset

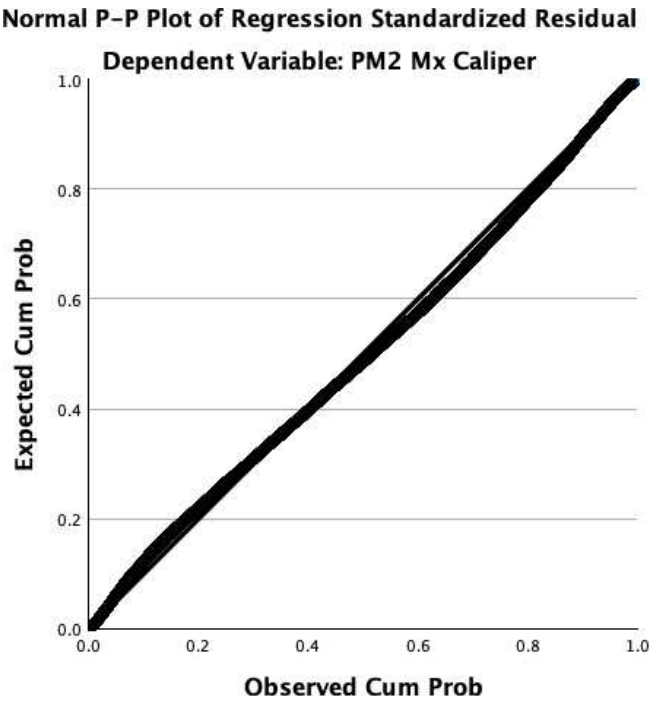


Figure 12 Normal probability plot of the residuals for the baseline dataset

Table 9 below shows the coefficients for the regression model. The most significant regression coefficients were basis weight and l* with coefficients of .73 and -.58, respectively.

Table 9 *Standardized regression coefficients for a model on the baseline data*

Variables	Standardized Coefficients
PM2 Stock flow (lpm)	.49
PM2 Basis Weight	.73
PM2 Sheet Width	.00
PM2 Speed	-.02
PM2 Headbox Temp	.03
PM 2 Headbox pH	-.05
PM2 Main Steam Pressure	.07
PM2 Steam Box Temp	.01
PM2 Steam Box Pressure (Kpa)	-.07
PM2 Steam Box Flow (Klb/hr)	.05
PM2 Mx Moist	-.07
PM2 Caliper spread (CD)	-.03
PM2 Moisture spread (CD)	.18
PM2 Dry Weight spread (CD)	-.04
PM2 Basis Weight spread (CD)	-.12
PM2 Stuff Pump Consistency	.07
PM2 Thick Stock flow (BDMT per Day)	-.33

PM2 %Broke	.28
PM2 Broke Flow(lpm)	-.02
PM2 Trim	.01
PM2 Caliper	.11
PM2 Porosity	.27
PM2 MD Tensile	.03
PM2 MD Tear	.35
PM2 CD Tear	-.14
PM2 Tear Ratio	-.09
PM2 MD Elongation	-.08
PM2 Burst	-.02
PM2 KSI	-.08
PM2 Opacity	.14
PM2 Brightness	.47
PM2 a*	-.14
PM2 b*	.46
PM2 PPS Top	.07
PM2 PPS Bottom	.03
PM2 MD TEA	-.01
PM2 MD TSI	-.05
PM2 CD TSI	-.08
PM2 TSI Ratio	-.00
PM2 TSO	-.04
Jet to Wire	-.07
l*	-.58

Table 10 shows the R and R^2 values for the model. As the R-value is close to 1, there is a strong linear association between the PM2 Mx_Caliper and the caliper predicted by the regression model. The high R-squared value also indicates that the model is a strong fit, as this expresses the proportion of variation accounted for by the regression model over and above the mean model. Adding the predictor variables into the regression model, compared to the mean model, explains 92.4 % of the variability of the response variable.

Table 10 *Model summary for the baseline regression model*

R	R Square	Adjusted R Square	Std. Error of the Estimate
.96	.92	.92	.77

Lastly, an ANOVA was run for the regression model. The null hypothesis for the ANOVA is that the regression model results are not statistically significantly different from the mean of the dependent variable. Thus, the alternative hypothesis is that the regression model results statistically differ from the mean model. The ANOVA results in Table 11 show a significance level of less than 0.05; therefore, the null hypothesis is rejected, so the regression model predicts that the machine caliper is statistically different (and better) from the mean model.

Table 11 *ANOVA results for the baseline regression model*

	Sum of Squares	df	Mean Square	F	Sig.
Regression	249103.06	42	5931.03	10062.03	<.001
Residual	20509.78	34795	.59		
Total	269612.84	34837			

The baseline dataset was then randomly split into two groups to ensure the above model does not overfit the data. The first group is a random sample of 70% of the baseline data for training a regression model. The other 30% is then used as a validation dataset on the regression model. The training dataset has created the model outlined in Table 12. The model looks similar to that in Table 9 and has an R^2 value of .93.

Table 12 *Standardized regression coefficients for a model on the baseline training data*

Variables	Standardized Coefficients
PM2 Stock flow (lpm)	0.53
PM2 Basis Weight	0.74
PM2 Sheet Width	0.00
PM2 Speed	-0.02
PM2 Headbox Temp	0.03
PM 2 Headbox pH	-0.05
PM2 Main Steam Pressure	0.07
PM2 Steam Box Temp	0.00
PM2 Steam Box Pressure (Kpa)	-0.06
PM2 Steam Box Flow (Klb/hr)	0.05
PM2 Mx Moist	-0.07
PM2 Caliper spread (CD)	-0.03

PM2 Moisture spread (CD)	0.18
PM2 Dry Weight spread (CD)	-0.05
PM2 Basis Weight spread (CD)	-0.11
PM2 Stuff Pump Consistency	0.08
PM2 Thick Stock flow (BDMT per Day)	-0.38
PM2 %Broke	0.27
PM2 Broke Flow(lpm)	0.00
PM2 Trim	0.01
PM2 Caliper	0.11
PM2 Porosity	0.28
PM2 MD Tensile	0.01
PM2 MD Tear	0.35
PM2 CD Tear	-0.14
PM2 Tear Ratio	-0.09
PM2 MD Elongation	-0.07
PM2 Burst	0.00
PM2 KSI	-0.09
PM2 Opacity	0.14
PM2 Brightness	0.50
PM2 a*	-0.14
PM2 b*	0.48
PM2 PPS Top	0.07
PM2 PPS Bottom	0.03
PM2 MD TEA	-0.01
PM2 MD TSI	-0.05
PM2 CD TSI	-0.08
PM2 TSI Ratio	0.00
PM2 TSO	-0.04
Jet to Wire	-0.07
l*	-0.60

Table 13 *Model summary for the baseline training data regression model*

R	R Square	Adjusted R Square	Std. Error of the Estimate
.96	.93	.93	.77

The above model was then used on the samples within the validation dataset to predict a value for the caliper. The predicted value of the caliper was then compared to the actual value of

PM2 Mx_Caliper for each sample. This comparison provided the below model summary in Table 14 with an R^2 value of .92. The R^2 value and the graph in Figure 13 show that the regression model created with the training dataset is proficient with the validation dataset and, thus, the model does not significantly overfit the data.

Table 14 *Model summary for the baseline training regression model on the baseline validation data*

R	R Square	Adjusted R Square	Std. Error of the Estimate
.96	.92	.92	.79

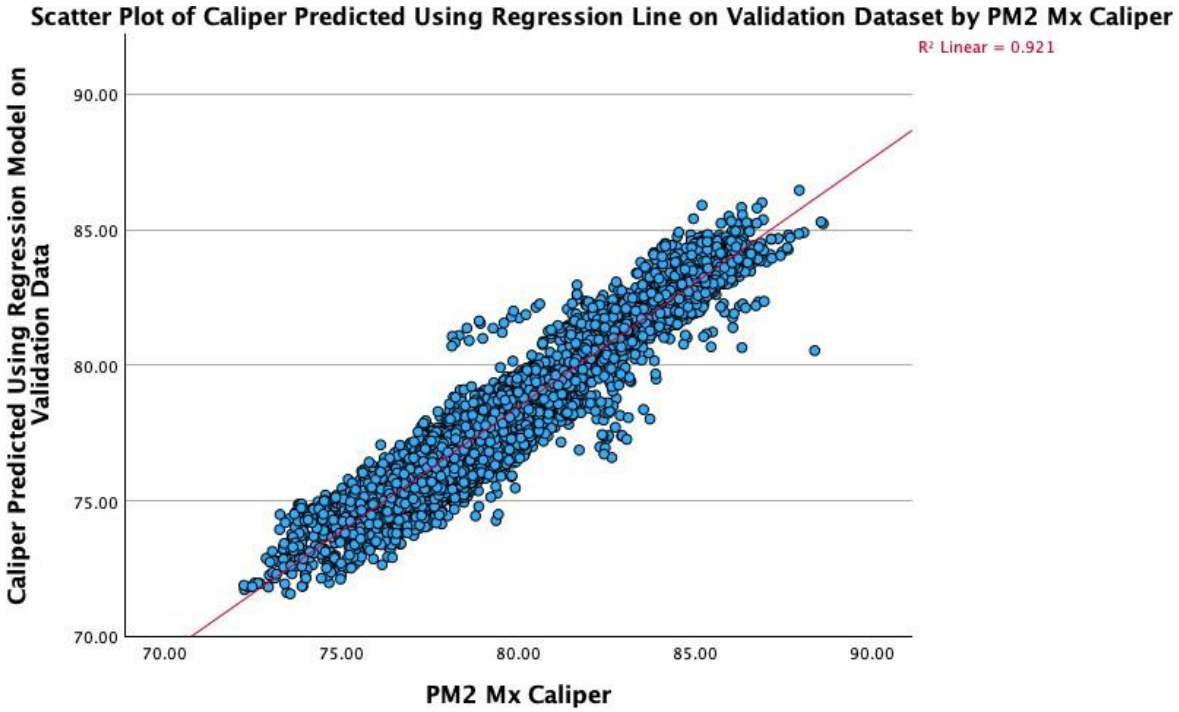


Figure 13 Scatterplot of the caliper predicted using the validation dataset vs. the actual caliper

Fir Dataset:

Next, regression analysis was run on the fir dataset. Figures 14-16 show a linear relationship, homoscedasticity of the residuals, and the residuals being drawn from a normal distribution with a mean close to zero and a standard deviation close to 1.

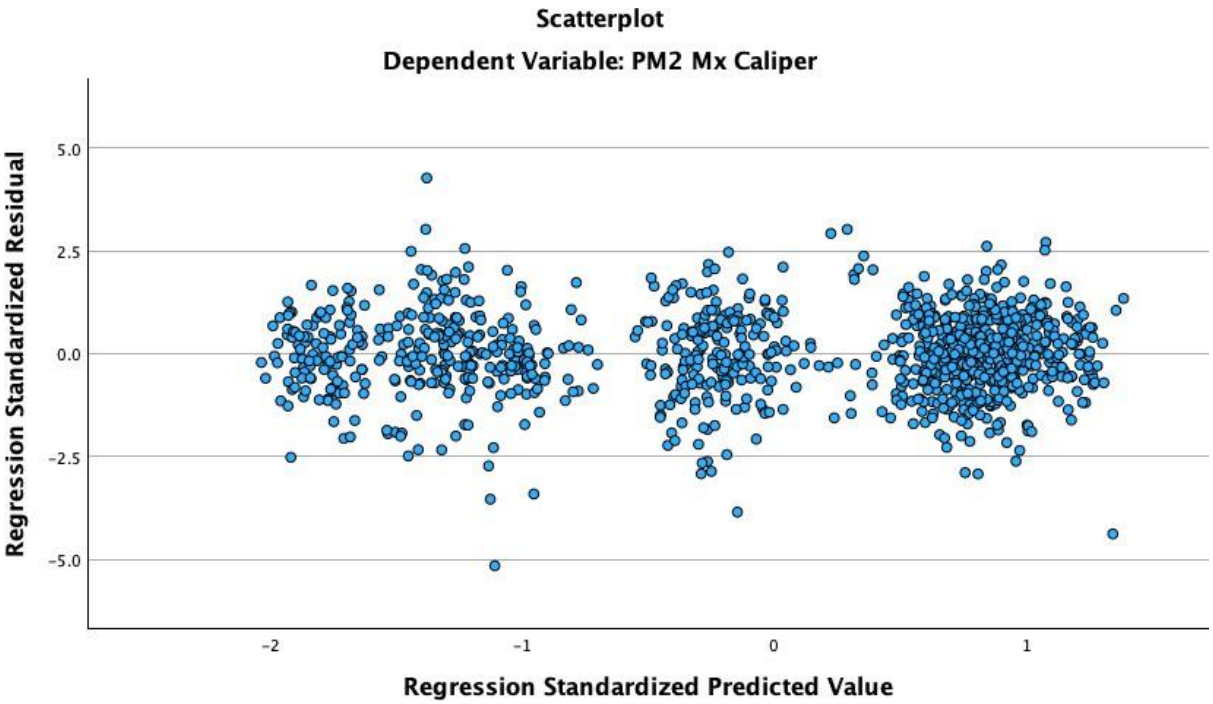


Figure 14 Scatterplot of the residual vs. predicted values for the fir dataset

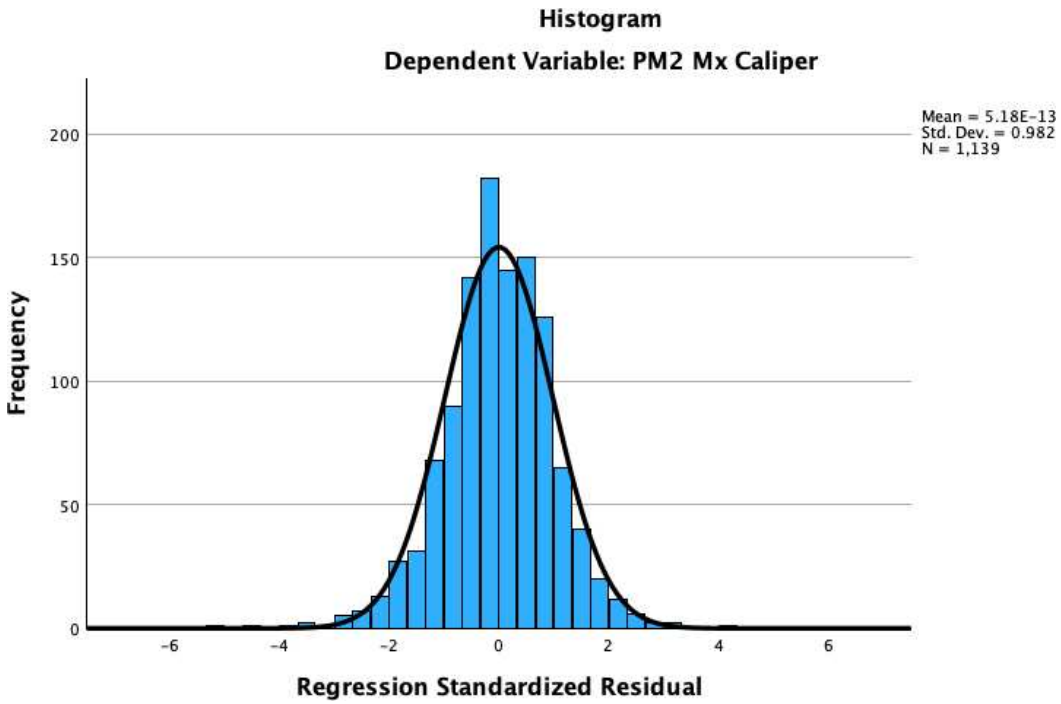


Figure 15 Histogram of the standardized residuals imposed with a normal curve for the fir dataset

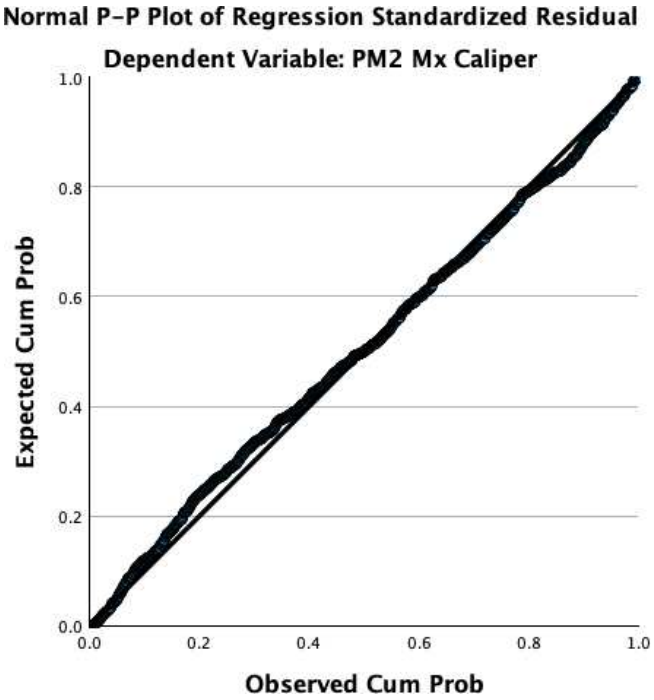


Figure 16 Normal probability plot of the residuals for the fir dataset

Table 15 below outlines the regression model coefficients that fit the fir dataset. Unlike the baseline dataset, the fir dataset has more variables with larger coefficients; these include PM2 KSI (-2.26), PM2 Burst (1.97), PM2 Basis Weight spread (CD) (-1.41), l* (-1.34), PM2 Brightness (.97), PM2 Dry Weight spread (CD) (.89). Table 16 shows that the model has an R-value close to 1 and an R^2 value of .98.

