Forintek Canada Corp.
Western Division
2665 East Mall
Vancouver, BC
V6T 1W5

A Statistical Method of Determining Check Depth in MPB Affected Logs

by

Laszlo Orbay
Research Scientist
Sawmilling Group

Prepared for

Forestry Innovation Investment

March 2006

Recipient Agreement No.: MBP – 06 – 01
FCC Project #: 5155

Laszlo Orbay
Project Leader

John Taylor
Reviewed

Peter Lister
Department Manager

Summary

Due to the current mountain pine beetle epidemic in BC, there now exists an extremely large amount of dead and dying lodgepole pine trees in the Interior of the Province. As sawmills would normally process the majority of these trees, the sawmilling industry is very interested in the affect these damaged trees will have on lumber recovery. The most significant problem is the checking that takes place as the trees dry out.

One method being used to investigate the lumber recovery from the damaged logs is that of simulation of the log breakdown process. To carry out any simulation studies, details of the checks are required. As well as the width and length, the depth of the check is also a very important characteristic. In this study, regression analysis was used to develop statistically significant equations that predict check depth using independent variables based on log and check dimensions.

Acknowledgements

Forintek Canada Corp. would like to thank its industry members, Natural Resources Canada (Canadian Forest Service), British Columbia, Alberta, Saskatchewan, Manitoba, Ontario, Quebec, Nova Scotia, New Brunswick, Newfoundland and Labrador, and the Yukon Territory, for their guidance and financial support for this research.

Forintek Canada Corp. acknowledge the assistance provided by the Province of British Columbia through the Forestry Innovation Investment Mountain Pine Beetle Program.

Forintek Canada Corp. would also like to acknowledge Tolko Industries Ltd., Quest Wood Division, Quesnel, B.C., for its support and cooperation throughout the project. In particular, the assistance of Ryan Lake, Troy Conolly and Derek Brears with selecting sample logs and collecting data.

The suggestions of Dr. Les Safranyik, research scientist in pathology, regarding check and log characteristics that might affect check-depth, were also helpful.

Table of Contents

List of Tables

List of Figures

1 Background

The Mountain Pine Beetle (MPB) epidemic in central BC is affecting vast amounts of Lodgepole pine timber. Once attacked, the affected trees quickly develop blue stain and eventually are killed. As the dead trees dry out, radial cracks develop that significantly reduce the volume and value of the lumber that can be recovered from the resource.

For those involved in the simulation and optimization of lumber breakdown processes, having information about how checks run along the length of MPB affected logs could go a long way towards lessening the lumber value recovery loss. By having detailed knowledge of both crack location and depth, log breakdown patterns could be adjusted to minimize losses in lumber volume and grade recovery. Today, several system integrators and machine vision developers are working on equipment to automatically detect the checks. Systems are emerging that will automatically inspect log ends and log sides to see if radial checks are present. However these systems can only measure the depth of the checks at the log ends.

It is known that the depth of the checks varies significantly along the length of the log due to variations in log moisture content. These changes in check depth can have an appreciable affect on the actual volume and value of the lumber produced, so are important for log breakdown optimization purposes. Consequently, this project was developed to provide a statistically valid regression equation estimating the check depth in MPB attacked logs based on features than can be readily measured on the log surface using emerging scanning and machine vision systems.

2 Objectives

The project objectives were to determine a statistically reliable regression equation for estimating the depth of radial surface checks along the length of logs and to determine the accuracy of estimation.

3 Staff

Laszlo Orbay,	Research Scientist, Sawmilling Group, Manufacturing Department
Vit Mlcoch,	Technologist, Lumber Drying Group, Manufacturing Department
Dr. Tony Kozak,	Professor Emeritus, Faculty of Forestry, UBC.

4 Materials and Study Method

In order to develop regression equations, data from MPB attacked logs was required. The project was broken into two phases. The first phase consisted of a small pilot study to obtain data for preliminary regression analysis purposes. A second phase was then completed to gather additional data that was used to refine the regression models.

For the pilot study phase, data on check width and log diameter was collected from sample logs taken from a BC interior mill that processes MPB attacked logs. As a first estimate, a linear regression analysis was carried out to predict check depth based on check-width (x_1) and log diameter (x_2). The diameter used was the average diameter of the log cross-section at the location where the check measurements were collected. The average diameter, x_2, was calculated from two diameter measurements ($x_2 = (D_1 + D_2) / 2$), as seen in Figure 1.

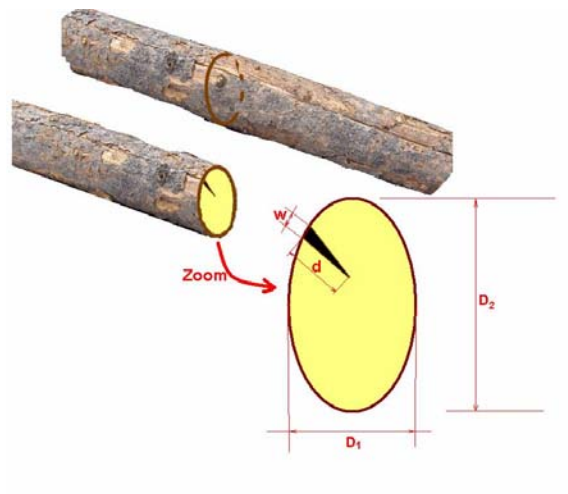


Figure 1: Log diameter, check-width and check–depth measurements.

For data collection, heavily damaged sample logs of beetle-killed wood were laid out at the log yard at the cooperating sawmill (Figure 2). The sample logs were crosscut at selected check locations and 77 sets of check-depth, check-width and log diameter measurements were taken. Figure 3 shows how check dimensions were characterized. The results of these measurements are given in Appendix 1.

Figure 2: ***Mountain pine beetle-killed sample logs laid out for measurements.***

For the pilot project, check depth was initially estimated using a simple linear regression equation developed from the log measurement data. For mathematic modeling purposes, the check depth, d, was considered the dependent variable that was to be estimated based on the independent variables check width, x_1 and average log diameter, x_2 (see Table 1). These variables were initially chosen because it was felt that they could be readily measured using currently available scanners and machine vision systems.

The resulting regression equation, Equation 1 in Table 3, provided a *statistically significant* estimate of the check depth. However, further analysis suggested that a non-linear relationship might further enhance the accuracy of the model. Based on the same observations as that of the pilot study, a two variable curvilinear equation was also developed (Equation 2, Table 3). This equation gave slightly better R^2, standard error and relative error results but also had a larger confidence interval (Table 4).

Following the pilot study, a second phase was carried out to try and further refine the regression equations. This work focused on finding additional independent variables that might improve check-depth predictions. Based on discussions with experts in wood anatomy, wood pathology and sawmilling, the following seven additional independent variables were evaluated. The rationale is shown in *italics*.

