

Article

Not peer-reviewed version

Wood Quality Assessment of Standing Stems When Measurements Do Not Line Up: A Knot Geometry Model for Spatially Misaligned Tree Data

[Udayalakshmi Vepakomma](#)^{*}, [Isabelle Duchesne](#), Magloire Loudegui Djimdou, [Arusharka Sen](#)

Posted Date: 19 March 2026

doi: 10.20944/preprints202603.1527.v1

Keywords: non-coincident observations; kernel density estimation; CT scanning; knot characteristics; knot modeling; log sort; white pine; red pine



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Wood Quality Assessment of Standing Stems When Measurements Do Not Line Up: A Knot Geometry Model for Spatially Misaligned Tree Data

Udayalakshmi Vepakomma ^{1,*†}, Isabelle Duchesne ², Magloire Loudegui Djimdou ³ and Arusharka Sen ³

¹ FPIInnovations, 570 Blvd. St Jean, Pointe-Claire, QC H9R 3J9

² Natural Resources Canada, Canadian Forest Service, Laurentian Forestry Centre, 1055 Du P.E.P.S. Street, P.O. Box 10380, Québec, QC G1V 4C7, Canada

³ Department of Statistics, Concordia University, 1455 Blvd. De Maisonneuve Ouest, Montreal, Quebec H3G 1M8

* Correspondence: udayavepa@gmail.com

† Present address: Independent Scientist, 6705 Kirwan, Cote St-Luc, QC H4V 1B6.

Abstract

Non-destructive assessment of wood quality in standing trees is increasingly essential for value-based harvesting, precision forestry, and large-scale monitoring. Yet most existing approaches rely on destructive sampling or plot-scale measurements that cannot be feasibly extended across extensive forest areas. A major barrier to scalable assessment is the spatial misalignment between external structural measurements and internal wood-quality responses, which introduces systematic bias when conventional regression methods are applied. We propose a generalized statistical framework called Regression based on Misaligned Covariates (RMC), that reconciles covariates and responses measured at separate spatial locations. By combining the Law of Total Expectation with kernel-based estimation of marginal relationships, RMC recovers height-dependent conditional means without requiring one-to-one spatial correspondence. The framework is demonstrated using two physiologically contrasting conifers, eastern white pine (*Pinus strobus*) and red pine (*Pinus resinosa*), using covariates that are directly measurable through non-destructive, remote-sensing-based structural characterization. RMC produced smooth, biologically interpretable estimates of knot volume along the bole for both species, even when trained on small datasets and using only a single structural covariate in addition to height. The resulting cumulative knot-volume profiles captured species-specific differences in knot accumulation and aligned with known patterns of crown architecture and stem form. RMC provides a scalable, non-destructive pathway for predicting internal wood quality from external structural measurements, supporting pre-harvest planning, log-sorting decisions, and long-term monitoring of stand development. Because the method is portable, parsimonious, and not species-specific, it offers a general solution for any ecological or remote-sensing application involving misaligned covariates and responses, extending well beyond wood-quality modelling.

Keywords: non-coincident observations; kernel density estimation; CT scanning; knot characteristics; knot modeling; log sort; white pine; red pine

1. Introduction

In North America, the assessment of standing wood quality has emerged as a critical early intervention strategy in modern forest management, marking a fundamental shift from traditional volume-based metrics to value-based recovery [1,2]. This paradigm shift is essential for optimizing the wood supply chain, ensuring the structural soundness of timber products, and maximizing

market value while simultaneously reducing waste and supporting sustainable silvicultural practices [3–5]. By implementing assessment protocols early in the production cycle, forest managers can facilitate the strategic segregation of timber by quality prior to harvest. For example, adding only simple visual parameters like number of branches in the lower part of the bole has significantly improved prediction of lumber yield [6,7]. This proactive approach enables the precise allocation of raw materials to their most economically viable end-uses, such as distinguishing high-stiffness structural lumber from pulpwood, thereby maximizing the return on investment for each individual stem [3,2].

Beyond immediate economic gains, early assessment is a prerequisite for meeting the rigorous mechanical standards for strength and stiffness required by the modern construction industry [1,4]. Instead of viewing timber as a generic product, managers must align forest growth with industry demands by predicting wood quality long before harvest. Predicting the quality of potential products within specific growing environments allows managers to tailor silvicultural prescriptions to optimize both the quantity and internal characteristics of the standing resource [8,9]. Because the standing tree represents the origin of the forest products value chain, it is the point at which wood quality is most susceptible to management influence [10]. Consequently, targeting wood quality during silvicultural planning creates value that extends beyond mere fiber extraction, allowing for more robust decision-making under varying environmental and management scenarios [3,8]. However, to realize these benefits across broad forest landscapes, it is necessary to move beyond plot-level sampling toward integrated, large-scale assessment methodologies

In recent years, LiDAR (Light Detection and Range) has evolved from a simple inventory tool and volume assessment to a sophisticated technology that either substitutes or complements traditional Non-Destructive Evaluation (NDE) by predicting internal traits via external structure and high-resolution geometry [2]. Although current resolutions are insufficient to detect individual branch scars, high resolution LiDAR have been successfully employed to delineate crown features, canopy density, bole characteristics such as crown base height (CBH), diameter along the bole, stem straightness (sweep) *etc.* These metrics serve as effective proxies for wood quality in various stand types [11–14]. Furthermore, airborne LiDAR (ALS) is seen to be effective for predicting site-level wood density and general stem dimensions across large territories [14]. However, ALS often suffers from significant occlusion near-nadir scanning angles and canopy interference, which prevents detailed characterization of the lower stem where the most valuable sawlogs are located [15]. Operating at lower altitudes, slower speed and multiple scanning perspectives, drone or UAV based LiDAR (ULS), with its significantly high density can penetrate the canopy more effectively to capture lower-stem morphology [13,16].

The size and distribution of knots are widely recognized as critical determinants of wood quality, directly influencing lumber grading, mechanical performance, and aesthetic value [2,17,18]. Although restricted to plot-level assessments, the proximity and perspective of Terrestrial (TLS) and Mobile Laser Scanning (MLS) enable the capture of millimeter-level point clouds. These high-density datasets allow for the accurate 3D reconstruction of woody structures, including the measurement of taper, sweep, and ovality [19,20], as well as the detection of branch scars and whorl patterns [21,22].

Recent studies have leveraged these external morphological features to infer internal characteristics; for instance, TLS data has been correlated with CT scans to achieve 62.5% accuracy in the automated quality grading of European beech sawlogs [23]. Furthermore, when integrated with resistance drilling, these scanners can predict wood basic density with moderate success of an R^2 of 0.51 [14]. While they provide high-resolution knot-level detail required for internal trait prediction, their limited spatial reach necessitates a bridge toward more scalable platforms for operational efficiency. One strategy for scaling wood quality predictions could be by leveraging ULS-derived crown and bole metrics as inputs for empirical models. These models, calibrated using meticulous ground-based data, allow for the translation of external structural geometry into broad-scale wood quality indicators.

Despite the high-resolution capabilities, spatial misalignment between external predictors and internal responses remains a fundamental bottleneck when integrating multi-source datasets [24]. In wood quality modeling, the covariates derived from LiDAR (such as external bole surface irregularities or branch metrics along the bole) and the response variables (such as internal knot geometry from destructive sampling or Computed Tomography - CT scans) are rarely evaluated at the same spatial coordinates. Despite efforts, this “change of support” problem introduces significant bias, as traditional regression techniques often assume a one-to-one spatial correspondence between variables [25–27]. Standard alignment methods frequently fail to account for the irregular 3D topology of the tree or the sensor-specific occlusions that shift the perceived location of a feature [28,29]. Consequently, there is a critical need for a generalized modeling framework capable of reconciling these disjointed data points without sacrificing the localized precision required for knot-level analysis or even log-segment level analysis.

The primary objective of this study is to develop a generalized modeling framework for cases where the covariate (e.g., external tree characteristics) and the response variable (e.g., knot geometry) are evaluated at disjointed spatial points. By leveraging the Law of Total Expectation in conjunction with Kernel Density Estimation, we propose a novel statistical method, called regression of misaligned covariates (RMC), that reconciles these misaligned datasets, enabling the prediction of internal wood quality attributes from external structural tree geometry.

The development of such a framework requires validation across contrasting physiological profiles; thus, red pine (*Pinus resinosa*) and eastern white pine (*Pinus strobus*) are utilized here as distinct prototypes of conifer physiology. Red pine, characterized by exceptional genetic and structural uniformity, serves as a stable baseline for quality-related assessments [30,31]. In contrast, eastern white pine presents significant geometric complexity due to its discrete branching architecture and susceptibility to terminal leader damage [32–34]. These biological stressors often result in morphological irregularities, including crooks, sweeps, and multi-leader crowns as well as positioning the species as a rigorous testbed for LiDAR-based algorithms designed to map knot distribution and volumetric recovery [35,36].

Through this dual-species validation, we illustrate the estimation of primary wood quality metrics in standing trees and demonstrate how these data can directly inform management decisions and industrial processing. By bridging the gap between raw spatial data and volumetric recovery, this research ultimately provides a robust, scalable pathway for assessing wood quality at both the individual tree and sub-tree levels across extensive forest blocks.

2. Materials and Methods

2.1. Regression of Misaligned Covariates (RMC)

2.1.1. Model Presentation

The goal is to predict the knot geometry, such as volume, diameter and size, using tree external characteristics such as log diameter, branch length and branch diameter measured along the height (or length) of the tree bole. However, in contrast with the usual regression framework, here we deal with *misaligned* covariates *i.e.*, when the data for the covariates and the response variables are collected at different times and at varying heights along the bole. Specifically:

- the knot geometry measurements (internal characteristics) are available at certain height points.
- the external tree characteristics are recorded at other, possibly different, height points.

Given Y as a response variable (indexed by 0) as tree external characteristics, X as a covariate (indexed by 1) as external characteristics, and H as the relative height (indexed by 2), the sample measurements could be denoted as:

$$\{(Y_i, H_i)\}_{i=1}^{i=n_{02}} \text{ and } \{(X_i, H_i)\}_{i=1}^{i=n_{12}}$$

The objective, hence, is to link Y to X , taking note that the Y_i 's to the X_i 's are misaligned *i.e.* are not available at the same height points. Based on the available data, one can then estimate

$$m_{02}(h) := \mathbb{E}(Y | H = h) \quad (1)$$

$$m_{12}(h) := \mathbb{E}(X | H = h) \quad (2)$$

The goal is to estimate its regression given as:

$$m_{012}(x, h) := \mathbb{E}(Y | X = x, H = h) \quad (3)$$

The proposed approach here is based on the well-known result that the expectation of conditional expectation is an unconditional expectation, or the law of total expectation, and given by the following equations:

$$\begin{cases} \mathbb{E}_X[\mathbb{E}(Y | X, h) | H = h] = \mathbb{E}[Y | H = h] \\ \mathbb{E}_{X,H}[\mathbb{E}(Y | X, H)] = \mathbb{E}[Y] \end{cases} \quad (4)$$

$$\quad (5)$$

For its estimation, a suitable finite-dimensional model for $m_{012}(x, h)$ could be chosen, for example a form like:

$$m_{012}(x, h) = a_0 + a_1x + a_2h + a_3x^2 + a_4h^2 + \dots \quad (6)$$

and solve for the unknowns $a_0, a_1, \dots$

Equations (4) and (5) can be written more precisely as,

$$\int \left(\int \frac{yf_{012}(y, x, h)dy}{f_{12}(x, h)} \right) \frac{f_{12}(x, h)dx}{f_2(h)} = \int \frac{yf_{02}(y, h)dy}{f_2(h)} \quad (7)$$

$$\int \left(\int \frac{yf_{012}(y, x, h)dy}{f_{12}(x, h)} \right) f_{12}(x, h)dx dh = \int \int yf_{012}(y, x, h)dy dx dh \quad (8)$$

where it is assumed that

- (Y, X, H) has joint density $f_{012}(y, x, h)$
- (Y, H) has joint density $f_{02}(y, h)$
- (X, H) has joint density $f_{12}(x, h)$
- H has marginal density $f_2(h)$,

and

$$\int f_{012}(y, x, h)dx = f_{02}(y, h) \quad (9)$$

$$\text{and } \int f_{012}(y, x, h)dx dh = f_0(y) \quad (10)$$

$f_0(y)$ being the marginal density of Y .