Table 15 *Standardized regression coefficients for a model on the fir data*

Variables	Standardized Coefficients
PM2 Stock flow (lpm)	-0.31
PM2 Basis Weight	0.08
PM2 Sheet Width	-0.11
PM2 Speed	0.79
PM2 Headbox Temp	-0.30
PM 2 Headbox pH	0.02
PM2 Main Steam Pressure	0.22
PM2 Steam Box Temp	-0.09
PM2 Steam Box Pressure (Kpa)	0.01
PM2 Steam Box Flow (Klb/hr)	0.01
PM2 Mx Moist	0.00
PM2 Caliper spread (CD)	-0.06
PM2 Moisture spread (CD)	0.74
PM2 Dry Weight spread (CD)	0.89

PM2 Basis Weight spread (CD)	-1.41
PM2 Stuff Pump Consistency	-0.02
PM2 Thick Stock flow (BDMT per Day)	0.29
PM2 %Broke	0.49
PM2 Broke Flow(lpm)	-0.31
PM2 Trim	-0.28
PM2 Caliper	0.21
PM2 Porosity	0.36
PM2 MD Tensile	0.16
PM2 MD Tear	-0.77
PM2 CD Tear	0.79
PM2 Tear Ratio	0.58
PM2 MD Elongation	-0.10
PM2 Burst	1.97
PM2 KSI	-2.26
PM2 Opacity	-0.12
PM2 Brightness	0.97
PM2 a*	0.08
PM2 b*	0.82
PM2 PPS Top	0.18
PM2 PPS Bottom	-0.13
PM2 MD TEA	-0.09
PM2 MD TSI	-0.59
PM2 TSI Ratio	0.14
PM2 TSO	0.41
Jet to Wire	-0.03
l*	-1.34

Table 16 *Model summary for the fir regression model*

R	R Square	Adjusted R Square	Std. Error of the Estimate
.99	.98	.98	.22

The ANOVA in Table 17 for the regression line has a significance level of less than .05, with the same hypothesis as the ANOVA for the baseline dataset. Thus, we reject the null hypothesis and conclude that the regression model predicts the caliper significantly differently from the mean model.

Table 17 *ANOVA results for the fir regression model*

	Sum of Squares	df	Mean Square	F	Sig.
Regression	2884.98	41	70.38	1508.07	<.001
Residual	51.58	1097	.05		
Total	2936.56	1138			

Spruce:

Finally, the same analysis was run using the spruce dataset. Figures 17-19 show a linear relationship, homoscedasticity of the residuals, and the residuals being drawn from a normal distribution with a mean close to zero and a standard deviation close to 1.

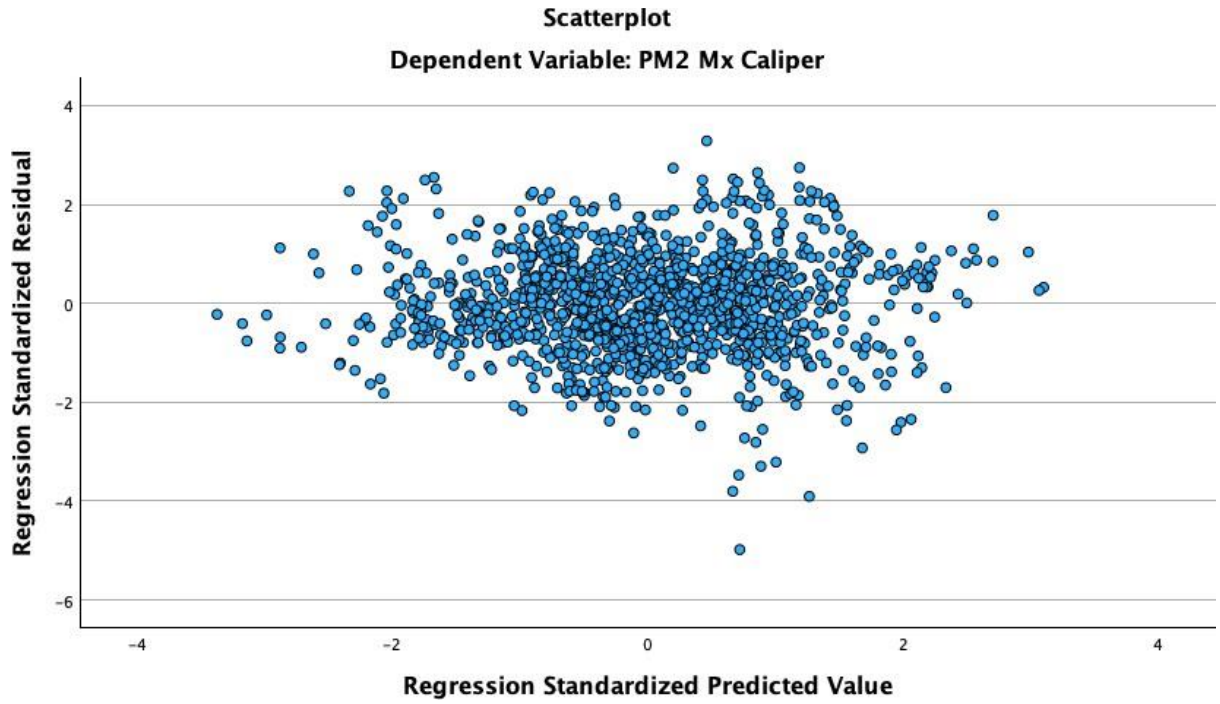


Figure 17 Scatterplot of the residual vs. predicted values for the spruce dataset

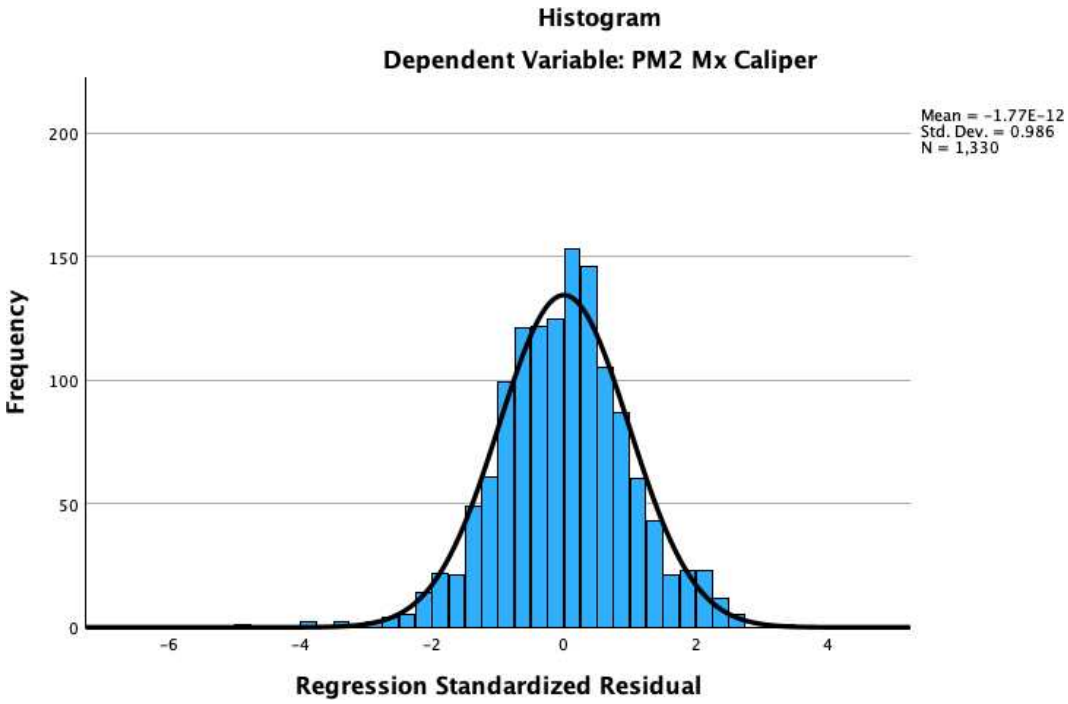


Figure 18 Histogram of the standardized residuals imposed with a normal curve for the spruce dataset

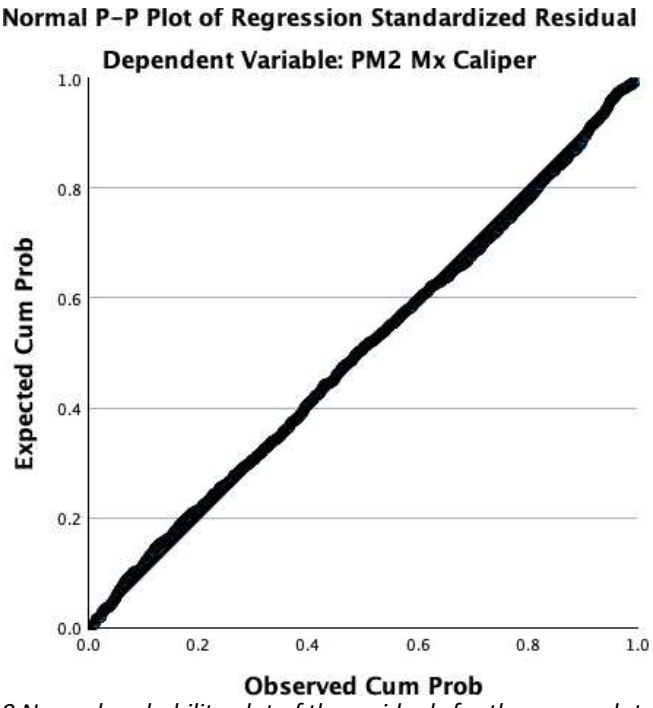


Figure 19 Normal probability plot of the residuals for the spruce dataset

The regression model coefficients are shown in Table 18 below. The following variables had large coefficients: PM2 Moisture spread (CD) (-2.87), PM2 Caliper spread (CD) (1.93), PM2

Dry Weight spread (CD) (1.86), PM2_Caliper (1.79), PM2 CD TSI (-1.77), PM2 a* (1.39), PM2 MD TSI (-1.20), PM2 Porosity (-1.13). The model summary in Table 19 indicates an R-value close to one and an R^2 value of .85.

Table 18 *Standardized regression coefficients for a model on the spruce data*

Variables	Standardized Coefficients
PM2 Basis Weight	0.34
PM2 Sheet Width	-0.07
PM2 Speed	0.40
PM2 Headbox Temp	-0.09
PM 2 Headbox pH	-0.01
PM2 Main Steam Pressure	0.29
PM2 Steam Box Pressure (Kpa)	-0.01
PM2 Steam Box Flow (Klb/hr)	0.04
PM2 Mx Moist	-0.03
PM2 Caliper spread (MD)	-0.67
PM2 Caliper spread (CD)	1.93
PM2 Moisture spread (CD)	-2.87
PM2 Dry Weight spread (CD)	1.86
PM2 Stuff Pump Consistency	-0.09
PM2 Thick Stock flow (BDMT per Day)	0.07
PM2 %Broke	-0.09
PM2 Broke Flow(lpm)	0.27
PM2 TMP flow (lpm)	0.14
PM2 Trim	-0.89
PM2 Caliper	1.79
PM2 Porosity	-1.13
PM2 MD Tensile	0.82
PM2 MD Tear	-0.32
PM2 Tear Ratio	-0.11
PM2 MD Elongation	0.02
PM2 Burst	0.21
PM2 a*	1.39
PM2 b*	0.62
PM2 PPS Top	-0.67
PM2 PPS Bottom	0.16

PM2 MD TEA	-0.90
PM2 MD TSI	-1.20
PM2 CD TSI	-1.77
PM2 TSO	-0.71
Jet to Wire	0.06
l*	-0.67

Table 19 *Model summary for the spruce regression model*

R	R Square	Adjusted R Square	Std. Error of the Estimate
.92	.85	.84	.30

The ANOVA was run with the same hypothesis as the other analysis, with the results in Table 20 below showing that the regression model statistically predicts the caliper differently from the mean model.

Table 20 *ANOVA results for the spruce regression model*

	Sum of Squares	df	Mean Square	F	Sig.
Regression	664.99	36	18.47	198.65	<.001
Residual	120.23	1293	.09		
Total	785.22	1329			

The regression model for the spruce dataset had the most variables with large coefficients compared to the other two datasets. The magnitude of the prominent coefficients in the spruce and fir datasets was more significant than those in the baseline dataset. No coefficients were identified as having a large magnitude in all three regression models. However, l* was close as it was identified in the models in the baseline and fir datasets and was the 12th largest in the spruce model. The regression models for the spruce and fir datasets had more variables arising as having large coefficients than the baseline dataset. In the regression models for the fir and spruce datasets, the only variable that arose as prominent in both was PM2 Dry Weight spread (CD).

3.4 Principal Component Analysis of Variables

PCA was run on all three datasets using a correlation matrix. The PCA results underwent

a varimax rotation to make interpreting the results easier. The varimax rotation maximizes the variance of the squared loading for each factor across the variables.

Baseline

Table 21 below shows that the first principal component accounts for 24.86% of the variance within the dataset. The second accounts for another 12.68%, and the third another 10.23%, for a total of 47.76% of the variance.

Table 21 *Total variance explained for the baseline dataset*

Component	Eigenvalues	% of Variance	Cumulative %
1	10.94	24.86	24.86
2	5.58	12.68	37.53
3	4.50	10.23	47.76
4	3.25	7.38	55.14
5	2.43	5.52	60.66
6	2.04	4.64	65.30
7	1.63	3.70	69.00
8	1.42	3.23	72.23
9	1.21	2.74	74.97
10	1.11	2.52	77.49
11	1.01	2.30	79.80
12	0.99	2.24	82.04
13	0.89	2.02	84.06
14	0.77	1.75	85.81
15	0.68	1.53	87.34
16	0.66	1.51	88.85
17	0.59	1.33	90.18
18	0.55	1.25	91.43
19	0.52	1.17	92.60
20	0.41	0.94	93.54
21	0.38	0.86	94.40
22	0.35	0.80	95.20
23	0.32	0.73	95.93
24	0.30	0.68	96.61
25	0.24	0.55	97.15
26	0.22	0.51	97.66
27	0.16	0.37	98.03
28	0.16	0.36	98.39
29	0.15	0.33	98.72
30	0.13	0.30	99.01
31	0.11	0.25	99.27

32	0.08	0.18	99.45
33	0.07	0.16	99.61
34	0.05	0.11	99.73
35	0.04	0.09	99.82
36	0.03	0.08	99.90
37	0.01	0.03	99.93
38	0.01	0.03	99.96
39	0.01	0.02	99.98
40	0.01	0.02	99.99
41	0.00	0.01	100.00
42	0.00	0.00	100.00
43	6.86E-5	.00	100.00
44	-8.14E-17	-1.85E-16	100.00

The eigenvalues from the principal components were then plotted on a scree plot. The scree plot in Figure 20 shows a change in elevation between 3-5 components. Using the results from Table 21 and Figure 20, plus the interest in visualizing the data, it was decided to keep three components for further analysis.