1. Moisture content at the tip of the check (close to the pith) - *At a given site and tree size, lower moisture content indicates longer time since death, probably resulting in a deeper check.*
2. Sapwood thickness at the cross-section of measurement – *Higher sapwood volume might indicate higher moisture content in the tree prior to successful attack; therefore, the rate of evaporation/drying might cause more severe checking.*
3. Check-spirality - *Indicates spiral grain that might be a result of poor growing conditions and consequently deeper and/or longer checks.*

4. Check-length along the log - *Longer check may indicate a longer time since death, probably deeper check.*
5. The number of checks at the cross-section of measurement - *Higher number of checks per cross-section presumably provides more ways to release the stresses and result in shallower checks.*
6. Log taper – *High taper might indicate a poor growing environment and consequently deeper checks.*
7. Bark thickness – *Thicker bark might indicate a higher growth rate and consequently deeper checks.*

A second group of sample logs with heavy checks were selected as good candidates for data collection and from these, 96 additional sets of measurements were taken with each set consisting of the 10 separate measurements as described in Table 1. This table also lists the nomenclature used in the regression analysis, as well as the units of measurement and tools used to collect the measurements. The actual measurement data, together with descriptive statistics are listed in Appendix 1. Certain data sets were not complete (see Appendix 1) so the various regression analyses used different independent variables and numbers of measurement groups for regression equation development, as shown in Table 2.

Table 1: **Variables used in the regression analysis.**

Dependent and independent variables	Log/Check characteristics	Tool used for measurement
d	Check-depth, mm	Digital caliper
x_1	Check-width, mm	Digital caliper
x_2	Log diameter, cm	Log caliper
x_3	Log taper, cm/m	Tape measure, log caliper
x_4	Check-spirality, degree/m	Tape measure, estimated angle
x_5	Sapwood depth, mm	Digital caliper
x_6	Bark thickness, mm	Digital caliper
x_7	Check-length, cm	Tape measure
x_8	Moisture content, %	Delmhorst moisture meter
x_9	Number of checks/cross section	Count

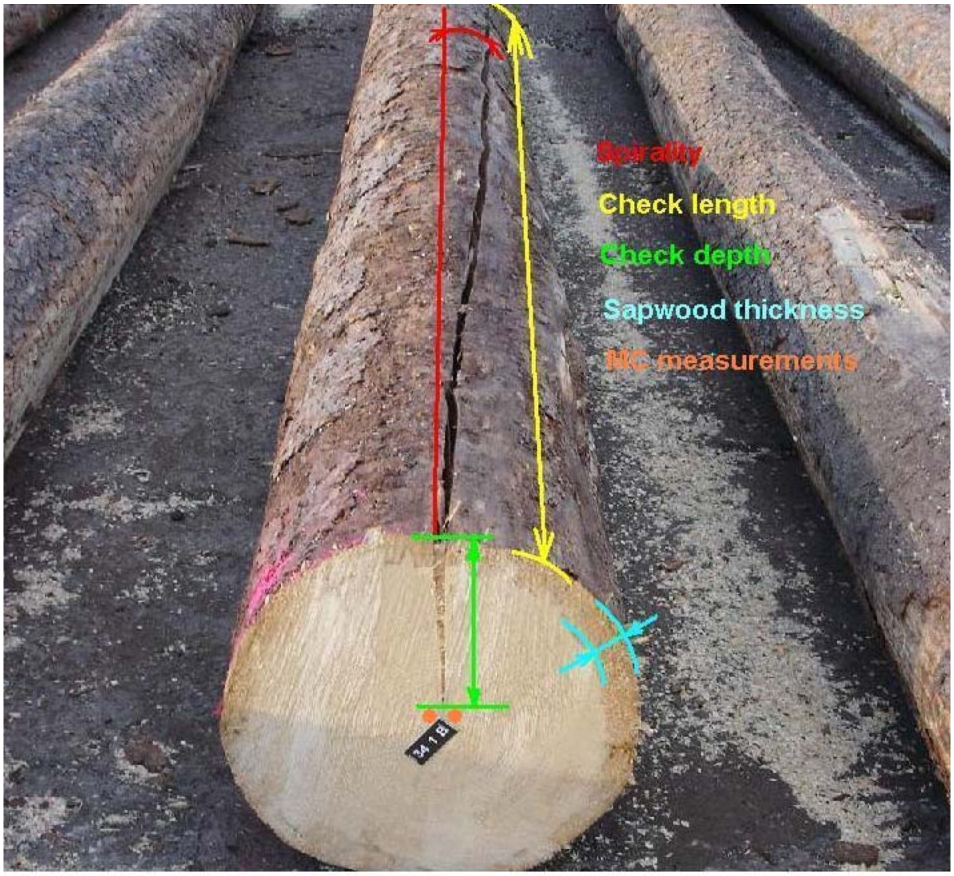


Figure 3: ***Check-spirality, check-length, sapwood-thickness and points of MC measurements.***

Table 2: ***Regression equation identification codes, the number of observations and types of independent variables.***

Regression equation (Table 3)	Number of observations	Variables used in the equation (See Table 1 for descriptions)
1	77	x_1, x_2
2	77	x_1, x_2, x_1^2, x_2^2, x_1x_2
3	38	x_1, x_2 ,…, x_9, x_1^2, x_2^2, x_3^2, x_5^2, x_7^2
4	88	Same as equation 3 but no x_6
5	88	Same as equation 3 but no x_6
6	173 (77+96)	x_1, x_2, x_1^2, x_2^2, x_1x_2
7	173	x_1, x_2, x_2^2
8	173	x_1, x_2, x_1^2
9	88	x_1, x_2 , x_3, x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3
10	88	x_2, x_3 , x_5, x_2^2, x_3^2, x_5^2, x_2x_3, x_2x_5, x_3x_5

The regression equations based on the expanded data sets were developed following standard statistical techniques:

- Creating scattergrams of individual independent variables versus a dependent variable to select sensible candidates for independent variables that significantly contribute to the variability of check-depth (dependent variable), as shown in Appendix 3.
- Selection of independent variables and the degree of the polynomial (curviness) by judging the scattergram.
- Calculation and testing for significance of regression equation coefficients.
- Calculating regression equation efficiency indicators (R^2, *Standard Error of Estimate*, *Relative Error, Bias, Trend, SEE Total, Confidence Interval* at the averages of independent variables – (see statistical background in Appendix 2).
- Final regression equation selection.

A glossary of statistical terms along with equations used in the analysis is given in Appendix 2 and the final regression equations that were developed are given in Table 3.

5 Results and Discussion

Table 3 shows all the regression equations that were found to be statistically significant at the level of α=0.05. The corresponding statistics that indicate the effectiveness of the equations are given in Table 4.

Table 3: The regression equations of the analysis.

Regression equation number	Regression equation
1	$\hat{d} = -2.08078 + 3.58633\ x_1 + 0.34661\ x_2$
2	$\hat{d} = -74.3682 + 3.61031\ x_1 + 1.04516\ x_2 - 0.00158\ x_2^2$
3	$\hat{d} = 27.54043 + 12.08590\ x_1 + 0.090008\ x_2 - 0.70185\ x_1^2 - 11.77139\ x_3^2$
4	$\hat{d} = 30.21910 + 7.93710\ x_1 - 0.34552\ x_1^2 + 0.004103\ x_2^2$
5	$\hat{d} = 3.02272 + 7.80801\ x_1 + 0.230782\ x_2 - 0.33893\ x_1^2 - 0.00242\ x_5^2$
6	$\hat{d} = -1.62119 + 0.47473\ x_2 - 0.35485\ x_1^2 - 0.00080626\ x_2^2 + 0.02908\ x_1\ x_2$
7	$\hat{d} = -18.02760 + 3.61835 x_1 + 0.53801 x_2 - 0.00067278\ x_2^2$
8	$\hat{d} = 8.39261 + 7.70394 x_1 + 0.21416\ x_2 - 0.34532\ x_1^2$
9	$\hat{d} = 37.42037 + 29.19641\ x_3 - 0.44554\ x_1^2 + 0.048915\ x_1\ x_2 - 6.01423\ x_1\ x_3$
10	$\hat{d} = 40.63523 + 0.9605 x_5 + 0.0048398 x_2^2 + 25.03962 x_3^2 - 1.35052 x_3 x_5$

See Table 1 for nomenclature

Table 4: Regression equation efficiency indicators.