2.1.2. Regression Estimation Approach

Cancelling $f_2(h)$ from Equation (7), one can write

$$\int \left(\int \frac{yf_{012}(y, x, h)dy}{f_{12}(x, h)} \right) \frac{f_{12}(x, h)dx}{f_2(h)} = \int \frac{yf_{02}(y, h)dy}{f_2(h)} \text{ for all } h \quad (11)$$

$$\Leftrightarrow \int m_{012}(x, h)f_{12}(x, h)dx = \int yf_{02}(y, h)dy \text{ for all } h \quad (12)$$

so that Equations (7) and (8) become

$$\left\{ \begin{aligned} \int m_{012}(x, h)f_{12}(x, h)dx &= \int yf_{02}(y, h)dy \\ \mathbb{E}[m_{012}(X, H)] &= \mathbb{E}(Y) \end{aligned} \right. \quad (13)$$

$$\quad (14)$$

Plugging in the corresponding kernel estimators on both sides we have

$$\int m_{012}(x, h)\hat{f}_{12}(x, h)dx = \int y\hat{f}_{02}(y, h)dy \quad (15)$$

$$\left\{ \begin{aligned} \frac{1}{n_{12}} \sum_{i=1}^{n_{12}} m_{012}(X_i, H_i) &= \frac{1}{n_{02}} \sum_{i=1}^{n_{02}} Y_i \end{aligned} \right. \quad (16)$$

where

$$\int \hat{f}_{12}(x,h)dx = \frac{1}{n_{12}b_n^2} \sum_{i=1}^{n_{12}} K\left(\frac{x-X_i}{b_n}\right) K\left(\frac{h-H_i}{b_n}\right) \text{ for } j = 1, \dots, n_{12} \tag{17}$$

$$\int \hat{f}_{02}(y,h)dx = \frac{1}{n_{02}b_n^2} \sum_{i=1}^{n_{02}} K\left(\frac{y-Y_i}{b_n}\right) K\left(\frac{h-H_i}{b_n}\right) \text{ for } j = 1, \dots, n_{02} \tag{18}$$

and $K(.)$ is a kernel. See, for example, [37] for an introduction to kernel smoothing.
The above equations (15) and (16) can be solved by choosing a suitable finite-dimensional model for

$$m_{012}(x,h)$$

as mentioned above, for example, assuming

$$m(x,h) = a_0 + a_1x + a_2h + a_3x^2 + a_4h^2 + \dots \tag{19}$$

Coefficients $a_0, a_1, \dots$, can be solved by creating as many equations as the number of coefficients by considering different values of h : $h_0, h_1, h_2, \dots$

To illustrate the proposed estimation method, the following subsection examines a tri-variate normal distribution of the variables of interest.

2.1.3. (Y,X,H) Follow a 3- Variate Normal

Suppose

$$\begin{pmatrix} Y \\ X \\ H \end{pmatrix} \sim \mathcal{N}_3\left(\begin{pmatrix} \mu_0 \\ \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma\right) \tag{20}$$

From the properties of multivariate Normal distributions, the expectations $m_{12}(h) = \mathbb{E}(X \mid H = h)$, $m_{02}(h) = \mathbb{E}(Y \mid H = h)$ and $m(x,h) = m_{012}(x,h) = \mathbb{E}(Y \mid X = x, H = h)$ are linear.

Thus we may write

$$\begin{aligned} m_{12}(h) &= \mathbb{E}(X \mid H = h) = b_0 + b_2h & (21) \\ \text{and } m_{02}(h) &= \mathbb{E}(Y \mid H = h) = c_0 + c_2h. & (22) \end{aligned}$$

And

$$m(x,h) = m_{012}(x,h) = \mathbb{E}(Y \mid X = x, H = h) = a_0 + a_1x + a_2h \tag{23}$$

Equations (4) and (5) then become

$$\begin{aligned} \begin{cases} a_0 + a_1(b_0 + b_2h) + a_2h = c_0 + c_2h, \text{ for all } h \\ a_0 + a_1\mu_1 + a_2\mu_2 = \mu_0 \end{cases} & (24) \\ & (25) \end{aligned}$$

This yields a system of three equations with three unknowns (a_0, a_1, a_2):

$$\begin{aligned} \begin{cases} a_0 + b_0a_1 = c_0 \\ b_2a_1 + a_2 = c_2 \\ a_0 + a_1\mu_1 + a_2\mu_2 = \mu_0 \end{cases} & (26) \\ & (27) \\ & (28) \end{aligned}$$

That can be represented in a matrix form as:

$$\begin{pmatrix} 1 & b_0 & 0 \\ 0 & b_2 & 1 \\ 1 & \mu_1 & \mu_2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} c_0 \\ c_2 \\ \mu_0 \end{pmatrix} \tag{29}$$

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 1 & b_0 & 0 \\ 0 & b_2 & 1 \\ 1 & \mu_1 & \mu_2 \end{pmatrix}^{-1} \begin{pmatrix} c_0 \\ c_2 \\ \mu_0 \end{pmatrix} \tag{30}$$

Given a sample,

$$\hat{\mu}_0 = \frac{1}{n_{02}} \sum_{i=1}^{n_{02}} Y_i \quad (31)$$

$$\hat{\mu}_1 = \frac{1}{n_{12}} \sum_{i=1}^{n_{12}} X_i \quad (32)$$

$$\hat{\mu}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} H_i = \frac{1}{n_{02} + n_{12}} \left(\sum_{i=1}^{n_{02}} H_i + \sum_{i=1}^{n_{12}} H_i \right) \quad (33)$$

where b_0, b_2, c_0 and c_2 can be estimated from Equations (21) and (22) using Ordinary Least Squares (OLS) technique. This should yield

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 1 & \hat{b}_0 & 0 \\ 0 & \hat{b}_2 & 1 \\ 1 & \hat{\mu}_1 & \hat{\mu}_2 \end{pmatrix}^{-1} \begin{pmatrix} \hat{c}_0 \\ \hat{c}_2 \\ \hat{\mu}_0 \end{pmatrix} \quad (34)$$

2.1.4. Generalisation of the Model

To demonstrate the flexibility of the proposed framework, three specific functional forms for $m(x, h)$ are evaluated and their estimations presented in this section.

Case 1: $m(x, h)$ is linear

Given $m(x, h)$ is linear,

$$m_{012}(x, h) = \mathbb{E}(Y | X = x, H = h) = a_0 + a_1 x + a_2 h \quad (35)$$

And equations (15) and (16) become:

$$\begin{cases} \int \int (a_0 + a_1 x + a_2 h) \hat{f}_{12}(x, h) dx = \int y \hat{f}_{02}(y, h) dy \\ \frac{1}{n_{12}} \sum_{i=1}^{n_{12}} (a_0 + a_1 X_i + a_2 H_i) = \frac{1}{n_{02}} \sum_{i=1}^{n_{02}} Y_i \end{cases} \quad (36)$$

$$\begin{cases} \int \int (a_0 + a_1 x + a_2 h) \hat{f}_{12}(x, h) dx = \int y \hat{f}_{02}(y, h) dy \\ a_0 + a_1 \bar{X}_{n_{12}} + a_2 \bar{H}_{n_{12}} = \bar{Y}_{n_{02}} \end{cases} \quad (37)$$

$$\quad (38)$$

$$\quad (39)$$

where

$$\bar{X}_{n_{12}} = \iint x \hat{f}_{12}(x, h) dx dh = \frac{1}{n_{12}} \sum_{i=1}^{n_{12}} X_i \quad (40)$$

$$\bar{H}_{n_{12}} = \iint h \hat{f}_{12}(x, h) dx dh = \frac{1}{n_{12}} \sum_{i=1}^{n_{12}} H_i \quad (41)$$

$$\bar{Y}_{n_0} = \int y \hat{f}_2(y) dy = \frac{1}{n_0} \sum_{i=1}^{n_0} Y_i \quad (42)$$

Taking $h = h_1$ and $h = h_2$ in the first equation (Equation 36), we have

$$\int \int (a_0 + a_1 x + a_2 h_1) \hat{f}_{12}(x, h_1) dx = \int y \hat{f}_{02}(y, h_1) dy \quad (43)$$

$$\int \int (a_0 + a_1 x + a_2 h_2) \hat{f}_{12}(x, h_2) dx = \int y \hat{f}_{02}(y, h_2) dy \quad (44)$$

$$a_0 + a_1 \bar{X}_{n_{12}} + a_2 \bar{H}_{n_{12}} = \bar{Y}_{n_{02}} \quad (45)$$

$$\begin{cases} a_0 \int \hat{f}_{12}(x, h_1) dx + a_1 \int x \hat{f}_{12}(x, h_1) dx + a_2 h_1 \int \hat{f}_{12}(x, h_1) dx = \int y \hat{f}_{02}(y, h_1) dy \\ a_0 \int \hat{f}_{12}(x, h_2) dx + a_1 \int x \hat{f}_{12}(x, h_2) dx + a_2 h_2 \int \hat{f}_{12}(x, h_2) dx = \int y \hat{f}_{02}(y, h_2) dy \end{cases} \quad (46)$$

$$\begin{cases} a_0 \int \hat{f}_{12}(x, h_1) dx + a_1 \int x \hat{f}_{12}(x, h_1) dx + a_2 h_1 \int \hat{f}_{12}(x, h_1) dx = \int y \hat{f}_{02}(y, h_1) dy \\ a_0 \int \hat{f}_{12}(x, h_2) dx + a_1 \int x \hat{f}_{12}(x, h_2) dx + a_2 h_2 \int \hat{f}_{12}(x, h_2) dx = \int y \hat{f}_{02}(y, h_2) dy \end{cases} \quad (47)$$

$$\begin{cases} a_0 \int \hat{f}_{12}(x, h_1) dx + a_1 \int x \hat{f}_{12}(x, h_1) dx + a_2 h_1 \int \hat{f}_{12}(x, h_1) dx = \int y \hat{f}_{02}(y, h_1) dy \\ a_0 \int \hat{f}_{12}(x, h_2) dx + a_1 \int x \hat{f}_{12}(x, h_2) dx + a_2 h_2 \int \hat{f}_{12}(x, h_2) dx = \int y \hat{f}_{02}(y, h_2) dy \end{cases} \quad (48)$$

$$\begin{cases} a_0 \hat{f}_2(h_1) + a_1 \int x \hat{f}_{12}(x, h_1) dx + a_2 h_1 \hat{f}_2(h_1) = \int y \hat{f}_{02}(y, h_1) dy \\ a_0 \hat{f}_2(h_2) + a_1 \int x \hat{f}_{12}(x, h_2) dx + a_2 h_2 \hat{f}_2(h_2) = \int y \hat{f}_{02}(y, h_2) dy \end{cases} \quad (49)$$

$$\begin{cases} a_0 \hat{f}_2(h_1) + a_1 \int x \hat{f}_{12}(x, h_1) dx + a_2 h_1 \hat{f}_2(h_1) = \int y \hat{f}_{02}(y, h_1) dy \\ a_0 \hat{f}_2(h_2) + a_1 \int x \hat{f}_{12}(x, h_2) dx + a_2 h_2 \hat{f}_2(h_2) = \int y \hat{f}_{02}(y, h_2) dy \end{cases} \quad (50)$$

$$\begin{cases} a_0 \hat{f}_2(h_1) + a_1 \int x \hat{f}_{12}(x, h_1) dx + a_2 h_1 \hat{f}_2(h_1) = \int y \hat{f}_{02}(y, h_1) dy \\ a_0 \hat{f}_2(h_2) + a_1 \int x \hat{f}_{12}(x, h_2) dx + a_2 h_2 \hat{f}_2(h_2) = \int y \hat{f}_{02}(y, h_2) dy \end{cases} \quad (51)$$

$$\begin{pmatrix} \hat{f}_2(h_1) & \int x \hat{f}_{12}(x, h_1) dx & h_1 \hat{f}_2(h_1) \\ \hat{f}_2(h_2) & \int x \hat{f}_{12}(x, h_2) dx & h_2 \hat{f}_2(h_2) \\ 1 & \bar{X}_{n_{12}} & \bar{H}_{n_{12}} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \int y \hat{f}_{02}(y, h_1) dy \\ \int y \hat{f}_{02}(y, h_2) dy \\ \bar{Y}_{n_{02}} \end{pmatrix} \quad (52)$$