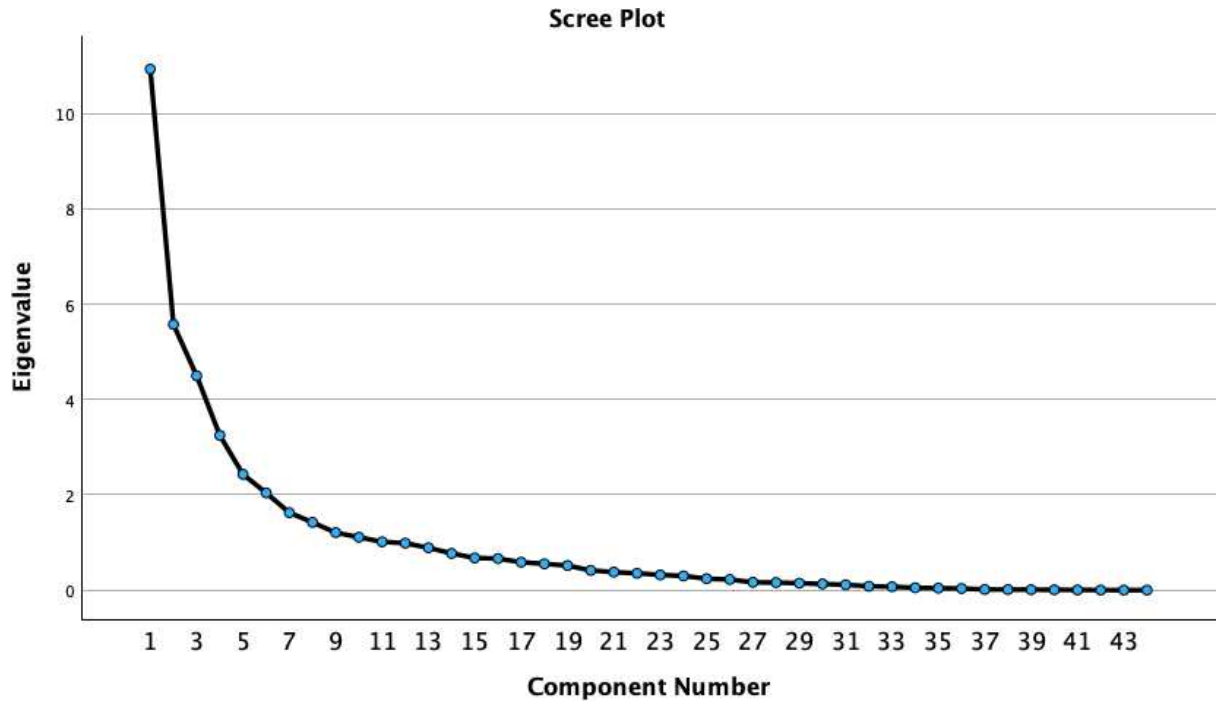


Figure 20 Scree plot depicting the eigenvalues of the principal components for the baseline dataset

Table 22 below shows each variable's loading scores on each component and rotated component, as well as the commonalities percentage of each variable when using the first three

components. With the first component explaining almost a quarter of the variance (24.86%), the variables with the highest loading on this component are of interest. As evident in the table below, the first ten variables weigh heavily on the first component and thus on the variation in the data. These variables include MD Tensile, MD Tea, Burst, KSI, MD Elongation, Basis Weight, Opacity, CD Tear, Porosity, and MD TSI. By examining the communalities column, using the first three components explains 58.9% of the variance in the PM2 Mx_Caliper variable.

Table 22 *PCA loadings and communalities for the baseline dataset*

	Component			Rotated Components			
Variables	1	2	3	1	2	3	Communalities
PM2 MD Tensile	0.94	0.11	0.05	0.87	0.22	-0.32	0.90
PM2 MD TEA	0.91	0.10	-0.03	0.81	0.19	-0.38	0.84
PM2 Burst	0.91	0.19	0.02	0.86	0.12	-0.33	0.86
PM2 KSI	0.90	0.26	0.13	0.92	0.08	-0.22	0.90
PM2 MD Elongation	0.79	0.07	-0.17	0.65	0.15	-0.47	0.66
PM2 Basis Weight	0.78	0.23	0.25	0.85	0.11	-0.06	0.73
PM2 Opacity	0.74	-0.07	-0.23	0.53	0.24	-0.51	0.60
PM2 CD Tear	0.69	0.37	0.26	0.83	-0.05	-0.01	0.68
PM2 Porosity	-0.68	0.19	0.09	-0.48	-0.37	0.37	0.50
PM2 MD TSI	0.61	-0.12	0.13	0.52	0.34	-0.14	0.40
PM2 MD Tear	0.57	0.52	0.24	0.77	-0.24	0.04	0.65
PM2 Main Steam Pressure	0.55	-0.07	0.16	0.50	0.28	-0.08	0.34
PM2 PPS Top	0.53	0.19	0.29	0.62	0.07	0.07	0.40
PM2 Caliper	0.52	0.29	0.34	0.68	-0.02	0.13	0.47
PM2 Mx Caliper	0.52	0.34	0.45	0.73	-0.04	0.24	0.59
PM 2 Headbox pH	-0.46	0.05	0.33	-0.26	-0.11	0.48	0.32
PM2 Steam Box Temp	-0.45	0.27	0.42	-0.14	-0.30	0.59	0.46
PM2 Mx Moist	0.06	0.02	0.03	0.07	0.01	0.00	0.00
PM2 %Broke	-0.38	0.79	-0.16	-0.07	-0.88	0.08	0.78
PM2 Broke Flow(lpm)	-0.32	0.77	-0.11	-0.02	-0.83	0.10	0.70
PM2 TMP flow (lpm)	0.51	-0.67	0.35	0.29	0.86	0.05	0.83
PM2 Stock flow (lpm)	0.51	-0.67	0.35	0.29	0.86	0.05	0.83
PM2 Thick Stock flow (BDMT per Day)	0.51	-0.67	0.35	0.29	0.86	0.05	0.83
PM2 Tear Ratio	0.05	0.55	0.04	0.27	-0.48	0.07	0.31
PM2 CD TSI	0.49	0.54	-0.02	0.63	-0.34	-0.16	0.54
PM2 TSI Ratio	-0.10	-0.53	0.10	-0.26	0.47	0.08	0.30
PM2 Stuff Pump Consistency	0.46	-0.50	-0.11	0.16	0.58	-0.33	0.47
PM2 Headbox Temp	0.41	-0.41	-0.02	0.19	0.51	-0.22	0.34
PM2 Basis Weight spread (CD)	0.16	0.40	0.02	0.30	-0.31	0.00	0.19

PM2 Dry Weight spread (CD)	0.26	0.40	0.02	0.39	-0.28	-0.05	0.23
PM2 Caliper spread (CD)	0.25	0.35	-0.03	0.34	-0.24	-0.09	0.18
PM2 Speed	0.04	-0.33	0.09	-0.07	0.33	0.04	0.12
PM2 Moisture spread (CD)	0.15	0.32	0.29	0.36	-0.18	0.24	0.21
PM2 Trim	-0.01	0.14	-0.08	0.02	-0.15	-0.06	0.02
PM2 a*	0.41	-0.02	-0.77	0.08	-0.03	-0.87	0.76
PM2 Steam Box Flow (Klb/hr)	-0.24	0.00	0.76	0.05	0.10	0.79	0.64
PM2 PPS Bottom	-0.27	0.20	0.62	0.06	-0.13	0.69	0.50
PM2 Steam Box Pressure (Kpa)	-0.58	-0.05	0.62	-0.30	0.01	0.79	0.72
l*	-0.15	0.03	0.60	0.08	0.07	0.61	0.38
Jet to Wire	-0.20	-0.06	0.49	-0.03	0.10	0.52	0.28
PM2 b*	0.09	-0.08	0.46	0.20	0.21	0.38	0.23
PM2 Brightness	-0.30	0.13	0.36	-0.08	-0.13	0.46	0.23
PM2 Sheet Width	0.06	0.14	-0.24	0.02	-0.16	-0.23	0.08
PM2 TSO	-0.10	-0.04	0.13	-0.06	0.03	0.15	0.03

Since three components were retained, a 3-D visualization can be created for the new space formed by the components as the axis, albeit hard to express on 2-D paper. Figure 21 below shows how the variables relate to the components. From the image below, it is interesting to see the variables clustered around PM2 Mx_Caliper, as they would have similar loading patterns that may exhibit variables with related patterns. However, by examining the figure and rotated loading scores from the table above, variables that appear to have similar loadings to PM2 Mx_Caliper in the rotated space include MD Tear and PM2_Caliper.

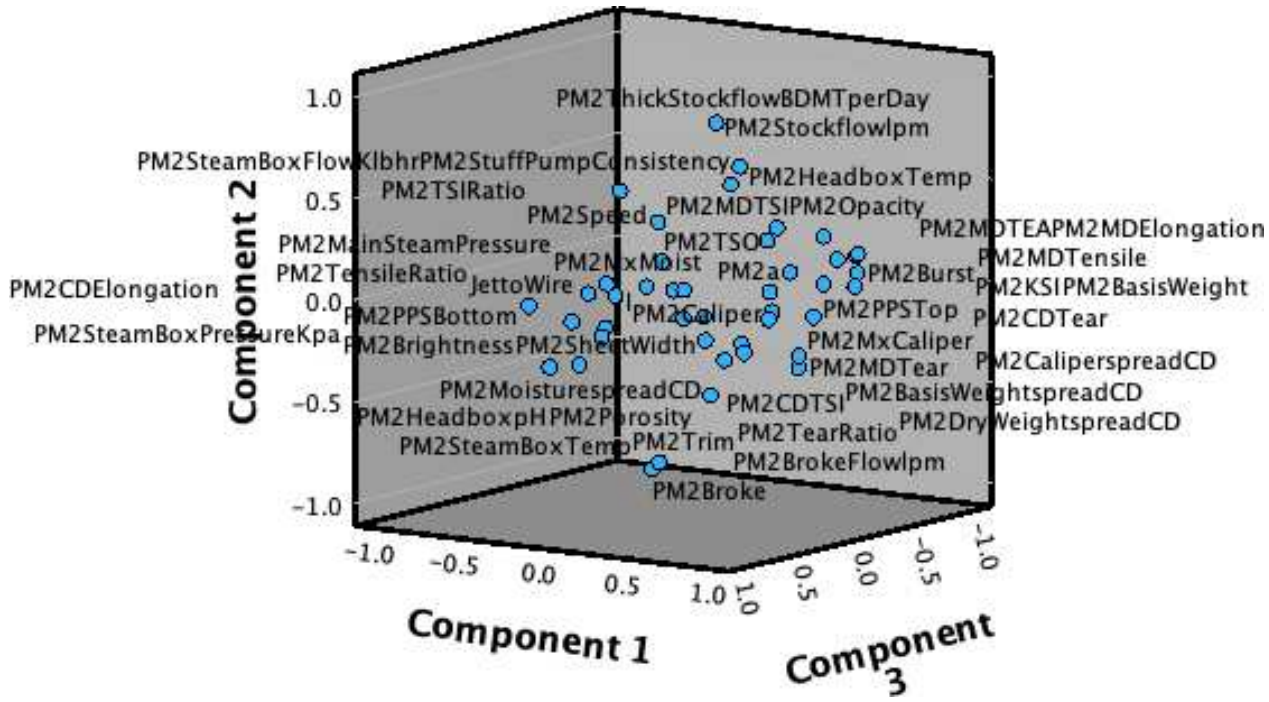


Figure 21 Component plot in the rotated space for the baseline dataset

Fir

PCA was then run on the fir dataset. Table 23 shows the eigenvalues, the percent of variance each component explains and the cumulative variance.

Table 23 *Total variance explained for the fir dataset*

Component	Eigenvalues	% of Variance	Cumulative %
1	17.51	39.79	39.79
2	7.29	16.56	56.35
3	4.41	10.02	66.36
4	2.16	4.92	71.28
5	1.95	4.43	75.71
6	1.77	4.03	79.74
7	1.50	3.40	83.14
8	1.07	2.43	85.57
9	1.04	2.35	87.93
10	0.91	2.07	89.99
11	0.68	1.55	91.54
12	0.54	1.24	92.77
13	0.51	1.16	93.93
14	0.45	1.03	94.96
15	0.39	0.88	95.84

16	0.33	0.75	96.59
17	0.32	0.72	97.31
18	0.22	0.50	97.81
19	0.19	0.43	98.24
20	0.15	0.35	98.59
21	0.13	0.30	98.89
22	0.11	0.26	99.14
23	0.09	0.19	99.33
24	0.06	0.13	99.47
25	0.05	0.10	99.57
26	0.04	0.09	99.66
27	0.04	0.08	99.74
28	0.03	0.06	99.80
29	0.02	0.05	99.85
30	0.02	0.04	99.89
31	0.01	0.03	99.92
32	0.01	0.03	99.95
33	0.01	0.02	99.97
34	0.01	0.01	99.98
35	0.00	0.01	99.99
36	0.00	0.00	99.99
37	0.00	0.00	100.00
38	0.00	0.00	100.00
39	0.00	0.00	100.00
40	0.00	0.00	100.00
41	0.00	0.00	100.00
42	6.60E-5	.00	100.00
43	9.58E-6	2.18E-5	100.00
44	1.20E-16	2.72E-16	100.00

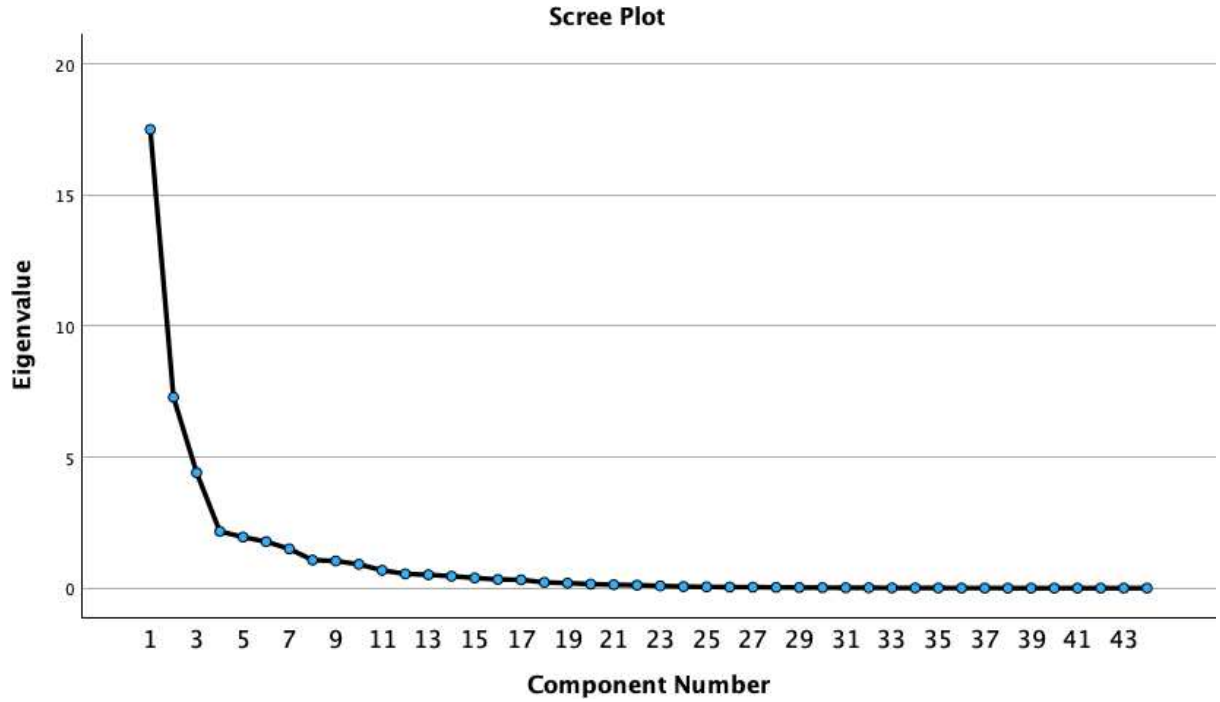


Figure 22 Scree plot depicting the eigenvalues of the principal components for the fir dataset

Using Table 23 and Figure 22 it is sufficient to represent the data with three principal components, which also allows for visualization in 3-D. The three components explain 66.35% of the total variance. The first component explains 39.79% of the variance and has many more variables than the baseline dataset that weigh heavily on this component. The first component also includes more variables with a heavy negative weight than the baseline data. 86.9% of PM2 Mx_Caliper's variance can be explained in the fir dataset, compared to only 58.9% in the baseline dataset.