Regr. equation number	Confidence interval at the average of independent variables, mm	Standard error of estimate, mm	Relative error, %	R^2	Number of observations
1	7.5469	16.62	19.81	0.58549	77
2	8.6029	15.50	18.48	0.6443	77
3	7.9671	8.47	10.66	0.6534	38
4	6.3252	12.57	14.22	0.5941	88
5	6.3949	12.54	14.19	0.6009	88
6	6.5726	15.61	18.05	0.5302	173
7	6.1574	15.84	18.33	0.5130	173
8	5.5077	15.92	18.41	0.5086	173
9	5.6700	11.86	13.42	0.6429	88
10	6.8172	15.42	17.45	0.3963	88

Note: *Computational formulas of Confidence interval, Standard error of estimate, Relative error and R^2 are in Appendix 2.*

Regression equations 1 and 2 (Table 3) were developed from the data collected in the initial pilot study. For Equation 1 all elements of the equation (intercept, x_1, x_2), were significant at the level of $\alpha=0.05$ (Table 3). Fifty-nine percent ($R^2 = 0.59$) of the change of the predicted variable, the check-depth, can be explained by the two independent variables. The *standard error of estimate* shows that 68% of check-depth measurements are within ± 16.62 mm of the surface of the regression equation.

Additional analysis of the pilot study data suggested that a non-linear relationship between check-depth, check-width and log diameter might further improve the accuracy of check-depth estimates. The best-fit curves in Figure 4 and 5 suggest non-linear trends. Consequently, the product of check width and diameter (x_1*x_2) and the second power of both (x_1^2 and x_2^2) were included in the analysis. Thus, the following two-variable non-linear equation was used as a starting point for further analysis:

$$\hat{d} = b_0 + b_1x_1 + b_2x_2 + b_3x_1x_2 + b_4x_1^2 + b_5x_2^2$$

where $\hat{d}$ is the estimated check-depth and b_i's are the independent variable coefficients to be determined.

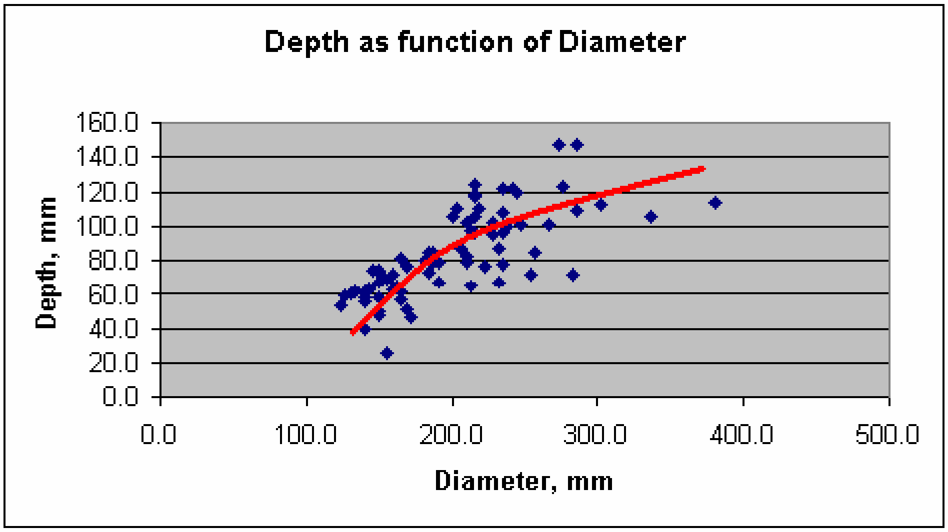


Figure 4: Change of check-depth as a function of diameter.

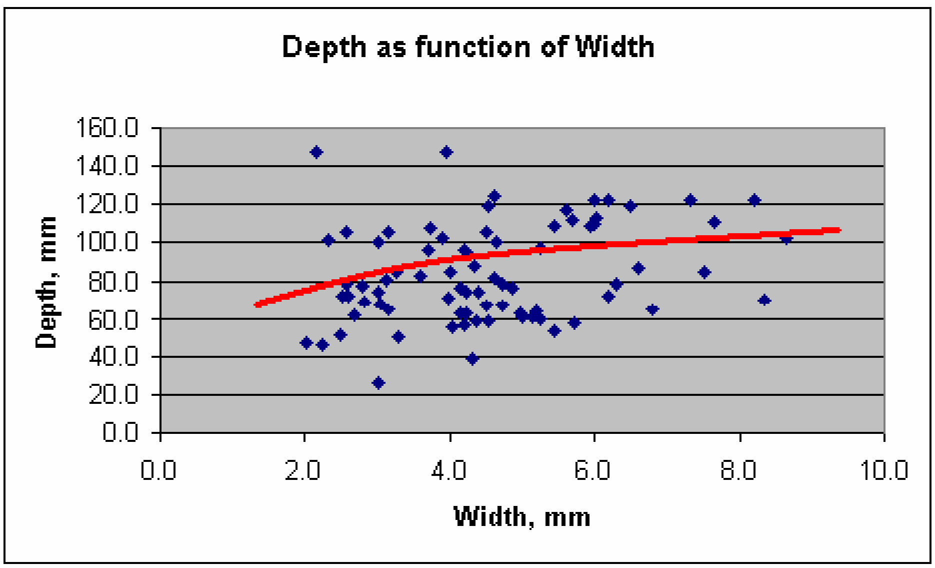


Figure 5: Change of check-depth as a function of check-width.

This analysis suggested that only four elements of the non-linear equation were significant (intercept, x_1, x_2 and x_2^2). The other two elements (x_1x_2 and x_1^2) were not significant and were therefore eliminated. Examination of the resulting regression equation, Equation 2 in Table 3 shows that log diameter affects the check depth more than check-width. As well, the corresponding R^2 value in Table 4 ($R^2 = 0.64$) shows that 64 percent of the variation of the dependent variable (check-depth) can be explained by the three significant elements (x_1, x_2 and x_2^2) . Though the *confidence interval* of Equation 2 is slightly larger than Equation 1, the *standard error of estimate* and the *relative error* (SE=15.50 mm and RE=18.48%) show less spread around the regression surface; hence Equation 2 provides a better estimate of check depth than Equation 1.