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \hat{f}_2(h_1) & \int x \hat{f}_{12}(x, h_1) dx & h_1 \hat{f}_2(h_1) \\ \hat{f}_2(h_2) & \int x \hat{f}_{12}(x, h_2) dx & h_2 \hat{f}_2(h_2) \\ 1 & \bar{X}_{n_{12}} & \bar{H}_{n_{12}} \end{pmatrix}^{-1} \begin{pmatrix} \int y \hat{f}_{02}(y, h_1) dy \\ \int y \hat{f}_{02}(y, h_2) dy \\ \bar{Y}_{n_{02}} \end{pmatrix} \quad (53)$$

where

$$\hat{f}_2(h_j) = \frac{1}{n_2 b_n} \sum_{i=1}^{n_2} K\left(\frac{h_j - H_i}{b_n}\right) \text{ for } j = 1, \dots, n_2 \quad (54)$$

$$\int x \hat{f}_{12}(x, h_j) dx = \frac{1}{n_{12} b_n} \sum_{i=1}^{n_{12}} X_i K\left(\frac{h_j - H_i}{b_n}\right) \text{ for } j = 1, \dots, n_{12} \quad (55)$$

$$\int y \hat{f}_{02}(y, h_j) dy = \frac{1}{n_{02} b_n} \sum_{i=1}^{n_{02}} Y_i K\left(\frac{h_j - H_i}{b_n}\right) \text{ for } j = 1, \dots, n_{02} \quad (56)$$

Case 2: $m(x, h)$ is quadratic

When $m(x, h)$ is quadratic,

$$m_{012}(x, h) = \mathbb{E}(Y | X = x, H = h) = a_0 + a_1 x + a_2 h + a_3 x^2 + a_4 h^2 + a_5 xh \quad (57)$$

And equations (15) and (16) become:

$$\int \int (a_0 + a_1 x + a_2 h + a_3 x^2 + a_4 h^2 + a_5 xh) \hat{f}_{12}(x, h) dx = \int y \hat{f}_{02}(y, h) dy \quad (58)$$

$$\left\{ \frac{1}{n_{12}} \sum_{i=1}^{n_{12}} (a_0 + a_1 X_i + a_2 H_i + a_3 X_i^2 + a_4 H_i^2 + a_5 X_i H_i) = \frac{1}{n_{02}} \sum_{i=1}^{n_{02}} Y_i \right. \quad (59)$$

$$\left\{ a_0 \hat{f}_2(h) + a_1 \int x \hat{f}_{12}(x, h) dx + a_2 h \hat{f}_2(h) + a_3 \int x^2 \hat{f}_{12}(x, h) dx + a_4 h^2 \hat{f}_2(h) + a_5 h \int x \hat{f}_{12}(x, h) dx = \int y \hat{f}_{02}(y, h) dy \right. \quad (60)$$

$$\left\{ a_0 + a_1 \bar{X}_{n_{12}} + a_2 \bar{H}_{n_{12}} + \frac{a_3}{n_{12}} \sum_{i=1}^{n_{12}} X_i^2 + \frac{a_4}{n_{12}} \sum_{i=1}^{n_{12}} H_i^2 + \frac{a_5}{n_{12}} \sum_{i=1}^{n_{12}} X_i H_i = \bar{Y}_{n_{02}} \right. \quad (61)$$

Taking $h = h_1$ and $h = h_2$ for the first equation, we have

$$\left\{ a_0 \hat{f}_2(h_1) + a_1 \int x \hat{f}_{12}(x, h_1) dx + a_2 h_1 \hat{f}_2(h_1) + a_3 \int x^2 \hat{f}_{12}(x, h_1) dx + a_4 h_1^2 \hat{f}_2(h_1) + a_5 h_1 \int x \hat{f}_{12}(x, h_1) dx = \int y \hat{f}_{02}(y, h_1) dy \right. \quad (62)$$

$$\left\{ a_0 \hat{f}_2(h_2) + a_1 \int x \hat{f}_{12}(x, h_2) dx + a_2 h_2 \hat{f}_2(h_2) + a_3 \int x^2 \hat{f}_{12}(x, h_2) dx + a_4 h_2^2 \hat{f}_2(h_2) + a_5 h_2 \int x \hat{f}_{12}(x, h_2) dx = \int y \hat{f}_{02}(y, h_2) dy \right. \quad (63)$$

$$\left\{ a_0 \hat{f}_2(h_3) + a_1 \int x \hat{f}_{12}(x, h_3) dx + a_2 h_3 \hat{f}_2(h_3) + a_3 \int x^2 \hat{f}_{12}(x, h_3) dx + a_4 h_3^2 \hat{f}_2(h_3) + a_5 h_3 \int x \hat{f}_{12}(x, h_3) dx = \int y \hat{f}_{02}(y, h_3) dy \right. \quad (64)$$

$$\left\{ a_0 \hat{f}_2(h_4) + a_1 \int x \hat{f}_{12}(x, h_4) dx + a_2 h_4 \hat{f}_2(h_4) + a_3 \int x^2 \hat{f}_{12}(x, h_4) dx + a_4 h_4^2 \hat{f}_2(h_4) + a_5 h_4 \int x \hat{f}_{12}(x, h_4) dx = \int y \hat{f}_{02}(y, h_4) dy \right. \quad (65)$$

$$\left\{ a_0 \hat{f}_2(h_5) + a_1 \int x \hat{f}_{12}(x, h_5) dx + a_2 h_5 \hat{f}_2(h_5) + a_3 \int x^2 \hat{f}_{12}(x, h_5) dx + a_4 h_5^2 \hat{f}_2(h_5) + a_5 h_5 \int x \hat{f}_{12}(x, h_5) dx = \int y \hat{f}_{02}(y, h_5) dy \right. \quad (66)$$

$$\left\{ a_0 + a_1 \bar{X}_{n_{12}} + a_2 \bar{H}_{n_{12}} + \frac{a_3}{n_{12}} \sum_{i=1}^{n_{12}} X_i^2 + \frac{a_4}{n_{12}} \sum_{i=1}^{n_{12}} H_i^2 + \frac{a_5}{n_{12}} \sum_{i=1}^{n_{12}} X_i H_i = \bar{Y}_{n_{02}} \right. \quad (67)$$

In matrix form

$$\begin{pmatrix} \hat{f}_2(h_1) \int x \hat{f}_{12}(x, h_1) dx & h_1 \hat{f}_2(h_1) \int x^2 \hat{f}_{12}(x, h_1) dx & h_1^2 \hat{f}_2(h_1) & h_1 \int x \hat{f}_{12}(x, h_1) dx \\ \hat{f}_2(h_2) \int x \hat{f}_{12}(x, h_2) dx & h_2 \hat{f}_2(h_2) \int x^2 \hat{f}_{12}(x, h_2) dx & h_2^2 \hat{f}_2(h_2) & h_2 \int x \hat{f}_{12}(x, h_2) dx \\ \hat{f}_2(h_3) \int x \hat{f}_{12}(x, h_3) dx & h_3 \hat{f}_2(h_3) \int x^2 \hat{f}_{12}(x, h_3) dx & h_3^2 \hat{f}_2(h_3) & h_3 \int x \hat{f}_{12}(x, h_3) dx \\ \hat{f}_2(h_4) \int x \hat{f}_{12}(x, h_4) dx & h_4 \hat{f}_2(h_4) \int x^2 \hat{f}_{12}(x, h_4) dx & h_4^2 \hat{f}_2(h_4) & h_4 \int x \hat{f}_{12}(x, h_4) dx \\ \hat{f}_2(h_5) \int x \hat{f}_{12}(x, h_5) dx & h_5 \hat{f}_2(h_5) \int x^2 \hat{f}_{12}(x, h_5) dx & h_5^2 \hat{f}_2(h_5) & h_5 \int x \hat{f}_{12}(x, h_5) dx \\ 1 & \bar{X}_{n_{12}} & \bar{H}_{n_{12}} & \frac{1}{n_{12}} \sum_{i=1}^{n_{12}} X_i^2 & \frac{1}{n_{12}} \sum_{i=1}^{n_{12}} H_i^2 & \frac{1}{n_{12}} \sum_{i=1}^{n_{12}} X_i H_i \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{pmatrix} = \begin{pmatrix} \int y \hat{f}_{02}(y, h_1) dy \\ \int y \hat{f}_{02}(y, h_2) dy \\ \int y \hat{f}_{02}(y, h_3) dy \\ \int y \hat{f}_{02}(y, h_4) dy \\ \int y \hat{f}_{02}(y, h_5) dy \\ \bar{Y}_{n_{02}} \end{pmatrix} \quad (68-73)$$

2.1.5. Performance Evaluation of the Models with Residual Analysis

Recall Equation (15):

$$\int m_{012}(x, h) \hat{f}_{12}(x, h) dx = \int y \hat{f}_{02}(y, h) dy \quad (15)$$

Taking $h = H_j$, for $j \in \{1, \dots, n_{02}\}$, and dividing by $\hat{f}_2(H_j)$, we get

$$\hat{m}(H_j) = \hat{\mathbb{E}}(m_{012}(X, H) | H = H_j) \quad (74)$$

$$= \frac{\int \hat{m}_{012}(x, H_j) \hat{f}_{12}(x, H_j) dx}{\hat{f}_2(H_j)} \quad (75)$$

$$= \frac{\frac{1}{n_{12} b_n} \sum_{i=1}^{n_{12}} \hat{m}_{012}(X_i, H_j) K\left(\frac{H_j - H_i}{b_n}\right)}{\frac{1}{n_2 b_n} \sum_{i=1}^{n_2} K\left(\frac{H_j - H_i}{b_n}\right)} \quad (76)$$

$$= \frac{n_2 \sum_{i=1}^{n_{12}} \hat{m}_{012}(X_i, H_j) K\left(\frac{H_j - H_i}{b_n}\right)}{n_{12} \sum_{i=1}^{n_2} K\left(\frac{H_j - H_i}{b_n}\right)} \quad (77)$$

as predicted values. Note that $\hat{m}(H_j)$ should be close to

$$\hat{m}(H_j) = \frac{\int y \hat{f}_{02}(y, H_j) dy}{\hat{f}_2(H_j)} \quad (78)$$

$$= \frac{\frac{1}{n_{02} b_n} \sum_{i=1}^{n_{02}} Y_i K\left(\frac{H_j - H_i}{b_n}\right)}{\frac{1}{n_2 b_n} \sum_{i=1}^{n_2} K\left(\frac{H_j - H_i}{b_n}\right)} \quad (79)$$

$$= \frac{n_2 \sum_{i=1}^{n_{02}} Y_i K\left(\frac{H_j - H_i}{b_n}\right)}{n_{02} \sum_{i=1}^{n_2} K\left(\frac{H_j - H_i}{b_n}\right)} \quad (80)$$

because of Equation (15).