Table 24 *PCA loadings and communalities for the fir dataset*

Variables	Component			Rotated Components			Communalities
	1	2	3	1	2	3	
PM2 Speed	0.97	-0.12	0.08	0.90	-0.37	-0.01	0.95
PM2 CD TSI	-0.96	0.14	-0.12	-0.92	0.34	0.05	0.96
PM2 TSI Ratio	0.95	-0.01	0.21	0.95	-0.21	0.03	0.95
PM2 Sheet Width	-0.94	0.06	-0.22	-0.94	0.22	0.03	0.94
PM2 Trim	-0.90	-0.07	-0.36	-0.97	0.04	-0.03	0.95
PM2 MD Tear	-0.89	-0.13	-0.12	-0.87	0.21	-0.18	0.83
PM2 Headbox Temp	0.87	0.32	0.18	0.88	-0.06	0.33	0.90
PM2 Main Steam Pressure	0.86	0.07	0.05	0.81	-0.27	0.16	0.75

PM2 KSI	-0.85	0.22	0.38	-0.61	0.74	-0.07	0.92
PM2 MD Elongation	-0.84	0.18	0.29	-0.63	0.65	-0.06	0.83
PM2 Thick Stock flow (BDMT per Day)	0.83	0.48	-0.04	0.76	-0.16	0.56	0.92
PM2 TMP flow (lpm)	0.83	0.48	-0.04	0.76	-0.16	0.56	0.92
PM2 Stock flow (lpm)	0.83	0.48	-0.04	0.76	-0.16	0.56	0.92
PM2 Burst	-0.82	0.26	0.35	-0.59	0.72	-0.01	0.87
PM2 Mx Caliper	0.80	-0.48	-0.08	0.66	-0.58	-0.30	0.87
PM2 Tear Ratio	-0.75	-0.17	-0.45	-0.88	-0.14	-0.07	0.80
PM2 Basis Weight spread (CD)	0.75	0.11	0.39	0.85	0.06	0.03	0.73
PM2 CD Tear	-0.74	-0.03	0.38	-0.52	0.59	-0.28	0.69
PM2 Dry Weight spread (CD)	0.72	-0.02	0.43	0.83	0.05	-0.10	0.70
PM2 MD TEA	-0.71	0.30	0.50	-0.42	0.82	-0.03	0.83
PM2 Caliper	0.71	-0.45	-0.01	0.61	-0.48	-0.31	0.70
PM2 Porosity	0.70	-0.53	-0.33	0.47	-0.77	-0.26	0.88
PM2 PPS Bottom	0.68	-0.28	-0.38	0.44	-0.70	-0.01	0.69
PM2 MD TSI	0.57	0.52	0.43	0.73	0.34	0.37	0.78
PM2 a*	-0.53	0.39	0.09	-0.42	0.45	0.25	0.43
PM 2 Headbox pH	-0.47	0.25	0.03	-0.40	0.32	0.16	0.29
PM2 Caliper spread (CD)	0.23	0.10	-0.10	0.17	-0.14	0.16	0.07
PM2 Basis Weight	0.03	-0.02	0.01	0.03	-0.01	-0.02	0.00
PM2 %Broke	0.06	-0.93	0.16	0.07	-0.29	-0.90	0.89
PM2 Broke Flow(lpm)	0.16	-0.91	0.19	0.17	-0.30	-0.88	0.89
PM2 b*	0.08	0.80	-0.32	-0.01	0.05	0.86	0.75
PM2 Stuff Pump Consistency	-0.03	0.79	-0.31	-0.11	0.09	0.84	0.72
l*	0.41	0.76	-0.31	0.29	-0.10	0.87	0.85
PM2 Moisture spread (CD)	0.16	0.62	0.48	0.38	0.59	0.39	0.64
PM2 Brightness	0.54	0.55	-0.23	0.43	-0.17	0.66	0.65
PM2 TSO	-0.22	-0.37	0.12	-0.17	0.02	-0.41	0.20
PM2 Steam Box Flow (Klb/hr)	0.21	0.31	-0.27	0.10	-0.17	0.41	0.21
PM2 MD Tensile	-0.32	0.41	0.75	0.05	0.91	0.02	0.84
PM2 Opacity	0.18	-0.27	0.71	0.44	0.39	-0.51	0.61
Jet to Wire	-0.02	0.12	0.52	0.21	0.48	-0.11	0.29
PM2 Steam Box Temp	-0.38	0.24	-0.38	-0.49	-0.06	0.33	0.35
PM2 PPS Top	-0.08	0.33	-0.36	-0.20	-0.13	0.43	0.25
PM2 Mx Moist	0.00	0.00	-0.11	-0.04	-0.09	0.05	0.01
PM2 Steam Box Pressure (Kpa)	-0.03	-0.01	0.07	0.00	0.06	-0.04	0.01

Figure 23 below shows the variable loadings plotted in the rotated component space. Like the baseline data, PM2_Caliper appears to be the variable most clustered by PM2 Mx_Caliper.

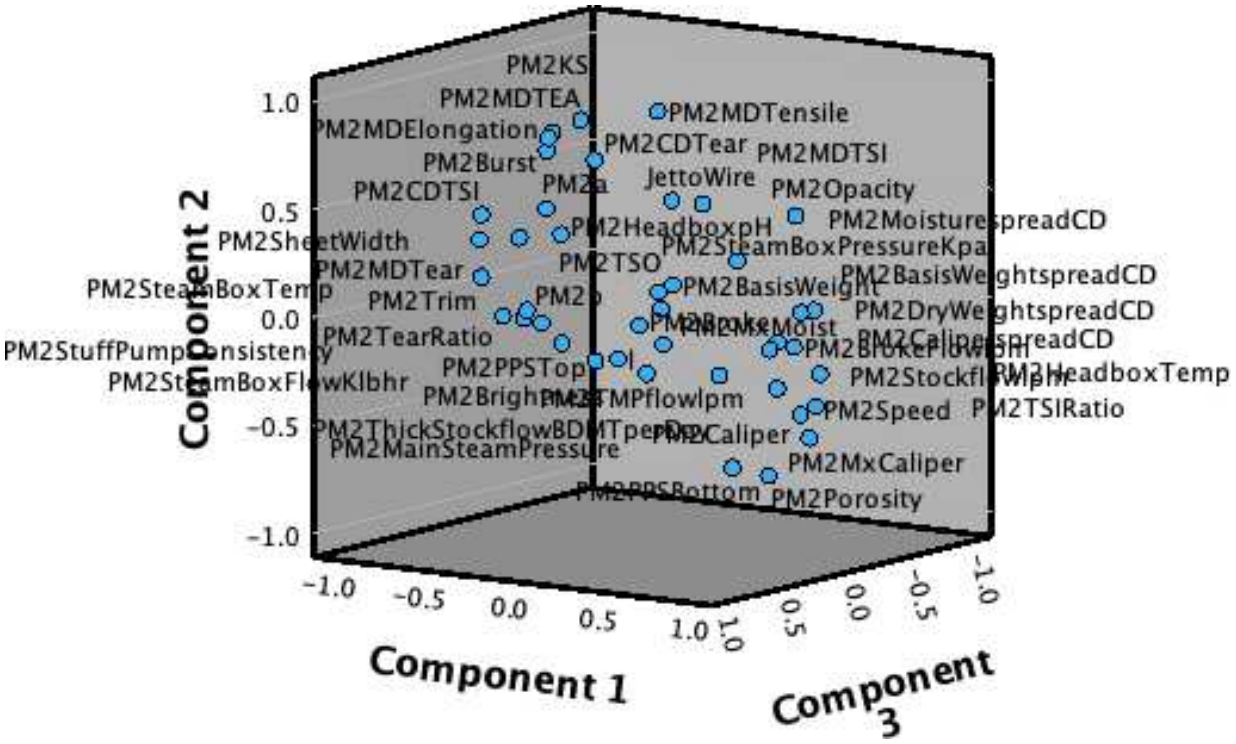


Figure 23 Component plot in the rotated space for the fir dataset

Spruce

Lastly, principal component analysis was run on the spruce dataset. Table 25 shows the total variance explained by each component. With the eigenvalues then being plotted on the scree plot. Three components are sufficient for the analysis, as they can be visualized and explain 52.58% of the total variance.

Table 25 *Total variance explained for the spruce dataset*

Component	Eigenvalues	% of Variance	Cumulative %
1	9.59	22.31	22.31
2	7.67	17.84	40.15
3	5.35	12.43	52.58
4	3.23	7.50	60.08
5	3.05	7.08	67.17
6	2.15	4.99	72.16
7	2.08	4.83	76.99
8	1.89	4.39	81.38
9	1.39	3.23	84.61
10	1.15	2.67	87.28
11	1.01	2.35	89.63
12	0.77	1.80	91.43

13	0.67	1.55	92.98
14	0.50	1.16	94.14
15	0.41	0.96	95.10
16	0.38	0.87	95.97
17	0.30	0.69	96.66
18	0.27	0.64	97.29
19	0.22	0.51	97.80
20	0.17	0.40	98.20
21	0.16	0.37	98.57
22	0.13	0.31	98.89
23	0.11	0.25	99.13
24	0.09	0.20	99.33
25	0.09	0.20	99.53
26	0.06	0.15	99.68
27	0.05	0.11	99.79
28	0.03	0.08	99.87
29	0.02	0.06	99.92
30	0.02	0.04	99.96
31	0.01	0.02	99.99
32	0.00	0.01	99.99
33	0.00	0.00	100.00
34	0.00	0.00	100.00
35	0.00	0.00	100.00
36	.000	.001	100.00
37	1.35E-14	3.14E-14	100.00
38	5.60E-15	1.30E-14	100.00
39	1.16E-15	2.69E-15	100.00
40	3.64E-16	8.47E-16	100.00
41	6.31E-17	1.47E-16	100.00
42	-3.93E-16	-9.13E-16	100.00
43	-9.27E-16	-2.16E-15	100.00

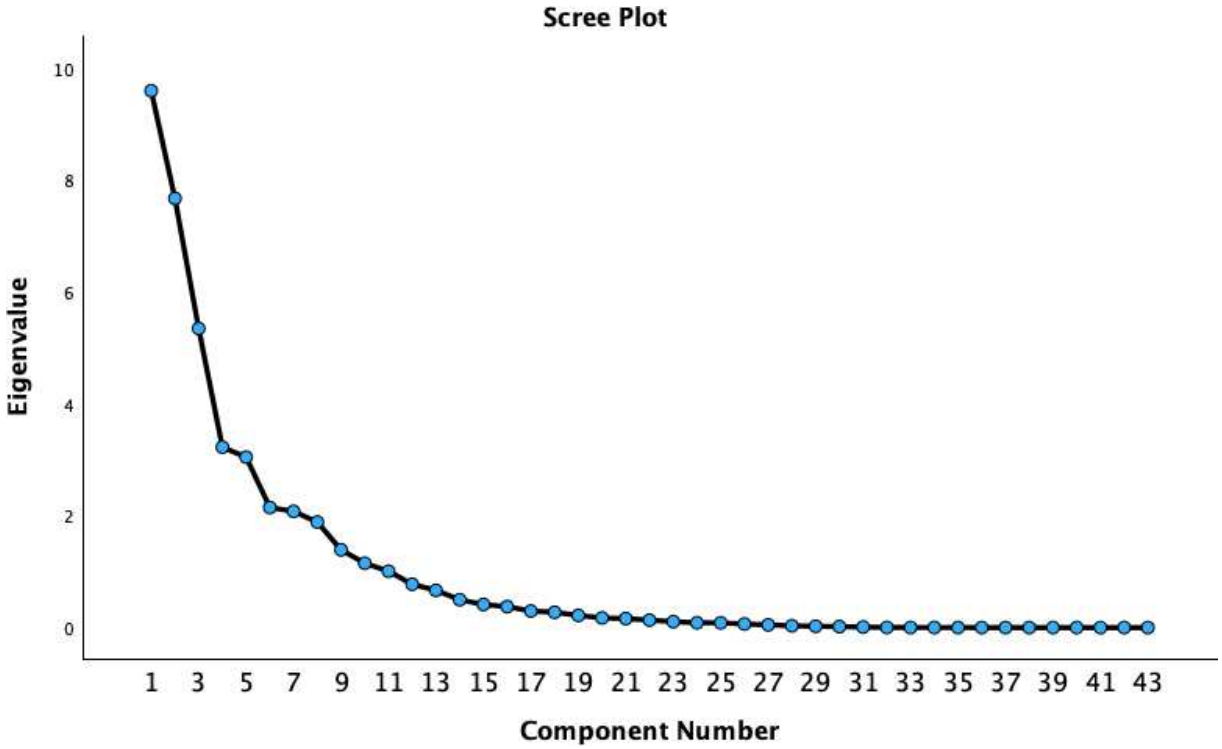


Figure 24 Scree plot depicting the eigenvalues of the principal components for the spruce dataset

In the spruce dataset, the first component explains 22.31% of the total variance. Similar to the baseline dataset, 11 variables heavily influence the first component. Of interest is that the top weightings of the first four variables are negative. These three components account for 46.7% of the variance in the PM2 Mx_Caliper variable.

Table 26 *PCA loadings and communalities for the spruce dataset*

	Component			Rotated Components			
Variables	1	2	3	1	2	3	Communalities
PM2 Brightness	-0.92	0.15	0.15	-0.90	0.10	-0.25	0.89
PM2 %Broke	-0.88	0.34	-0.17	-0.95	0.11	0.13	0.92
l*	-0.88	0.23	0.14	-0.87	0.16	-0.19	0.84
PM2 Broke Flow(lpm)	-0.87	0.38	-0.14	-0.94	0.15	0.12	0.91
PM2 Thick Stock flow (BDMT per Day)	0.82	0.01	0.35	0.84	0.27	-0.12	0.80
PM2 Stock flow (lpm)	0.82	0.01	0.35	0.84	0.27	-0.12	0.80
PM2 TMP flow (lpm)	0.82	0.01	0.35	0.84	0.27	-0.12	0.80
PM2 CD Tear	0.76	-0.09	-0.20	0.73	-0.09	0.29	0.62
PM2 MD TEA	0.68	0.30	-0.30	0.56	0.18	0.54	0.64
PM2 Stuff Pump Consistency	0.67	-0.17	-0.05	0.69	-0.09	0.10	0.48
PM2 MD Elongation	0.66	0.35	-0.08	0.56	0.34	0.37	0.56
PM2 MD Tensile	0.53	0.29	-0.50	0.39	0.05	0.68	0.62

PM2 KSI	0.53	-0.41	-0.27	0.56	-0.43	0.15	0.52
PM2 Trim	0.49	0.29	-0.03	0.40	0.29	0.26	0.33
PM2 TSO	0.48	0.27	-0.16	0.39	0.20	0.37	0.33
PM2 Dry Weight spread (CD)	0.40	0.30	0.29	0.35	0.45	-0.03	0.33
PM2 Steam Box Flow (Klb/hr)	0.30	-0.28	-0.26	0.32	-0.34	0.15	0.23
PM2 Basis Weight	-0.11	-0.04	0.08	-0.09	-0.01	-0.11	0.02
PM2 Mx Moist	-0.07	-0.02	0.06	-0.06	0.00	-0.08	0.01
PM2 PPS Bottom	-0.21	-0.85	0.22	-0.01	-0.65	-0.64	0.82
PM2 Opacity	0.22	0.85	-0.03	0.04	0.74	0.47	0.77
PM2 Tear Ratio	-0.28	0.82	-0.11	-0.45	0.62	0.43	0.77
PM2 Speed	-0.12	0.78	0.22	-0.24	0.76	0.16	0.66
PM2 TSI Ratio	0.15	-0.66	0.54	0.35	-0.28	-0.75	0.76
PM2 Main Steam Pressure	-0.09	0.65	-0.05	-0.22	0.53	0.33	0.44
PM2 PPS Top	-0.23	-0.62	-0.11	-0.12	-0.61	-0.25	0.45
PM2 Headbox Temp	0.43	0.59	0.16	0.32	0.64	0.23	0.56
PM2 MD Tear	0.36	0.59	-0.26	0.20	0.42	0.58	0.55
PM2 Burst	0.34	-0.51	-0.25	0.41	-0.53	0.04	0.44
PM2 Caliper spread (CD)	-0.16	0.41	0.24	-0.20	0.46	-0.05	0.25
PM2 Caliper spread (MD)	-0.35	0.37	0.11	-0.40	0.33	0.01	0.27
PM2 b*	-0.01	0.25	-0.04	-0.05	0.19	0.15	0.06
PM2 CD TSI	-0.17	0.41	-0.81	-0.36	-0.08	0.86	0.86
PM 2 Headbox pH	0.03	-0.13	-0.73	-0.03	-0.48	0.57	0.55
PM2 Moisture spread (CD)	0.33	0.49	0.66	0.30	0.79	-0.27	0.78
PM2 Mx Caliper	-0.14	0.12	0.66	-0.08	0.42	-0.54	0.47
PM2 Porosity	-0.07	-0.48	0.64	0.11	-0.10	-0.79	0.65
PM2 MD TSI	-0.04	-0.51	-0.57	-0.01	-0.72	0.23	0.58
PM2 a*	0.31	0.12	-0.53	0.20	-0.13	0.57	0.39
PM2 Basis Weight spread (CD)	0.31	0.45	0.50	0.27	0.67	-0.15	0.55
PM2 Caliper	-0.15	-0.06	0.44	-0.08	0.15	-0.44	0.22
PM2 Sheet Width	0.05	0.24	0.26	0.03	0.34	-0.10	0.13
PM2 Steam Box Pressure (Kpa)	-0.02	-0.02	0.04	-0.01	0.00	-0.05	0.00

Figure 25 below shows the variable loadings in the rotated component space for the spruce trial. Like the other two datasets, PM2_Caliper clusters are nearest to PM2 Mx_Caliper.