Using the additional data collected after the pilot study, more complex regression models were developed that included seven additional check and log characteristics (x_3, x_4, …, x_9) thought to potentially affect check-depth. Repeating the process described above the following equation was selected as an appropriate starting point for the elimination procedure and to find the significant equation elements:

$$\hat{d} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_6 + b_7x_7 + b_8x_8 + b_9x_9 + b_{10}x_1^2 + b_{11}x_2^2 + b_{12}x_3^2 + b_{13}x_5^2 + b_{14}x_7^2$$

From the seven newly introduced independent variables, the only statistically significant, additional independent variable was the square of log taper (x_3^2), this led to Equation 3 in Table 4. In this last step, the number of measurement groups was limited by the availability of independent variable x_6 (bark thickness). Consequently, Equation 3 used only 38 of the additional 96 measurement groups (Appendix 1).

Since independent variable x_6 in Equation 3 was not a significant variable it was eliminated from the next phase of the analysis. This led to eighty-eight groups of measurements (from the 96 groups listed) that were used in the next phase of the analysis, with the following starting equation:

$$\hat{d} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_7 + b_7x_8 + b_8x_9 + b_9x_1^2 + b_{10}x_2^2 + b_{11}x_3^2 + b_{12}x_5^2 + b_{13}x_7^2$$

Regression equations 4, 5 and 9 are the resulting equations. Among the added independent variables there were only two that were found to be slightly significant: taper and sapwood depth (x_3 in Equation 9 and x_5 in Equation 5). It is interesting to note that the moisture content of the logs (x_8, measured at the crack tip) and the number of radial checks (x_4) were not found to be a significant indicators of check depth for the sample of logs tested.

To make use of the full number of measurement groups, the next phase of the analysis concentrated on the two original variables, check-width and log diameter (x_1, x_2), this gave 173 groups; 77 preliminary and 96 additional. Three combinations of polynomial elements were considered:

x_1, x_2, x_1x_2, x_1^2 and x_2^2,
x_1, x_2 and x_2^2,
x_1, x_2 and x_1^2,

leading to equations 6, 7 and 8 respectively.

The regression equations developed so far were based on data collected in the Interior of B.C. under dry conditions. Descriptive statistics of the moisture content measurements show the dry state of the sample logs (Table 5). This may explain the lack of significance of moisture content in the analysis.

Table 5: Descriptive statistics of moisture content.

Upper boundary of MC (%)	Frequencies	Cumulative %
MC ≤ 16	7	7.37
MC ≤ 18	23	31.58
MC ≤ 20	26	58.95
MC ≤ 22	25	85.26
MC ≤ 24	4	89.47
MC > 24	10	100.00
Average MC:	19.6%	
Range of MC:	[14.5%, 28.3%]	
Standard deviation of MC's:	2.9	

Discussions with representatives from industry and preliminary experimentation with sample logs having severe checking suggested that check-width (openings) is sensitive to changes in atmospheric conditions. Even sample logs that have been below the moisture content limit of 19% (the moisture content limit of dry lumber) might expand and cause check openings to decrease when exposed to an increase in moisture. This might limit the utility of the recommended regression equations (with variable $x_{1)}$ to certain geographical regions and seasons of the year. Consequently, additional regression analysis, that excluded check-width, was carried out to see if other variables could still predict check-depth reasonably well.
The last segment of the analysis focused on x_2, x_3 and x_5 with the following starting equation:

$$\hat{d} = b_0 + b_1x_2 + b_2x_3 + b_3x_5 + b_4x_2^2 + b_5x_3^2 + b_6x_5^2 + b_7x_2x_3 + b_8x_2x_5 + b_9x_3x_5$$

and resulted in equation 10. Among its efficiency indicators, the R^2 fell to 0.40. However, the other indicators showed that, where the moisture content of the checked logs had subsequently increased and the check widths were difficult to measure with the scanning equipment, Equation 10 could be used to predict check-depth.

Having derived the regression equations in Table 3, the arising question is which regression equation should be used? Unfortunately, choosing the "best" regression equation is not an exact science. Both statistical and practical application aspects should be taken into account in the final decision. From the viewpoint of statistics one should choose the most effective equation. When choosing the "best" regression equation, provisions were made to exclude subjectivity as much as possible. Ranking rules were created based on simultaneous considerations of prediction efficiency indicators of *confidence interval* width (estimation accuracy), *standard error of estimate*, relative error, R^2 and Bias (Appendix 5). Based on a comparative evaluation of the equations of Table 3, Equation 5 was found to be the "best" equation and is recommended for industrial use under "dry" conditions:

$$\hat{d} = 3.02272 + 7.80801x_1 + 0.237082x_2 - 0.33893x_1^2 - 0.00242x_5^2$$

There are regression equations with efficiencies close to each other. Therefore, practical considerations, such as the cost and complexity of a machine vision system required to detect the required check-width (x_1) and sapwood depth (x_5) must be included in final decision.
Under moist conditions (MC > 22%), when the check can still be located by automated vision systems, the recommended equation is:

$$\hat{d} = 40.63523 + 0.9605x_5 + 0.0048398{x_2}^2 + 25.03962{x_3}^2 - 1.35052x_3x_5$$

Where x_2 is the average log diameter, x_3 is the log taper and x_5 is the sapwood depth.

6 Conclusions

Based on the limited data collected from the dry logs examined in this study, two regression equations are recommended. For dry logs where check widths can be readily measured, check depth should be estimated from the equation

$$\hat{d} = 3.02272 + 7.80801x_1 + 0.237082x_2 - 0.33893{x_1}^2 - 0.00242{x_5}^2 \qquad \text{(Equation 5)}$$

where x_1 is the check width, x_2 is the average log diameter at the check location and x_5 is the sapwood thickness.

However, for wet logs, or in cases where check width cannot be readily measured, check depth can be estimated from

$$\hat{d} = 40.63523 + 0.9605x_5 + 0.0048398{x_2}^2 + 25.03962{x_3}^2 - 1.35052x_3x_5 \qquad \text{(Equation 10)}$$

where x_2 is the average log diameter at the check location, x_3 is the log taper and x_5 is the sapwood thickness.

These equations are both statistically significant and can provide check-depth estimates for log breakdown optimization programs reasonably well. Their *confidence intervals*, indicating check-depth prediction accuracies, are 6 and 7 mm for Equations 5 and 10 respectively. However, it is not known how their prediction accuracy affects the efficiency of log breakdown optimization. Hence investigation on how log breakdown optimization programs respond to check measurement and hence to check modeling accuracy is recommended.

It is important to note that the regression equations developed in this study are based on measurements from a limited number of logs and may not be representative of the wider population of MPB affected logs. The logs considered in this study also had relatively narrow ranges of measured characteristics. For example, log diameters were between 12 and 38 cm, check-width was between 1 and 14mm and taper and sapwood were between 0 to 14cm/m and 13 to 66mm respectively. The regression equations developed from data obtained from these logs may not be suitable for predicting check depth for all MPB affected logs, nor for logs having diameter, moisture content taper and sapwood dimensions outside of the ranges represented by the logs used in the study. If extended use of check-depth prediction is needed, i.e., outside of the range of measurements this study was based on, then collecting additional data and conducting additional regression analysis is necessary. Additional regression analysis might show that developing separate regression equations for different segments of independent variables could improve prediction efficiency.

In addition, the range of moisture content measurements based on which the regression analysis was carried out is fairly narrow: 85% of moisture content measurements were below 22%. For these conditions, the recommended equation works well. However, under what moisture conditions the equation starts making unacceptable errors is not known. Therefore, further investigation into this matter should be considered.