The residuals can then be computed as

$$r_j = Y_j - \hat{m}(H_j), \text{ for } j \in \{1, \dots, n_{02}\}. \quad (81)$$

and Mean Squared Error as

$$\mathbf{MSE} = \frac{1}{n_{02}} \sum_{j=1}^{n_{02}} r_j^2 = \frac{1}{n_{02}} \sum_{j=1}^{n_{02}} (Y_j - \hat{m}(H_j))^2 \quad (82)$$

2.2. Demonstration Data Set

2.2.1. Destructive Sampling and External Tree Characterization

More than 40 candidate trees were initially identified through random selection on the LiDAR map of two unmanaged sites within the mixed wood stands of the Petawawa Research Forest (PRF) in Ontario, Canada (Figure 1). Of these, 18 trees were excluded due to decay or major structural defects, including forking, crooks, rotten branches, dead tops, and stem lean greater than 5°. The final sample consisted of 14 white pine aged (113 - 131 years) and 8 red pine trees (aged 70 - 128 years), which were manually felled using two approaches: (1) full-tree felling followed by delimbing on the ground (Figure 2a), and (2) delimbing the standing tree, lowering the branches to the ground with a crane, and subsequently felling the bole (Figure 2b). While the second method provided greater

control and minimized damage, the first method resulted in breakage of approximately 5–10% of the branches.

After felling, all branches were organized to ensure that each limb remained intact and correctly oriented, and the bole and branches were positioned on a level surface for measurement but close to the stump. Stump characteristics, including stump height at the lowest and highest points and stump diameter were recorded. A reference line was then marked along the stem from the cut base to the tip (Figure 2c), and total tree length of the felled part was measured. Stump height was noted as the average height from the ground at the highest and lowest elevation. Total tree height is defined as the total of stump height and total tree length. The height to the base of the live crown (crown base height, CBH) is defined as the position of the lowest living whorl containing at least one live branch and with continuous whorls above it. Stem diameter was recorded at 1-m intervals up to the height of the first living branch.



Figure 1. Local and characteristics of the site where sample trees were felled.

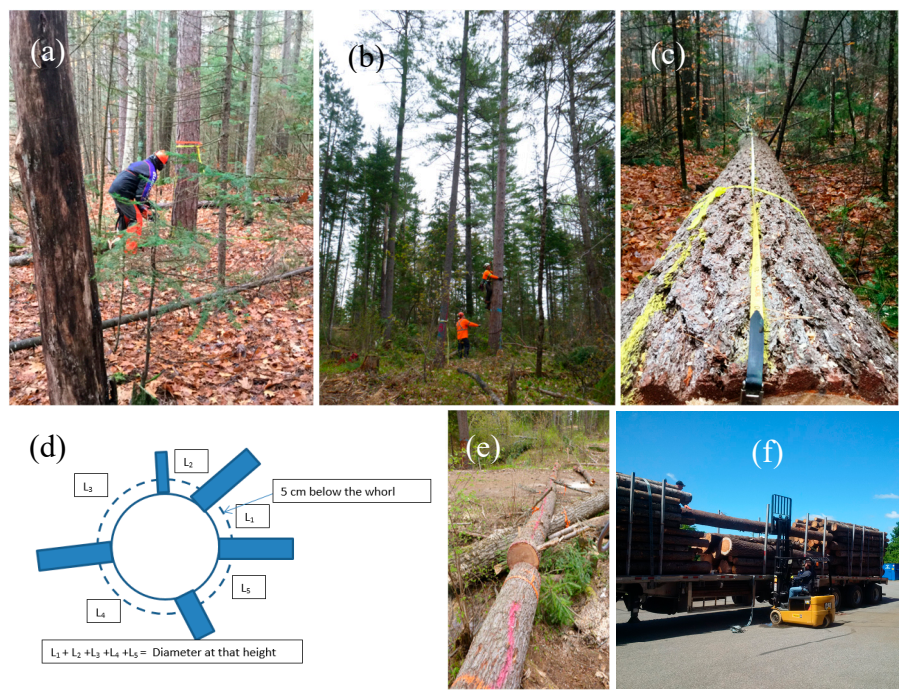


Figure 2. Destructive sampling and external tree measurement. (a) Felling of the whole tree prior to delimbing (b) Arborist carefully delimbing the branches before felling (c) Reference line along the bole for measurement (d) sample guideline - measuring diameter at a whorl (e) 5-m log segments of the bole after completing measurements (f) Logs carefully transported to the lab for CT scanning of internal log attributes.

For each whorl, starting from the last living branch to the tree apex, its position from the base of the stem was recorded. A sample of how the stem diameter at each recorded whorl was measured is shown in Figure 2d. Sum of the length between the branches of the whorl exactly 5 cm below the whorl was calculated as the diameter of the stem at the whorl. For every branch within each whorl, measurements included position relative to the reference line, branch length, branch diameter, branch inclination, and vitality status (dead or live). Finally, the stem was divided into 5-m sections for further bucking and evaluation. A small notch was made along the reference line at each 5-m mark (needed to keep track of the original orientation of each log within each tree for stem reconstruction, after which the stem was cut into logs and each segment was numbered.

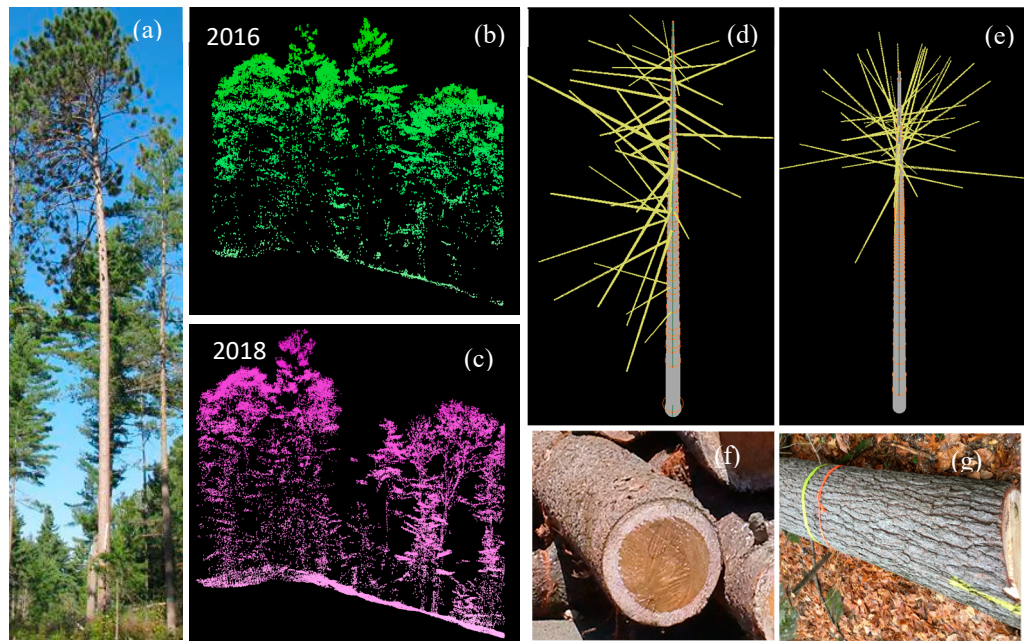


Figure 3. Validation and comparative visualization of the felled tree external tree measurements (a) Field photo of red and white pines in the test site (b) and (c) Bi-temporal lidar point clouds showing the structural state of a felled white pine tree (d) and (e) Scaled digital reconstruction of white and red pine sample felled tree respectively using the developed architectural model based on precise manual field measurements (f) and (g) Segmented white pine log showing the cross-section and side profile respectively.

To facilitate visualization and validation of the detailed measurements collected from the felled trees, a digital tree-reconstruction tool was developed in-house using OpenGL (Figure 3d,e). The tool visualizes the tree segments to the scale. Each tree was validated against the visualization model and bi-temporal lidar data that was flown before and after the harvest. Due either to the 50-cm width constraint during log CT scanning or to branch losses of 5 - 10% resulting from the felling process, four white pine and two red pine trees were excluded from further analysis.

Finally, a total of 58 white pine and 20 red pine 5m logs were sent for log reconstruction and knot characterisation. Because the height of the trees was variable, the number of logs and their lengths varied between trees. Over 97% of the logs had a length in the 4.5–5.5 m range and 2–7 logs were cut per tree (Table 1).

2.2.2. Knot Characterization

Following harvesting at the PRF, the segmented logs were transported to the Institut national de la recherche scientifique (INRS) in Quebec City for internal characterization. The knot measurement protocol was executed in two primary stages: initial X-ray Computed Tomography (CT) scanning, followed by automated feature extraction and measurement using the CT2Opti software suite.

X- Ray CT Scanning and Image Acquisition

Log sections were scanned using a Siemens Somatom Definition 128 AS+ medical CT scanner (Siemens-Healthineers, Erlangen, Germany) available at the INRS. To accommodate the scanner’s 2.1 m length limit, each 5 m log was halved (Figure 4a) and scanned in two passes (Figure 4b). CT images were reconstructed at 1 mm intervals along the longitudinal axis with a 2 mm slice thickness. This 1 mm overlap was intentional, to ensure maximum knot detection across consecutive frames [38]. Scanning parameters included an X-ray voltage of 120 kV, a current of 275 mAs, a pitch of 0.5, and a variable field of view yielding a 0.6 mm x 0.6 mm pixel resolution. Image reconstruction utilized the I70f Safir 3 filter, producing approximately 2,500 images per 2.5 m section.

Table 1. Summary external characteristics of the felled sample trees considered for modeling.

Measured Characteristic	White pine		Red Pine	
	Mean	Range	Mean	Range
Number of subject trees	10		5	
Tree age (yr)	124	111-131	100	70 - 128
Density (kg / m3)	360.19	312.99 - 624.13	380.68	325.77 - 450.24
DBH (cm)	43.2	26.5 -51.8	40	35.0 - 47.7
Total Height (m)	27.54	24.7 - 31.79	25.22	22.07 - 29.92
Stump height (cm)	31.8	10.0 - 70.7	34.3	19.0 - 44.4
Topping length at 9.1 cm diameter (m)	3.17	1.37-4.35	3.2	3.07-3.45
Height at topping (m)	24.36	20.8 - 27.99	21.99	19.0 - 26.47
Height to the last living branch (m)	10.05	3.9 - 16.16	11.62	5.2 - 16.42
Total number of dead branches	111	59 - 228	80	18 - 121
Total number of live branches	70	22 - 199	49	34 - 63
Number of whorls	181	108 - 427	129	67 - 162
Crown diameter (m)	6.11	2.93 - 8.76	5.54	2.86 - 7.6
Branch diameter (cm)	4.45	3.28 - 6.29	3.82	3.24 - 4.27
Branch length (cm)	215.4	45.52 - 336.52	211.8	200.8 - 232.08

Branch inclination (o)	110	107 - 140	120	25 - 160
Number of logs	5	4 - 7	5	4 - 6
Log diameter (cm)	29.4	10.3 - 64	24.63	9.2 - 42.2
Log length (m)	4.28	2.5 - 5.05	4.33	1.37 - 5.04

The resulting Hounsfield units (gray tones) were utilized as a proxy for wood density [39]. While knots typically appear brighter due to higher density relative to clear wood [40], sapwood and wet pockets can exhibit similar density profiles, potentially obscuring contrast. This challenge was addressed using the specialized extraction pipeline detailed below.

Knot Extraction and Measurements with CT2Opti

Internal features were extracted using CT2Opti and Optitek 10, a specialized image-processing suite (please consult [38,41] for details) that combines morphological filtering, thresholding, and edge detection. The extraction followed a refined pipeline that distinguished real knots from false positives by requiring objects to originate near the pith and extend radially toward the log surface. The three-step pipeline can be described as:

- *Candidate Identification*: Where potential knot sections were identified in each 2D image
- *3D Reconstruction*: Where individual sections were grouped into 3D objects by connecting adjacent pixels with similar characteristics across the image stack and
- *Biological Filtering*: Real knots were distinguished from false positives (e.g., wet pockets) by applying a knot model requiring objects to originate near the pith and extend radially toward the log surface.

This process enabled the detailed 3D reconstruction of diverse log profiles, ranging from large-diameter white pine butt logs with significant flare (Figure 4d) and low-tapered intermediary sections (Figure 4e) to complex crown logs with smaller diameters and higher knot densities (Figure 4f–g). High-resolution visualizations further demonstrate the characteristic anatomical expansion of knot size as branches grow from the pith toward the outer bark (Figure 4h).