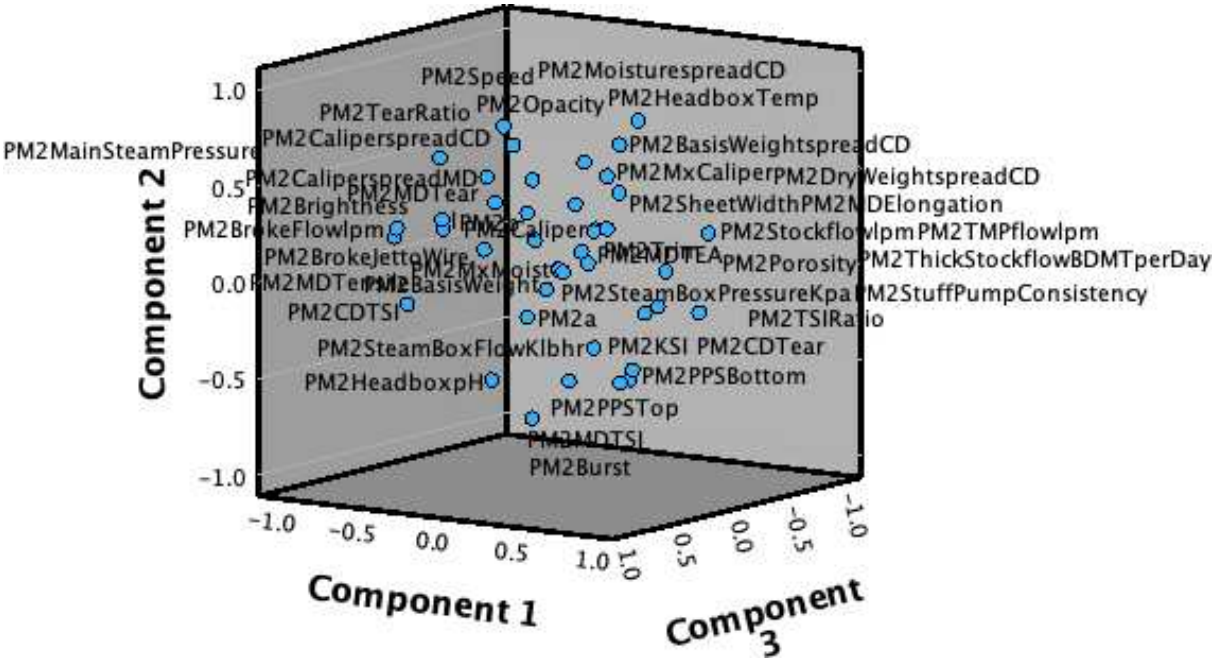


Figure 25 Component plot in the rotated space for the spruce dataset

The results can be easily compared since three components were retained in each analysis. Firstly, keeping three components in the fir dataset explained 20-40% more of the variance in the caliper than in the other two datasets. Also, in the fir dataset, the first component accounted for the most variance compared to the other two. While comparing the first component of the three datasets, the PM2 Mx_Caliper had the largest loading factor in the fir dataset. MD Tea and CD Tear appeared significant among all three datasets in the first components.

3.5 Neural Network Analysis of Variables

The final analysis tool used on the three datasets was neural networks. Several neural networks were trained to predict the machine caliper (response) based on the other variables (predictor), just as with linear regression. The network shape was chosen to allow visualization of how the neural network transforms the input space to obtain the prediction.

The number of predictor variables used was restricted to variables that appeared as easily controlled or arose as interesting in the other analysis, which include the following: 'PM2

Caliper', 'PM2 Stock flow (lpm)', 'PM2 Basis Weight', 'PM2 Sheet Width', 'PM2 Speed',' PM2 Headbox Temp', 'PM 2 Headbox pH', 'PM2 Main Steam Pressure', 'PM2 Steam Box Temp', 'PM2 Steam Box Pressure (Kpa)', 'PM2 Steam Box Flow (Klb/hr)',' PM2 Thick Stock flow (BDMT per Day)', 'PM2 Broke Flow(lpm)', 'PM2 MD Tensile', 'PM2 CD Tear', 'PM2 KSI',' PM2 Dry Weight spread (CD)',' PM2 Burst'.

The three datasets (Fir, Spruce, and Baseline) were randomly pooled and partitioned into training/validation sets. Four neural networks, N_{16}, N_8, N_4 and N_2, were trained to predict the PM2 Mx_Caliper from the remaining variables. The four networks have different shapes and parameter complexity: network N_k has k+3 layers, with an input layer of size 19 (the number of variables), then k layers each of size k, followed by a layer of size 2, then an output layer of size 1. For example, the sequence of layer sizes for N_{16} is (19,16,16,16,16, 16,16,16,16, 16,16,16,16, 16,16,16,16, 2, 1). A layer of size two was used for the second-to-last layer to visualize how the network transforms the input space before making the final prediction. All networks used rectified linear units (ReLU) as the activation function. The loss function used was the mean squared error between y and y_p. Like in the regression section, the validation data was used to ensure the model trained was not overfitting.

In all four network models, N_{16}, N_8, N_4 and N_2, the second last layer outputs a vector X_{-2} of dimension 2. Therefore, for each of these network models, N, the second last step in computing y=N(X) for any input point X in the input can be derived. The point X_{-2} results from transforming the original point X in an n-dimensional space down to a point in 2-space. Because X_{-2} is a vector with two coordinates, it can be visualized as a point in the plane. The collection of all such points X_2 for each point in the training sample is a visualization of how the neural network has transformed the original points X in R^n to points X_{-2} in the plane as part of the learned process for predicting y. Each point in this visualization is colour-coded to incorporate

additional information. The results for N_{16} are shown and discussed below. The visualizations relating to the smaller networks (N_8, N_4 and N_2) are included in Appendix B for comparison.

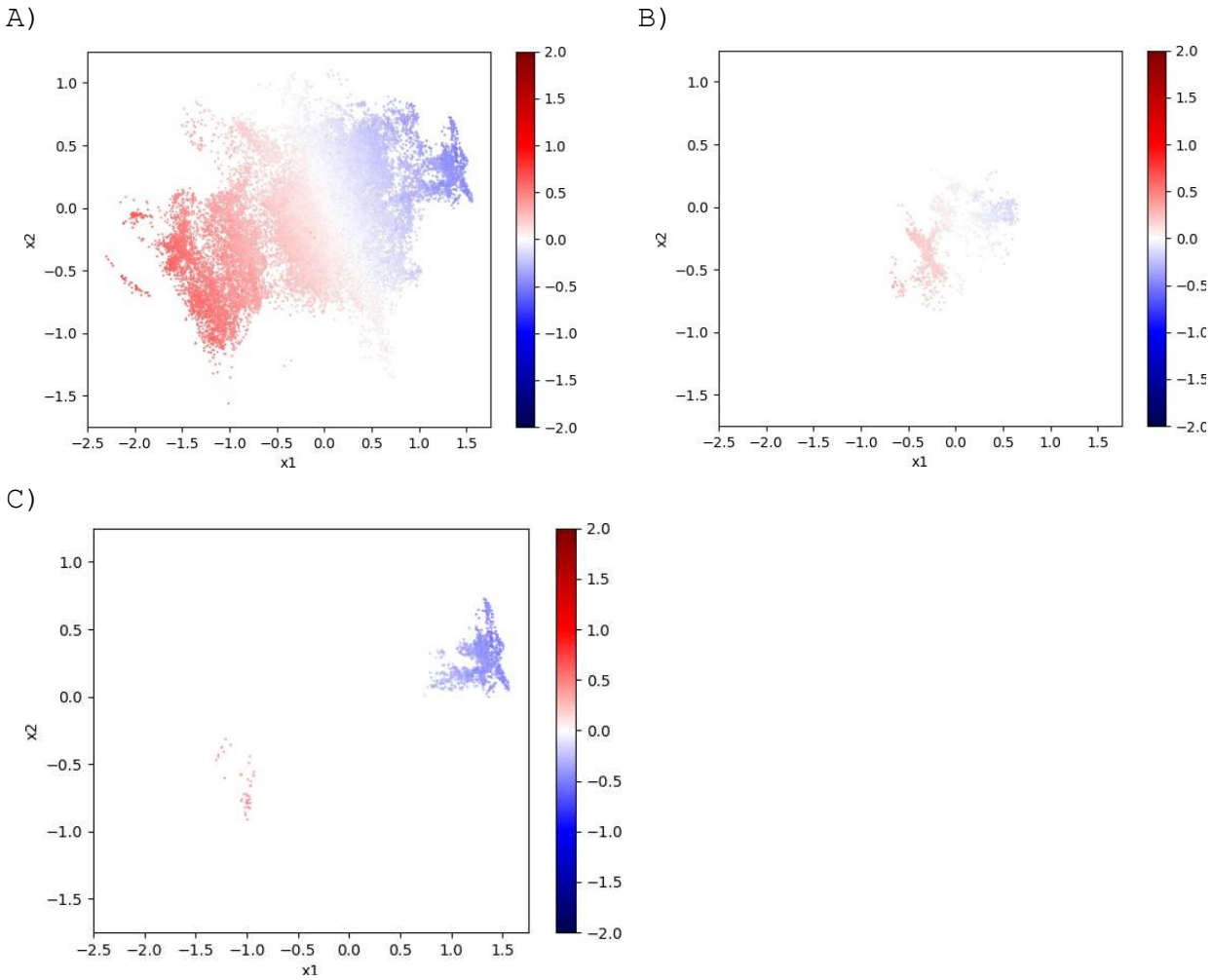


Figure 26 Neural Network transformation of the actual machine caliper in the given state for the A) Baseline, B) Fir, and C) Spruce datasets

Figure 26 above is a visualization of the points X_{-2}, resulting from transforming each point X in the input. Each point in the visualization is coloured according to PM2 Mx_Caliper for the three datasets, Baseline (Figure 26A), Fir (Figure 26B), and Spruce (Figure 26C). The almost perfectly smooth transition of colours indicates that the model has transformed the original space X down to R^2 so that states with the same output caliper are grouped along parallel lines.

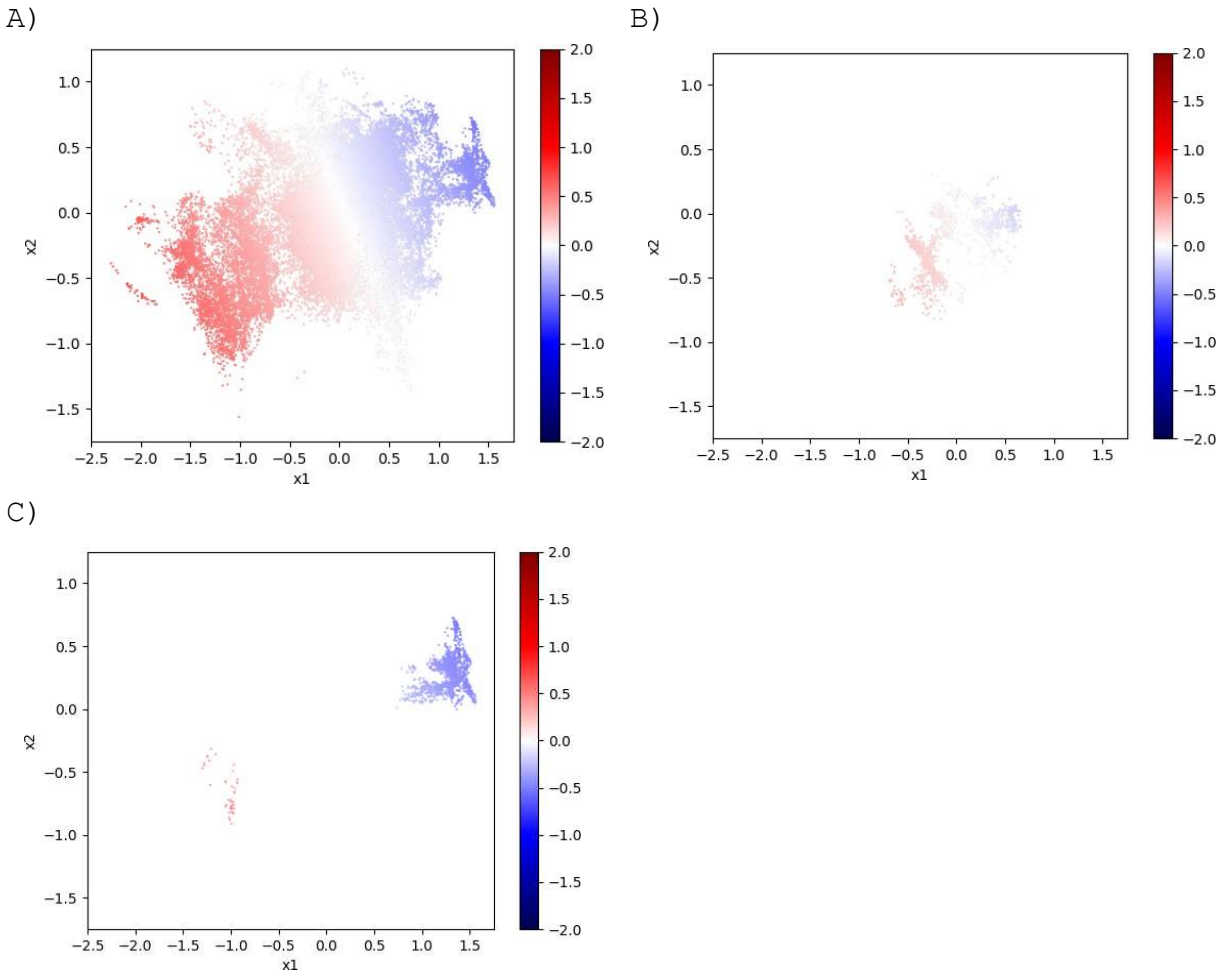


Figure 27 Neural network visualization of the predicted caliper for A) Baseline, B) Fir, and C) Spruce

Next, consider Figure 27, where the colour of each point X_{-2} is the caliper predicted by the network $y_{pred} = N(X)$ for the three datasets. These images show a perfect gradient transition from dark red in the bottom right corner to dark blue in the upper right. This is because the following (and final) network layer outputs the prediction y_{pred} using an affine transformation. The level curves of this affine transformation are parallel lines. Notice that Figure 27 looks very similar to Figure 26, which indicates that the model's prediction y_p for the caliper will be very close to the actual value PM2 Mx_Caliper. This is consistent with the model R^2 value being very close to 1.