7 Recommendations

The recommendations below are related to two topics. A brief discussion follows.

1. The interaction of log breakdown optimization system components: accuracy of check measurements that machine vision systems can provide, check modeling and sawing optimization software.
2. Atmospheric conditions and their impact on check-width.

7.1 Discussion

The interaction of log breakdown optimization system components: accuracy of check measurements that machine vision systems can provide, check modeling and sawing optimization software.

The recommended regression equations require machine vision systems that can provide check-width and sapwood thickness measurements. (This study assumed the availability of machine vision systems that can provide these measurements accurately). How accurate must the measurements be to be acceptable? Studies about the importance of accurate measurements focusing on their ultimate use in log breakdown optimization should be carried out for the following reason:

Check measurement accuracy affects check model accuracy and log breakdown optimization. These studies should investigate how log breakdown optimization programs respond to check measurement and hence to check modeling accuracy. Thus, sawing simulation studies comparing optimized sawing decisions based on true check shape and location versus optimized sawing decisions based on modeled check shape and location are recommended.

The results of these studies could set accuracy requirements for machine vision systems. For example, consider a machine vision system that operates with 2 mm error in check-width measurement and therefore causes 4 -12 mm error in check-depth prediction. The question is, if the log breakdown decision is significantly different with this error from the decision based on a check model without error, then how much is the loss in value recovery due to incorrect sawing decision?

Atmospheric conditions and their impact on check-width

Observations have suggested that check width can change with changing log moisture content. When the wood is in the drying stage and MC gets below the fibre saturation point, it starts shrinking and consequently checking. However, if exposed to moisture for any length of time, the MC of the wood can increase and the checks can start closing. Unfortunately, checks that have closed are still there and will cause losses in lumber recovery. As a result, narrow check openings might mask deep checks and in extreme cases, checks might not be detectable by machine vision systems. Therefore, under certain conditions, check-width may not be a good indicator of check-depth.

Moisture content measurements from this study were in a narrow range. Eighty-five percent of moisture content measurements were below 22%. So Equation 5 is recommended for use under these “dry” conditions. However, we do not know the exact boundaries of MC conditions above which Equation 5 is not applicable. So searching for answers to the following questions is recommended. How does check-width respond to changing atmospheric conditions (changing humidity, rain) and how they can be quantified?

Appendix 1: Measurements and their descriptive statistics.

Table App 1-1: Two groups of measurements: original (pilot study) and additional measurements.

Sequence	d	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
	Check-depth, mm	Check-width, mm	Diameter, mm	Taper, cm/m	Spirality, dgr/m	Sap-wood thickness, mm	Bark thickness, mm	Check-length, cm	Moisture Content, %	Number of checks per cross-section
Original 77 groups of measurements										
1	72.7	6.2	184.2							
2	110.0	6.0	203.2							
3	102.2	8.7	228.6							
4	122.3	8.2	235.0							
5	122.0	7.3	241.3							
6	112.4	5.7	301.6							
7	67.3	4.5	190.5							
8	119.2	4.5	244.5							
9	46.5	2.3	171.5							
10	122.2	6.0	235.0							
11	82.7	3.6	209.6							
12	71.7	2.5	282.6							
13	107.3	3.7	235.0							
14	106.0	4.5	336.6							
15	87.5	4.3	206.4							
16	118.8	6.5	215.9							
17	78.8	6.3	187.3							
18	108.7	5.5	285.8							
19	96.1	4.2	235.0							
20	117.5	5.6	215.9							
21	110.7	7.6	219.1							
22	84.8	4.0	257.2							
23	61.0	5.0	165.1							
24	38.9	4.3	139.7							
25	72.2	2.6	158.8							
26	55.5	4.0	139.7							
27	25.7	3.0	155.6							
28	62.8	4.2	158.8							
29	100.0	4.7	238.1							
30	70.4	8.3	152.4							
31	74.6	3.0	149.2							
32	102.6	3.9	209.6							
33	65.0	6.8	161.9							
34	147.4	2.2	285.8							
35	124.0	4.6	215.9							
36	122.6	6.2	276.2							
37	108.3	5.9	235.0							
38	100.8	3.0	247.7							
39	77.7	2.8	235.0							
40	96.5	3.7	215.9							
41	113.2	6.0	381.0							
42	95.4	4.2	228.6							
43	67.6	3.1	231.8							
44	71.6	4.0	254.0							
45	65.0	3.2	212.7							
46	86.8	6.6	231.8							

Forintek Canada Corp.

	d	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
Sequence	**Check-depth, mm**	**Check-width, mm**	**Diameter, mm**	**Taper, cm/m**	**Spirality, dgr/m**	**Sap-wood thickness, mm**	**Bark thickness, mm**	**Check-length, cm**	**Moisture Content, %**	**Number of checks per cross-section**
47	147.0	4.0	273.1							
48	101.2	2.3	266.7							
49	68.9	2.8	155.6							
50	78.6	2.6	190.5							
51	85.0	3.3	187.3							
52	61.5	2.7	133.4							
53	78.8	4.7	209.6							
54	56.3	4.2	165.1							
55	64.1	5.2	142.9							
56	51.4	2.5	168.3							
57	49.9	3.3	149.2							
58	60.5	5.1	130.2							
59	58.2	4.5	149.2							
60	74.7	4.4	146.1							
61	62.6	5.0	133.4							
62	58.7	4.4	127.0							
63	59.9	5.3	139.7							
64	52.9	5.4	123.8							
65	62.8	4.2	142.9							
66	57.8	5.7	139.7							
67	105.6	3.1	200.0							
68	80.6	3.1	181.0							
69	74.5	4.2	149.2							
70	81.3	4.6	165.1							
71	85.1	7.5	184.2							
72	76.2	4.2	222.3							
73	105.5	2.6	215.9							
74	76.0	4.9	168.3							
75	97.4	5.2	212.7							
76	47.2	2.0	149.2							
77	67.3	4.7	149.2							
				Additional 96 groups of measurements						
1	67.7	1.9	225	0.45	3.5	21	3.4	430	26.8	1
2	85.7	6.1	240	0.64	8.0	28	1.3	100	20.7	1
3	88.0	4.1	208	0.64	24.0	26	1.7	250	14.5	1
4	73.4	2.7	268	0.53	20.0	26	3.4	120	27.5	1
5	62.6	2.3	200	0.74	16.7	32	2.9	300	20.7	1
6	45.1	1.7	174	0.74	25.0	29	3.2	180	20.5	1
7	71.2	2.4	268	0.62	1.4	22	4.7	430	23.8	1
8	69.2	2.7	200	0.62	5.5	21	2.5	220	24.7	1
9	57.2	1.4	180	0.62	•	18	4.1	•	24.5	1
10	75.6	2.8	213	0.46	8.6	30	1.2	140	23.8	1
11	84.8	2.8	210	0.76	11.4	22	2.5	220	18.8	1
12	59.2	4.0	188	0.76	•	28	4.3	•	17.1	1
13	64.6	2.7	188	0.51	2.8	16	2.4	460	16.5	1
14	96.0	4.1	295	0.35	12.3	35	2.9	1060	20.5	1
15	97.3	4.6	265	0.35	•	30	2.9	•	16.5	1
16	98.1	6.4	246	0.35	•	37	2.9	•	15.0	1
17	80	3.0	240	0.61	8.2	41	2.6	61	18.6	1
18	77	3.6	215	0.61	12.3	40	2.6	244	17.7	1
19	76	4.7	230	0.59	1.4	46	3.1	366	19.7	1
20	82	4.5	195	0.59	10.9	43	2.4	457	17.4	1
21	70	4.0	300	0.85	9.8	34	3.6	30	21.4	1
22	89	3.7	270	0.85	4.1	37	3.7	488	19.3	1