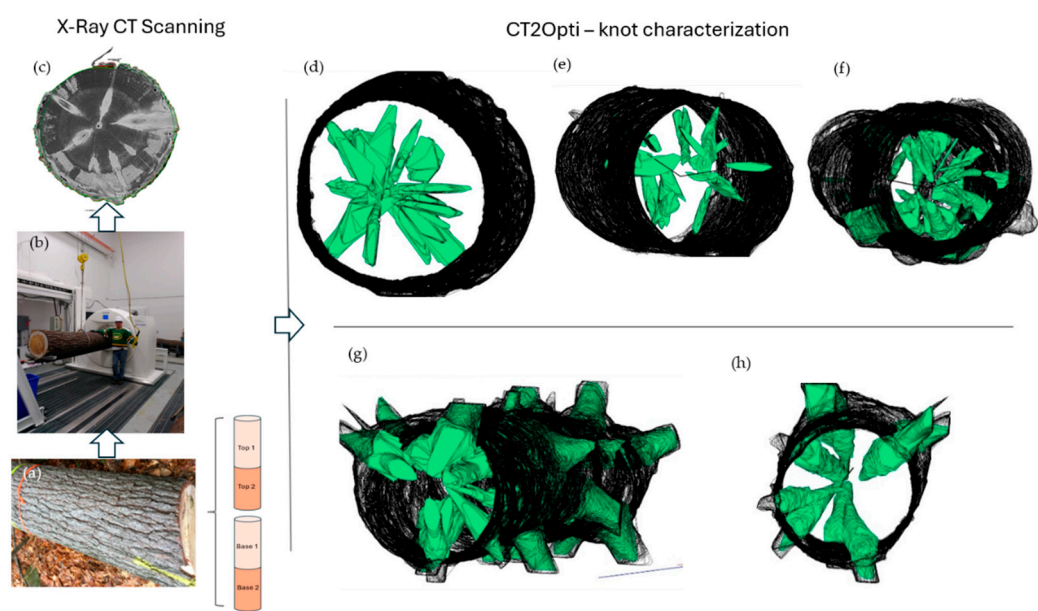


Figure 4. Knot extraction using CT2Opti and visualization using Optitek: (a) Segmenting a 5 m log to meet scanner size constraints; (b) X-ray CT scanning of a white pine log at INRS (Photo credit: INRS); (c) a typical X-ray image with contrasting (denser) knots; (d)–(h) Visualized samples in Optitek 10 (after CT2Opti knot extraction): (d) 2.3 m white pine butt log ($LED = 48.73$ cm, $Volume = 330.39$ dm³); (e) 2.02 m intermediary white pine log ($LED = 42.13$ cm, $Volume = 259.96$ dm³); (f) 1.84 m white pine crown log $LED = 17.87$ cm, $Volume = 37.33$

dm³); (g) 2.34 m red pine crown log (*LED* = 25.98 cm, *Volume* = 67.36 dm³); (h) Close-up of a whorl from log illustrating the radial increase in knot size from pith to surface.

Following manual review and digital reassembly along the pith-line, to reconstruct the complete tree architecture, high resolution metrics were measured for each sample with care. Metrics extracted by CT2Opti includes Total number of knots; Total knot volume (cm³) along with Average knot volume, Minimum knot volume and Maximum knot volume; Average knot area (cm²) along with Minimum knot area and Maximum knot area (found at the surface of the log), and total log volume (cm³). Table 2 summarises the measurements extracted from the felled white and red pine logs. In addition to these metrics, K/T index was also computed. K/T is defined as the proportion of knot volume over the total volume of the reconstructed tree, which provides insights of between tree or species variations [42].

Table 2. Summary internal characteristics of the felled sampled trees considered for modeling.

Species	TreeId	#Knots	Knot - Volume (cm ³)				Knot Area (cm ²)			LogVolume	K/T*100
			Total	Average	Min	Max	Average	Min	Max		
WhitePine	WP1	345	12606.33	45.27	0.19	277.83	15.39	0.22	54.04	1885519.03	0.67
	WP2	329	13011.17	48.35	0.08	332.31	14.87	0.14	55.84	2255631.28	0.58
	WP6	237	7878.58	33.40	0.06	568.29	12.48	0.12	63.91	1092508.77	0.72
	WP8	249	4139.20	20.84	0.01	159.05	7.11	0.05	49.77	498720.45	0.83
	WP9	254	10360.23	37.86	0.07	299.63	10.09	0.14	50.37	1491214.65	0.69
	WP11	331	9027.89	27.13	0.09	348.16	11.49	0.16	62.03	1395931.70	0.65
	WP12	435	12196.71	41.78	0.06	361.86	14.30	0.09	96.15	20280238.00	0.06
	WP5	361	17660.22	48.92	0.05	1564.05	17.27	0.06	160.58	2148683.20	0.82
	WP7	369	26603.19	72.10	0.06	676.09	20.66	0.11	100.48	1876483.12	1.42
	WP14	406	23265.39	57.30	0.00	1050.82	10.61	0.01	117.77	2325245.90	1.00
RedPine	RP1	259	13620.45	19.46	0.01	273.42	12.50	0.01	57.24	986707.53	1.38
	RP2	290	17377.75	67.07	0.05	359.33	18.77	0.06	61.33	1454018.80	1.20
	RP4	184	6504.10	40.99	0.06	163.86	14.51	0.07	41.99	915729.20	0.71
	RP5	417	17393.11	44.03	0.02	428.59	12.72	0.05	62.03	1666761.00	1.04

¹ WP stands for white pine and RP stands for red pine; All volumes are indicated in cm³.

2.3. Application of the Model – Estimating Knot Characteristics of White and Red Pines

2.3.1. Regression Strategy and Data Preparation

As the covariates varied between the lower and upper portions of the tree, the Regression of Misaligned Covariates (RMC) was estimated independently for the Clear Stem (CS) and the Living Crown (LC) across both species, white pine (WP) and red pine (RP). In this framework, trees were grouped by species. To evaluate model performance, a leave-one-tree-out cross-validation approach was adopted: for each species, one tree was sequestered as the evaluation (test) dataset, while the remaining pooled trees served as the training dataset.

To account for varying total heights among the pooled samples, the vertical coordinate was transformed into a relative altitude (*RelAltitude*) variable. This dimensionless index ranges from 0 to 1, where:

- For the CS: 0 represents the base of the bole and 1 represents the highest point of the clear stem.
- For the LC: 0 represents the transition from the clear stem and 1 represents the topping height.

Prior to modeling, every tree was validated by ensuring the total bole length matched the cumulative lengths of the CT scanned logs following the knot characterization process.

Variable Selection: The internal characteristics serving as response variables (Y) included Knot Volume (*KnotVolume*), Surface Knot Diameter (*SurfaceKnotDiameter*), and Knot Size (*KnotSize*)—the latter derived from the surface diameter assuming a circular knot geometry. The external covariates

utilized for the LC included Log Diameter (*LogDiameter*), Branch Length (*BranchLength*), and Branch Diameter (*BranchDiameter*), while the CS model relied primarily on *LogDiameter*.

2.3.2. Model Application to Different Pine Species

As previously described, given a small set of felled trees, a Leave-One-Tree-Out cross-validation approach was employed to train the RMC model. Data partitioning was strictly controlled at the tree level rather than the individual knot level. This was to ensure independence between training and validation sets. In this demonstration, knot volume was chosen as the response as it connects statistical estimation directly to cumulative defect risk along the bole, allowing model performance to be evaluated in terms that are immediately relevant to log sorting and wood quality assessment. In addition, covariates that are directly estimable from LiDAR point clouds are chosen, *viz.*, *BranchLength* and *BranchDiameter* within the live crown, and *LogDiameter* for the clear stem, to facilitate seamless integration of the proposed RMC framework with operational remote-sensing workflows and enable stand-level prediction of knot volume without reliance on destructive measurements.

Discrete knot volume predictions were integrated along the length of the bole to generate cumulative response profiles characterizing the rate of defect accumulation from the stump to the tree top. These profiles provide a spatially explicit basis for identifying optimal bucking points and log-quality thresholds.

2.3.3. Model Formulations and Software

Four distinct variations of the generalized model were implemented to evaluate different structural assumptions regarding the relationship between the response (Y), the covariates (X), and the relative height (H):

- **norm:** A parametric approach assuming (Y, X, H) follows a trivariate normal distribution.
- **linear:** A model assuming the regression function $m(x, h)$ is linear in both x and h .
- **quad:** A model assuming $m(x, h)$ follows a quadratic polynomial relationship in x and h .
- **cubic:** A model assuming $m(x, h)$ follows a cubic polynomial relationship in x and h .

It is to be noted that despite the implementation of a cubic model, this study focuses only on norm, linear and quad models to maintain the intended scope.

Performance of the models was evaluated using the mean squared error derived in the section above. All modeling and statistical analyses were performed using R (v.2026.01.0; [43]).

3. Results and Discussion

3.1. Summary of the Demonstration Dataset

The demonstration dataset comprised mature timber, with most sampled trees exceeding 111 years of age; specifically, WP ranged from 111 to 131 years and RP) from 70 to 128 years, with only a single specimen (RP4) representing a younger age class at 70 years (Table 1). All samples adhered to a 50 cm diameter restriction necessitated by the physical aperture limits of the CT scanner. The WP specimens exhibited an average diameter at breast height (DBH) of 43.2 cm and an average total height of 27.54 m, while RP specimens averaged 40.0 cm in DBH and 25.22 m in height. Anatomical complexity varied significantly between the species: WP trees averaged 181 whorls per tree compared to 129 for RP. Out of a total of 2,104 recorded branches (1,590 WP; 514 RP), approximately 60% of them were identified as dead.

The comprehensive internal wood characterisation spanned a total bole length of 155 m for WP and 87 m for RP, resulting in the characterization of 3,316 and 1,150 knots, respectively (Table 2). Of these knots, over 90% of the surface diameters were below 8 cm, and 90% of knot sizes were below 20 cm² for both species. However, knot volume exhibited greater inter-species variation: 90% of the knots in RP were below 150 cm³, whereas the 90th percentile for WP extended up to 400 cm³ (Table 2).

3.2. Process of RMC to the Demonstration Dataset

Figure 5 illustrates the implementation of the RMC framework in R, predicting internal knot volume from externally measured log diameter across the entire bole of a sample RP specimen. It documents a two-step estimation strategy: estimate the one-covariate conditional means $m_{0,2}(h) = E[Y | H = h]$ and $m_{1,2}(h) = E[X | H = h]$ from training data, then solve the integral equations that recover the two-covariate regression $m_{0,1,2}(x, h) = E[Y | X = x, H = h]$ under a chosen finite-dimensional model (normal, linear, quadratic, cubic, etc.). In this example, the training phase utilized three RP trees, comprising a total of 1,107 *RelAltitude*, 361 *LogDiameter* and 726 *KnotVolume* observations. *LogDiameter* (x) and *KnotVolume* (y), are not aligned in space, yet both exhibit distinct correlations with the shared spatial index, *RelAltitude* (h).

The resulting predicted surfaces demonstrate the framework’s ability to reconcile these disjointed variables into a unified 3D modeling space. Aside from the data range, the distributions of the training and test variables are largely consistent. While *KnotVolume* exhibits significant noise in both datasets, most observations remain below 70 cm³. Within this range, all candidate models, apart from the quad, successfully track the central density of the measured data across the tree’s relative height. Quantitative prediction performance, as indicated by the MSE, shows that the norm) and linear models yield results closely aligned with the measured data and the optimized m-hat estimator.