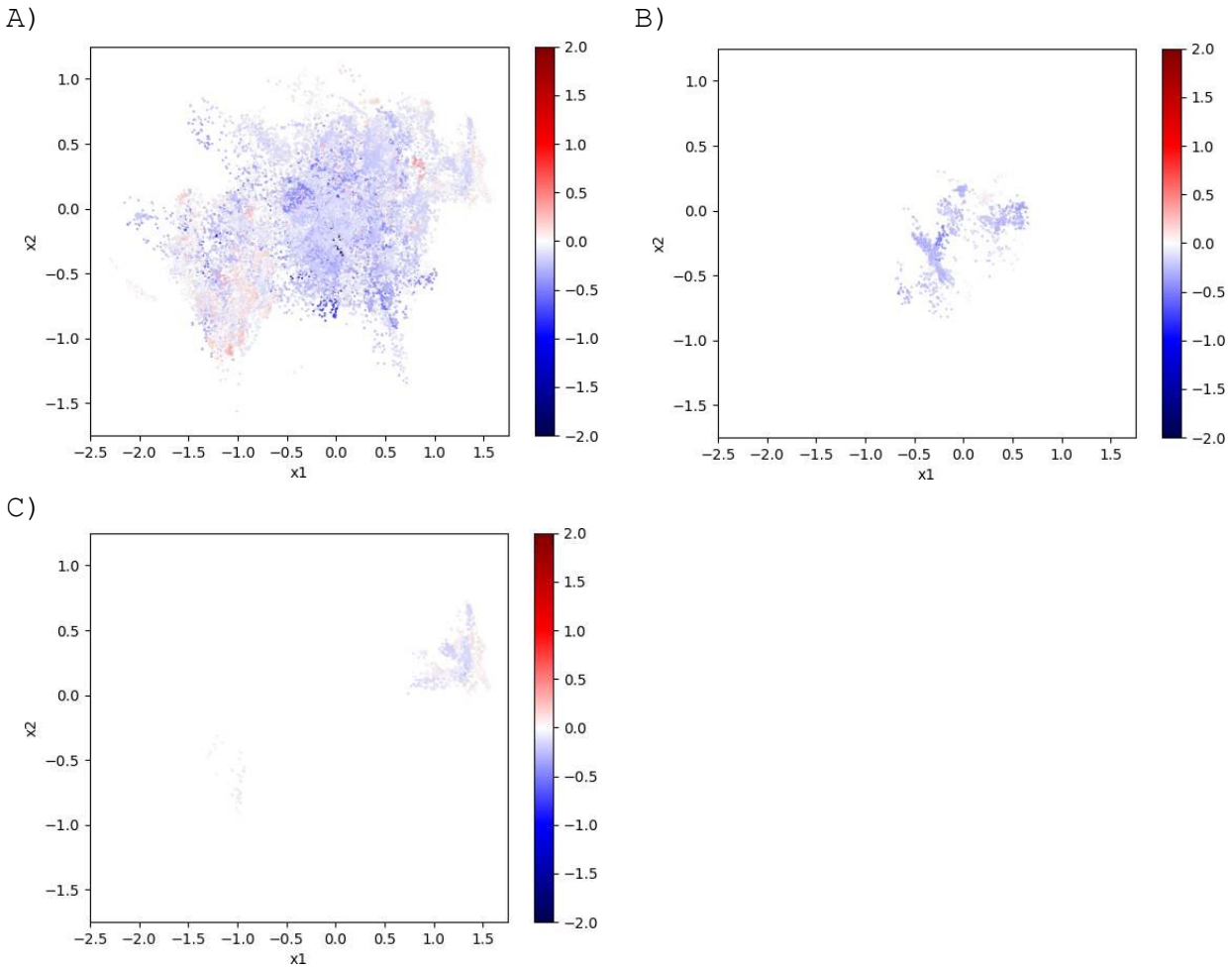


Figure 28 Visualization of how well the machine caliper matches the desired caliper for the A) Baseline, B) Fir, and Spruce

Figure 28 shows the data as in the previous examples, but now points are coloured using the value of the difference PM2_Caliper-PM2 Mx_Caliper. Dark red means the desired caliper was two standard deviations larger than the actual, and dark blue means the desired caliper was two standard deviations less. Darker colours indicate sample points where the output paper caliper did not align closely with the desired caliper. Observe that, as a rule, the PM2 _Caliper is always smaller than the PM2 Mx_Caliper, as seen by Figure 28(A), being predominantly blue in colour, and that is especially true in the Fir dataset. This aligns with what was seen earlier in this

thesis. Also note Figure 28C, representing the spruce dataset, is a lighter blue colour compared to the fir, which is consistent with the findings of the regression analysis earlier.

Appendix B contains the diagrams above for the remaining three neural networks. It is worth mentioning that the training of a neural network results in a local minimum of an error function, and there is no unique local minimum. A different training session on the same dataset might result in a different local minimum which would result in different diagrams for the same dataset. This is also why the direction of gradient change for the different networks is not always the same. However, there is a smooth gradient of parallel lines in all cases because the final mapping is an affine map from R^2 to R.

The final layer of the neural network takes a point in layer X_{-1} to the final prediction. For $N_{16,}$ the MSE of this prediction was 0.0182, indicating less than a 2 percent error in predicting the (normalized, squared) PM2 Mx_Caliper from the machine state. Table 27 below shows the MSE of the residuals for the other three neural networks. The lowering of the MSE by using larger networks (more training parameters) on the same data indicates there are beneficial nonlinear relationships between the predictor and predicted variables.

Table 27 *Mean squared error values for the remaining neural networks*

Networks	MSE of the residuals
N_8	0.002
N_4	0.004
N_2	0.005

In conclusion, feed-forward neural networks proved an effective tool for predicting machine calipers from the array of variables used. They provided an effective tool for visualizing the high-dimensional dataset.

Chapter 4: CONCLUSIONS

Given three datasets from Corner Brook Pulp and Paper, the above analysis was completed to better understand relationships between variables within the paper-making process using various analytical and predictive tools.

The analyses showed that the difference between desired and measured caliper in the fir dataset varied less than in the other two datasets, but the actual machine caliper was consistently different from the desired caliper. An ANOVA showed a statistically significant effect of species on the caliper difference. These statistics showed that the spruce dataset had the lowest average difference between the desired and actual caliper. Of all three datasets, the fir dataset had the most variables throughout all the analyses that showed significance relative to the caliper. When looking solely at this analysis, using only fir wood would be advantageous to paper makers if the machine caliper could be controlled to be more in line with the desired caliper; otherwise, the spruce dataset appears beneficial, as the machine caliper was closest to the desired caliper more than the other datasets. These findings agree with other literature, such as Drost et al. (2003), that species influences paper quality.

With more control over the variables than the species mixture, it can be seen that certain variables may influence the process. In the baseline dataset, basis weight was the only variable that showed an influence on the caliper in correlation, regression, and principal component analysis (PCA). MD tensile and CD tear showed an impact in correlation and PCA. Neither correlation nor PCA had any variables that arose in common with the regression analysis. In the fir dataset, KSI, dry weight spread, and burst were variables recurring in all three analyses. The following variables arose in the fir dataset's correlation and principal component analysis: Sheet width, speed, main steam pressure, trim, porosity, MD tear, CD tear, MD elongation, PPS bottom, MD

tea, CD TSI, and TSI ratio. The only variable that arose in both PCA and regression for the fir dataset was the basis weight spread. In the spruce dataset, no variables appeared in the correlation analysis, and there were no variables that arose in common among any of the analyses. Throughout all three datasets, basis weight, CD tear, and KSI repeatedly arose as influential variables in the analysis.

This study also investigated how neural networks could supplement traditional statistical methods. The neural networks provided a more flexible way to explore patterns and relationships within the data. As seen in this thesis, traditional statistics often require predefined assumptions about the data and relationships, such as linearity or independence, compared to the neural networks section, which captures natural patterns in the data without the researcher's influence. Similar to correlation analysis's investigation of simple linear associations, neural networks identified more intricate patterns between the variables. Neural networks adaptively learned relationships amongst the variables with no prior assumptions, including no knowledge of the underlying relationships, compared to linear regression, which required meeting an array of assumptions. Lastly, PCA provided a useful linear dimension reduction, whereas neural networks found an alternative latent space. Researchers can enhance their analytical analysis by combining neural networks with traditional statistics, achieving both interpretability from traditional statistics and predictive power from neural networks.

Due to the conciseness of this thesis, much further and alternative research could be done. With any mechanical process, there are many influences on the outcome. As for the papermaking process, it has been seen that many factors, some human-controlled and some natural, influence the final caliper of the paper. Further research could investigate a tighter control on wood fibre by creating an experimental design for wood selection, including things other than species, such as age, geographic origin, wood density, part of a tree, etc. Other variables within the process that

the mill did not provide could also be investigated. Future investigations could also explore the relationship among variables, including factors like the time delay of some variables' influence on the caliper and the interactions of variables.

Overall, this study adds to the literature of influences on sheet caliper in the papermaking industry. There is a need to consider species in the process and the impact of machine variables on the desired outcome. The study also adds to the use of neural networks as an investigative tool during research.

References

Asuero, A. G., Sayago, A., & Gonzalez, A. G. (2006). The correlation coefficient: An overview. *Critical reviews in analytical chemistry*, *36*(1), 41-59.

Bajpai, P. (2018). Brief description of the pulp and papermaking process. In Biotechnology for pulp and paper processing (pp. 9–26).

Barnes, T. J. (1998). A history of regression: actors, networks, machines, and numbers. *Environment and Planning A*, *30*(2), 203-223.

Bebis, G., & Georgiopoulos, M. (1994). Feed-forward neural networks. *Ieee Potentials*, *13*(4), 27-31.

Bengio, Y. (2012, June). Deep learning of representations for unsupervised and transfer learning. In Proceedings of ICML workshop on unsupervised and transfer learning (pp. 17–36). JMLR Workshop and Conference Proceedings.

Bennett, D. A. (2001). How can I deal with missing data in my study? *Australian and New Zealand journal of public health*, *25*(5), 464-469.

Biermann, C. J. (1996). *Handbook of pulping and papermaking*. Elsevier.

Bro, R., & Smilde, A. K. (2014). Principal component analysis. Analytical methods, 6(9), 2812-2831.

Casson, R. J. & Farmer, L. D. (2014). Understanding and checking the assumptions of linear regression: a primer for medical researchers. Clinical & Experimental Ophthalmology, 42(6), 590–596.

Chu, D., Forbes, M., Backström, J., Gheorghe, C., & Chu, S. (2011). Model predictive control and optimization for papermaking processes. *Advanced model predictive control*, 309-341.

Cunningham, J. P., & Ghahramani, Z. (2015). Linear dimensionality reduction: Survey, insights, and generalizations. The Journal of Machine Learning Research, 16(1), 2859–2900.

Drost, C., Ni, Y., & Shewchuk, D. (2003). Effect of mature and juvenile wood from five wood species

on kraft pulp strength. *Pulp & paper Canada*, *104*(11), 33-36.

Fradkov, A. L. (2020). Early history of machine learning. *IFAC-PapersOnLine*, *53*(2), 1385-1390.

Gogtay, N. J., & Thatte, U. M. (2017). Principles of correlation analysis. *Journal of the Association of Physicians of India*, *65*(3), 78-81.

Håkansson, M. (2014). *Connecting Process Variables to Product Properties in Papermaking: A Multivariate Approach* (Doctoral dissertation, Luleå tekniska universitet).

Holmstad, R., Antoine, C., Silvy, J., Costa, A. P., & Antoine, J. (2001). Modelling the paper sheet structure according to the equivalent pore concept. *Proceedings from the Cost Action E11–Characterisation methods for fibres and paper*, *15*(2001), 25.

Jolliffe, I. T. (2002). Springer series in statistics. *Principal component analysis*, *29*, 912.

Joutsimo, O. (2004). *Effect of mechanical treatment on softwood kraft fiber properties*. Helsinki University of Technology.

Koivo, H. N. (2009). Automation And Control of Pulp and Paper Processes. *Control Systems, Robotics, and Automation.*

Kumari, K., & Yadav, S. (2018). Linear regression analysis study. *Journal of the practice of Cardiovascular Sciences*, *4*(1), 33-36.

Lee Rodgers, J., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. The American Statistician, 42(1), 59–66.

Li, J. (1994). *Adaptive control of sheet caliper on paper machines* (Doctoral dissertation, University of British Columbia).

Lu, X. (1999). *Print mottle of wood-containing paper, the effect of fines and formation* (Doctoral dissertation).

Marklund, A., Hauksson, J. B., Edlund, U., & Sjöström, M. (1998). Prediction of strength parameters for softwood kraft pulps. *Nordic Pulp & Paper Research Journal*, *13*(3), 211-219.

Mertens, O., Gurr, J., & Krause, A. (2017). The utilization of thermomechanical pulp fibers in WPC: A

review. *Journal of Applied Polymer Science*, *134*(31), 45161.

Mishra, S. P., Sarkar, U., Taraphder, S., Datta, S., Swain, D., Saikhom, R., ... & Laishram, M. (2017). Multivariate statistical data analysis-principal component analysis (PCA). International Journal of Livestock Research, 7(5), 60-78.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.

Norcliffe, G., & Bates, J. (1997). Implementing lean production in an old industrial space: Restructuring at Corner Brook, Newfoundland, 1984–1994. *Canadian Geographer/Le Géographe canadien*, *41*(1), 41-60.

Osborne, J. W., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical assessment, research, and evaluation*, *8*(1).

Pikulik, I. (2011). Papermaking. Pulp and Paper Technical Association, Canada.

Piovani, J. I. (2008). The historical construction of correlation as a conceptual and operative instrument for empirical research. *Quality & Quantity*, *42*(6), 757-777.

Poole, M. A., & O'Farrell, P. N. (1971). The assumptions of the linear regression model. *Transactions of the Institute of British Geographers*, 145-158.

Rättö, P., & Rigdahl, M. (1998). Platen press and calendering experiments with paper-estimating the final deformation in the thickness direction. *Nordic Pulp & Paper Research Journal*, *13*(3), 186-190.

Ringnér, M. (2008). What is principal component analysis?. *Nature biotechnology*, *26*(3), 303-304.

Rodionova, O., Kucheryavskiy, S., & Pomerantsev, A. (2021). Efficient tools for principal component analysis of complex data—A tutorial. Chemometrics and Intelligent Laboratory Systems, 213, 104304.

Sainani, K. L. (2015). Dealing with missing data. *PM&R*, *7*(9), 990-994.

Shiina, K. (2016). Origin of the correlation coefficient and its multifaceted interpretation. *Jpn Psychol*

Rev, *59*(4), 415-444.

Shlens, J. (2003). A tutorial on principal component analysis: derivation, discussion and singular value decomposition. Mar, 25(1), 16.

Stanton, J. M. (2001). Galton, Pearson, and the peas: A brief history of linear regression for statistics instructors. *Journal of Statistics Education*, *9*(3).

Samuel, M., & Okey, L. E. (2015). The relevance and significance of correlation in social science research. International Journal of Sociology and Anthropology Research, 1(3), 22–28.

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*, *126*(5), 1763-1768.

Seel, M. (2015). Statistics: A way to support or undermine one's arguments? NU ideas, 4(2), 49–56.

Smook, GA. (1992). Handbook for pulp and paper technologists. Joint textbook committee of the paper industry of the United States and Canada, 425p

Soley-Bori, M. (2013). Dealing with missing data: Key assumptions and methods for applied analysis. Boston University, 4(1), 1–19.

Statistics Canada. (2021). Census Profile, 2021 Census: Long Range Mountains Federal Electoral District. Retrieved March 2025 from: https://www12.statcan.gc.ca/census-recensement/2021/dppd/prof/details/page.cfm?Lang=E&SearchText=Corner%20Brook&DGUIDlist=2021S05100204&GENDERlist=1,2,3&STATISTIClist=1&HEADERlist=0

Tabachnick, B. G., & Fidell, L. S. (2001). Using Multivariate Statistics (4th ed.). Boston, MA: Allyn and Bacon

Tarca, A. L., Carey, V. J., Chen, X. W., Romero, R., & Drăghici, S. (2007). Machine learning and its applications to biology. *PLoS Computational Biology*, *3*(6), e116.

Tripepi, G., Jager, K. J., Stel, V. S., Dekker, F. W., & Zoccali, C. (2011). How to deal with continuous and dichotomic outcomes in epidemiological research: linear and logistic

regression analyses. *Nephron Clinical Practice*, *118*(4), c399-c406.

Ververis, C., Georghiou, K., Christodoulakis, N., Santas, P., & Santas, R. (2004). Fibre dimensions, lignin and cellulose content of various plant materials and their suitability for paper production. *Industrial crops and products*, *19*(3), 245-254.

Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, *2018*.

Warner, B., & Misra, M. (1996). Understanding neural networks as statistical tools. *The american statistician*, *50*(4), 284-293.