	d	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
Sequence	**Check-depth, mm**	**Check-width, mm**	**Diameter, mm**	**Taper, cm/m**	**Spirality, dgr/m**	**Sap-wood thickness, mm**	**Bark thickness, mm**	**Check-length, cm**	**Moisture Content, %**	**Number of checks per cross-section**
23	57	1.5	287	0.88	10.9	66	4.2	91	18.7	2
24	69	2.5	287	0.88	13.1	66	4.2	152	18.5	2
25	90	6.5	258	0.88	6.6	65	2.4	152	16.1	1
26	88	8.8	235	0.88	19.7	64	2.2	457	16.7	1
27	108	8.2	285	0.67	3.6	45	4.3	274	20.4	1
28	98	8.4	273	0.72	7.5	58	3.0	335	18.8	1
29	81	3.9	245	0.72	27.9	56	4.1	305	21.0	1
30	71	2.4	215	0.72	43.7	51	3.2	183	17.3	1
31	103	5.8	293	0.71	8.2	42	1.8	427	20.4	1
32	70	2.6	240	0.71	6.6	50	3.2	76	22.0	1
33	55	4.2	250	0.60	32.8	42	0.5	30	18.6	1
34	107	7.6	220	0.60	9.8	45	1.0	305	17.5	1
35	83	4.4	243	0.62	1.7	55	3.9	117	18.9	1
36	83	3.9	228	0.62	0.0	55	2.3	168	20.6	1
37	83	7.7	210	0.62	2.3	50	1.6	442	15.7	1
38	84	4.2	273	0.83	13.1	46	3.0	30	20.6	1
39	86	6.8	245	0.83	26.8	46	1.6	168	17.5	1
40	84	3.9	268	0.72	27.7	13	1.9	488	25.0	1
41	83	3.8	223	0.72	7.6	21	1.7	105	18.5	1
42	79	6.1	263	1.33	13.3	29	2.6	600	16.9	1
43	96	5.7	375	1.53	2.0	30	2.5	254	•	1
44	80	0.7	335	0.89	0.0	39	•	457	21.6	1
45	76	2.0	280	0.89	0.0	32	•	274	14.7	1
46	107	10.8	288	0.60	2.5	35	•	122	19.9	1
47	67	4.8	250	0.60	1.3	41	•	229	20.6	1
48	115	5.3	365	1.21	4.1	58	•	244	20.1	1
49	64	1.5	290	1.21	2.3	60	•	213	18.1	1
50	96	4.5	290	0.81	9.4	45	•	213	17.4	1
51	84	3.7	240	0.81	22.7	48	•	396	16.1	1
52	89	4.4	280	0.81	26.2	29	•	305	18.6	1
53	159	12.7	373	0.84	17.7	26	•	508	19.1	1
54	127	12.0	330	0.84	•	27	•	•	17.7	1
55	84	5.3	285	1.98	4.1	54	•	122	17.6	1
56	108	4.5	365	1.13	5.5	45	•	274	20.4	1
57	89	1.0	295	1.13	24.6	41	•	61	22.3	1
58	93	4.7	280	0.64	5.5	32	•	366	20.9	1
59	99	3.4	290	2.14	8.2	48	•	61	17.6	1
60	82	1.8	378	0.97	42.7	31	•	610	24.6	2
61	110	5.3	378	0.97	•	31	•	610	24.6	2
62	125	7.1	318	0.97	0.0	30	•	610	21.5	1
63	98	2.7	323	0.76	98.4	40	•	91	20.2	2
64	104	4.8	323	0.76	6.0	40	•	335	19.5	2
65	80	4.4	275	0.76	22.4	44	•	335	18.6	1
66	140	4.3	340	1.01	26.2	27	•	152	25.3	2
67	121	3.2	340	1.01	24.6	27	•	122	25.2	2
68	110	7.1	333	0.76	0.0	28	•	213	20.7	1
69	69	4.5	285	0.76	2.1	32	•	488	19.6	1
70	84	7.5	305	1.29	2.3	35	•	347	19.3	1
71	79	3.9	235	0.81	14.6	17	•	137	19.5	1
72	57	2.9	190	0.81	6.8	27	•	74	17.4	1
73	86	13.5	250	0.60	2.1	36	•	244	18.3	1
74	86	6.5	213	0.60	0.0	35	•	274	16.4	1
75	88	8.5	248	0.85	6.8	34	•	74	16.8	1

	d	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
Sequence	**Check-depth, mm**	**Check-width, mm**	**Diameter, mm**	**Taper, cm/m**	**Spirality, dgr/m**	**Sap-wood thick-ness, mm**	**Bark thick-ness, mm**	**Check-length, cm**	**Mois-ture Content, %**	**Number of checks per cross-section**
76	125	5.9	328	0.89	0.0	49	•	107	21.4	1
77	89	1.9	273	0.89	0.0	52	•	107	23.7	1
78	100	8.4	233	0.52	0.0	18	•	122	18.9	1
79	64	3.5	200	0.52	0.0	21	•	61	15.2	1
80	111	4.6	260	1.13	17.9	39	•	168	18.3	1
81	75	4.8	248	0.86	•	23	•	•	20.1	1
82	75	5.6	205	0.86	23.4	25	•	427	16.9	1
83	75	3.6	185	0.40	10.9	18	•	46	15.4	1
84	110	7.6	305	0.76	11.9	36	•	168	18.6	1
85	105	6.1	258	0.76	3.3	40	•	152	17.7	1
86	98	5.7	325	0.85	1.6	53	•	305	21.5	1
87	105	6.9	273	0.85	6.6	45	•	305	16.7	1
88	100	4.1	365	1.41	21.9	47	•	411	28.3	1
89	118	3.2	283	1.41	19.7	37	•	229	17.6	1
90	69	3.1	330	0.89	1.8	30	•	274	21.3	1
91	94	4.3	275	0.89	8.2	27	•	61	20.7	1
92	99	8.8	283	0.89	28.1	27	•	107	18.6	1
93	92	4.9	228	0.89	29.5	27	•	152	19.5	1
94	100	4.5	270	1.53	0.0	36	•	610	20.5	1
95	129	6.7	330	0.89	33.5	34	•	119	18.8	1
96	80	3.6	275	0.89	9.8	32	•	51	15.6	1
Avg	86.5	4.7	238.1	0.8	12.2	36.9	2.8	258.4	19.6	1.1
Min	25.7	0.7	123.8	0.4	0.0	13.3	0.5	30.5	14.5	1.0
Max	159.0	13.5	381.0	2.1	98.4	66.0	4.7	1060.0	28.3	2.0
StD	22.5	2.1	61.1	0.3	14.0	12.3	1.0	179.7	2.9	0.3

Note: *• - Not available.*

Appendix 2: Glossary of statistical terms

Bias – is the average of differences between actual measurements and predicted values of the dependent variable:

$$B = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{n}$$

Where:
y_i is the i-th actual measurement of the dependent variable,
$\hat{y}_i$ is the predicted value of the dependent variable, calculated from the regression equation by substituting in the i-th group of independent variable measurements.