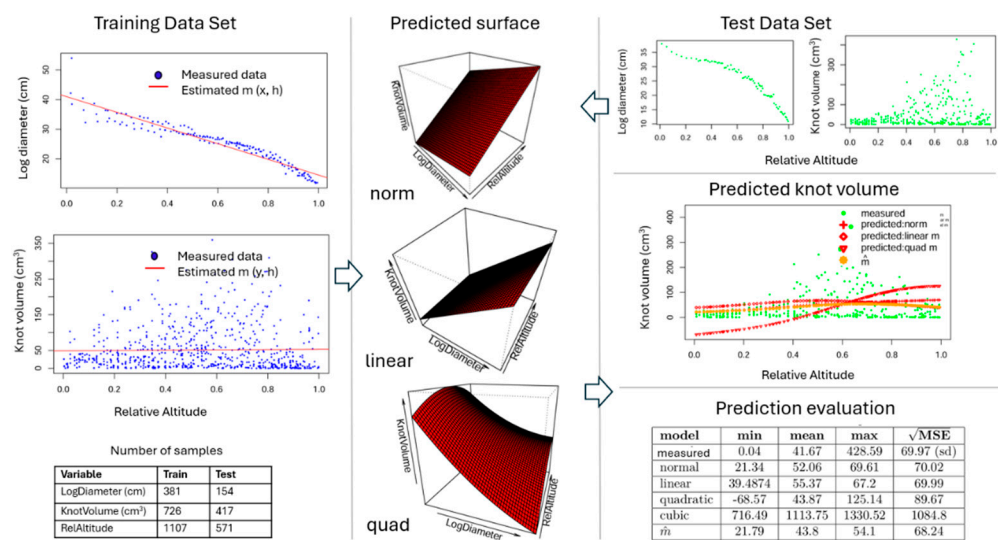


Figure 5. Illustration of the proposed workflow with a sample prediction of knot volume along the clear stem of felled 125-year-old red pine tree. Left: training data for log diameter and knot volume versus relative altitude with fitted conditional means $m(x, h)$ and $m(y, h)$. Center: predicted knot-volume surfaces from three candidate models (normal, linear, quadratic). Right: test data and predicted knot volumes from each model; bottom table summarizes sample sizes and prediction metrics (min, mean, max, $\sqrt{MSE}$). See Methods for model definitions and evaluation metrics.

3.3. Application of RMC to Two Different Species

One sample tree from each of the species was randomly selected (*viz.*, WP11 and RP5) to evaluate application of RMC to the two pine species. Both are matured timber of similar ages, of 125 (WP11) and 128 years (RP5), but exhibit distinct structural differences that test the model’s flexibility (Table 3). While RP5 reached a greater height (29.92 m) compared to WP11 (26.46 m), WP11 has a longer clear stem (10.3 m vs 7.75 m). In the living crown part, RP5 has higher number of branches as well as lower proportion of dead branches. As a result, RP5 has a higher number of knots, and almost twice the amount of knot volume compared to WP11. The higher variance in knot volume (up to 428.60

cm³) and the distinct K/T ratios between the two trees provide a comprehensive dataset to evaluate the predictive accuracy of the RMC framework across different pine species.

Table 3. Comparative summary of test tree characteristics for RMC application.

Variable	WP11	RP5
Age (yr)	125	128
Total tree length (m)	26.46	29.92
Total bole length	23.53	26.1
Canopy Base Height (m)	16.16	14.01
Crown length (m)	12.54	14.1
Clear Stem (m)	10.3	7.75
# Branches (Alive + Dead)	106 (65% D)	155 (35% D)
Range LogDiameter (m)	9.4 - 45	9.5 - 38
Range BranchLength (m)	39 - 441	37 - 461
Range BranchDiameter (cm)	0.5 – 7.8	1 – 7.2
# knots	331	417
Range of knot diameter	0.1 - 8.5	0.1 - 20.2
Range of knot size	0.008 - 56.75	0.008 - 320.47
Total knot volume (mm3)	9027.84	17393.1
Range of knot volume	0.09 – 348.16	0.02 – 428.6
LogVolume	1395932	1666761
K / T	0.006	0.01

Table 4. Summary sample size of the training and test (prediction) dataset used for WP11 and RP5.

Response	Covariates	WP11		RP5	
		Train	Test	Train	Test
<i>KnotVolume - ClearStem</i>	<i>LogDiameter</i>	32	10	26	8
	<i>KnotVolume</i>	173	100	124	88
	<i>RelAltitude</i>	205	110	150	96
<i>KnotVolume - Crown</i>	<i>BranchLength</i>	474	85	180	56
	<i>KnotVolume</i>	2130	215	620	329
	<i>RelAltitude</i>	2604	300	782	385
<i>KnotVolume - Crown</i>	<i>BranchDia</i>	127	56	438	78
	<i>KnotVolume</i>	396	244	1660	141
	<i>RelAltitude</i>	523	300	2098	291

Table 4 details the partitioning used across the two primary physiological segments (clear stem and live crown). In the clear stem, the dataset maintains a high concentration of knot observations relative to the primary covariate (*LogDiameter*) to effectively model the spatial misalignment between disjointed measurements. Reflecting the greater morphological complexity of the upper bole, the living crown had significantly larger training sets, for instance, training data comprised up to 2,130 knot observations for white pine, and 620 for red pine, accounting for the increased density and volumetric variance of the branch base.

The resulting estimated regressions after applying the RMC framework, which characterize these spatial relationships for both species, are presented below and illustrated in Figure 6.

Estimated regressions:

(a) White pine – Clear stem with *LogDiameter*:

$$\hat{m}(x, h) = \hat{m}_{012}(x, h) = \hat{\mathbb{E}}(Y \mid X = x, H = h) = 5.69 * 10^{+01} + 4.47 * 10^{-02}x + 2.59h \text{ (nrmal)}$$
$$\hat{m}_{012}(x, h) = 7.82 * 10^{+01} - 5.93 * 10^{-01}x + 8.41h \text{ (linear)}$$
$$\hat{m}_{012}(x, h) = 1.68 * 10^{+02} + 1.49 * 10^{+01}x - 2.01 * 10^{+02}h - 3.79 * 10^{-01}x^2 + 2.23 * 10^{+02}h^2 - 3.51xh \text{ (quad)}$$

(b) White pine – live crown with *BranchLength*

$$\hat{m}(x, h) = \hat{m}_{012}(x, h) = \hat{\mathbb{E}}(Y | X = x, H = h) = 6.53 * 10^{+01} - 5.28 * 10^{-01}x - 4.05 * 10^{+01}h \text{ (normal)}$$

$$\hat{m}_{012}(x, h) = -3.20 * 10^{+01} + 2.52x + 3.41h \text{ (linear)}$$

$$\hat{m}_{012}(x, h) = 7.14 * 10^{+01} - 2.20x + 1.59 * 10^{+01}h + 4.01 * 10^{-02}x^2 - 3.76h^2 - 1.76xh \text{ (quad)}$$

(c) White pine – live crown with *BranchDiameter*

$$\hat{m}(x, h) = \hat{m}_{012}(x, h) = \hat{\mathbb{E}}(Y | X = x, H = h) = 1.11 * 10^{+02} - 1.01 * 10^{+01}h \text{ (normal)}$$

$$\hat{m}_{012}(x, h) = 8.19 * 10^{+01} - 6.04x - 2.85h \text{ (linear)}$$

$$\hat{m}_{012}(x, h) = 4.79 * 10^{+01} + 6.04 * 10^{+01}x + 4.60 * 10^{+01}h - 1.04 * 10^{+01}x^2 - 2.25 * 10^{+01}h^2 - 2.36 * 10^{+01}xh \text{ (quad)}$$

(d) Red pine – Clear stem:

$$\hat{m}(x, h) = 1.96 * 10^{+01} + 4.83 * 10^{-01}x + 2.16 * 10^{+01}h \text{ (norm)}$$

$$\hat{m}_{012}(x, h) = -1.64 * 10^{+01} + 2.25x - 3.71h \text{ (linear)}$$

$$\hat{m}_{012}(x, h) = 3.15 * 10^{+01} - 2.95x + 8.87h + 9.04 * 10^{-02}x^2 - 3.86 * 10^{+01}h^2 + 1.98xh \text{ (quad)}$$

(e) Red pine – live crown with *BranchLength*

$$\hat{m}(x, h) = \hat{m}_{012}(x, h) = \hat{\mathbb{E}}(Y | X = x, H = h) = -9.34 + 3.49 * 10^{-01}x - 2.74 * 10^{+01}h \text{ (normal)}$$

$$\hat{m}_{012}(x, h) = 7.88 * 10^{+01} - 9.02 * 10^{-02}x - 3.18h \text{ (linear)}$$

$$\hat{m}_{012}(x, h) = 8.65 * 10^{+01} - 9.58 * 10^{-01}x + 1.19 * 10^{+01}h + 2.87 * 10^{-03}x^2 - 1.24h^2 + 2.25 * 10^{-02}xh \text{ (quad)}$$

(f) Red pine – live crown with *BranchDiameter*

$$\hat{m}(x, h) = \hat{m}_{012}(x, h) = \hat{\mathbb{E}}(Y | X = x, H = h) = 2.11 * 10^{+02} - 3.86 * 10^{+01}x - 3.37 * 10^{+01}h \text{ (normal)}$$

$$\hat{m}_{012}(x, h) = 1.08 * 10^{+02} - 2.40 * 10^{+01}x + 4.31 * 10^{+01}h \text{ (linear)}$$

$$\hat{m}_{012}(x, h) = 8.17 * 10^{+01} - 2.03 * 10^{+01}x - 4.21 * 10^{+01}h + 1.74x^2 + 5.98 * 10^{+01}h^2 - 2.52xh \text{ (quad)}$$

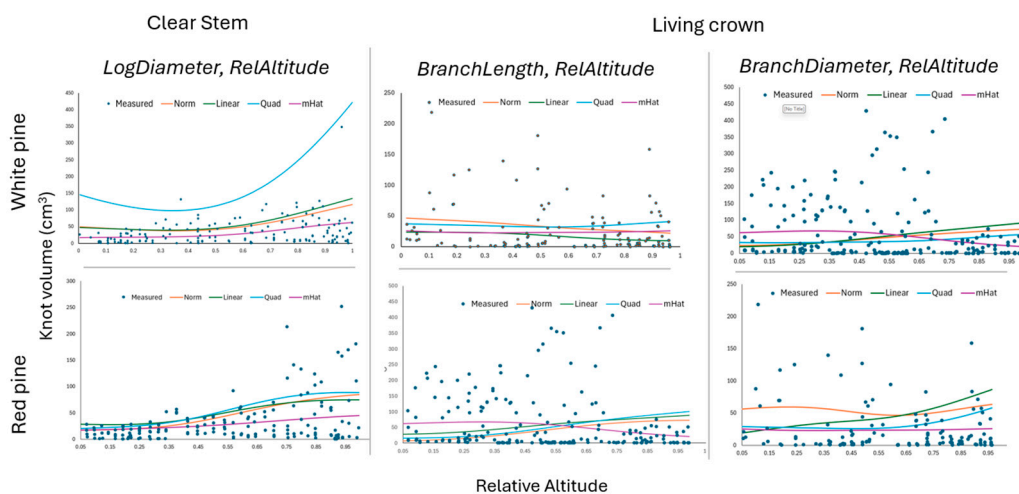


Figure 6. Predicting knot volume of a white and red pine trees – in two different physiological sections: clear stem and living crown. Panels show knot volume (cm³) versus relative altitude for white pine (top row) and red pine (bottom row). Columns show relationships with *LogDiameter* (left), *BranchLength* (center), and *BranchDiameter* (right). Blue points are measured values; colored lines are fitted models: Normal (orange), Linear (green), Quadratic (blue), and mHat (purple).

Table 5. Summary of predictive performance of the models in estimating knot volume (cm³) of white and red pine specimens.