Appendix A: Chart of Variables

Number	Variable	Variable type in Baseline Data	Variable type in Fir Dataset	Variable type in the Spruce Dataset
1.	Stock flow (lpm)	Continuous	Continuous	Continuous
2.	Wet End Break	Appears Binary between zero and 1	Constant at zero or "No Good Data"	Constant at zero or "No Good Data"
3.	Dry End Break	Appears Binary between zero and 1	Constant at zero or "No Good Data"	Constant at zero or "No Good Data"
4.	Basis Weight	Continuous	Continuous	Continuous
5.	Spec. #	Discrete	Discrete	Discrete
6.	Sheet Width	Continuous	Continuous	Continuous
7.	Speed	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
8.	Headbox Temp	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
9.	Headbox pH	Continuous	Continuous	Appears to change Discretely
10.	Main Steam Pressure	Continuous	Continuous	Continuous
11.	Steam Box Temp	Continuous	Continuous	Constant at 115.7519531
12.	Steam Box Pressure (Kpa)	Continuous	Continuous	Continuous
13.	Steam Box Flow (Klb/hr)	Continuous	Continuous	Continuous
14.	Mx Moist	Continuous	Continuous	Continuous
15.	Caliper spread (CD)	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
16.	Moisture spread (CD)	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
17.	Dry Weight spread (CD)	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
18.	Basis Weight spread (CD)	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely

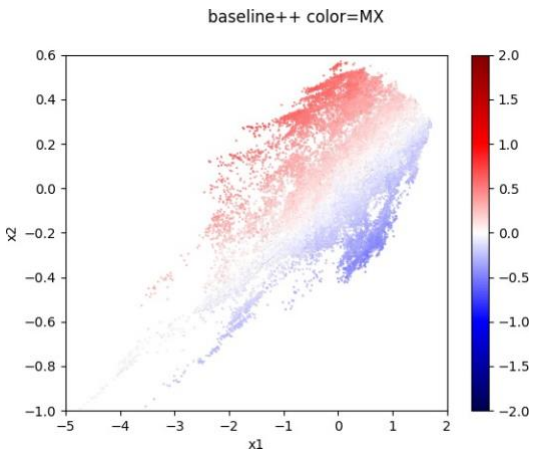
19.	Stuff Pump Consistency	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
20.	Thick Stock flow (BDMT per Day)	Continuous	Continuous	Continuous
21.	%Broke	Appears to change Discretely	Continuous	Appears to change Discretely
22.	Broke Flow(lpm)	Continuous	Continuous	Continuous
23.	TMP flow (lpm)	Continuous	Continuous	Continuous
24.	Trim	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
25.	Porosity	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
26.	MD Tensile	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
27.	CD Tensile	Stays Constant at 2.62	Stays Constant at 2.619999886	Stays constant at 2.619999886
28.	Tensile Ratio	Stays Constant at 0.88	Stays constant at 0.879999995	Stays constant at 0.879999995
29.	MD Tear	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
30.	CD Tear	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
31.	Tear Ratio	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
32.	MD Elongation	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
33.	CD Elongation	Stays constant at 0.21	Stays constant at 0.209999993	Stays constant at 0.209999993
34.	Burst	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
35.	KSI	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely

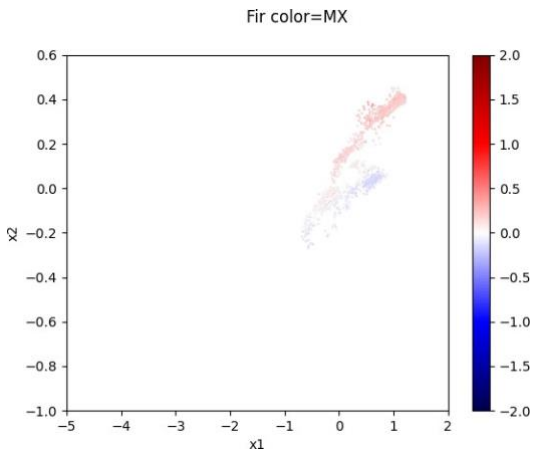
36.	Opacity	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
37.	Brightness	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
38.	a*	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
39.	b*	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
40.	PPS Top	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
41.	PPS Bottom	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
42.	MD TEA	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
43.	CD TEA	Stays constant at 10.01	Stays constant at 10.01000023	Stays constant at 10.01000023
44.	MD TSI	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
45.	CD TSI	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
46.	TSI Ratio	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
47.	TSO	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
48.	Jumbo	Discrete, reference number	Discrete, reference number	Discrete, reference number
49.	Jet to Wire	Continuous	Continuous	Continuous
50.	l*	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely
51.	Mx Caliper	Continuous	Continuous	Continuous
52.	Caliper	Appears to change Discretely	Appears to change Discretely	Appears to change Discretely

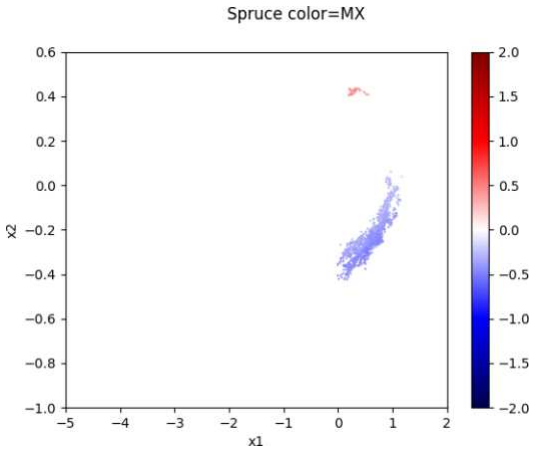
Appendix B: Images pertaining to N_8, N_4, and N_2 Neural Networks

N8

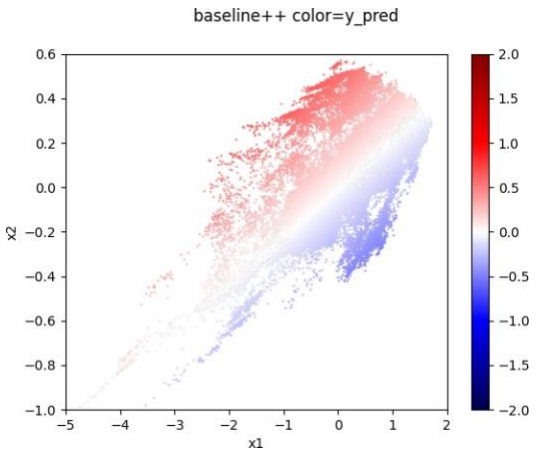
Transformed points of the original space for N8

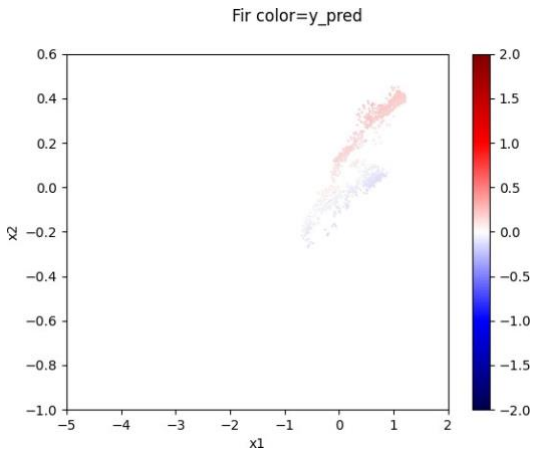


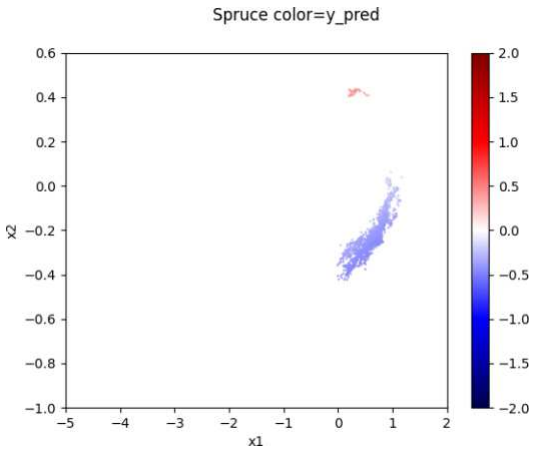




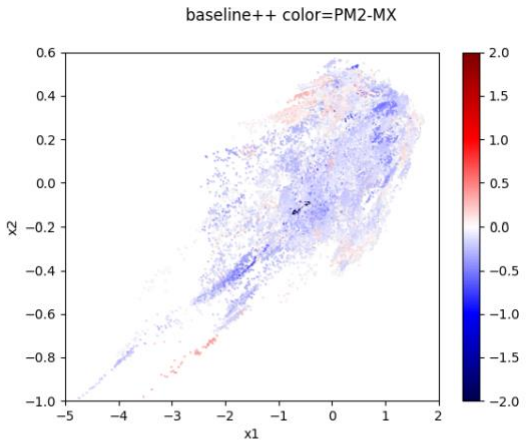
Caliper predicted by the neural network for N8

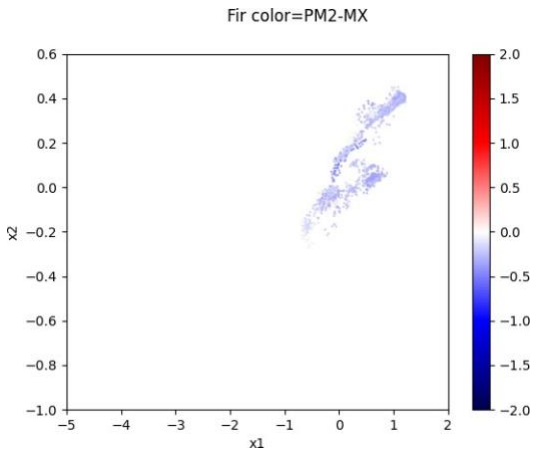


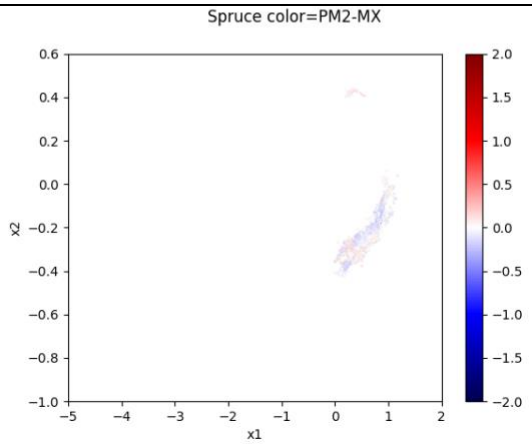




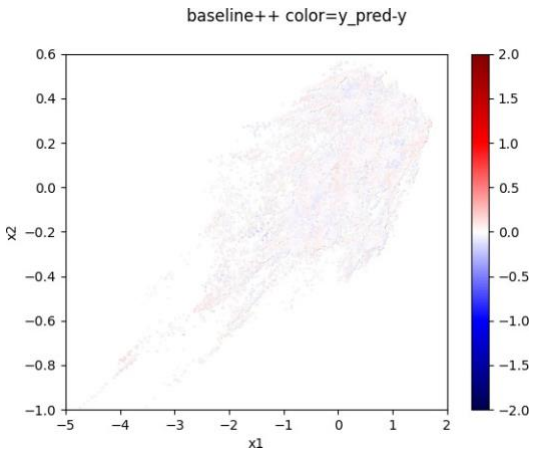
PM2_Caliper-MX_Caliper for N8

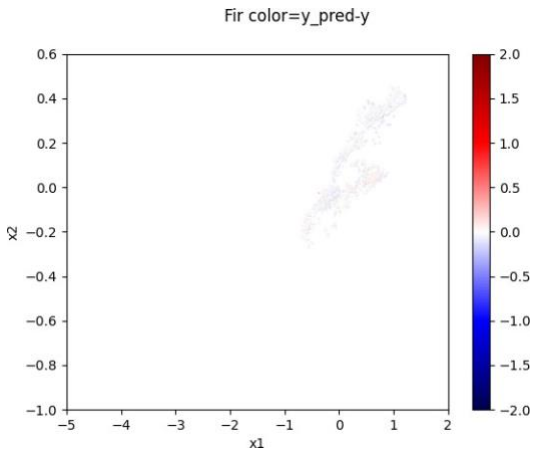


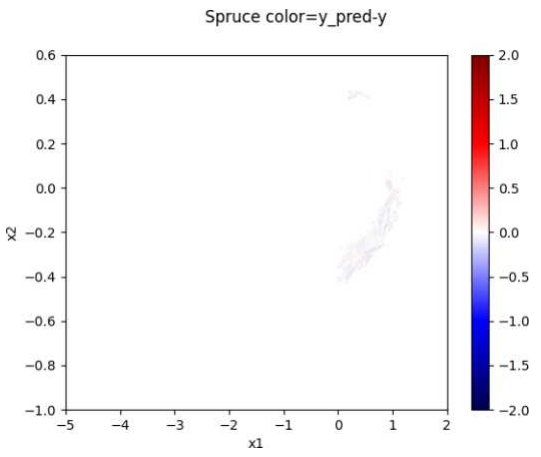




Difference in predicted value of y and y for N8



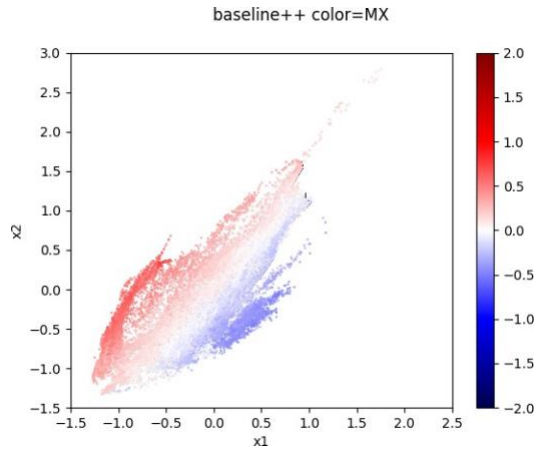
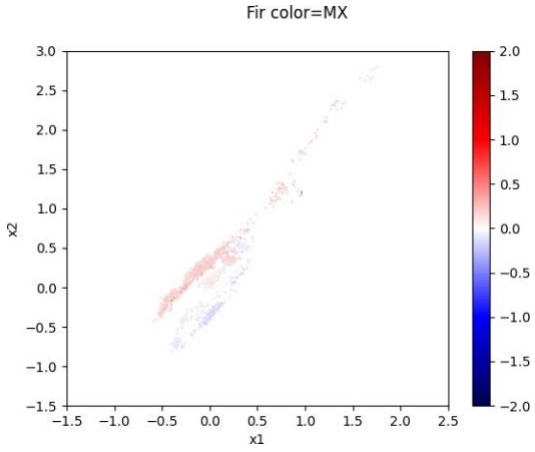
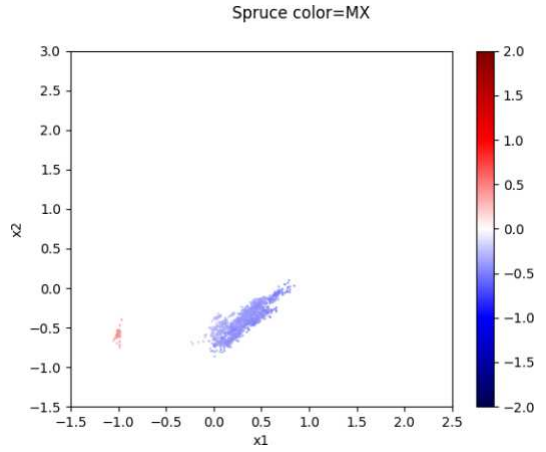




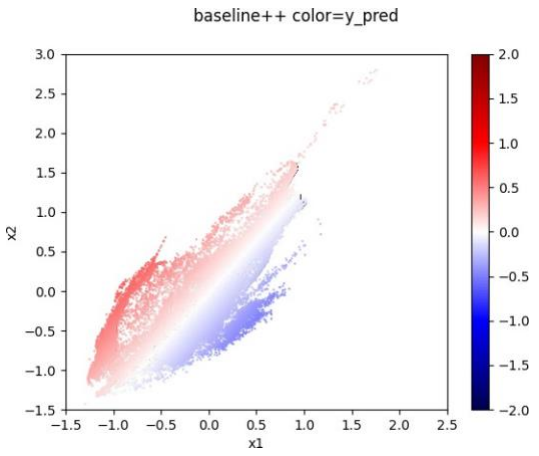
Regression summary statistics for N8

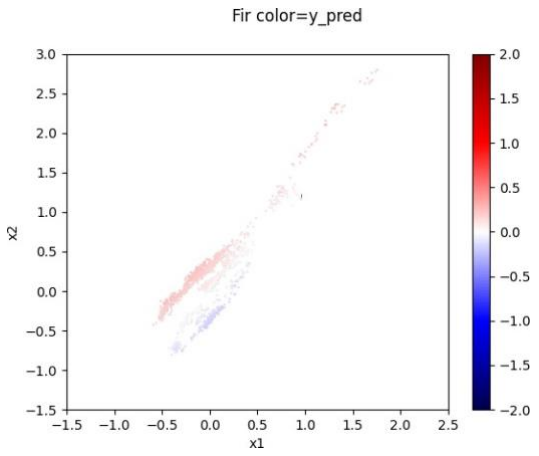
R^2	0.83
MSE of residuals	0.007

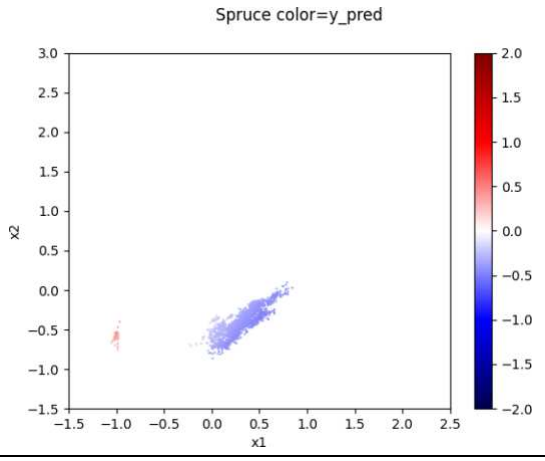
N4

Transformed points of the original space for N4	
	
	