Statistically significant equation – A regression equation is statistically significant if the independent variables contribute significantly to the variation of the dependent variable. Usually an F or t-test is carried out to determine if the contribution is significant.

R^2 – is an indicator of what proportion of the variability of dependent variable can be explained by the independent variables. The formula for this calculation is:

$$R^2 = \frac{SS_{\mathrm{Re}\,gr}}{SS_{Total}}$$

Where:
SS_{Regr} is the regression sum of squares,
SS_{Total} is the total sum of squares.

Their computational formulas are in the following description.

Confidence interval of the regression equation – is the measure of prediction accuracy, i.e., the spread of regression surfaces resulting from repeated measurements.

The accuracy of regression equations is determined by the upper and lower limits of the confidence intervals of the estimation. They can be thought of as two surfaces being half a confidence interval away from the surface of the regression equation. As an example, Figure 4 shows the confidence interval of a regression equation estimating check-depth as a function of check-width and log diameter.

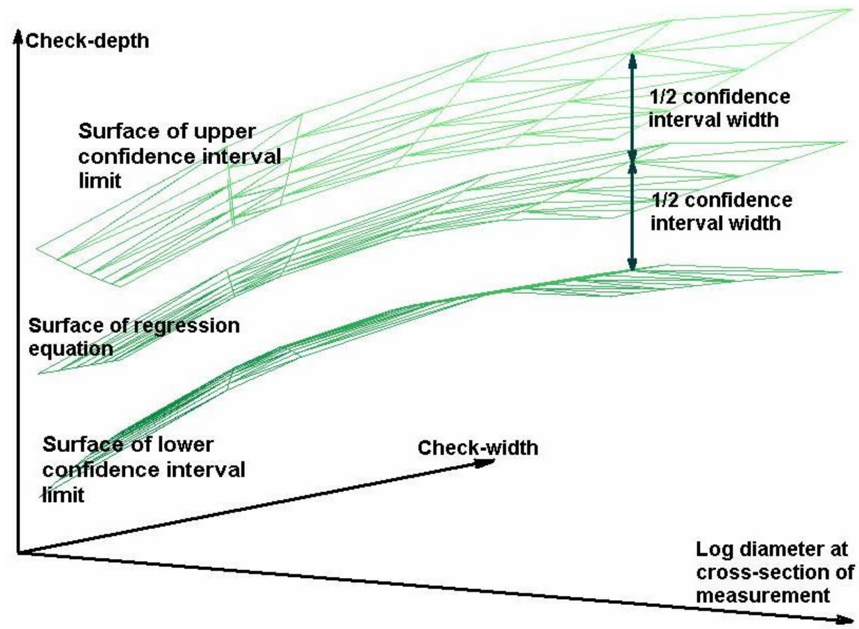


Figure App 2-1: Predicted check-depth values with upper and lower confidence limits.

Narrower confidence intervals indicate less spread of the predicted values calculated by the repeated equations and hence, more accurate estimation. An important feature of the confidence interval is that it changes with the independent variables, however, it is always the smallest at the location determined by the average of independent variables.

As an example, Appendix 4 shows the predicted $\hat{d}$ (check-depth) values with the corresponding widths of confidence intervals for 36 combinations of the two independent variables—check-width and log diameter—of regression Equation 2. The formula for calculating the co-ordinates of points determining the confidence interval limit surfaces of a multiple curvilinear regression equation, with 3 independent variables, is the following:

$$CInt = \hat{y}_{x_{1k},x_{2k},x_{3k}} \pm t_{\alpha/2_{(n-m-1)}} \sqrt{MS_{Res} * W}$$

Where:

$\hat{y}_{x_{1k},x_{2k},x_{3k}}$ is the predicted value calculated from the regression equation for the values of x_{1k}, x_{2k}, and x_{3k} independent variables.

$t_{\alpha/2(n-m-1)}$ is the tabulated *t* value with given *α/2* probability and *n-m-1* degrees of freedom,

n is the number of observations,

m is the number of independent variables,

n-m-1 is the degrees of freedom of the residual sum of squares,

MS_{Res} is the mean square residual or residual variance.

Where:

$$MS_{\mathrm{Re}s} = \frac{SS_{\mathrm{Re}s}}{n-(m+1)}$$

$$SS_{\mathrm{Re}s} = SS_{Total} - SS_{\mathrm{Re}gr}$$

$$SS_{Total} = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

$$SS_{\mathrm{Re}gr} = b_1 SP_{x_1 y} + b_2 SP_{x_2 y} + b_3 SP_{x_3 y}$$

$$SS_x = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

$$SP_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}$$

And:

$$W = \frac{1}{n} + \left[x_{1k} - \bar{x}_{1,} x_{2k} - \bar{x}_{2,} x_{3k} - \bar{x}_3\right] * \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} * \begin{bmatrix} x_{1k} - \bar{x}_1 \\ x_{2k} - \bar{x}_2 \\ x_{3k} - \bar{x}_3 \end{bmatrix}$$

Where:

- $\bar{x}_1$ is the average of x_{1i}'s,
- $\bar{x}_2$ is the average of x_{2i}'s,
- $\bar{x}_3$ is the average of x_{3i}'s,
- $\bar{x}_{1k}$ is a given point of x_{1i} for which the confidence interval is calculated.
- $\bar{x}_{2k}$ is a given point of x_{2i}, for which the confidence interval is calculated.
- $\bar{x}_{3k}$ is a given point of x_{3i}, for which the confidence interval is calculated.

The $\begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{32} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}$ matrix is the inverse of the $\begin{bmatrix} SS_{x1} & SP_{x1x2} & SP_{x1x3} \\ SP_{x1x2} & SS_{x2} & SP_{x2x3} \\ SP_{x1x3} & SP_{x2x3} & SS_{x3} \end{bmatrix}$ matrix,

SS_{x_i} is the sum of squares, i=1,2,3,…n
$SP_{x_i x_j}$ is the sum of products, i=1,2,3,…n; j=1,2,3,…n
i is the row index,
j is the column index.