	Clear Stem- LogDiameter, RelAltitude					Living Crown - BranchLength, RelAltitude				Living Crown - BranchDia, RelAltitude			
	model	min	mean	max	√MSE	min	mean	max	√MSE	min	mean	max	√MSE
White pine - WP11	measured	0.19	31.66	348.16	38.71(sd)	0.08	30.8	250.95	41.09(sd)	0.04	49.37	428.59	83.92(sd)
	normal	37.86	62.98	115.54	51.79	16.97	46.71	84.84	41.92	22.73	44.95	72.48	87.95
	linear	36.07	69.14	134.58	57.74	27.78	50.56	75.12	42.84	19.5567	49.45	90.29	90.12
	quadratic	97.78	193.15	421.14	190.36	19.71	54.63	89.02	46.06	32.07	37.89	55.75	86.14
	m̂	17.81	34.2	62.44	37.78	17.46	29.41	44.81	38.37	21.01	53.32	67.17	82.3
Red pine- RP5	measured	0.09	24.22	218.52	38.03(sd)	0.04	49.37	428.59	83.92(sd)	0.09	24.22	218.52	38.03(sd)
	normal	21.68	31.21	46.64	37.95	5.83	38.94	71.92	90.69	46.41	54.58	63.44	48.76
	linear	9.68	16.2	23.84	38.23	27.144	56.05	88.42	89.17	17.0183	50.5	86.16	52.76
	quadratic	31.81	35.22	40.64	39.69	15.08	48.99	99.62	92.56	25.55	34.9	57.7	41.92
	m̂	22.58	23.99	25.96	37.96	21.01	53.32	67.17	82.3	22.58	23.99	25.96	37.96

It should be noted that the m-hat estimator (Nadaraya-Watson estimator) provides a theoretical ceiling (a theoretical maxima and a benchmark) for predictive accuracy by utilizing known response (y) values. It represents the lowest possible error achievable for a given data structure. In practice, implementation for unknown timber relies on the proposed parametric models that utilize only the covariates (here they are externally measurable variables from LiDAR like log diameter, branch length and height).

In this application, m-hat performs well across all variable combinations, species and physiological sections justifying a good benchmark (Figure 6 and Table 5). It yields low or closest MSE to measured deviation (sd), for example 37.98 vs 38.71 and 37.95 vs 38.05 in the clear stem of WP11 and RP5 respectively, with conservative spread avoiding extreme overpredictions at high values. Similarly in the living crown, the model performance is comparable for both sets of covariates, *BranchLength* (38.37 vs 41.09 and 82.3 vs 83.92 for WP11 and RP5 respectively) and *BranchDiameter* (82.3 vs 83.92 and 37.96 vs 38.03 for WP11 and RP5 respectively), indicating its robustness in complex structural heterogeneity well. This also emphasizes its role as a strong operational benchmark for LiDAR-based standing-tree wood-quality estimation.

Table 6. Comparative performance metrics of various models relative to m-hat in terms of MSE and mean bias.

	Model	Clear Stem		Living Crown			
		<i>LogDiameter, h</i>		<i>BranchLength, h</i>		<i>BranchDiameter, h</i>	
		Ratio	MeanBias	Ratio	MeanBias	Ratio	MeanBias
White pine - WP11	normal	1.37	28.78	1.09	17.30	1.07	-8.37
	linear	1.53	34.94	1.12	21.15	1.10	-3.87
	quad	5.04	158.95	1.20	25.22	1.05	-15.43
Red pine - RP5	normal	1.00	7.22	1.10	-14.38	1.28	30.59
	linear	1.01	-7.79	1.08	2.73	1.39	26.51
	quad	1.05	11.23	1.12	-4.33	1.10	10.91

Within the clear stem section, the predictive performance of the proposed models varied between the species (Tables 5 and 6). Both the linear and normal models closely approximated the trajectory of the m-hat benchmark across the bole, with linear model having a slightly higher bias at higher altitudes in WP11 compared to RP5 (Figure 6). Comparing their predictive performance, both models have lower bias and consistently near-parity ratio against that of m-hat, establishing their reliability for predicting knot volume from external *LogDiameter* measurements. Conversely, the quadratic (Quad) model exhibited a consistent upward bias at higher *RelAltitude* and five times higher error compared to m-hat. This sensitivity to extreme values was particularly pronounced in the morphologically complex white pine (WP11) relative to the more uniform red pine (RP5).

Predictions in the crown section show higher overall error across all models due to increased data variance, yet the hierarchy of performance remains consistent (Figure 6, Tables 5 and 6). For WP11, *BranchLength* as well as *BranchDiameter* all models have near-parity with the m-hat, especially normal performed within 10% of m-hat, but showed higher mean bias in estimating *KnotVolume* when using *BranchLength*. For RP5, the quadratic model performed surprisingly well in the crown using *BranchDiameter* (41.92 vs. 37.96 for m-hat and low mean bias), suggesting that for this specific species and section, a slightly curved relationship better captures the volume of larger branch bases. However, linear model performs well with *BranchLength* as the predictor closest to m-hat with a comparatively low mean bias (Tables 5 and 6).

3.4. Management Implications – Turning Data into Decisions for Bucking and Log Sort

Profiles of the cumulative estimated knot volume along the bole height for the proposed and benchmark models (WP11 and RP5) are illustrated in Figure 7. Vertical markers indicate the heights at which 25%, 50%, and 75% of the total knot volume are exceeded. These profiles provide a high-resolution map of internal wood quality for individual standing trees, identifying transition points and accumulation rates that allow foresters to move beyond external visual assessments toward more precise, volume-based grading.

For both species, the 25% threshold identifies the portion of the bole most likely to yield high-quality butt logs with minimal knot impact. When this threshold lies above the clear stem height, the first log can be confidently assigned to sawlog or veneer classes. The 50% threshold marks a transition zone where knot influence becomes substantial; logs cut above this height are increasingly likely to require downgrading to lower-grade sawlogs and lumber. Finally, the 75% threshold suggests that the upper bole sections contribute little additional high-quality material and are better directed to pulp or biomass streams.

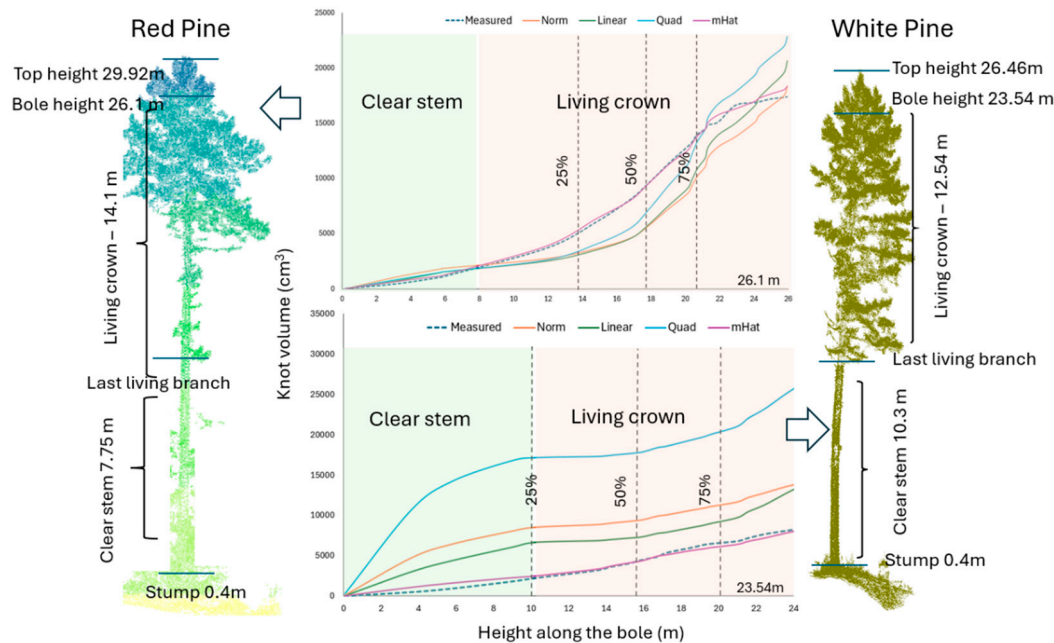


Figure 7. Cumulative knot volume along the bole for red pine and white pine. Cumulative knot volume is plotted as a function of height from the stump, expressed as a proportion of total knot volume per tree. Vertical markers indicate the heights at which 25%, 50%, and 75% of total knot volume are exceeded. The clear stem–living crown boundary is shown in different colors for each species. Curves represent measured values (and model estimates where applicable). These cumulative thresholds illustrate how rapidly knot volume accumulates with height and provide a direct basis for evaluating log quality and bucking decisions along the stem.

Except for the quadratic model in WP11, all proposed models effectively captured the accumulation trends observed in the benchmark and measured data profiles. Although the clear stem of WP11 (10.3 m) is longer than that of RP5 (7.75 m), WP11 accumulates the first 25% of its knot volume at a significantly faster rate. Beyond this point, accumulation in WP11 becomes gradual, whereas RP5—with its more complex branch structure—adds knot volume rapidly after the 50% threshold. Specifically, while WP11 reaches the 25% mark at 10.7 m (near the clear stem top), RP5 does not reach this same threshold until 13.8 m.

Species-specific differences in these threshold positions provide a framework for differentiated sorting strategies. WP11's later accumulation of most of its knot volume offers a greater proportion of high-grade lower logs. Conversely, the earlier accumulation trends in RP5 imply that a more conservative bucking height is required for premium products. Ultimately, these cumulative curves allow managers to predefine bucking heights that balance value recovery and defect risk, improving consistency and efficiency in industrial log sorting.

4. Discussion

This study introduces a generalized statistical framework, Regression of Misaligned Covariates, for reconciling covariates and responses that are measured at different spatial locations - a long-standing challenge in wood-quality modelling and in the integration of multi-source forest data. By combining the Law of Total Expectation with kernel-based estimation of marginal relationships, the RMC approach provides a principled way to recover conditional mean structure even when the underlying datasets are spatially misaligned. This directly addresses the “change of support” problem that arises when LiDAR-derived external attributes and destructively sampled internal knot properties are not observed at the same coordinates.

Applying RMC to eastern white pine and red pine demonstrates the flexibility of the framework across contrasting physiological architectures. Red pine, with its uniform crown structure, provided a stable baseline, while eastern white pine with its irregular branching and susceptibility to structural disturbances, served as a rigorous test of robustness. Across both species, RMC produced smooth, biologically interpretable patterns of knot volume along the bole, capturing species-specific differences in the rate and height at which knot volume accumulates.

The study also demonstrates simplicity of its application and operational practicality. The results highlight that the method does not require species-specific parameterization or complex covariate sets, history or stand conditions. In fact, the model performed well using only height and a single LiDAR-derivable covariate. Notably, the framework produced stable estimates even when trained on a very small dataset, for instance only four red pine trees were used to train and predict RP5 knot characteristics, suggesting that RMC can extract meaningful structure from limited data, though performance will naturally improve as larger datasets become available.

The cumulative knot-volume profiles derived from RMC also illustrate the direct management value of the framework. Expressing knot accumulation as a function of height provides a clear link to operational decisions such as bucking, log sorting, and product allocation. Thresholds marking where 25%, 50%, and 75% of total knot volume occur allow managers to identify the heights at which log quality begins to decline, assess the likely grade of the first and second logs, and anticipate the proportion of the stem suitable for sawlog, dimension lumber and/or studwood, or pulp. Because these thresholds are expressed as cumulative proportions, they can be compared across species, stands, and management regimes, making the approach broadly applicable in both industrial and silvicultural contexts.

Beyond wood-quality modelling, RMC represents a general framework for any application involving misaligned covariates and responses. Many ecological, environmental, and remote-sensing datasets share this structure—for example, soil properties measured at sparse points paired with canopy metrics from LiDAR, or wildlife observations paired with environmental covariates collected at different scales. The RMC formulation is therefore not limited to forestry; it provides a transferable solution for a broad class of misalignment problems where traditional regression approaches fail.

There are, however, several avenues for further work. First, the choice of kernel bandwidth (b_n) plays a central role in shaping the smoothness and accuracy of the marginal estimates that feed into the RMC solution. A careful, data-driven selection of b_n —potentially through cross-validation or adaptive bandwidth strategies—will be important for optimizing performance in different forest conditions. Second, the present study demonstrates the framework using two covariates (height and one structural attribute) to illustrate the core idea. Extending RMC to three or more covariates ($H, X_1, \dots, X_p$) is a natural next step and would allow the integration of richer LiDAR-derived metrics such as branch angle, crown width, or local curvature of the stem. Finally, although the method performed well on small datasets, broader validation across larger samples, additional species, and diverse stand structures will help refine its operational reliability.