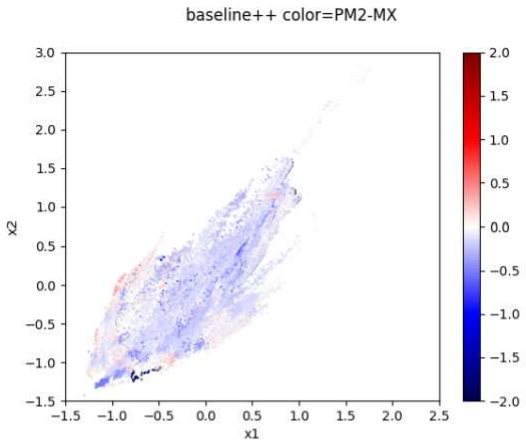
Caliper predicted by the neural network for N4

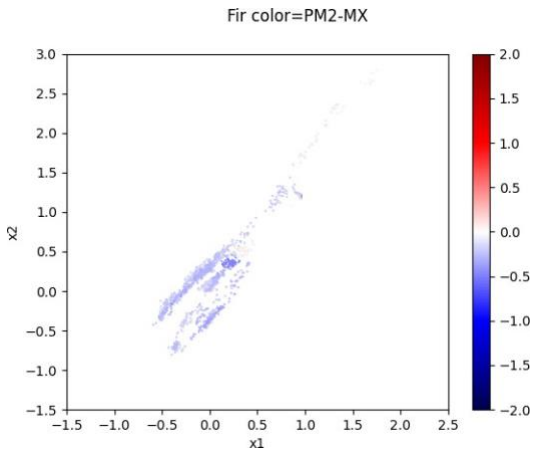


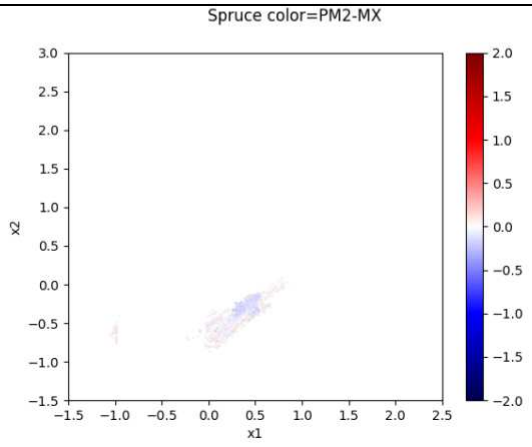




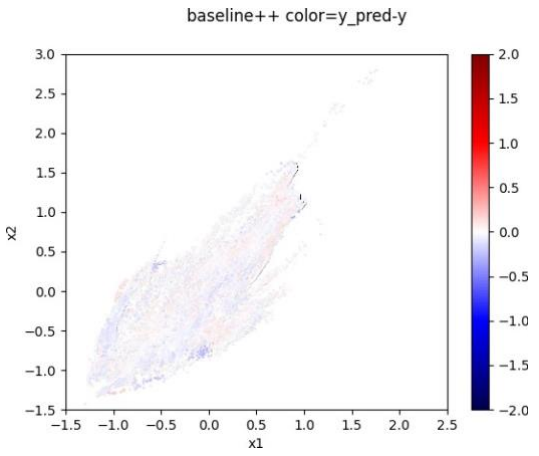
PM2_Caliper-MX_Caliper for N4

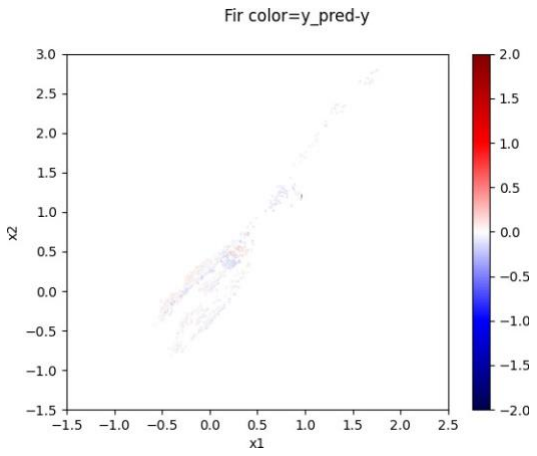


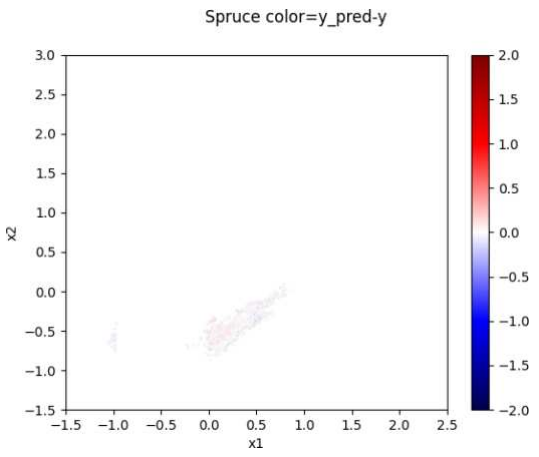




Difference in predicted value of y and y for N4



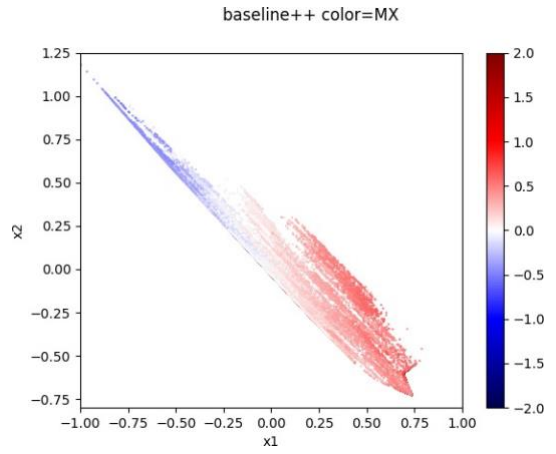
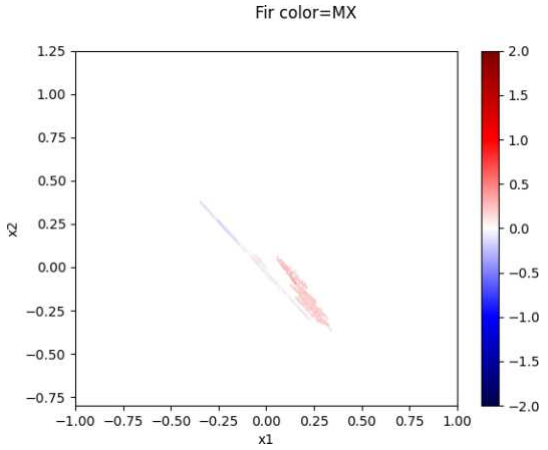
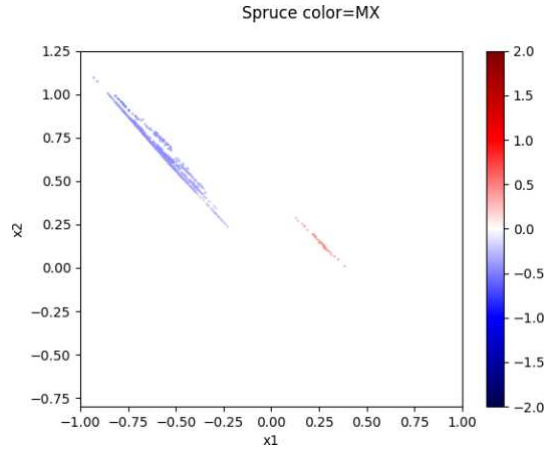




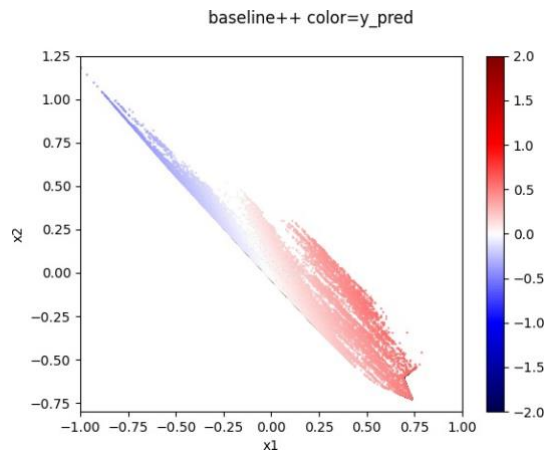
Regression summary statistics for N4

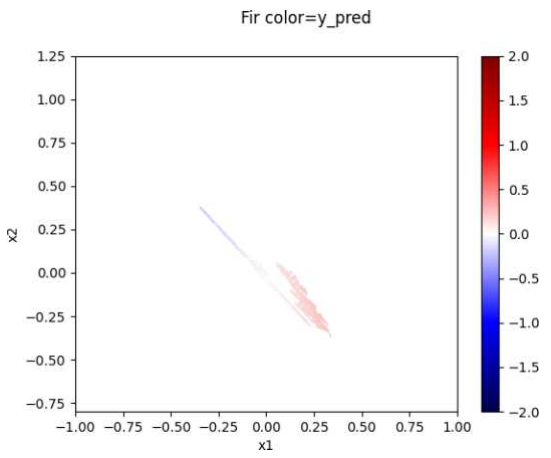
R^2	0.83
MSE of residuals	0.008

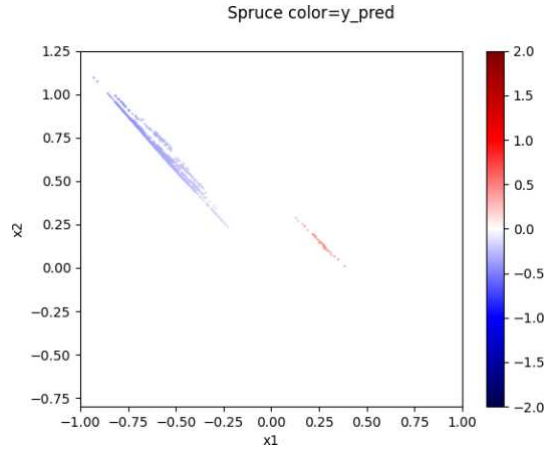
N2

Transformed points of the original space for N2	
	
	

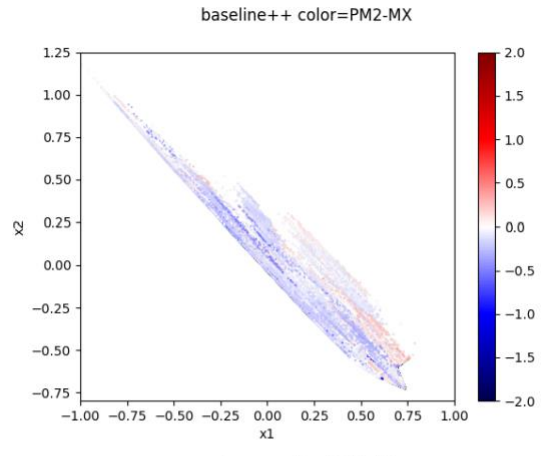
Caliper predicted by the neural network for N2

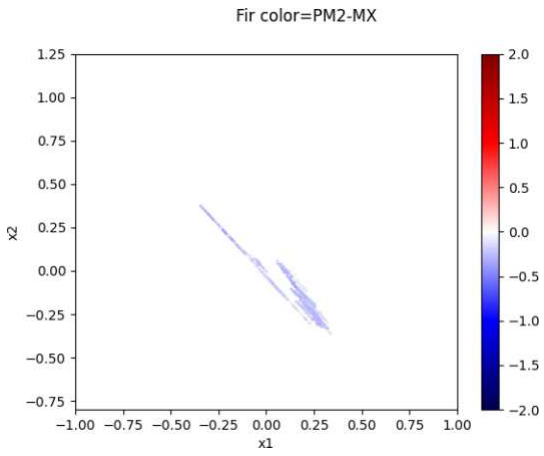


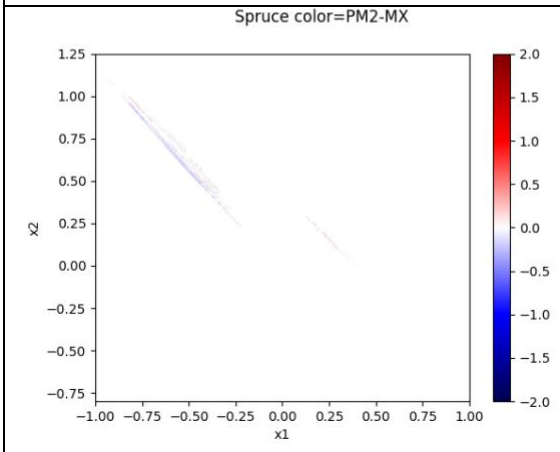


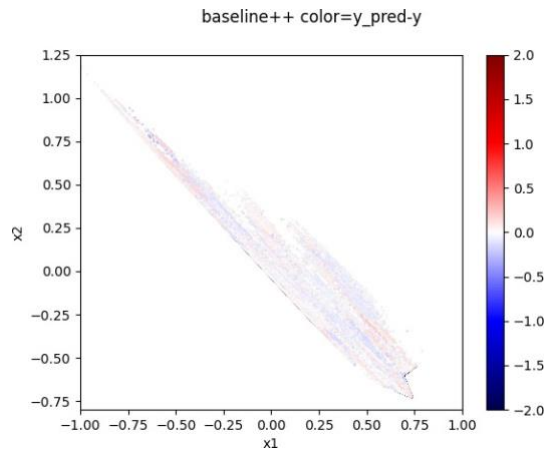
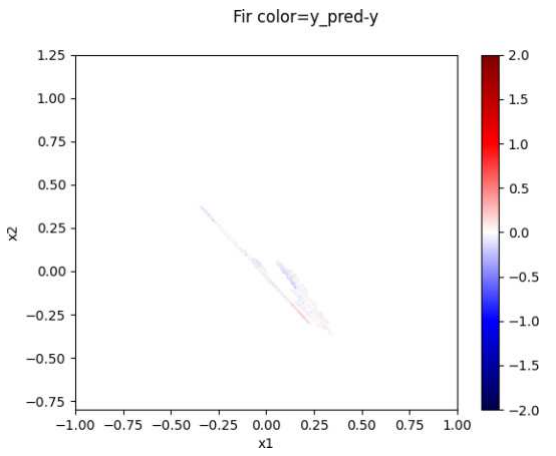
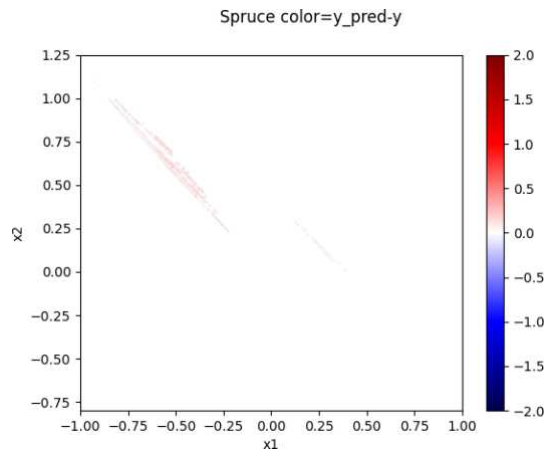


PM2_Caliper-MX_Caliper for N2







Difference in predicted value of y and y for N2	
	
	

Regression summary statistics for N2

R^2	0.83
MSE of residuals	0.008