Standard error of estimate of the regression equation – is the measure of the spread of the actual (measured) dependent variables around the surface determined by the regression equation. Its various computational formulae are as follows:

Standard error of estimate = SEE $= \sqrt{MS_{\mathrm{Re}\,s}}$ or $\sqrt{-\frac{\left(\sum_{i=1}^{n}(y_i - \hat{y}_i)\right)^2}{n}}$

%SEE = Standard error of estimate as percent of the average of dependent variable:

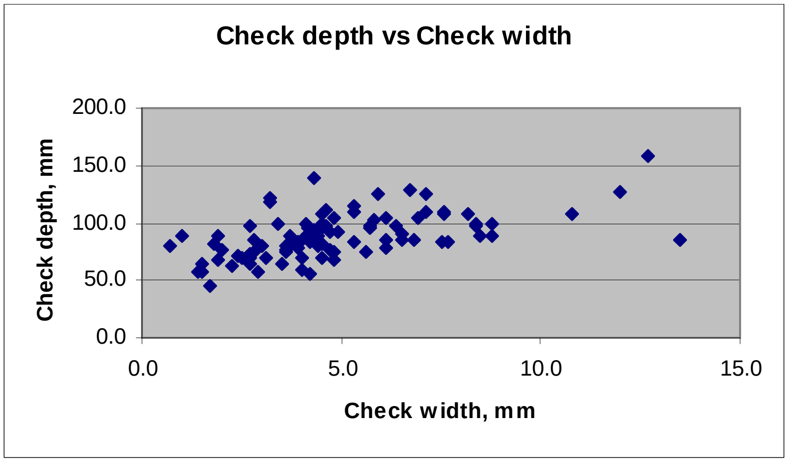
$$\%SEE = \frac{SEE}{\bar{y}}100$$

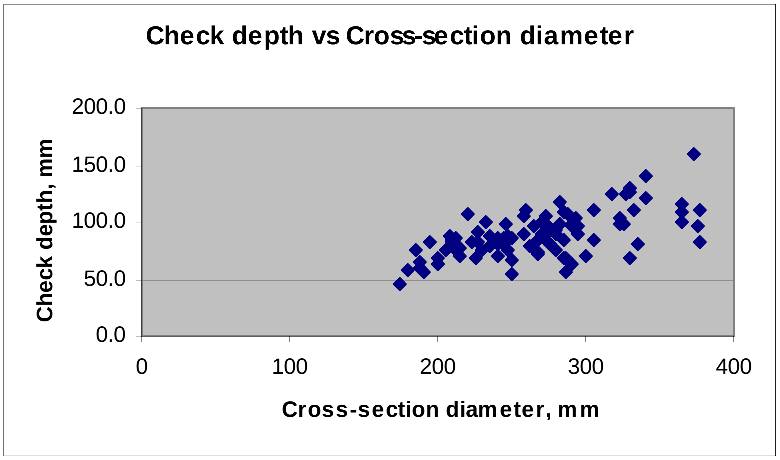
Where
$\bar{y}$ = average of check-depth measurements.

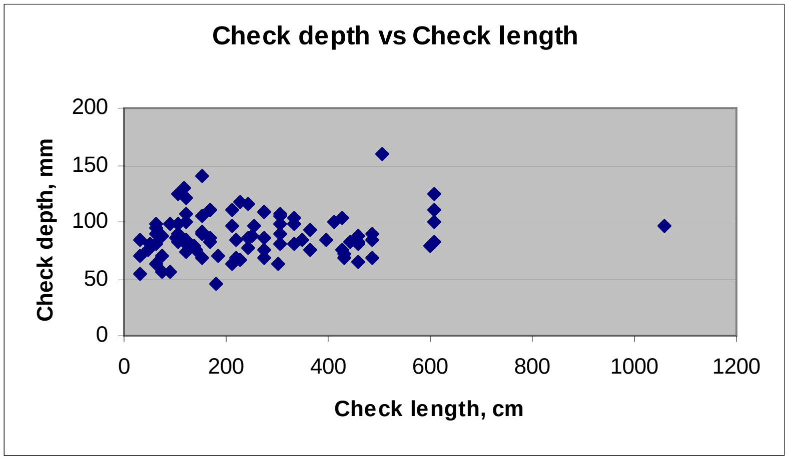
This is also called a **Relative error**.

Trend – A noticeable trend of the biases around the regression surface.

Appendix 3: Scattergram examples demonstrating the relationship of dependent and individual independent variables.






Check depth vs Check length
Check depth, mm
200
150
100
50
0
0
200
400
600
800
1000
1200
Check length, cm

Appendix 4: Predicted check-depths and their confidence intervals as function of check-width and diameter.

Check-width, mm	Diameter, mm	Predicted $\hat{d}$, mm	Width of confidence interval, mm
2.5	130	43.8	17.4
2.5	180	71.5	12.2
2.5	200	80.4	12.7
2.5	250	97.1	14.5
2.5	300	105.8	19.4
2.5	350	106.6	34.6
3.5	130	47.4	15.6
3.5	180	75.1	9.2
3.5	200	84.0	9.8
3.5	250	100.7	11.8
3.5	300	109.4	17.3
3.5	350	110.2	33.3
4.5	130	51.0	15.1
4.5	180	78.7	8.0
4.5	200	87.6	8.6
4.5	250	104.3	10.5
4.5	300	113.0	16.2
4.5	350	113.9	32.7
5.5	130	54.6	16.0
5.5	180	82.4	9.3
5.5	200	91.2	9.6
5.5	250	107.9	11.2
5.5	300	116.6	16.5
5.5	350	117.5	32.7
6.5	130	58.2	18.0
6.5	180	86.0	12.2
6.5	200	94.8	12.4
6.5	250	111.5	13.4
6.5	300	120.2	17.9
6.5	350	121.1	33.4
8.5	130	65.5	24.2
8.5	180	93.2	20.1
8.5	200	102.1	20.1
8.5	250	118.7	20.4
8.5	300	127.5	23.4
8.5	350	128.3	36.4
Confidence interval at the average of independent variables:			**8.6029**

Appendix 5: Ranking of regression equations

Equations were ranked from various aspects of prediction efficiency: 1) lack of fit statistics: *%Bias, %SEE,* and *Trend,* (calculated for 5 frequency classes of each independent variable); and 2) fit statistics: R^2, *SEE* and *CL* (Confidence interval of the regression equation) calculated for the whole equation– see Glossary of statistical terms, Appendix 1. The total of ranking points is a combination of individual rankings and helped to select the "best" regression equation. The lower the total score the higher the ranking.

Excluding x_6 increased the number of measurement groups to 88. The analysis resulted in three different equations and these are compared in Table App 5 – 1.

Table App 5- 1: Ranking of Equations 4, 5 and 9.

Equation	%Bias	%SEE	Trend	R^2	SEE	CL	Total
4	2.5	1.5	3	3	3	2	15.0
5	1	1.5	1	2	2	3	10.5
9	2.5	3	2	1	1	1	10.5

Using only the independent variables x_1 and x_2 increased the number of measurement groups to 173. This led to three more equations that are compared in Table App 5 – 2.

Table App 5 – 2.: Ranking of Equations 6, 7 and 8.

Equation	%Bias	%SEE	Trend	R^2	SEE	CL	Total
6	1	1	2	1	1	3	9
7	2.5	2.5	2	2	2	2	13
8	2.5	2.5	2	3	3	1	14

The two best equations from the above comparisons are now compared along with Equation 10 (the one without the check width as an independent variable). As there are two ties in the first table this led to the following two tables being generated.

Table App 5 – 3: Ranking of Equations 5, 6 and 10.

Equation	%Bias	%SEE	Trend	R^2	SEE	CL	Total
5	1	1	1	1	1	1	6.0
6	2.5	2.5	2	2	2	2	14.0
10	2.5	2.5	3	3	3	3	16.0

Table App 5 – 4: Ranking of Equations 6, 9 and 10.

Equation	%Bias	%SEE	Trend	R^2	SEE	CL	Total
6	2	2	2	2	3	2	13
9	3	1	2	1	1	1	9
10	1	3	2	3	2	3	14