Overall, RMC provides a portable and adaptable modelling framework that can be applied to any species or region without requiring species-specific calibration. Because the covariates used here are directly measurable from terrestrial or airborne LiDAR, the approach enables non-destructive prediction of internal wood quality in standing trees. This positions RMC as a practical tool for pre-harvest planning, value-based harvesting, and long-term monitoring of stand development - supporting more precise and sustainable forest management.

5. Conclusions

This study presents RMC as a general and scalable solution to the persistent problem of misaligned covariates and responses in forestry and beyond. By validating the framework across two physiologically contrasting conifer species, we demonstrate that reliable, biologically interpretable estimates of knot volume can be recovered even from small datasets and using only minimal LiDAR-derived covariates. The resulting cumulative profiles translate directly into operational guidance for bucking, sorting, and value-based harvesting, while also enabling non-destructive assessment of standing-tree wood quality. Because RMC is not tied to species-specific parameters and can incorporate additional covariates as needed, it is readily portable across regions, forest types, and remote-sensing platforms. As multi-source data become increasingly central to forest monitoring, RMC offers a flexible pathway for integrating external structural measurements with internal wood-quality attributes, supporting both pre-harvest decision-making and long-term sustainable management.

Author Contributions: Conceptualization, UV; methodology, UV, ID, AS, MD.; software, MD.; validation, MD, UV.; formal analysis, UV.; investigation, UV, MD, AS.; resources, UV, ID, AS.; data curation, UV, ID, MD.; writing—original draft preparation, UV, MD, ID, AS.; writing—review and editing, UV; visualization, UV.; supervision, UV, ID, AS.; project administration, ID, UV.; funding acquisition, ID, UV. All authors have read and agreed to the published version of the manuscript.

Funding: The research was funded by multiple sources. Udayalakshmi Vepakomma received financial support from Natural Resources Canada under the Transformative Technologies contribution agreement with FPInnovations and the industry members of FPInnovations, and Isabelle Duchesne obtained funding from the Forest Innovation Program of Natural Resources Canada (via the Canadian Wood Fibre Centre).

Acknowledgments: The authors would like to thank the Petawawa Research Forest (Canadian Forest Service, Natural Resources Canada) for providing access to the study area and supporting our field trials. We are grateful to Jacques Lirette and Josianne Guay (FPInnovations) for their essential support and assistance with field assessments, tree harvesting, and meticulous data processing. Special thanks go to Steve Vallarand (FPInnovations) for his technical expertise in modifying CT2Opti to meet the specific requirements of this project, to Airu Ji for assistance with knot characterization, to Luc Bedard (FPInnovations) for his valuable support on Optitek, and to Pierre Francus and Louis-Frédéric Daigle (INRS) for providing CT scanning expertise. We also thank Denis Cormier (FPInnovations) for his thorough review of the manuscript and for providing insightful suggestions throughout the project. During the preparation of this work, the authors used Gemini and Copilot to improve clarity and brevity of the text, especially in the methods section. After using this tool, the authors

reviewed and edited the content thoroughly as needed and take full responsibility for the content of the publication.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

RMC	Regression of Misaligned Covariates
RP	Red pine
WP	White Pine
LiDAR	Light Detection and Ranging
NDE	Non-Destructive Evaluation
ALS	Airborne LiDAR Scanning
TLS	Terrestrial LiDAR Scanning
ULS	UAV borne LiDAR Scanning
UAV	Unmanned Aerial Vehicle
CT	X-Ray? Computed Tomography
PRF	Petawawa Research Forest
INRS	Institut national de la recherche scientifique
CS	Clear Stem
LC	Living Crown
DBH	Diameter at Breast Height
CBH	Canopy Base Height
MSE	Mean-Squared Error

References

1. Wang, X.; Ross, R.J. 2022. In-forest wood quality assessments: NDT technologies. USDA Forest Service.
2. Nocetti, M.; Brunetti, M. 2024. Advancements in standing-tree visual wood-quality assessment: A review. *Forests*, 15, 943.
3. Rudnicki, M.; Wang, X.; Ross, R.J.; Allison, R.B.; Perzynski, K. 2017. *Measuring wood quality in standing trees: A review*. USDA Forest Service, FPL-GTR-248.
4. Schimleck, L.; Dahlen, J.; Apiolaza, L.A.; et al. 2019. Non-destructive evaluation techniques for wood property variation. *Forests*, 10, 728.
5. Schimleck, L.; Dahlen, J.; Mora, C.; et al. 2025. Maps of within-tree wood property variation in North American conifers. *Canadian Journal of Forest Research*, 55, 1–14.
6. Clark, A. III; McAlister, R.H. 1998. Visual tree grading systems for estimating lumber yields in southern pine. *Forest Products Journal*, 48(10), 59–67.
7. Trincado, G.; Burkhart, H. 2008. A model of knot shape and volume in loblolly pine. *Wood and Fiber Science*, 40(4), 634–646.
8. Burkhart, H.E.; Tomé, M. 2012. *Modeling Forest Trees and Stands*. Springer, Heidelberg.
9. Barrette, J.; Achim, A.; Auty, D. 2023. Impact of intensive forest management practices on wood quality from conifers: Literature review and reflection on future challenges. *Current Forestry Reports*, 9, 101–130.
10. Drew, D.M.; Downes, G.M.; Seifert, T.; et al. 2022. Progress and applications in wood quality modelling. *Current Forestry Reports*, 8, 317–332.
11. Holmgren, J.; Persson, Å. 2004. Identifying species of individual trees using airborne laser scanning. *Remote Sensing of Environment*, 90, 415–423.
12. Popescu, S.C.; Zhao, K. 2008. A voxel-based LiDAR method for estimating crown base height. *Remote Sensing of Environment*, 112, 767–781.
13. Vepakomma, U.; Cormier, D. 2019. Valuing forest stands using UAV-based LiDAR. *ISPRS Archives*, XLII-2/W13, 643–647.
14. Li, T.; Shen, X.; Zhou, K.; Cao, L. 2024. Estimating tree structure and wood density of Ginkgo biloba using TLS and resistance drilling. *Remote Sensing*, 17(1), 99.

15. Alvites, C.; Marchetti, M.; Lasserre, B.; Santopuoli, G. 2022. LiDAR as a tool for assessing timber assortments: A systematic literature review. *Remote Sensing*, 14(18), 4466.
16. Vepakomma, U.; Cormier, D.; Hansson, L.; Talbot, B. 2023. Remote sensing at local scales for operational forestry. In *Boreal Forests in the Face of Climate Change*. Springer.
17. Zolotarev, F.; Eerola, T.; Lensu, L.; et al. 2020. Modelling internal knot distribution using external log features. *Computers and Electronics in Agriculture*, 179, 105795.
18. Miao, Z.; Zhao, X.; Jiang, Y.; et al. 2026. Integrating taper and knot-length models for Korean pine. *Journal of Forestry Research*, 37, 37.
19. An, Z.; Froese, R.E. 2023. Tree stem volume estimation from terrestrial LiDAR point cloud by unwrapping. *Canadian Journal of Forest Research*, 53(1), 60–70. <https://doi.org/10.1139/cjfr-2022-0153>
20. Vandendaele, B.; Martin-Ducup, O.; Fournier, R.A.; Pelletier, G. 2024. Evaluation of mobile laser scanning scenarios for automated wood volume estimation. *Canadian Journal of Forest Research*, 54(6), 774–792.
21. Stängle, S.M., Brüchert, F., Kretschmer, U., Spiecker, H. and Sauter, U.H., 2014. Clear wood content in standing trees predicted from branch scar measurements with terrestrial LiDAR and verified with X-ray computed tomography. *Canadian Journal of Forest Research*, 44(2), pp.145-153.
22. Pehkonen, M.; Vastaranta, M.; Holopainen, M.; et al. 2025. Identification of branch whorls and sawlogs using TLS and deep learning. *Forestry*, 98(5), 712.
23. Morhart, C.; Schindler, Z.; Frey, J.; et al. 2024. Limitations of estimating branch volume from TLS. *European Journal of Forest Research*, 143, 687–702.
24. Balestra, M.; Marselis, S.; Sankey, T.T.; et al. 2024. LiDAR data fusion to improve forest attribute estimates: A review. *Current Forestry Reports*, 10, 281–297. <https://doi.org/10.1007/s40725-024-00223-7>
25. Tho, Z.Y.; Hui, F.K.C.; Welsh, A.H.; Zou, T. 2024. Cokrig-and-Regress for spatially misaligned environmental data. *arXiv*. <https://doi.org/10.48550/arXiv.2310.05548> (doi.org in Bing)
26. Gotway, C.A. and Young, L.J. 2002. Combining incompatible spatial data. *Journal of the American Statistical Association*, 97(458), 632–648.
27. Gelfand, A.E., Zhu, L. and Carlin, B.P. 2001. On the change of support problem for spatio-temporal data. *Biostatistics*, 2(1), 31–45.
28. Calders, K.; Newnham, G.; Burt, A.; et al. 2015. Nondestructive estimates of above-ground biomass using terrestrial laser scanning. *Methods in Ecology and Evolution*, 6(2), 198–208.
29. Raumonen, P., Kaasalainen, M., Åkerblom, M., Kaasalainen, S., Kaartinen, H., Vastaranta, M., Holopainen, M., Disney, M. and Lewis, P., 2013. Fast automatic precision tree models from terrestrial laser scanner data. *Remote Sensing*, 5(2), pp.491-520.
30. Fowler, D.P.; Morris, R.W. 1977. Genetic diversity in red pine: Evidence for low heterozygosity. *Canadian Journal of Forest Research*, 7(2), 349–357.
31. Gilmore, D.W.; Palik, B.J. 2006. *A Synthesis of Red Pine Silviculture in the Great Lakes Region*. USDA Forest Service, GTR NC-262.
32. Rapraeger, E.F. 1939. Development of branches and knots in western white pine. *Journal of Forestry*, 37(3), 239–245.
33. Mottet, M.J.; Daoust, G.; Zhang, S.Y. 2006. Impact of white pine weevil on Norway spruce lumber properties. *Forestry Chronicle*, 82(6), 834–843.
34. Dreibelbis, S.R.; Germain, R.H.; Smith, W.B. 2025. Longer crowns and fewer knots: Managing for high-quality eastern white pine. *Journal of Forestry*, 123, 321–338.
35. Côté, J.F., Fournier, R.A., Frazer, G.W. and Niemann, K.O., 2012. A fine-scale architectural model of trees to enhance LiDAR-derived measurements of forest canopy structure. *Agricultural and forest meteorology*, 166, pp.72-85.
36. Klockow, P.A., Putman, E.B., Vogel, J.G., Moore, G.W., Edgar, C.B. and Popescu, S.C., 2020. Allometry and structural volume change of standing dead southern pine trees using non-destructive terrestrial LiDAR. *Remote Sensing of Environment*, 241, p.111729.
37. Wand, M.P.; Jones, M.C. 1995. *Kernel Smoothing*. Chapman & Hall.

38. Belley, D.; Duchesne, I.; Vallerand, S.; Barrette, J.; Beaudoin, M. 2019. Computed tomography scanning of internal log attributes prior to sawing increases lumber value in white spruce and jack pine. *Canadian Journal of Forest Research*, 49, 1516–1524.
39. Freyburger, C.; Longuetaud, F.; Mothe, F.; Constant, T.; Leban, J.-M. 2009. Measuring wood density using X-ray CT. *Annals of Forest Science*, 66, 804.
40. Boutelje, J.B. 1966. On the anatomical structure, moisture content, density, shrinkage, and resin content of the wood in and around knots in Swedish pine and spruce. *Svensk Papperstidning*, 69(1), 1–10.
41. Vallerand, S.; Belley, D.; Duchesne, I.; Beaudoin, M. 2011. *Utilisation d'Images CT pour la Modélisation 3D de Billes Réelles*. ForêtValeur.
42. Ji, A.; Cool, J.; Duchesne, I. 2021. Using CT-reconstructed logs to predict knot characteristics and tree value. *Forests*, 12, 720.
43. R Core Team. 2026. R: A Language and Environment for Statistical Computing. R Foundation, Vienna.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.